

Bioestadística para não estatísticos

Base para interpretações científicas



Terceira Edição / Primeira Edição: 2008 / 2004





ELSEVIER
MASSON

© 2007 Elsevier Doyma, S.L.
Es una publicación MASSON
Travessera de Gràcia, 17-21 - 08021 Barcelona (España)



Fotocopiar es un delito (Art. 270 C.P.)

Para que existan libros es necesario el trabajo de un importante colectivo (autores, traductores, dibujantes, correctores, impresores, editores...). El principal beneficiario de ese esfuerzo es el lector que aprovecha su contenido.

Quien fotocopia un libro, en las circunstancias previstas por la ley, delinque y contribuye a la «no» existencia de nuevas ediciones. Además, a corto plazo, encarece el precio de las ya existentes.

Este libro está legalmente protegido por los derechos de propiedad intelectual. Cualquier uso fuera de los límites establecidos por la legislación vigente, sin el consentimiento del editor, es ilegal. Esto se aplica en particular a la reproducción, fotocopia, traducción, grabación o cualquier otro sistema de recuperación de almacenaje de información.

ISBN: 978-84-458-1782-7

Depósito Legal: M. 36.412 - 2007
Composición y compaginación: Fotoletra, S.A.
Impreso en España por Gráficas Muriel, S.A.

A todos los que piensan diferente.

«¿Cómo podía saberlo? ¿Cómo podía comprobarlo? Cualquier estudiante puede hacer experimentos durante su clase de física y comprobar si determinada hipótesis científica es cierta. Pero el hombre, dado que sólo vive una vida, nunca tiene la posibilidad de comprobar una hipótesis mediante un experimento y por eso nunca llega a averiguar si debía haber seguido sus sentimientos o no.»

MILAN KUNDERA

Erik Cobo Valeri

Profesor Titular,
Departamento de Estadística
e Investigación Operativa,
Universitat Politècnica de Catalunya

José Antonio González Alastrué

Profesor Titular,
Departamento de Estadística
e Investigación Operativa,
Universitat Politècnica de Catalunya

Pilar Muñoz Gracia

Profesora Titular,
Departamento de Estadística
e Investigación Operativa,
Universitat Politècnica de Catalunya

Joan Bigorra Llosas

Doctor en Medicina;
Especialista en Farmacología Clínica;
Director de la Fundació Clínic

Cristina Corchero García

Investigadora,
Departamento de Estadística
e Investigación Operativa,
Universitat Politècnica de Catalunya

Francesc Miras Rigol

Investigador,
Departamento de Estadística
e Investigación Operativa,
Universitat Politècnica de Catalunya

Albert Selva O'Callaghan

Doctor en Medicina;
Especialista en Medicina Interna;
Redactor jefe de *Medicina Clínica*

Sebastià Videla Ces

Doctor en Medicina;
Especialista en Farmacología Clínica;
Asesor Fundació Lluita Contra La Sida

CLÍNICO

La inferencia estadística constituye el núcleo del método científico en biomédicina. Por ello, tanto los investigadores como sus destinatarios precisan conceptos estadísticos. Los primeros, para diseñar el estudio de mayor calidad con el menor número de casos y en el menor período de tiempo; y los segundos, para poder asimilar críticamente los resultados aportados por los primeros.

Como revisores metodológicos de *Medicina Clínica*, los profesores Erik Cobo, Pilar Muñoz y José Antonio González conocen qué necesidades tiene la clínica de la estadística. Y han incluido a los doctores Joan Bigorra, Albert Selva y Sebastià Videla, para velar por su comprensión y su adecuación clínica. El resultado es un texto que muestra la utilidad de la estadística para descifrar qué información aportan los datos y cuáles son sus puntos críticos.

La participación de revisores metodológicos de *Medicina Clínica* en una obra que presenta las bases estadísticas del método científico es motivo de orgullo para nuestra revista.

MIQUEL VILARDELL
Catedrático de Universidad
Departamento de Medicina Interna
Universitat Autònoma de Barcelona
Editor de Medicina Clínica

ESTADÍSTICO

Nuestra primera misión, como profesores de Estadística, consiste en formar profesionales de la estadística que potencien la investigación formal en equipos multidisciplinares. A este objetivo se dirige nuestra oferta de los títulos de Graduado, Master y Doctorado en Estadística.

En cambio, este libro se dirige a usuarios de la Estadística y permite, por tanto, cerrar el ciclo, ya que son aquellos a los que van dirigidos los resultados de la investigación quienes deben incorporarla en su profesión.

Los autores presentan de forma intuitiva y práctica los conceptos estadísticos necesarios para que un profesional del campo de la salud valore e interprete los resultados de un estudio científico. Los ejemplos y ejercicios prácticos incluidos deben permitir al lector asentar conocimientos e iniciar su aplicación profesional y científica.

Por todo ello, es una satisfacción que profesores del departamento de Estadística e Investigación Operativa de la UPC hayan realizado esta obra.

MANUEL MARTÍ RECOBER
Catedrático de Universidad
Departamento de Estadística
e Investigación Operativa
Universitat Politècnica de Catalunya

La variabilidad es quizá lo único constante en nuestro mundo. En este libro pretendemos ofrecer unas guías para tenerla en cuenta. ¿Qué significado y utilidad tiene el pronóstico de un paciente cuando la predicción no es exacta? ¿Cómo afirmar que el tabaco mata cuando conocemos tantas excepciones? O, dicho en positivo, ¿cómo cuantificar el efecto de las intervenciones terapéuticas o preventivas en presencia de variabilidad? El profesional sanitario construye su ojo clínico a base de horas de trabajo. La estadística ofrece conceptos que pueden facilitar este aprendizaje. Por poner un ejemplo, ante el residente que descubre en un paciente los síntomas típicos de una enfermedad rara, su instructor le pedirá prudencia mientras le avisa que «es más fácil que sea la presentación anómala de una enfermedad frecuente, que la presentación típica de un fenómeno raro». Detrás de esta afirmación está el teorema de Bayes, concepto estadístico desarrollado por el abate de este nombre mientras perseguía demostrar científicamente la existencia de Dios.

Así pues, en este libro ofrecemos al profesional sanitario una serie de conceptos e ideas en las que basar su conocimiento y sus acciones. Como toda la Ciencia, no se pretende que estos conceptos sean verdades absolutas y definitivas, que no vayan a ser nunca refutadas. Aunque Einstein mostró las limitaciones de las teorías de Newton,

éstas siguen siendo la base con la que construimos nuestros edificios. Y se aguantan. En consecuencia, la Ciencia pretende tan sólo ofrecer ideas que puedan ser útiles para representar, esquematizar y reproducir nuestro entorno. Y cuanto más parsimoniosas o simples sean estas ideas, mejor y más veces serán utilizadas. Qué mejor imagen que la maqueta de nuestra querida Sagrada Familia para ilustrar que estas ideas y conceptos son una mera construcción nuestra para modelar el entorno.



Una *ley* científica pretende *modelar* el mundo, no que le *obedezca*.

Por ello, el concepto principal de la estadística es la variabilidad. Al modelarla, empieza a ser una idea positiva y deja de ser una molestia. Desde Darwin, se sabe que es necesaria para la evolución: es decir, que sin variabilidad, no estaríamos aquí. La estadística aporta teoremas que conectan variabilidad e indepen-

dencia con información. Por primera vez, un modelo científico establece que lo diferente no es necesariamente malo.

Para desarrollar los conceptos estadísticos básicos, el libro incluye tres instrumentos didácticos: ejercicios resueltos, objetivos docentes y direcciones de aplicaciones de internet para el aprendizaje. Como en otros cuerpos de conocimiento, los conceptos estadísticos necesitan asentarse con la ayuda de ejercicios. Una vez realizados y antes de pasar al tema siguiente, conviene que el lector se asegure de haber asimilado una cantidad razonable de los objetivos docentes. Finalmente, el lector interesado puede utilizar las aplicaciones electrónicas para observar cómo se comportan las herramientas estudiadas.

Todos estos conceptos serán útiles al lector si desarrolla su espíritu científico crítico. Preguntábamos al inicio cómo valorar la capacidad pronóstica o terapéutica en presencia de variabilidad. El ejercicio definitivo que muestre la adquisición por el estudioso de estas herramientas consiste en leer informes y artículos de investigación biomédica aplicando estas ideas.

Parte del material de esta obra procede de: los apuntes del curso «Bioestadística para no estadísticos: Principios para interpretar un estudio científico»;

las asignaturas impartidas por los autores en las facultades de Informática y de Matemáticas y Estadística de la UPC; diversos artículos sobre metodología en *Medicina Clínica*, y la monografía *Investigación Clínica y Bioestadística: una visión multidisciplinar con aplicaciones en estudios de VIH/sida*, editada juntamente con los doctores Sebastià Videla y Bonaventura Clotet.

Para acabar, queremos hacer explícitas las contribuciones de los autores. Erik Cobo escribió un primer borrador, que ha sido sucesivamente revisado y completado por Pilar Muñoz y José Antonio González. El doctor Sebastià Videla ha orientado el contenido, ha incluido ejemplos y ha colaborado en la redacción. Los estadísticos Cristina Corchero y Francesc Miras han aportado el formato, elaborado las tablas y repasado los ejercicios. Los doctores Joan Bigorra y Albert Selva han revisado la legibilidad clínica del texto. Los participantes en las tres primeras ediciones del curso «Principios para interpretar un estudio científico» (<http://www-fme.upc.edu/bioestadistica>) han sugerido mejoras. Un especial agradecimiento merece el doctor Juanjo Cabanillas, del Instituto de Formación Novartis, por su apoyo a lo largo de todo el proyecto.

Capítulo 1:

Introducción

Presentación

Este capítulo introduce las bases de la comunicación científica. En la primera parte se sientan los principios para interpretar los valores que toman las variables y así desarrollar la capacidad de razonamiento numérico o cuantitativo: a partir de la definición de medida, se razona sobre el significado de un valor observado; por ejemplo, 78 kg al hablar de peso.

En la segunda parte se presentan las guías de publicación consensuadas internacionalmente.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Delante de una variable, distinguirá entre escala de medida nominal, ordinal y de intervalo.
- Delante de una variable, dirá que está en escala ordinal si puede ordenar sus categorías.
- Delante de una variable, dirá que está en escala de intervalo si dispone de unidad de medida.
- Delante de una variable, distinguirá entre la variable observada y su objetivo de medida.
- Delante de una variable, se preguntará si es válida (en el sentido de medir el objeto de su medida).
- Delante de una variable, se preguntará si es fiable (en el sentido de obtener resultados similares en determinaciones repetidas).
- Identificará Buenas Prácticas Clínicas con el respeto de los derechos de los participantes y con la fiabilidad de los datos.
- Conocerá la existencia de recomendaciones internacionales para el desarrollo y el informe de estudios clínicos.
- Conocerá la recomendación CONSORT para la publicación de ensayos clínicos en revistas biomédicas.
- Conocerá las normas ICH como referente para el desarrollo de intervenciones terapéuticas.

El proceso de cuantificar la observación. Escalas de medida

En el lenguaje literario es excitante percibir, para una misma frase, diferentes perspectivas, matices o significados. De la misma forma, un amigo nos hará reír con un «doble sentido».

Historieta



El amigo Carlos está obsesionado por una calvicie que sólo él percibe. Al llegar al grupo de amigos, Ramón le dice: «¡Cuánto tiempo sin verte el pelo!». Y mientras se apagan las risas, Miguel Ángel le aconseja en tono confidente: «Carlos, esto del pelo deberías quitártelo de la cabeza».

Por el contrario, una característica fundamental del proceso científico es establecer un lenguaje común, con idéntico significado para cualquier observador. Es decir, «llamar a las cosas por su nombre».

Nota técnica



En el tema de escalas de medida va a seguirse la línea de Stevens (1,2).

Escala nominal

Clasificar consiste en agrupar los objetos estudiados en categorías: dos objetos de la misma categoría deben ser equivalentes entre sí y diferentes de los objetos de otra categoría.

Definición



La escala **nominal** clasifica las unidades de acuerdo con ciertas características, de forma que dos unidades de la misma categoría son iguales entre sí y diferentes de las restantes categorías.

Ejemplo 1.1



Una burda y primera clasificación de las personas las dividiría en enfermos y sanos. Dos enfermos compartirán ciertas características comunes que los hacen diferentes de los sanos.

Ejercicio 1.1



Proponga ejemplos de otras clasificaciones posibles.

Nótese que si todas las unidades fueran iguales, todas pertenecerían a la misma categoría y el proceso de clasificación no tendría ninguna utilidad. Es decir, si no existiera variabilidad, no nos plantearíamos ni clasificar ni medir.

Ejemplo 1.2

Los códigos de identificación personal pertenecen a la escala nominal. Incluso aunque estén formados por números, como lo hace el DNI. Nótese que los números del DNI o del teléfono no tienen auténtico significado de número: no significa nada tener un número de teléfono mayor; el DNI sólo «clasifica» y, por tanto, está en escala nominal.

Escala ordinal

En el ejemplo anterior de enfermo y sano, la inmediata ambición del clínico y del científico es poder matizar más y, por ejemplo, establecer grados de intensidad: leve, moderado y grave. Igual que antes, dos unidades de la misma categoría serán iguales entre sí y diferentes de las restantes categorías. Pero ahora, además, se puede establecer una **relación de orden** y decir que grave es más que moderado. Y como moderado es más que leve, entonces grave también es más que leve.

Definición

La escala **ordinal** cumple las propiedades de la escala nominal y, además, permite ordenar las categorías.

Ejemplo 1.3

La reacción a la prueba de la tuberculina puede valorarse en una escala ordinal que va de 0 a 3 cruces (0 / + / ++ / +++).

Ejercicio 1.2

Proponga algún otro ejemplo de variable en escala ordinal.

Escala de intervalo

Una vez establecido un orden, la siguiente ambición del científico será comparar las diferencias entre categorías sucesivas. En el ejemplo de la tuberculina, ¿existe el mismo salto de + a ++, que de ++ a +++? Si todos los saltos tuvieran el mismo significado, se podría hablar de una misma **unidad de medida**, lo que permitiría comparar diferentes intervalos y decir, por ejemplo, que la diferencia entre + y +++ es mayor que la diferencia entre 0 y +. Mientras no se disponga de esta unidad de medida, los «saltos» tendrán diferente significado: no será lo mismo, por ejemplo, pasar de + a ++, que de ++ a +++.

Definición

La escala de **intervalo** cumple las propiedades de la escala ordinal y, además, dispone de unidad de medida.

En la escala de intervalo, las categorías se han convertido en números que disfrutan de una unidad que se aplica por igual a todos ellos: ya tienen significado de número. Como todos contienen las mismas unidades, se pueden restar entre ellos, lo que permite, por ejemplo, comparar la amplitud de varios intervalos, lo que da nombre a la escala.

Ejemplo 1.4



Se puede decir que entre dos cuerpos, uno a 19 °C y otro a 20 °C, hay la misma diferencia que entre uno a 29 °C y otro a 30 °C. O incluso, que el intervalo entre 10 °C y 20 °C es 5 veces mayor que entre 30 °C y 32 °C.

Ejercicio 1.3



Proponga algún otro ejemplo de variable en escala de intervalo.

Escala de razón

Historieta



Dice: «¿Qué tiempo hace?». Y contesta: «Según el termómetro, ni frío, ni calor: 0 °C».

Cuando hay unidad de medida es interesante preguntarse si existe un **cero absoluto**. Es decir, si el valor 0 de la escala tiene significado de «absoluta ausencia de ...». En el ejemplo de la temperatura en grados centígrados, ¿significa 0 °C ausencia absoluta de temperatura? ¡Cuánto tuvo que avanzar la física para poder contestar a esta pregunta! Si se puede afirmar que existe cero absoluto, entonces se dice que se está en escala de razón y es posible también comparar proporciones.

Ejemplo 1.5



Se puede decir que un cuerpo que está a 400° Kelvin tiene el doble de temperatura que un cuerpo a 200 °K.

Para los propósitos clínicos habituales, la distinción entre escala de intervalo y de razón es irrelevante.

Ejercicio 1.4



La variable «¿tiene cefalea?» admite las categorías «nunca», «a veces», «muchas veces» y «siempre». ¿En qué escala de medida se encuentra?

Ejercicio 1.5

¿En qué escala de medida se encuentra la variable «número de cigarrillos que fuma al día»?

Escalas de medida y tipos de variables

La tabla 1-1 resume las propiedades que corresponden a las escalas de medida. Estas propiedades son acumulativas, ya que para tener una propiedad superior es necesario cumplir la anterior. Así, una variable en escala de intervalo, además de poseer unidad constante, tiene sus valores ordenados.

Nota



Más adelante se verá que, en el momento de escoger un tipo de análisis estadístico, se puede optar por renunciar a las propiedades superiores y utilizar uno que corresponda a las inferiores. Por ejemplo, se verá que la edad tiene unidad de medida y permite calcular la media, pero también se pueden hacer categorías (joven, adulto...) y calcular frecuencias.

Escala	Propiedades
Nominal	Equivalencia
Ordinal	Orden
Intervalo	Unidad
Razón	Cero absoluto

Tabla 1-1 Tipos de escala y propiedades acumulativas

Nota técnica



Otra clasificación divide las variables en **cualitativas** y **cuantitativas**. Las primeras hacen referencia a la escala nominal, y las últimas a la de intervalo con una unidad que permite comparar cantidades. La escala ordinal puede tener un papel más ambiguo y corresponderse con ambas, ya que las propiedades de orden podrían aplicarse a categorías o cualidades (como la clase social) o a expresiones cuantitativas (como el número de puntos obtenidos en un *score* como el índice de Apgar).

Nota técnica



Otra división muy habitual consiste en clasificar las variables en **discretas** o **continuas**. Un recuento (el número de hermanos, por ejemplo) es una variable discreta ya que sólo puede tomar un número limitado de valores. En cambio, la altura, que en teoría puede tomar cualquier valor, es una variable continua. Suele considerarse a las variables en escala nominal como discretas, mientras que las otras dos escalas pueden ser tanto discretas como continuas.

Teoría representativa de la medida

Ejemplo 1.6



¿Podemos utilizar la edad como aproximación del grado de madurez de una persona? Veamos, para empezar, en qué escala de medida podemos considerar que se encuentra la variable «edad». Si se habla de la edad que figura en el DNI,

Ejemplo 1.6 (Cont.)

si se pretende utilizar la edad para aproximarse al grado de madurez del sujeto, se podría discutir si representa el mismo incremento en madurez pasar de 2 a 3 años que de 42 a 43, con lo que se estaría poniendo en entredicho la escala de intervalo. Pero, además, se podrían encontrar ejemplos de personas con menos años pero más maduras, con lo que se pondría en entredicho la propiedad de orden. Finalmente, incluso se podría argumentar que no tienen la misma madurez dos individuos de la misma edad, con lo que no se tendría la propiedad de equivalencia y no se podría considerar que la edad es una medida de la madurez. Pero, por otro lado, puede ser útil observar la edad de una persona para considerar qué comportamiento esperamos de ella. Así pues, si no se quieren perder estas posibilidades que ofrece la edad, conviene redefinir el proceso de medida.

Las escalas nominal, ordinal y de intervalo corresponden a una visión «operativa» de la medida: se define una variable por la forma de medirla. Esta visión permitiría definir «el cociente de inteligencia (CI)», como la variable con la que se cuantifica la inteligencia, pero nunca permitiría definir el concepto de inteligencia en sí mismo.

Definición

Medida es el proceso que conecta un concepto con una variable latente y ésta, con variables observables.

Es decir, existe por un lado un atributo latente que no es directamente observable (por ejemplo, la inteligencia) y por otro lado, una o varias variables que pretenden cuantificar dicho atributo (por ejemplo, el CI). El CI será tanto mejor medida de inteligencia cuanto más intensa sea su relación con la misma y menor dependencia tenga con otros factores.

La teoría representativa de la medida es estadística en el sentido de que acepta cierta variabilidad en los resultados. Dos individuos que obtengan exactamente la misma puntuación en una prueba de inteligencia no tienen por qué ser exactamente idénticos, pero cabe esperar que sean más similares que dos individuos con valores alejados.

Dos propiedades son especialmente deseables para poder decir que la variable observable (el CI en el ejemplo) es una buena medida del concepto o variable latente (la inteligencia en sí misma): la **validez** y la **fiabilidad**. Si una variable es válida y fiable, la variabilidad de sus valores depende exclusivamente de la variabilidad de su objeto de medida (el **concepto latente**): por no depender de otros conceptos, no tendrá error sistemático y se dirá que es válida; y por no tener error aleatorio de medida, se dirá que es fiable o repetible.

Definición

Se dice que una variable mide de manera **válida** un concepto representado por una variable latente si está relacionada con esta variable latente y sólo con ella.

Se dice que una variable mide de manera **fiable** un concepto si las variaciones de esta variable sólo están relacionadas con variaciones en el concepto.

La validez hace referencia a la ausencia de error sistemático y la fiabilidad a un error aleatorio razonablemente pequeño. Así, validez implica que se esté valorando el concepto y nada más: que la medida es «pertinente», y que variaciones en la variable comporten variaciones en la medida. Por su parte, la fiabilidad requiere obtener valores próximos en mediciones repetidas en el mismo individuo en condiciones similares.

Ejemplo 1.7

Los logros sanitarios en la cantidad de vida están permitiendo que cada vez sea más relevante la calidad de vida. Desde el punto de vista del paciente, la calidad de vida es su objetivo. Es pertinente y relevante. En una palabra, es una medida válida. Pero para un clínico es una variable algo incómoda, ya que un mismo paciente puede tener oscilaciones en su puntuación de la calidad de vida que no son explicables por variaciones en sus parámetros clínicos. Dicho de otra manera, determinaciones repetidas de la calidad de vida de un paciente, que teóricamente está en las mismas condiciones, no resultan en la misma puntuación. Es decir, no es una medida fiable en su sentido técnico.

Lectura

Si quiere saber más sobre las posibilidades de la conceptualización estadística del proceso de medida puede consultar, por ejemplo, Bollen (3) o Hand (4).

Nota técnica

Estas escalas de medida no se deben tomar como un procedimiento automático para decidir el análisis estadístico, vease Velleman (5).

Ejercicio 1.6

La variable «recuento de linfocitos CD4» suele emplearse en el seguimiento de pacientes con sida. ¿En qué escala de medida se encuentra? Si se decidiera utilizarla como indicador de la evolución, ¿qué opina de su validez y de su fiabilidad?

Ejercicio 1.7

El proceso de aprendizaje universitario, como unos estudios de Medicina, pretende que aquellos que lo finalicen sean capaces de ejercer como profesionales. ¿En relación a la validez y fiabilidad, qué le parece el examen MIR comparado con, por ejemplo, la observación de su trabajo delante de un paciente real?

Ejercicio 1.8

Cierto fármaco tiene un efecto directo sobre la presión arterial. Se ha decidido hacer un ensayo clínico para demostrar su efecto y se discute si medirlo en dicha presión (al cabo de cierto plazo) o en la aparición de complicaciones vasculares (en un plazo muchísimo más largo). El responsable de Sanidad argumenta que el objetivo auténtico es prevenir el riesgo vascular, pero el investigador argumenta que la aparición de accidentes vasculares es muy caprichosa y variable. ¿Quién se preocupa por la validez y quién por la fiabilidad?

Guías internacionales consensuadas

Para mejorar la calidad de la investigación y de los informes que la documentan, la administración, la industria y las revistas científicas han desarrollado varias iniciativas.

Normas de Buena Práctica en Estudios Clínicos

Para garantizar la autenticidad de los datos obtenidos, de su análisis y de su comunicación, se deben seguir una serie de principios y procedimientos conocidos como normas de Buena Práctica en Estudios Clínicos (BPC).

Definición

La **BPC** es la denominación que ha recibido un conjunto internacional de normas éticas y de calidad científica que deben regir el diseño, realización, registro de datos y comunicación de ensayos clínicos (EC).

Lectura

Real Decreto 223/2004, de 6 de febrero (6), por el que se regulan los ensayos clínicos con medicamentos. «Especial mención merece la obligación de aplicar las normas de buena práctica clínica a la planificación, realización, registro y comunicación de todos los ensayos clínicos que se realicen en España, como conjunto de requisitos éticos y científicos de calidad reconocidos a escala internacional y como garantía de la protección de los derechos, la seguridad y el bienestar de los sujetos del ensayo, así como la fiabilidad de sus resultados.»

Nota técnica

Otros términos que también resultan descriptivos del verdadero sentido de BPC son: Buena Práctica de Investigación Clínica, utilizado por la Asociación Británica de la Industria Farmacéutica, o Buena Práctica de EC, utilizado por el Consejo Nórdico de Medicina. Todos ellos son preferibles a normas de Buena Práctica Clínica, que no contextualiza bien su entorno de aplicación.

El cumplimiento de dichas normas garantiza que se han protegido los derechos, bienestar y confidencialidad de los sujetos participantes en el estudio.

En todos los países, los medicamentos deben ser autorizados por las Autoridades Sanitarias antes de su comercialización. Si se observa la definición de BPC, éstas adquieren mayor importancia cuando la información obtenida de un EC se va a utilizar como documentación para el registro de un fármaco y su posterior comercialización.

Comentario

El EC es la herramienta metodológica de investigación clínica requerida por las Autoridades Reguladoras para demostrar la eficacia de un tratamiento y autorizar su uso comercial. Por eso la BPC está referida a los EC. No obstante, cualquier otro estudio de investigación clínica debería también realizarse bajo esta filosofía de trabajo. Prueba de ello es la regulación de los estudios postautorización, que no deben ser necesariamente EC.

Ejercicio 1.9

¿Cuáles son los objetivos principales que pretenden garantizar las BPC?

Directrices ICH (International Conference of Harmonisation)

Para evitar la heterogeneidad de requisitos de las Autoridades Reguladoras de diferentes países, las asociaciones de la industria farmacéutica y autoridades de Estados Unidos, la Unión Europea y Japón fundaron en 1989 la Conferencia Internacional de Armonización de los Requisitos Técnicos para el Registro de Productos Farmacéuticos para Uso Humano.

Ejercicio de Navegación

Entre en la página que se indica a continuación y vea las diferentes directrices expuestas, con especial atención en la ICH E6.

<http://www.ich.org>

Lectura



La Conferencia Internacional de Armonización organiza reuniones con grupos técnicos para preparar directrices (guidelines) consensuadas para la investigación y desarrollo de fármacos (7). Estas actividades se desarrollan en interés de los consumidores y de la salud pública con el fin de desarrollar, de la manera más eficiente posible, medicamentos seguros y eficaces. Con ello se evitan duplicaciones innecesarias de los EC en humanos y se minimiza la experimentación animal, sin comprometer los requisitos normativos necesarios para demostrar eficacia y seguridad de los nuevos medicamentos. Además, también persigue que estas directrices sean aceptadas por las autoridades reguladoras y sean seguidas por la industria farmacéutica como método común, facilitando así la evaluación y el registro de nuevos medicamentos en el denominado territorio ICH.

La tabla 1-2 lista 10 directrices relevantes para realizar y documentar EC.

Código	Tema
E1	Tiempo de exposición y número de caso para valorar la seguridad clínica de fármacos
E2	Gestión de datos sobre seguridad y tolerabilidad
E3	Informe de estudios clínicos
E4	Información sobre relación dosis-respuesta
E5	Factores étnicos en estudios clínicos foráneos
E6	Buenas prácticas en estudios clínicos
E7	Población geriátrica
E8	Consideraciones generales para EC
E9	Aspectos estadísticos en el diseño de EC
E10	Elección del grupo control

Tabla 1-2 Directrices ICH sobre estudios clínicos.

Los documentos que se aceptan en la Conferencia Internacional de Armonización deben ser implementados en la normativa de cada país.

Lectura



Desde 1996 se dispone del documento ICH de BPC (ICH E6 Good Clinical Practice, tabla 1-3) que proporciona un estándar para la realización de EC, válido para la Unión Europea, Japón y Estados Unidos, que pretende facilitar la aceptación mutua de los resultados de los EC por parte de las autoridades sanitarias responsables del registro de medicamentos en dichos países (8).

Capítulo 1.	Glosario de términos referentes a investigación clínica con medicamentos
Capítulo 2.	Principios de Buena Práctica Clínica
Capítulo 3.	Responsabilidades, funciones y procedimientos de actuación de los Comités Éticos
Capítulo 4.	Responsabilidades del investigador
Capítulo 5.	Responsabilidades del promotor
Capítulo 6.	Protocolo
Capítulo 7.	Manual del investigador
Capítulo 8.	Documentos esenciales para la realización de un EC

Tabla 1-3 Índice del contenido de la directriz ICH E6 sobre BPC

Recuerde



Las guías ICH ofrecen un marco de referencia para el desarrollo y la investigación de nuevas intervenciones terapéuticas.

Guías de las revistas biomédicas

Las revistas más prestigiosas del campo biomédico han coordinado sus esfuerzos para ofrecer recomendaciones a sus autores que mejoren la calidad de sus presentaciones y permitan al lector evaluar apropiadamente la relevancia de los resultados presentados.

Se verá a lo largo de este curso que la estadística ofrece diferentes herramientas para obtener pruebas empíricas sobre el rendimiento de alternativas terapéuticas, diagnósticas y pronósticas.

Recuerde



Diferentes objetivos médicos requieren diferentes técnicas estadísticas.

Así, el método por excelencia para demostrar las mejoras producidas por una intervención es el ensayo clínico; pero para establecer el pronóstico de una determinada enfermedad se recurre a la regresión; y para evaluar la calidad de un procedimiento diagnóstico, a estimar las probabilidades de acierto.

Lectura



The Revised CONSORT (9) Statement for Reporting Randomized Trials: Explanation and Elaboration. Hay en la actualidad abundantes datos que demuestran que la calidad de la publicación de los ensayos clínicos controlados y realizados con asignación aleatoria (ECA) está por debajo de los niveles considerados óptimos. En estudios de carácter metodológico publicados recientemente se señala que su deficiente publicación y diseño se asocian a la aparición de sesgos en la estimación de los efectos de los tratamientos evaluados. Este tipo de error sistemático está deteriorando gravemente los EC. Por lo que su eliminación se ha convertido en un objetivo prioritario. El error sistemático en los EC refleja ciencia de bajo nivel que amenaza los principios éticos. Un grupo de científicos y editores elaboró la declaración CONSORT (Consolidated Standards of Reporting Trials) con el objetivo de incrementar la calidad en la redacción y publicación de los EC. En esta declaración se incluye una lista de ítems y diagrama de flujo que pueden utilizar los autores para la publicación de un EC. Muchas de las principales revistas médicas y de los grupos editoriales internacionales más importantes han adoptado la declaración CONSORT, que facilita el estudio e interpretación críticos de los EC, ofreciendo a los autores una guía para mejorar la publicación de los resultados obtenidos en sus estudios.

Ejercicio 1.10



Los 22 puntos de la guía CONSORT (tabla 1-4) requieren combinar habilidades y conocimientos tanto clínicos como estadísticos. Diga 3 que sean eminentemente clínicos y 3 estadísticos.

Ejercicio 1.11

¿Cuál es la diferencia más relevante entre las ICH y la CONSORT?

La tabla 1-4 muestra los puntos de la guía CONSORT que han de incluirse en el informe de un ensayo clínico aleatorizado, mientras que la figura 1-1 presenta el diagrama de flujo del progreso de los individuos a través de las fases del mismo.

Lectura



En diciembre de 2005, Medicina Clínica publicó un número extraordinario con las principales listas de comprobación para autores, revisores y editores de revistas médicas. (10)

	Punto n.º	Descriptor	Informado en pág. n.º
Título y resumen	1	Cómo se asignaron los pacientes a las intervenciones (p. ej. «asignación aleatoria» o «aleatorizado»).	
Introducción			
Antecedentes	2	Antecedentes científicos y razón de ser del estudio	
Métodos			
Participantes	3	Criterios de selección de pacientes y ámbito y lugares en que se recogieron los datos	
Intervenciones	4	Detalles precisos de las intervenciones pretendidas para cada grupo y cómo y cuándo efectivamente se administraron	
Objetivos	5	Objetivos e hipótesis específicos	
Resultados	6	Definición clara de las medidas de los resultados principal y secundarios y, cuando proceda, de cualquier método utilizado para mejorar la calidad de las medidas (p. ej., observaciones múltiples, entrenamiento de evaluadores)	
Tamaño muestral	7	Cómo se determinó el tamaño muestral y, cuando proceda, explicación de cualquier análisis intermedio y reglas de la interrupción del ensayo	
Aleatorización	8	Método utilizado para generar la secuencia de asignación aleatoria, incluyendo detalles sobre cualquier restricción (p. ej., bloques)	
Asignación oculta	9	Método utilizado para implementar la secuencia de asignación aleatoria (p. ej., contenedores de medicación numerados, aleatorización centralizada por teléfono), y clarificar si la secuencia se mantuvo oculta hasta el momento de la asignación	
Implementación	10	Quién generó la secuencia de asignación, quién incluyó a los participantes y quién los asignó a los grupos	
Cegado (enmascaramiento)	11	Si los participantes, quienes administraron las intervenciones o quienes evaluaron los resultados, conocían o no la intervención asignada. Si procede, cómo se evaluó el éxito del enmascaramiento	
Métodos estadísticos	12	Métodos estadísticos empleados para comparar los grupos en el resultado principal; métodos utilizados en análisis adicionales, como análisis de subgrupos o análisis ajustados	
Resultados			
Flujo de participantes	13	Flujo de participantes en cada fase (se recomienda un diagrama). Específicamente, para cada grupo, documentar los números de participantes asignados aleatoriamente, que recibieron el tratamiento pretendido, que completaron el protocolo del estudio y a los que se incluyó en el análisis del resultado principal. Describir las desviaciones del protocolo planificado y los motivos	
Reclutamiento	14	Fechas que limitan los períodos de reclutamiento y de seguimiento	
Datos basales	15	Características demográficas y clínicas basales en cada grupo	

Tabla 1-4 Lista de comprobación para publicar un ensayo clínico aleatorizado (11)

	Punto n.º	Descriptor	Informado en pág. n.º
Números analizados	16	Número de participantes (denominador) de cada grupo incluidos en cada análisis, y si el análisis se realizó «por intención de tratar». Expresar los resultados mediante números absolutos cuando sea factible (p. ej., 10/20 en lugar del 50%)	
Resultados y estimación	17	Para cada resultado principal y secundario, un resumen de resultados por grupo y el efecto estimado y su precisión (p. ej., intervalo de confianza del 95%)	
Análisis complementarios	18	Considerar la multiplicidad, e informar sobre cualquier otro análisis realizado, incluidos análisis de subgrupos y análisis ajustados, indicando los preespecificados y los exploratorios	
Eventos adversos	19	Todos los eventos adversos o efectos colaterales importantes en cada grupo de intervención	
Discusión Interpretación	20	Interpretación de los resultados, teniendo en cuenta las hipótesis del estudio, las fuentes de sesgo o imprecisión potenciales y los peligros asociados a la multiplicidad de análisis y de variables	
Generalización	21	Generalización (validez externa) de los hallazgos del ensayo	
Evidencia actual	22	Interpretación general de los resultados en el contexto de la evidencia	

Tabla 1-4 Lista de comprobación para publicar un ensayo clínico aleatorizado (Cont.)

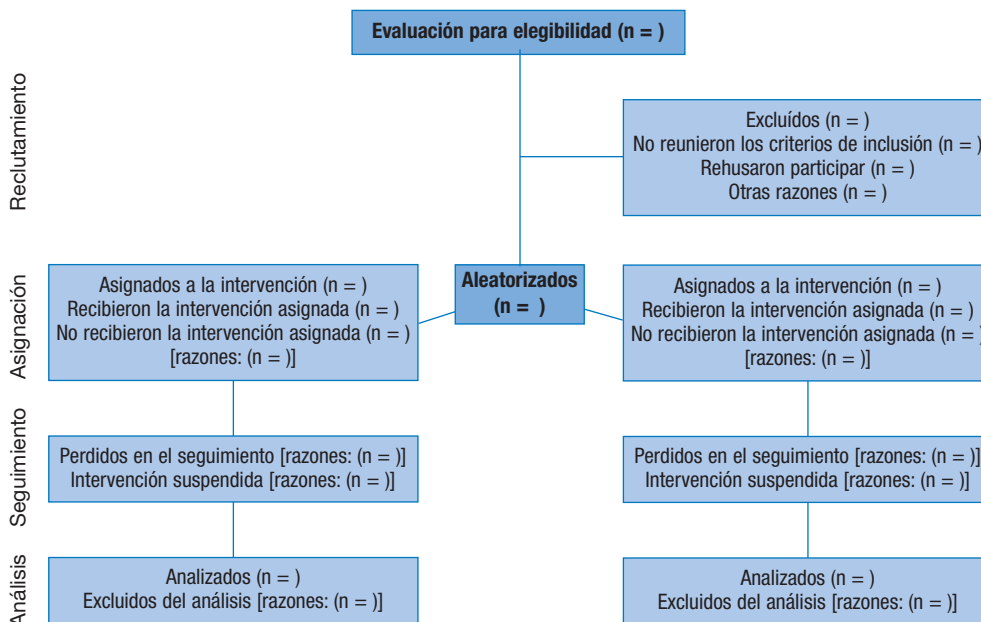


Figura 1-1 CONSORT: diagrama de flujo del progreso de los individuos a través de las fases de un estudio aleatorizado.

Ejercicio 1.12

Algunas de las «n» que figuran en el diagrama de flujo (fig. 1-1) de la CONSORT reflejan pérdidas que podrían comprometer la credibilidad de los resultados y que se estudian con detalle más adelante. Ahora, para empezar diga, para un estudio ideal, ¿cuáles de estas pérdidas le gustaría que fueran cero?

Soluciones a los ejercicios

1.1 La clasificación más habitual de los seres humanos es en masculino o femenino. Otro ejemplo posible de clasificación serían las especialidades médicas, siempre que no exista solapamiento entre las categorías. Para que sea una auténtica clasificación, las unidades deben pertenecer siempre a una categoría y solamente a una. (La teoría de los conjuntos borrosos –*fuzzy sets*– relaja este requisito y permite que una unidad pertenezca a un conjunto con cierta probabilidad.)

1.2 La clase social, en alta, media y baja es otro ejemplo de escala ordinal. Un *score*, tipo test de Apgar, entre 0 y 10, también cumple las propiedades de orden: un recién nacido con valor de 10 está mejor que otro con 9, y así sucesivamente.

1.3 El peso de un paciente está en escala de intervalo: la diferencia entre un sujeto con 65 y otro con 70 kg es la misma que la que existe entre otros dos, de 85 y 90 kg.

1.4 «¿Tiene cefalea?» está claramente en escala ordinal.

1.5 El número de cigarrillos que se fuma al día es más problemático. Para un comercial de una empresa tabacalera, tiene claramente unidad de medida y está en escala de intervalo, ya que su empresa obtiene el mismo beneficio cuando un fumador pasa de 20 a 21 cigarrillos al día, que cuando pasa de 0 a 1 cigarrillo al día. En cambio, para un neumólogo de Sanidad es muy diferente el incremento de riesgo que supone pasar de 0 a 5 cigarrillos/día que de 20 a 25. El profesional sanitario estará más cómodo con representaciones del tipo «no fumador», «fumador moderado» y «fumador severo». Ambos profesionales (el comercial y el neumólogo) están utilizando la misma variable observada (el número de cigarrillos por día) para «acercarse» a dos variables subyacentes (o latentes) muy diferentes: el beneficio comercial de la empresa o el riesgo potencial del paciente.

1.6 Desde el punto de vista de escala de medida, el recuento de CD4 posiblemente estaría en una escala cuantitativa de intervalo, con un mismo significado del incremento al pasar de 150 a 200 que de 550 a 600. Una primera dificultad aparece si el aparato de medida precisa un valor mínimo, pongamos 20, para poder detectar los linfocitos. Si fuera así, se trataría de una variable «censurada», en la que todos los valores inferiores a 20 han sido reconvertidos en un único valor «no detectado». De esta forma, se dispondría de una variable parcialmente de intervalo y parcialmente nominal u ordinal. Otra dificultad es si esta variable se pretende utilizar como indicadora de una variable subyacente, no directamente observable, como podría ser la evolución de ese paciente ante su enfermedad. ¿Aún significa lo mismo un incremento de 150 a 200 que de 550 a 600? ¿O de 375 a 425? Posiblemente no. Todo apunta a que debemos movernos desde la visión operativa de la medida hacia la visión representativa y preguntarnos, no por la escala, sino por la validez y la fiabilidad. Esta última será posiblemente alta en el sentido de que, si se repite la determinación de CD4, se obtienen valores similares. Pero esta fiabilidad será no tan alta si lo que se pretende que sea similar es la evolución, por lo que deberá matizarse también cómo se define la fiabilidad. En cuanto a la validez, se trata de estudiar cómo ayudan los

valores de CD4 a predecir esta evolución, lo que puede estudiarse, por ejemplo, con la ayuda de los términos sensibilidad y especificidad estudiados más adelante.

1.7 El examen MIR es menos válido, ya que mide la capacidad de contestar unas preguntas, no la de actuar profesionalmente. En cambio, es mucho más fiable, en el sentido de que si se repite la evaluación de un mismo individuo (con otras preguntas) se obtendrán puntuaciones mucho más similares, sea quien sea el evaluador, que si se cambia el paciente-caso o el examinador.

1.8 El responsable de Sanidad, que quiere medir el impacto en el auténtico objetivo sanitario, se preocupa por la validez. El investigador desea una variable más fiable, que tenga menor variabilidad aleatoria y permita, de esta forma, una muestra de menor tamaño.

1.9 Los derechos de los participantes y la fiabilidad de los datos.

1.10 Por ejemplo, las más clínicas son: 2, 4, 21 y 22. Y las más estadísticas son: 1, 5, 7, 8, 12, 15 y 18. El resto combinan habilidades de ambas disciplinas.

1.11 Las ICH son recomendaciones de las agencias reguladoras para facilitar la obtención de la autorización de uso comercial de una nueva intervención, usualmente farmacológica. La CONSORT, en cambio, recoge recomendaciones de las revistas científicas para la publicación de EC.

1.12 Todas. Como se verá más adelante, las pérdidas después de la asignación de la intervención (excluidos del análisis pérdidas de seguimiento y no recibieron la intervención) atentan a la validez interna y comparabilidad de los resultados, mientras que las pérdidas previas (rechazaron participar) atentan a la validez externa y generalidad.

Capítulo 2:

Estadística descriptiva

Presentación

En este capítulo se introducen los estadísticos y los gráficos más adecuados a cada escala de medida, así como las medidas de posición relativa de las unidades. Todo ello se aplica a la descripción de los participantes en un estudio.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Definirá media, mediana, moda, desviación típica, cuartiles y desviación intercuartil.
- Delante de los resultados de un estudio, se preguntará por el grado de dispersión de una variable.
- A partir de la desviación típica y de la media construirá un intervalo aproximado que contenga los valores observados.
- Interpretará la desviación típica como el promedio de las diferencias con la media.
- Interpretará un valor tipificado como la distancia a la media expresada en número de desviaciones típicas.
- Identificará como estadísticamente raro (extremo) un valor que se distancie de la media más de 2 (o 3) desviaciones típicas.
- En distribuciones asimétricas recurrirá a la distancia intercuartil en lugar de a la desviación típica.
- Usará los diagramas de barras y gráficos de sectores para representar variables cualitativas y variables discretas.
- Usará histogramas y diagramas de caja (*box-plot*) para representar variables cuantitativas.
- Observará si los ejes de los gráficos están completamente indicados.
- Deducirá, a partir de un *box-plot*, los valores de los cuartiles.

Estadístico más adecuado para cada escala de medida

Veamos a continuación cómo la escala de medida puede ayudar a escoger el estadístico con el que se resumirá el conjunto de los datos. Empezaremos con las medidas de posición central, que informan sobre cómo son las observaciones prototípicas.

Estadísticos de tendencia central

Si las variables están en escala nominal, el parámetro más relevante para caracterizar su distribución es la probabilidad de las categorías más repetidas. En algunas ocasiones, para resumir estas variables, se las representa por su categoría más frecuente, estadístico que se conoce por el nombre de **moda**.

Recuerde



La moda representa a la categoría que más se repite.

Lectura



Martín et al. (12), al describir los pacientes de su estudio, dice: «Los tumores de estadio II fueron los más frecuentes (55,5%)». Nótese que dan la moda pero que, además, concretan a cuántos casos representa.

Si las variables están en la escala ordinal, es posible utilizar aquellas medidas que se basan en la posibilidad de ordenar las observaciones. En general, usan las probabilidades acumuladas, que suman las de las categorías anteriores o menores. Así, si se desea situar alrededor de qué valor se encuentran los valores observados, se puede recurrir a la **mediana** o valor del individuo por debajo del cual se encuentra el 50% de las unidades.

Existen otras muchas medidas basadas en el orden de las observaciones. Los percentiles dividen la muestra en 100 partes, los deciles en 10, los quintiles en 5 y los cuartiles en 4. Conviene notar que, para dividir la muestra en cuatro partes, bastan tres cuartiles.

Ejemplo 2.1



La edad de los pacientes incluidos en un estudio tiene la distribución que se muestra en la figura 2-1. El percentil 20 deja por debajo el 20% de las observaciones, igual que el 2.º decil y el 1.º quintil. Todos ellos toman, por tanto, el mismo valor: 43 años.

Ejercicio 2.1



La mediana, ¿a qué percentil corresponde? ¿Y a que cuartil?

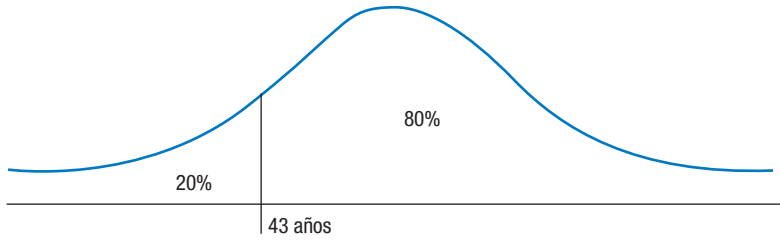


Figura 2-1 El percentil 20 y el decil 2 son 43 años.

Ejercicio 2.2



¿A qué percentil corresponde el cuartil 1? ¿Y el cuartil 2? ¿Y el cuartil 3?

Si las variables están en escala de intervalo, entonces sus valores pueden sumarse, ya que todas ellas se basan en una misma unidad de medida que tiene el mismo significado, independientemente de dónde se haya obtenido: es lo mismo 1 cm apartado por un individuo de 180 cm que 1 cm de otro individuo de 150 cm. Así, para conocer el centro de la distribución puede recurrirse al promedio o **media**: se suman los valores obtenidos en todas las observaciones y se reparten entre el número total de casos.

Lectura



El grupo para la Asistencia Médica Integrada Continua de Cádiz (13), al describir sus resultados, sostiene: «La media de pruebas solicitadas por paciente es [...] menor [...] que las del grupo control».

Ejemplo 2.2



Estudieemos la media con la ayuda de un ejemplo. Se ha preguntado, a los 5 últimos pacientes que han entrado en la consulta, por el número de parejas que han tenido en los últimos 48 meses. Han contestado que 1, 3, 4, 5 y 7 parejas, respectivamente.

Dado que la suma total de parejas es 20, el promedio «que le corresponde a cada uno» es de 4 parejas:

$$\begin{array}{r}
 1 \\
 + 3 \\
 + 4 \\
 + 5 \\
 + 7 \\
 \hline
 \end{array}$$

$$\text{suma} = \sum_{i=1,5} X_i = 20$$

De donde el promedio o media es: $\sum_{i=1,5} X_i / n = 20/5 = 4$

Nota técnica

$\sum_{i=1,5} X_i$ representa la suma de los valores que la variable X toma en los individuos 1 a 5. Simboliza el «sumatorio desde $i = 1$ hasta $i = 5$ de X sub i ».

Con un promedio de 4 parejas por paciente, un investigador descuidado, que se olvidara de la riqueza de la variabilidad y de la existencia de diferencias entre las unidades, podría decir que *cada uno* de estos 5 pacientes ha tenido 4 parejas en los últimos 48 meses. ¡Qué sorpresa para el de 1 pareja!. Y qué forma de decir mentiras. La tabla 2-1 muestra cuánto valen estas mentiras. Nótese que su suma es igual a 0.

Dicen ellos	Se les asigna	Mentira resultante
1	4	+3
3	4	+1
4	4	0
5	4	-1
7	4	-3
Suma	20	0

Tabla 2-1 Mentira o error resultante si se interpreta que cada paciente tiene exactamente el valor de la media

Estadísticos de dispersión

La media representa el centro de la distribución, pero ¿hasta qué punto representa a cada individuo? Sería ingenuo creer que todas las observaciones se sitúan en la media. Además, esta simplicidad implicaría perder toda la información contenida en su diversidad. Por ello, la siguiente medida de interés consiste en estudiar cuál es la distancia que suelen tener las observaciones respecto a ese centro que representa la media.

Definición

La **desviación típica** o desviación estándar (DE) representa el alejamiento prototípico con el centro.

Ejemplo 2.2 (Cont.)



Si a cada uno de ellos se le dice que ha tenido 4 parejas, las mentiras respectivas serán +3, +1, 0, -1 y -3. Ahora bien, como el investigador descuidado es, además, terco, insiste en que su cálculo es acertado ya que la suma de las mentiras da 0 y, por tanto, su mentira promedio es también 0. La media, como centro de gravedad de la variable, tiene

Ejemplo 2.2 (Cont.)



esta propiedad: se compensan los desvíos positivos con los negativos. Para poder valorar la «mentira promedio», se elevan estas distancias al cuadrado antes de sumarlas:

Dicen ellos	Se les asigna	Mentira resultante	Mentira ²
1	4	+3	9
3	4	+1	1
4	4	0	0
5	4	-1	1
7	4	-3	9
Suma	20	0	20

Tabla 2-2 Cuadrado de la mentira si se interpretara que cada paciente tiene el valor medio

Ahora, la suma de las mentiras cuadradas es 20 parejas². Si las parejas que han tenido entre todos se reparten «equitativamente» en los 5 casos, se observa una «mentira² promedio» de 4 parejas², cálculo conocido por el nombre de **varianza**. Para eliminar ese engorroso «cuadrado», se hace la raíz cuadrada, de donde se obtiene que la mentira prototípica es de 2 parejas. Este valor representa, pues, la distancia o desvío (con la media) típico de todas las observaciones. Por esta razón recibe el nombre de desviación típica.

Ejemplo 2.3



Uso de la media y de la desviación típica. Cien niños tratados han tenido fiebre durante una media de 3 días. La desviación típica (o estándar) ha sido de 1 día. Se están describiendo los resultados obtenidos en la muestra: el centro se ha situado en 3 días y los niños se alejaban de este centro, en promedio, 1 día (se entiende que se alejaban por arriba y por abajo).

Para interpretar si la desviación típica es grande o pequeña es útil el siguiente truco. Como promedio de todas las distancias, quiere decir que habrá distancias mayores y distancias menores, que se equilibrarán mutuamente. Así, para «compensar» un valor que coincida exactamente con la media, es decir, que tenga un desvío igual a 0, se necesita otro valor que tenga un desvío que sea el doble de la desviación típica.

Ejemplo 2.4



Si la media de la fiebre era de 3 días y la desviación típica de 1 día, puede interpretarse que, para tener un desvío promedio de 1 día, los casos se distanciarán aproximadamente entre 0 y 2 días de la media. Así, la distancia máxima con

Ejemplo 2.4 (Cont.)

la media será, en este cálculo aproximado, de 2 días. Por tanto, en general, los niños han tenido fiebre entre 1 y 5 días. (Nota: éste es un cálculo aproximado que más adelante se afinará teniendo en cuenta la forma de la distribución.)

Ejemplo 2.5

Soriano et al. (14). «La edad media (DE) de los 11 pacientes con infección de prótesis total de cadera era de 69 (10) años [...]. Se interpreta que el centro de la distribución está en 69 años. Pero esto no significa que todos los pacientes tengan 69 años, sino que están a su alrededor. La distancia o desviación típica que mantienen con el centro vale 10. Esta cifra representa el alejamiento “típico” de 69. Así, algunos casos estarían más cerca y otros más lejos. De manera aproximada, puede decirse que un caso que está justo en el centro (y tiene una distancia de 0) se compensa con un caso que tiene una distancia que dobla la desviación típica (20 años). Así, en esta primera aproximación, cabe imaginar que estos 11 pacientes tienen edades comprendidas entre los 49 y los 89 años.»

Comentario

Un cálculo mental aproximado de la desviación típica, en una variable con distribución simétrica, consiste en dividir entre 2 la distancia entre el valor más alto (o el más bajo) y la media.

Ejercicio 2.3

El personal de cierto hospital camina a una velocidad media de 3 km/h, siendo los extremos de velocidad 2 y 4 km/h aproximadamente. ¿Qué valor cree que puede tener la desviación típica?

Recuerde

La varianza es el promedio de las distancias con la media elevadas al cuadrado. La desviación típica es su raíz cuadrada y valora el promedio de las distancias con la media: representa la distancia típica o esperada de una observación con la media.

Recuerde (Cont.)

La desviación típica muestral se representa por S . En Medicina Clínica se representa por DE (desviación estándar) y en las revistas inglesas por SD (standard deviation).

Ejercicio 2.4

Los 21 pacientes tenían una edad media (DE) de 82 (8) años. Interprete la media y la desviación típica. ¿Entre qué márgenes aproximados cabe esperar que fluctúe la edad de estos pacientes?

La desviación típica es el estadístico por excelencia para valorar las dispersiones, pero requiere que exista escala de intervalo.

Comentario

Se ha visto que existe escala de intervalo cuando hay unidad de medida. Es decir, cuando un aumento de una unidad siempre significa lo mismo. Esta situación es verosímil cuando la variable es simétrica. Si, por ejemplo, se estudia la variable salario, ¿significa lo mismo un aumento mensual de 100 € para quien gana 500 € que para quien gana 5.000 €? Posiblemente tampoco significa lo mismo un aumento de las GOT de 10 a 40 que de 110 a 140.

Comentario

Las variables salario y GOT tienen una marcada asimetría, con una cola muy larga en el extremo superior (fig. 2-2). En esta situación, la desviación típica pierde sentido, ya que no puede interpretarse de la misma forma en ambas colas de la distribución.

Por ejemplo, una persona que gane poco, se distanciará del salario promedio menos que una persona que gane mucho. Por tanto, un mismo estadístico, la desviación típica, no puede representar bien desvíos que son diferentes a ambos lados de la media.

Recuerde

Para poder interpretar, con la misma desviación típica, distancias por encima como por debajo de la media, se requiere que la distribución sea simétrica.

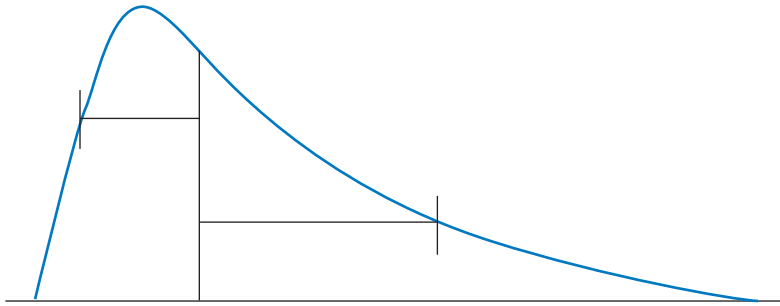


Figura 2-2 Si la distribución es asimétrica, la desviación típica no puede representar simultáneamente los desvíos superiores e inferiores.

Si las variables son muy asimétricas puede recurrirse también a los cuartiles. Asimismo, para valorar la dispersión en la escala ordinal es muy útil la diferencia entre el primer y el tercer cuartil, conocida como **distancia intercuartil**.

La tabla 2-3 muestra los estadísticos de tendencia central y de dispersión que pueden aplicarse en las diferentes escalas de medida, así como las propiedades mínimas que requiere cada estadístico. Por ejemplo, la media sólo puede ser utilizada en escala de intervalo, pero la moda puede ser empleada en cualquier escala.

Escala	Propiedades	Tendencia central	Dispersión
Nominal	Equivalencia	Moda	
Ordinal	Orden	Mediana	Distancia intercuartil
Intervalo	Unidad	Media	Desv. típica = $\sqrt{\text{Varianza}}$

Tabla 2-3 Estadísticos apropiados según la escala de medida

Ejercicio 2.5



a) Suponga que se ha medido la presión arterial sistólica a 5 pacientes: 115, 117, 124, 135 y 142 mmHg. Sin hacer el cálculo, diga qué valor aproximado le parece correcto para la media:

- i) 115 mmHg
- ii) 125 mmHg
- iii) 135 mmHg

b) Suponga ahora que el resultado observado en los 5 pacientes ha sido 100, 125, 130, 135 y 160 mmHg, con una media de 130 mmHg. Sin hacer el cálculo, diga qué valor aproximado le parece correcto para la desviación típica:

- i) 5 mmHg
- ii) 20 mmHg
- iii) 35 mmHg

Ejercicio de Navegación



Entre en la página que se indica al final del párrafo, dentro del apartado Statistics → Stastoscope. Este *applet* calcula los estadísticos de interés. Existen dos opciones: introducir manualmente un conjunto de datos o simular un conjunto de datos con una determinada media y desviación estándar. Introduzca manualmente algunos datos e intente adivinar los valores de la media y de la desviación típica.

<http://www.stat.duke.edu/sites/java.html>

Comentario



En las sociedades industriales predominaba el paradigma de la uniformidad, hasta el punto de que las diferencias con el patrón estándar, con la media, recibían el nombre de desvíos. En la sociedad de la información se abre paso el paradigma biológico de la diversidad, y las diferencias empiezan a ser consideradas un valor positivo y los ordenadores intentan imitar las redes neuronales para acercarse a la inteligencia natural.

Historieta



Demos pues la bienvenida a la diversidad y olvidemos las connotaciones negativas del término desviación. Un término de connotaciones menos negativas, especialmente en el ejemplo de las parejas, podría ser «diversión típica». Seguiremos buscando...

Nota técnica



El cálculo de la varianza presentado ha dividido la suma de las distancias cuadradas por el número de observaciones, pero puede verse que una de las observaciones no estaba aportando ninguna distancia. Si tuviéramos sólo una observación se podría estimar la media, pero no la dispersión. El hecho de estimar la media y la desviación típica en la misma muestra implica gastar una pieza de información, lo que se denomina «perder un grado de libertad». El estadístico más habitual para el cálculo de la varianza divide entre «n-1» (número de casos menos uno) en lugar de entre «n». Los libros de estadística matemática explican sus ventajas.

Definición



$$\text{Varianza muestral } S^2 = \frac{\sum (x_i - \bar{x})^2}{(n - 1)}$$

$$\text{Desviación típica muestral } S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n - 1)}}$$

$$\text{Fórmulas abreviadas } S^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1} = \frac{\sum x_i - \bar{x}^2 \cdot n}{n - 1}$$

donde x_i representa el valor de la observación i -ésima y \bar{x} la media muestral.

Ejercicio 2.6

- a) Suponga ahora que el resultado observado en los 5 pacientes ha sido 100, 110, 120, 130 y 140 mmHg. Calcule la media, la varianza y la desviación típica.
- b) Suponga que se ha medido la presión arterial sistólica al mismo paciente 5 veces en la última visita, habiendo observado 125, 128, 130, 132 y 135 mmHg. Calcule la media, la varianza y la desviación típica.
- CONSEJO: hágalo con una hoja de cálculo.

Comentario

La desviación típica del segundo enunciado es muy inferior, ya que sólo incluye las oscilaciones debidas a las fluctuaciones intracaso, que pueden ser debidas a cambios en el individuo o a errores en el proceso de medida. En el primer caso, aparecen las fluctuaciones a causa de las diferencias entre individuos. Nótese que la variabilidad entre casos es mayor que la variabilidad intracasos. Si esta última fuera mayor tendríamos una baja repetibilidad de los valores, lo que indicaría que la determinación es poco fiable y cuestionaría la utilidad del proceso de medida.

Medidas de posición relativa de los individuos

La existencia de diferencias representa información. El hecho de que seamos diferentes nos permite distinguirnos. Para ello, puede resultar muy útil conocer cuál es la posición de una unidad respecto a otras unidades de su entorno.

Ejemplo 2.6

Vamos a visitar a un amigo al que hemos conocido en un chat de internet. Pongamos que vive en un poblado de África y que, para identificarlo, nos dice que él mide 170 cm. A medida que nos acercamos a su poblado nos entran dudas sobre si seremos capaces de reconocerlo. ¿Cuál debe ser la altura típica de su poblado? Podría ser que fueran muy altos. O todo lo contrario. Saber la media de la altura puede ser una gran ayuda. Pongamos que en su poblado dicha media sea de 150 cm. Por tanto, consideraremos como «altos» a todos los que midan más de 150 cm y «bajos» a los que midan menos. Ahora ya sabemos que tenemos que mirar hacia los altos, ya que nuestro conocido tiene una distancia positiva de 20 cm con la media del poblado. Ahora bien, podría ser que en dicho poblado existiera una gran dispersión y nuestro conocido pasara desapercibido dentro de los altos. O podría ser que todos los habitantes estuvieran muy cerca de la media y nuestro co-

Ejemplo 2.6 (Cont.)



nocido enseguida resaltara. Ahora queremos saber cuánto vale la desviación típica. Si fuera de 20 cm, nuestro conocido sería alto, pero sin destacar entre los altos: sería un «alto típico». En cambio, si la desviación típica fuera de 2 cm, sabemos que la altura de nuestro conocido resaltará mucho entre las de sus vecinos.

Definición



El procedimiento estadístico de tipificar o estandarizar el valor de una variable consiste en restarle la media y dividirlo por la desviación típica.

$$z = \text{desvío tipificado} = \frac{\text{valor observado} - \text{media}}{\text{desviación típica}}$$

Valores de z alrededor de 1 o -1 representan distancias típicas al valor central. Valores cercanos a 0 representan valores muy próximos al centro de la distribución. Y valores de z mayores que 2 (o menores que -2) representan individuos que se están alejando más del doble de lo que se aleja el individuo típico.

Ejemplo 2.7



Si la desviación típica del poblado de nuestro amigo africano es de 20 cm, el desvío tipificado de nuestro amigo vale 1:

$$z_1 = \frac{170 - 150}{20} = 1$$

En cambio, si la desviación típica del poblado fuera 2 cm, el desvío tipificado de nuestro amigo sería 10:

$$z_2 = \frac{170 - 150}{2} = 10$$

Comentario



Regla «a ojo de buen cubero». Hemos visto que si la desviación típica representa la distancia promedio, quiere decir que por cada caso que coincida con la media, que no se aleje nada, habrá un caso que se aleje 2 desviaciones típicas.

Ejemplo 2.7 (Cont.)



El desvío tipificado de nuestro amigo de 1 en el poblado de desviación típica de 20 indica que es un alto típico. En cambio, el desvío de 10 (correspondiente al hipotético poblado con una desviación de 2 cm) indica que nuestro amigo tiene una altura atípica, extraordinariamente alto. Desde un punto de vista estadístico, se trata de un caso «faro» o **extremo**.

Recuerde

Un caso que se aleje más de 2 desviaciones típicas está **fuera de la banda (outlier)** y puede considerarse como extremo en una primera aproximación.

Ejemplo 2.8

Los «errores» en la duplicación del ADN introducen ciertas variaciones que se traducen en individuos de diferentes características. La evolución de las especies se produce porque el entorno selecciona a las unidades mejor adaptadas. La selección natural precisa, por tanto, de la existencia de variabilidad.

Ejercicio 2.7

En cierta población, el colesterol total tiene una media de 200 mg/dl y una desviación típica de 50 mg/dl. Un paciente con colesterol de 175, ¿qué desvío tipificado le corresponde? ¿Cómo interpreta este valor? ¿Y para un paciente con 350 mg/dl?

Ejemplo 2.9

Serían ejemplos de observaciones extremas, un individuo que midiera más de 210 cm (criterio univariante) y otro de 180 cm que pesara 55 kg (criterio bivariante).

Conviene distinguir entre situaciones imposibles (p. ej., 300 cm) o situaciones raras pero posibles (p. ej., 227 cm). Un *outlier* alerta sobre posibles errores de transcripción, o posibles contaminaciones de la muestra, pero no es ninguna prueba definitiva de dato erróneo, por lo que se deben consultar y revisar estas anomalías. No se aconseja eliminar un caso por criterios de «rareza» estadística.

Digamos, para terminar, que la variabilidad no tiene por qué ser necesariamente molesta. Al contrario, puede ser fuente de información y de mejora.

Ejemplo 2.10

Ciertas rutinas de programación generan, al azar, muchas posibles soluciones de un problema. Luego se mejoran, se seleccionan y se vuelve a añadir ruido para reiniciar este pequeño ciclo.

Ejercicio 2.8

La variable RFS tiene una media de 400 y una desviación típica de 150. Defina criterios para detectar datos «sospechosos» en las semanas 0, 6, 12 y 24 del estudio. ¿Qué hará con estos casos?

Ejercicio 2.9

Si consulta al investigador que generó los datos, ¿cuándo le parece más oportuno?

Descripción de los participantes en un estudio

Para que el lector pueda apreciar hasta qué punto los resultados de un estudio pueden ser aplicados en su propio entorno, los autores de artículos científicos deben describir las condiciones en las que han sido recogidos los datos y las características de sus unidades estadísticas, sean pacientes, voluntarios sanos, determinaciones analíticas o muestras de tejido. Las recomendaciones CONSORT para informe de ensayos clínicos, en su ítem 15 dicen:

Lectura

The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration (9). «La información basal se presenta de manera eficiente en una tabla (tabla 2-4). En lo que se refiere a las variables continuas, tal como el peso corporal o la presión arterial, se debería indicar la variabilidad de los datos junto con los valores medios. Las variables continuas pueden ser resumidas en cada grupo mediante los valores correspondientes a la media y a la desviación estándar. En los casos en los que los datos continuos presentan una distribución asimétrica, un planteamiento preferible puede ser el de presentar los datos correspondientes a la mediana y al rango de percentiles (quizá, los percentiles 25 y 75). El error estándar y el intervalo de confianza no son apropiados para describir la variabilidad debido a que ambos son parámetros estadísticos de tipo inferencial más que descriptivo. Las variables constituidas por un número pequeño de categorías ordenadas (como los estadios I a IV de la enfermedad) no deben ser consideradas como variables continuas; en estos casos, es necesario presentar los números y las proporciones correspondientes a cada categoría.»

Ejercicio 2.10

¿Cómo representaría los resultados de éstas variables?

- Glicemia en ayuno en personas sanas
- Transaminasas en enfermos
- Grado de cardiopatía (nivel I a IV) según NYA
- Presión arterial sistólica

Características	Grupo vitaminas (n = 141)	Grupo placebo (n = 142)
Edad media \pm SD, y	28,9 \pm 6,4	29,8 \pm 5,6
Fumadores, n (%)	22 (15,6)	14 (9,9)
Índice de masa corpora media \pm SD, kg/m ²	25,3 \pm 6,0	295,6 \pm 5,6
Tensión arterial media \pm SD, mmHg		
Sistólica	112 \pm 15	110 \pm 12
Diastólica	67 \pm 11	68 \pm 10
Paridad, n (%)		
0	90 (65)	87 (61)
1	39 (28)	42 (30)
2	9 (6)	8 (6)
>2	2 (1)	5 (4)
Enfermedad concomitante, n (%)		
Hipertensión idiopática	10 (7)	7 (5)
Lupus o síndrome antifosfolipídico	4 (3)	1 (1)
Diabetes	2 (1)	3 (2)

Tabla 2-4 Ejemplo de tabla con las características iniciales, clínicas y demográficas, de los grupos en comparación (9)

Comentario



Observe que esta guía dice que el error estándar y los intervalos de confianza (todavía no estudiados) no sirven para describir las condiciones iniciales de los casos.

Ejemplo 2.11



Bobes et al. (15). «Descripción de la muestra: Las características basales de los 168 sujetos incluidos en el estudio (52 pacientes estables, 116 pacientes inestables) se describen en la tabla 2-5. Los pacientes fueron en su mayoría mujeres (el 85 y 82%, respectivamente), con una media (DE) de edad de 47 (12) y 45 (13) años, respectivamente, y nivel de estudios primario. En ambos grupos, la mayoría de pacientes estaba en situación laboral activa (el 35 y el 47%), si bien también fue importante el porcentaje de amas de casa incluidas (el 29 y el 35%). El diagnóstico mayoritario fue el trastorno depresivo mayor de episodio único (el 31 y el 20% en pacientes estables e inestables, respectivamente) o recidivante (el 33 y 42%,

Variables	Pacientes estables (n = 52)	Pacientes inestables (n = 116)
Edad (años), media (DE)	47,5 (12,1)	45,2 (12,8)
Sexo		
Varones	8 (15,4)	21 (18,3)
Mujeres	44 (84,6)	94 (81,7)
Nivel de educación		
Sin estudios	3 (5,9)	8 (7,0)
Estudios primarios	33 (64,7)	72 (62,6)
Estudios secundarios	9 (17,6)	19 (16,5)
Estudios universitarios	6 (11,8)	16 (13,9)
Situación laboral		
Trabaja fuera de casa	18 (34,6)	53 (47,3)
Parado	2 (3,8)	7 (6,3)
Jubilado	2 (3,8)	3 (2,7)
Incapacidad laboral o invalidez permanente	13 (25,0)	9 (8,0)
Ama de casa	15 (28,8)	39 (34,8)
Estudiante	2 (3,8)	1 (0,9)
Diagnóstico (código DSM-IV)		
Trastorno depresivo mayor, episodio único (296,2)	16 (30,8)	23 (19,8)
Trastorno depresivo mayor, recidivante (296,3)	17 (32,7)	49 (42,2)
Trastorno distímico (300,4)	12 (23,1)	20 (17,2)
Trastorno adaptativo con depresión (309,0)	7 (13,5)	24 (20,7)
Tiempo de evolución del trastorno		
0-3 meses	7 (14,3)	35 (30,7)
4 meses-1 año	14 (28,6)	32 (28,1)
> 1 año	28 (57,1)	47 (41,2)
Gravedad del trastorno		
Un poco enfermo	2 (3,8)	7 (6,0)
Levemente enfermo	22 (42,3)	84 (72,4)
Moderadamente enfermo	22 (42,3)	24 (20,7)
Gravemente enfermo	6 (11,5)	1 (0,9)
Entre los casos más graves de la enfermedad		

Tabla 2-5 Características sociodemográficas y clínicas de los pacientes en estudio

Ejemplo 2.11 (Cont.)



respectivamente). En el grupo de pacientes estables, el siguiente diagnóstico en importancia fue el de trastorno distímico (23%), mientras que para el grupo de pacientes inestables fue el de trastorno adaptativo con depresión (21%). En ambos grupos, la mayoría de los pacientes presentó una duración del trastorno superior a un año (el 57 y el 41%, respectivamente), y la gravedad del trastorno, en función de la impresión clínica global (ICG), fue moderada o grave en más de la mitad de los casos (el 54% en el grupo de pacientes estables y el 93% en el grupo de pacientes inestables).>

Comentario



Se trata de una descripción de los casos observados: cómo eran y cómo evolucionan. Queda pendiente por aclarar qué información (y cuánta) aportan estos pacientes sobre cómo cabe esperar que evolucionen otros casos futuros de las mismas características.

Gráficos según escala de medida

Veamos a continuación cómo el tipo de variable y la escala de medida pueden ayudar a escoger el gráfico con el que se resumirá el conjunto de los datos. En los capítulos sucesivos se irán presentando los gráficos más adecuados para cada tipo de análisis.

Lectura



González et al. (16). «Cuando las relaciones entre variables son complejas, los procesos temporales juegan un papel primordial y el componente aleatorio enmascara los procesos en estudio, entonces la representación gráfica deviene una herramienta imprescindible. La biomedicina, en su sentido más amplio, desde las actividades de investigación hasta las de asistencia o de gestión, es un ámbito con estas características y donde los gráficos, bien utilizados, permiten una aproximación nueva y enriquecedora a la información disponible.»

Variables discretas

Los dos gráficos más apropiados para la representación de este tipo de variables son el **gráfico de sectores** y los **diagramas de barras**.

Gráfico de sectores

Consiste en un círculo segmentado en sectores de tamaño proporcional a la frecuencia de cada uno de los valores de la variable. Este gráfico es apropiado cuando los valores de la variable, preferiblemente pocos, son excluyentes entre ellos.

Ejemplo 2.11 (Cont.)



Bobes et al. (15). Si se representa la variable tiempo de evolución del trastorno para cada grupo de pacientes se obtiene la figura 2-3.

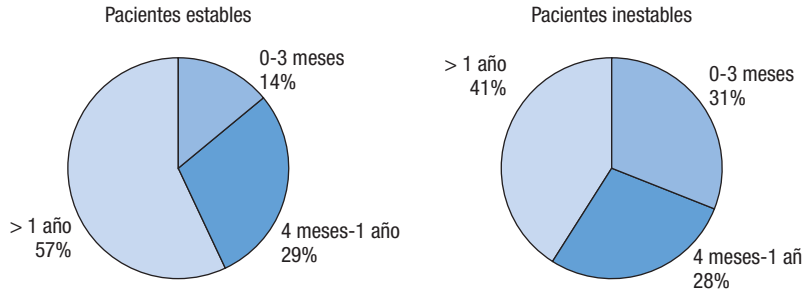


Figura 2-3 Tiempo de evolución de pacientes, estables e inestables.

Diagrama de barras

Este tipo de gráfico se emplea para variables nominales, ordinales y cuantitativas discretas. Consiste en un eje de coordenadas en el que se colocan los distintos valores de la variable en el eje horizontal, con un rectángulo cada uno de ellos de altura proporcional a la frecuencia del valor. En el eje vertical se presenta la escala que va desde 0 hasta, como mínimo, la frecuencia del valor modal.

Ejemplo 2.12



Ramos et al. (17). Hay que destacar las notables diferencias en la proporción poblacional de tratados entre las distintas recomendaciones de tratamiento de la hipercolesterolemia que se observa en la figura 2-4.

Comentario



Para que el gráfico proporcione una correcta impresión visual el eje de ordenadas debe empezar en 0. De no ser así, debe resaltarse para alertar al lector.

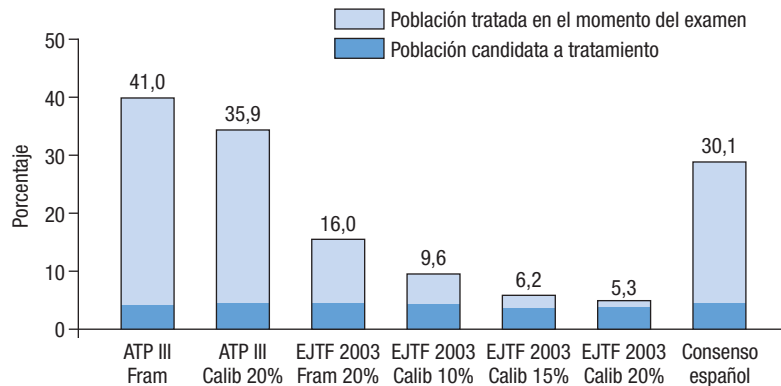


Figura 2-4 Tratamiento de la hipercolesterolemia.

VARIABLES CONTINUAS

En el caso de las variables continuas, existen multitud de gráficos, entre los que presentamos el **histograma** y el *box-plot*.

HISTOGRAMA

El histograma es una extensión del diagrama de barras que dibuja los rectángulos unidos entre sí, indicando de este modo que existe continuidad en los valores de las variables. Un histograma es, por tanto, un gráfico de variable continua dividida en intervalos de los que se eleva un rectángulo con área proporcional a su frecuencia.

COMENTARIO



Obsérvese que lo que es proporcional es el área, no la altura, lo que permite intervalos de diferente amplitud.

Habitualmente, los intervalos son de igual amplitud.

EJEMPLO 2.13



Figura 2-5. Distribución de los valores de tiopurina metiltransferasa (TPMT) en los pacientes con hepatitis autoinmune. Gisbert et al. (18).

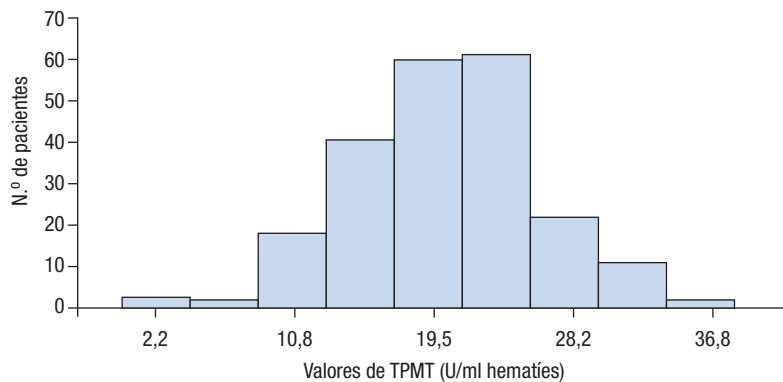


Figura 2-5 Distribución de los valores de tiopurina metiltransferasa (TPMT) en los pacientes con hepatitis autoinmune. Gisbert et al. (18)

EJERCICIO DE NAVEGACIÓN



Entre en la página que se indica al final del párrafo, dentro del apartado Distributions → Histograms. En este *applet* se encuentran representados en un histograma datos de las erupciones de un géiser. Se permite cambiar la amplitud de los intervalos, ¿puede cambiar la interpretación de los resultados según la amplitud del intervalo escogida?

<http://www.stat.duke.edu/sites/java.html>

A partir de un histograma pueden construirse otros tipos de gráficos. Por ejemplo, los gráficos de línea consisten en unir los puntos medios de todos los intervalos contiguos mediante una recta, construyendo así un polígono de frecuencias.

Box-plot o diagrama de caja

En este gráfico (fig. 2-6) se representan los cuartiles. Un diagrama de caja o *box-plot* consta de un rectángulo cuya longitud es igual a la amplitud intercuartil, con una recta en su interior que representa la mediana; siendo los cuartiles 1 y 3, los límites inferior y superior de la caja. Por fuera de ésta, se dibujan dos rectas que, sin que tengan una longitud superior a una vez y media el rango intercuartil, llegan hasta el valor mínimo o máximo de la distribución.

Este gráfico es muy útil, entre otros aspectos, para detectar rápidamente los valores extremos o atípicos (*outliers*), que el *box-plot* define como los individuos que se ubican por encima o por debajo de un rango y medio intercuartil, es decir, fuera de las dos rectas complementarias a la caja.

Ejercicio de Navegación



Entre nuevamente en la página que se indica al final del párrafo, dentro del apartado Statistics → Statoscope. Vuelva a simular conjuntos de datos y observe el *box-plot* correspondiente a cada uno de ellos.

<http://www.stat.duke.edu/sites/java.html>

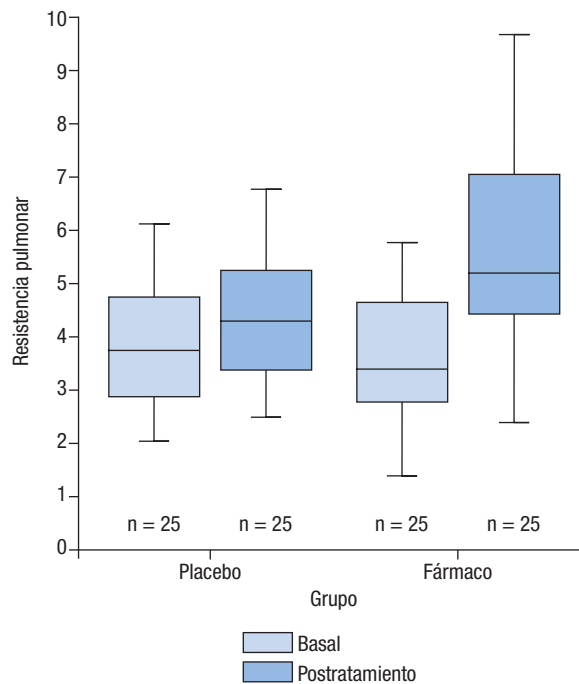


Figura 2-6 Ejemplo de *box-plot* o diagrama de caja. González et al. (16).

Soluciones a los ejercicios

2.1 La mediana se corresponde con el percentil 50 y el cuartil 2.

2.2 El cuartil 1 se corresponde con el percentil 25; el cuartil 2, con el 50 y el cuartil 3, con el 75.

2.3 Si podemos aceptar que alguien que camina muy despacio va a 2 km/h y alguien muy rápido a 4 km/h, cabe esperar una desviación típica próxima al valor 0,5 km/h, dado que la mitad de $4 - 2 = 2$ es 1 es 0,5.

2.4 El doble de la desviación típica es 16, que restado y sumado de 82, da 66 y 98. Se trata de una población anciana (82 años) pero que cubre un amplio margen, ya que fluctúa entre 66 y 98.

2.5 a) ii) De hecho, el valor exacto es 126,6 mmHg.

b) ii) De hecho, el valor exacto es 21,5 mmHg.

2.6 a) Media = 120 mmHg; varianza $S^2 = \frac{\sum (x_i - \bar{x})^2}{(n - 1)} = 1000 / 4 = 250 \text{ mmHg}^2$,

luego la desviación típica es $S = \sqrt{250 \text{ mmHg}^2} \approx 15,8 \text{ mmHg}$.

b) Media = 130 mmHg; varianza $S^2 = 58 / 4 \approx 14,5 \text{ mmHg}^2$, luego la desviación típica es

$S = \sqrt{14,5} \approx 3,8 \text{ mmHg}$.

2.7 Al paciente con un valor de 175 mg/dl le corresponde un desvío típico de $-0,5$, lo que indica que está ligeramente por debajo, ya que es negativo pero no alcanza la distancia habitual ($= 1$) que guardan los valores bajos con la media. El paciente con un valor de 350 mg/dl tiene un desvío típico de $+3$, lo que indica que está muy por encima, ya que su distancia es 3 veces mayor que la distancia típica de todos los que están por encima. Estadísticamente, se trataría de un caso extremo.

2.8 Con esta media y esta desviación típica, la regla de «buen cubero» (aproximada) dice que los casos deberían estar comprendidos entre:

Valores = media \pm 2 desviación típica = $400 \pm 2 \cdot 150 \approx 400 \pm 300 = [100, 700]$

Así, los valores que fueran inferiores a 100 o superiores a 700 serían sospechosos de acuerdo con este criterio univariante. Con un criterio bivariante, podría establecerse como sospechoso a un paciente que sufriera variaciones de su CD4 superiores al, por ejemplo, 50%.

Estos casos deberían ser contrastados con mucho cuidado, de acuerdo con su historia clínica, a la búsqueda de posibles errores de transcripción. Si no se encuentran errores, el valor debe darse por bueno.

2.9 Por supuesto, lo más próximo al momento en el que se generó el dato. De lo contrario, puede llegar a ser imposible verificarlo.

2.10 *a)* Media y desviación típica, ya que por experiencia previa cabe esperar una distribución simétrica.

b) Mediana y cuartiles 1 y 3 (o percentiles 25 y 75, que son lo mismo), ya que no parece simétrica.

c) Frecuencias y porcentajes de cada nivel I–IV.

d) Media y desviación típica, ya que parece simétrica.

Y recuerde que hay que informar, siempre, del número total, n , de casos.

Capítulo 3:

Probabilidad

Presentación

No es ningún secreto descubrir que muchos fenómenos son inciertos. A las mentes más deterministas les gusta pensar que esta incertidumbre es el resultado de nuestra falta de conocimiento: si se conocieran todas las fuerzas que actúan en un momento dado, se podría predecir exactamente el resultado. El modelo estadístico divide estas causas en dos grandes grupos: el primero lo forman un limitado número de fuerzas con entidad suficiente como para que su efecto pueda ser conocido y modelado de forma determinista; el segundo grupo, en cambio, es ilimitado y con influencia reducida: son tantas, y con efectos tan pequeños, que la única manera de modelar el resultado de su influencia es mediante la teoría de la combinatoria y de la probabilidad.

En este capítulo se introducen, con la ayuda de ejemplos, conceptos necesarios para interpretar los resultados de una prueba diagnóstica o la capacidad predictiva de un indicador.

A su vez, mediante el uso de la probabilidad condicionada se introduce al alumno en el concepto de riesgo y en las medidas de comparación de riesgos. Para aquellas situaciones en las que varía el tiempo de seguimiento, se introducen las tasas.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Interpretará el riesgo como una probabilidad.
- Interpretará la probabilidad de A condicionada a B como la probabilidad de A en el conjunto de casos que cumplen B.
- Distinguirá entre $P(A|B)$ y $P(B|A)$.
- Distinguirá entre sensibilidad, especificidad y valores predictivos.
- Definirá sensibilidad, especificidad y valores predictivos en términos de probabilidad condicionada.
- Para interpretar los resultados de una prueba diagnóstica, se preguntará por la prevalencia de la enfermedad en estudio.
- Sabrá que el teorema de Bayes también puede usarse con *odds*.
- Sabrá calcular un riesgo.
- Cuando los tiempos de seguimiento o el nivel de exposición varíen, calculará la tasa.
- Distinguirá entre riesgo y tasa.
- Distinguirá entre riesgo y *odd*.
- Sabrá calcular e interpretar el riesgo relativo, el atribuible y el *odds ratio*.
- En un estudio que fije los totales de enfermos y controles, usará el *odds ratio* y no el riesgo relativo ni el atribuible.

Probabilidad

Introducción a la probabilidad

Definición



La **probabilidad**, desde el punto de vista subjetivo, es el grado de certeza que se posee de un suceso.

Comentario



Se han propuesto varias definiciones de probabilidad. La primera definía la probabilidad como el «cociente entre casos favorables y casos posibles». Dada su circularidad (exige equiprobabilidad) enseguida se propuso sustituirla por «la frecuencia de aparición de un suceso». En el fondo, ambas definiciones son métodos para asignar valores a la probabilidad, es decir, para estimarla y conocer sus valores, pero no definiciones formales de lo que es en sí misma la probabilidad. En el siglo pasado se propusieron, casi al mismo tiempo, dos definiciones muy interesantes. Una de ellas, que era abstracta y axiomática, permitió un desarrollo formidable de toda la teoría de probabilidad. La otra definición, la subjetiva, permite expresar en términos de probabilidad tanto el grado de creencia en una teoría o en una afirmación científica, como la simple expectativa de la ocurrencia de un suceso. Su influencia en los conceptos de ciencia y de conocimiento aún está por desarrollarse plenamente.

Una representación gráfica del riesgo o probabilidad del suceso A , $P(A)$, es la siguiente (fig. 3-1), donde su valor concreto es directamente el cociente entre el área sombreada (A) y el área total (Ω).

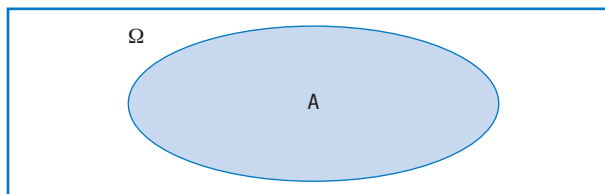


Figura 3-1 Probabilidad del suceso A .

Ejemplo 3.1



En 1978 se registraron 66 muertes por leucemia en una población de 890.575 individuos.

$$P = \text{muertes por leucemia} / \text{población total} = 66 / 890.575 = 0'0000741$$

Si se estima la probabilidad de morir por leucemia mediante el estimador proporción, se obtiene un riesgo de 74 muertes por millón de habitantes.

Recuerde



El riesgo es una probabilidad, por lo que no tiene unidades de medida y en el denominador están incluidos todos los casos del numerador.

Probabilidad condicionada

Existen infinidad de factores que pueden modificar esta probabilidad («las fuerzas con entidad suficiente»). Para considerarlas es preciso realizar la siguiente definición.

Definición



Probabilidad del suceso A condicionado al suceso B es la probabilidad de aparición del suceso A en los casos que cumplen también la característica B. Se representa por $P(A|B)$.

En el gráfico siguiente (fig. 3-2) su cálculo se obtiene mediante el cociente del área común a ambos sucesos ($A \cap B$) y el área del suceso B. Así, la probabilidad de A condicionada a B, $P(A|B)$, se puede entender como una reducción de la población en estudio: en lugar de considerar todos los casos de Ω , se reduce a los casos que cumplen la característica B.

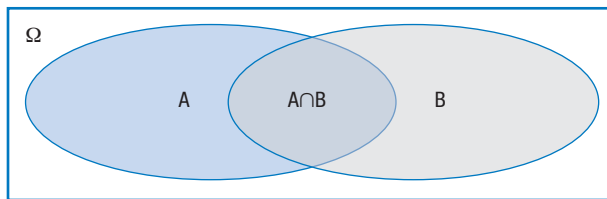


Figura 3-2 Probabilidad del suceso A condicionado al suceso B.

Recuerde



*$P(A|B)$ representa la probabilidad del suceso A «dentro» del total de casos que cumplen B.
 $P(A \cap B)$ representa la probabilidad de que ocurran simultáneamente los sucesos A y B.*

Definición



$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{Probabilidad de ser, a la vez, A y B}}{\text{Probabilidad de ser B}}$$

Ejercicio 3.1



¿Qué sería en el gráfico $P(B|A)$? ¿Cuál sería su definición?

Aplicación de la probabilidad al diagnóstico

Lectura



Bossuyt et al. (19) ponen como ejemplo de una buena definición de objetivos en un estudio de capacidad diagnóstica: «Objetivo: determinar la sensibilidad y la especificidad de la colonografía tomográfica computarizada aplicada a la detección de los pólipos y el cáncer colorrectales tomando la colonoscopia como estándar de referencia». Y añaden: «Los estudios de precisión diagnóstica tienen una estructura básica común. Se evalúan una o más pruebas con el objetivo de detectar o predecir una cierta condición. La condición objetivo puede ser una enfermedad específica, el estadio de una enfermedad, un nivel de salud o cualquier otra condición del paciente. [...] En este contexto, “prueba” se refiere a cualquier método que permita obtener información adicional sobre el nivel de salud del paciente. Entre estas pruebas están los análisis de laboratorio, las técnicas de imagen, las determinaciones funcionales, el estudio anatomopatológico, la historia clínica y la exploración física. En un estudio de precisión diagnóstica, la prueba evaluada (denominada en este caso prueba índice) se aplica a una serie de individuos. Los resultados obtenidos con la prueba índice se comparan con los correspondientes al estándar de referencia, obtenidos en los mismos individuos. En este contexto, el estándar de referencia es el mejor método existente para establecer la presencia o ausencia de la condición objetivo y puede constar de una única prueba o una combinación de métodos y técnicas, incluyendo el seguimiento clínico de los individuos evaluados. El término de precisión se refiere al grado de concordancia entre los resultados obtenidos con la prueba índice y los obtenidos con el estándar de referencia. La precisión diagnóstica se puede expresar de distintas maneras, como el binomio sensibilidad-especificidad, la razón de verosimilitud, el cociente de posibilidades diagnósticas y el área bajo la curva ROC [receiver-operating characteristic]».

Por simplicidad, supóngase que se debe realizar el proceso diagnóstico de una sola enfermedad con dos únicos posibles estados, enfermo (**E**) y sano (**S**), y que se dispone de un único indicador con dos posibles valores, positivo (+) y negativo (-).

Recuerde



La primera dificultad consiste en definir qué variable mide perfectamente, sin error, el estado enfermo / sano; y qué variable representa el resultado del test positivo / negativo.

Definición


Sensibilidad (*Sens*): tendencia o propensión de los enfermos a dar positivo (en esta prueba).

Especificidad (*Esp*): tendencia o propensión de los sanos a dar negativo.

Valor predictivo positivo (*VP+*): confianza o credibilidad de un resultado positivo (de esta prueba).

Valor predictivo negativo (*VP-*): confianza o credibilidad de un resultado negativo.

Ejemplo 3.2


Una prueba sería extremadamente sensible si, aplicada a un conjunto de enfermos, casi el 100% dan positivo. Y un resultado tendrá un elevado valor predictivo positivo si casi el 100% de los que dan positivo están realmente enfermos. ¡Parece lo mismo, pero no lo es! La probabilidad condicionada, como se verá, ayuda a distinguirlos.

Ejemplo 3.3


Gisbert et al. (20) obtienen los siguientes resultados: «se incluyó a 117 pacientes con enfermedad de Crohn (EC), 72 con colitis ulcerosa (CU) y 2 con colitis indeterminada. Un paciente con EC (0,9%) y 6 con CU (8,3%) presentaron positividad en la determinación de PANCA. La sensibilidad, la especificidad, los valores predictivos positivo y negativo de los PANCA para el diagnóstico de CU (en comparación con la EC) fueron del 8, el 99, el 86 y el 64% respectivamente, lo que indica la necesidad de estandarizar la metodología. Conclusiones: la sensibilidad para el diagnóstico de CU en los pacientes con enfermedad inflamatoria intestinal es muy baja, si bien la especificidad es muy elevada».

Utilizando la probabilidad condicionada, se pueden definir estos conceptos con mucha mayor formalidad, aclarando sus diferencias y similitudes. De hecho, aparecen cuatro probabilidades condicionadas muy interesantes: $P(E|+)$ o probabilidad de enfermo si ha dado positivo, $P(+|E)$ o probabilidad de positivo en los enfermos, $P(-|S)$ o probabilidad de negativo en los sanos y $P(S|-)$ o probabilidad de sano en los negativos. Y también pueden ser interpretados usando la proporción muestral en lugar de la probabilidad poblacional: $P(S|+)$ o proporción de positivos que están sanos, $P(+|E)$ o proporción de enfermos que dan positivo, $P(-|S)$ o proporción de sanos que dan negativo y $P(S|-)$ o proporción de negativos que están sanos.

Ejercicio 3.2

Haga corresponder cada uno de estos términos con los conceptos anteriores.

Ejercicio 3.3

Intente explicar con sus propias palabras qué significan la especificidad y el valor predictivo negativo.

En la sensibilidad y en la especificidad, el condicionante, o punto de salida, es la realidad (enfermo o sano); mientras que el condicionado, o punto de llegada, es el indicador (positivo o negativo). Así, ambas van de la realidad al resultado: siguen el planteamiento «racionalista» de ir de la causa a la consecuencia. Los valores predictivos, en cambio, van al revés: condicionan o parten del indicador y llegan al estado real del paciente. Por ello contestan la pregunta «empirista» del diagnóstico: dado que este paciente ha presentado estos síntomas, signos e indicadores, ¿está enfermo? O mejor: ¿qué probabilidades tiene de estar realmente enfermo sabiendo que ha dado positivo, $P(E|+)$? ¿O de estar sano si ha dado negativo, $P(S|-)$?

Ejemplo 3.4

La sensibilidad, la especificidad y los valores predictivos de la tabla 3-1, que se muestra a continuación son:

	+	-	TOTAL	
Enfermo	94	38	132	Sens = $P(+ E) = 94/132 \approx 0,712 = 71,2\%$
Sano	215	653	868	Esp = $P(- S) = 653/868 \approx 0,752 = 75,2\%$
TOTAL	309	691	1.000	VP+ = $P(E +) = 93/309 \approx 0,304 = 30,4\%$
				VP- = $P(S -) = 653/691 \approx 0,945 = 94,5\%$

Tabla 3-1 Probabilidades diagnósticas en una muestra con un 13,2% de enfermos

A pesar de que los valores de sensibilidad y especificidad son muy similares, los valores predictivos se alejan considerablemente entre sí: la probabilidad de que un paciente que dé negativo esté sano (VP-) es alta, pero la probabilidad de que un paciente que dé positivo esté enfermo (VP+) es baja. La razón es muy simple: ya que hay más sanos (86,8%) que enfermos (13,2%), la segunda fila pesa más que la primera.

Ejemplo 3.5



Tal y como puede observarse en la tabla 3-2, si cambiamos la P(E) de 0,132 a 0,75, pero mantenemos una sensibilidad y una especificidad iguales a las anteriores, los valores predictivos cambian.

	+	-	TOTAL	
Enfermo	534	216	750	Sens = P(+/E) = 537/750 ≈ 0,712 = 71,2%
Sano	62	188	250	Esp = P(-/S) = 188/250 ≈ 0,752 = 75,2%
TOTAL	596	404	1.000	VP+ = P(E/+) = 534/596 ≈ 0,896 = 89,6%
				VP- = P(S/-) = 188/404 ≈ 0,465 = 46,5%

Tabla 3-2 Probabilidades diagnósticas en una muestra con un 75% de enfermos

No debe olvidarse cuál es la proporción de enfermos en la población origen de los datos. Si es muy baja, la proporción de enfermos seguirá siendo baja después del resultado de la prueba, lo que llevará a un valor predictivo positivo (proporción de enfermos en los positivos) muy bajo y a un valor predictivo negativo (proporción de sanos en los negativos) muy alto. Un observador poco atento se podría olvidar de las condiciones de salida y esperar que, si un 70% de los enfermos dan positivo, pues aproximadamente un 70% de los positivos estarán enfermos.

La sensibilidad y la especificidad son el resultado de un mecanismo fisiológico o patológico y, por tanto, suelen ser extrapolables de una población de pacientes a otra. En cambio, los valores predictivos también dependen de la frecuencia de la enfermedad en la población, por lo que variarán mucho de unas condiciones a otras.

Recuerde



La sensibilidad y la especificidad pueden ser universales, pero los valores predictivos dependen de la frecuencia de la enfermedad en la población en estudio.

Ejemplo 3.6



Suponga dos poblaciones diferentes: 1) la formada por los toxicómanos y homosexuales de una cierta prisión, y 2) la formada por los médicos de un cierto hospital. Suponga un resultado positivo a la misma prueba del sida de dos miembros de cada una de estas dos comunidades. ¿Tienen es-

Ejemplo 3.6 (Cont.)

tos dos individuos las mismas probabilidades de padecer sida? La respuesta es no: delante de un resultado positivo a una prueba del sida, el valor predictivo positivo o probabilidad de estar realmente enfermo habiendo dado positivo es mucho mayor en el primer caso. Más adelante se verá una exposición formal de este hecho aparentemente paradójico.

Comentario

Todos estos fenómenos paradójicos tienen menor repercusión cuanto menos aleatorio es el fenómeno en estudio. Si la sensibilidad y la especificidad son ambas del 100%, no hay duda de que un positivo está enfermo y un negativo está sano.

Ejercicio 3.4

Construya la tabla 2×2 del ejemplo anterior con los números correspondientes a esta nota (sensibilidad = especificidad = 100%) y compruebe que la afirmación realizada en la nota es cierta (suponga que de 1000 casos estudiados, estaban enfermos 132 y 868, sanos). ¿Se cumple que ambos valores predictivos son del 100%?

Ejercicio 3.5

Mirando a la tabla construida en el ejercicio anterior, para que el valor predictivo positivo sea perfecto ($VP+ = 100\%$), ¿qué se necesita que sea del 100%, la sensibilidad o la especificidad?

Ejercicio 3.6

En clínica se dice que un signo es patognomónico cuando su presencia asegura definitivamente la enfermedad que ayuda a diagnosticar. ¿Alguna de las probabilidades diagnósticas es del 100%?

Ejercicio 3.7

Intente explicar con sus propias palabras lo que miden la sensibilidad, la especificidad y ambos valores predictivos.

Ejercicio 3.8

Repita los cálculos, suponiendo que se habían obtenido dos muestras de 1.000 casos, una para enfermos y otra para sanos. Tabla 3-3.

	+	-	TOTAL
Enfermo	712	288	1.000
Sano	248	752	1.000
TOTAL	960	1.040	

Tabla 3-3 Probabilidades diagnósticas en dos muestras de 1.000 casos

En el ejercicio anterior ambas filas pesan lo mismo, y los valores predictivos se han parecido más a la sensibilidad y especificidad. Pero este peso similar de ambas filas se ha fijado artificialmente, por lo que estos valores predictivos sólo se corresponderían con una realidad en la que ambas filas pesaran lo mismo. Cuando el número de enfermos y de sanos está fijado por muestreo, el cálculo de los VP no puede hacerse con el procedimiento anterior y debe usarse el que se explica en el siguiente apartado.

Recuerde



La mayor parte de diseños para calcular la capacidad de una prueba diagnóstica se basan en una muestra de enfermos obtenida de un determinado servicio y una muestra de sanos obtenida por otro medio. Dado que el número de casos en cada fila lo decide el investigador, no se pueden calcular directamente los valores predictivos.

Teorema de Bayes

El teorema de Bayes permite resolver el problema anterior, ya que incorpora la información sobre la proporción de enfermos que hay en la población en estudio. Esquemáticamente, consiste en combinar los valores de sensibilidad y especificidad con esta proporción de enfermos para obtener los valores predictivos (fig. 3-3).

Veamos en primer lugar la fórmula del teorema de Bayes.

Definición



Sean A y B dos eventos cualesquiera (por ejemplo A=ser rubio y B=ser nórdico)

Teorema de Bayes:
$$P(B|A) = \frac{P(A|B) P(B)}{P(A)}$$

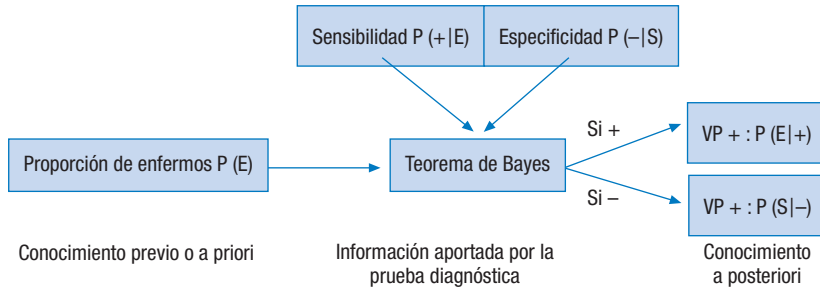


Figura 3-3 Esquema del teorema de Bayes.

Nótese que el teorema de Bayes permite invertir los términos de condicionante y condicionado: a partir de $P(A|B)$ se obtiene $P(B|A)$, y por tanto, a partir de $P(+|E)$ se llega a $P(E|+)$.

Nota técnica



Demostración a partir de la definición de probabilidad condicionada

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B) \cdot P(B)$$

$$\text{Y entonces } P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Comentario



El monje Bayes quería demostrar la existencia de Dios (D). Partía de que el ser humano existe [$P(H)=1$] y de que Dios, por su propia definición, crearía la humanidad [$P(H|D) = 1$]. Su pregunta era: ¿cuál es la probabilidad de que exista Dios, sabiendo que el ser humano existe, [$P(D|H)$]? Y Bayes, para contestar a esta pregunta, desarrolló el teorema que lleva su nombre. La respuesta que alcanzó fue:

$$P(D|H) = \frac{P(H|D) \cdot P(D)}{P(H)} = \frac{1 \cdot P(0)}{1}$$

Bayes observó que todo dependía de sus probabilidades iniciales de creer en Dios. Si él, a priori ya creía en Dios, $P(D)=1$ entonces, habiendo observado que existía el hombre, también $P(D|H)=1$. En cambio, si a priori no lo hacía: $P(D)=0$, tampoco a posteriori; tras observar al hombre: $P(D|H)=0$. Por lo que su teorema no servía para demostrar la existencia de Dios. Thomas Bayes, al no conseguir sus objetivos, olvidó sus resultados, que sólo fueron publicados póstumamente por el prior que le sucedió.

Nota técnica



La fórmula del teorema de Bayes se puede expresar también mediante la descomposición del denominador A en los dos sucesos $A \cap B$ y $A \cap B^c$ que lo componen (B^c representa el contrario o negación de B; si, por ejemplo, B representa enfermo, B^c representará sano):

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A \cap B) + P(A \cap B^c)} =$$

$$= \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)}$$

Recuerde



Una expresión alternativa del teorema de Bayes es:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)}$$

Ejemplo 3.7



Las sensibilidades y especificidades obtenidas en el ejercicio 3.8 se quieren aplicar en un entorno en el que hay una $P(E) = 0,132$. El valor predictivo positivo será:

$$P(E|+) = \frac{P(+|E) \cdot P(E)}{P(+|E) \cdot P(E) + P(+|S) \cdot P(S)}$$

$$= \frac{0,712 \cdot 0,132}{0,712 \cdot 0,132 + 0,248 \cdot 0,868} \approx 0,304$$

Ejercicio 3.9



Compruebe que sabe calcular el valor predictivo negativo.

El teorema de Bayes permite, a partir de la proporción de enfermos, $P(E)$, de una población y de la sensibilidad y especificidad de una prueba, obtener los valores predictivos que se pueden aplicar a un paciente determinado.

Comentario



En ocasiones, se habla de **falsos positivos** (FP) y de **falsos negativos** (FN), que representan los complementarios o contrarios de la especificidad y sensibilidad, respectivamente. Es decir, si la sensibilidad es del 90%, quiere decir que un 10% de los enfermos dan un resultado negativo que es falso.

Ejemplo 3.8

Una prueba diagnóstica para la diabetes tiene FP de 4% y FN del 5%. Si la prevalencia de la diabetes en la población donde se usa es del 7%, ¿cuál es la probabilidad de que sea diabético un individuo en el que la prueba dé positivo? ¿Y de que no lo sea uno en el que dé negativo?

$$p(+|S) = 0,04 \rightarrow p(-|S) = 1 - 0,04 = 0,96 = \text{Esp}$$

$$p(-|E) = 0,05 \rightarrow p(+|E) = 1 - 0,05 = 0,95 = \text{Sens}$$

$$p(E) = 0,07 \rightarrow p(S) = 1 - 0,07 = 0,93$$

$$\begin{aligned} P(E|+) &= \frac{P(+|E) \cdot P(E)}{P(+|E) \cdot P(E) + P(+|S) \cdot P(S)} = \\ &= \frac{0,95 \cdot 0,07}{0,95 \cdot 0,07 + 0,04 \cdot 0,93} = 0,641 = \text{VP}^+ \end{aligned}$$

$$\begin{aligned} P(S|-) &= \frac{P(-|S) \cdot P(S)}{P(-|S) \cdot P(S) + P(-|E) \cdot P(E)} = \\ &= \frac{0,96 \cdot 0,93}{0,96 \cdot 0,93 + 0,05 \cdot 0,07} = 0,996 = \text{VP}^- \end{aligned}$$

El teorema de Bayes permite ir actualizando la información de que se dispone: su fórmula «mezcla» la información previa, disponible a priori $P(E)$, con los nuevos resultados (+ o -).

Recuerde

Información a priori + *nueva información* = *Información a posteriori*

Riesgos y tasas**Riesgo**

En un sentido amplio, el riesgo es la probabilidad de que algo ocurra. En investigación clínica, es el resultado desfavorable de una actividad, intervención o exposición, especialmente referido a la probabilidad de que aparezca un fenómeno adverso concreto.

Lectura

En el caso de estudios sin beneficio terapéutico, se acepta como lícito el hecho de que los voluntarios se encuentren sometidos a riesgo mínimo o insignificante, según la FDA. Se acepta como tal la probabilidad de entre 1 y 100 por mil de sufrir una complicación menor y de entre 10 y 1.000 por millón de sufrir una complicación grave (21, p. 56).

Ejemplo 3.9



Suponga que el diagnóstico de una enfermedad (Y) y la existencia de un Factor de Riesgo (X) sólo pueden tomar dos valores (es decir, son «dicotomías») que son: presente (+) y ausente (-). Así, por ejemplo, Y+ representará que tiene la enfermedad y X- que no está expuesto al factor de riesgo. El riesgo absoluto de sufrir la enfermedad en estudio sería P(Y+). El riesgo entre los expuestos sería: P(Y+ | X+). El riesgo entre los no expuestos sería P(Y+ | X-). Es habitual presentar simultáneamente estas dos dicotomías en forma de una tabla 2 x 2, en las que los marginales representan los totales de cada variable. Así, en la tabla 3-4 puede leerse que de los 1.000 casos estudiados, 15 presentaban la enfermedad, de los que 7 estaban expuestos y 8 no.

Ahora se pueden calcular los riesgos habituales con las fórmulas de probabilidad condicionada:

Riesgo: $P(Y+) = 15 / 1.000 = 0,015$

Riesgo en los Expuestos: $P(Y+|X+) = 7 / 132 \approx 0,053$

Riesgo en los No-Expuestos: $P(Y+|X-) = 8 / 868 \approx 0,009$

	Y+	Y-	TOTAL
X+	7	125	132
X-	8	860	868
TOTAL	15	985	1.000

Tabla 3-4 Presencia de la enfermedad Y y el factor de riesgo X en 1.000 casos

Ejercicio 3.10



Fumaz et al. (22) comparan dos grupos de pacientes tratados, uno con efavirenz (EFV, n = 51) y otro con inhibidores de la proteasa (PI, n = 49), habiendo observado respectivamente 36 y 9 acontecimientos adversos relacionados con el sistema nervioso central. Elabore la tabla 2 x 2 y calcule los riesgos respectivos.

Tasa

En muchas ocasiones, se observa a los pacientes bajo diferentes circunstancias. La más habitual es que diversos casos hayan sido estudiados durante un diferente tiempo de seguimiento. Por ello, no es lo mismo que el evento de interés lo haya experimentado un caso seguido durante 1 mes que uno seguido durante 2 años. Conviene considerar esta nueva información a la hora de calcular los riesgos. Para ello, se debe incluir en el denominador el tiempo total de seguimiento de cada caso, con lo que el

riesgo así calculado es el cociente entre un número de casos y una suma de tiempos de seguimiento, por lo que ya no se trata de una probabilidad, que eran casos (posibles) divididos entre casos (totales). Por ello, se habla de tasas. En los temas específicos de supervivencia se estudian con más detalle.

Definición



Tasa es la medida de la frecuencia de un evento expresada de forma relativa al tiempo.

Ejemplo 3.10



Regidor et al. (23). En 1998 se produjeron en España 360.511 defunciones, lo que supone una tasa de mortalidad de 915,7 por 100.000 habitantes en un año de seguimiento. En la tabla 3-5 aparecen ordenadas jerárquicamente, en virtud del número de fallecimientos, las 12 causas de muerte estudiadas. Esas causas ocasionaron el 78% de todos los fallecimientos ocurridos en España ese año. Los tumores malignos fueron la principal causa de muerte, con una tasa de mortalidad de 227,7 por 100.000 habitantes-año, lo que supone el 24,9% de todas las defunciones; las enfermedades del corazón figuraban en segundo lugar, con una tasa de 207,7 por 100.000 y el 22,7% de los fallecimientos, y las enfermedades cerebrovasculares en el tercero, con una tasa de 96,8 por 100.000 y el 10,6% de las defunciones.

Odd

Los países de tradición anglosajona también usan una forma alternativa a la probabilidad para expresar los resultados inciertos. Si la probabilidad expresa los casos a favor divididos entre todos los casos posibles, las *odds* hablan de los casos a favor divididos entre los casos en contra.

Ejemplo 3.11



Así, mientras nosotros diríamos que cierto caballo tiene 7 números sobre 8 de ganar una carrera, los anglosajones podrían decir también que los números de este caballo están 7 a 1, indicando que tienen 7 a favor y 1 en contra.

Comentario




En los ambientes de apuestas (pelota vasca, carreras de galgos, etc.) se emplea un término equivalente a las odds. Se dice, por ejemplo, que las apuestas por cierto pelotari están 7 momios a 1. ¿Alguien se atreve a traducir odd por momio?

Causas de muerte (CIEB.* revisión)	Defunciones	Tasas brutas	Mortalidad proporcional (%)	Porcentaje de cambio en la tasa ajustada de mortalidad	
				1995-1998	1980-1988
Todas las causas	360.511	915,7	100,0	-3,2*	-20,2*
Cáncer (140-208)	89.665	227,7	24,9	-2,4*	7,7*
Enfermedades del corazón (390-398, 410-429)	81.768	207,7	22,7	-1,8*	-26,4*
Enfermedad cerebrovascular (430-438)	38.121	96,8	10,6	-12,0*	-52,8*
Enfermedad pulmonar obstructiva crónica (490-796)	17.768	45,1	4,9	1,1 NS	30,8*
Accidentes no intencionales (E800-E949)	13.122	33,3	3,6	0,1 NS	-21,6*
Diabetes mellitus (250)	9.533	24,2	2,6	-3,4***	-15,1*
Neumonía e influenza (480-487)	8.491	21,6	2,4	3,9***	-52,6*
Cirrosis y otras enfermedades crónicas del hígado (571)	6.246	15,9	1,7	-14,3*	-45,3*
Nefritis, síndrome nefrótico y nefrosis (580-589)	5.566	14,1	1,5	-1,6 NS	-16,1*
Aterosclerosis (440)	4.717	12,0	1,3	-26,5*	-81,2*
Enfermedad de Alzheimer (331,0)	3.551	9,0	1,0	42,9*	2.111,6*
Suicidio (E950-E959)	3.261	8,3	0,9	0,1 NS	53,8*

*p < 0,001; **p < 0,01; ***p < 0,05; NS: no significativo.

Tabla 3-5 Principales causas de muerte en España en 1998. Número de defunciones, tasas de mortalidad por 100.000 habitantes, mortalidad proporcional y porcentajes de cambio en la mortalidad ajustada por edad de 1995 a 1998 y de 1980 a 1998




Definición

La *odd* de A es la probabilidad de que se presente el suceso A dividida por la probabilidad de que no se presente A.


$$Odd(A) = \frac{P(A)}{P(\text{no } A)}$$

Ejercicio 3.11



¿Cuánto valen la probabilidad y la *odd* de sacar un «3» en el lanzamiento de un dado?

Ejemplo 3.12



Veamos cómo se calculan las *odds* en los mismos datos del ejemplo 3.9 (tabla 3-4):

Odd expuestos: $O(Y+ | X+) = 7 / 125 \approx 1 / 18 \approx 0,056$

Odd no expuestos: $O(Y+ | X-) = 8 / 860 \approx 1 / 107 \approx 0,0093$

En los expuestos, la enfermedad aparece en 1 caso por cada 18 en los que no aparece. En cambio, en los no expuestos, la enfermedad aparece en 1 caso por cada 107 en los que no.

Nota técnica

Si la probabilidad de enfermedad es muy pequeña, la probabilidad de estar sano será muy próxima a 1, por lo que la *odd* tendrá un valor muy similar a la probabilidad:

$$Odd(\text{enfermo}) = \frac{P(\text{enfermo})}{P(\text{sano})} \approx \frac{P(\text{enfermo})}{1} = P(\text{enfermo})$$

Ejercicio 3.12

El dolor lumbar o la gripe son enfermedades comunes en el sentido de que a lo largo de la vida es fácil padecerlas al menos en una ocasión. Pongamos que sus probabilidades respectivas son 0,5 y 0,8. Calcule sus odds. La esclerosis múltiple, en cambio, es muy poco frecuente. Pongamos que la probabilidad de padecerla a lo largo de la vida es de 0,001 (uno por mil). Calcule la *odd*.

Recuerde

En el caso de enfermedades «raras», el riesgo y la odd dan resultados muy similares.

Ejemplo 3.12 (Cont.)

La *odd* en los expuestos vale 0,056 [$O(Y+ | X+) = 7/125 \approx 0,056$], valor muy similar al 0,053 anterior del riesgo en los expuestos.

En los no expuestos la similitud es aún mayor, ya que la *odd* vale 0,0093 por 0,0092 el riesgo.

Un aspecto importante a tener en cuenta es la forma en la que se han recogido los datos, ya que si, se ha *forzado* la recogida de datos para que tenga mayor representación cierto tipo de casos, algunos de los cálculos pueden no ser correctos.

Comentario

De la misma forma que los valores predictivos no podían calcularse si el diseño implicaba dos muestras, una de enfermos y una de sanos, el siguiente ejercicio evidencia que tampoco pueden calcularse los riesgos.

Ejercicio 3.13



En una cierta población, se ha recogido información en 1.000 casos sobre dos variables, la exposición al riesgo del tabaco (fumador: X+ y no fumador: X-) y su evolución posterior (bronquitis: Y+ y no bronquitis: Y-). A partir de los resultados de la tabla 3-6A, calcule los riesgos y las odds.

Repita los cálculos en la tabla 3-6B, en la que los datos se han obtenido de dos muestras de 1000 casos, una de fumadores y otra de no fumadores.

Repita una vez más, tabla 3-6C, suponiendo ahora que las dos muestras de 1.000 casos corresponden una a bronquíticos y otra a no bronquíticos.

	Y+	Y-	TOTAL
X+	94	38	132
X-	215	653	868
TOTAL	309	691	1.000

Tabla 3-6 A Datos sobre una única muestra de 1.000 casos

	Y+	Y-	TOTAL
X+	712	288	1.000
X-	248	752	1.000
TOTAL	960	1.040	

Tabla 3-6 B Datos sobre dos muestras de 1.000 casos, una de fumadores (X+) y otra de no fumadores (X-)

	Y+	Y-	TOTAL
X+	304	55	359
X-	696	945	1.641
TOTAL	1.000	1.000	

Tabla 3-6 C Datos sobre dos muestras de 1.000 casos, una de pacientes con bronquitis (Y+) y otra sin (Y-)

Recuerde

Si se fija por diseño el número total de casos con la enfermedad y el número de casos sin la enfermedad, ya no pueden calcularse ni los riesgos ni las odds de desarrollar la enfermedad (ni en los expuestos ni en los no expuestos).

Odds diagnóstica

Una de las ventajas de trabajar con odds en lugar de con probabilidades es que el teorema de Bayes se simplifica mucho. Como explica la nota técnica siguiente, conocer las odds de enfermo a sano tras la obtención de un valor positivo del indicador diagnóstico consiste en multiplicar la **razón de verosimilitud** (*likelihood ratio*), de un resultado positivo (sensibilidad dividida entre el complementario de la especificidad) por la *odd* previa de enfermo a sano en un entorno determinado.

Nota técnica

Utilizando probabilidades condicionadas se obtuvo que:

$$P(E|+) = P(+|E) P(E)/P(+) \quad \text{y} \quad P(S|+) = P(+|S) P(S)/P(+)$$

Recordando la definición de la *odd* ($E = P(E)/P(S)$), y aplicándola a las probabilidades condicionadas anteriores, se pueden obtener las odds a posteriori entre enfermo (E) y sano (S), una vez conocido el resultado positivo de la prueba:

$$\underbrace{\frac{P(E|+)}{P(S|+)}}_{\substack{\text{Odds} \\ \text{a} \\ \text{posteriori}}} = \frac{P(+|E) P(E)/P(+)}{P(+|S) P(S)/P(+)} = \underbrace{\frac{P(+|E)}{P(+|S)}}_{\substack{\text{Razón de} \\ \text{verosimi-} \\ \text{litud (RV)}}} \cdot \underbrace{\frac{P(E)}{P(S)}}_{\substack{\text{Odds} \\ \text{a} \\ \text{priori}}}$$

Ejemplo 3.4 (Cont.)

La razón de verosimilitud (RV) de un resultado positivo vale

$$\frac{P(+|E)}{P(+|S)} = \frac{0,712}{1-0,752} \approx 2,87$$

Si se sabe que en cierto servicio sanitario hay 1 enfermo por cada 3 sanos (odds a priori), una vez haya dado positivo el indicador diagnóstico, habrá aproximadamente 1 enfermo por cada 1 sano (*odds a posteriori* = *RV* × *odds a priori* = $2,87 \times 1/3 = 0,96 \rightarrow 0,96$ enfermos por cada 1 sano ≈ 1 enfermo por cada sano).

Ejercicio 3.14



Suponga que el resultado de la prueba en el ejemplo anterior ha sido negativo, ¿Cuánto vale la RV de enfermo para un resultado negativo? ¿Cuánto valen las *odds* a posteriori de enfermo para un resultado negativo?

Nota técnica



Si se desea trabajar de forma aditiva, se puede recurrir a tomar logaritmos

$$\underbrace{\log(Odds_{posteriori})}_{\text{Información a posteriori}} = \underbrace{\log(RV)}_{\text{Función SOPORTE}} + \underbrace{\log(Odds_{priori})}_{\text{Información a priori}}$$

Es decir, los *logodds* a priori más la información aportada por el resultado empírico («función soporte»), proporcionan los *logodds* a posteriori.

Ya se ha dicho que las *odds* («casos a favor» / «casos en contra») son una forma alternativa de presentar la probabilidad («casos a favor» / «casos totales») que suele emplearse en la cultura anglosajona en los entornos de apuestas. Ahora se ha visto que conduce a cálculos más simples y directos, lo que representa su principal ventaja. En el libro de Guyat et al. (24) pueden verse numerosos ejemplos de su aplicación en la práctica clínica.

Incidencia y prevalencia

A continuación se definen dos medidas básicas de frecuencia de morbilidad.

Definición



La incidencia se refiere al número de casos nuevos de una enfermedad en una población durante un período de tiempo determinado.

Ejemplo 3.13



Cohn et al. (25). La incidencia del criterio de valoración fue un 13,2% más baja con Valsartán que con placebo (riesgo relativo, 0,87).

Definición

La prevalencia define el número de individuos que presenta una determinada característica o enfermedad en una población y en un momento de tiempo determinado.

Ejemplo 3.14

Martín et al. (26). El asma es un grave problema de salud en los países industrializados, donde supone una de las enfermedades crónicas más frecuentes, sobre todo en la infancia. En Europa, la prevalencia varía ampliamente entre los distintos países, con cifras que oscilan entre un 8% en el Reino Unido y un 2% en Grecia. También existen amplias diferencias entre distintas zonas del mismo país, encontrándose en España cifras entre el 5 y el 1%.

Ejemplo 3.15

Ricart et al. (27). Recientemente, la American Diabetes Association (ADA) concluyó que no se deben realizar estudios para detectar la diabetes mellitus gestacional (DMG) en el grupo de mujeres gestantes con bajo riesgo para desarrollar esta enfermedad. El objetivo de este trabajo es determinar en una población española la prevalencia de DMG en un grupo de gestantes con bajo riesgo [...]. Diseño y métodos: revisión de una cohorte de 2.262 gestaciones (2.085 caucásicas), controlada desde 1990 a 1998. Se estudia la prevalencia y las características de las gestantes de bajo riesgo. Resultados: la prevalencia de DMG fue del 15%. Doscientas setenta y cuatro mujeres (12,1%) se catalogaron de bajo riesgo de desarrollar una DMG, 13 de las cuales (4,7%) presentaban DMG, que contrasta con el 16,6% en el resto de las mujeres estudiadas.

Comentario

Prevención: cualquier intervención que reduzca el riesgo de que una enfermedad o trastorno afecte a un individuo, que interrumpa o detenga su progreso o evite la muerte. Podemos distinguir:

- Prevención **primaria:** cualquier intervención dirigida a individuos sanos sin la enfermedad cuya aparición se quiere evitar.
- Prevención **secundaria:** desde un punto de vista diagnóstico, intervenciones dirigidas a detectar precozmente una enfermedad.
- Prevención **terciaria:** intervención preventiva en pacientes que ya han padecido una enfermedad relacionada con la rehabilitación y la mejora de la calidad de vida.

Ejemplo 3.16



- De la prevención primaria: vacunaciones, modificación de factores de riesgo como el tabaco para evitar el cáncer o la utilización del ácido acetilsalicílico para prevenir el infarto de miocardio en sujetos sanos.
- De la prevención secundaria: la utilización de la mamografía para la detección del cáncer de mama o la prueba de Papanicolaou para el cáncer de cérvix.
- De la prevención terciaria: tratamiento fisioterapéutico y rehabilitador en pacientes que han padecido un accidente cerebrovascular.

Medidas de asociación en tablas 2 × 2

Una vez estimado el riesgo de desarrollar una enfermedad en un grupo determinado puede ser interesante la comparación de estos riesgos entre dos grupos de diferentes características.

Ejemplo 3.9 (Cont.)



Habíamos visto en el **ejemplo 3.9**, al hablar de probabilidad condicionada, que el riesgo o probabilidad en los expuestos era de un 5,3% [$P(Y+|X+) = 7 / 132 \approx 0,053$], mientras que en los no expuestos era del 0,9% [$P(Y+|X-) = 8 / 868 \approx 0,009$] (tabla 3-4).

Estos dos números se pueden comparar matemáticamente mediante la resta o la división. Así, el riesgo atribuible consiste en hacer la diferencia entre el riesgo en los expuestos y los no expuestos, mientras que el riesgo relativo es su cociente.

Lectura



CONSORT (9): «Para cada criterio de valoración principal y secundario. [...] En respuestas binarias, la medida del efecto podría ser el cociente de riesgo (riesgo relativo), el odds ratio o el riesgo atribuible; en lo relativo a los datos del tiempo de supervivencia, la medida del efecto podría ser la razón de tasas o la diferencia en la supervivencia mediana. En los datos continuos la medida del efecto suele ser la diferencia en las medias. [...]

En lo que se refiere a los datos binarios y supervivencia, también puede ser útil la expresión de los resultados en forma del número necesario de pacientes que se han de tratar para la obtención de efecto beneficioso (NNTB, number needed to treat for benefit) o de efecto adverso (NNTH, number needed to treat for harm).»

Riesgo atribuible

Definición



El *riesgo* de una enfermedad *atribuible* a una exposición es la diferencia entre el riesgo en los expuestos y el mismo riesgo en los no expuestos.

Ejemplo 3.16 (Cont.)



En el ejemplo anterior, la diferencia entre 0,053 y 0,009 es 0,044; es decir, expresado en porcentajes, un 4,4%.

El riesgo atribuible puede tomar el valor máximo de 1 (o 100%) en el caso de que todos los expuestos desarrollaran la enfermedad y, en cambio, no lo hiciera ninguno de los no expuestos. El valor 0 representa que el riesgo es el mismo en ambos grupos.

Nota



En teoría, sería posible observar valores negativos hasta un mínimo de -1 (-100%) en el caso de que el factor en estudio protegiera de la enfermedad en lugar de favorecer su aparición, aunque siempre es más fácil de interpretar si los resultados se expresan en sentido positivo, para lo que basta con intercambiar las definiciones de expuesto y no expuesto.

Nota técnica



El nombre de riesgo atribuible es muy ambicioso, ya que da a entender que este incremento del riesgo viene exclusivamente originado por la exposición al factor, para lo que es necesaria la existencia de una relación de causa-efecto. Establecer esta relación precisa toda una serie de consideraciones que se estudian más adelante y que van más allá de la simple medida de asociación entre dos variables. Mientras tanto, el concepto riesgo debe valorarse únicamente a nivel predictivo: los expuestos tienen ese mayor riesgo que los no expuestos.

Recuerde



El RA valora la diferencia de riesgos. Si la relación fuera causal, podría decirse que la exposición añade o suma el riesgo RA al riesgo de los no expuestos.

Riesgo relativo

Definición



El riesgo de una enfermedad en los expuestos *relativo* a los no expuestos es el cociente entre el riesgo en los expuestos y el mismo riesgo en los no expuestos.

Ejemplo 3.12 (Cont.)



En el **ejemplo 3.12** (cont. del 3.9), la razón entre 0,053 y 0,009 es prácticamente 6, lo que indica que los expuestos tienen un riesgo casi 6 veces superior a los no expuestos.

El riesgo relativo pretende evaluar cuánto se multiplica la probabilidad de desarrollar la enfermedad. Puede tomar el valor máximo de infinito (∞) en el caso de que ningún no expuesto desarrollara la enfermedad y, en cambio, sí que lo hiciera alguno de los expuestos. El valor 1 representa que el riesgo es el mismo en ambos grupos. Valores inferiores a 1 indicarían un factor protector.

Recuerde



El RR valora la razón de riesgos. Si la relación fuera causal, podría decirse que la exposición multiplica el riesgo RR veces.

Ejemplo 3.17



Banegas et al. (28). En la tabla 3-7 se muestra cómo van aumentando los RR de muerte cardiovascular. Nótese que el grupo «<120/80» se toma como referencia y, por tanto, RR = 1.

Presión arterial (mmHg)	Varones				Mujeres			
	%	Cardiopatía isquémica RR	Enfermedad cerebrovascular (RR)	Todas las causas (RR)	%	Cardiopatía isquémica RR	Enfermedad cerebrovascular (RR)	Todas las causas (RR)
<120/80	20,1	1	1	1	25,6	1	1	1
120-129/80-84	18,1	1,3	1,4	1,2	16	1,4	1,5	1,2
130-139/85-89	17,2	1,6	1,9	1,3	16,9	1,7	1,9	1,3
140-155/90-99	29,1	2,5	2,5	1,6	21,8	2,6	2,7	1,5
160-169/100-109	17,2	3,4	4,4	2,2	10,4	3,5	4,2	1,9
>180/110	3,3	4,5	6,3	3,4	3,3	4,5	6,2	3,4
TOTAL	100				100			

Tabla 3-7 Prevalencias de presión arterial en España y RR de muerte cardiovascular

Odds ratio

Definición



La *odds ratio* es el cociente entre las *odds* en los expuestos y las mismas *odds* en los no expuestos.

Ejemplo 3.12 (Cont.)



En el **ejemplo 3.12** (continuación del 3.9), la razón entre 0,056 y 0,009 es 6,2, lo que indica que la *odd* en los expuestos es 6,2 veces superior a la *odd* en los no expuestos.

Si la enfermedad es poco frecuente, la *odds ratio* es similar al riesgo relativo. En los ejemplos previos, 6,2 y 6, respectivamente. Algunas traducciones proponen llamarla riesgo relativo aproximado. Ambas medidas de asociación se interpretan de forma idéntica.

Recuerde



Riesgo relativo y odds ratio se interpretan igual.

Ejercicio 3.15



Calcule RR, RA y OR en las tablas del ejercicio 3.13. A pesar de que se ha dicho que no tenía sentido calcular los riesgos ni las *odds* en la tercera tabla, haga también en ella todos estos cálculos y observe qué sucede con el valor del *odds ratio*.

Una gran ventaja de la *odds ratio* sobre las medidas basadas en riesgos es que puede ser calculado en cualquier tabla 2×2 , independientemente de cuál haya sido el plan de muestreo.

Nota técnica



Ello es así porque, de la misma forma que se definió la OR como el cociente entre expuestos y no expuestos de las *odds* enfermo/sano $[(a/b)/(c/d)]$, también podría haberse definido como el cociente entre enfermos y sanos de las *odds* expuesto/no expuesto $[(a/c)/(b/d)]$. Dado que ambas definiciones son equivalentes, la *odds ratio* se puede utilizar en cualquier tabla de dos filas y dos columnas, independientemente del plan de muestreo: $(a/b)/(c/d) = (a/c)/(b/d) = ad/bc$

Nota técnica



Mientras que el RA descansa en un modelo aditivo (los riesgos se suman), en el RR subyace un modelo multiplicativo. Por ejemplo, por el hecho de fumar, el riesgo o probabilidad de tener bronquitis puede «sumarse» o «multiplicarse». El RA invita a pensar que ciertos casos siempre tendrán bronquitis (los que aparecen en los no expuestos), mientras que, por el hecho de fumar aparecen casos «nuevos». En cambio, el RR invita a pensar que cada caso tiene una cierta probabilidad de tener bronquitis (la proporción en los no expuestos), mientras que los fumadores tienen otra probabilidad multiplicada por RR.

Recuerde



La odds ratio tiene la gran ventaja de que se puede utilizar en cualquier estudio, independientemente del plan de muestreo. Esta propiedad no la tienen ni el riesgo atribuible ni el relativo.

Ejercicio 3.16



Imagine un diseño casos-contrroles, en los que se escoge una muestra de enfermos (casos) y una muestra de sanos (contrroles), por lo que se deja fija la variable enfermo / sano. ¿Cuáles de las medidas anteriores pueden aplicarse?

Ejercicio 3.17

¿Qué relación existe entre el riesgo relativo y la *odds ratio*?

Número necesario de casos tratados

Definición



El número necesario de pacientes a tratar para evitar un caso (NNT, *number needed to treat/to be treated*) expresa el esfuerzo que hay que realizar para prevenir un caso de enfermedad o la muerte. Se calcula mediante el inverso de la reducción absoluta del riesgo ($1/RA$).

Ejemplo 3.12 (Cont.)



En el **ejemplo 3.12**, si asignamos 1.000 casos a $X+$, cabe esperar que 53 casos desarrollen $Y+$. Si, en cambio, estos mismos 1.000 casos se asignan a $X-$, cabría esperar sólo 9, con un «ahorro» de 44 casos por 1.000 «reassignados». Si se necesitan 1.000 para bajar 44, el número necesario para evitar un caso (NNT) sería 22,7, es decir, 23 casos.

Nota técnica



Esta medida es muy interpretable, pero hay que tener en cuenta que sus propiedades estadísticas son pobres (es muy poco estable de una muestra a otra).

Si en lugar de basarnos en probabilidades o riesgos nos basamos en tasas, estas medidas se definen de forma similar.

Soluciones a los ejercicios

3.1 Dado que $P(B|A)$ debe representar la probabilidad del suceso B dentro del total de casos que cumplen A, se trata de dividir la probabilidad de ser a la vez A y B, es decir $P(A \cap B)$, por la probabilidad de ser A, es decir $P(A)$. En resumen, tiene el mismo numerador que $P(A|B)$, pero cambia el denominador.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Gráficamente sería el cociente entre la zona $A \cap B$ y la zona A

- 3.2 $\text{Sens} = P(+|E)$ Proporción de positivos en el conjunto de enfermos.
 $\text{Esp} = P(-|S)$ Proporción de negativos en el conjunto de sanos.
 $\text{VP+} = P(E|+)$ Proporción de enfermos en el conjunto de positivos.
 $\text{VP-} = P(S|-)$ Proporción de sanos en el conjunto de negativos.

3.3 Una prueba sería extremadamente específica si, aplicada a un conjunto de sanos, casi el 100% dan negativo. Y un resultado tendrá un elevado valor predictivo negativo si casi el 100% de los que dan negativo están realmente sanos.

3.4 La tabla correspondiente es:

	+	-	TOTAL
Enfermo	132		132
Sano		868	868
TOTAL	132	868	1.000

$$\text{VP+} = P(E|+) = 132 / 132 = 1,00 = 100\%$$

$$\text{VP-} = P(S|-) = 868 / 868 = 1,00 = 100\%$$

3.5 Para que $\text{VP+} = 100\%$ se requiere que los casos positivos provengan todos de los enfermos, es decir, que ningún sano dé positivo. Por tanto, se requiere que la especificidad sea del 100% para tener un VP+ del 100%. (Similarmente, para que el VP- sea del 100% se requiere que la sensibilidad sea del 100%.)

3.6 La especificidad y el valor predictivo positivo.

3.7 Compare sus respuestas con las definiciones anteriores del inicio del punto 1.4. Discútalas con un colega y vuelva a comprobarlo con las definiciones anteriores.

$$3.8 \text{ Sens} = P(+|E) = 712/1000 = 0,712 = 71,2\%$$

$$\text{Esp} = P(-|S) = 752/1000 = 0,752 = 75,2\%$$

$$\text{VP+} = P(E|+) = 712/960 \approx 0,742 = 74,2\%$$

$$\text{VP-} = P(S|-) = 752/1040 \approx 0,723 = 72,3\%$$

$$3.9 \quad P(S|+) = \frac{P(+|S) \cdot P(S)}{P(+|S) \cdot P(S) + P(+|E) \cdot P(E)} = \frac{0,248 \cdot 0,868}{0,248 \cdot 0,868 + 0,712 \cdot 0,132} = 0,696$$

3.10 La tabla correspondiente es:

	AA: Y+	NO AA: Y-	TOTAL
EFV	36	15	51
PI	9	40	49
TOTAL	45	55	100

Riesgo en los expuestos a EFV: $P(Y+|EFV) = 36 / 51 \approx 0,70 = 70\%$

Riesgo en los expuestos a PI: $P(Y+|PI) = 9 / 49 \approx 0,18 = 18\%$

3.11 La probabilidad de sacar un «3» vale $1/6 = 0,167 = 16,7\%$. La *odd* respectiva es $1/5 = 0,2 = 20\%$

3.12 *Odd* (dolor lumbar) = $P(\text{dolor lumbar}) / P(\text{no dolor lumbar}) = 0,5 / 0,5 = 1$ (las *odds* de padecer dolor lumbar a lo largo de la vida están «1 a 1»).

Odd (gripe) = $0,8 / 0,2 = 4$ (las *odds* de gripe están 4 a 1: por cada persona que no padecerá gripe, hay 4 que si la tendrán).

Odd (esclerosis múltiple) = $0,001 / 0,999 = 0,001001001 \approx 0,001$ (las *odds* de esclerosis múltiple están 1 por 1.000).

3.13 Los resultados que figuran a continuación muestran que los riesgos y las *odds* de la última tabla no coinciden con los anteriores. ¿Qué ha pasado? Nótese que la variable respuesta (bronquitis) de la tercera tabla ya no depende de las observaciones, pues, por diseño, se ha dejado fija: se ha elaborado una tabla que tiene, porque así lo hemos querido, la mitad de pacientes con bronquitis y la mitad sin bronquitis. Estos datos, obtenidos de dos muestras de la variable respuesta, ya no sirven para calcular ni los riesgos ni las *odds* de desarrollar una bronquitis. Nótese que en la segunda tabla, en la que había una muestra de fumadores y una de no fumadores, sí que podían calcularse los riesgos y las *odds* condicionados a fumador o a no fumador.

	Riesgos			Odds		
	Tabla 1	Tabla 2	Tabla 3	Tabla 1	Tabla 2	Tabla 3
En los expuestos: X+	0,71	0,71	0,85	2,47	2,47	5,53
En los no expuestos: X-	0,25	0,25	0,42	0,33	0,33	0,74

3.14 Para calcular la RV de un resultado negativo procederemos:

$$RV = \frac{P(-|E)}{P(-|S)} = \frac{1 - 0,712}{0,752} = 0,38$$

Con la misma proporción de enfermos tenemos que:

$$\text{Odds a posteriori} = RV \cdot \text{Odds a priori} \rightarrow 0,38 \cdot 1/3 \approx 0,13$$

3.15 Como cabía esperar, la tercera tabla, en la que se había dejado fijo el número de casos con y sin la enfermedad, no permite calcular los riesgos, ni el relativo ($2,87 \neq 2,00$) ni el atribuible ($0,46 \neq 0,42$). La sorpresa es que sí que permite calcular la *odds ratio*, ya que se obtienen los mismos valores (7,50).

	Riesgos			Odds		
	Tabla 1	Tabla 2	Tabla 3	Tabla 1	Tabla 2	Tabla 3
En los expuestos: X+	0,71	0,71	0,85	2,47	2,47	5,53
En los no expuestos: X-	0,25	0,25	0,42	0,33	0,33	0,74
Riesgo atribuible	0,46	0,46	0,42			
Riesgo relativo	2,87	2,87	2,00			
Odds ratio				7,50	7,50	7,50

3.16 Por lo dicho anteriormente, sólo la *odds ratio*. El riesgo relativo y el riesgo atribuible no tienen valor en este tipo de diseños, ya que su valores dependerán del número de controles que se haya decidido seleccionar para cada caso.

3.17 Son estimadores de la relación entre dos variables dicotómicas, una supuesta respuesta y una supuesta causa. Cuando la proporción de la respuesta es muy pequeña, dan valores muy parecidos. Se interpretan igual.

Capítulo 4:

Uso de la distribución normal

Presentación

Hemos visto que, en Ciencias de la Vida, la variabilidad es la norma. Eso implica saber aceptar que ciertas distancias con el valor central son, por definición, «normales», en el sentido de no patológicas. Y por tanto, hay que aprender a distinguir qué distancias, por su magnitud, ya no deben ser aceptadas y pueden ser sospechosas de patológicas. La distribución del Gauss-Laplace, llamada normal, es muy útil para una gran cantidad de variables. En este capítulo, con la ayuda de ejercicios de dificultad progresiva, el lector se habituará al uso de la tabla de la distribución normal y aprenderá a distinguir hasta qué punto una observación puede considerarse, o no, «normal».

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Sabrá manejar las tablas de la distribución normal para convertir valores observados en percentiles.
- Sabrá manejar las tablas de la distribución normal para convertir percentiles en valores observados.
- Sabrá reconocer si una variable puede ser representada por la distribución normal.
- Interpretará un valor fuera de bandas como poco frecuente.
- A partir de pares de valores de sensibilidad y especificidad, dibujará la curva característica (ROC) para evaluar un indicador diagnóstico cuantitativo.

DISTRIBUCIÓN NORMAL

La probabilidad no sólo aparece en variables con dos posibles valores como las estudiadas hasta ahora. A continuación se expone cómo el modelo *normal* de Gauss-Laplace permite representar la distribución de variables cuantitativas.

La distribución normal (fig. 4-1) tiene la conocida forma de campana o montaña, simétrica alrededor de la media (μ) y con la desviación típica (σ) marcando la distancia que separa la media del punto de máxima pendiente o de inflexión de la curva.

Recuerde: μ (mu) y σ (sigma) representan los parámetros **media** y **desviación típica**; σ^2 representa la **varianza**.

Nota técnica



Interpretación física. La media representa el centro de gravedad, es decir, aquel punto que permitirá aguantar en equilibrio, la distribución. La varianza representa la inercia, es decir, la resistencia en hacer girar la distribución alrededor de la media.

Recuerde



En la distribución normal, la media μ (centro) y la desviación típica σ (distancia con la media del punto de máxima pendiente) tienen pleno sentido.

Es sorprendente cómo este modelo matemático consigue reproducir con bastante exactitud la distribución empírica de un buen número de variables biológicas.

Nota técnica



Decir que «una variable biológica sigue la distribución normal» es un abuso de lenguaje. Sería más correcto decir que, utilizando el modelo normal, se consigue reproducir de forma muy notable el comportamiento empírico de dicha variable. La distribución normal es un modelo, no una verdad absoluta.

La distribución normal asume que la variable en estudio es el resultado de la actuación de muchos fenómenos independientes y con igual influencia.

Ejemplo 4.1



En sus inicios, fue utilizada para representar la distribución de los errores de medida. Pero no los errores groseros, pocos y evidentes; sino los muchos, pequeños e inapreciables errores que acompañan a ciertos procedimientos de medida como la balanza de fiel.

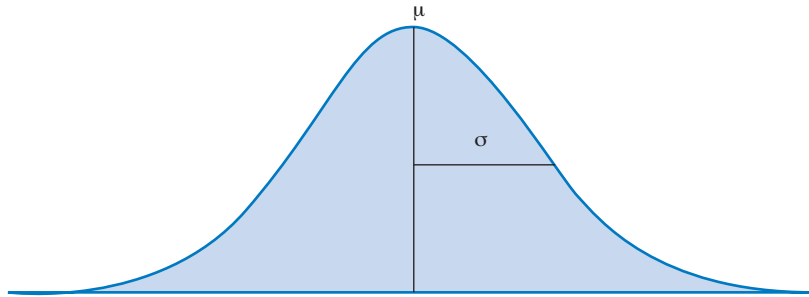


Figura 4-1 Representación de la distribución normal con media μ y desviación estándar σ .

Las leyes de la combinatoria muestran que la probabilidad de que todos estos pequeños fenómenos actúen en el mismo sentido, generando valores extremos, es muy pequeña. En general, estos efectos se compensan unos con otros y los valores se acercan a una cierta media que representa los efectos más «sólidos», de mayor envergadura.

Ejemplo 4.2



La altura de los varones adultos y sanos de una determinada población puede aproximarse, razonablemente bien, por la distribución normal. Para decir que es normal, ha sido preciso primero especificar la edad, el género y la población, ya que éstas son variables que podrían originar diferencias notables, remarcables. Si, por ejemplo, se mezclan ambos géneros, la distribución resultante tendrá dos montañitas o jorobas que definen los intervalos modales de hombres y mujeres.

La dispersión de los valores de la distribución normal es, por tanto, el resultado de establecer un modelo sobre el elevado número de fenómenos con muy pequeña influencia. Éstos son tantos y tan pequeños que no aportan información y representan el «ruido». Su media, en cambio, representa cierta tendencia que puede ser el resultado de otros fenómenos de mayor envergadura.

Recuerde



La media μ de la distribución normal representa la señal «relevante»; y la desviación típica, el ruido «irreproducible».

Ejercicio de Navegación

Entre en la página que se indica a continuación y observe, con la ayuda de la simulación que realiza la aplicación («La máquina de Galton»), que la distribución resultante de dejar caer unas bolas sobre clavos separadores (que las van distribuyendo al azar) es la distribución normal.

<http://www.rand.org/methodology/stat/applets/clt.html>

Recuerde

La notación $N(\mu, \sigma)$ indica que una variable sigue la distribución normal con media μ y desviación típica σ .

Ejemplo 4.3

La altura de los varones adultos sanos es $N(170 \text{ cm}, 8 \text{ cm})$.

Uso de las tablas de la distribución normal

La utilidad de la distribución normal reside en que permite cuantificar la proporción de observaciones que se encuentran a cierta distancia de la media (fig. 4-2). Por ejemplo, si se toma una vez hacia arriba y una vez hacia abajo el valor de la desviación típica, se incluye el 68% de las observaciones. Y si en lugar de hacer una vez el valor de la desviación típica, se toma dos veces dicho valor, se incluye el 95% de las observaciones.

Por ello, es muy útil para construir intervalos en los que cabe esperar que se encuentre un determinado porcentaje de las unidades.

Así, la distribución normal permite establecer una correspondencia entre los valores de una variable y el porcentaje de unidades comprendidas entre estos dos valores. Lo que permite dos usos recíprocos:

1) Dada una probabilidad, buscar un valor: **ir de los porcentajes a los valores**; cierto porcentaje (p. ej., 95%) viene delimitado por ¿qué valores de la variable? (p. ej., 150 y 170 cm)

2) Dado un valor, buscar una probabilidad: **ir de los valores a los porcentajes**; ciertos valores de la variable (p. ej., 150 y 170 cm) ¿qué porcentaje de unidades comprenden? (p. ej., 95%)

Ejemplo 4.4

De la utilidad 1 (fig. 4-3): Podríamos desear conocer la proporción de MIR que sacan más de 240 puntos.

De la utilidad 2 (fig. 4-4): ¿Cuál es límite de la glucemia que deja por encima el 5% de los individuos sanos?

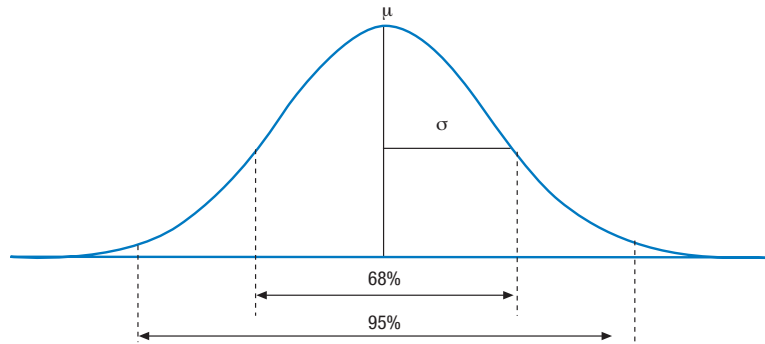


Figura 4-2 Representación de la distribución normal con media μ y desviación estándar σ , con las regiones que contienen el 68 y el 95% de las observaciones.

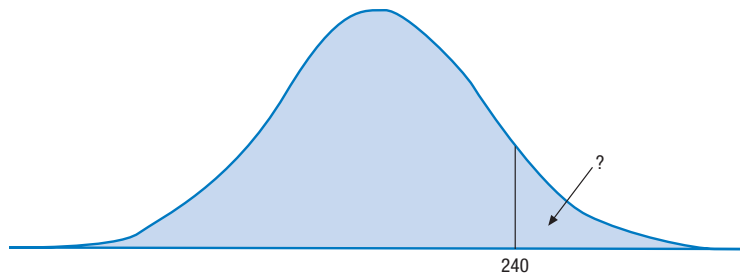


Figura 4-3 ¿Cuál es la proporción de casos que obtiene más de 240 puntos?

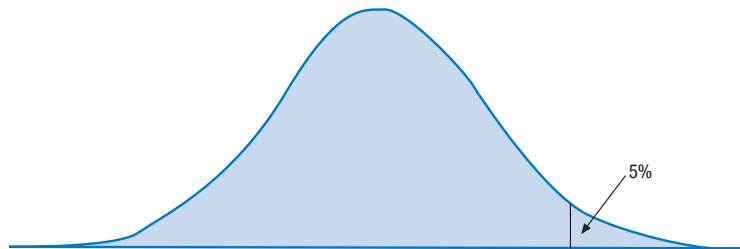


Figura 4-4 ¿Qué valor deja por encima el 5% de la distribución?

Ejemplo 4.5



De la utilidad 1: Un paciente tiene en cierta prueba, índice o escala de medida una puntuación de 112 unidades. Este valor no aporta nada a un inexperto en dicha prueba, pero sí que lo haría decirle que ocupa el percentil 70, es decir, que un 70% de casos tienen puntuaciones inferiores.

De la utilidad 2: Conocidos los valores de la distribución de los individuos sanos de un cierto indicador bioquímico, se pueden calcular los valores de referencia que delimitan el 95% de los individuos sanos.

Para resolver estos ejemplos, los estadísticos matemáticos han realizado los cálculos necesarios y los han puesto en una tabla. Dado que diferentes valores de la media y de la desviación típica resultan en diferentes valores de los intervalos y de los porcentajes, deberían hacerse tantas tablas como posibles combinaciones de valores de la media y de la desviación típica. Para poder usar una única tabla, se puede recurrir al desvío tipificado descrito en el apartado «Descripción de los participantes en un estudio» del capítulo 2.

$$z = \text{desvío tipificado} = \frac{\text{Valor} - \text{Media}}{\text{desviación típica}} = \frac{x - \mu}{\sigma}$$

Recuérdese que esta nueva variable tiene media 0 (valores negativos representarán observaciones por debajo de la media) y desviación típica 1 (una observación prototípica se aleja de la media, por arriba o por debajo, en una unidad). Asimismo, se necesita relacionar los valores de esta variable con los porcentajes. Esto es lo que hace la tabla: proporciona el valor por el que debe ser multiplicada la desviación típica para obtener el porcentaje deseado.

Un dilema es qué porcentajes poner en la tabla: ¿los centrales que quedan dentro?, ¿los que quedan fuera?, ¿por debajo? o ¿por encima? En función del uso que se hará de las tablas convendrá poner unos u otros. En la tabla 4-1, α representa el porcentaje de casos que quedan fuera.

α	0,001	0,01	0,05	0,10	0,20	0,32
$\alpha/2$	0,0005	0,005	0,025	0,05	0,10	0,16
Z	3,29	2,58	1,96	1,64	1,28	1

Tabla 4-1 Valores seleccionados de la distribución normal tipificada (Z)

Ejercicio 4.1



En la distribución normal tipificada, Z, ¿qué proporción de casos quedan por encima de $-1,96$ y por debajo de $+1,96$?

Ejemplo 4.6



Por debajo de $-1,96$ y por encima de $+1,96$ queda un 5% de unidades.

Recuerde



En la normal tipificada, Z, «1,96» (o redondeado: «2») es el valor que contiene el 95% de las observaciones.

Ejercicio 4.2



- a) ¿Entre qué valores de la distribución normal tipificada se encuentran el 99% de las observaciones?
- b) ¿Qué valores contienen el 90%?
- c) ¿Qué valor deja por encima el 5%? ¿Y por debajo?

La tabla 4-2 es más completa que la tabla 4-1 y sirve para encontrar más valores. Para poderlos contener, necesita abarcar más de una fila. El primer valor de la tabla, 0,5000, se corresponde con la fila 0,0_ y la columna _,_0, indicando que, por encima de $z = 0,0_ + _,_0 = 0,00$ se encuentra el 50% de la distribución. En la última columna de la primera fila se comprueba que, por encima de $z = 0,0_ + _,_9 = 0,09$ se encuentra el 46,41% de los casos. El valor siguiente a $z = 0,09$ que muestra la tabla 4-2 es 0,10, en la primera columna de la segunda fila $z = 0,1_ + _,_0 = 0,10$ que deja por encima el 46,02% de los casos (fig. 4-5).

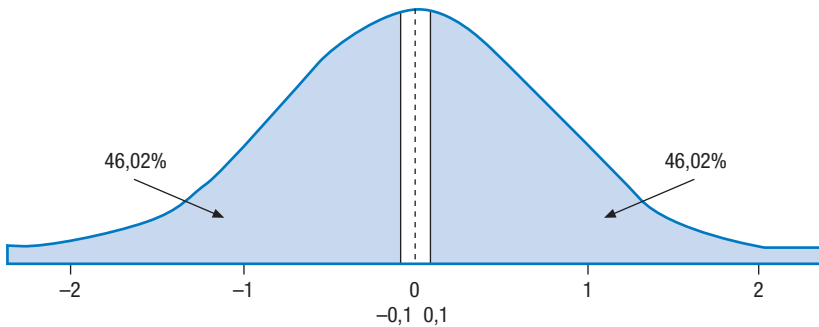


Figura 4-5 En la tabla 4-2 puede verse, en la 1.ª columna y 2.ª fila, que por encima de $Z = 0,10$ queda un 46,02% de los casos.

Comentario

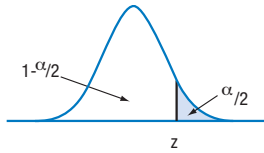


Dado que para hacer corresponder este valor con su porcentaje se necesitarían dos filas larguísimas, por cuestiones de edición, al llegar al 0 se parten las filas, se ponen por debajo y así queda en forma de tabla, más fácil de imprimir en un libro.

Ejemplo 4.7



¿Qué proporción de casos están por encima de $z = 1,66$? Es decir, ¿cuál es la probabilidad de que $Z > 1,66$? Se descompone el número en $1,6 + 0,06$, y se busca en la celda que une la fila del 1,6 y la columna del 0,06: el resultado es 0,0485.



z	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,9	4,8E-05	4,6E-05	4,4E-05	4,2E-05	4,1E-05	3,9E-05	3,7E-05	3,6E-05	3,4E-05	3,3E-05
4,0	3,2E-05	3,0E-05	2,9E-05	2,8E-05	2,7E-05	2,6E-05	2,5E-05	2,4E-05	2,3E-05	2,2E-05
4,5	3,4E-06	3,2E-06	3,1E-06	3,0E-06	2,8E-06	2,7E-06	2,6E-06	2,4E-06	2,3E-06	2,2E-06
5,0	2,9E-07	2,7E-07	2,6E-07	2,5E-07	2,3E-07	2,2E-07	2,1E-07	2,0E-07	1,9E-07	1,8E-07
5,5	1,9E-08	1,8E-08	1,7E-08	1,6E-08	1,5E-08	1,4E-08	1,4E-08	1,3E-08	1,2E-08	1,1E-08
6,0	9,9E-10	9,3E-10	8,8E-10	8,2E-10	7,7E-10	7,3E-10	6,8E-10	6,4E-10	6,0E-10	5,7E-10

Tabla 4-2 Distribución NORMAL estandarizada. Áreas de cola hacia la derecha

En la fila 1,9_ y columna _,_6 puede leerse que, por encima de $Z = 1,9_ + _,_6 = 1,96$ queda el 2,5% de la distribución.

Dada la simetría de la distribución normal, la tabla 4-2 también proporciona los límites por la izquierda. Así, por debajo de $-0,10$ también hay el 46,02% de los casos.

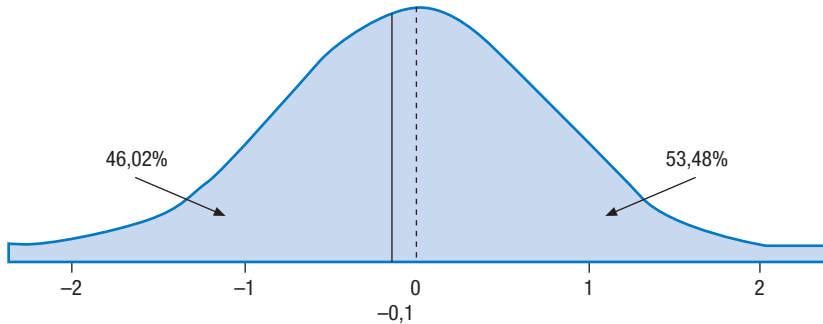


Figura 4-6 Por debajo de 0,1 hay el 46,02% de las observaciones.

Ejemplo 4.7 (Cont.)



¿Qué proporción de casos están por debajo de $-1,38$? Es decir, ¿cuál es la probabilidad de que Z sea menor que $-1,38$? Según la tabla 4-2, a 1,38 (fila 1,3_ y columna _,_8) le corresponde 0,0838.

Asimismo, pueden obtenerse los valores complementarios restando el observado de 100 (fig. 4-6).

Ejemplo 4.8



Así, por encima de $-0,1$, se encuentra el 53,48% de los casos.

$$100 - 46,02 = 53,48$$

Ejercicio 4.3



Compruebe que sabe reproducir con la tabla 4-2 los resultados de los ejercicios 4.1 y 4.2, que se obtuvieron con la tabla 4-1.

Nota técnica



Los textos de estadística suelen expresar la frase «por encima de 0,1 se encuentra el 46,02% de la distribución normal tipificada Z » de manera más compacta, como, por ejemplo: $P(Z > 0,1) = 0,4602$.

Ejemplo 4.9

¿Qué valor Z de la distribución normal deja por encima el 29% de los casos? En la tabla 4-2, las celdas que más se parecen a este valor son 0,2912 y 0,2877. Si la diferencia entre 29 y 29,12% puede despreciarse, la respuesta es que por encima de $Z = 0,55$ se encuentra el 29% de los casos.

Recuerde

Siempre que utilice unas nuevas tablas observe detalladamente el gráfico y el ejemplo para concretar qué valor se representa en la tabla.

Un «truco» que funciona bastante bien es recordar el valor «mágico» que define el 95% central de las observaciones, que es el 1,96 y buscarlo en las tablas para comprobar que se están interpretando correctamente (fig. 4-7). En la tabla puede verse que el valor 1,96 se encuentra en la fila y columna encabezadas por 0,00 y 0,05, la suma de ambos valores es el valor representado por α (proporción de casos que queda por debajo de $-1,96$ y por encima de $+1,96$).

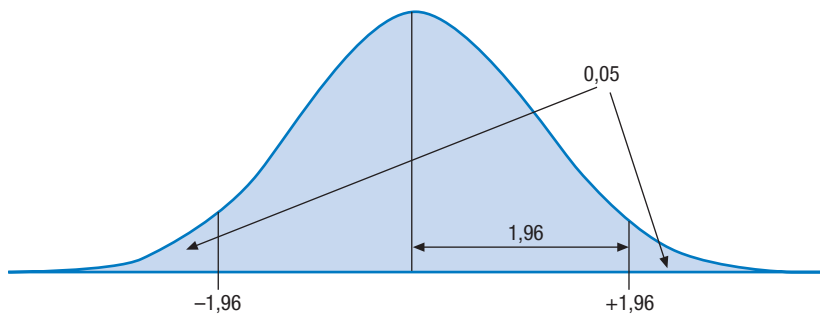


Figura 4-7 $-1,96$ y $+1,96$ dejan por fuera el $0,05 = 5\%$ de los casos.

Ejemplo 4.10

De la utilidad 1: dado un valor, buscar la probabilidad que delimita: la puntuación del examen MIR sigue una DN de media $\mu = 200$ puntos y desviación típica $\sigma = 20$ puntos ¿Qué proporción de casos se sitúan por encima de 240 puntos?

$$Z = \frac{X - \mu}{\sigma} = \frac{240 - 200}{20} = 2$$

Según las tablas, la proporción de alumnos que sacan más de 240 puntos es del 2,28%.

Ejemplo 4.11

De la utilidad 2, dada una probabilidad, buscar el valor que la delimita: La HCM (hemoglobina corpuscular media) en sangre sigue una DN de media $\mu = 30$ y desviación típica $\sigma = 2$. ¿Qué límites de normalidad que incluyan el 95% de los individuos sanos se pueden proponer? En tablas, se vuelve a encontrar el valor $Z = 1,96$, pero antes de usarla hay que darle la vuelta a la fórmula anterior

$$Z = \frac{X - \mu}{\sigma} : \text{se convierte, como } z \text{ puede ser positiva o negativa, en:}$$

$$X = \mu + Z\sigma \quad \text{y} \quad X = \mu - Z\sigma$$

$$\text{por lo que: } X = \mu + Z\sigma = 30 + 1,96 \cdot 2 = 33,92 \approx 34$$

$$X = \mu - Z\sigma = 30 - 1,96 \cdot 2 = 26,08 \approx 26$$

y los límites serán 26 y 34.

Ejercicio 4.4

Cierto estimulador tiene un umbral que varía de un voluntario sano a otro. Su distribución es aproximadamente normal con una media de 5 voltios y una desviación típica de 0,5.

- El 95% de los voluntarios tienen un umbral que se sitúa entre ___ y ___ voltios.
- En el 95% de los voluntarios, el umbral se sitúa por encima de ___ voltios.
- En el 95% de los voluntarios, el umbral se sitúa por debajo de ___ voltios.
- El 90% de los voluntarios tienen un umbral que se sitúa entre _ y _ voltios.
- En el 84% de los voluntarios, el umbral se sitúa por encima de ___ voltios.
- En el 84% de los voluntarios, el umbral se sitúa por debajo de ___ voltios.
- ¿Cuál es la probabilidad de que el umbral supere 6,3 voltios?
- ¿Cuál es la probabilidad de que un voluntario tenga un umbral entre 4,5 y 5,5?

Uso de la distribución normal con aplicaciones informáticas

Por cuestiones de espacio, la tabla 4-2 sólo incluye algunos valores seleccionados, lo que puede ocasionar pequeños errores de aproximación. Muchas aplicaciones informáticas permiten obtener muchos más valores. Una hoja de cálculo, Excel por ejemplo, permite obtener directamente la proporción de casos por debajo de un cierto valor. Por ejemplo, la función:

= DISTR. NORM. ESTAND. (z)

devuelve la proporción de casos por debajo de z en el caso de una distribución normal tipificada. Así, si introducimos $Z = 1,96$, dará 0,9750021.

A su vez, la función:

= DISTR. NORM. ESTAND. INV. (probabilidad)

hace la función inversa: devuelve el valor de Z a partir de la probabilidad. Si se le introduce probabilidad = 0,975, proporciona $Z = 1,95996398$.

La función:

= DISTR. NORM. (x; media; desviación estándar; 1)

devuelve la proporción de casos por debajo de X en el caso de una distribución normal con la media y la desviación típica especificadas. Así, si introducimos $X = 1,96$, media = 0 y desviación estándar = 1 proporciona el valor anterior de 0,9750021.

Finalmente:

= DISTR. NORM. INV. (probabilidad; media; desviación estándar)

devuelve la inversa de la función anterior. Si se introduce probabilidad = 0,975, media = 0; desviación estándar = 1, proporciona $x = 1,95996398$.

Ejercicio 4.5



En unidades del Sistema Internacional, el cloruro plasmático tiene unos límites de «normalidad» de 95 y 105 mmol/l.

- ¿Es posible que una persona sana supere estos límites?
- ¿Cuál cree usted que es el valor de la media y de la desviación típica de esta variable en los «normales»?
- ¿Existe alguna condición (premisa) para este cálculo?
- Para la ferritina, estos límites son 15-200 mg/l ¿Cómo se imagina su distribución?

Ejercicio 4.6*

Busque variables relacionadas con su trabajo que presumiblemente sigan una distribución normal.

Ejercicio 4.7*

Invente aplicaciones «útiles» para las variables del punto anterior. Invente condiciones o situaciones en las que sea razonable que las variables del ejercicio anterior dejen de seguir una distribución normal.

*No incluye solución al final del capítulo.

Aplicaciones al diagnóstico

Hasta este momento se ha hecho hincapié en pruebas diagnósticas llamadas cualitativas, es decir, aquellas que sólo admiten dos posibles resultados: positivo y negativo. En este caso, la definición de sensibilidad y especificidad, así como sus complementarias, es inmediata y unívoca.

En los test donde hay varios resultados numéricos posibles, llamados cuantitativos, la definición de los términos anteriores no es inmediata sino convencional, ya que dicha definición requiere establecer un límite o umbral que separe el conjunto de resultados en dos grupos, positivo y negativo.

Ejemplo 4.12



Por ejemplo, los resultados de un test que mide la concentración de glucosa en plasma, en condiciones basales. Dichos resultados, expresados en mg/dl, pueden ser muy variados: 50, 75, 110, 128, 165, 192, etc. Ninguna cifra de éstas es, por sí misma, positiva ni negativa. Ahora bien, si en virtud de conocimientos fisiológicos y epidemiológicos, entre otros, se considera que las cifras inferiores a 100 definen un resultado negativo, y las superiores, positivo, entonces la situación se ha hecho similar a los test cualitativos.

Recuerde



Si el resultado del indicador diagnóstico es un número, es necesario establecer un punto de corte.

Las curvas ROC (Receiver Operating Characteristic curves)

Ejemplo 4.13



El ejercicio 4.4 dice que el umbral de estimulación de los voluntarios sanos sigue una $N(5, 0,5)$. Supongamos, además, que en cierto tipo de enfermos sigue una $N(6, 0,5)$ (fig. 4-8).

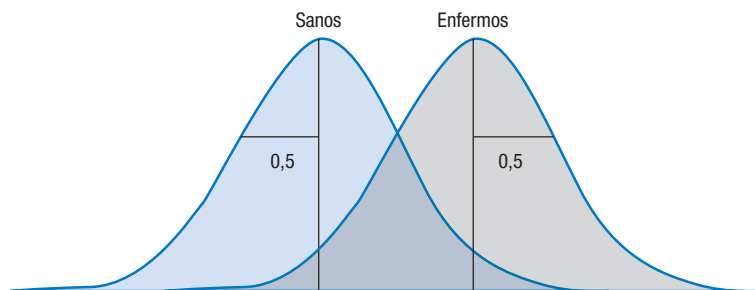


Figura 4-8 Distribución del umbral de estimulación en sanos y enfermos.

Ejemplo 4.13 (Cont.)

Si el criterio diagnóstico se establece en 5,5 (fig. 4-9), los valores de sensibilidad y especificidad serán:

$$\text{sens} = P(+|E) = P(y > 5,5 \mid \text{Enfermo}) = P\left(z > \frac{5,5 - 6}{0,5}\right) \\ = P(z > -1) \approx 84,13\%$$

$$\text{esp} = P(-|S) = P(y < 5,5 \mid \text{Sano}) = P\left(z < \frac{5,5 - 5}{0,5}\right) \\ = P(z < 1) \approx 84,13\%$$

En cambio, si el criterio se hubiera establecido en 5,2 (fig. 4-10), serían:

$$\text{sens} = P(+|E) = P(y > 5,2 \mid \text{Enfermo}) = P\left(z > \frac{5,2 - 6}{0,5}\right) \\ = P(z > -1,6) \approx 94,52\%$$

$$\text{esp} = P(-|S) = P(y < 5,2 \mid \text{Sano}) = P\left(z < \frac{5,2 - 5}{0,5}\right) \\ = P(z < 0,4) \approx 65,54\%$$

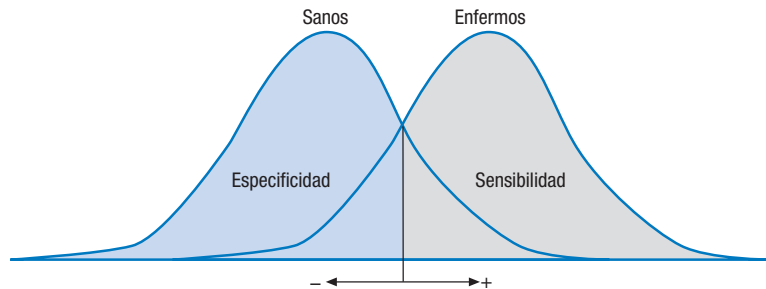


Figura 4-9 Sensibilidad es la proporción de la curva de enfermos que queda por encima del criterio diagnóstico y especificidad la de sanos que queda por debajo.

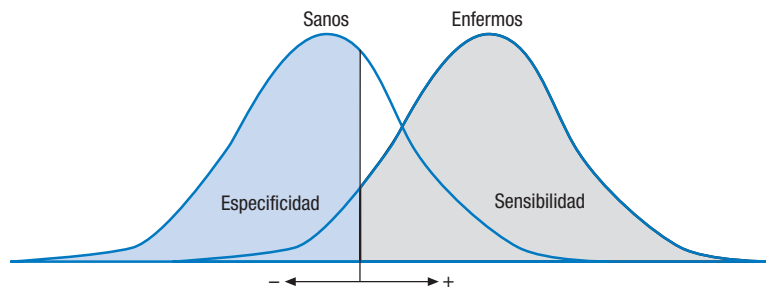


Figura 4-10 Al desplazar el criterio diagnóstico hacia la izquierda aumenta la sensibilidad y disminuye la especificidad.

Moviendo el punto de corte se cambian los valores de especificidad y sensibilidad. Si se desea aumentar la sensibilidad, la especificidad disminuye. Y viceversa. Nótese que habrá tantos «pares» de valores de sensibilidad y especificidad como posibles puntos de corte. Cada indicador diagnóstico tiene unos pares de valores de sensibilidad y especificidad que le «caracterizan».

Ejemplo 4.14



Everitt (29). «Examine los siguientes valores de 1 (claramente sano) a 5 (claramente enfermo) para 50 sujetos sanos y 50 enfermos [tabla 4-3]. Si el valor 5 se usa como punto de corte para identificar casos de enfermedad, entonces la sensibilidad se calcula como $8/50 = 0,16$; y la especificidad, como $49/50 = 0,98$. En cambio, usar el valor 4 como punto de corte supone una sensibilidad de $27/50 = 0,54$ y una especificidad de $41/50 = 0,82$.»

Valor	1	2	3	4	5	TOTAL
Sanos	4	17	20	8	1	50
Enfermos	3	3	17	19	8	50

Tabla 4-3 Número de casos, sanos y enfermos, con cada valor de la escala

Ejercicio 4.8



Calcule la sensibilidad y la especificidad si el punto de corte fuera 3, 2 o 1.

Definición



La curva característica (ROC: Receiver Operating Characteristic Curve) contiene todos los posibles pares de sensibilidad y especificidad de un indicador diagnóstico.

Nota técnica



El término «*receiver operating characteristic*» proviene de las telecomunicaciones y analiza la capacidad de un receptor de señales para discriminarlas correctamente. Una posible traducción sería curva característica de la operatividad del receptor.

Lectura



Everitt (29). «Curvas características (curvas ROC): grafico de la sensibilidad de un test de diagnóstico frente al complementario de la especificidad según varía el punto de corte que indica que un test es positivo. A menudo se usa para elegir entre varios test en competencia, aunque el procedimiento no tiene en cuenta la prevalencia de la enfermedad que se estudia.»

Nota técnica



El área bajo la curva ROC se interpreta como la probabilidad de, seleccionados al azar 1 sano y 1 enfermo, que el primero tenga valores menos patológicos que el segundo.

Ejemplo 4.15



El doctor Manuel Callis, en su tesis doctoral, predice el grado de afectación abdominal en enfermos de Hodking sin necesidad de recurrir a la cirugía (laparoscopia). Para ello, ha obtenido la siguiente regresión logística:

$$\log(p/(1-p)) = 0,85 + 0,04 \cdot X_1 - 13,9 \cdot X_2 + 1,14 \cdot X_3 + 1,93 \cdot X_4 + 1,53 \cdot X_5$$

en la que las diferentes X indican variables que pueden ser obtenidas en una visita clínica habitual (análisis, radiografías, etc.). La tabla 4-4 indica los valores de sensibilidad y especificidad que se obtienen escogiendo diferentes puntos de corte mediante esta regresión logística (p. ej., escogiendo como punto de corte una predicción de afectación abdominal de 0,10, se obtiene una sensibilidad de 0,95 y una especificidad de 0,22).

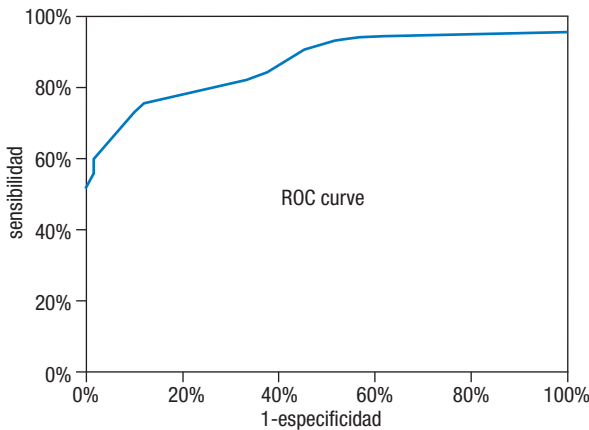
Utilizando Excel, se genera el gráfico de la curva ROC (fig. 4-11) con las instrucciones del cuadro 4-1.

Punto de corte	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
Sensibilidad	0,95	0,93	0,90	0,83	0,76	0,73	0,59	0,59	0,49
Especificidad	0,22	0,49	0,65	0,86	0,86	0,90	0,98	0,98	1,00

Tabla 4-4 Pares de sensibilidad y especificidad del ejemplo 4-15

En la barra de herramientas seleccionar ->Insertar->Gráfico
 Seleccionar la opción XY (Dispersión)
 Elegir el gráfico
 Dispersión con líneas suavizadas y sin marcadores de datos
 Seleccionar la opción para introducir una serie
 En *Agregar* introducimos los datos correspondientes al gráfico:
Nombre: nombre del gráfico, ROC curve
Valores eje X: Seleccionar el rango de valores (sensibilidad) de la hoja de Excel
Valores eje Y: Seleccionar el rango de valores (especificidad) de la hoja de Excel (especificidad) de la hoja de Excel
 Seguir avanzando con *Siguiente* hasta finalizar la construcción del gráfico.

Cuadro 4-1 Instrucciones de Excel para generar una curva ROC.



Especificidad	Sensibilidad
22%	95%
49%	93%
65%	83%
86%	76%
90%	73%
98%	59%
100%	49%

Figura 4-11 Curva ROC.

© Elsevier. Es una publicación MASSON. Fotocopiar sin autorización es un delito.

Nota técnica



Este clínico le pide a usted asesoría para escoger el punto de corte más adecuado para decidir si existe o no existe afectación abdominal, ¿qué nuevo análisis le recomendaría Vd. para responder a esta pregunta? Debería realizarse un análisis de decisión y considerar la matriz de beneficios/pérdidas (¿qué «cuesta» un enfermo al que se declara sano?, ¿y un sano al que se le declara enfermo? ¿Qué «vale» un sano al que se le declara como tal? ¿Y un enfermo?). El gran reto de la economía de la salud es precisamente convertir en una misma «moneda» las consecuencias de no detectar a un enfermo y de tratar a un sano. Una vez establecidos estos «valores» para una cierta población, es fácil, usando sensibilidad y especificidad, decidir la estrategia «óptima».

Soluciones a los ejercicios

4.1 Dada la simetría de la distribución normal, la proporción de casos por encima de $-1,96$ y la proporción de casos por debajo de $1,96$ es la misma. Tal y como puede verse en la tabla 4-1, el valor $1,96$ deja fuera el 5% de los casos, así que, por encima de $-1,96$ se encuentran el 97,5% de los casos, así como por debajo de $1,96$.

4.2 a) Para acotar el 99% de las observaciones, debemos mirar la fila $\alpha/2$ de la tabla 4.1, ya que en este caso repartimos la probabilidad del 1% entre las dos colas dada la simetría de la distribución. Marcando el límite en $0,005 = 0,5\%$, se obtiene que el 99% de las observaciones se encuentran entre los valores $-2,58$ y $2,58$.

b) Análogamente, marcando el límite en $0,05 = 5\%$, se obtiene que el 90% de las observaciones se encuentran entre los valores $-1,64$ y $1,64$.

c) En esta ocasión, debemos fijarnos en la fila $\alpha/2$ de la tabla 4-1 para descubrir qué valor deja por encima el 5% y qué valor deja por debajo el 5%; en el segundo caso, la tabla 4-1 nos muestra que el valor es $1,64$, así que para el primer caso será $-1,64$.

4.3 La fila que empieza en «1,9_» se une a la columna encabezada por «_,_6» en el valor $0,025$, lo que significa que por encima de $1,96$ hay el 2,5% de la distribución. La fila que empieza en «2,5_» y la columna «_,_8» se unen en $0,0049$, aproximadamente 0,5%.

La fila que empieza en «1,6_» y la columna «_,_4» proporcionan $0,0505$, aproximadamente 5%.

4.4 a) En las tablas de la distribución normal se encuentra que el valor $1,96$ deja fuera el 5% y delimita el 95% de los casos. Por tanto, se debe multiplicar la desviación típica por este número y el resultado sumarlo y restarlo de la media:

$$\text{Valores} = \text{media} \pm 1,96 \text{ desviación típica} = 5 \pm 1,96 \cdot 0,5 \approx 5 \pm 1 = [4, 6]$$

b) Ahora se trata de dejar en un extremo el 5%, lo que equivale a dejar en dos extremos iguales al 10% y a contener el 90%. El valor encontrado en las tablas es $1,645$.

$$\text{Valor} = \text{media} - 1,64 \cdot \text{desviación típica} = 5 - 1,64 \cdot 0,5 = 5 - 0,82 = 4,18$$

$$\text{c) Valor} = \text{media} + 1,64 \cdot \text{desviación típica} = 5 + 1,64 \cdot 0,5 = 5 + 0,82 = 5,82$$

$$\text{d) Valores} = \text{media} \pm 1,64 \cdot \text{desviación típica} = 5 \pm 1,64 \cdot 0,5 = 5 \pm 0,82 = [4,18, 5,82]$$

e) Ahora se trata de dejar en un extremo el 16%, lo que equivale a dejar en dos extremos iguales al 32% y a contener el 68%. El valor encontrado en las tablas es $0,99$.

$$\text{Valor} = \text{media} - 0,99 \text{ desviación típica} = 5 - 0,99 \cdot 0,5 \approx 5 - 0,5 = 4,5$$

f) y, por simetría,

$$\text{Valor} = \text{media} + 0,99 \text{ desviación típica} = 5 + 0,99 \cdot 0,5 \approx 5 + 0,5 = 5,5$$

g) Ahora el problema se resuelve en el orden inverso; primero se tipifica la variable:

$$z = \text{desvío tipificado} = \frac{\text{valor} - \text{media}}{\text{desviación típica}} = \frac{6,3 - 5}{0,5} = 2,6$$

En la tabla 4-2, el valor $2,6$ deja por encima $0,0047 \approx 0,005 = 0,5\%$. Por tanto, la probabilidad de que un caso supere la cifra $6,3$ es menor del 0,5%.

h) Calculemos primero las probabilidades de no alcanzar $5,5$ y $4,5$ voltios y luego las restaremos entre ellas.

$$P(X > 5,5) = P\left(Z > \frac{5,5 - 5}{0,5}\right) = P(K > 1) = 0,1587$$

$$P(X < 5,5) = 1 - P(X > 5,5) = 1 - 0,1587 = 0,8413$$

$$P(X < 4,5) = P(X > 5,5) = 0,1587 \text{ (por simetría)}$$

$$P(4,5 < X < 5,5) = P(X < 5,5) - P(X < 4,5) = 0,8413 - 0,1587 = 0,6826$$

Por lo tanto, tenemos una probabilidad del 68% de que el umbral se sitúe entre los 4,5 y los 5,5 voltios.

4.5 a) Convendría estudiar cómo se han definido estos límites. Dado que (con pequeña probabilidad) puede haber personas sanas que tengan valores muy alejados, suelen definirse estos límites de forma que incluyan el 95% de los sanos. Por tanto, es posible que una persona sana supere estos límites.

b) A partir de estas cifras, si se asume la forma de montaña simétrica de la normal, la media sería el punto central, 100, y la desviación típica, la mitad de la distancia de los extremos, 2,5.

c) Que la variable siga la distribución normal.

d) Parece difícil imaginar una distribución simétrica para la ferritina. El cálculo anterior no sería correcto. A veces, transformar logarítmicamente estas variables positivas permite descubrir detrás una forma de ¡montaña simétrica!

4.8 Si se define como positivo con valores iguales o superiores a 3, habrá $17 + 19 + 8 = 44$ enfermos que dan positivo y $1 + 8 + 20 = 29$ sanos. Por tanto, la sensibilidad será de $44/50 = 88\%$ y la especificidad será $(50 - 29)/50 = 42\%$.

Para el valor 2 serán $\text{sens} = 47/50 = 94\%$ y $\text{esp} = 4/50 = 8\%$.

Para el valor 1, dado que todos los casos se declaran como positivos, la sensibilidad será del 100% y la especificidad del 0%.

Capítulo 5:

Principios de inferencia

Introducción

¿Qué información proporciona, a un clínico de Barcelona, los resultados obtenidos en un estudio previo realizado en Boston? La evolución de estos casos de Boston se puede conocer perfectamente, sin error. Pero esos casos ya han «evolucionado», no tiene interés predecir una evolución que ya ha sucedido. En cambio, sería muy interesante poder aplicar estos resultados pasados a unos nuevos casos. ¿Cómo hacerlo?

La inferencia estadística, para incorporar al conocimiento teórico la información empírica, define los conceptos de muestra y población. Los valores obtenidos en las muestras permitirán estimar, con un cierto error cuantificable, los parámetros que caracterizan al conjunto de la población. En resumen, la estadística permite cuantificar, tanto la información (o «señal») aportada por los datos, como el error aleatorio (o «ruido») que implica el proceso de generalización.

En este capítulo se exponen los conceptos fundamentales de inferencia estadística. Se introduce la oscilación aleatoria de los valores obtenidos en muestras y cómo la estadística cuantifica esta oscilación. Para ello, se trabaja con el indicador estadístico más usual: el promedio o media muestral. Se estudia cómo oscila, cuánto oscila, alrededor de qué valor oscila y qué forma adopta esta oscilación.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Delante de una variable, se planteará la necesidad de resumir y representar los datos.
- Delante de un estudio, identificará la unidad o caso.
- Delante de un estudio, se preguntará por los posibles errores de inferencia.
- Delante de un estudio, se preguntará por la población a la que pueden hacer referencia sus resultados.
- Delante de un estudio, distinguirá entre estadístico, estimador y parámetro.
- Sabrá que la muestra aleatoria permite disponer de estimadores insesgados.
- Conocerá la distribución del estimador media.
- Distinguirá entre desviación típica y error típico.
- Interpretará el error típico como el error esperado al decir que el parámetro toma el valor observado en la muestra.
- Conocerá cómo el incremento del tamaño muestral reduce la incertidumbre de la estimación.

Introducción a la inferencia estadística

¿Qué es la inferencia estadística?

Si, por ejemplo, se desea estimar el tiempo de crecimiento de un cierto tejido, se pueden utilizar dos procedimientos. El primero, teórico, consiste en *deducir* este tiempo a partir de los tiempos de división de sus células. El segundo procedimiento, empírico, consiste en *inducirlo* a partir de un número limitado de casos. Ahora bien, ¿hasta qué punto unas pocas pruebas permiten establecer leyes generales sobre el crecimiento de estos tejidos? O mejor, ¿cuánta información están realmente aportando? En los puntos sucesivos se estudia cómo la inferencia estadística formaliza este proceso, lo que requiere: *a)* definir; *b)* cuantificar, y *c)* acotar los riesgos que conlleva.

Definición



La **inferencia** pretende generalizar la información contenida en unos datos a un cuerpo de conocimiento más amplio.

Historieta



Dos amigos caminan por el Pirineo y, al ver un caballo, uno de ellos comenta: «No sabía que los caballos de La Cerdaña fueran marrones y con las patas anchas». Su amigo, que es lógico, le responde: «Perdona, lo que no sabías es que en La Cerdaña hay, por lo menos, un caballo marrón de patas anchas».

Lectura



Hasta hace relativamente poco tiempo, los filósofos se han quejado de la falta de herramientas técnicas que permitan este salto de las partes al todo (Chalmers, 30; Bunge, 31). Para Hume, la inferencia era simplemente imposible, y para Rusell, la inducción seguía siendo un problema de lógica no resuelto. A mediados del siglo pasado, Popper aportó un punto de vista algo más optimista: «Sólo la refutación de una teoría puede ser inferida a partir de datos empíricos y esta inferencia es puramente deductiva». Hoy en día, ya está plenamente aceptado que, en estudios bien diseñados, y ejecutados, la metodología estadística hace posible la inferencia, siempre que se acepten ciertos riesgos.

Respuestas que ofrece la inferencia estadística

Veamos algunas preguntas que pueden ser contestadas con la ayuda de la metodología estadística. El ejemplo más sencillo estudiaría la distribución de una sola variable: ¿cuál es el valor de monóxido de carbono en el aire espirado por fumadores jóvenes? O bien, ¿cuál es la distribución de los valores de homocisteína plasmática en

pacientes con lupus eritematoso? Nótese que si no existiera variabilidad —si la cantidad de monóxido de carbono espirado fuera siempre la misma—, la inferencia sería inmediata: con una observación sería suficiente para conocer el comportamiento de todas ellas. La metodología estadística considera y cuantifica la diversidad entre las unidades.

Ejemplo 5.1



No es necesario hacer un estudio estadístico para conocer la distribución de la variable «número de cerebros» que tiene cada uno de los habitantes de una ciudad. Cada ciudadano tiene un cerebro y sólo uno: así de fácil.

En cambio, sería terriblemente aburrido «decir toda la verdad» sobre la altura de una muestra de 23 pacientes: el primer caso mide 164 cm, el segundo, 173 cm; el tercero 168; ...y el vigésimo tercero, 192.

Ejercicio 5.1



Suponga que, en el ejemplo anterior, por no aburrir, decide hacer un resumen de los datos, ¿qué información le gustaría que este resumen le proporcionara?

Ejercicio 5.2

Proponga otro ejemplo en el que también sea conveniente hacer un resumen estadístico de los datos.

Conocer la distribución de una variable permitirá al equipo clínico realizar de forma científica el diagnóstico, el tratamiento y/o pronóstico de un paciente.

Ejemplo 5.2



Si se conoce cuál es la distribución de la variable tiempo de convalecencia tras cierta enfermedad, el médico de cabecera puede «adelantar» al paciente cuántos días tendrá sus facultades mermadas. Con la media, le dirá al paciente cuál es su valor esperado. Y con la desviación típica, cuál es el error esperado: cuánto cabe esperar que un paciente típico se aleje de esa media.

Población, muestra e individuo

El reto de la inferencia estadística consiste en generalizar un fenómeno observado en unos cuantos casos a todas las observaciones posibles. Para ello, las primeras definiciones que se deben establecer son las de población, muestra y unidad.

Definición

Población: Conjunto de todos los elementos, que cumplen ciertas propiedades comunes, entre los que se desea estudiar un determinado fenómeno.

Muestra: Subconjunto de la población que es estudiado y a partir del cual se sacan conclusiones sobre las características de la población.

Unidad (individuo o caso): Es cada uno de los elementos que componen la muestra y la población.

Población, muestra y unidad se contienen progresivamente, a la manera de las muñecas rusas. La población contiene la muestra y la muestra, a su vez, las unidades. La diferencia es que, conceptualmente, hay un número ilimitado de muestras y de individuos. La población, sin embargo, es única, y representa al conjunto que deseamos conocer.

Ejemplo 5.3

Costa et al. (32). Se invitó a participar en el estudio, de manera consecutiva, a todas las personas que acudieron al Centro de Extracciones del Hospital Clínic i Universitari de Barcelona, desde diferentes servicios, para la realización de una prueba de tolerancia oral a la glucosa (PTOG).

Cilla et al. (33). El grupo de población considerado en el estudio incluyó a mujeres que tuvieron un primer parto después de septiembre de 1989 y un segundo parto entre 2 y 8 años después en la Maternidad del Hospital Nuestra Señora de Aránzazu de San Sebastián (Guipúzcoa).

Por su parte, las unidades no tienen por qué ser «individuos». Pueden ser hospitales, comarcas o visitas clínicas. Es muy importante definir con sumo cuidado estas unidades, ya que se podría llegar a conclusiones diferentes.

Ejemplo 5.4

Cierto facultativo presume de tener un razonable promedio de 7 pacientes por hora. Pero la asociación de usuarios ha preguntado a todos sus pacientes y ha obtenido un promedio de 9. ¡Y pudiera ser que todos digan la verdad, sin trampa!

Pongamos que este profesional tiene 3 horas de visita. En una de ellas ve a las primeras visitas, a razón de 3 por hora. En otra, recibe a las segundas visitas, 6 por hora. Y en la restante hora recibe las demás visitas, 12 por hora. Este facultativo ha definido como unidad del estudio la «hora de visita»: el promedio de 3, 6 y 12 es, efectivamente, 7 pacientes por hora.

Ejemplo 5.4 (Cont.)

$$\sum_{i=1..3} \frac{X_i}{n} = \frac{3+6+12}{3} = 7$$

donde $\sum_{i=1..3}$ indica la suma de los casos 1 a 3

Los usuarios, en cambio, han definido como unidad a cada uno de ellos, de forma que la respuesta «3» la obtienen 3 veces, la respuesta «6», 6 veces y «12», 12 veces. Y el promedio es, efectivamente, 9 pacientes:

$$\sum_{i=1..21} \frac{X_i}{n} = \frac{3+3+3+6+6+6+6+6+6+6+12+\dots+12}{21} = 9$$

Posiblemente, la primera definición represente mejor la pregunta del clínico (¿qué promedio de pacientes visito yo por hora?) y la segunda la del usuario (¿cuánto suele durar mi visita?).

Ambas definiciones son correctas y válidas. Pero no son intercambiables y en cada estudio debe estar muy clara cuál es la unidad que se ha definido. Así, diferentes objetivos requieren diferentes cálculos, todos ellos lícitos y correctos, pero que no deben confundirse: siempre debe quedar bien clara la unidad del estudio.

Ejercicio 5.3

El colegio de odontólogos ha realizado un estudio aleatorio entre los pacientes de sus consultas en la semana anterior. De 1.000 fichas analizadas, 500 habían tenido una visita el año anterior, por lo que concluyen que un 50% de la población acude al dentista cada año. ¿Qué opina? ¿Se puede conocer la frecuencia de visitas al odontólogo en la población general a partir de una muestra obtenida en las consultas?

Ejercicio 5.4

Los centros sanitarios de la Seguridad Social suelen realizar una encuesta de satisfacción a sus usuarios, cuyos resultados suelen ser altamente positivos. ¿Qué le lleva al defensor del pueblo a realizar una encuesta en la población general? (Pista: defina la unidad de ambos estudios y medite sus diferencias.)

Ejercicio 5.5

Para estimar la infección nosocomial, puede hacerse un estudio seleccionando algunos de los pacientes que ingresan o bien seleccionando algunas de las camas ocupadas en el hospital. Asumiendo que los pacientes que están ingresados más tiempo son los que tienen mayor probabilidad de desarrollar esta infección, ¿cuál de los dos estudios dará cifras más altas de infección nosocomial?

Recuerde

Estudie siempre con mucho cuidado cómo se definen las unidades. Dos estudios, para poder ser comparados, requieren la misma definición de las unidades.

Estadísticos, estimadores y parámetros

En el estudio de la información disponible, la inferencia estadística afronta el reto de abarcar un «universo» más amplio que los «pocos casos» disponibles.

Definición

Los indicadores que se calculan en las muestras reciben el nombre de **estadísticos**.

Los indicadores de la población, que estamos interesados en conocer, reciben el nombre de **parámetros**.

Recuerde

Un parámetro hace referencia a un valor de la población, mientras que un estadístico lo hace de la muestra.

Por ejemplo, la media puede representar al «parámetro» media cuando hablamos del centro de gravedad (o **esperanza**) de una distribución poblacional, o al «estadístico» media cuando nos referimos al **promedio** de una serie de valores calculado en una muestra.

Ejemplo 5.5

Supóngase que la probabilidad de que un paciente con anticuerpos del sida tarde, en ciertas condiciones, más de 2 años en desarrollar los primeros síntomas es 0,50. Es decir, expresado en porcentajes, del 50%. Esta probabilidad es un parámetro que resume las expectativas del paciente y que representa una característica intrínseca de la enfermedad. Por otro lado, se han estudiado, en esas condiciones,

Ejemplo 5.5 (Cont.)

una muestra de 25 pacientes de esa población y 15 de ellos han superado los 2 años. Este resultado de $15/25 = 0,60$ (60%) representa la proporción, que es el estadístico o valor observado en la muestra.

Ejercicio 5.6

Proponer un ejemplo similar con la media en lugar de la probabilidad.

El reto de la inferencia estadística es conocer los parámetros, que caracterizan al todo de la población, a partir de los estadísticos, obtenidos en una muestra.

Definición

Cuando un estadístico de una muestra se usa para conocer el valor de un parámetro de la población recibe el nombre de **estimador**.

Nota técnica

Si usted dispone de los datos de toda la población, es decir, si las conclusiones de su estudio se aplicarán únicamente a estos casos y no desea poderlas aplicar a otros datos diferentes, usted no necesita saber qué es la inferencia estadística. Pero tenga cuidado al hablar: no podrá establecer ninguna ley «universal» que vaya más allá de sus propios datos.

Cada muestra es fugaz, en el sentido de ser irrepitable y, en el fondo, irrelevante en sí misma. Una vez terminado el seguimiento de los pacientes de la muestra y cumplidas las responsabilidades sanitarias con ellos, el interés científico se centrará en conocer qué dicen estos casos sobre los pacientes futuros.

Ejemplo 5.6

Las encuestas electorales, a partir de unos pocos miles de entrevistados, intentan conocer la tendencia de unos cuantos millones: el auténtico interés está en lo que votará toda la población. La importancia que tienen los pocos entrevistados es su capacidad para informar sobre la distribución poblacional de esta variable.

Recuerde

Se puede acceder al estadístico observado en la muestra, pero el auténtico objetivo es el parámetro de la población.

Definición

La inferencia estadística es el proceso formal de analizar y cuantificar la información empírica («**evidencia**» o pruebas) que el estimador proporciona del parámetro.

Es tan importante distinguir si se trata de valores muestrales o poblacionales que se les dará diferente símbolo en un caso o en otro. Incluso, la media puede recibir el nombre de esperanza cuando se trata del parámetro poblacional y de promedio cuando es el valor obtenido en la muestra. De la misma forma, el estimador muestral de la probabilidad recibe el nombre de proporción (tabla 5-1).

	Parámetro (θ) (Población)	Estadístico ($\hat{\theta}$) (Muestra)
Media	$\mu = E(X)$ esperanza	\bar{X} o m promedio
Desviación típica	σ	S
Probabilidad	π probabilidad	p proporción

Tabla 5-1 Símbolos utilizados para distinguir parámetros y estadísticos

Así, una proporción observada en una muestra informa sobre la probabilidad de la población. Pero ¿cuánta información aporta? ¿Se debe «creer» mucho o poco que el valor poblacional se acerca al valor del estadístico observado? La teoría de probabilidad permite valorar la cantidad de información que un estadístico (la proporción muestral P) aporta sobre el valor desconocido del parámetro (la probabilidad π), auténtico objetivo de nuestro estudio.

A continuación se estudia la distribución de los estimadores a lo largo de todas las posibles muestras. Así, se podrá saber cuánto oscila un estadístico de una muestra a otra y se podrá, por lo tanto, proponer medidas que cuantifiquen la cantidad de información que un estadístico, observado en una muestra concreta, proporciona sobre el parámetro de la población. Por simplicidad, veamos esta distribución en el caso de la muestra aleatoria simple (MAS).

Muestra aleatoria**Definición**

Muestra aleatoria simple (MAS) es aquella en la que: a) todos los elementos de la población tienen la misma probabilidad de pertenecer a ella, y b) cualquier combinación de n elementos tiene la misma probabilidad de pertenecer a la muestra.

Nota técnica

Todos los elementos de la muestra tienen la misma distribución, ya que vienen de la misma población.

Ejemplo 5.7

Imaginemos: a) la población infinita de todos los posibles pacientes de una enfermedad; b) un procedimiento aleatorio por el que se seleccionan de forma independiente $n = 1.000$ pacientes de esta población.

Contraejemplo 5.8

Una asociación profesional con 25.000 afiliados decide hacer un estudio para conocer qué proporción de ellos han recibido malos tratos en su trabajo. Diseña una muestra aleatoria de 2.000 afiliados a los que les envía un cuestionario, que contestan sólo 500. Se puede saber que los 2.000 representan a los 25.000, pero se desconoce a quién representan estos 500 y, por tanto, qué información aportan sobre el total de la población.

La definición de la población a la que se desea aplicar los resultados puede cambiar la consideración de la muestra.

Contraejemplo 5.9

Supongamos: a) la población finita de los 80 pacientes de una enfermedad determinada de un centro hospitalario; b) un proceso aleatorio de selección de 20 pacientes diferentes. Nótese que, al ser un muestreo sin reemplazamiento, al eliminar un paciente cada vez, la población de los pacientes susceptibles de ser seleccionados va variando, con lo que la variable aleatoria no tiene la misma distribución para cada uno de los elementos de la muestra.

Ejemplo 5.10

En el fondo, el objetivo del estudio del **contraejemplo 5.9** no puede ser conocer cómo se comportan estos 80 pacientes (tema vital para ellos y para el centro que los atiende, pero sin ningún interés para el resto de pacientes). El objetivo del estudio debe ser más ambicioso, de manera que se puedan beneficiar los pacientes de otros centros.

Ejemplo 5.10 (Cont.)

Ahora, por un lado, la situación se simplifica, ya que eliminar un elemento de esta población infinita prácticamente no modifica su distribución. Pero, por otro lado se complica, ya que debe tenerse en cuenta que los casos estudiados (sean 20 o sean 80) no son una muestra aleatoria de la población de todos los pacientes con la misma enfermedad. ¿Hasta qué punto los resultados son extrapolables?

Recuerde

Caso, muestra y población no se definen por separado, de forma aislada. Haga siempre la definición conjunta.

Volvamos a la definición de MAS. También resalta que la información aportada por las diferentes unidades debe ser independiente entre sí. Es decir, el valor obtenido en una observación no aporta información sobre el valor de otras observaciones. Este «no aportar información» debe entenderse como que la distribución de las restantes variables es la misma sea cual sea el valor observado.

Ejemplo 5.11

Sigamos con el **ejemplo 5.7** de pacientes con una enfermedad. Cada uno de los elementos de la muestra aporta exactamente la misma información sobre la población: que cierto paciente tenga un valor elevado no implica que el paciente siguiente ni cualquier otro deba tenerlo ni más alto ni más bajo.

Contraejemplo 5.12

En un estudio multicéntrico, ¿puede creerse que el resultado de un paciente de un centro no aporta información sobre el resultado de otro paciente del mismo centro? O por el contrario, ¿es más razonable pensar que los resultados obtenidos en pacientes de un mismo centro son más similares que los de pacientes de centros diferentes? Si es este último caso, la variable centro es una variable importante que debe ser tomada en cuenta en el análisis posterior.

Recuerde

En una MAS: a) las unidades se escogen al azar; b) todas ellas tienen la misma probabilidad de ser escogidas, y c) todas las posibles combinaciones de elementos tienen la misma probabilidad de configurar la muestra.

Papel de la inferencia estadística en los procesos científico y técnico

La capacidad de la estadística para inferir formalmente desde unos pocos datos de la muestra a la totalidad de la población ha permitido un progreso espectacular en todas las ciencias. Hoy en día, se acepta como modelo de razonamiento científico el contenido en el siguiente esquema (cuadro 5-1). Nótese que este modelo integra los razonamientos deductivos con los inductivos. Los primeros son necesarios, por ejemplo, para diseñar la recogida de datos. Los inductivos, por su parte, se requieren para generalizar las observaciones obtenidas en unos cuantos elementos.

1. Descubrir el problema a investigar
2. Documentar y definir el problema o hipótesis
3. Deducir consecuencias contrastables de las hipótesis
4. Diseñar la observación o la experimentación
5. Recoger los datos
6. Analizar los datos mediante inferencia estadística
7. Establecer las conclusiones
8. Integrar las conclusiones en el cuerpo de conocimiento

Cuadro 5-1 Pasos del método científico.

Comentario



Las profesiones científicas se van aproximando progresivamente al esquema de trabajo del método científico (fig. 5-1). Tras un análisis de la situación de partida, el planificador establece una estrategia (hipótesis) para alcanzar unos objetivos. Y se marca unos indicadores numéricos que le permitirán ir verificando (contraste empírico) hasta qué punto ha alcanzado estos objetivos. Se realiza el trabajo de producción (recogida de datos): en un médico, podría ser visitar a los pacientes; en un profesor, impartir la materia. Luego, deben evaluarse los resultados para medir el grado de obtención de los objetivos.

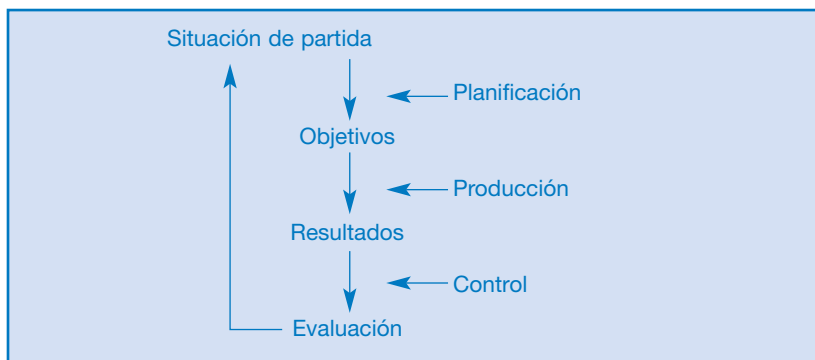


Figura 5-1 Esquema de trabajo del método científico-técnico.

Posibles errores en la inferencia estadística

En los libros de estadística es fácil encontrar definiciones como la siguiente: «Se entiende por **muestra aleatoria** aquella en la que todos los elementos de la población tienen una probabilidad conocida de pertenecer a la muestra. Si esta probabilidad es la misma para cualquier combinación de k elementos, se habla de **muestra aleatoria simple**». A partir de esta definición, la teoría de probabilidad permite cuantificar la información y el ruido aportados por dicha muestra aleatoria. Sin embargo, diferentes hechos hacen irreal la muestra aleatoria. En primer lugar, los individuos tienen derecho a rechazar su participación en el estudio, o incluso a abandonarlo en cualquier momento. En segundo lugar, no se dispone de definiciones operativas de las poblaciones; por ejemplo: no hay ningún listado con todos los pacientes de una determinada enfermedad. Todos estos fenómenos —no aleatorios— pueden provocar distorsiones no aleatorias, llamadas **sesgos**.

Resultados (%)		Previsiones	
		EL PERIÓDICO 23/10/1989 n = 9.524 + 2.000 (%)	LA VANGUARDIA 23/10/1989 n = 3.262 (%)
PSOE	39,6	40,5	41,5
PP	25,8	19,1	25,0
CIU	5,0	4,9	4,5
IU	9,1	10,3	7,5
CDS	7,9	8,5	6,5
Margen		±1	±2

Tabla 5-2 Prospección de voto y resultados electorales de octubre de 1989

Lectura



¿Hasta qué punto debemos creernos las previsiones electorales que se publican en diferentes medios? A continuación, y respecto a las elecciones generales de octubre de 1989, figuran los resultados reales (parámetros poblacionales) junto a las previsiones (estimaciones basadas en muestras) publicadas por El Periódico de Catalunya y por La Vanguardia (tabla 5-2). El segundo diario, que se comprometía con un margen menos ambicioso (2%) ha cumplido. En cambio, el primero ha fallado en dos ocasiones: a pesar de que prometía un margen máximo de 1 punto, para el PP se ha distanciado en 6,7 puntos y para IU en 1,2 puntos. La lectura de las «fichas técnicas» de ambos estudios permite encontrar posibles explicaciones a estas diferencias: que si el tipo de entrevista (personal o telefónica); que si la selección de los casos; que si los días en que se ha realizado la encuesta; que si no se tiene en cuenta la profesión;

Lectura (Cont.)



que si hay 2.000 encuestas adicionales en Cataluña, donde el PP suele estar bajo; etc. Nótese que estas explicaciones se basan en argumentos sociológicos, no estadísticos. Se podría haber argumentado que, por mala suerte, la estimación proporcionada por El Periódico se había alejado del auténtico valor. Éste sería un argumento probabilístico, basado en el error aleatorio debido al azar del muestreo.

La lectura anterior ilustra que en todo procedimiento de muestreo existen dos prototipos de errores: los debidos exclusivamente a las fluctuaciones del azar o **errores aleatorios** y todos los demás, conocidos como **errores sistemáticos** o **sesgos**. La estadística ayuda a cuantificar la magnitud de los primeros. Eliminar o acotar los segundos es una responsabilidad que la estadística comparte con la disciplina objeto del estudio, en nuestro caso, la Medicina. En el ejemplo anterior, han sido razonamientos políticos o sociológicos los que sugerían la existencia de posibles sesgos. En los estudios clínicos, el profesional sanitario debe razonar si las condiciones en las que se ha realizado el estudio le permiten negar la existencia de sesgos.

Recuerde



Al pasar de la muestra a la población, en el proceso inferencial hay dos posibles fuentes de errores: los aleatorios que la Estadística le ayudará a cuantificar; y los sistemáticos, o sesgos, cuya posible existencia debe usted estudiar a la luz de sus conocimientos clínicos.

Poblaciones implicadas en la inferencia estadística

Así pues, un buen estudio debe, en primer lugar, cuantificar la magnitud del error aleatorio. En segundo lugar debe justificar que la magnitud de este error es razonable para los objetivos del estudio. Y en tercer lugar, debe defender la ausencia de sesgos, que podrían haberse producido por desviaciones de la aleatoriedad de la muestra. Si se dan estas condiciones, en epidemiología suele decirse que el estudio es válido. Ahora bien, ¿válido para qué conclusiones, las de los autores del estudio o las de aquellos que desean aplicarlo en una nueva población? Puede ser ilustrativo hacer las siguientes definiciones de poblaciones, progresivamente más amplias.

Definición



Kleinbaum et al. (34) ofrecen las siguientes definiciones:

Población origen de la muestra (*actual population*) o población muestreada: es aquella población imaginaria de la que se hubiera obtenido, por extracción aleatoria pura, la muestra. Excluye, por lo tanto, aquellos casos que se niegan a participar en el estudio, o los que no figuran correctamente en el censo utilizado para generar la selección, o los que fallecen, etc.

Población objetivo o diana (*target population*): es aquella población a la que se desea poder inferir los resultados del estudio.

Poblaciones externas (*external populations*): son cada una de las posibles poblaciones más amplias a las que se puede desear, en algún momento futuro, o por otros investigadores, generalizar los resultados.

Una vez más, se puede recurrir a las muñecas rusas para resaltar que la población externa contiene a la población objetivo y que esta última contiene, a su vez, a la población origen de la muestra (fig. 5-2).

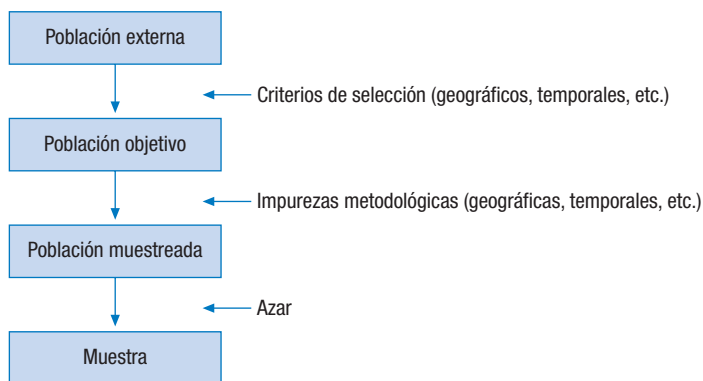


Figura 5-2 Esquema de las poblaciones que intervienen en un estudio.

Ejemplo 5.14



Los ensayos clínicos suelen realizarse con el criterio de inclusión de que la edad debe estar entre 18 y 65 años. Así, los pacientes menores y mayores nunca forman parte de la población objetivo. Quizá sí que formen parte de la externa. En este caso, los argumentos para poder creer que los resultados del estudio también les aplican son ajenos a la inferencia estadística.

Lectura



Dal-Ré et al. (35). «En general, la inclusión de enfermos en ensayos clínicos es un problema más importante de lo que los propios investigadores piensan, y siempre resulta más difícil de lo que en un principio se planeó [...]. En la mayor parte de los ensayos clínicos, muchos de los centros participantes son incapaces de seleccionar el número de enfermos según el calendario acordado; esto se debe a dos razones fundamentales: la selección de los investigadores participantes y las características del protocolo [...]. Las normas de buena práctica clínica de la Unión Europea especifican que los investigadores deben aportar datos de los pacientes que hubieran cumplido los criterios de selección de enfermos señalados en el protocolo, de forma que se pueda asegurar una adecuada tasa de selección.»

Ejercicio 5.7



A partir de un artículo de investigación imaginario o real (p. ej.: «Resultados a los 12 meses de un programa de deshabituación tabáquica en un centro de atención primaria») defina: población externa, población objetivo, población muestreada y muestra (fig. 5-3).

De acuerdo con las definiciones anteriores, se pueden aclarar varios términos de uso habitual.

Definición



Sesgo es toda diferencia entre el valor del parámetro en la población origen de la muestra y el valor del mismo parámetro en la población objetivo. **Error aleatorio** es la oscilación de los estadísticos obtenibles en las posibles muestras (siempre centrados en el parámetro de la población origen de la muestra). En epidemiología (Kleinbaum et al., 34) se dice que un estudio es **preciso** si este error aleatorio es razonablemente pequeño. Se dice que un estudio tiene **validez** si es preciso y no tiene sesgos. Se habla de **validez interna** si no tiene sesgos para con la población objetivo, y de **validez externa** si la ausencia de sesgos abarca también a la población externa.

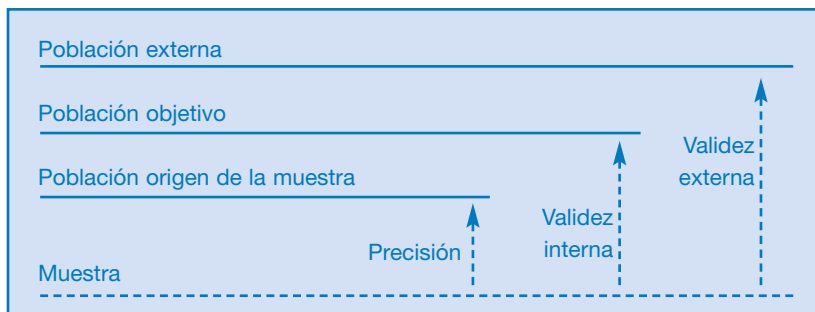


Figura 5-3 Relación entre las propiedades deseables de inferencia y las poblaciones a las que se aplican (Kleinbaum et al. (34)).

Comentario



Si le hacen una pregunta tan vaga como «¿es representativa su muestra?», puede contestar, seguro y con aplomo, que sí. Siempre hay una población origen para la que su muestra será representativa. Las preguntas de interés son: ¿representa su muestra a «cierta» población objetivo? ¿Cómo viene definida ésta? El riesgo de error aparece al redactar las conclusiones donde puede quererse llegar más lejos de lo que permiten los datos. Recuerde que entre los derechos humanos figura el de no participar en una encuesta o en un estudio clínico o epidemiológico, lo que hace inverosímil la muestra aleatoria «pura».

Nota técnica

Seleccionar una muestra «al azar» es un proceso riguroso, muy distinto de seleccionarla «al tuntún».

Ejercicio 5.8

Es bien conocida la tendencia de los autores y los editores de revistas a publicar solamente los resultados que son estadísticamente significativos. ¿Puede esta actitud provocar un sesgo?

Ejercicio 5.9

Los textos médicos suelen estar basados en artículos científicos escritos desde centros de atención terciaria. ¿Puede este hecho provocar un sesgo?

Recuerde

La inferencia estadística sólo cuantifica la magnitud del error aleatorio.

Si un número elevado de casos rehúsa participar en el estudio, el investigador debe dejar clara esta situación y discutir hasta qué punto compromete o invalida las conclusiones.

Estadístico media muestral**Distribución del estadístico media muestral \bar{X}** **Recuerde**

El cálculo de la media muestral o promedio obtenido en la muestra es bien conocido: $\bar{X} = \sum x_i / n$

Se dijo en el primer capítulo que para usar la media es necesario que la variable esté en escala de intervalo, que exista una unidad de medida que signifique siempre lo mismo.

Historieta

Bienvenido a Valle Sillytont

Fundado en 1.986

Altura 1.191 metros

823 habitantes

TOTAL 4.000

Lectura



Mediavilla et al. (36). La media del colesterol LDL (mmol/l) en los pacientes incluidos es $\bar{X}_{incl} = 3,33$ y la de los pacientes excluidos es $\bar{X}_{excl} = 3,49$.

Detengámonos un segundo: la media muestral ¿tiene distribución? ¿Qué significa esto? La pregunta es si el «estadístico media muestral» es una constante o, por el contrario, se trata de una variable que debe ser caracterizada por su distribución. Es fácil imaginar que la media muestral varía de una muestra a otra, tomando diferentes valores.

Ejemplo 5.14



En el estudio anterior, si se obtienen dos muestras, las medias \bar{X} de colesterol LDL serán algo diferentes, incluso aunque se tratara de casos de la misma población.

Ejemplo 5.15



Seleccionamos al azar $n = 100$ pacientes con hipertensión y calculamos la media muestral o promedio de sus valores del colesterol LDL. Este valor será diferente si obtenemos otra muestra de 100 pacientes y volvemos a calcular su media.

Contraejemplo 5.16



Seleccionamos al azar un paciente y le preguntamos 100 veces por su edad (¡pobre hombre! ¿Qué paciencia! ¿Y qué pensará de nosotros?) y hacemos el promedio. Ahora, cabe esperar que obtengamos el mismo valor si calculamos este promedio en otras 100 preguntas sobre su edad (se asume que el paciente es muy paciente, claro). Al no haber variabilidad en la edad, la media tampoco varía.

Ejemplo 5.17



Al paciente anterior en lugar de preguntarle la edad se le determina 100 veces el colesterol LDL por procedimientos independientes. Ahora, las medias podrían variar por un posible error de medida. Nótese que, en este caso, la variabilidad que se cuantifica hace referencia a la población de las diferentes mediciones en un mismo paciente, no a la población de pacientes. Es decir, cuantifica la variabilidad intra-paciente en lugar de la entre-paciente.

Así, las medias irán variando de una muestra a otra. Si se desea utilizar el estadístico promedio como estimador del parámetro poblacional «esperanza», esta variabilidad inducirá a errores, lo que, por supuesto, nunca es deseable. Ahora bien, ¿se pueden cuantificar estos errores? O, lo que es más importante, ¿se puede limitar su

magnitud? Para responder a estas dos preguntas cruciales, se debe primero contestar a otras más sencillas:

1. ¿Alrededor de qué valor varían? (Es decir, ¿cuál es su centro?)
2. ¿Varían mucho o poco alrededor de este valor? (Es decir, ¿cuál es su dispersión?)
3. ¿Qué forma tiene su distribución?

Nótese que, una vez aceptado que el estadístico promedio o media muestral tiene una cierta distribución, las dos primeras preguntas se reducen a conocer la media poblacional y la varianza o la desviación típica del estadístico. Veámoslas sucesivamente.

Centro de la distribución del estadístico \bar{X}

El centro poblacional del estadístico media muestral \bar{X} recibe el nombre de *esperanza de \bar{X}* y se representa por $E(\bar{X})$. Se sabe que, si la muestra es aleatoria simple, la esperanza de \bar{X} coincide con la esperanza de X que ya denotamos por $E(X)$ o μ .

Fórmula



$$E(\bar{X}) = E(X) = \mu$$

Recuerde



El centro de \bar{X} coincide con el de X . Si vamos obteniendo MAS de 100 pacientes de sus niveles de colesterol LDL, la media de las medias de todas las muestras ($E(\bar{X})$) coincide con la media poblacional de los niveles de colesterol LDL ($E(X) = \mu$).

Ejercicio 5.10



¿Es deseable esta situación? ¿Qué utilidad puede tener este hecho?

En resumen, se sabe que el conjunto de todas las medias de todas las posibles muestras aleatorias se encuentra situado precisamente en el centro de la variable que se está estudiando. Si se utilizara la media de la muestra \bar{X} para conocer la media de la población $\mu = E(X)$, existiría el consuelo de que el conjunto de todas las posibles muestras «apuntan» en la dirección correcta. Los errores podrán ser tanto por exceso como por defecto. Pero dado que el centro de la variable \bar{X} coincide con el objetivo de la estimación, el conjunto de los posibles errores positivos y negativos está equilibrado. Por lo menos, los errores no tienen ningún «favorito».

No sesgo: primera propiedad de un estimador. \bar{X} es insesgado

Definición



Se dice que un **estimador** es **insesgado** si el centro de su distribución a lo largo de todas las posibles muestras coincide con el parámetro que está estimando.

Nota técnica

Demostración de que \bar{X} no tiene sesgo al estimar $E(X)$ en el caso de MAS.

$$E(\bar{X}) = E(\sum x_i/n) = [E(\sum x_i)]/n = [\sum E(x_i)]/n = [\sum E(X)]/n = [n E(X)]/n = E(X)$$

Ejemplo 5.18

Se desea estimar la evolución de los ingresos de los médicos colegiados. Cada año, se pregunta a 30 de ellos, seleccionados al azar, por su salario. Las estimaciones que se obtengan se distribuirán alrededor del auténtico valor. Si se repitiera el muestreo infinitas veces, la mitad de estas muestras tendría valores que inflarían la realidad y la otra mitad valores inferiores. Pero, en general, estaría siendo equilibrado.

Contraejemplo 5.19

No sería correcto extrapolar estos resultados a los enfermeros, cuyos salarios son inferiores. Si se hiciera, se estaría cometiendo un sesgo igual a la diferencia entre las medias de los salarios de ambos colectivos.

Analogía

Sean dos lanzadores con arco que apuntan a sus respectivas dianas (fig. 5-4). El lanzador de la izquierda tiene un sesgo hacia la izquierda y arriba, mientras que el de la derecha está centrado.

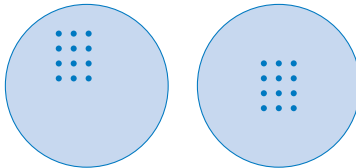


Figura 5-4 El arquero de la izquierda tiene sesgo, pero no el de la derecha.

Ejemplo 5.20

Dos informáticos han diseñado dos experimentos para conocer el rendimiento en tiempo de un nuevo algoritmo que proponen para decodificar el ADN. El primer informático analiza muestras del cromosoma 21, que es más corto, mientras que el segundo selecciona muestras de todos los cromosomas. Las posibles muestras del primero tendrán medias muestrales, cuyo centro, $E(\bar{X})$, estará por debajo de la media poblacional, μ . Las del segundo informático estarán centradas en la auténtica media poblacional (fig. 5-5).

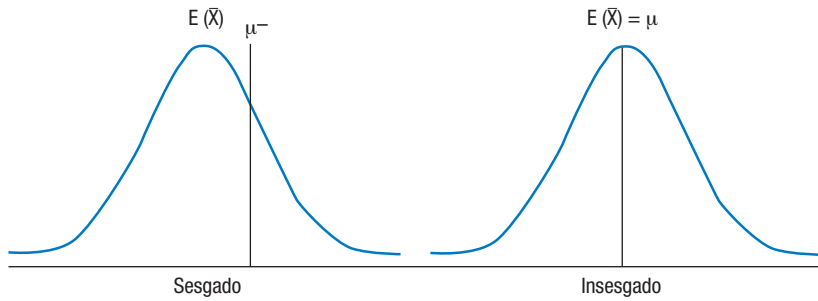


Figura 5-5 Estimador sesgado ($E(\bar{X}) \neq \mu$) e insesgado ($E(\bar{X}) = \mu$).

Recuerde



Si la muestra es aleatoria, el promedio muestral es insesgado.

Recuerde



Si la muestra NO es aleatoria, la estadística NO puede garantizar ausencia de error sistemático.

En el fondo, se está diciendo que podría ser peor, que podría ser que las estimaciones estuvieran apuntando en dirección incorrecta. Parece un pobre consuelo, ya que las estimaciones, aunque se distribuyan alrededor del parámetro de interés, no acaban de acertar. De hecho, si se ha trabajado con una variable continua (con tantos decimales como se desee), nunca coincidirá el valor estimado con el parámetro de interés: siempre habrá un error que no puede olvidarse.

Por ello, hay que preguntar: ¿es la magnitud de este error tolerable? Ello conduce a la siguiente pregunta: ¿cómo se puede medir o cuantificar este error? A continuación se estudia la información que la varianza de \bar{X} aporta sobre este error.

Dispersión de la distribución del estadístico \bar{X}

Ejercicio 5.11



¿Recuerda el cálculo de la media y de la varianza de una muestra?

$$\text{REPASO: } S^2 = \sum_{i=1, n} (x_i - \bar{X})^2 / (n - 1)$$

$$S^2 = [\sum x_i^2 - (\sum x_i)^2/n] / (n-1) \text{ (más eficiente)}$$

Practique ambas fórmulas, a mano y con la calculadora, para el ejemplo sencillo de $n = 5$ alumnos que contestan que en su familia son 1, 2, 3, 4 y 5 hermanos.

Varianza del estadístico media muestral: $V(\bar{X})$. ¿Cuánto vale? ¿Qué significa? Se sabe que la dispersión de \bar{X} es directamente proporcional a la dispersión de X e inversamente proporcional al tamaño n de la muestra.

Fórmula

$$V(\bar{X}) = V(X)/n$$

Nota técnica

Demostración en el caso de MAS.

$$V(\bar{X}) = V(\sum x_i / n) = [V(\sum x_i)] / n^2 = [\sum V(x_i)] / n^2 = [\sum V(x)] / n^2 = [nV(x)] / n^2 = V(x) / n$$

↑ independientes
 ↑ idénticamente distribuidas

Nótese que la relación entre la variabilidad de las medias muestrales, $V(\bar{X})$, y el tamaño muestral, n , es inversa. Ello implica que cuanto mayor sea el tamaño n de la muestra, menor será la variabilidad de las medias muestrales. Lo que coincide con todo lo que ya sabíamos: cuanto mayor es el tamaño de la muestra, mayor credibilidad tienen los resultados.

Lectura

Una sola flor no indica primavera.

Ejemplo 5.21

Supóngase que se están tomando muestras de la altura de niños. La variabilidad de las posibles medias muestrales será mayor si tomamos muestras de tamaño $n = 3$ que si son de tamaño $n = 1.000$.

Ejercicio 5.12

¿Es coherente esta situación?, ¿qué utilidad puede tener?

Contraejemplo 5.22

No se tendrá más información si se mide 1.000 veces al mismo niño. Para que una nueva observación aporte el máximo de información deberá ser lo más independiente posible de la información aportada por las observaciones previas.

Tampoco es una gran sorpresa que cuanto mayor sea la variabilidad de una variable, mayor será el grado de oscilación de la media de una muestra a otra.

Ejemplo 5.23



Supóngase que se están tomando muestras de la altura de niños. La variabilidad de las posibles medias muestrales será mayor si los niños tienen edades comprendidas entre los 5 y los 15 años, que si todos tienen 8 años.

Ejemplo 5.24



Suponga que los ingresos de los titulados de una facultad aumentan con el tiempo que pasa desde que dejan la universidad. Si es así, la dispersión de X será mayor si se estudia el conjunto de todos los titulados, que si se estudia solamente los titulados en un cierto año. En consecuencia, si se obtienen muestras de la población total, las medias muestrales, \bar{X} , fluctuarán más que las muestras de un solo curso.

Recapitulación: Se ha visto que la media muestral \bar{X} oscila de una muestra a otra alrededor del parámetro $\mu = E(X)$ que se desea estimar. ¿Qué significa la varianza de \bar{X} ? Si es el promedio de las distancias (cuadradas) con el centro, representa el error (cuadrado) que se cometería en cada muestra al estimar el parámetro poblacional a partir del valor de \bar{X} obtenido. Así, la varianza es el promedio de todos los posibles errores (cuadrados).

Recuerde



La varianza de \bar{X} proporciona el promedio de los errores al cuadrado.

Por lo tanto, la varianza de \bar{X} informa sobre el promedio del error «cuadrado». Dado que este cuadrado es incómodo a la hora de interpretar los resultados, de la misma manera que se obtenía la desviación típica de una variable mediante la raíz de su varianza, ahora, a través de la raíz de la varianza de \bar{X} , se obtiene el error típico de \bar{X} .

Nota técnica



Por error típico de \bar{X} se entiende el error esperado de la media muestral \bar{X} al estimar el parámetro μ .

FórmulaError típico de \bar{X}

$$\sigma_{\bar{x}} = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

Ejercicio 5.13

Suponga ahora que está interesado en conocer el promedio de hermanos de las familias de los alumnos del ejercicio 5.11. Si se considera a esta muestra de $n = 5$ como una muestra aleatoria representativa de todas las familias, ¿qué error cabe esperar que tiene la media observada en la muestra al estimar la media poblacional?

Recuerde

El error típico de la media \bar{X} es la desviación típica de la variable en estudio dividida entre la raíz del número de casos.

Ejemplo 5.25

La siguiente frase «los 100 niños tratados han tenido fiebre durante una media de 3 días; el error típico (o estándar) ha sido de 0,1 día» hace inferencia a los valores de la población: se afirma que, al aplicar este tratamiento en niños de las mismas características, la media de duración de la fiebre es de 3 días y que el error esperado al decir que la media poblacional es de 3 días, es de 0,1 día.

Debe quedar claro que se trata de un «error», por tanto con connotación negativa. Nótese que mientras el término desviación típica no debería tener ninguna connotación, ni positiva ni negativa, ahora el error típico ya deja claro desde el primer momento que se trata de algo negativo, no deseable: el error que cabe esperar que se cometa al estimar el parámetro media poblacional a partir del estimador media muestral.

Ejemplo 5.26

La altura de las mujeres adultas tiene una distribución Normal de media $\mu = 165$ cm, y desviación típica $\sigma = 7$ cm. Que la desviación típica sea de 7 cm no es ni bueno ni malo, simplemente refleja una situación natural. (Nótese que esta diversidad es, para un ecólogo, fuente de riqueza; mientras que para un fabricante de camisas representa un reto que

Ejemplo 5.26 (Cont.)

superar.) En cambio, si para estimar la altura media μ de las mujeres se calcula la media en una muestra de $n = 100$ mujeres, el error típico que conlleva la estimación es:

$$\sigma_{\bar{x}} = \sqrt{V(\bar{X})} = \sigma_x / \sqrt{n} = 7 \text{ cm} / \sqrt{100} = 0,7 \text{ cm}$$

Este valor del error típico dice que la imprecisión de \bar{X} al estimar μ es de 0,7 cm.

Nota

De la misma manera que \bar{X} es el estimador de μ , s es el estimador de σ y $S_{\bar{x}}$ el de $\sigma_{\bar{x}}$. Como en general no se conocerá σ , el error típico que se emplea es $S_{\bar{x}}$.

Ejercicio 5.14

Se estima en 4,4 puntos el incremento en la calidad de vida de la semana 0 a la 24 en 43 pacientes. Si la desviación típica observada ha sido de 1,2 puntos, ¿cuánto vale el error típico? Interprete el resultado.

El error *típico* habla del error esperado o promedio, ya que el error exacto que se comete en una muestra concreta permanece desconocido y puede ser más grande o más pequeño.

Nota

Formalmente no se puede interpretar el error típico como el promedio de los errores (es la raíz cuadrada del promedio de los errores cuadrados), pero a nivel práctico, decir que representa el error promedio o esperado es una buena aproximación.

Es muy importante resaltar que este error típico puede hacerse tan pequeño como se quiera: simplemente se trata de ir aumentando el tamaño de la muestra.

¿Desviación típica o error típico?

«¿Y qué debo utilizar, la desviación típica o el error típico?»

Esta pregunta no tiene razón de ser, ya que no son medidas alternativas para un mismo objetivo: la desviación típica es una medida descriptiva de cómo son los casos, mientras que el error típico es una medida del error asociado a un proceso inferencial. Así, se puede usar la desviación típica cuando se describen las condiciones y los casos en los que se ha hecho el estudio (al inicio de «resultados»); y se puede usar el error típico cuando se desea inferir (desde la muestra a la población) el efecto observado. Esta inferencia permitirá a otros científicos utilizar nuestros resultados.

Recuerde

La desviación típica es una medida de dispersión que sirve para describir los datos (¿cómo son mis casos?), mientras que el error típico es una medida del error de estimación al hacer inferencia (¿qué incertidumbre o ruido lleva asociado mi atrevido salto de la muestra a la población?).

Nota técnica

Desviación típica es «*standard deviation*» mientras que error típico es «*standard error*».

Lectura

«Ítem 15 (9): las variables continuas pueden resumirse para cada grupo a través de la media y de la desviación típica. Los errores típicos y los intervalos de confianza no son adecuados para describir la variabilidad; son más bien estadística inferencial que descriptiva.

Ítem 17 (9): resultados y estimación. Para cada variable principal y secundaria, un resumen de resultados para cada grupo y la estimación del tamaño del efecto y su precisión (p. ej., 95% de intervalo de confianza).

Es muy importante resaltar que la relación entre el tamaño de la muestra y el error típico no es inversamente proporcional. Sí que es inversa, pero hay una raíz de por medio. Por tanto, si aumentando el tamaño muestral se quiere disminuir el error típico a la mitad, se deberá multiplicar por cuatro el esfuerzo en recoger datos.

Ejemplo 5.27

El nivel de plaquetas en pacientes de una determinada enfermedad tiene una varianza de 2.500 unidades². Si, para conocer su valor medio, se obtiene una muestra de 25 pacientes, la varianza de la variable promedio \bar{X} en 25 pacientes vale:

$$V(\bar{X}) = V(X) / n = 2.500 \text{ u}^2 / 25 = 100 \text{ u}^2$$

En consecuencia, el error (cuadrado) que cabe esperar de nuestra observación es de 100 unidades (cuadradas). Asimismo, el error típico de la estimación de la media en los 25 pacientes es:

$$\sigma_{\bar{x}} = \sqrt{V(\bar{X})} = \sqrt{100 \text{ u}^2} = 10 \text{ u}$$

Es decir, el valor de la media que obtengamos estará afectado por un error típico de estimación de 10 unidades.

Ejemplo 5.27 (Cont.)

En cambio, si se aumenta la muestra de 25 a 100 casos, la varianza de la media muestral \bar{X} y el correspondiente error típico valen:

$$V(\bar{X}) = V(X) / n = 2.500 u^2 / 100 = 25 u^2$$

$$\sigma_{\bar{X}} = \sqrt{V(\bar{X})} = \sqrt{25 u^2} = 5 u$$

A pesar de tener una muestra cuatro veces más cara, el ruido de estimación ha bajado a la mitad.

Recuerde

Si desea estimar un parámetro y dispone de un estimador insesgado, el error típico (SE: standard error) de este estimador le informa del error esperado al afirmar que el valor del parámetro poblacional coincide con el valor del estimador obtenido en su estudio.

Estabilidad del conjunto

El hecho de que el error típico de la media se vaya haciendo más pequeño a medida que aumenta el tamaño muestral indica una cierta estabilidad de los grupos que se contraponen a la variabilidad de las unidades.

Recuerde

La variabilidad de los individuos contrasta con la regularidad del conjunto.

Ejemplo 5.28

Pongamos que la probabilidad de nacer varón sea 1/2. El próximo nacimiento de Barcelona tiene esta probabilidad de ser varón. Pero, o bien será niño o bien será niña, pero no será mitad niño y mitad niña. En cambio, podemos tener la tranquilidad de que el próximo año nacerán alrededor de un 50% de niños y un 50% de niñas en Cataluña. No le pediremos a un político que elabore un plan de contingencia por si, por azar, durante unos años sólo nacen bebés de uno de los dos géneros.

Otro ejemplo similar lo tenemos con las quinielas o la lotería. A una persona en concreto o le toca o no le toca. La incertidumbre es absoluta. Pero el que organiza los juegos tiene la confianza de que le tocará a un cierto tanto por ciento. Y esta confianza será tanto mayor cuanto mayor sea el número de personas que apuestan.

Más propiedades de los estimadores

De la misma forma que la esperanza de \bar{X} sirvió para definir el concepto de estimador insesgado, se puede utilizar el error típico de \bar{X} para definir propiedades interesantes de los estimadores.

Definición



Se dice que un estimador es **convergente** si, a medida que crece el tamaño de la muestra, se va acercando progresivamente al valor del parámetro que se desea estimar.

Ejemplo 5.29



\bar{X} es un estimador convergente ya que $V(\bar{X})$ disminuye a medida que aumenta n .

No se insiste más en esta definición que resalta la deseada propiedad de que cuanto mayor sea la muestra, mejor será la estimación.

Definición



Entre dos estimadores insesgados, se dice que es más **eficiente** el que tiene un error típico menor.

Analogía



Imaginemos dos lanzadores con arco, ambos insesgados. El de la izquierda tiene una mayor dispersión alrededor de la diana, por lo que es menos eficiente (fig. 5-5).

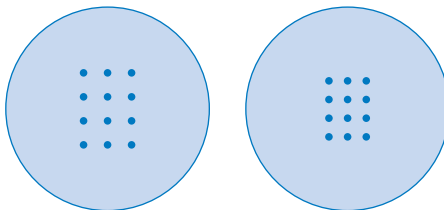


Figura 5-5 Ambos arqueros son insesgados, pero el de la derecha es más eficiente porque su error es más pequeño.

Nota técnica

El término **eficiente** tiene una clara connotación económica, en el sentido de que su relación calidad / coste será más favorable. El estimador más eficiente proporciona más información (tiene menos ruido o error aleatorio) para un mismo tamaño muestral (mismo coste); o también puede obtener la misma cantidad de información con una muestra más pequeña (menor coste).

Ejemplo 5.30

Dos investigadores han diseñado dos experimentos para comparar la biodisponibilidad de dos preparaciones farmacéuticas alternativas. El primero ha obtenido dos muestras de voluntarios, administrando a cada una, una de las dos formulaciones. Luego compara las medias de las dos muestras ($\bar{X}_1 - \bar{X}_2$). El segundo investigador ha recogido la información de las dos preparaciones en un único grupo, y calcula la media de las diferencias (\bar{X}_D), eliminando, de esta forma, la variabilidad debida al voluntario. Los gráficos muestran que, siendo ambos experimentos insesgados, el segundo es más eficiente (fig. 5-6).

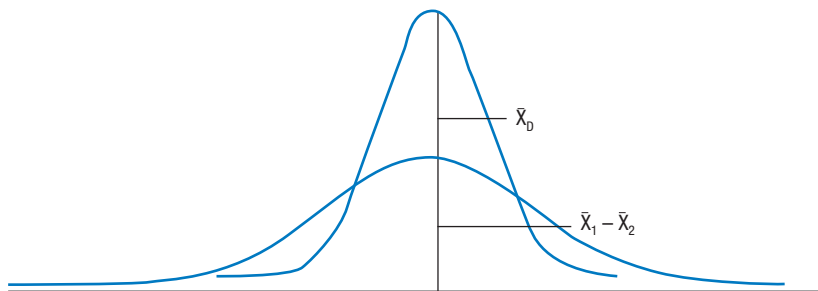


Figura 5-6 Ambos estimadores, $\bar{X}_1 - \bar{X}_2$ y \bar{X}_D , son insesgados pero \bar{X}_D es más eficiente.

Comentario

La estadística permite cuantificar los errores aleatorios. Si conviene que la MAS (Muestra Aleatoria Simple) deje de ser «simple», ningún problema, un profesional de la estadística le ayudará a obtener correctamente el valor del error típico de estimación. Pero si la muestra deja de ser «aleatoria», debe realizarse un detallado estudio sustantivo y estadístico de las consecuencias de las desviaciones de la aleatoriedad. Por estudio sustantivo se entiende el que emplea los conocimientos de la disciplina subyacente, sean clínicos, biológicos, epidemiológicos, etc.

Estimación puntual

Definición



Al valor observado de un estimador en una muestra se le denomina estimación puntual del parámetro.

Ejemplo 5.31



En una muestra aleatoria de 9 personas, la presión arterial sistólica (PAS) ha tenido una media muestral \bar{X} igual a 120 mmHg y una desviación típica muestral (S) de 12 mmHg. Así, la estimación puntual de la PAS media en esta población ha sido de 120 mmHg. El error típico de esta estimación se puede cifrar en:

$$S_{\bar{x}} = S / \sqrt{n} = 12 \text{ mmHg} / \sqrt{9} = 4 \text{ mmHg}$$

Por lo tanto, hay una señal de 120 mmHg que está afectada por un ruido de 4 mmHg.

Recuerde



El error típico informa del error esperado o promedio, pero el error exacto en una muestra concreta permanece desconocido, pudiendo ser inferior o superior.

Ejercicio 5.15



El descenso de la PAS tras la administración de un fármaco en una muestra de 16 pacientes ha tenido una media de 12 mmHg y una desviación típica de 8 mmHg. Calcule el error típico e interprete los resultados.

Ejercicio 5.16

Si hubiera deseado que el error típico hubiera sido de 1 mmHg, ¿cuántos casos hubiera necesitado (desviación típica de 8 mmHg)?

Forma de la distribución del estadístico \bar{X}

Ya se ha dicho que las posibles medias muestrales \bar{X} se distribuyen alrededor de la media poblacional $\mu = E(X)$ con una distancia promedio que cuantifica el error típico $\sigma_{\bar{x}}$. Ahora bien, ¿qué forma tiene la distribución de \bar{X} ?

Por las leyes de combinatoria y probabilidad, en general la muestra contendrá valores próximos a la media poblacional. También es posible obtener valores alejados, si bien será menos frecuente, ya que, por ejemplo, la probabilidad de obtener una muestra aleatoria que sólo contenga valores muy superiores a la media, y resulte en un valor positivo extremo de \bar{X} , es pequeña. En realidad, cuanto más se aleje una \bar{X}

del valor de la μ , menos probable es observarla. Todos estos razonamientos ilustran que \bar{X} se distribuye con una forma de campana o de montaña como la de la ley normal de Gauss-Laplace.

Recuerde



La media muestral \bar{X} se distribuye siguiendo la ley Normal Gauss-Laplace.

La distribución Normal aparece en variables que sean el resultado de muchos factores o fuerzas que actúen independientemente y con influencias similares. Y eso es precisamente lo que es una media muestral \bar{X} , ya que cada observación de la muestra contribuye con el mismo peso o influencia. Queda por aclarar qué significa «muchos»: ¿cuántos casos se necesitan para que la distribución del promedio de una muestra se acerque a la ley normal?

Ejercicio de Navegación



Entre en la página que se indica a continuación y observe, con la ayuda de la simulación que realiza la aplicación, cómo se comporta la media de una muestra a lo largo de muchas muestras y compruebe cómo se acerca la media muestral \bar{X} a la poblacional μ al crecer n .

http://www.ruf.rice.edu/~lane/stat_sim/sampling_dist/index.html

Nota



Este número puede ser muy inferior si la distribución original de la variable que se está estudiando ya se aproxima a la forma de la distribución normal. De hecho, si X es normal, para $n = 1$, \bar{X} ya es normal.

Nota técnica



El teorema del Límite Central (TLC) establece que si se toman muestras de tamaño n , de una población de media μ y desviación típica σ , a medida que crece n , la distribución de \bar{X} se aproxima a la ley Normal con media μ y desviación típica σ/\sqrt{n} .

Ejercicio 5.17



A partir de lo dicho, ¿la distribución de la variable \bar{X} cambia de forma cuando crece el tamaño de la muestra? ¿Y la de X ?

Ejercicio 5.18

¿Cómo cambia la forma de la distribución de la variable \bar{X} cuando crece el tamaño de las muestras?

Ejemplo 5.32

La edad de los pacientes incluidos en un estudio sigue una distribución uniforme (aplanada, con el mismo número de casos en todas las franjas de edad). Si se toman muestras de tamaño $n = 30$ y se calcula la media muestral \bar{X} de la edad, su distribución también será, a niveles prácticos, normal.

Ejemplo 5.33

La presión arterial sistólica en los adultos sanos tiene una distribución que se asemeja bastante a la ley normal. Si se toman muestras de tamaño $n = 3$ y se calcula la media muestral \bar{X} de la PAS, la distribución de esta media será, a niveles prácticos, normal.

Recuerde

Las condiciones para poder creer que el promedio obtenido en una muestra sigue una distribución normal son, o bien muestra grande ($n \geq 30$) o bien distribución normal de la variable en estudio.

Aplicación práctica de la distribución de la media muestral \bar{X}

Se ha visto que la distribución normal permite construir intervalos que contengan un determinado porcentaje de unidades o casos. Ahora, la variable en estudio es \bar{X} , por lo que, utilizando la distribución normal, se pueden construir intervalos que contengan un deseado porcentaje de las medias que se podrían obtener en todas las posibles muestras.

Ejemplo 5.34

La glucosa en sangre (X) sigue una distribución normal de media 100 y desviación típica 10

$$X \rightarrow N(100\text{mg/ml}, 10\text{mg/ml})$$

Se desean construir los siguientes intervalos:

- i) el intervalo que contiene el 95% de las unidades de la población;
- ii) el intervalo que contiene el 95% de las posibles \bar{X} de muestras de tamaño $n = 9$; y
- iii) el intervalo que contiene el 95% de las posibles \bar{X} de muestras de tamaño $n = 100$.

Los tres intervalos son similares en el sentido de que deben contener el 95% de las unidades y dejar fuera el 5% ($\alpha = 0,05$).

Ejemplo 5.34 (Cont.)

Y son distintos en el sentido de que se refieren a unidades totalmente diferentes, con distribuciones también diferentes. En el primer intervalo, las unidades son individuos, pacientes o casos, mientras que en el segundo y tercer ejemplo se trata de medias muestrales que se podrían obtener si se repitiera indefinidamente el proceso de tomar muestras de $n = 9$ y $n = 100$ de estos individuos.

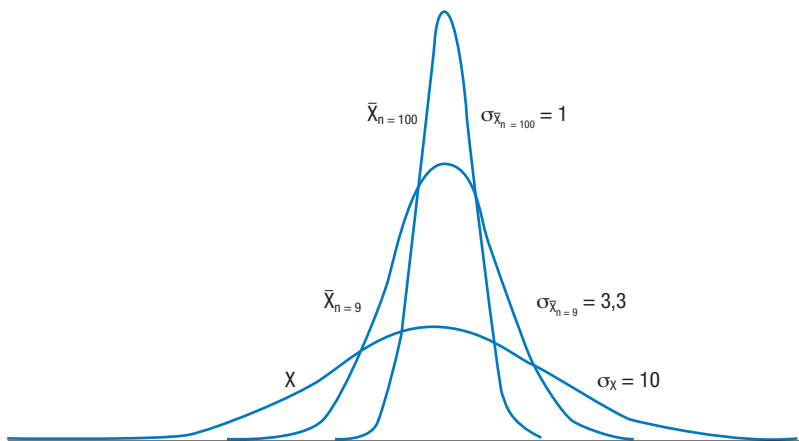


Figura 5-7 Distribución de las variables X , $\bar{X}_{n=9}$ y $\bar{X}_{n=100}$.

Ejemplo 5.34 (Cont.)

Todas estas distribuciones (fig. 5-7) seguirán la ley normal: al ser normal la distribución de la glucosa en los casos, también lo es la distribución de la media \bar{X} , sea cual sea el número de casos. Todas tienen, también, la misma media. Pero cambia la dispersión: para el primer ejemplo, se trata de la desviación típica de la variable original, la glucosa en sangre, 10 mg/ml; mientras que para los restantes dos ejemplos, se trata del error típico, debiéndose dividir la desviación típica por la raíz del número de casos respectivos:

- i) $X \rightarrow N(100 \text{ mg/ml}, 10 \text{ mg/ml})$
- ii) $\bar{X}_{n=9} \rightarrow N(100, 10/\sqrt{9}) = N(100, 10/\sqrt{9}) = N(100 \text{ mg/ml}, 3,33 \text{ mg/ml})$
- iii) $\bar{X}_{n=100} \rightarrow N(100, 10/\sqrt{100}) = N(100, 10/\sqrt{100}) = N(100 \text{ mg/ml}, 1 \text{ mg/ml})$

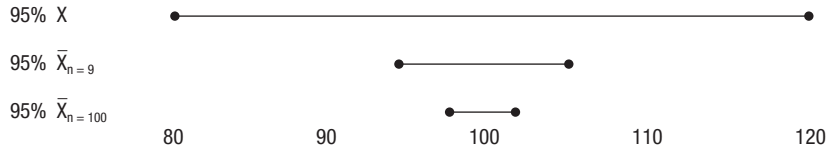


Figura 5-8 Intervalos que contienen el 95% de las observaciones de las variables X , $\bar{X}_{n=9}$ y $\bar{X}_{n=100}$.

Ejemplo 5.34 (Cont.)



Los límites de los intervalos (fig. 5-8) se pueden calcular utilizando las tablas de la normal tipificada: $Z_{0,025} = 1,96$

i) Intervalo que contiene el 95% de las glucemias individuales, X

$$\mu \pm Z_{\alpha/2} \sigma = 100 \pm 1,96 \cdot 10 = 100 \pm 19,6 = [80,4, 119,6]$$

ii) Intervalo que contiene el 95% de las medias ($\bar{X}_{n=9}$) de las infinitas muestras de $n=9$ individuos

$$\mu \pm Z_{\alpha/2} \sigma/\sqrt{n} = 100 \pm 1,96 \cdot 10/3 = 100 \pm 6,53 = [93,47, 106,53]$$

iii) Intervalo que contiene el 95% de las medias ($\bar{X}_{n=100}$) de las infinitas muestras de $n=100$ individuos

$$\mu \pm Z_{\alpha/2} \sigma/\sqrt{n} = 100 \pm 1,96 \cdot 10/10 = 100 \pm 1,96 = [98,04, 101,96]$$

Nota



El ejemplo anterior muestra, una vez más, la mayor variabilidad de las muestras de menor tamaño, ya que la distribución de los casos individuales puede considerarse como la distribución de las muestras de tamaño $n = 1$.

Este ejemplo muestra cómo, utilizando la distribución normal, se puede conocer, a partir de los valores poblacionales de la media [$E(X)=\mu$] y la varianza [$V(X)$], dónde estarán los valores de la media muestral (\bar{X}) en el 95% de las posibles muestras. Puede ser interesante, pero tiene poca utilidad práctica, ya que el problema habitual es, conocidos los estimadores muestrales de la media (\bar{X}) y la varianza (S^2), ¿qué se sabe de la media poblacional $E(X) = \mu$? Esta interesante pregunta se contesta en el siguiente capítulo.

Ejercicio 5.19



Los psicólogos miden la inteligencia por el CI, que sigue, en la población adulta general, una $N(100,15)$. Es decir, $\mu = 100$ y $\sigma = 15$. Si se recolectaran muchas muestras de tamaño $n = 9$ y en cada muestra j se calculara su media \bar{X}_j ;

Ejercicio 5.19 (Cont.)

- a) ¿Cómo variarían las medias \bar{X}_j de las muestras?
- b) En una facultad de Medicina se ha recogido una muestra de tamaño $n = 9$ y se ha observado $\bar{X}_j = 104$, ¿se trata de 1) un ejemplo aceptablemente típico; o 2) especialmente afortunado, muy cerca de μ ; o 3) tan raro y alejado que se sospecha que estos alumnos no son de aquella población?
- c) Repetir los dos apartados anteriores, pero con $n = 25$ y $n = 225$.

Soluciones a los ejercicios

5.1 Debe desearse tener información sobre el centro y sobre la dispersión de la distribución, sobre ambos. Usualmente, la comunidad científica enfoca su interés en los valores centrales para poder resumir o representar los casos. Pero hay que hacer el esfuerzo de cuantificar también el grado de dispersión, e incluso conocer la forma de la distribución.

5.2 Cualquier ejemplo es válido. También sería terriblemente aburrido «decir toda la verdad» sobre la carga viral. Nótese, en cambio, que no lo sería sobre el género («53 fueron del género masculino y 47 del femenino»). Y quizá tampoco sobre el número de infecciones oportunistas («2.523 casos no presentaron ninguna; 48 tuvieron una; 7, dos y 1 caso, tres infecciones»).

5.3 En el estudio de los odontólogos, una unidad es una visita a la consulta, mientras que en la población general, una unidad es un habitante. Como hay habitantes que van al dentista más veces que otros, los primeros estarán sobrerrepresentados en un estudio en el que se seleccionen «visitas». Nótese que aquellos que nunca van al dentista tienen una probabilidad nula de ser seleccionados. En definitiva, el estudio de los dentistas ofrecerá cifras más altas que el de la población general.

5.4 Los estudios de satisfacción hospitalaria se basan en las altas hospitalarias, mientras que el estudio del defensor se basó en ciudadanos. Igual que en el ejercicio anterior, aquellos ciudadanos que van menos a los centros públicos tienen una probabilidad menor de ser seleccionados. En resumen, la satisfacción entre las «altas» hospitalarias será mayor que la satisfacción de los habitantes.

5.5 Lo mismo que antes: si los pacientes que desarrollan infecciones nosocomiales permanecen más tiempo en el centro y la selección se hace a partir de las «camas ocupadas», las cifras de infección serán mayores.

5.6 Por ejemplo, en un estudio del perfil lipídico en 41 pacientes con HIV tratados, la media del colesterol total fue de 4,51 mmol/l. Se trata de la media obtenida en la muestra (promedio muestral), y que informa sobre el valor de la media en la población (esperanza poblacional), que es desconocida.

5.7 La muestra serían los casos seleccionados para el estudio; la población muestreada, aquella de la cual se hubiera obtenido por meros mecanismos aleatorios, la muestra; la población objetivo, todos los fumadores de esa región sanitaria; y la población externa, los fumadores de cualquier región sanitaria.

5.8 Se trata del conocido sesgo de publicación: al publicar sólo lo que ha resultado significativo, se da menor oportunidad a difundir resultados «negativos». (Este argumento se complementará en el capítulo 7.)

5.9 No, si se pretende utilizar los resultados para centros de atención terciaria. Pero sí, si se desea utilizarlos en centros donde la severidad de los casos tratados pueda ser menor.

5.10 Si se desea utilizar a la media muestral (\bar{X}) para acercarse a la media poblacional ($E(X)$ o μ) es bueno que la distribución de \bar{X} se disponga alrededor del auténtico valor de μ . Dicho al revés: sería peor que se distribuyera alrededor de cualquier otro valor. Y cuanto más alejado de μ estuviera este otro valor, peor.

5.11 Media $\bar{X} = \sum_i x_i / n = (1+2+3+4+5) / 5 = 3$

Varianza $S^2 = \sum_{i=1,n} (x_i - \bar{X})^2 / (n - 1) =$

$$= [(1 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 + (4 - 3)^2 + (5 - 3)^2] / 4 = 10/4 = 2,5$$

$$S^2 = [\sum x_i^2 - (\sum x_i)^2 / n] / (n - 1) =$$

$$= [(1^2+2^2+3^2+4^2+5^2) - (1+2+3+4+5)^2/5] / 4 = [55 - 225/5] / 4 = 10/4 = 2,5$$

$$S = \sqrt{S^2} = \sqrt{2,5} \approx 1,58$$

Es decir, la media muestral es 3 hermanos; la varianza muestral, 2,5 hermanos² y la desviación tipo muestral es de aproximadamente 1,6 hermanos. Podemos imaginar que la distancia (o desvío) de una familia típica con la media es de 1,6 hermanos.

5.12 Es coherente: cuantos más casos se tiene, se dispone de más información y, por tanto, hay menos error aleatorio y menos riesgos del muestreo. Es útil (y, por consiguiente, deseable) en el sentido de que un mayor esfuerzo en la recolección de datos se ve recompensado por una menor oscilación de las estimaciones.

5.13 $S_{\bar{X}}^2 = S^2/n = 2,5/5 = 0,5$; $S_{\bar{X}} = \sqrt{S_{\bar{X}}^2} = \sqrt{0,5} \approx 0,707$

Si se afirma que la media de la población es de 3 hermanos, el error esperado al hacer esta afirmación es de 0,7 hermanos.

5.14 $S_{\bar{X}} = \sqrt{S_{\bar{X}}^2} = S/\sqrt{n} = 1,2/\sqrt{43} = 0,18$.

La señal obtenida ha sido 4,4 y el error asociado que lleva esta señal es de 0,18.

5.15 $S_{\bar{X}} = \sqrt{S_{\bar{X}}^2} = S/\sqrt{n} = 8/\sqrt{16} = 2$.

La señal obtenida ha sido de 12 mmHg y el error asociado que lleva esta señal es de 2 mmHg.

5.16 Si se desea que $S_{\bar{X}} = 1 \rightarrow 8/\sqrt{n} = 1 \rightarrow n = 8^2 = 64$.

Una vez más se ve que si se desea reducir a la mitad el ruido de la muestra anterior, hay que multiplicar por 4 el tamaño muestral ($64 = 16 \cdot 4$).

5.17 A medida que crece el tamaño muestral, lo que va cambiando de forma es la distribución de la variable media muestral \bar{X} . La distribución de los valores observados, es decir, lo que se llama la distribución de la variable en estudio es siempre la misma para todos los casos, haya 3, 50 o 1.000. Si no tiene claro que lo que va cambiando es la distribución de la media muestral \bar{X} repita los ejercicios de la última navegación.

5.18 Su centro, la esperanza, no cambia, pero sí que lo hace la dispersión y también puede hacerlo la forma. La dispersión, cuantificada por el error típico, se va haciendo más pequeña a medida que crece el tamaño muestral (la reducción es proporcional a \sqrt{n}). La forma, en el caso de variables que no siguen una distribución normal, se aproxima cada vez más a la de esta distribución (en el caso de variables que siguen la ley Normal, ya tiene esta distribución para cualquier n).

$$5.19 \text{ a) } V(\bar{X}) = V(X) / n = 15^2/9 = 25 \text{ u}^2 \rightarrow S_{\bar{X}} = 5 \text{ u}$$

La variabilidad de las medias muestrales es la tercera parte de la variabilidad de la variable

$$\text{b) } [\bar{X}_i - E(X)]^2 = [104 - 100]^2 = 4^2 \text{ u}^2 \text{ cifra «aceptablemente típica»}$$

ya que su valor esperado era 5^2 u^2

$$\text{c) Si } n = 25; V(\bar{X}) = V(X) / n = 15^2/25 = 9 \text{ u}^2 \rightarrow S_{\bar{X}} = 3 \text{ u}$$

$$[\bar{X}_i - E(X)]^2 = [104 - 100]^2 = 4^2 \text{ u}^2 \text{ cifra «aceptablemente típica»,}$$

ya que su valor esperado es 3^2 u^2

$$\text{Si } n = 225 V(\bar{X}) = V(X) / n = 15^2/225 = 1^2 \text{ u}^2 \rightarrow S_{\bar{X}} = 1 \text{ u}$$

$$[\bar{X}_i - E(X)]^2 = [104 - 100]^2 = 4^2 \text{ u}^2 \text{ no es una cifra «aceptablemente típica»,}$$

ya que su valor esperado es 1 u^2 (se sospecha que éstos no son de aquella población).

Capítulo 6:

Estimación por intervalo

Presentación

En este capítulo el lector aprenderá a construir e interpretar intervalos de confianza, que informan sobre los valores razonables de los parámetros de acuerdo con la información muestral. Dado que el parámetro representa un valor poblacional, se pretende conocer verdades absolutas y dar respuestas universales. Verdades universales, aunque reducidas a las condiciones y características del estudio.

En resumen, los intervalos de confianza que el lector aprenderá en este capítulo permiten cuantificar el conocimiento.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Percibirá la necesidad de calcular intervalos de confianza para presentar sus resultados.
- Distinguirá entre error típico de estimación e intervalo de confianza.
- Interpretará la mayor amplitud de un intervalo de confianza como menor cantidad de información empírica.
- Interpretará el intervalo de confianza como el rango de valores del parámetro compatibles con la información muestral.
- Sabrá calcular el intervalo de confianza de una media en muestras grandes o de variables con distribución normal.
- Sabrá que, si utiliza el estimador de la varianza en la estimación de la media, precisará emplear la distribución t de Student.
- Sabrá calcular el tamaño muestral necesario para estimar una media con una precisión predeterminada.
- Sabrá calcular el intervalo de confianza de una proporción en muestras grandes.
- Sabrá calcular el tamaño muestral necesario para estimar una proporción con una precisión predeterminada.
- Sabrá calcular el intervalo de confianza de un riesgo atribuible en muestras grandes.
- Sabrá calcular el intervalo de confianza de un riesgo relativo en muestras grandes.
- Sabrá calcular el intervalo de confianza de una *odds ratio* en muestras grandes.
- En su valoración de la calidad de un original científico, exigirá que incluya intervalos de confianza de los resultados relevantes.
- Cuantificará la cantidad de información aportada por un original científico por (el inverso de) la amplitud del intervalo de confianza de su objetivo principal.

Estimación por intervalo

En el capítulo 5 se propuso usar el valor de la media muestral como estimador puntual del correspondiente parámetro poblacional, lo que venía avalado por ser la media muestral un estimador insesgado. También se vio el uso de la distribución normal para saber, a partir de los valores poblacionales de la media $[E(X) \text{ o } \mu]$ y la varianza $[V(X)]$, dónde se encontrarían el 95% de las posibles medias muestrales obtenibles en infinitas muestras. Pero la pregunta de interés práctico va justo en la dirección contraria: conocidos los estimadores muestrales de la media (\bar{X}) y la varianza (S^2), ¿qué se sabe sobre la media poblacional $E(X) = \mu$?

Ejemplo 6.1



Dado que se conoce la distribución de las oscilaciones de las medias \bar{X} de una muestra a otra, se puede construir un intervalo que contenga el porcentaje $1 - \alpha$ de dichas medias \bar{X} alrededor de su media poblacional μ (fig. 6-1):

Intervalo $1 - \alpha$ de $\bar{X}_n = \mu \pm Z_{1-\alpha/2} \sigma/\sqrt{n}$

Para que este intervalo contenga el 95% de las medias muestrales, el valor de la distribución normal debe ser

$Z_{\alpha/2} = Z_{0,025} = 1,96$

Intervalo 95% de $\bar{X}_n = \mu \pm 1,96 \sigma/\sqrt{n}$

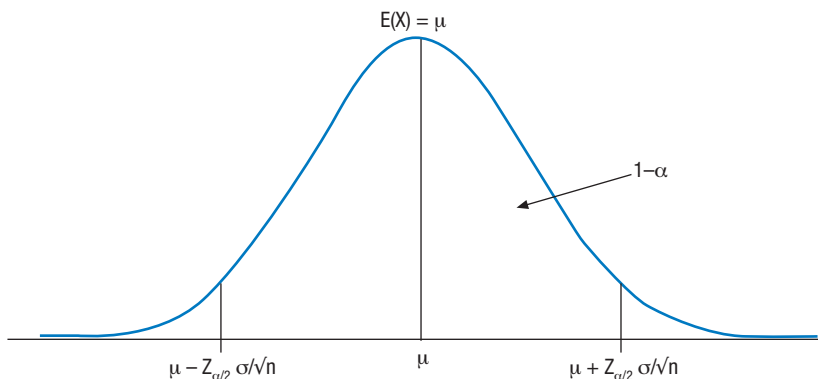
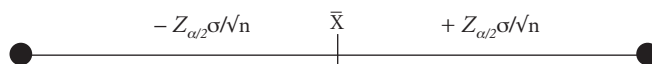


Figura 6-1 Distribución del estimador \bar{X} alrededor del parámetro $E(X) = \mu$. Una proporción $1 - \alpha$ de las posibles medias muestrales \bar{X} está incluida entre los límites indicados.

Comentario



En el gráfico de la fig. 6-1, la distancia es exactamente:



Pero el problema real es el contrario: se conoce \bar{X} y se quiere estimar μ . Se desea disponer de un intervalo que informe, con una certeza cuantificable, dónde se encuentra el valor del parámetro que se estima. Para construirlo, se añaden los valores $\pm Z_{\alpha/2} \sigma / \sqrt{n}$ alrededor de la media muestral \bar{X} .

El gráfico 6-2 muestra el resultado de añadir esta distancia alrededor de 7 posibles medias \bar{X}_i observadas en 7 muestras. Nótese que en las medias observadas en las muestras 1 a 5 (\bar{X}_1 a \bar{X}_5), el intervalo dibujado pasa por encima del punto que representa el valor del parámetro μ : estos intervalos contienen el parámetro de interés. Lo mismo sucedería con todas las medias muestrales contenidas entre los límites v , w , que son el 95% central de las posibles medias muestrales.

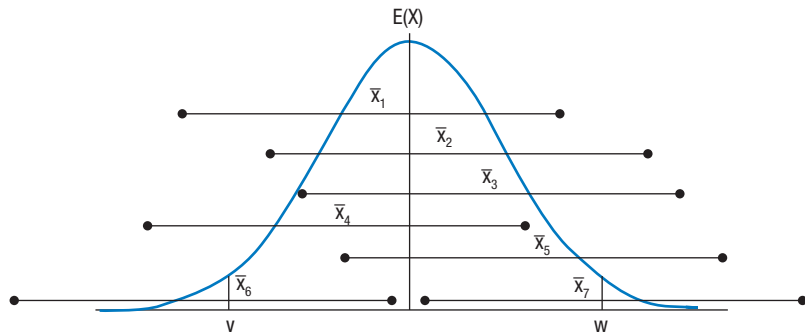


Figura 6-2 Siete posibles medias muestrales y sus respectivos IC. Las 5 primeras \bar{X} conducen a IC que incluyen el parámetro $E(X)$, pero no las 2 últimas.

En cambio, los intervalos de las medias de las muestras 6 y 7 (\bar{X}_6 y \bar{X}_7) no contienen el parámetro. Representan a ese $5\% = \alpha$ de posibles muestras que fallarían en su apreciación. El intervalo así construido tiene, por lo tanto, un 95% ($1 - \alpha$) de posibilidades de contener el parámetro poblacional, por lo que recibe el nombre de **intervalo de confianza del 95% ($1 - \alpha$)**. Figura 6-2.

Este « $1 - \alpha$ » representa el porcentaje de ocasiones en las que se desea que el intervalo obtenido contenga el parámetro de interés. Normalmente se acepta que el intervalo sea del 95%, pero si en una situación particular se desea aumentar la **cobertura** al 99% o al 99,9%, simplemente se trata de sustituir el 1,96 anterior por los correspondientes valores de tablas (2,58 y 3,29).

Comentario



Observe que, si desea aumentar la cobertura, precisa ampliar el intervalo.

Definición



El **intervalo de confianza $1 - \alpha$ ($IC_{1-\alpha}$)** de $\mu = E(X)$ conocida σ es:
 $IC_{1-\alpha} \mu = \bar{X} \pm Z_{\alpha/2} \sigma_{\bar{X}} = \bar{X} \pm Z_{\alpha/2} \sigma / \sqrt{n}$

Nota técnica

El intervalo estudiado se basa en el teorema del Límite Central, que establece las condiciones de convergencia hacia la distribución normal centrada y reducida $N(0,1)$ del estadístico

$$\hat{Z} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \rightarrow N(0,1)$$

Este estadístico \hat{Z} parte de la media muestral \bar{X} , la centra (al restarle su media poblacional) y la reduce (al dividirla por su dispersión poblacional). Es decir, aplica constantes a la variable \bar{X} : cambia la escala y el origen de la v.a. pero no modifica en absoluto la forma de su distribución.

Comentario

Para poder afirmar que la variable media muestral \bar{X} , una vez centrada y reducida, sigue una normal centrada y reducida es preciso previamente que X sea normal, lo que sucede cuando o bien n es grande ($n \geq 30$) o bien X es normal.

Recuerde

Además de conocer la dispersión poblacional σ , la fórmula anterior requiere que: $n > 30$ o $X \rightarrow N$.

Ejemplo 6.2

(Prestado del control de calidad y de la vida misma): La asociación de usuarios (ASU) sospecha que las gasolineras no sirven la cantidad pactada. Por ley, se acepta que el dispensador tenga un error $\sigma = 10$ ml por cada litro que expende. En una muestra al azar de $n = 100$ pedidos de 1 litro (¡qué poco suspicaz el dependiente!), la media observada ha sido $\bar{X} = 995$ ml. El intervalo de confianza del 95% de μ vale:

$$\begin{aligned} IC_{95\%} \mu &= \bar{X} \pm Z_{\alpha/2} \sigma/\sqrt{n} = \\ &= 995 \pm 1,96 \cdot 10/\sqrt{100} = \\ &= 995 \pm 1,96 = \\ &= [993,04, 996,96] \end{aligned}$$

Por tanto, se cree, con una confianza del 95%, que la auténtica media poblacional de esta máquina está entre 993 ml y 997 ml.

Ejemplo 6.3

Se sabe que la glucemia en mmol/l tiene una desviación típica igual a 1. En una muestra de 9 pacientes, la media ha sido de 5.

$$IC_{95\%} \mu = \bar{X} \pm Z_{\alpha/2} \sigma/\sqrt{n} = 5 \pm 1,96 \cdot 1/\sqrt{9} \approx 5 \pm 2/3 \approx [4,33, 5,67]$$

Se cree, con una «fuerza» del 95%, que el auténtico valor poblacional se encuentra entre estos límites.

Nota técnica

Un 95% de confianza significa que cada 20 estudios que se realicen, 19 contendrán el parámetro de interés y 1 no lo hará.

Ejercicio de Navegación

En la página que se indica a continuación, con la ayuda de la simulación que realiza la aplicación, observe los intervalos que aparecen en la estimación de un parámetro concreto:

http://www.ruf.rice.edu/~lane/stat_sim/conf_interval/index.html

Ejercicio 6.1

Sin cambiar la confianza, ¿cómo podría reducir el intervalo del ejemplo 6.3 a la mitad?

Ejercicio 6.2

Con los datos del ejemplo 6.3, calcule el IC para una confianza del 99%.

Ejercicio 6.3

Al final, ¿el $IC_{95\%}$ contiene o no contiene μ ?

Ejercicio 6.4

El intervalo de confianza del 99% (elija una):

- incluye el 99% de las medias poblacionales
- incluye el 99% de las medias muestrales
- incluye la media poblacional el 99% de las ocasiones
- incluye la media muestral el 99% de las ocasiones.

Ejercicio 6.5

Con un intervalo de confianza ($1 - \alpha = 95\%$) podemos afirmar que (elija una):

- a) el 95% de los casos están dentro del intervalo
- b) si se repitiera el proceso, el 95% de los casos estarían dentro del intervalo
- c) hay una probabilidad del 5% de que el parámetro no esté en el intervalo
- d) hay una confianza del 95% de que el parámetro esté en el intervalo.

Comentario

En la estadística clásica, habitual, el parámetro es una constante, no una variable aleatoria. Por ello, se evita hablar de un intervalo de probabilidad del parámetro y se usa el término de confianza. La perspectiva bayesiana, en cambio, sí que acepta que el conocimiento sobre el parámetro pueda expresarse en términos de probabilidad. Desde la perspectiva frecuentista habitual sólo puede usarse probabilidad en lugar de confianza si queda claro que las variables aleatorias son los extremos del intervalo. En otras palabras, no se dice que entre los límites a y b del intervalo se encuentre un parámetro «flotante» con alta probabilidad, como si a y b fueran fijos, sino que el procedimiento del IC garantiza con alta probabilidad que el parámetro esté entre los dos valores aleatorios a y b .

Nótese que esta fórmula para calcular el $IC_{95\%}\mu$ utiliza σ , lo que implica que para poder estimar la media poblacional necesita conocer previamente la varianza de la variable. Esta situación puede darse en alguna ocasión, pero no es, ni mucho menos, una situación general.

Recuerde

Esta fórmula para estimar μ sólo sirve si previamente se conoce σ .

Ejemplo 6.4

Se sabe que la distribución de cierto parámetro sanguíneo sigue una distribución $N(\mu, \sigma)$. Por un cambio del procedimiento analítico, se incrementan sus valores en una cierta constante K . Se puede asumir que el nuevo valor siga una $N(\mu', \sigma)$, que tenga una media desconocida y una varianza conocida.

Distribución τ de Student

Introducción a la t de Student

Ahora ya no nos basaremos en: $\hat{Z} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \rightarrow N(0,1)$

sino en: $\hat{Z} = \frac{\bar{X} - \mu}{\sqrt{s^2/n}} \rightarrow t_{n-1}$

Sustituir el parámetro σ por el estadístico S implica sustituir una constante, que tiene un único valor, por una variable aleatoria, que tiene toda una distribución de valores. Así, sustituir σ por S implica pagar el precio de recurrir a una nueva distribución: la t de Student.

Lectura



Gosset era un estadístico (37) que trabajaba para la cervecera Guinness realizando estudios de calidad de sus cervezas. Por ejemplo, de su contenido de alcohol. Para poder detectar y rechazar los lotes de cerveza que no cumplieran con las especificaciones deseadas, él había aceptado el coste de rechazar un 5% de los lotes que sí que las cumplieran. Y rechazaba aquellas muestras en las que el valor resultante se encontraba por fuera de los límites $-1,96, +1,96$.

Pronto empezó a sospechar que estaba desechando demasiados lotes de cervezas. Y se aperció de que por fuera de los límites $-1,96, +1,96$ se encontraban más del 5% de los lotes correctos. Cayó en la cuenta de que S era un estadístico y no un parámetro y propuso una distribución algo más aplanada que la normal, en la que observó que rechazaba el $\alpha\%$ deseado de lotes correctos. Esta distribución se aplanaba y alejaba más de la distribución normal cuanto menos precisa fuera la estimación S , es decir, cuanto más pequeña fuera la muestra, lo que solía ser su caso, ya que, al tener que destruir las unidades que analizaba, solía usar muestras de 5 o 6 unidades.

Como no le permitieron firmar con su nombre el artículo en el que publicaba sus resultados, lo firmó con el seudónimo de «estudiante», razón por la que se conoce la distribución que propuso como la distribución t de Student.

Nota técnica



Los grados de libertad o información «neta» de una muestra viene dada por el número de observaciones (independientes) menos las preguntas que previamente ha debido contestar. Por ejemplo, si para calcular S^2 en una muestra de n casos primero se ha debido estimar 1 parámetro, μ , mediante \bar{X} , los grados de libertad que tiene esta estimación de la varianza son « $n - 1$ ».

Nota técnica



Un sistema de n ecuaciones (piezas de información) con k incógnitas tiene $n - k$ grados de libertad.

Ejemplo 6.5



Uso de la tabla 6-1. Si se desea contener el 95% de las observaciones y dejar un 5% repartido simétricamente en los dos lados (α), debe buscarse en la columna del 0,05. Si además se dispone de una muestra de 15 casos, debe buscarse en la fila de los 14 grados de libertad, donde se encuentra el valor 2,145.

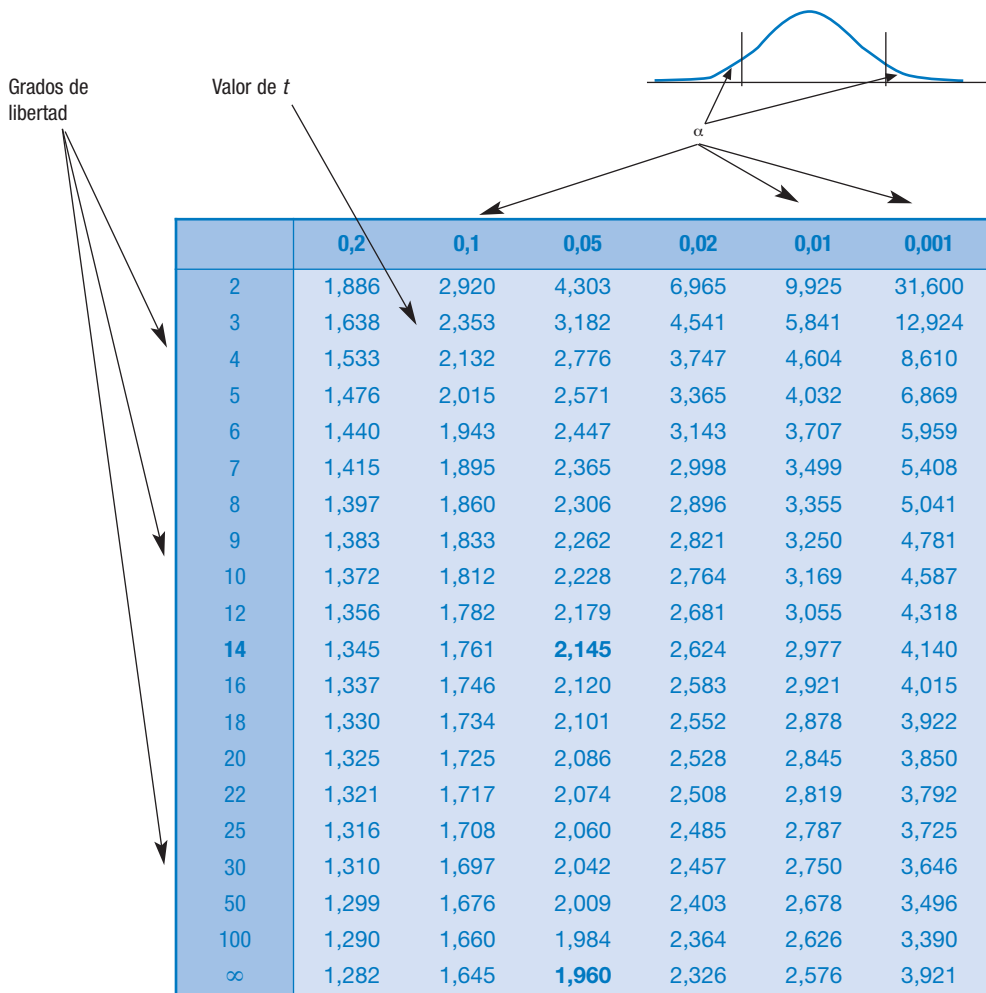


Tabla 6-1 Tabla bilateral de la t de Student-Fisher.

Ejemplo 6.6

Si se dispusiera de una muestra de 500 casos, dado que la tabla no lo contiene se debería utilizar el valor de la fila anterior o posterior. Este último ofrece, precisamente, el conocido valor 1,96. Por eso, en muestras grandes, suele utilizarse directamente el valor de la distribución normal.

Comentario

Por brevedad, la tabla no contiene más que una serie limitada de valores. Puede recurrirse a cualquier hoja de cálculo o programa de estadística para obtener los valores exactos. Por ejemplo, para 499 grados de libertad, Excel proporciona el valor 1,971 para un 5% bilateral.

Intervalo de confianza de μ usando S

Dada la simetría de la distribución t de Student, el intervalo de confianza $1 - \alpha$ de μ será:

$$IC_{1-\alpha} \mu = \bar{x} \pm t_{n-1, \alpha/2} S_{\bar{x}} = \bar{x} \pm t_{n-1, \alpha/2} S/\sqrt{n}$$

Ejemplo 6.7

El tiempo utilizado en la atención al paciente sigue una distribución normal. Para conocer el tiempo medio empleado en este servicio, se han recogido 20 observaciones que han tardado, en minutos, $\bar{X} = 34$ y $S = 2,3$. Como el enunciado dice que X es normal, se puede realizar el cálculo usando la t de Student.

[Nota: con Excel se obtiene «= distr.t.inv(0,05; 19)» = 2,093.]

$$IC_{0,95} \mu = \bar{x} \pm t_{19, 0,025} S/\sqrt{n} = 34 \pm 2,093 \cdot 2,3 / \sqrt{20} \approx 34 \pm 1,08 = [32,92, 35,08]$$

Se cree, con una confianza del 95%, que la media poblacional del tiempo de atención se sitúa entre 32,92 y 35,08 min.

Recuerde

La amplitud del intervalo hace referencia a nuestra ignorancia sobre dónde se encuentra el único y auténtico valor de la media, no a que esta media poblacional pueda ser más de un valor ni que esté oscilando.

Premisas para estimar μ sin conocer σ

Lectura



Para referirse al término inglés assumptions, diferentes autores utilizan diferentes vocablos: asunciones, hipótesis previas necesarias, requisitos, condiciones de aplicación, etc. Nosotros usaremos «premisas».

Para poder afirmar que el estadístico t siga una distribución t de Student con $n - 1$ grados de libertad, la premisa es que la variable en estudio sigue una distribución normal. Ahora bien, aunque no sea normal, si el tamaño muestral crece, la estimación S^2 de σ^2 es mejor, acercándose al valor real, por lo que la sustitución de σ^2 por S^2 tiene menores implicaciones. Por esta razón, aunque la variable estudiada no sea normal, en tamaños muestrales grandes puede asumirse que el estadístico t se acerca a la normal.

Recuerde



Para poder usar la t de Student se requiere o bien distribución normal de la variable en estudio, o bien tamaño grande.

Lectura



¿Qué significa tamaño grande? ¿Por qué unos autores dicen 20, otros 30 y otros 100? ¿Hay algún número mágico que cambie tanto la forma de la distribución? No, se trata de una aproximación sucesiva y que, recuérdese, sucede antes cuanto más se asemeje a la normal la distribución de la variable en estudio. Además, también es anterior para el estadístico Z (que usa σ) que para el estadístico t (que usa S).

Recuerde



Tamaño «grande» suele considerarse a partir de 30 casos.

Así pues, se sabe cómo inferir los resultados de la muestra a la población si se dispone de una variable con distribución normal o bien si la muestra es suficientemente grande. Estas fórmulas deben servir para solucionar la gran mayoría de las situaciones.

Comentario



La distribución normal de Gauss-Laplace es la distribución de los errores de medida, o también, la distribución de la variabilidad natural. En cualquier caso, variabilidad no explicable por causas mayores. Por tanto, se trabajará con la distribución normal cuando se pueda creer que ya se han considerado (o medido o controlado) las fuentes importantes de variación. Por ello, encontrar una distribución normal permite creer, al menos tentativamente, que las fuentes importantes de variabilidad ya han sido identificadas y, por tanto, sólo quedan las fuentes menores, no identificables, que se consideran aleatorias. En otras palabras: por un lado, se controlan las fuentes importantes de variación; y por otro, se cuantifican, con la ayuda de la distribución normal, las fuentes menores.

Ejercicio 6.6



En una muestra de 100 pacientes con infarto, se han valorado las GOT a las 12 h. La media ha sido de 80 y la desviación típica de 120. Haga un $IC_{95\%}$ de la media.

Se ha justificado un tamaño muestral mayor que 30 para poder utilizar una fórmula estadística. Dado que existen fórmulas y procedimientos alternativos, el tamaño muestral no debe fijarse por este criterio, sino por la cantidad de información que se desea finalmente disponer.

Ejercicio 6.7



Asumiendo que la desviación típica poblacional de las GOT es de 120 u, ¿cuántos casos se necesitan para:

- tener un error típico de estimación de la media poblacional igual a 12 u?
- tener una amplitud (incertidumbre) total del $IC_{95\%}$ de la media poblacional igual a 12 u?

Lectura



En el caso de que no se disponga de una muestra grande ni de una variable con distribución normal se puede recurrir a dos grandes grupos de soluciones: a) recurrir a procedimientos estadísticos que no requieren esta distribución (cálculos exactos o por re-muestreo, principalmente), y b) transformar la variable para conseguir su normalidad. Existen varias transformaciones que funcionan muy bien en la práctica. Para variables positivas (como «el tiempo hasta...» o «el nivel de GOT») la transformación logarítmica suele corregir su habitual asimetría y conseguir distribuciones muy parecidas a la normal. Por otro lado, si se dispone de un recuento de fenómenos raros, de baja probabilidad, que suelen seguir una distribución de Poisson, la transformación raíz cuadrada puede funcionar bien.

Intervalo de confianza de la probabilidad π

Se estudia a continuación el caso de variables dicotómicas, como por ejemplo, el hecho de padecer, o no, cierto acontecimiento adverso (AA). La mejor forma de resumir esta variable es mediante la proporción p de pacientes que han experimentado dicho AA. Esta proporción p obtenida en la muestra permitirá estimar la probabilidad π de que un nuevo paciente de las mismas características presente dicho AA.

Nota técnica



p es un estimador insesgado de π ya que $E(p) = \pi$. Y es convergente, ya que su varianza depende inversamente del tamaño muestral: $V(p) = \pi \cdot (1 - \pi)/n$. Es decir, si definimos el error típico de estimación como la raíz de la variable $V(p)$, la dispersión del estimador de la probabilidad va disminuyendo a medida que aumenta el tamaño de la muestra, de forma proporcional a la raíz del tamaño muestral.

Recuerde



El error típico del estimador p cuantifica su distancia

esperada al parámetro π y vale. $\sqrt{\frac{\pi(1-\pi)}{n}}$

Aunque se puede proceder a un procedimiento «exacto» basado directamente en el cálculo de probabilidades, es más cómodo recurrir a la aproximación a la distribución normal de la distribución del estimador p , $p \rightarrow N(\pi, \pi \cdot (1 - \pi)/n)$.

Comentario



Formalmente, el número o recuento de casos de una muestra aleatoria que tienen una cierta característica (y por tanto, también, la proporción observada) es una variable que sigue la distribución binomial. Pero la binomial puede aproximarse de forma muy razonable mediante la distribución normal, lo que hace muy cómodo el cálculo del intervalo de confianza de π .

Ejercicio de Navegación



En los programas de simulación puede verse que la aproximación de la binomial a la normal es tanto mejor cuanto mayor es el número de observaciones y más alejado de 0 y de 1 está el valor de π . Compruébelo en:

http://www.ruf.rice.edu/~lane/stat_sim/Normal_approx/index.html

Nota técnica



Observe que, en una binomial, debería dar los mismos resultados, una estimación de la probabilidad π de éxito que de la probabilidad $1 - \pi$ de fracaso. O de la proporción poblacional o de hombres o de mujeres. Por ello, π y $1 - \pi$ tienen un papel simétrico, por lo que la condición de que π no sea muy pequeña también aplica a $1 - \pi$.

Recuerde



Se aceptan como **condiciones de aplicación** de la aproximación normal que el tamaño muestral sea grande y la probabilidad π y $1 - \pi$, no extrema
 $[\pi \cdot n \geq 5$ y $(1 - \pi) \cdot n \geq 5]$

Utilizando la distribución normal, el cálculo del IC es casi idéntico al de μ :

$$IC_{1-\alpha} \pi = P \pm Z_{\alpha/2} \sigma_p = P \pm Z_{\alpha/2} \sqrt{[\pi \cdot (1 - \pi)]/n}$$

Comentario



¿Recuerda aquella situación paradójica en la que para estimar μ era necesario conocer σ ? Pues ahora se ha superado: se necesita conocer π para poder estimar la variabilidad de p que, a su vez, es necesaria para poder estimar π . Hay dos posibles soluciones.

1) La primera solución viene de que el producto $\pi \cdot (1 - \pi)$ tiene un máximo cuando $\pi = 0,5 = 1 - \pi$ (tabla 6-2).

Se puede, por lo tanto, adoptar una actitud conservadora y decir que, en una muestra de tamaño n , la dispersión del estadístico p vale, como mucho: $\sigma_p = \sqrt{[\pi(1 - \pi)]/n} = \sqrt{[0,5(1 - 0,5)]/n} = 0,5/\sqrt{n}$

Por lo que el cálculo del intervalo de confianza $1 - \alpha$ de π es:

$$IC_{1-\alpha} \pi = p \pm Z_{\alpha/2} \sigma_p = p \pm Z_{\alpha/2} \sqrt{[0,5 \cdot (1 - 0,5)]/n} = p \pm Z_{\alpha/2} 0,5/\sqrt{n}$$

2) La segunda solución consiste en sustituir π por p , tal como se hizo con σ^2 por S^2 . Ahora, el cálculo del intervalo de confianza $1 - \alpha$ de π es:

$$IC_{1-\alpha} \pi = p \pm Z_{\alpha/2} \sigma_p = p \pm Z_{\alpha/2} \sqrt{[p(1 - p)]/n}$$

Recuerde



En la práctica, para calcular el intervalo de confianza de π , se sustituye, en la fórmula del error típico de estimación, π por $0,5$ o por p :

$$IC_{1-\alpha} \pi = p \pm Z_{\alpha/2} \sigma_p = p \pm Z_{\alpha/2} \sqrt{[0,5 \cdot (1 - 0,5)]/n}$$

$$IC_{1-\alpha} \pi = p \pm Z_{\alpha/2} \sigma_p = p \pm Z_{\alpha/2} \sqrt{[p(1 - p)]/n}$$

π	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$1 - \pi$	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1
$\pi (1 - \pi)$	0,09	0,16	0,21	0,24	0,25	0,24	0,21	0,16	0,09

Tabla 6-2 Ejemplo ilustrativo de que el máximo de $\pi (1 - \pi)$ es para $\pi = 0,5$.

Ejemplo 6.8

Si se lanza 100 veces una moneda al aire y se observan 56 caras:

$$IC_{95\%} \pi = p \pm Z_{\alpha/2} \sqrt{[0,5 \cdot 0,5/n]} = 0,56 \pm 1,96 \cdot 0,5/\sqrt{100} \approx 0,56 \pm 0,10 = [0,46, 0,66]$$

Por lo que se cree, con una confianza del 95%, que la probabilidad de que salga cara en esta moneda es uno de los valores comprendidos entre 0,46 y 0,56.

Y de acuerdo con el segundo procedimiento:

$$IC_{95\%} \pi = p \pm Z_{\alpha/2} \sqrt{[p(1 - p)/n]} = p \pm Z_{\alpha/2} \sqrt{[0,56 \cdot 0,44 / n]} = 0,56 \pm 0,10 = [0,46, 0,66]$$

Puede verse que, al nivel de precisión habitual, ambos procedimientos conducen a un intervalo idéntico.

Nota técnica

Se da esta coincidencia de resultados porque, en este ejemplo, p se encuentra muy cerca de 0,5. Si se estuviera estimando un fenómeno más raro, con una π alejada de 0,5, la concordancia entre ambos procedimientos sería menor.

Comentario

Puede decirse que $\sqrt{(0,5 \cdot 0,5/n)} = 0,5/\sqrt{n}$ es el valor del error típico de p en la situación de máxima indeterminación. Tiene la ventaja de que, dado cierto tamaño muestral, se dispone del mismo valor para cualquier proporción que se desee estimar. Por lo tanto, si se realiza una encuesta con muchas preguntas o si se está estudiando una variable con varias categorías (p. ej., en la intención de voto) se puede usar el mismo valor de σ_p para cada una de ellas.

Ejercicio 6.8

Se dispone de una población, pongamos que infinita, de preguntas tipo test. Para un examen, se seleccionan al azar 30 preguntas y un alumno contesta bien 18 de ellas. Como el interés del evaluador es conocer la proporción de preguntas de la población (no de esta muestra de 30 preguntas) que conoce este alumno ¿qué se sabe sobre la proporción poblacional de preguntas que conoce el alumno?

Comentario

Una variable dicotómica del tipo «cura o no cura» es la que más simplifica la evolución del paciente y, en consecuencia, se trata de la que obtiene menos información. Por ello, para un mismo número de casos, el intervalo de confianza tendrá una amplitud que sorprenderá por su magnitud.

Recuerde

Los intervalos de confianza de una probabilidad (o una proporción poblacional) suelen ser amplios: requieren una n más elevada para poder proveer una información razonable.

Ejercicio 6.9

En un megaensayo, de los primeros 160 cuadernos de recogida de datos (CRD), 34 presentan una infracción mayor del protocolo en la primera visita. Calcule el $IC_{95\%}$ de la probabilidad de que un CRD tenga esta condición.

Ejercicio 6.10

¿Qué amplitud máxima tiene el $IC_{95\%}(\pi)$ de la proporción poblacional de pacientes con AA si $n = 100$? ¿Y si $n = 400$? ¿Y si $n = 2.500$? ¿Y si $n = 10.000$?

Ejercicio 6.11

¿Qué relación hay entre la amplitud del intervalo de confianza de π y n ?

Ejercicio 6.12

De un total de 100 médicos, 40 prescriben cierto fármaco. Calcule el $IC_{95\%}$ de la proporción poblacional de médicos que lo prescriben. ¿Algún comentario sobre cómo deberían haber sido seleccionados estos médicos?

Ejercicio 6.13

Situándonos en el caso de mayor variabilidad o incertidumbre ($\pi = 1 - \pi = 0,5$), ¿cuántos casos se necesitan para:

- estimar una proporción con un error típico de 0,05?
- estimar una proporción con un $IC_{95\%}$ de amplitud total de 0,05?

Intervalos de confianza de las medidas de asociación

Riesgo atribuible (RA)

Se definió el riesgo atribuible como la diferencia entre la probabilidad de que un caso expuesto al factor desarrolle la enfermedad y la misma probabilidad en un caso no expuesto al factor (diferencia de riesgo entre expuestos y no expuestos).

	Y+	Y-	TOTAL
X+	7	125	132
X-	8	860	868
TOTAL	15	985	1000

Tabla 6-3 Presencia de la enfermedad Y y el factor de riesgo X en 1.000 casos.

Ejemplo 6.9



La tabla 6-3 reproduce la tabla 3-4, en la que la estimación muestral p de la probabilidad en los expuestos era 5,3% [$P(Y+|X+) = 7/132 \approx 0,053$] mientras que en los no expuestos era 0,9% [$P(Y+|X-) = 8 / 868 \approx 0,009$].

La diferencia entre 0,053 y 0,009 es 0,044, es decir, expresado en porcentajes, un 4,4%.

Definición



El **error típico del RA** es $\sqrt{p_1 \cdot (1 - p_1)/n_1 + p_2 \cdot (1 - p_2)/n_2}$
 $IC_{95\%} RA = RA \pm Z_{\alpha/2} \sigma_{RA} = RA \pm Z_{\alpha/2} \sqrt{p_1 \cdot (1 - p_1)/n_1 + p_2 \cdot (1 - p_2)/n_2}$

Nota técnica



La varianza de la estimación de la diferencia de dos probabilidades (estimadas en muestras independientes) se convierte en la suma de las varianzas de la estimación de cada una. Esto se explica porque ambos ruidos de estimación se añaden en el momento de querer estimar la diferencia.

El **requisito** (Agresti, 38) para poder aplicar esta fórmula es que el tamaño muestral sea grande. Por dar unas cifras «mágicas» de referencia, las frecuencias de las celdas de la tabla 2×2 deberían ser superiores a 3 y el tamaño total de la tabla, a 100.

Ejemplo 6.10




En los datos del ejemplo, el $IC_{95\%} RA$ vale
 $IC_{95\%} RA = RA \pm Z_{\alpha/2} \sqrt{p_1 \cdot (1 - p_1)/n_1 + p_2 \cdot (1 - p_2)/n_2} =$
 $= 0,044 \pm 1,96 \sqrt{0,053 \cdot 0,947/132 + 0,009 \cdot 0,991/868} \approx$
 $= 0,044 \pm 1,96 \cdot 0,0198 = 0,044 \pm 0,0388 = [0,0051, 0,0826] \approx$
 $\approx [0,5\%, 8,3\%]$

Y se concluye, por tanto, que los expuestos al factor tienen un riesgo entre 0,5 y 8,3% superior.

	Y+	Y-	TOTAL
X+	94	38	132
X-	215	653	868
TOTAL	309	691	1.000


Tabla 6-4 Datos para los ejercicios 6.14, 6.15 y 6.16.

Nota técnica



No se ha dicho que la exposición al factor aumenta entre un 0,5% y un 8,3% el riesgo para evitar el uso de frases con connotación causal. Véase el tema sobre diseño de investigaciones para aclarar las circunstancias en las que es posible hacer afirmaciones causales.

Ejercicio 6.14




Con los datos de la tabla 6-4, calcule el IC_{95%} RA.

Riesgo Relativo (RR)


Se definió el riesgo relativo como el cociente entre la probabilidad de que un caso expuesto al factor desarrolle la enfermedad y la misma probabilidad en un caso no expuesto al factor (razón entre el riesgo en los expuestos y en los no expuestos).

Ejemplo 6.11




Siguiendo con los datos del **ejemplo 6.9**, la razón entre 0,053 y 0,009 vale 5,7538, es decir, que el riesgo relativo observado es casi 6 veces superior en los expuestos.

Definición



$IC_{95\%} \ln(RR) = \ln(RR) \pm Z_{\alpha/2} \sigma_{\ln(RR)} =$
 $= \ln(RR) \pm Z_{\alpha/2} \sqrt{[(1 - p_2)/n_2 p_2 + (1 - p_1)/n_1 p_1]}$
 El **requisito** para aplicar esta fórmula es, como antes, tamaño muestral grande.

Nota técnica



Este cálculo es ahora más complejo. Dada la asimetría del RR (que oscila entre 0 y 1 para riesgos inferiores en los expuestos y entre 1 e infinito para riesgos superiores), es preciso hacer previamente la transformación logarítmica natural (neperiana) para poder aprovechar la simetría resultante. La varianza del logaritmo del RR tiene ahora la misma interpretación en cualquier sentido.

Nota técnica

La fórmula de la varianza del logaritmo del RR no es inmediata. Es la suma de las varianzas de los logaritmos de las proporciones que son, a su vez, la varianza de la binomial dividida por el cuadrado de la proporción.

Ejemplo 6.12

En los datos del **ejemplo 6.9**, el $RR = 5,7538$

$$\ln(RR) = \ln(5,7538) = 1,7499$$

$$\begin{aligned} IC_{95\%} \ln(RR) &= \ln(RR) \pm Z_{\alpha/2} \sqrt{[(1-p_2)/n_2 p_2 + (1-p_1)/n_1 p_1]} \\ &= 1,7499 \pm 1,96 \sqrt{[0,947/132 \cdot 0,053 + 0,991/868 \cdot 0,009]} \approx \\ &= 1,7499 \pm 1,96 \cdot 0,5090 = 1,7499 \pm 0,9977 \\ &= [0,7521, 2,7476] \end{aligned}$$

Así, se puede afirmar que el valor de $\ln(RR)$ aumenta entre 0,75 y 2,75, lo que resulta prácticamente imposible de interpretar: ¿Qué significa un aumento de $\ln(RR)$ igual a 2,75?

Para facilitar la interpretación se deshace el logaritmo:

$$IC_{95\%} RR = \exp[IC_{95\%} \ln(RR)] = [e^{0,7521}, e^{2,7476}] \approx [2,1, 15,6]$$

Por lo que se concluye que los expuestos tienen un riesgo que es entre 2,1 y 15,6 veces superior, lo que resulta más fácil de interpretar: sea cual sea el riesgo en los no expuestos, en los expuestos toma un valor entre 2,1 y 15,6 veces superior.

Comentario

Una vez más se ha evitado la connotación causal de frases como «la exposición al factor aumenta el riesgo entre 2,1 y 15,6 veces», o «el hecho de estar expuestos multiplica el riesgo entre 2,1 y 15,6 veces».

Nota técnica

$\exp[Y] = e^Y$ indica el número $e = 2,7183$ elevado al número Y . Las operaciones matemáticas \exp y \ln (logaritmo natural o neperiano) son inversas: $e^{\ln(y)} = Y$; $\ln(e^Y) = Y$. El lector no debe desanimarse por la aparición de unos logaritmos a los que no está habituado. Piense que son tan sólo un instrumento para dar simetría a los RR y poder calcular con el mismo procedimiento ambos límites. Permiten, por tanto, un cálculo más simple. Si el «pánico al logaritmo» persiste, practique un poco las funciones \exp y \ln de su calculadora o de su hoja de cálculo.

Comentario

Nótese que el intervalo del RR es asimétrico alrededor de la estimación puntual 5,75.

Ejercicio 6.15



Con los datos del ejercicio 6.14 calcule el IC del RR.

Odds Ratio (OR)

Se definió la *odds ratio* como el cociente entre las *odds* (o razones si/no) de desarrollar la enfermedad entre los expuestos y los no expuestos.

Ejemplo 6.13



Siguiendo con los datos de la tabla 6.3, las *odds* respectivas son 0,056 y 0,009 y su razón vale 6,0200, es decir, que la razón enfermo/sano es 6 veces superior en los expuestos.

Como con el riesgo relativo, la asimetría del OR aconseja emplear la transformación logarítmica.

Comentario



La fórmula de la varianza del logaritmo de la OR es la suma de las inversas de las frecuencias observadas en las cuatro casillas de la tabla 2 x 2: $1/a + 1/b + 1/c + 1/d$.

Nota técnica



Se obtiene asumiendo estimaciones de Poisson independientes en las cuatro celdas.

Definición



$IC_{95\%} \text{Ln(OR)} = \text{Ln(OR)} \pm Z_{\alpha/2} \sigma_{\text{Ln(OR)}}$
 $= \text{Ln(OR)} \pm Z_{\alpha/2} \sqrt{1/a+1/b+1/c+1/d}$
 El **requisito** es tamaño muestral grande.

Ejemplo 6.14



En los datos del ejemplo, la $OR = (7/125)/(8/860) = 6,0200$
 $\text{Ln(OR)} = \text{Ln}(6,0200) = 1,7951$
 $IC_{95\%} \text{Ln(OR)} = \text{Ln(OR)} \pm Z_{\alpha/2} \sqrt{1/a+1/b+1/c+1/d} =$

Ejemplo 6.14 (Cont.)

$= 1,7951 \pm 1,96 \sqrt{[1/7+1/125+1/8+1/860]} \approx$
 $= 1,7951 \pm 1,96 \cdot 0,5263 = 1,7951 \pm 1,0316 = [0,763, 2,826]$
 Y para facilitar la interpretación se deshace el logaritmo:
 $IC_{95\%} OR = \exp[IC_{95\%} \ln(OR)] = [e^{0,763}, e^{2,826}] \approx [2,1, 16,9]$
 Por lo que se concluye que los expuestos tienen una razón enfermo/sano que es entre 2,1 y 16,9 veces superior.

Nota técnica

Como siempre, se ha evitado hablar de efecto causal con frases como «el factor multiplica la razón enfermo / sano entre 2,1 y 16,9 veces».

Comentario

El intervalo de la OR también es asimétrico alrededor de la estimación puntual 6,02.

Observe que los IC del RR y del OR son muy similares.

Lectura

Serra-Prat et al. (39). «Si agrupamos las distintas categorías de la variable origen en dos categorías (autóctonos e inmigrantes), observamos una asociación estadísticamente significativa entre el déficit de yodo y el origen (OR = 2,88; IC_{95%}, 1,33- 6,12).»

Ejercicio 6.16

Con los datos del ejercicio 6.14 calcule el IC del OR.

Resumen de intervalos de confianza

Los intervalos de confianza constituyen la herramienta de inferencia más relevante y fácilmente comunicable. Tanto es así, que las revistas biomédicas más importantes aconsejan basar la presentación de los resultados del estudio en intervalos de confianza.

Lectura



CONSORT (9). «Para cada resultado primario y secundario, un resumen de los resultados de cada grupo y de la magnitud estimada del efecto y de su precisión (p. ej., intervalo de confianza del 95%).

»Explicación de cada resultado. Los resultados del estudio se deben presentar en forma de resumen dentro de cada grupo (p. ej., la proporción de participantes con o sin una complicación concreta, o bien el valor medio y la desviación estándar de los distintos parámetros), junto con la diferencia observada entre los distintos grupos, denominada magnitud del efecto. En parámetros de carácter binario, la magnitud del efecto podría ser el riesgo relativo, la odds ratio o la diferencia de riesgos; en lo que se refiere a los datos de supervivencia, la magnitud del efecto podría ser el cociente de tasas de riesgo (hazard rate ratio) o la diferencia en los tiempos de supervivencia mediana; finalmente, con respecto a los datos continuos, la magnitud del efecto se determina generalmente a través de la diferencia entre los valores medios. Es necesaria la presentación de los intervalos de confianza para la comparación de los distintos grupos. Los resultados obtenidos en el estudio se muestran con mayor claridad en una tabla similar a la 6-5.

»En lo que se refiere a los valores de los distintos parámetros, los autores deben proporcionar los intervalos de confianza para indicar la precisión (incertidumbre) de la estimación. [...] Este parámetro es especialmente útil si las diferencias no son significativas, ya que permite valorar si el resultado descarta una diferencia clínica importante. A pesar de que los valores de P se pueden ofrecer junto a los intervalos de confianza, no es correcto publicar los resultados únicamente en forma de valores de P.

Es necesaria la presentación de los resultados correspondientes a todos los criterios de valoración, principal y secundarios, no solamente de los análisis que muestran significación estadística.»

Criterio de valoración	Grupo Etanercept (n = 30)	Grupo placebo (n = 30)	Diferencia (95% CI)	Valor de P
Principal (%)				
Respuesta a las 12 semanas	26 (87)	7(23)	63(44-83)	<0,001
Secundario (%)				
Pacientes que satisfacen los criterios ACR				
ACR20	22(73)	4(13)	60(40-80)	<0,001
ACR50	15(50)	1(3)	47(28-66)	<0,001
ACR70	4(13)	0	13(1-26)	0,04

Tabla 6-5 Ejemplo de informe de resumen de resultados para cada grupo de estudio.

Ejercicio 6.17



Pongamos que se define el fracaso escolar (FE) como el hecho de no terminar los estudios dentro del plazo previsto más un año de margen (posibles valores: SÍ/NO). Se dispone de un posible predictor dicotómico de FE: notas de entrada (PAU) superiores (S) o inferiores (I) a la media de dicho centro.

- Invente una tabla 2×2 con relación entre FE y PAU.
- Calcule el RA.
- Calcule el RR.
- Calcule el OR y los límites de su intervalo de confianza. (Para agilizar los cálculos y garantizar su exactitud, puede programarlos, p. ej., con la ayuda de Excel.)

	ECV+	ECV-
≥ 25 h	111	87
≤ 10 h	231	261

Tabla 6-6 Enfermedades de la columna vertebral según horas delante del ordenador.

Ejercicio 6.18



El comité de una empresa del sector servicios ha solicitado una compensación económica extraordinaria para los empleados que pasan mucho tiempo delante del ordenador, alegando que este hecho genera enfermedades de la columna vertebral (ECV). Por ello, se ha encargado al servicio médico que se pronuncie sobre este tema.

Vd. forma parte del equipo investigador que debe pronunciarse sobre este tema. Han recogido información sobre ECV en todos los trabajadores de la empresa y comparan los datos de aquellos que pasan más de 25 h a la semana delante del ordenador con los que pasan menos de 10 h. Los datos figuran en la tabla 6-6.

- Usted debe elegir entre una medida de asociación para comparar los riesgos de ambos grupos. Discuta con sus compañeros de equipo, a partir de la nota técnica de la página 65, qué implican los modelos aditivo y multiplicativo que subyacen detrás del riesgo atribuible y del riesgo relativo.
- Calcule el RA.
- Calcule el RR.
- Finalmente han decidido utilizar la medida de asociación más habitual: la *odds ratio*. Calcúlelo junto con su intervalo de confianza al 95%.

Ejercicio 6.18 (Cont.)



e) Interprete el resultado anterior. En concreto, ¿se sostiene que la probabilidad de ECV es la misma en ambos grupos?

Ejercicio 6.19

¿Racismo? En la tabla 6-7 figuran datos —tomados de Bishop et al. (40)— sobre la promulgación de la pena de muerte (P: SÍ/NO) en función de la raza (blanco/negro) del acusado (A) y de la víctima (V).
Estime por intervalo el la *odds ratio* de la relación marginal o global entre la pena de muerte y la raza de la víctima. Interprete el resultado.

	Pena de muerte: Sí		Pena de muerte: No		
	A: Blanco	A: Negro	A: Blanco	A: Negro	
V: Blanco	19	11	V: Blanco	132	52
V: Negro	0	6	V: Negro	9	97

Tabla 6-7 Pena de muerte en función de la raza de la víctima (V) y del acusado (A).

Ejercicio 6.20



En un Centro de Atención Primaria han realizado un experimento para poner a prueba un nuevo tratamiento contra las molestias gástricas «posprandiales» (tras comer). Para ello, han realizado un experimento en el que han asignado al azar: o bien el tratamiento en estudio, o bien un placebo idéntico. El paciente, cuando le aparecía el dolor, en su domicilio, tomaba el tratamiento y anotaba en un cuadernillo si había o no había desaparecido según su propio criterio.
A pesar de que la asignación había sido realizada al azar, la «mala suerte» (o quizás una asignación no enmascarada) quiso que los pacientes quedaran desequilibrados en cuanto a una variable muy importante: el nivel de dolor que sentían al inicio, en el momento en que decidían tomar la medicación; los pacientes con dolor suave fueron mayoritariamente del grupo placebo, mientras que los pacientes con dolor fuerte pertenecían al grupo tratado. En la tabla 6-8 aparecen los resultados obtenidos.
Calcule la *odds ratio* y su intervalo de confianza (95%) para cada una de las tablas: dolor suave, dolor fuerte y todos.

	Dolor suave		Dolor fuerte		Todos	
	Cura	No cura	Cura	No cura	Cura	No cura
Tratados	100	5	650	650	750	655
Placebos	550	80	10	200	560	280

Tabla 6-8 Evolución (cura/no cura) para cada grupo de tratamiento según el nivel basal del dolor.

Soluciones a los ejercicios

6.1 La amplitud del intervalo es lo que en la fórmula va detrás del «±». Por ello, la amplitud depende de 3 valores: $Z_{\alpha/2}$, σ y n . Por el enunciado, no podemos cambiar la confianza γ , por tanto, $Z_{\alpha/2}$ deberá quedar igual. Así pues, sólo se **dispone de σ y de n** para hacer más estrecho el intervalo. Dejemos, por ahora, σ , ya que es la dispersión del fenómeno en estudio, y centrémonos en n . Como está dentro de una raíz cuadrada, para conseguir que el intervalo de confianza sea la mitad de amplio, hay que multiplicar por 4 el tamaño muestral.

$$6.2 \text{ IC}_{99\%} \mu = \bar{X} \pm Z_{0,975} \sigma/\sqrt{n} = 5 \pm 2,576 \cdot 1/\sqrt{9} = 5 \pm 2,576/3 \approx [4,14, 5,86]$$

6.3 No puede saberse si el intervalo contiene realmente μ . Si se repite indefinidamente este procedimiento, el $1 - \alpha$ % de las ocasiones contendrá μ , pero no se puede saber para una vez concreta.

6.4 La respuesta correcta es la *c*), ya que el IC se hace alrededor de la media muestral observada \bar{X} para tener una alta confianza de contener a la (única) media poblacional μ desconocida.

6.5 El IC se no hace referencia a los casos, sino a los parámetros desconocidos, por ello, las respuestas posibles son la *c*) o la *d*), si bien es más correcto formalmente hablar de confianza que de probabilidad (lea el «comentario» que sigue al ejercicio para más explicaciones).

6.6 Dado que la muestra es de 100 casos, no es necesario preguntarse si la distribución es normal (lo que es una suerte, ya que las GOT son positivas, pero una desviación típica mayor que la media implicaría valores negativos en una distribución simétrica como la normal).

$$\text{IC}_{95\%} \mu = \bar{X} \pm t_{99, 0,05} S/\sqrt{n} \approx 80 \pm 1,984 \cdot 120/\sqrt{100} \approx 80 \pm 24 \approx [56, 104]$$

6.7 a) Si $\sigma/\sqrt{n} = 12$ y $\sigma = 120 \rightarrow n = 100$

b) Si la amplitud total debe ser 12, cada lado debe ser 6.

$$\text{Si } \pm Z_{0,975} \sigma/\sqrt{n} = \pm 6; 1,96 \cdot 120/\sqrt{n} = 6; \rightarrow n = (1,96 \cdot 120/6)^2 = 1.536,64 \rightarrow n = 1.537 \text{ (ambos asumen que se conoce } \sigma \text{ y por eso recurren a } z)$$

$$6.8 \text{ IC}_{95\%} \pi = p \pm Z_{\alpha/2} \sqrt{[p(1-p)/n]} = p \pm Z_{\alpha/2} \sqrt{[0,6 \cdot 0,4/30]} \approx 0,60 \pm 0,18 = [0,42, 0,78]$$

Parece que, con 30 preguntas, se sabe de este alumno menos de lo que parecía: sólo se sabe que la proporción poblacional de preguntas que conoce este alumno es algún valor entre el 42 y el 78%. Evidentemente, convendría no basar toda la evaluación del alumno en esta prueba.

$$[\text{Condiciones de aplicación: } 0,42 \cdot 30 = 12,6 > 5 \text{ y } (1 - 0,78) \cdot 30 = 6,6 > 5]$$

$$6.9 \text{ IC}_{95\%} \pi = p \pm Z_{\alpha/2} \sqrt{[p(1-p)/n]} = 0,2125 \pm Z_{\alpha/2} \sqrt{[0,2125 \cdot 0,7875/160]} \approx \approx 0,2125 \pm 0,0634 = [0,1491, 0,2759] \approx [15\%, 28\%]$$

$$[\text{Condiciones de aplicación: } 0,15 \cdot 160 = 24 > 5]$$

6.10 Amplitud máxima $IC_{95\%}\pi \rightarrow \pm 1,96\sqrt{[0,5 \cdot 0,5/n]} =$

$$a) n=100 \rightarrow \pm 1,96\sqrt{[0,5 \cdot 0,5/100]} = \pm 1,96 \cdot 0,05 = \pm 0,098 \approx \pm 10\%$$

$$b) n=400 \rightarrow \pm 1,96\sqrt{[0,5 \cdot 0,5/400]} = \pm 1,96 \cdot 0,025 = \pm 0,049 \approx \pm 5\%$$

$$c) n=2.500 \rightarrow \pm 1,96\sqrt{[0,5 \cdot 0,5/2.500]} = \pm 1,96 \cdot 0,01 = \pm 0,0196 \approx \pm 2\%$$

$$d) n=10.000 \rightarrow \pm 1,96\sqrt{[0,5 \cdot 0,5/10.000]} = \pm 1,96 \cdot 0,005 = \pm 0,0098 \approx \pm 1\%$$

6.11 La amplitud del intervalo es inversamente proporcional a la raíz del tamaño muestral. Como en el caso de la media muestral, para disminuir la incertidumbre a la mitad, es necesario aumentar el tamaño muestral cuatro veces.

$$6.12 \quad IC_{95\%}\pi = p \pm Z_{\alpha/2}\sqrt{[p(1-p)/n]} = 0,40 \pm Z_{\alpha/2}\sqrt{[0,40 \cdot 0,60/100]} \approx 0,40 \pm 0,096 \approx [0,304, 0,496] \approx [30\%, 50\%]$$

[Condiciones de aplicación: $0,3 \cdot 100 = 30 > 5$]

Atención: no ha dicho nada de que se trate de una muestra aleatoria. Recuerde que el IC del error típico de estimación sólo tiene en cuenta los errores aleatorios, pero no los sistemáticos.

$$6.13 \quad \text{Si } \sigma_p = \sqrt{[\pi(1-\pi)/n]} = \sqrt{[0,5 \cdot 0,5/n]} = 0,05 \rightarrow n = 100$$

$$\text{Si } \pm Z_{0,975} \sigma_p = \pm 0,25; 1,96 \cdot \sqrt{[0,5 \cdot 0,5/n]} = 0,25; \rightarrow n = (1,96 \cdot 0,5/0,025)^2 = 1.536,64 \rightarrow n = 1.537$$

$$6.14 \quad IC_{95\%}RA = RA \pm Z_{\alpha/2}\sqrt{[p_1 \cdot (1-p_1)/n_1 + p_2 \cdot (1-p_2)/n_2]} =$$

$$= 0,4644 \pm 1,96\sqrt{(0,712 \cdot 0,288/132) + (0,248 \cdot 0,752/868)} \approx$$

$$= 0,4644 \pm 1,96 \cdot 0,0420 = 0,4644 \pm 0,0824 = [0,3820, 0,5468] \approx [38,2\%, 54,7\%]$$

Por lo que puede afirmarse que los expuestos al factor presentan entre un 38 y 55% más de riesgo.

$$6.15 \quad RR = 0,7121/0,2477 = 2,875 \rightarrow \text{Ln}(RR) = 1,0560$$

$$IC_{95\%}\ln(RR) = \text{Ln}(RR) \pm Z_{\alpha/2}\sqrt{[(1-p_2)/n_2 p_2 + (1-p_1)/n_1 p_1]} =$$

$$= 1,0560 \pm 1,96\sqrt{[0,2879/132 \cdot 0,7121 + 0,7523/868 \cdot 0,2477]} \approx$$

$$= 1,0560 \pm 1,96 \cdot 0,0810 = 1,0560 \pm 0,1588 = [0,8973, 1,2148]$$

$$IC_{95\%}RR = \exp[IC_{95\%}\log(RR)] = [e^{0,8973}, e^{1,2148}] \approx [2,45, 3,37]$$

Por lo que se concluye que los expuestos tienen un riesgo que es entre 2,45 y 3,37 veces superior.

Nótese que, a pesar de tener una muestra total de tamaño 1.000, al igual que en el ejemplo basado en la tabla 6-3, la amplitud del $IC_{95\%}$ es mucho más reducida, desde 2,45 hasta 3,37, en lugar de desde 2,1 hasta 15,6. Esta menor ignorancia se debe, en parte, a que la enfermedad Y+ es ahora más frecuente (309 casos en lugar de 15), aportando más información. (La amplitud del IC depende, sobre todo, del número de casos en la celda de la tabla que tiene menos casos.)

$$6.16 \quad \text{En los datos del ejemplo, el } OR = (94/38) / (215/653) = 7,5131 \rightarrow \text{Ln}(OR) = 2,0166$$

$$IC_{95\%}\ln(OR) = \text{Ln}(OR) \pm Z_{\alpha/2}\sqrt{(1/a + 1/b + 1/c + 1/d)} =$$

$$= 2,0166 \pm 1,96\sqrt{[1/94 + 1/38 + 1/215 + 1/653]} \approx$$

$$= 2,0166 \pm 1,96 \cdot 0,2077 = 2,0166 \pm 0,4071 = [1,6096, 2,4237]$$

$$IC_{95\%}OR = \exp[IC_{95\%}\ln(OR)] = [e^{1,6096}, e^{2,4237}] \approx [5,0, 11,3]$$

Por lo que se concluye que los expuestos tienen una razón enfermo/sano que es entre 5,0 y 11,3 veces superior.

	FE: NO	FE: SI
PAU: S	200	10
PAU: I	100	100

Tabla 6-9 Posible ejemplo de relación entre notas de entrada (PAU) superiores (S) o inferiores (I) a la media y fracaso escolar (FE).

6.17 a) La tabla 6-9 muestra un posible ejemplo.

b) $RA = (200 / (200 + 10)) - (100 / (100 + 100)) \approx 0,452$

c) $RR = (200 / (200 + 10)) / (100 / (100 + 100)) \approx 1,905$

d) $OR = (200 \cdot 100) / (100 \cdot 10) = 20$

$\ln(OR) = \ln(20) \approx 3$

$V(\ln(OR)) = a^{-1} + b^{-1} + c^{-1} + d^{-1} = 200^{-1} + 10^{-1} + 100^{-1} + 100^{-1} = 0,125$

$IC_{95\%}(\ln(OR)) = 3 \pm 1,96 \cdot 0,125 \approx 3 \pm 0,69 = [2,31,3,69]$

$IC_{95\%}OR = [\exp(2,31), \exp(3,69)] \approx [10,07, 40,05]$

6.18 a) Ambos parten del principio de que una proporción de casos desarrollan la ECV, independientemente de su exposición al ordenador. Pero difieren en que el riesgo atribuible considera que por el hecho de estar expuesto, aparecen nuevos casos, diferentes a los anteriores, que desarrollan también la enfermedad. En cambio, el riesgo relativo considera que el hecho de estar expuesto aumenta, en una cierta persona, la probabilidad de desarrollar ECV. Es decir, en el atribuible se suman dos grupos de casos, mientras que en el relativo, lo que se modifica es la probabilidad de cada caso.

b) $RA = (111 / (111+87)) - (231/(231 + 261)) \approx 0,091$

c) $RR = (111 / (111+87)) / (231/(231 + 261)) \approx 1,194$

d) $OR = 111 \cdot 261 / (87 \cdot 231) \approx 1,442$

$\ln(OR) \approx 0,366$

$V(\ln(OR)) = (1/111) + (1/261) + (1/87) + (1/231) = 0,029$

$SE(\ln(OR)) \approx 0,1693$

$IC_{95\%} \ln(OR) = \ln(or) \pm 1,96 \cdot SE(\ln(or)) \approx (0,034, 0,698)$

$IC_{95\%} OR = \exp(0,034, 0,698) \approx (1,034, 2,009)$

e) No, dado que el IC excluye el valor de no relación, podemos rechazar la independencia entre el grado de exposición al ordenador y la presencia de ECV. Otro tema es si se trata de relación causal, ya que es un estudio transversal y no puede distinguirse qué variable sigue a qué variable. (En los capítulos 11 y 12 se trata este tema con detalle.)

6.19 En los datos globales (tabla 6-10), sin tener en cuenta otras variables, la disparidad «PENA MUERTE = SÍ / PENA MUERTE = NO» es entre 1,16 y 7,15 superior cuando la víctima es de raza blanca que cuando es de raza negra.

Victima	Blanco	Negro
Pena: Sí	30	6
Pena: No	184	106
$OR = (30 \cdot 106)/(184 \cdot 6) = 2,88$		

$\ln(OR) = \ln(2,88) \approx 1,06$

$V(\ln(OR)) = a^{-1} + b^{-1} + c^{-1} + d^{-1} =$

$= 30^{-1} + 106^{-1} + 184^{-1} + 6^{-1} \approx 0,22$

$IC_{95\%} \ln(OR) \approx 1,06 \pm 1,96\sqrt{0,22} \approx 1,06 \pm 0,91 = [0,15, 1,97]$

$IC_{95\%} OR \approx [\exp(0,15), \exp(1,97)] \approx [1,16, 7,15]$

Tabla 6-10 Datos globales, sin distinguir la raza del acusado, de la tabla 6-7.

	Dolor suave	Dolor fuerte	Todos
OR	2,909	20,000	0,573
Ln (OR)	1,068	2,996	-0,558
Var (Ln[OR])	0,224	0,108	0,008
L.I. (Ln[OR])	0,140	2,351	-0,736
L.S. (Ln[OR])	1,996	3,640	-0,380
L.I. (OR)	1,150	10,500	0,479
L.S. (OR)	7,360	38,095	0,684

Tabla 6-11 Cálculo de todos los $IC_{95\%}$ (OR) de la tabla 6-8

6.20 Utilizando los intervalos de confianza al 95% (tabla 6-11), la interpretación sería la siguiente: en los que tienen dolor suave, el fármaco tiene un efecto positivo moderado, aumenta la disparidad «cura / no cura» entre 1,15 y 7,36 veces; en los que tienen un dolor fuerte, el efecto es más marcado: aumenta la razón «cura / no cura» entre 10,50 y 38,10 veces. En cambio, al realizar el estudio global con todos los casos, se llega a la conclusión contraria: el fármaco disminuye las posibilidades de curarse entre 0,48 y 0,68 veces.

Este ejemplo muestra hasta qué punto un efecto no homogéneo de la intervención junto con un diseño deficiente, que no equilibra variables importantes, pueden complicar la interpretación de los resultados. Los capítulos 11 y 12 se dedican a estos puntos.

Capítulo 7:

Prueba de significación y contraste de hipótesis

Introducción

Este capítulo formaliza la respuesta a dos preguntas diferentes pero relacionadas: «¿Qué sé?» (inferencia) y «¿Qué hago?» (decisión). Se define la prueba y el valor de P en el entorno de la evidencia empírica o inferencia sobre conocimiento. Por su parte, los riesgos α y β y el contraste de hipótesis se enmarcan en la decisión entre dos acciones alternativas. Finalmente, se distingue entre planteamientos para poner a prueba diferencias y planteamientos de equivalencia.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Percibirá la mayor dificultad de interpretar una prueba de significación o un contraste de hipótesis que un intervalo de confianza.
- Sabrá que el intervalo de confianza excluye aquellos valores del parámetro que hubieran sido rechazados en una prueba de significación.
- Sabrá que la hipótesis precede a la recogida de datos.
- En una prueba de significación, escribirá en la hipótesis H la negación de su objetivo de investigación.
- Identificará valor de P con inferencia y lo interpretará como magnitud de evidencia en contra de la hipótesis H .
- En una prueba de significación, si no rechaza H , dirá: «Nada se opone a aceptar H » o «no se han encontrado diferencias» en lugar de «no existen diferencias».
- No interpretará el valor de P como el riesgo de haberse equivocado.
- Escribirá $P < 0,001$ en lugar de $P = 0,000$.
- Interpretará el contraste de hipótesis entre H_0 y H_1 como una decisión entre dos acciones alternativas (A_0 y A_1).
- En un contraste, distinguirá entre riesgo α y riesgo β .
- Identificará riesgos α y β con la frecuencia de decisiones erróneas.
- Calculará la prueba de una media para el caso de muestras grandes o de variables con distribución normal.
- Distinguirá entre un planteamiento para demostrar diferencias y otro para establecer equivalencia.
- Distinguirá entre equivalencia, no inferioridad y no superioridad.
- Interpretará el margen de equivalencia como diferencias no relevantes.
- Tendrá curiosidad por el planteamiento bayesiano.

Objetivos

A la vista de la información aportada por la muestra, las dos principales preguntas de la inferencia estadística son: ¿qué valores del parámetro son creíbles? Y ¿se puede negar cierto valor del parámetro? La primera, mediante intervalos de confianza, se resolvió en el capítulo anterior; la segunda se expone en éste.

Comentario



Dado que el parámetro representa un valor poblacional, la inferencia pretende alcanzar modelos universales o absolutos. Estos modelos deben ser interpretados de acuerdo con las condiciones y características del estudio (quizá las conclusiones sean válidas sólo para adultos voluntarios sanos), pero son universales en el sentido de que se aplican a toda su población. En resumen, ambos procedimientos (intervalos de confianza y prueba de significación) son ambiciosos, ya que pretenden formalizar el conocimiento científico.

La pregunta que responden los intervalos de confianza (¿qué valores son creíbles?) engloba, de alguna manera, a la pregunta de la prueba de significación (¿se puede negar cierto valor?). Los intervalos de confianza aportan más información y son más fáciles de entender, asimilar y explicar. ¿Qué interés ofrece, entonces, poner a prueba una hipótesis? Pues, simplemente, que puede ser la auténtica pregunta de interés.

Ejemplo 7.1



Saber si un fármaco es más eficaz que otro puede reducirse a conocer si la diferencia de sus medias en la respuesta de interés es o no es exactamente el valor 0. Por tanto, poder negar el valor 0 implica haber demostrado que un producto es más eficaz que otro.

Esta prueba sobre una hipótesis puede abordarse desde el punto de vista de la inferencia (¿qué sé?) o de la decisión (¿qué hago?), y se ven, respectivamente, en la prueba de significación, PS, y en el contraste de hipótesis, CH.

Inferencia: prueba de significación

Se desea **poner a prueba** o confrontar una hipótesis previa H con la información que proporcionan los datos.

Ejemplo 7.2



Desde hace algún tiempo, un residente se juega a cara o cruz las guardias que coinciden con las fiestas familiares. Su compañero lanza su moneda y... ¡siempre gana!, de for-

Ejemplo 7.2 (Cont.)

ma que la confianza en él se va deteriorando, hasta que un día se plantea estudiar formalmente si la moneda está apañada. Así, el problema consiste en analizar si es cierta la: Hipótesis $H : \pi = 0,5$ (moneda correcta), donde π representa la probabilidad de cara que se desea contrastar. Obsérvese que, para demostrar que la moneda está cargada, se pone en H que es correcta. Si se lanza $n = 100$ veces la moneda y se observa la proporción p de caras, se dispondrá de cierta «evidencia» empírica que conviene asimilar. Supóngase que se observa una proporción $p = 0,63 = 63\%$. Este resultado conduce a creer, de alguna manera, que la moneda está «cargada»: que no es cierto que $\pi = 0,5$. En cambio, si el resultado fuera $p = 0,52 = 52\%$, se consideraría «compatible» con que la moneda no esté cargada. Cuanto más se aleje p de $0,5$, hay más información en contra de la hipótesis H de que la moneda es correcta.

Comentario

La aleatoriedad asociada a este proceso hace que no esté libre de riesgos. Es posible que una moneda perfecta, no cargada, genere una observación de 63 caras en 100 lanzamientos. Y, de forma recíproca, también es posible que una moneda que no sea perfecta y que tenga una probabilidad de cara diferente de 0,5 genere una muestra con un 50% de caras.

Comentario

Por supuesto, se podría abordar el problema desde un punto de vista físico y, dando por bueno el conocimiento actual de esta ciencia, estudiar la composición de la moneda, su centro de gravedad, su circunferencia, etc. Ahora bien, sea cual sea la respuesta del estudio teórico, siempre conviene estudiar qué dicen los datos, no sea que convenga revisar el modelo teórico.

La hipótesis H establece una condición sobre el parámetro poblacional que se desea poner a prueba, confrontando ésta con la información que proporcionan los datos. Esta información se «condensa» en un estadístico apropiado, que fluctúa aleatoriamente según una distribución dependiente del verdadero valor del parámetro. Cuando H es correcta, la distribución es conocida, y el estadístico se localizará de forma previsible o, lo que es lo mismo, no se alejara mucho de una zona determinada. Por tanto, cuanto más lejos se encuentre el estadístico de la zona asociada a H , menos verosímil se presenta ésta, y más credibilidad cobra la posibilidad de que el estadístico proceda de otra distribución, es decir, de parámetros distintos al que determina H .

Definición

La prueba de significación es una técnica de inferencia estadística para juzgar si una propiedad que, se supone, cumple una población es compatible con lo observado en una muestra de la misma.

Ejercicio 7.1

En relación con la prueba de significación (elija una):

- Se desea conocer el valor de cierto parámetro.
- Se construye una hipótesis sugerida por los datos.
- Se busca «evidencia» (pruebas) a favor de la hipótesis H que se desea demostrar.
- Ninguna de las anteriores es correcta.

Ejercicio 7.2

Escriba la hipótesis para contrastar si un nuevo tratamiento es mejor que uno clásico:

- H : el nuevo tratamiento no es mejor que el clásico.
- H : el nuevo tratamiento es mejor que el clásico.
- H : el rendimiento del nuevo tratamiento supera al clásico.
- Ninguna de las anteriores es correcta.

Así pues, se requiere un proceso formal que permita «incorporar» la información muestral o «evidencia» empírica. Este proceso debe ser transparente, en el sentido de ser reproducible por un segundo investigador.

Valor de P

Un procedimiento consiste en calcular el valor de P (P value) o probabilidad de que se presente un valor del estadístico más alejado de la hipótesis H que el observado.

Recuerde

Si el valor de P (P value) es pequeño, se dice que H es inverosímil.

Recuerde

El valor de P puede interpretarse como «cuán inverosímil es el resultado observado si H fuera cierta» o «hasta qué punto el resultado observado es probabilísticamente compatible con H ». Lo que suele interpretarse, cuando P es pequeña, como que hay «suficiente evidencia o pruebas en contra de H » para creer que el resultado es «estadísticamente significativo».

Mecánica de la prueba de significación

La prueba de significación se basa en el siguiente proceso formal (fig. 7-1):

- i) Se escoge una variable, que valora el objetivo del estudio.
- ii) Se elije un diseño y un estadístico que resuma la variable en la muestra.
- iii) Se especifica una hipótesis (H) que se desea poder rechazar.
- iv) Se especifica la distribución del estadístico bajo H y las premisas necesarias.
- v) Se realiza el experimento y se calculan los valores del estadístico y del valor de P.
- vi) Si el valor de P es muy pequeño se dice que H es inverosímil.
- vii) Se reporta el estimador con su IC_{95%}.

A continuación se explica este proceso para el caso de una probabilidad.

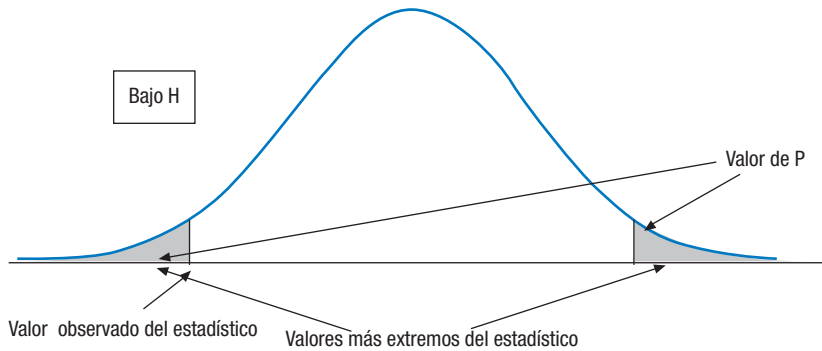


Figura 7-1 El valor de P indica la probabilidad de observar valores del estadístico igual o más extremos que el observado, en el caso de que H sea cierta.

Prueba de significación de una probabilidad

Se pueden utilizar los conocimientos sobre la distribución de la proporción p observada en una muestra para poner a prueba una hipótesis H sobre una probabilidad poblacional π .

Si la probabilidad π de la población origen de la muestra es 0,5, se escribe $H : \pi = 0,5$

Recuerde



Si la probabilidad poblacional de cierto evento es π , la distribución de la proporción observable p en muestras de tamaño n puede aproximarse por una distribución normal centrada alrededor de π con varianza $\pi (1-\pi) / n$:
 $p \rightarrow N (\pi, \pi (1 - \pi) / n)$

Comentario



Recuerde que las condiciones de aplicación de esta fórmula son que la muestra sea grande y la probabilidad π no extrema:
 $\pi \cdot n \geq 5$ y $(1 - \pi) \cdot n \geq 5$

Comentario



Observe que, como π vendrá especificado por la hipótesis, ahora no es necesario sustituirlo por su estimación muestral p ni por el valor del producto máximo $\pi(1 - \pi)$.

En el ejemplo anterior de la moneda, con una muestra de $n = 100$,

i) Variable: resultado cara o cruz

ii) Estimador: proporción p de caras

iii) Hipótesis H : $\pi = 0,5$ (moneda correcta)

iv) Si H es cierta: $p \rightarrow N(\pi, \pi(1 - \pi) / n) = N(0,5, 0,05^2)$

Premisas: muestra grande $\pi \cdot n > 5$ y $(1 - \pi) \cdot n > 5$

v) Caso a) con $n = 100$ se observan 63 caras

$$p = 63 / 100 = 0,63 = 63\%$$

$$\hat{Z} = (p - \pi) / \sqrt{\{\pi(1 - \pi) / n\}} = (0,63 - 0,5) / \sqrt{\{0,5 \cdot 0,5 / 100\}} \approx 0,13 / 0,05 = 2,6$$

En la tabla 4-2 se obtiene que al valor 2,6 le corresponde una probabilidad unilateral de 0,0047, por lo que el valor bilateral de P es $0,0094 < 0,01$.

vi) Por tanto (fig. 7-2), como el valor de P (o probabilidad de observar un valor de p tan o más alejado de H : $\pi = 0,5$) es $P < 0,01$, se considera inverosímil que π es 0,5 con un valor de $P < 0,01$.

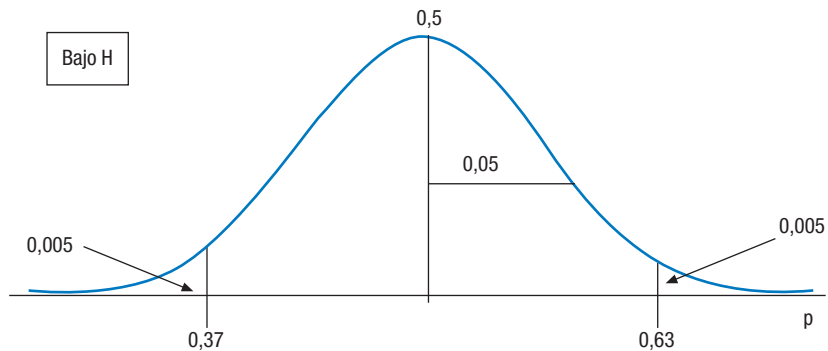


Figura 7-2 Bajo H : $\pi = 0,5$ y con una muestra $n = 100$, la distribución de p , proporción observada en la muestra, es Normal $(0,5, 0,05^2)$. Si se observan 63 caras, $p = 0,63$ y la probabilidad de observar 63 o más caras es de 0,005, que junto a la de observar 63 o más cruces (27 o menos caras) hace $P = 0,01$.

$$\text{vii) } IC_{95\%}: p \pm Z_{\alpha/2} \sigma_p \approx 0,63 \pm 1,96 \cdot 0,05 \approx 0,63 \pm 0,10 = 0,53, 0,73$$

La «auténtica» proporción de cara π se encuentra entre 53 y 73% (fig. 7-2).

v) Caso b) con $n = 100$ se observan 52 caras

$$\pi = 52/100 = 0,52 = 52\%$$

$$\hat{Z} = (p - \pi) / \sqrt{\{\pi(1 - \pi) / n\}} = (0,52 - 0,5) / \sqrt{\{0,5 \cdot 0,5 / 100\}} \approx 0,02 / 0,05 = 0,4$$

La tabla 4-2 proporciona, para el valor 0,4, una probabilidad unilateral de 0,3446, a la que corresponde $P = 0,6892$ bilateral.

vi) Por tanto (fig. 7-3), valor de $P = \text{Prob}(p(\text{cara}) \geq 0,52 \text{ y } p(\text{cara}) \leq 0,48) \approx 0,69$

Como el valor de P es grande, no es inverosímil, o más formalmente: «nada se opone a aceptar H ».

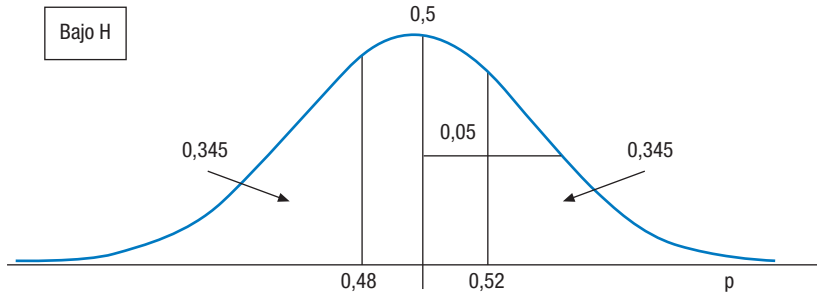


Figura 7-3 Si se observan 52 caras, $p = 0,52$ y la probabilidad de observar 52 o más caras es de 0,345, que junto a su simétrica (observar 52 o más cruces = 48 o menos caras) hace $P = 0,69$.

vii) $IC_{95\%}: p \pm Z_{\alpha/2} \sigma_p = 0,52 \pm 1,96 \cdot 0,05 \approx 0,52 \pm 0,10 = 0,42, 0,62$
 La «auténtica» proporción de cara π se encuentra entre 42 y 62%.

Ejercicio 7.3



Una serie de 400 pacientes con sida han recibido, en diferentes etapas de su seguimiento, dos tratamientos diferentes, A y B que son, a priori, potencialmente similares. Ahora, se les ha preguntado cuál prefieren, y un 58% ha optado por el A. Se desea saber si es razonable creer que son igualmente preferidos. Escriba todos los pasos del proceso.

Ejercicio 7.4

Repita el ejercicio 7.3 asumiendo que las preferencias por A han sido 53%.

Un programa informático, al redondear, proporciona un nivel de significación con muchos ceros (0,000...) que parecería indicar un resultado imposible, de probabilidad nula. Como ello no es así, se expresa de la siguiente forma: $P < 0,001$.

Recuerde

Nunca escriba $P = 0,000$; en su lugar, ponga $P < 0,001$.

Ejercicio 7.5

Los usuarios de una biblioteca llevan años protestando por las prestaciones del sistema de búsqueda disponible para realizar sus consultas. Los responsables de la biblioteca deciden valorar la posibilidad de cambiar el sistema. Durante el período de prueba, han realizado un experimento comparando ambos sistemas mediante una escala que mide la satisfacción de los usuarios. Hacen la prueba anterior de comparación de preferencias y resumen sus resultados con la siguiente frase: *el nuevo sistema genera mayor satisfacción en los usuarios ($P < 0,01$)*. ¿Cuál o cuáles de las siguientes afirmaciones son ciertas?:

- Se rechaza la hipótesis H_0 de que la satisfacción es igual en ambos grupos.
- Suponiendo que ambos sistemas generen la misma satisfacción, la probabilidad de haber obtenido un resultado tan o más extremo que el observado es menor del 1%.
- Creemos que el resultado observado refleja una diferencia poblacional, en el conjunto de todos los posibles casos, del nivel de preferencias.
- La proporción de casos más satisfechos con el sistema antiguo que con el nuevo es menor del 1%.
- Cuando se dice que el nuevo sistema es mejor se tiene una probabilidad de error menor de 0,01.

Prueba de significación de una media

La aplicación del mecanismo anterior a una variable continua en la que se desea contrastar una hipótesis sobre su media es muy similar.

Ejemplo 7.3

Por ejemplo, se quiere poner a prueba si la media μ de la respuesta Y es una cierta media μ_H especificada en la hipótesis H_0 .

Si la media μ de la población origen de la muestra es cierta media μ_H preespecificada, se escribe: $H_0 : \mu = \mu_H$

Conocida σ , ya se sabe que \bar{y} será normal cuando o bien n sea grande ($n \geq 30$) o bien Y sea normal. Por tanto, bajo $H_0 : \hat{Z} = (\bar{y} - \mu_H) / (\sigma/\sqrt{n}) \rightarrow N(0,1)$.

\bar{y} , como antes, puede calcularse el valor de P y creer H_0 inverosímil si P es pequeña.

Nota técnica



$$P = P[|z| > |\hat{Z}|] = P[|z| > |(\bar{y} - \mu_H) / (\sigma/\sqrt{n})|]$$

Ejemplo 7.4



¿Recuerda el ejemplo 6.2, de la ASU para demostrar que las gasolineras estaban poniendo menos gasolina de la que cobraban? Se resolvió con un IC, pero ¿se puede demostrar que están timando? En una muestra aleatoria de 100 envases, aceptando $\sigma = 10$ ml, se debe tomar una decisión sobre si $\mu = 1.000$, habiendo observado una media $\bar{y} = 997$ ml.

i) Variable: contenido real en envases de 1.000 ml

ii) Estimador: promedio muestral \bar{y}

iii) H: $\mu_H = 1.000$ ml

iv) Se usará el estadístico $\hat{Z} = (\bar{y} - \mu_H) / (\sigma/\sqrt{n})$

que bajo H tiene una distribución Normal: $\hat{Z} \rightarrow N(0,1)$ siempre que (premisas) o bien la muestra sea grande ($n \geq 30$) o bien Y sea normal

v) Cálculo del estadístico: $\hat{Z} = (997 - 1.000) / (10/\sqrt{100}) = -3$
 $P = \text{Prob}[|z| > |-3|] < 0,0027$ (valor de Excel)

$P = 0,0027$, $\mu = 1.000$ ml es muy inverosímil

vi) Conclusión práctica: es poco verosímil que se esté dispensando la cantidad especificada.

vii) IC_{95%}: $\bar{y} \pm Z_{\alpha/2} \sigma/\sqrt{n} = 997 \pm 1,96 \cdot 10 / \sqrt{100} \approx 997 \pm 2$
 $= [995, 999]$

La auténtica media μ de cantidad servida se encuentra entre 995 y 999 ml. Nos están timando, aunque a nivel individual, la cantidad es pequeña. La pequeña amplitud del IC_{95%} muestra que se dispone de mucha información.

Ejercicio 7.6



Los CD4 de una población de pacientes con sida tienen una desviación típica de 100. Se desea superar una media de 500 unidades. Se aplica una terapia experimental a una muestra de 36 casos de esta población y se obtiene una media muestral de 505. ¿Se puede afirmar que la media poblacional está por encima de 500?

Si, como es más usual, la varianza poblacional σ^2 es desconocida, se recurre a su estimador muestral S^2 y a la distribución t de Student. En este caso se usa el estadístico: $\hat{t} = (\bar{y} - \mu_0) / (S/\sqrt{n})$, que sigue una distribución t de Student con $n - 1$ grados de libertad, siendo «n» el número de casos.

Ejemplo 7.5

En 9 voluntarios sanos se ha estudiado la diferencia d entre los tiempos de respuesta a un estímulo visual y auditivo, habiéndose observado, $\bar{d} = 6,71$ y $S = 6,0$. Asumiendo que $d \rightarrow N$, ¿se puede aceptar que $E(d) = \mu = 0$, lo que implica que la respuesta a ambos estímulos es idéntica?

Solución:

i) Variable: diferencia entre el tiempo de respuesta a los estímulos visual y auditivo

ii) estimador: media de las diferencias \bar{d}

iii) H: $E(d) = \mu_H = 0$

iv) estadístico referencia $\hat{t} = (\bar{d} - \mu_H) / (S/\sqrt{n})$

Distribución bajo H : $\hat{t} \rightarrow t_{n-1} = t_8$

Premisa: d normal

v) Cálculos $\hat{t} = (6,71 - 0) / (6/\sqrt{9}) = 3,355$

$P = \text{Prob}[|t| > |3,355|] = 0,01$

vi) Como $P = 0,01$; H : $\mu_H = 0$ es poco inverosímil

Conclusión práctica: la igualdad entre ambos estímulos es dudosa.

vii) IC_{95%}: $\bar{d} \pm t_{n-1, \alpha/2} S/\sqrt{n} = 6,71 \pm 2,306 \cdot 6/\sqrt{9} \approx 6,71 \pm 4,612 \approx [2,10, 11,32]$

La auténtica diferencia entre la respuesta media a ambos estímulos se encuentra entre 2,10 y 11,32.

Ejercicio 7.7

La satisfacción de los usuarios se mide por una escala que puntúa entre 0 y 100 cuya distribución se asemeja razonablemente a la normal. El objetivo de un servicio sanitario es conseguir satisfacciones por encima de 70. En una muestra al azar de 16 usuarios se ha observado una media de 79 y una desviación típica de 12. ¿Se puede afirmar que la media poblacional está por encima de 70?

Los estadísticos « \hat{z} » y « \hat{t} » como cociente señal / ruido

Los estadísticos \hat{z} y \hat{t} tienen una estructura muy similar. El numerador representa la distancia entre el valor de la muestra y el parámetro μ de la población. Y el denominador informa del error típico de \bar{y} , ya que como μ es un parámetro de la población (forma parte de la pregunta), no tiene error aleatorio de muestreo.

Ejemplo 7.4 (Cont.)

En el **ejemplo 6.2** del control de calidad de las gasolineras, si se desea saber si el surtidor cumple con las especificaciones (μ), este numerador representa la señal que proporciona la muestra: cuánto se desvía o distancia de la media especificada en la hipótesis. Se ha observado un valor de -3 .

Ejemplo

Por otro lado, la oscilación de \bar{y} explicable por el muestreo aleatorio puede cuantificarse en $\sigma_{\bar{y}} = \sigma / \sqrt{n} = 1$. Por tanto, el cociente «señal/ruido» vale -3 , lo que indica que la señal observada es negativa y 3 veces superior al error aleatorio medido por el error típico.

Recuerde

Tanto el estadístico \hat{z} como \hat{t} pueden ser vistos como un cociente señal / ruido.

Aun así, conviene reportar el valor exacto de P .

Pruebas de significación e intervalos de confianza

Digamos otra vez que las pruebas de significación (PS) y los intervalos de confianza (IC) son dos herramientas de inferencia: ambas permiten pasar de la muestra a la población. Mientras la prueba de significación hace una pregunta concreta o «cerrada» sobre el valor del parámetro en la población (¿es $\pi = 0,5$?), el intervalo de confianza hace una pregunta «abierta», (¿cuál es el valor de π ?). Se podría argumentar que el intervalo es una herramienta positiva que dice cuáles son los posibles valores del parámetro compatibles con la muestra observada, mientras que la prueba de significación es una herramienta negativa.

Ejemplo 7.6

Recuperemos el ejemplo de las 52 caras en 100 lanzamientos de una moneda. El intervalo de confianza del auténtico valor de la probabilidad de cara era:

$$IC_{0,95} \pi = p \pm Z_{\alpha/2} \sqrt{p(1-p)/n} = 0,52 \pm 1,96 \sqrt{[0,52 \cdot 0,48 / 100]} \\ \approx 0,52 \pm 0,10 = [0,42, 0,62]$$

Se cree, con una confianza del 95%, que esta moneda tiene una probabilidad de cara, π , situada entre el 42 y el 62%. Este resultado coincide con el de la prueba de significación que, con una $P = 0,69$, no permitía rechazar la hipótesis de que la probabilidad de cara era 0,5.

En el caso de observar 63 caras el IC es:

$$IC_{0,95} \pi = p \pm Z_{\alpha/2} \sqrt{p(1-p)/n} = 0,63 \pm 1,96 \sqrt{[0,63 \cdot 0,37 / 100]} \\ \approx 0,63 \pm 0,095 = [0,535, 0,725]$$

Por lo que ahora se cree, con una confianza del 95%, que esta probabilidad de cara, π , es alguno de los valores comprendidos entre el 53,5 y el 72,5%. Dado que este intervalo excluye el valor 0,5, se llega a la misma conclusión que con la prueba de significación, que había arrojado un valor de $P = 0,01$.

Las conclusiones de la prueba de significación y del intervalo de confianza coinciden.

Se pueden utilizar los intervalos de confianza para poner a prueba la hipótesis de interés, ya que los valores del parámetro excluidos del IC generarían PS con valores de P poco verosímiles.

Ejercicio 7.8



En 100 pacientes con sida el intervalo de confianza al 95% de la media μ del recuento de CD4 va de 375 a 500. Si se plantearan las dos pruebas de significación siguientes con $\alpha = 0,05$: (A) $H_A: \mu = 400$; (B) $H_B: \mu = 350$ las conclusiones serían:

- nada se opone a aceptar ambas H ;
- se rechazan ambas H ;
- nada se opone a aceptar H_A y se rechaza H_B ;
- se rechaza H_A y nada se opone a aceptar H_B .

El intervalo de confianza completa la información de la prueba de significación y ayuda a interpretarla, ya que informa sobre los valores plausibles del parámetro.

Nota técnica



En el caso de rechazar una hipótesis H (p. ej., $\pi = 0,5$), la conclusión formal de la prueba de significación sería que H es inverosímil sin decantarse hacia ninguno de los dos lados. Pero, a nivel práctico, el intervalo de confianza permite conocer, no sólo el lado, sino también los valores razonables del parámetro.

Cuando no se rechaza H , el intervalo de confianza permite distinguir entre poca (IC muy amplio) y mucha información (IC estrecho).

Recuerde



Utilice siempre IC. Incluso si el objetivo principal es la prueba de significación, conviene acompañar sus resultados con un IC. Si se rechaza H , porque se dice dónde se cree que está el parámetro. Si no se rechaza H , porque se cuantifica la información de que se dispone.

Comentario



Cuando no se puede rechazar la hipótesis, la prueba de significación concluye: «No es inverosímil: no hay evidencia en contra de H ». Pero ello puede ser, bien por falta de evidencia para establecer algo existente (¿muestra pequeña, diseño deficiente, análisis pobre, etc.?), o bien porque realmente no hay nada que ver.

Recuerde

*En PS, ausencia de pruebas no es prueba de ausencia.
En PS, ausencia de evidencia no es evidencia de ausencia.*

Lectura

Las recomendaciones para los autores de revistas biomédicas (9) anteponen el uso de IC al de PS: «A pesar de que los valores de P se pueden ofrecer junto a los intervalos de confianza, los resultados no deben recoger únicamente los valores de P» (CONSORT, ítem 17).

Pruebas de significación unilaterales y bilaterales

Hasta el momento, sólo se han planteado pruebas **bilaterales** (también llamadas de dos colas), como el ejemplo de la moneda, que se consideraba defectuosa tanto si salían demasiadas caras como demasiado pocas. En consecuencia, la región crítica de la hipótesis se sitúa a ambos lados (fig. 7-4):

$$H : \pi = 0,5$$

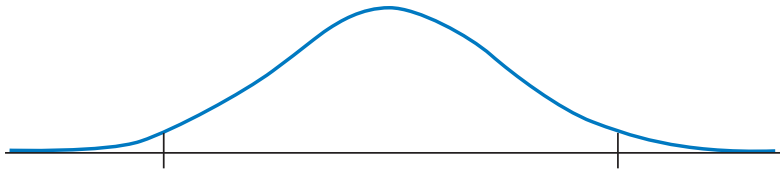


Figura 7-4 Las pruebas bilaterales contemplan los casos en contra de H en ambas colas.

Pero se pueden plantear también pruebas de una sola cola. En el ejemplo de la gasolinera, en el que se quería detectar si había tимо, ¿qué se puede concluir si la media observada se situaba por encima de la media teórica? ¿Que regalan gasolina? En esta situación, tiene más sentido una prueba **unilateral por la izquierda** (fig. 7-5):

$$H : \mu \geq \mu_0$$

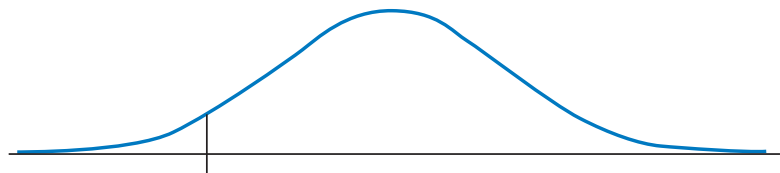


Figura 7-5 Las pruebas unilaterales por la izquierda contemplan los casos en contra de H en el lado izquierdo.

Y, de forma simétrica, si se estudia cómo aumenta la respuesta al aumentar la dosis, podría tener más sentido una prueba **unilateral por la derecha** (fig. 7-6):

$$H: \mu \leq \mu_0$$

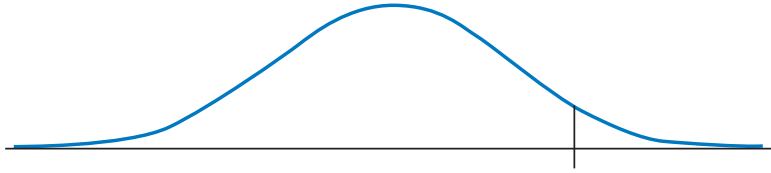


Figura 7-6 Las pruebas unilaterales por la derecha contemplan los casos en contra de H en el lado derecho.

Recuerde



En el caso de pruebas unilaterales, debe considerarse sólo una cola.

Resaltemos dos aspectos relevantes:

1) Al concentrar todo el nivel de significación en un lado, se hace algo mayor la región crítica de esa cola, por lo que una hipótesis unilateral o bilateral, ¡puede cambiar las conclusiones!

2) El signo igual (acompañado, ahora, por el desigual) sigue figurando en H .

Recuerde



La hipótesis H es el punto de salida (no el de llegada), y condiciona todo el experimento: la hipótesis H debe establecerse previamente a la recogida de datos.

Ejercicio 7.9



Repita los ejercicios 7.6 y 7.7 bajo un planteamiento unilateral.

Ejercicio 7.10

Se desea resolver la prueba $H: \mu \leq 0$ mediante un estadístico que sigue una distribución normal $(0,1)$. El resultado de la prueba ofrece $z = -2$, por lo que se concluye (elija una):

- que la media poblacional es 0;
- que la media poblacional es mayor que 0 ($P = 0,0227$);
- que la media poblacional es menor que 0 ($P = 0,0227$);
- hay una probabilidad del 97,37% de que la media poblacional sea 0;
- nada se opone a aceptar la hipótesis de que la media es igual o inferior a 0.

Decisión: contraste de hipótesis

El contraste de hipótesis es un instrumento para tomar una decisión manteniendo controlados los riesgos de error.

Ejemplo 7.7



(Prestado de un examen para informáticos de la profesora Monique Becue.) Para creer si cierto «garabato» es un 8, un programa de reconocimiento de patrones mide la curvatura izquierda (Y), cuya distribución tiene una media de 12 u si se trata de un «8», y una media superior si se trata de una «B». Se sabe que la distribución de Y es normal y que $\sigma = 3$ u. Si se está dispuesto a aceptar que un 5% de ochos (8) sean reconocidos como «bes» (B), ¿a partir de qué valor se dirá que se trata de una «B»? Es decir: ¿cuál es el límite de la región crítica?

i) $H : \mu = 12$ u (se trata de un 8)

ii) Se usará el estadístico $\bar{Z} = (\bar{y} - \mu_H) / (\sigma/\sqrt{n})$

Distribución bajo $H : \bar{Z} \rightarrow N(0,1)$

Premisas: dado que $n = 1$, Y debe ser normal

iii) Regla rechazo. Con $P = 0,05$, se rechazará H si $\hat{Z} > Z_{\alpha} = 1,645$.

iv) Cálculo del límite: $\mu_0 + Z_{\alpha} (\sigma/\sqrt{n}) = 12 + 1,645 \cdot 3 = 16,935$

Si $y > 16,935$ se «cree» que no se trata de un 8.

Supóngase ahora que se conoce que la distribución de las «B» es $N(21,3)$. Aceptando como límite de decisión $y = 16,935$, se desea calcular la probabilidad de que una B sea reconocida como un 8 (fig. 7-7). Ahora se dispone de dos situaciones hipotéticas, entre las que se debe escoger.

$H_0 : \mu = 12$ (se trata de un 8)

$H_1 : \mu = 21$ (se trata de una B)

Prob [$y \leq 16,935$ condicionado a $Y \rightarrow N(21,3)$] =

= $P [Z \leq (16,935 - 21) / 3] =$

= $P (Z \leq -1,355) \approx 0,0885$

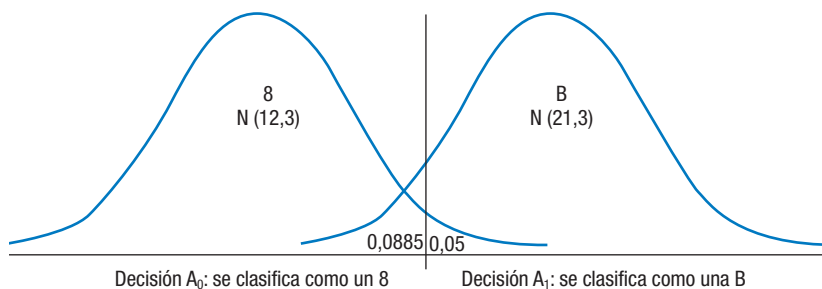


Figura 7-7 Si el valor observado supera el límite 16,935 se clasifica el garabato como B (A_1) y en caso contrario como 8 (A_0).

Comentario



En lo que sigue emplearemos A_0 y A_1 (acción 0, acción 1) para resaltar la acción que implica el contraste de hipótesis. Sea A_0 «conservadora» y A_1 «innovadora». Para tomar la acción A_1 hace falta rechazar H_0 .

Definición



Un contraste de hipótesis permite elegir entre dos acciones alternativas.

Se han identificado dos actuaciones erróneas y se han cuantificado los riesgos respectivos:

$$P(\text{concluir } B \mid \text{realidad } 8) = 0,05$$

$$P(\text{concluir } 8 \mid \text{realidad } B) = 0,088$$

Un organismo científico, como la Sociedad Española de Neurología, la revista *Medicina Clínica* o la Cochrane Collaboration, está interesado en lo que científicamente se sabe y, por tanto, en realizar intervalos de confianza o pruebas de significación. En cambio, un órgano ejecutivo, como una agencia reguladora del medicamento o un comité que elabora protocolos, debe proponer decisiones, acciones concretas.

Ejemplo 7.8



Fisher (41) y Hill (42) mantuvieron posiciones distintas en cuanto a la evidencia disponible sobre los efectos del tabaco. Sea cual sea esta evidencia, a un responsable de Salud Pública, lo que le concierne es, a la luz de dicha información, cuál debe ser su actuación. Greenland (42b) recuerda que un organismo de Salud Pública debe actuar y debe, por tanto, tomar decisiones: ante humo en un bosque, la acción pertinente es enviar bomberos, no científicos para averiguar si debajo del humo hay fuego. A nivel personal, por ejemplo, un fumador debe valorar las consecuencias de los dos «errores» posibles: *i*) que decida seguir fumando, pero tenga razón Hill y él mismo sea de la proporción de casos que desarrollan el cáncer hacia los 50 años; o *ii*) que decida no fumar, pero tenga razón Fisher y no se «ahorre» dicha enfermedad. Cada uno debe valorar qué consecuencias tiene cada posible situación.

Historieta



Supongamos que quien debe decidir si fuma, o no, se plantea minimizar su sentimiento de estupidez. Para eso debe pensar con cuál de los dos posibles errores se sentirá más estúpido: i) ¡Qué lástima! podría haber vivido 30 años más; ii) ¡Qué lástima!, podría haber fumado.

Límites de significación

El límite de significación a partir del cual se rechaza H_0 tiene un equivalente en la escala de los estadísticos, \hat{Z} o \hat{t} . En la escala Z, los límites que corresponden a $P = 0,05$ son $-1,96$ y $+1,96$. En la t de Student, dependerá de los grados de libertad.

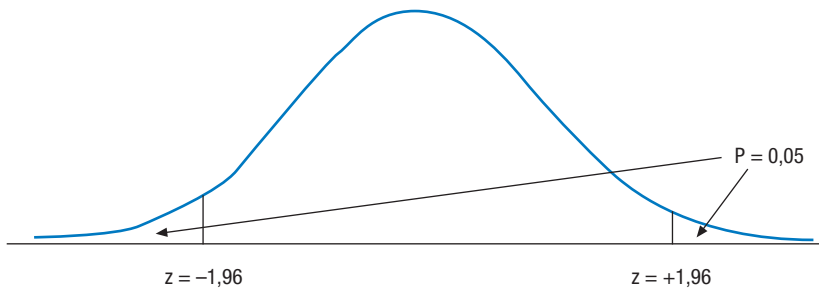


Figura 7-8 Es equivalente preguntarse si $P < 0,05$ o si \hat{Z} es mayor que 1,96 o menor que $-1,96$.

Ejercicio 7.11



¿Cómo se habría realizado la comparación con Z en los ejercicios 7.3, 7.4 y 7.6? ¿Y con t en el 7.7?

Recuerde



La prueba de significación contesta la pregunta «qué creo» y el contraste de hipótesis, «qué hago».

Errores tipo I y II. Riesgos α y β

En el contraste de hipótesis se definen dos tipos de errores.

Definición



El *error de primera especie* o *tipo I* consiste en tomar la acción alternativa (A_1) cuando era cierta H_0 .
Tomar A_1 | es cierta H_0

Ejemplo 7.9

Concluir que es una B cuando en realidad es un 8 es un error tipo I.

Definición

El error de *segunda especie* o *tipo II* consiste en tomar la acción nula (A_0) cuando es cierta H_1
Tomar A_0 | es cierta H_1

Ejemplo 7.9 (Cont.)

Concluir que es un 8 cuando en realidad es una B es un error tipo II.

Definición

Las probabilidades correspondientes de cometer errores de primera y de segunda especie reciben el nombre de *riesgos* α y β :
Riesgo $\alpha = P(\text{Tomar } A_1 | \text{ es cierta } H_0)$
Riesgo $\beta = P(\text{Tomar } A_0 | \text{ es cierta } H_1)$

Ejemplo 7.10

Riesgo $\alpha = P(\text{Decidir es una B} | \text{ en realidad es un 8})$
Riesgo $\beta = P(\text{Decidir es un 8} | \text{ en realidad es una B})$
De esta manera, α representa la proporción de 8 que serán identificados como B y β su recíproco.

Ejemplo 7.11

Un laboratorio farmacéutico propone a una agencia reguladora del medicamento un ensayo clínico para contrastar H_0 (misma eficacia que control) frente H_1 (eficacia mayor que el control = Δ). Si se rechaza H_0 , el fármaco se pondrá en el mercado (A_1). El riesgo α sería la proporción de medicamentos iguales que el control (H_0) que son finalmente puestos en el mercado (A_1). A su vez, el riesgo β , es la proporción de medicamentos que superan el control en un valor Δ (H_1) que no llegan al mercado (A_0).

Ejercicio 7.12



Un proveedor entregaba un reactivo con un tiempo de reacción medio de 100 ms y desviación típica de 10 ms. Ahora, ofrece uno nuevo con parámetros $m = 50$ ms y $s = 5$ ms. Sea H_0 : $m = 100$ ms y $s = 10$ ms (reactivo viejo); H_1 : $m = 50$ ms y $s = 5$ ms (reactivo nuevo); A_0 : rechazar el lote y A_1 : aceptar el lote. El riesgo α de cometer un error de primera especie es:

- a) la probabilidad de que el reactivo sea nuevo;
- b) la probabilidad de aceptar un lote (A_1) de reactivos viejos (H_0) [ante un reactivo viejo, la probabilidad de decidir que es de los nuevos];
- c) la probabilidad de rechazar un lote (A_0) de reactivos nuevos (H_1) [ante un reactivo nuevo, la probabilidad de decidir que es de los viejos];
- d) todas son falsas.

Ejercicio 7.13

En un contraste de hipótesis, si H_0 es cierta, es posible (elija una):

- a) cometer dos errores, el de tipo I y el de tipo II;
- b) sólo se puede producir el de tipo I;
- c) sólo se puede producir el de tipo II;
- d) ninguno, ya que H_0 es cierta.

Definición



La potencia (tabla 7-1) de un estudio es $1 - \beta$ o probabilidad de decidir A_1 cuando es cierta H_1 :

Potencia = $1 - \beta = P(\text{Decidir } A_1 \mid \text{es cierta } H_1)$

Tipos de errores y riesgos		Decisión	
		A_0	A_1
Realidad	H_0		Tipo I (riesgo α)
	H_1	Tipo II (riesgo β)	Potencia = $1 - \beta$

Tabla 7-1 Resumen de tipos de errores y riesgos

Intervalos de confianza, pruebas de significación y contraste de hipótesis

En un intervalo de confianza, IC, el nivel de confianza α se decide a priori. En un contraste de hipótesis, CH, también, y se opta por aquel diseño y estadístico que minimiza β , que también se establece a priori. Así, en el entorno de IC y CH, lo único que tiene valor y debe, por tanto, ser reportado son los valores de α y β decididos a priori. En cambio, en la prueba de significación, PS, el valor de P es un resultado ob-

tenido al final del experimento y el nivel de evidencia que aporta en contra de H sería diferente ante un valor de $P = 0,023$ o de $P < 0,001$, por lo que se recomienda reportar el valor de P exacto.

Resumen



*En IC se debe informar del valor de α fijado a priori.
En PS se debe reportar el valor exacto obtenido de P .
En CH se debe informar de los valores de α y β fijados a priori.
La misma concordancia en el cálculo que existe entre IC y PS se aplica también a CH. En cambio, los resultados de cada técnica deben interpretarse de acuerdo con sus objetivos.*

Resumen



*IC, PS y CH difieren en objetivos:
IC, estimar valores del parámetro
PS, aportar evidencia en contra de H
CH, decidir entre A_0 y A_1 minimizando los riesgos α y β
Pero coinciden en su mecánica:
IC $(1 - \alpha)$: $\mu \in (\bar{y} \pm 1,96 \cdot \sigma/\sqrt{n})$
PS (P): considerar inverosímil H si $\bar{y} \notin (\mu_H \pm 1,96 \cdot \sigma/\sqrt{n})$
CH (α, β): decidir A_0 si $\bar{y} \in (\mu_0 \pm 1,96 \cdot \sigma/\sqrt{n})$*

Nota técnica



En algunas situaciones, la estimación de la varianza del estimador no es la misma bajo H , H_0 , H_1 o en la muestra observada, por lo que IC, PS y CH no coincidirán plenamente. Por ejemplo, en el caso de π y p la amplitud de los intervalos puede diferir:
PS (p): aceptar si $p \in \pi_H \pm 1,96 \sqrt{[\pi_H(1 - \pi_H) / n]}$
IC $_{1-\alpha}$: $\pi \in p \pm 1,96 \sqrt{[p(1 - p) / n]}$
En el modelo lineal (comparación medias, regresión, etc.) sí que coinciden.


Interpretación errónea de P y α

Nota técnica




Es importante remarcar que en la tabla 7-1 los riesgos α y β representan probabilidades condicionadas a la fila, no a la columna. Es decir, la probabilidad de una conclusión dada la realidad de una hipótesis. Nótese que las filas representan valores del parámetro, que es una constante, mientras que las columnas representan zonas en las que se sitúa el estadístico, que sí que es una variable aleatoria.
Así, en un contraste de hipótesis (H_0 frente a H_1) para tomar una decisión (A_0 frente a A_1), α y β representan la proporción o frecuencia de decisiones erróneas a largo plazo. En el ejemplo 7.11, α es la proporción de fármacos iguales que el control (H_0) que a largo plazo son puestos en

Nota técnica




el mercado (A_1) y β , la proporción de fármacos que superan el control en un valor Δ (H_1) que no son puestos en el mercado (A_0).
 Nótese que en una prueba de significación, P (y su máximo aceptado, 0,05) indica el nivel de evidencia en contra de H , mientras que en un contraste de hipótesis α y β indican la frecuencia de decisiones erróneas.

Recuerde



P en la prueba de significación es medida de información empírica («evidencia») en contra de H , mientras que α y β en el contraste de hipótesis cuantifican la frecuencia de decisiones erróneas.

Nota técnica



En la tabla 7-1, α se lee en la primera fila como la proporción de veces en que se toma la decisión A_1 del total de ocasiones en que se parte de H_0 . Otra cosa muy distinta sería la lectura en la segunda columna de la proporción de casos provenientes de H_0 del total de ocasiones en que se ha tomado la acción A_1 . Si en el ejemplo 7-11 $\alpha = 0,05$ y $\beta = 0,10$, ambos bilaterales, un 5% (primera fila) de fármacos no eficaces (H_0) llegarán al mercado (A_1) y un 10% (segunda fila) de fármacos de eficacia Δ (H_1) no llegarán al mercado (A_0), pero la proporción (segunda columna) de fármacos no eficaces entre los que llegan al mercado es desconocida.

La tabla 7-2 expone los términos que pueden emplearse para informar el resultado de una PS o un CH.

PRUEBA DE SIGNIFICACIÓN		
Si el valor de P es...	Grande (p. ej. 0,634)	Pequeño (p. ej. 0,0001)
H es...	verosímil	inverosímil
La diferencia...	es explicable por...	no es explicable por...
La diferencia...	no es estadísticamente significativa	si es estadísticamente significativa
A nivel práctico...	no hemos logrado demostrar que la moneda está cargada	creemos que la moneda está cargada
CONTRASTE DE HIPÓTESIS		
Si el estadístico se sitúa en...	Región de aceptación	Región crítica
	Se acepta H_0	Se rechaza H_0
	Se toma la acción A_0	Se toma la acción H_1

Tabla 7-2 La prueba de significación y el contraste de hipótesis en palabras

Recuerde

Ni el riesgo α ni el valor de P pueden resumirse por «la probabilidad que tengo de haberme equivocado».

Ejercicio 7.14

¿Cuál o cuáles son correctas?

- a) El valor P es la probabilidad de equivocarse.
- b) El valor P es la probabilidad de equivocarse al rechazar H_0 .
- c) El valor P es la probabilidad de equivocarse al aceptar H_0 .
- d) El valor P es la probabilidad de observar el resultado actual (o más discrepante) si fuera cierta H_0 .
- e) El riesgo α es la probabilidad de equivocarse;
- f) El riesgo α es la probabilidad de equivocarse al rechazar H_0 .
- g) El riesgo α es la probabilidad de equivocarse al aceptar H_0 .
- h) El riesgo α es la frecuencia esperada de ocasiones en las que siendo cierta H_0 tomaremos la decisión (errónea) A_1 .
- i) El riesgo β es la probabilidad de equivocarse.
- j) El riesgo β es la probabilidad de equivocarse al rechazar H_0 .
- k) El riesgo β es la probabilidad de equivocarse al aceptar H_0 .
- l) El riesgo β es la frecuencia esperada de ocasiones en las que siendo cierta H_1 tomaremos la decisión (errónea) A_0 .

Ejemplo 7.12

La celebración final de carrera ha sido magnífica. A las 5 de la mañana los amigos se despiden, pero uno de ellos decide seguir la juerga y le pide al taxista que le lleve a una buena partida de póquer. Tras pasar los controles típicos, que su amigo creía cosa de película, consigue entrar en un 5.º piso de la calle Enrique Granados, donde se sienta a una mesa y empieza a perder dinero. Sus rivales no paran de sacar magníficas jugadas. Tanto, que él calcula que, asumiendo que no hacen trampas, la probabilidad de esos resultados (o incluso mejores) es de tan sólo una entre un millón. ¿Qué hace? Por supuesto, deja de jugar. El valor de $P = 0,000001$ le permite rechazar la hipótesis de que no le hacen trampas.

Ejemplo 7.13

En la celebración de las Navidades, un joven investigador vuelve del hospital Mount Sinai para visitar a su familia y acaban jugando al póquer con idénticos resultados que en el ejemplo anterior. A pesar de que este investigador calcu-

Ejemplo 7.13 (Cont.)

la el mismo nivel de significación anterior (asumiendo que no hacen trampas, esos resultados o mejores sólo ocurren 1 vez por millón), sigue jugando confiado, ya que no se plantea la posibilidad alternativa, de que su familia le haga trampas. Por lo que dice, «caramba, qué mala suerte tengo hoy».

Nota técnica

La estadística bayesiana lamenta que la solución de los dos ejemplos anteriores no tenga en cuenta toda la información contenida en el enunciado. Antes de empezar a jugar, el primer titulado ya podía sospechar que le harían trampas, pero no el segundo. Para poder calcular, a partir de los resultados muestrales, la probabilidad de que una hipótesis sea cierta, es preciso recurrir a una formalización del conocimiento científico previo: antes de los datos que actualmente se están analizando, ¿qué se sabía sobre este tema?, ¿qué se sabía sobre el valor del parámetro? Si se acepta representar el nivel de incertidumbre previa en forma de probabilidades sobre los diferentes valores del parámetro, ya se tienen los elementos necesarios para actualizar la información científica mediante el teorema de Bayes.

Ejercicio de Navegación

Referencias críticas sobre el abuso de las pruebas de significación, así como enlaces a paginas web aplicadas, y un *applet* muy instructivo, pueden encontrarse en:
<http://www.stat.duke.edu/~berger/p-values.html>

Recuerde

La prueba de significación, el contraste de hipótesis y el intervalo de confianza se concentran en la información aportada por los datos actuales, pero no la «suman» a la información previa.

Sólo el contraste de hipótesis permite «aceptar H_0 »

La PS no especifica ninguna hipótesis alternativa y, por tanto, no tiene definida ninguna medida análoga al riesgo β . En consecuencia, la PS no tiene ningún argumento para defender la credibilidad de H , su única hipótesis.

Nota técnica



El riesgo β puede delimitarse cuando el contraste de hipótesis tiene, como en el ejemplo del 8 y la B, la forma:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$$

Pero si la prueba de significación es de la forma:

$$H : \mu = \mu_H$$

Entonces la definición de una medida análoga al riesgo β bajo todos los posibles $\mu \neq \mu_H$ llevaría al cálculo de infinitos riesgos β , tantos como valores posibles de μ . Pero para valores próximos a μ_H ese riesgo tiende hacia $1 - \alpha$ (fig. 7-9).

La prueba de significación sólo define una hipótesis H. Este planteamiento «asimétrico» conduce a una conclusión asimétrica: si el valor de P es pequeño, se considera H inverosímil. En cambio, si P es grande, «nada se opone a aceptar H ».

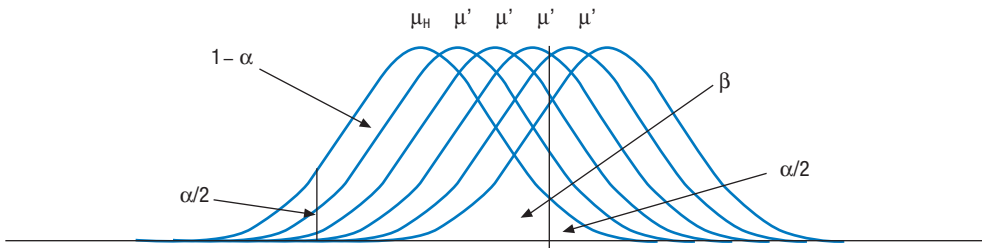


Figura 7-9 Si no hay hipótesis alternativa cerrada, el riesgo beta no está acotado.

Recuerde



En PS, «ausencia de pruebas» no es «prueba de ausencia». En PS diga «no se han detectado diferencias» en lugar de «no existen diferencias». El CH, al tener acotados α y β , permite tomar ambas decisiones.

Ejercicio 7.15



El laboratorio Yotambién S.L., para demostrar que su genérico es tan eficaz como el de la compañía Los primeros S.A., realiza un ensayo en el que compara ambos productos. Supóngase que obtiene un valor de $P = 0,23$, ¿puede concluir que ambos productos tienen la misma eficacia?

Conviene ir con mucho cuidado con las palabras que se utilizan para explicar las conclusiones de una prueba de significación. La tabla 7-2 resume algunas de las más habituales. Nótese la asimetría de la conclusión a la que se llega en ambas regiones: mientras en la zona crítica se afirma que se rechaza H («se ha demostrado la culpabilidad del acusado»), en la zona de aceptación no hay afirmaciones rotundas («absuelto por falta de pruebas»).

Ejercicio 7.16



¿Es alguna/s de las siguientes falsa?:

- En la PS se buscan evidencias en contra de H.
- El CH permite tomar ambas decisiones, A_0 y A_1 .
- Tanto P como α cuantifican áreas de las distribuciones de probabilidad, pero miden aspectos distintos.
- En la PS debe reportarse el valor exacto del valor de P .
- En el CH debe reportarse el valor previo α , usualmente, 0,05.
- En la PS, si $P > 0,05$, nada se opone a aceptar H.
- Una ventaja del CH es que permite decidir A_0 y A_1 .
- Una ventaja del CH es que cuantifica β .

Ejercicio 7.17

La prueba de significación es conservadora en el sentido de que se declara ??? la hipótesis hasta que no haya clara evidencia en su contra:

- ??? = cierta
- ??? = falsa
- La prueba de significación no es conservadora.
- Todas son incorrectas.

Ejercicio 7.18

En un estudio para comparar dos tratamientos (H_0 : son iguales) $P = 0,341$, ¿cuál o cuáles son ciertas?

- Nada se opone a aceptar H_0
- No existen diferencias
- No se han detectado diferencias
- La probabilidad de que sean diferentes es 0,341.

Lectura



Karl Popper (43) ha contribuido a incorporar los avances estadísticos en la epistemología o metodología científica. De acuerdo con esta asimetría de la conclusión de una prueba de significación, afirmó que lo único que se puede hacer con una teoría científica es ponerla a prueba y rechazarla en el caso de que encontremos pruebas en su contra, pero que nunca se podrá demostrar que sea cierta y constituya la última palabra de la ciencia en ese punto. Así, Popper dice que el criterio para establecer el estatus científico de una teoría es su refutabilidad o su testabilidad: «Para ser colocados en el rango de científicos, los enunciados o sistemas de enunciados deben ser susceptibles de entrar en conflicto con observaciones posibles», lo que es conocido como problema de la demarcación. Así, una teoría científica es más fuerte cuando más falseable es, es decir, cuanto más fácilmente podría demostrarse su falsedad (caso de ser falsa).

Interpretación del contraste de hipótesis

Desde un punto de vista formal, disponer de dos hipótesis simples, cada una con un único valor, permite definir muchas propiedades interesantes para escoger el «mejor» estadístico. Los libros clásicos de estadística matemática exponen la teoría desarrollada por Pearson y Newman sobre el contraste de dos hipótesis simples. Fisher se centra en la inferencia sobre una hipótesis, por lo que sólo puede cuantificar P y sólo puede rechazar H en lo que él llama prueba de significación.

Recuerde



La prueba de significación es un instrumento de inferencia; si la P es pequeña, Fisher recomienda modificar nuestras opiniones sobre la veracidad de H . El contraste de hipótesis es un instrumento de decisión que permite acotar los riesgos de tomar acciones erróneas. En la prueba de significación hay que reportar el valor exacto de P , por ejemplo, 0,0015; mientras que en el contraste de hipótesis hay que informar sobre las frecuencias α y β de errores a largo plazo, por ejemplo, $\alpha = 0,05$ y $\beta = 0,020$.

Historieta



Si quiere que los estadísticos no influyan en la reunión, póngalos en número par.

Detrás de las dos decisiones diferentes en las partidas de póquer de los ejemplos 7.12 y 7.13, la del garito y la de Navidad, también hay dos consecuencias muy diferentes. Levantarse de la partida de póquer del garito (sin ofender, claro) no debería conllevar consecuencias. Por su parte, seguir la partida con nuestra madre no conlleva pérdidas: incluso, en el caso de que sea cierto que hace trampas, «el dinero se queda en casa».

Ejemplo 7.14



Ya se ha dicho que Gosset era un estadístico que trabajaba en la cervecera Guinness, en su departamento de control de calidad, donde se planteaba la decisión de aceptar o rechazar una barrica de cerveza. Además de los riesgos α y β , debía considerar los costes por desechar una barrica correcta y por poner en el mercado una que no lo era.

Recuerde



El proceso de decisión, además de los riesgos de error, debe valorar también sus consecuencias, el coste que se paga por cada decisión errónea y el premio que se obtiene con las decisiones correctas.

Lectura



Uso del paracaídas para prevenir el fallecimiento y el traumatismo grave relacionados con la exposición a la fuerza de la gravedad: una revisión sistemática de los ensayos clínicos efectuados con asignación aleatoria y control [Smith et al. (44), publicado en el número bufo de Navidad]

Objetivos. Determinar la efectividad de los paracaídas para la prevención de los traumatismos graves relacionados con la exposición a la fuerza de la gravedad.

Diseño. Revisión sistemática de los ensayos clínicos efectuados con asignación aleatoria y control.

Fuentes de datos. Medline, Web o Science, Embase y las bases de datos de la Cochrane Library; sitios web apropiados y listas de bibliografía recogidas en Internet.

Selección de los estudios. Estudios en los que se demuestran los efectos del uso del paracaídas en situaciones de caída libre.

Criterio principal de valoración. Fallecimiento o traumatismo grave, definido como un traumatismo con una puntuación de gravedad > 15.

Resultados. No ha sido posible identificar ningún ensayo clínico efectuado con asignación aleatoria y control relativo al uso del paracaídas en situaciones de caída libre.

Conclusiones. Al igual que ocurre con muchas otras intervenciones para prevenir problemas de salud, la efectividad del paracaídas no ha sido objeto de una evaluación rigurosa a través de ensayos clínicos realizados con asignación aleatoria. Los defensores de la medicina basada en la evidencia han criticado la adopción de intervenciones evaluadas únicamente a través de datos de observación. Consideramos que sería de gran utilidad que los defensores más radicales de la medicina basada en la evidencia diseñaran y llevaran a cabo un ensayo clínico sobre la efectividad del paracaídas que fuera: enmascarado, controlado con placebo y con diseño cross-over, en el que cada caso es sometido a las dos intervenciones en comparación.

Ejemplo 7.15



Es bien conocido que aunque un tratamiento puede haber demostrado un cierto efecto positivo en una variable de interés, sus costes pueden aconsejar antes otra intervención sanitaria más eficiente, en el sentido de que una misma «inversión» origine un mayor «retorno», valorado en términos de salud.

Ejemplo 7.16



La pregunta de quien desea optimizar el empleo de los recursos sanitarios es: la inversión de este euro, ¿dónde me genera un mayor retorno en términos de salud?

Nota técnica

El contraste de hipótesis es el primer instrumento de la teoría de la decisión, que constituye toda una rama de la estadística y es ampliamente utilizada en otras disciplinas, como por ejemplo, la economía, donde los «costes» y los «premios» son fácilmente expresables en una única escala. El diagnóstico y el tratamiento son dos ejemplos de acciones médicas que podrían beneficiarse de las aportaciones de la teoría de la decisión.

Ejercicio 7.19

¿Cuáles de las siguientes frases son de inferencia y cuáles de decisión?

- el riesgo es mayor en pacientes de tipo A;
- el riesgo disminuye a la mitad si se adoptan las medidas X;
- la obesidad abdominal es el componente de síndrome metabólico de mayor prevalencia en mujeres;
- el valor predictivo de la escala de Z implica que debería utilizarse en el futuro para clasificar a este tipo de enfermos;
- si hay dos o menos factores de riesgo presentes y la PAS ≥ 160 o la PAD < 100 (siendo PAS < 180 y PAD < 110), conviene intentar cambios en el estilo de vida durante varios meses y luego, si se mantiene, tratamiento farmacológico.

Puede ser razonable esperar que el efecto de una intervención sea el mismo en diferentes condiciones (país, entorno de atención al paciente, raza, etc.). Y también la capacidad predictiva de un indicador. Pero no es en absoluto razonable esperar que las consecuencias de una decisión se valoren igual en diferentes entornos. Por ejemplo, el «valor» del mismo coste de un medicamento puede diferir de un país a otro.

Recuerde

El proceso de decisión implica una valoración de las consecuencias, que tienen connotaciones locales y es más difícilmente extrapolable que la mera inferencia de conocimiento.

Ejercicio adicional

Encuentre un original científico reciente que, en su discusión, vaya de la interpretación de los resultados de inferencia a la decisión ulterior sin considerar formalmente el proceso de decisión, sus riesgos y sus consecuencias en el entorno en el que propone la acción.

Nota técnica

Una ventaja del planteamiento bayesiano es que, de forma natural, conduce desde la inferencia científica hasta la toma de decisiones práctica. Permite, pues, combinar la cotidiana toma de decisiones de un profesional con la adquisición científica de conocimiento.

En otras palabras, hace transparente el paso de la investigación epidemiológica a las decisiones de salud pública o de la investigación farmacológica a las decisiones de las agencias de tecnología sanitaria.

Equivalencia

Usualmente interesa establecer «diferencias».

Ejemplo 7.17

Por ejemplo, «el riesgo de sida es mayor en toxicómanos por vía parenteral», o bien «el nuevo tratamiento es mejor que el clásico».

Pero también puede interesar establecer «equivalencia».

Definición

Un tratamiento es **equivalente** a otro si la diferencia de sus efectos no alcanza un cierto valor Δ que hace relevantes las consecuencias.

Ejemplo 7.18

Se desea establecer que: $\Delta_1 < \text{efecto} < \Delta_2$.
 Δ_1 y Δ_2 delimitan el intervalo de equivalencia.

cercanos a la igualdad, que no son relevantes desde el punto de vista práctico.

Recuerde

Ambos límites deben ser rechazados para poder establecer equivalencia.

Ejemplo 7.19

Se desea establecer, de forma simétrica, que $|\text{efecto}| < \Delta$.

El concepto de equivalencia es más amplio que el de la estricta igualdad, pues incluye también aquellos valores, cercanos a la igualdad, que no son relevantes desde el punto de vista práctico.

La figura 7-10 representa todos los posibles valores de la diferencia entre las dos medias de interés: $\mu_1 - \mu_2$.

Para establecer equivalencia se debe demostrar que las diferencias no alcanzan ni superan los dos límites especificados. Y ello puede hacerse mediante un intervalo de confianza que deberá quedar comprendido entre estos límites, lo que equivale a realizar dos pruebas que deberán rechazar ambos límites.

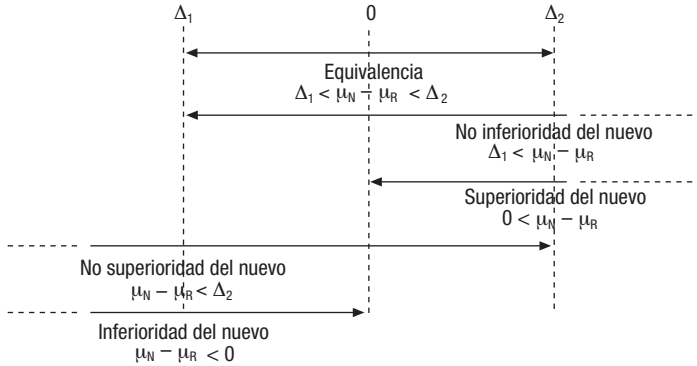


Figura 7-10 Definición de los conceptos de superioridad, equivalencia y no inferioridad.

Ejemplo 7.20



La figura 7-11 muestra 3 estudios en los que se concluiría equivalencia y 3 estudios en los que no.

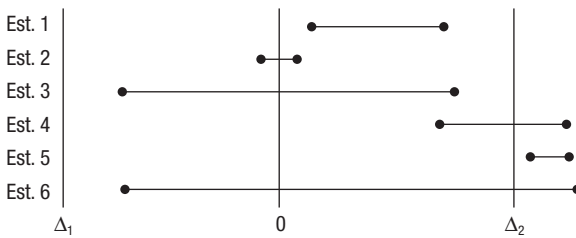


Figura 7-11 Los estudios 1 a 3, que excluyen Δ_1 y Δ_2 , permiten concluir equivalencia.

Ejercicio 7.20



Si en el ejemplo 7.20 de los datos de la figura 7-11 se hubiera hecho el CH para decidir diferencias, ¿en qué estudios de los anteriores se concluiría que los tratamientos son diferentes? Razone posibles discordancias.

Definición

Un tratamiento es *no inferior* a otro si éste no le supera en un cierto valor Δ que hace relevantes las consecuencias.

Recuerde

Los planteamientos de «no inferioridad» y «no superioridad» sólo consideran un límite Δ .

Los planteamientos de equivalencia y de no inferioridad han surgido en las agencias reguladoras del medicamento, que deben decidir si autorizan que un nuevo producto comparta un espacio ocupado previamente por otro producto.

Recuerde

Equivalencia y no inferioridad son conceptos de decisión y, por tanto, de contraste de hipótesis.

Tanto la no inferioridad como la no superioridad se establecen cada una mediante un contraste unilateral. Y la equivalencia puede establecerse mediante el uso simultáneo de ambos, por lo que el procedimiento que se utiliza recibe el nombre de contraste doblemente unilateral (CDU) o *Two-one-sided test*.

Ambos contrastes del CDU suelen realizarse con riesgo $\alpha = 0,05$, y el riesgo α global del CDU se mantiene en $0,05$. Si el IC se calcula con una confianza $1 - 2\alpha$ ($0,90$, si $\alpha = 0,05$), coincidirán las conclusiones del CDU con las del IC.

Nota técnica

Aunque la estimación por intervalo se realiza con una confianza $1 - 2\alpha = 0,90$, el criterio de decisión basado en dicho IC tendrá un riesgo $\alpha = 0,05$. Ello es así porque ambos límites de no equivalencia (que definen las dos H_0) no pueden ser simultáneamente ciertos; y, por tanto, los riesgos α con que se ha contrastado cada una no necesitan ser sumados.

Ejemplo 7.21

Un nuevo (N) antiinflamatorio tiene una tolerabilidad superior a cierto producto clásico de referencia (R). Interesa poder demostrar que sus niveles de eficacia son parecidos para poder compartir el mercado. La eficacia se mide por la proporción de casos en los que desaparece el dolor a los 30 min. Ambos fármacos serán equivalentes en eficacia si las proporciones de desaparición del dolor no difieren en más de un 8%. El intervalo de confianza (90%) de la dife-

Ejemplo 7.21 (Cont.)

rencia de ambas proporciones oscila entre -6 y $+3\%$. Dado que no alcanza los límites de la no equivalencia, se puede rechazar esta última ($\alpha = 0,05$) y autorizar (A_1) la comercialización de N.

Ejercicio 7.21

Mediante un diseño en que todos los casos pasan por ambos tratamientos, se ha obtenido en $n = 20$ casos el valor de presión arterial diastólica (PAD) tras 3 meses con el tratamiento de referencia y con el nuevo. Se ha obtenido la variable diferencia entre ambas PAD. Se ha establecido el límite de no equivalencia clínica de un hipotensor en ± 10 mmHg. Los resultados han sido $\bar{X}_D = 3$ y $S_D = 10$. Calcule el intervalo de confianza y decida si existe equivalencia.

Ejercicio 7.22

Decidir en el ejercicio anterior si existe equivalencia mediante el doble contraste de hipótesis unilateral.

Ejercicio 7.23

(El mismo hipotensor, pero menos casos). Repita los ejercicios 21 y 22, mediante IC y CH, asumiendo que los resultados han sido: $n = 5$; $\bar{X}_D = 3$ y $S_D = 10$.

Ejemplo 7.22

(Muy técnico) Para el establecimiento de equivalencia en biodisponibilidad (o bioequivalencia) se suele requerir que el cociente de los niveles en sangre entre R y N se encuentre entre $0,8$ y $1,25$; es decir, que ni R puede estar al 80% de N ($80\% = 4/5$), ni N puede estar al 125% de R ($125\% = 5/4$). En concreto, se pide que la media geométrica de dichos cocientes esté entre ambos valores o, lo que es lo mismo, que la media aritmética de la diferencia entre ambos logaritmos se sitúe entre $\ln(0,8) = -0,223$ y $\ln(1,25) = 0,223$. Así, se trabajará con la «diferencia de los logaritmos naturales» (DL), que se corresponde con el logaritmo de los cocientes que se desea mantener, en promedio, entre los dos valores requeridos. En un diseño de datos apareados los resultados han sido: $n = 12$, $\bar{X}_{DL} = 0,1$ y $S_{DL} = 0,2$.

Mediante $IC_{90\%}$:

$$\mu_{LR-LN} \in X_{LR-LN} \pm t_{1-2\alpha} S_{LR-LN} / \sqrt{n}$$

$$\mu_{LR-LN} \in 0,1 \pm t_{1,0'90} 0,2 / \sqrt{12}$$

$$\mu_{LR-LN} \in (-0,004, 0,204)$$

Ejemplo 7.22 (Cont.)

Mediante $PDU_{\alpha = 0,05}$:

$$\begin{cases} H_{0A} : \mu_{LR-LN} \leq -0,223 \\ H_{1A} : \mu_{LR-LN} > -0,223 \end{cases}$$

$$t_1 = (0,1 - (-0,223)) / (0,2/\sqrt{12}) = 5,595 \rightarrow P < 0,001$$

$$\begin{cases} H_{0B} : \mu_{LR-LN} \geq 0,223 \\ H_{1B} : \mu_{LR-LN} < 0,223 \end{cases}$$

$$t_2 = (0,1 - 0,223) / (0,2/\sqrt{12}) = -2,130 \rightarrow P \approx 0,028$$

Por lo que tanto el IC como la PDU permiten concluir la equivalencia de ambos productos.

Nota técnica

Los planteamientos de equivalencia que se han resuelto en estos ejemplos hacen referencia a la equivalencia en media. Ello implica que un paciente tiene los mismos valores esperados bajo ambos productos en comparación y, por tanto, ambos preparados o productos son igualmente aconsejables para un nuevo paciente (equivalencia poblacional o prescribibilidad). Para que dos preparados se puedan intercambiar en un paciente ya tratado (equivalencia individual o intercambiabilidad) es necesario, además, que no exista interacción entre el preparado y el paciente, es decir: que la diferencia (quizá nula) entre ambos preparados sea la misma para todos los pacientes. La demostración de esta condición ha sido exigida por algunos, lo que ha resultado en una mayor dificultad para la salida al mercado de productos genéricos.

Soluciones a los ejercicios

7.1 La respuesta correcta es la *d*. La *a* es incorrecta, ya que «conocer el valor del parámetro» es el objetivo de la estimación, quizás por intervalo de confianza, no del contraste de hipótesis. La *b* no es correcta, ya que en la prueba de significación la hipótesis forma parte del enunciado del problema y debe siempre ser previa a la obtención de los datos. La *c* no es correcta, ya que se buscan pruebas en contra de la hipótesis *H* que se desea rechazar.

7.2 La respuesta correcta es la *a* ya que debe situarse en *H* aquello que se desea rechazar para así demostrar su complementario.

7.3 El proceso formal de decisión es el siguiente:

- i) Variable: preferencia por A o por B
- ii) Estadístico: proporción p que prefieren A
- iii) Hipótesis *H*: $\pi = 0,5$ (ambos fármacos tienen igual preferencia)
- iv) Si *H* es cierta: $p \rightarrow N(\pi, \pi(1 - \pi)/n) = N(0,5, 0,025^2)$.

Premisas: muestra grande $\pi \cdot n > 5$ y $(1 - \pi) \cdot n > 5$

v) $\hat{z} = (p - \pi) / \sqrt{\pi(1 - \pi)/n} = (0,58 - 0,50) / \sqrt{0,5 \cdot 0,5/400} = 0,08 / 0,025 = 3,2$

vi) El valor 0,0007 correspondiente a $\hat{z} = 3,2$ se encuentra en la tabla 4-2 en la fila que empieza por 3,2_ y la columna encabezada por _,_0. Si se le suma su simétrico (para $Z < -3,2$), se obtiene $P = 0,0014$. Por ello, puede rechazarse, con $P = 0,0014$ que ambos tratamientos sean iguales: A y B no tienen la misma preferencia. [Nota, en la tabla 4-1, en cambio, el valor máximo de tablas es 2,58 que deja por fuera 0,01. Con esta tabla, el valor reportado sería $P < 0,01$]

vii) $IC_{95\%}: \pi \pm Z_{\alpha/2} \sigma_p = 0,58 \pm 1,96 \cdot 0,025 \approx 0,58 \pm 0,05 = [0,53, 0,63]$

La «auténtica» preferencia π por A se encuentra entre 53 y 73%.

7.4 v) $\hat{z} = (p - \pi) / \sqrt{\pi(1 - \pi)/n} = (0,53 - 0,50) / \sqrt{0,5 \cdot 0,5/400} = 0,03 / 0,025 = 1,2$

vi) El valor 0,1151 correspondiente a $\hat{z} = 1,2$ se encuentra en la tabla 4-2 en la fila que empieza por 1,2_ y la columna encabezada por _,_0. Si se le suma su simétrico (para $Z < -1,2$), se obtiene $P = 0,2302$. Por ello, dado $P = 0,2302 < 0,05$, nada se opone a aceptar que ambos tratamientos tienen la misma preferencia.

vii) $IC_{95\%}: p \pm Z_{\alpha/2} \sigma_p = 0,53 \pm 1,96 \cdot 0,025 \approx 0,53 \pm 0,05 = [0,48, 0,58]$

7.5 Las tres primeras son correctas, la cuarta es una tontería que no tiene nada que ver y la quinta es un error muy habitual de interpretación del valor de P , que cuantifica la probabilidad de unos resultados condicionando a una cierta hipótesis, no la probabilidad de que sea cierta una hipótesis condicionando a unos resultados. Más adelante se insistirá en esta distinción.

7.6 $\hat{z} = (\bar{y} - \mu_0) / (s/\sqrt{n}) = (505 - 500) / \sqrt{(100^2/36)} = 0,3$

El valor 0,3821 correspondiente a $\hat{z} = 0,3$ se encuentra en la tabla 4-2 en la fila que empieza por 0,3_ y la columna encabezada por _'_0. Si se le suma su simétrico (para $Z < -0,3$), se obtiene $P = 0,7642$. Por ello, nada se opone a aceptar que la media es de 500. Es decir, no se ha logrado demostrar que $\mu > 500$

$IC_{95\%}: \bar{y} \pm Z_{\alpha/2} \sigma/\sqrt{n} = 505 \pm 1,96 \cdot 100 / \sqrt{36} \approx 505 \pm 33 = [472, 538]$

Se sabe, con una confianza del 95%, que μ está entre 472 y 538

$$7.7 \hat{t} = (\bar{y} - \mu_0) / (S/\sqrt{n}) = (79 - 70) / \sqrt{(12^2/16)} = 3$$

Debería buscarse en la tabla 6-1 el valor correspondiente a una t de Student con 15 grados de libertad. Se encuentra que el valor 2,977 de 14 grados de libertad deja por fuera 0,01. La probabilidad que deja fuera el valor 3 será $P < 0,01$ y, por tanto, se ha logrado demostrar que $\mu > 70$. [Nota: Excel proporciona el valor exacto $P = 0,00897274$, pero es más legible $P < 0,01$]

7.8 La correcta es la respuesta c .

7.9 Ahora cambian los niveles P de significación, que deben dividirse por 2: y se obtiene 0,3821 y 0,0045, por lo que las conclusiones no cambian. Nótese que este planteamiento unilateral es más adecuado en estos dos ejemplos.

7.10 Es correcta la respuesta e , ya que se trata de una prueba unilateral cuya H incluye el 0 y todos los valores negativos. Dado que el estadístico se sitúa en H , la única conclusión posible en una prueba de significación es «nada se opone a aceptar H ». Nótese que en un contraste de hipótesis, ambos riesgos, α y β , están acotados, lo que permite tomar ambas decisiones (en una PS , sólo podía rechazarse H).

7.11 Nótese que, al pasar de PS a CH , el enunciado debería pedir una decisión.

En 7.3, al ser $\bar{z} = 3,2 > Z_{\alpha/2} = 1,96$, se rechaza H .

En 7.4, al ser $\bar{z} = 1,2 < Z_{\alpha/2} = 1,96$, nada se opone a aceptar H .

En 7.6, al ser $\bar{z} = 0,3 < Z_{\alpha/2} = 1,96$, nada se opone a aceptar H .

En 7.7, al ser $\bar{z} = 3,0 > t_{14, \alpha/2} = 2,145$, se rechaza H .

7.12 Es correcta la respuesta b .

7.13 Es correcta la respuesta b .

7.14 Efectivamente, las correctas son las tres largas d , h , e l : ¡es peligroso abreviar!

7.15 No, los resultados de su experimento lo único que le dicen es que, asumiendo que los dos productos sean iguales, la probabilidad de obtener unos resultados como los suyos (o más extremos) no es muy pequeña. Por tanto, no puede demostrar que H sea falsa, lo que no equivale a haber demostrado que H sea cierta. Por ello, no puede afirmar que tengan la misma eficacia. Más adelante se estudia cómo puede demostrarse la equivalencia entre dos productos.

7.16 Todas son ciertas.

7.17 La correcta es la respuesta a , si bien formalmente la frase más correcta sería: «por el momento, se acepta tentativamente H , ya que no existen evidencias en su contra».

7.18 Son correctas la a , y la c .

7.19 a y c son claramente inferencia; así como d y e , decisión. b hace inferencia sobre las consecuencias de una decisión.

7.20 Se rechazaría la H de estricta igualdad en los estudios 1, 4 y 5. Nótese que el estudio 1 tiene un IC razonablemente estrecho, que le permite concluir tanto equivalencia (porque excluye Δ_1 y Δ_2) como diferencias (porque excluye 0); es decir, los dos tratamientos no son estrictamente iguales, pero sus diferencias no alcanzan el criterio de relevancia. El estudio 6, en cambio, tiene un IC tan amplio, aporta tan poca información, que no le permite ni rechazar la estricta igualdad ni el límite de relevancia clínica. Los restantes estudios no presentan estas paradojas: el 2 y el 3 no consiguen rechazar la estricta igualdad y sí logran establecer equivalencia (aunque el 2 tiene un IC más estrecho que implica que se dispone de mucha información); y el 4 y el 5 consiguen rechazar la estricta igualdad y no logran establecer equivalencia. Nótese que el estudio 5 no incluye el margen de equivalencia Δ_2 , pero que se sitúa al lado de la no equivalencia (lo que coincide con el planteamiento unilateral).

7.21 $IC_{90\%}: \mu_D \in \bar{x}_D \pm t_{1-2\alpha} \cdot S_D/\sqrt{n}$

$$\mu_D \in 3 \pm t_{19,0,90} \cdot 10/\sqrt{20} \rightarrow \mu_D \in 3 \pm 1,729 \cdot 10/\sqrt{20} \rightarrow \mu_D \in (-0,866, 6,866).$$

Luego la media de las diferencias entre las presiones de ambos hipotensores está entre $-0,866$ (el de referencia consigue presiones más bajas en media: gana por $0,866$ mmHg) y $+6,866$ (el nuevo consigue presiones más bajas: gana por $6,866$ mmHg). Luego la diferencia entre ambos fármacos está entre los límites -10 y $+10$: se ha establecido equivalencia.

7.22 CH: $PDU_{\alpha=0,05} \begin{cases} H_{0A}: \mu_D \leq -10 \\ H_{1A}: \mu_D > -10; & t_1 = (3 - (-10))/(10/\sqrt{20}) = 5,814 > 1,645 = z_{0,05} \\ H_{0B}: \mu_D \geq 10 \\ H_{1B}: \mu_D < 10; & t_2 = (3 - 10)/(10/\sqrt{20}) = -3,130 < -1,645 = z_{0,95} \end{cases}$

Con riesgo $\alpha = 0,05$, se aconseja actuar autorizando el genérico.

7.23 $IC_{90\%}: \mu_D \in \bar{X}_D \pm t_{1-2\alpha} \cdot S/\sqrt{n}$

$$\mu_D \in 3 \pm t_{4,0,90} \cdot 10/\sqrt{5} \rightarrow \mu_D \in 3 \pm 2,132 \cdot 10/\sqrt{5} \rightarrow \mu_D \in (-6,534, 12,534).$$

Ahora, el $IC_{90\%}$ sobrepasa el dintel superior que marca la no equivalencia y, por tanto, no se puede defender que haya equivalencia.

CH: $PDU_{\alpha=0,05} \begin{cases} H_{0A}: \mu_D \leq -10 \\ H_{1A}: \mu_D > -10 & t_1 = (3 - (-10))/(10/\sqrt{5}) = 2,907 \rightarrow P \approx 0,022 < 0,05 = \alpha \\ H_{0B}: \mu_D \geq 10 \\ H_{1B}: \mu_D < 10 & t_2 = (3 - 10)/(10/\sqrt{5}) = -1,565 \rightarrow P \approx 0,096 > 0,05 = \alpha \end{cases}$

Asimismo, aunque la primera prueba aún permite afirmar que la media de las diferencias está por encima de -10 , la segunda no ha permitido establecer que esté por debajo de $+10$. Por tanto, no se ha podido demostrar que la media de las diferencias de ambos hipotensores esté entre -10 y $+10$. Por tanto, la acción debe ser no autorizar el genérico (A_0).

Capítulo 8:

Comparación de dos grupos

Presentación

Este capítulo expone la popular prueba t de Student para comparar 2 medias, de la que se ofrece una interpretación intuitiva. Se explica por qué el proceso de inferencia estadística requiere un grupo control y las ventajas de concentrar el estudio en el análisis de la media. Finalmente, mediante el método de aparear los datos, se insiste en el concepto de eficiencia estadística.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Percibirá la necesidad de disponer de un grupo control.
- Percibirá la necesidad de que los grupos en comparación difieran sólo en el tratamiento.
- Interpretará la diferencia entre medias como la magnitud del efecto en cada caso.
- Interpretará la amplitud de un intervalo de confianza como la ignorancia sobre el valor del parámetro, no como la variabilidad entre las unidades.
- Interpretará la «razón t » como un cociente «señal/ruido».
- A partir de las medias y de las varianzas de cada grupo, sabrá calcular la prueba t de Student de comparación de dos medias con datos independientes.
- Percibirá las ventajas de que se cumplan las premisas necesarias para realizar la prueba t .
- A partir de la media y de la varianza de la diferencia, sabrá calcular la prueba t de Student de comparación de dos medias con datos apareados.
- Sabrá que el apareamiento permite aumentar la precisión de la estimación del efecto del tratamiento.
- Decidirá que dos muestras están apareadas si los valores de las parejas son similares.

La importancia de tener un control

En los capítulos anteriores se ha visto cómo inferir los resultados desde una muestra a una población. Se ha aprendido a realizar intervalos de confianza y pruebas de hipótesis sobre la media o la probabilidad. Pero sólo se ha expuesto el estudio de una muestra, lo que limita las posibilidades de comparar, por ejemplo, resultados entre dos procedimientos. Incluso en la evaluación de un solo procedimiento, conviene tener un grupo control que proporcione una referencia a la comparación.

Ejemplo 8.1



Un profesor, preocupado por la gran «bolsa» de repetidores que tiene su asignatura, introduce una novedad pedagógica en su docencia: un programa informático de autoaprendizaje que permite al alumno ir comprobando su nivel de progreso en la resolución de problemas. Para medir el efecto de dicho programa, evalúa a los repetidores al inicio y al final del curso y estima, definiendo como variable respuesta Y , la diferencia de conocimientos entre el final y el inicio. El $IC_{95\%}$ de la media de este incremento es de entre 2 y 3 puntos, por lo que concluye que su sistema tiene esa eficacia.

Falso. O, por lo menos, no se sostiene, ya que existen muchas otras variables que pueden ser explicación alternativa a estas diferencias. Especialmente, todo lo que los repetidores hayan estudiado por otros métodos. O incluso, ¿cómo garantizar que el segundo examen no era más fácil que el primero?

Por otro lado, la pregunta de interés práctico es, en general, la elección entre dos opciones alternativas. Puede tener interés, a nivel teórico, conocer el rendimiento o el coste de un sistema; pero, si no hay alternativa, se utilizará el único sistema disponible, por lo que la pregunta no tiene interés práctico. Al menos, la opción alternativa debería ser no hacer nada: o bien implementar el sistema nuevo o bien dejarlo todo igual. Una vez se han definido las dos posibilidades que se desean comparar, un buen diseño de recogida de datos debe informar sobre ambas.

Recuerde



La pregunta más interesante incluye la comparación con una alternativa de control.

Ejemplo 8.1 (Cont.)

El profesor anterior ha seleccionado a los repetidores que cumplían ciertos criterios de homogeneidad y en los que creía que dicho sistema aún podía tener efecto porque no eran situaciones especiales (p. ej., era la primera vez que repetían). A estos estudiantes les solicita permiso para participar en una experiencia docente y requiere su compromiso para seguir las normas de la misma. Los que aceptan son repartidos al azar en dos grupos, uno con acceso al sistema nuevo y el otro, sin. Ambos grupos pasan, en idénticas condiciones, la evaluación inicial y la final, que son corregidas de forma automática. Define como variable de interés el incremento de conocimiento, definido como la diferencia entre la evaluación final y la inicial. El profesor calcula el intervalo de confianza de la diferencia entre los que tienen acceso al sistema respecto a los que no lo tienen. Igual que antes (pero comparando con el otro grupo), el IC_{95%} de $\mu_1 - \mu_2$ muestra en el grupo experimental un incremento que supera al del grupo control entre 2 y 3 puntos. Este resultado es mucho más sólido. Se cree, con una confianza del 95%, que utilizar este programa mejorará entre 2 y 3 puntos el rendimiento de los repetidores.

Comentario

«Más sólido» pero no definitivo, porque (aunque menos que antes) aún se podría encontrar algún punto delicado. Por ejemplo, si el efecto ha sido realmente debido al sistema de autoaprendizaje o puede ser explicado por la motivación extra de sentirse «agraciado» en el reparto de los grupos. En este último caso, la mejora observada no se reproduciría en el futuro, al incorporar el sistema al aprendizaje habitual.

Recuerde

Al comparar dos tratamientos, conviene que el tratamiento recibido sea la única diferencia entre los dos grupos.

Ejemplo 8.2

Un servicio de traumatología decide evaluar el efecto de implantar una nueva prótesis sustitutiva tras la intervención. Para valorarlo, comparan la media de la calidad de vida de unos pacientes del año anterior (previa a la nueva prótesis) con la de otros pacientes del año siguiente. Los resultados son estadísticamente significativos y favorecen

Ejemplo 8.2 (Cont.)

al nuevo planteamiento. ¿Han demostrado que el nuevo sistema es mejor? No, aunque sean resultados alentadores no son, ni mucho menos, definitivos. Tres grupos de críticas serían las siguientes:

- 1) ¿Son comparables las características de los pacientes de los dos años?
- 2) ¿Han habido otros cambios de un año a otro? ¿Los profesionales han mejorado sus habilidades? ¿Se han introducido también otras terapias?
- 3) ¿Qué significa esta calidad de vida? ¿Puede haber sido valorada de forma más generosa en el segundo año?

Nótese que estos tres grupos de preguntas están buscando terceras variables que puedan ser una explicación alternativa de las diferencias encontradas y, por lo tanto, estén sesgando la comparación en el sentido de generar grupos que difieren en algo más que la intervención en estudio.

Recuerde

Al comparar dos tratamientos, las tres preguntas clave son:

- 1) *¿Son iguales los grupos al inicio?*
- 2) *¿Se ha podido introducir alguna diferencia durante el período experimental?*
- 3) *La evaluación, ¿ha podido introducir otras diferencias?*

Ejercicio 8.1

¿Cuáles son las respuestas ideales (SÍ/NO) a estos tres grupos de preguntas?

Ejercicio 8.2

Proponga un diseño que permita evaluar los efectos de la novedad terapéutica del ejemplo 8.2.

Recuerde

Dos intervenciones que se deseen comparar deben tener:

- 1) *Dos grupos iniciales iguales.*
- 2) *El seguimiento idéntico.*
- 3) *La misma evaluación.*

La importancia de estudiar la medias

Si se dispone de una variable con distribución normal, el análisis habitual es la comparación de las medias en los dos grupos de tratamiento. Pero ¿por qué se comparan las medias si luego se aplicará el nuevo tratamiento a las unidades? ¿Qué información aportan estas medias sobre lo que pasa en cada unidad? Veamos las razones.

Supóngase que el procedimiento nuevo que se desea comparar con el estándar tiene un efecto Δ que es el mismo en todas las unidades. El resultado de aplicar el nuevo procedimiento provocará una traslación de la distribución original exactamente igual a Δ , como muestra la figura 8-1.

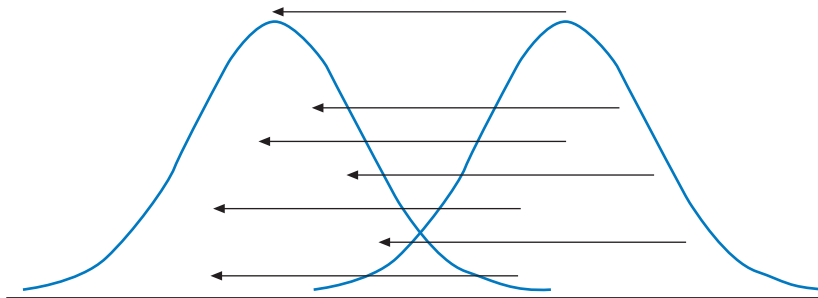


Figura 8-1 Comparación de dos distribuciones normales. Efecto constante, aditivo.

Ejemplo 8.3



Cierto fármaco antihipertensivo desciende la presión sistólica en exactamente 10 mmHg. Este descenso es el mismo en todos los pacientes, sea cual sea su nivel anterior de presión. Los pacientes siguen teniendo diferentes valores de presión sistólica, pero porque ya eran diferentes antes del tratamiento, no porque ésta haya incluido variabilidad en el proceso. Nótese que, sea cual sea la forma de la distribución de la presión sistólica antes del tratamiento, esta distribución se trasladará exactamente 10 unidades, pero conservando la misma forma y la misma dispersión.

Recuerde



Si el efecto Δ es el mismo en todas las unidades, basta con estimar la diferencia entre las medias de ambas poblaciones para conocer el efecto que se dará en cada unidad de esa población.

Nota técnica

Puede demostrarse que la manera más **eficiente** (más precisa) de comparar la posición de dos variables con distribución normal es, precisamente, mediante la comparación de sus medias. De aquí la importancia de comparar las medias: si se asume que la forma de ambas distribuciones es idéntica, es el procedimiento estadístico más eficiente para estimar un efecto único que se observará en cada una de las unidades.

Una consecuencia muy importante de este efecto común en todos los casos es que la distribución es la misma bajo el procedimiento estándar y bajo el nuevo. Por lo tanto, también será igual su varianza.

Nota técnica

Esta situación de igual varianza recibe el nombre técnico de **homocedasticidad**. Permite creer que la diferencia entre las medias está estimando el efecto en cada unidad de la población.

Ejercicio 8.3

Al comparar el efecto de dos tratamientos A y B, se obtiene un intervalo del 95% de confianza de las diferencias de sus medias que va de 5 a 8. Se cree, con una confianza del 95% que:

- a) La auténtica diferencia de medias poblacionales se encuentra entre 5 y 8.
- b) El efecto diferencial de A respecto a B consiste en descender entre 5 y 8 unidades más.
- c) En el 95% de las medias muestrales la diferencia se encuentra entre 5 y 8.
- d) En el 95% de los pacientes, la diferencia del efecto está entre 5 y 8.

Ejercicio 8.4

En un estudio comparativo entre dos terapias alternativas A y B, el $IC_{95\%}$ de la diferencia de las medias de los dos grupos dice que el descenso de la carga viral ha sido entre 10 y 100 copias superior en el grupo tratado con A. Esto significa que (elija una):

- a) En los casos en los que el tratamiento A ha sido más favorable, en el grupo A se ha conseguido un descenso de 100 unidades más que en el grupo B, mientras que en los menos favorables ha sido de 10 copias.
- b) En los respondedores, la ventaja de A ha sido de 100 copias, mientras que en los no respondedores, ha sido de 10 copias.

Ejercicio 8.4 (Cont.)

- c) A desciende más que B un valor único y común para todos los casos, pero que se desconoce y debe ser uno de los comprendidos entre 10 y 100 copias.
d) Ninguna de las anteriores es correcta.

Recuerde

La amplitud del intervalo de confianza refleja el grado de ignorancia, no la variabilidad del efecto.

Nota técnica

En el entorno de comparación de medias no es correcto hablar de «respondedores» y «no respondedores». Si se sospecha que puede haber dos grandes grupos de pacientes, los que responden y los que no (respuesta dicotómica), un análisis correcto podría ser la comparación de la proporción de respondedores entre los grupos.

Resumen

Si la distribución en los dos grupos tiene la misma forma y la misma dispersión, la comparación de dos poblaciones se reduce a la comparación de sus medias.

Contraejemplo 8.4

Supóngase ahora que el efecto del tratamiento consiste en reducir un 20% la presión sistólica, de forma que a un paciente con 100 mmHg se la baja a 80 mmHg y a uno de 150 mmHg se la baja a 120 mmHg. Ahora, este efecto «proporcional» no respeta la variabilidad de las observaciones, ya que provoca mayor descenso cuanto mayor es la presión inicial, como puede verse en la figura 8-2.

Ahora, la diferencia entre medias ya no representa el efecto en cada unidad, lo que podría complicar mucho la interpretación: como el efecto dependería del valor inicial debería estimarse por separado según la severidad de la presión inicial. Afortunadamente, las matemáticas acuden en nuestra ayuda: una simple transformación logarítmica (la misma que define el pH) permitirá convertir este modelo multiplicativo («descenso proporcional») en un modelo aditivo en el que poder utilizar la diferencia de las medias.

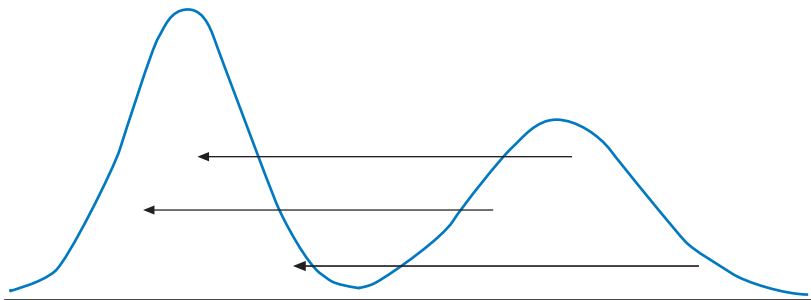


Figura 8-2 Comparación de dos distribuciones normales. Efecto no constante.

Comparación de dos medias en muestras independientes. Varianzas iguales

Veamos ahora cuál es la solución numérica de la comparación de dos medias. Para empezar, se asume que se conocen las varianzas y que las distribuciones de las dos variables en comparación son normales. Se asume también que los procesos de muestreo de los tratamientos en comparación son aleatorios e independientes.

Para estudiar la **diferencia entre las medias poblacionales** μ_1 y μ_2 se recurre al estadístico $(\bar{y}_1 - \bar{y}_2)$ o **diferencia de las medias muestrales**.

¿Qué **distribución** sigue este estadístico?: **normal**, ya que es una combinación de variables normales. Sabiendo la forma, queda por especificar su centro y su dispersión: ¿qué valen su esperanza y su varianza? La **esperanza** de la diferencia de las medias muestrales es directamente la diferencia de sus esperanzas o medias poblacionales:

$$E(\bar{y}_1 - \bar{y}_2) = E(\bar{y}_1) - E(\bar{y}_2) = \mu_1 - \mu_2$$

La diferencia entre las medias muestrales $(\bar{y}_1 - \bar{y}_2)$ estima, sin sesgo, cuánto vale la diferencia de las medias poblacionales $(\mu_1 - \mu_2)$.

La **varianza** de la diferencia de las medias muestrales se convierte, en esta situación de muestras independientes, en la suma de las varianzas de las medias:

$$V(\bar{y}_1 - \bar{y}_2) = V(\bar{y}_1) + V(\bar{y}_2) = \sigma^2/n_1 + \sigma^2/n_2$$

Comentario



La varianza de una resta es la suma de las varianzas. Se puede imaginar este resultado si se tiene en cuenta que se trata de variables aleatorias independientes y, para conocer su diferencia, las oscilaciones aleatorias de cada variable contribuyen al ruido de estimación. Por ejemplo, en un primer experimento, la media muestral del grupo 1 puede oscilar hacia arriba, y la del grupo 2, independiente del anterior, hacia abajo. En un segundo experimento, puede ser al revés o similar... Ambos errores aleatorios se añaden en la estimación de la diferencia de las medias poblacionales $\mu_1 - \mu_2$.

Recuerde



La diferencia de las medias muestrales sigue una distribución normal cuyo centro es la diferencia de las medias poblacionales y cuya dispersión es la suma de la oscilación de cada media.

$$(\bar{y}_1 - \bar{y}_2) \rightarrow N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$$

Una vez conocida la distribución de la diferencia de las medias muestrales, ya se puede proceder de forma similar a como se hacía para el caso de una muestra.

Si se asume que se conocen las varianzas poblacionales, se puede tipificar esta diferencia de medias en forma de cociente señal/ruido:

$$\hat{z} = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

Y se puede realizar inferencia, bien mediante un intervalo de confianza:

$$IC_{1-\alpha}(\mu_1 - \mu_2): (\mu_1 - \mu_2) \in (\bar{y}_1 - \bar{y}_2) \pm Z_{\alpha/2} \cdot \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$$

bien mediante un contraste de hipótesis:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 - \mu_2 = \Delta \end{cases}$$

Bajo H_0 :

$$\hat{z} = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2)}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \rightarrow N(0,1)$$

Como usualmente σ_1 y σ_2 son desconocidas, deberán ser estimadas. Si se puede asumir igualdad de varianzas ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) entonces, S_1^2 y S_2^2 serán estimadores del mismo parámetro σ^2 , por lo que pueden ser combinados en un **estimador único ponderado** por la información que aporta cada uno:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

Nota técnica



Si ambas muestras son del mismo tamaño, este estimador S^2 se reduce al promedio de las varianzas de cada muestra:

$$S^2 = \frac{(S_1^2 + S_2^2)}{2}$$

Este estimador promedia el cuadrado de todas las desviaciones con la media, dándoles el mismo peso vengan del grupo que vengan. Para cada grupo, utiliza su media.

Finalmente, se llega al conocido estadístico llamado «razón t o t ratio»:

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow N(0,1)$$

En el numerador está la diferencia entre las medias muestrales, o *señal* proporcionada por los datos. En el denominador está la raíz de la varianza del estimador que figura en el numerador, es decir, el error típico de esta señal: su *ruido* aleatorio.

Recuerde



La razón t es el cociente entre la señal y el ruido.

El estadístico *razón t* sigue una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad. Las premisas necesarias son: 1) variable, con distribución normal; 2) varianzas iguales, y 3) muestras independientes.

Nota técnica



Aunque la variable original en estudio no siga la distribución normal, la aproximación se considera razonable si el tamaño de ambas muestras supera los 30 casos. Cuanto más se asemeje a la distribución normal la variable en estudio más pequeño es este número requerido de casos.

Ejercicio 8.5



- El estadístico «razón t » sigue una t de Student con
- $n_1 + n_2 - 2$ grados de libertad.
 - $n - 1$ grados de libertad.
 - 30 grados de libertad.
 - Este estadístico no sigue una t de Student.

Ejercicio 8.6



- Para que el estadístico «razón t » siga una t de Student con $n_1 + n_2 - 2$ grados de libertad, las premisas necesarias son:
- MAS e independientes.
 - Normalidad de la variable original.
 - Homocedasticidad o igualdad de varianzas.
 - Todas son correctas.

Ejemplo 8.5



Para tomar la decisión de sustituir el tratamiento 2 por el tratamiento 1, se ha realizado un ensayo clínico ($n_1 = 50$ y $n_2 = 100$) cuyos resultados han sido: $\bar{y}_1 = 24$ e $\bar{y}_2 = 21$, siendo $s_1 = 8$ y $s_2 = 6$. Presuponiendo que las varianzas poblacionales son iguales, ¿se puede considerar que ambos tratamientos inducen la misma calidad de vida?

Solución:

- 1) $\begin{cases} H_0 : \mu_1 = \mu_2 \text{ (misma calidad de vida)} \\ H_1 : \mu_1 - \mu_2 = \Delta \text{ (El tratamiento 1 es superior en } \Delta \text{ unidades)} \end{cases}$

- 2) **Estadístico:**

$$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- 3) **Distribución** bajo $H_0 : \hat{t} \rightarrow t_{n_1 + n_2 - 2}$

Premisas: Y_1, Y_2 normales;

$\sigma^2_1 = \sigma^2_2$ MAS e independientes

- 4) **Regla decisión:** rechazar H_0 ($\alpha = 0,05$)

$$\text{si } \hat{t} > t_{n_0 + n_0 - 2, 0,975} = 1,976$$

- 5) **Cálculo:** $S^2 = (8^2 \cdot 49 + 6^2 \cdot 99) / 148 \approx 44$

$$\hat{t} = (24 - 21) / \sqrt{[(44/50) + (44/100)]} \approx 2,61$$

La señal es más del doble del ruido aleatorio.

$$P[|t_{n_1 + n_2 - 2}| > 2,61] < 0,01$$

- 6) **Decisión:** dado que $2,61 > 1,976$

se rechaza que $\mu_1 = \mu_2$ con riesgo $\alpha = 0,05$

- 7) $IC_{95\%}(\mu_1 - \mu_2) = (\bar{y}_1 - \bar{y}_2) \pm 1,976 \sqrt{(S^2/n_1 + S^2/n_2)}$
 $= 3 \pm 1,976 \cdot 1,149 = 3 \pm 2,27 = [0,73, 5,27]$

- 8) **Conclusión** práctica: el tratamiento 1 tiene una calidad de vida entre 0,73 y 5,27 puntos mayor.

Sustituiremos el tratamiento 2 por el tratamiento 1.

Recuerde



La «razón t » o cociente señal/ruido empieza a ser relevante a partir de 2 (la señal debe doblar al ruido).

Ejercicio 8.7



Llano et al. (45) comparan dos momentos evolutivos diferentes (SI y NSI) en grupos de pacientes ($n_1 = 56$, $n_2 = 75$) de sida. Las medias de CD4 han sido 125 y 329 y las desviaciones típicas, 155 y 223. Calcule el estimador de la varianza común y estime por intervalo la diferencia media entre ambos tratamientos.

Nota técnica



El estadístico «razón t» es muy importante. Considerémoslo desde otro punto de vista.

$$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2)}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Ahora es el producto de dos cantidades. La primera representa, como antes, la señal. Pero ahora, al dividirla por la desviación típica ha perdido sus unidades: está estandarizada o tipificada. La segunda cantidad representa la «magnitud» de la muestra.

Así, la razón «t» es la **señal tipificada** multiplicada por la **cantidad de información** que aporta el diseño muestral.

Nota técnica



Para un tamaño muestral total N prefijado ($N = n_1 + n_2$), el máximo de esta «magnitud» es para tamaños muestrales iguales ($n_1 = n_2$). Es decir, si se dispone de n unidades, el diseño de dos muestras independientes más eficiente consiste en hacerlas de igual tamaño.

Recuerde



Si dispone de un total de N unidades, el diseño óptimo, que conduce al IC más estrecho, dispone N/2 en cada grupo.

Comparación de dos medias en muestras independientes. Varianzas diferentes

Si no se puede asumir que las varianzas sean iguales aparecen dificultades.

Comentario



La primera dificultad es que la diferencia entre las medias ya no puede representar al efecto «común» en cada uno de los casos. Lo único que puede representar es cierto efecto promedio y pierde utilidad, ya que aparece incertidumbre sobre cuál será el efecto en una unidad.

La segunda dificultad es más teórica: resulta que el estadístico, análogo al anterior, que ahora utilizaríamos:

$$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

ya no sigue una t de Student, pues el denominador no es ahora una sola variable, sino dos.

La mejor solución consiste en encontrar una transformación de Y, en la que las varianzas sean iguales (y la distribución, normal), ya que se resuelven, a la vez, varios problemas.

Nota técnica

Podría darse la situación en la que el efecto en estudio no fuera aditivo (una constante que se suma o resta a cada observación) sino proporcional, multiplicativo como en el contraejemplo 8.4 del tratamiento que reducía la presión en un 20%. Ahora, la constante que permite comparar ambas distribuciones debería obtenerse mediante un cociente en lugar de una resta. Vimos que tomar logaritmos permite convertir ese cociente en una resta. Además, debido a que la normal aparece cuando la variable en estudio es la suma, no el producto, de pequeños efectos, al aplicar logaritmos, la distribución resultante suele aproximarse mejor a una normal.

Nota técnica

Existen muchas soluciones al problema de cuál es la distribución del estadístico señal/ruido en caso de heterocedasticidad. Permiten calcular el nivel P de significación y el intervalo de confianza, pero recuerde que la interpretación ya no será tan directa, porque el valor estimado en el IC ya no puede representar el efecto en cada paciente, sino a un cierto efecto «promedio».

Comparación de dos medias con muestras dependientes o apareadas

Supóngase que, para comparar 2 tratamientos, A y B, los pacientes en que se comparan son los mismos. Es decir: en cada unidad se obtiene información del rendimiento de ambos tratamientos. Así, se consigue eliminar, en la comparación de los tratamientos, el efecto que podría producir el hecho de comparar diferentes pacientes. Cada unidad proporciona información sobre la diferencia del efecto de ambos tratamientos y puede definirse una nueva variable D, que sea esta diferencia.

$$D_i = Y_{iA} - Y_{iB}$$

Comentario

La comparación se hace «dentro» de cada unidad: de observar diferencias entre las medias de ambos tratamientos, no podrían ser explicadas por diferencias entre los pacientes. En otras palabras, se suprime la posibilidad de que diferencias entre las unidades con las que se evalúa cada tratamiento, sean una explicación alternativa que impida poder atribuir a los tratamientos las diferencias observadas.

Es instructivo descomponer la varianza σ^2 de la variable que mide el efecto en dos componentes: las diferencias entre los distintos pacientes (**variabilidad entre unidades** o efecto de las unidades) y las divergencias entre diferentes determinaciones en un mismo paciente, representando la **variabilidad intra unidad**, que incluye también el error de medida.

En un diseño apareado la comparación se hace dentro de cada unidad, por lo que se elimina la varianza debida a las unidades, lo que hace al diseño más eficiente.

Comentario



Dado que puede esperarse de un proceso de medida razonable que el error que se comete al determinar el auténtico valor sea muy inferior a las diferencias entre las unidades, el diseño con datos apareados elimina la fuente de variación más importante (entre). Como ejemplo, supongamos la determinación de la altura de una persona con un procedimiento muy poco preciso: la cinta métrica «de costurera». Es poco preciso porque determinaciones repetidas pueden proporcionar valores diferentes. Aun así, las diferencias entre personas serán siempre mucho mayores que las variaciones entre los diferentes valores obtenidos en la misma persona por determinaciones repetidas.

El procedimiento estadístico que se utiliza en datos apareados es muy sencillo si se define la variable diferencia. De esta manera, la prueba de la hipótesis de que ambos tratamientos son iguales se reduce a la prueba de conformidad de una media de una muestra, de la que ya se vieron ejemplos en el capítulo anterior.

Ejemplo 8.6



En 6 pacientes se han estudiado dos tratamientos A, B. Se desea decidir si tienen medias iguales. Los resultados han sido:

		\bar{y}_j	S_j^2	S^2
Y_{IA}	23,05 39,06 21,72 24,47 28,56 27,58	27,407	39,421	42,023
Y_{IB}	20,91 37,21 19,29 19,95 25,32 24,07	24,458	44,625	

La solución **errónea**, mediante la prueba de datos independientes, sería:

$$\hat{t} = \frac{(\bar{Y}_1 - \bar{Y}_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(27,407 - 24,458)}{\sqrt{42,023 \left(\frac{1}{6} + \frac{1}{6} \right)}} \approx \frac{2,948}{3,743} \approx 0,788$$

siendo $42,023 \approx (39,421 + 44,625) / 2$

Como $0,788 < 2,228 = t_{10, 0,05}$ nada se opone a aceptar H_0 en esta solución errónea.

La solución **correcta** de la prueba de datos dependientes o apareados empieza calculando la diferencia $D_i = Y_{IA} - Y_{IB}$:

Ejemplo 8.6 (Cont.)



La solución **correcta** de la prueba de datos dependientes o apareados empieza calculando la diferencia $D_i = Y_{iA} - Y_{iB}$:

		\bar{Y}_j	S_j^2	S^2
Y_{iA}	23,05 39,06 21,72 24,47 28,56 27,58	27,407	39,421	42,023
Y_{iB}	20,91 37,21 19,29 19,95 25,32 24,07	24,458	44,625	
$D_i = Y_{iA} - Y_{iB}$	2,14 1,85 2,43 4,52 3,24 3,51	2,948	1	

$$\hat{t} = (\bar{D} - \mu_0) / (s_D / \sqrt{n}) = 2,948 / \sqrt{1/6} = 2,948 / 0,408 = 7,225$$

Como $7,225 > 2,571 = t_{5, 0,005}$ se rechaza H_0 .

A pesar de que hay la misma señal (la media de las diferencias es la diferencia de las medias), ahora el cociente señal/ruido toma un valor muy superior. La varianza de la variable diferencia (1) es muy inferior a la varianza anterior (42,023), que indica el beneficio de hacer un diseño con datos apareados: a pesar de que el numerador (*señal*) es el mismo, el denominador (*ruido*) es muy inferior en la segunda. Ello provoca que la primera «solución», la errónea, no pueda llegar a rechazar la hipótesis nula, pero que la segunda, a pesar de tener la misma señal, 2,948, encuentre pruebas en contra de la hipótesis nula y la pueda rechazar.

Recuerde



Controlar el efecto unidad, estudiando los tratamientos en los mismos casos conlleva dos ventajas metodológicas: a) evita que diferencias entre unidades puedan ser explicación alternativa de las diferencias observadas; y b) aumenta la eficiencia de la estimación al reducir el error aleatorio.

Nota técnica



La varianza de la estimación de la diferencia entre medias incluye un término que resta la covarianza entre ellas. Si se trataba de muestras independientes, esta covarianza se anulaba.

$$V(\bar{y}_1 - \bar{y}_2) = V(\bar{y}_1) + V(\bar{y}_2) - 2Cov(\bar{y}_1, \bar{y}_2)$$

Esta covarianza representa el efecto individuo, la variabilidad «entre casos». Cuanto más similares son las dos determinaciones y_1, y_2 en cada individuo, mayor es la ventaja por utilizar un diseño apareado.

Los ejemplos de datos apareados que se han presentado se referían todos al estudio de dos tratamientos en las mismas unidades. Pero pueden plantearse diseños con un menor grado de dependencia o apareamiento. Por ejemplo, y de más a menos: en hermanos gemelos; en hermanos; en dos pacientes de la misma edad, género y antecedentes; en dos pacientes de la misma edad, etc. Ello puede hacer difícil la decisión de si los datos están o no apareados.

Comentario



¿En qué circunstancias se puede creer que un diseño es de datos apareados? La primera pista es el tamaño de la muestra de las dos variables en estudio, ya que para que sean datos apareados se necesita que $n_1 = n_2$. La segunda pista es que exista un nexo que permita calcular las diferencias. Y en tercer lugar, la pregunta clave es si esta conexión entre los pares de observaciones está eliminando alguna variabilidad en el modelo, alguna diferencia natural entre las unidades estudiadas. Dicho de otra manera: ¿cabe esperar que las dos observaciones de la misma pareja sean más similares entre sí que otras dos observaciones seleccionadas al azar? Es decir, conocido el valor de la primera observación, ¿es posible acercarse, adivinar el valor de la segunda? Más formalmente: ¿vale 0 su correlación?

La figura 8-3 muestra la elevada **correlación** entre las variables Y_A e Y_B del ejemplo anterior. En efecto, a mayores valores de la variable Y_A , le corresponden mayores valores de la variable Y_B . El hecho de que todas las observaciones estén por debajo de la diagonal (bisectriz, dado el escalado) muestra gráficamente que los valores de Y_B son cada uno de ellos menores que los de Y_A .

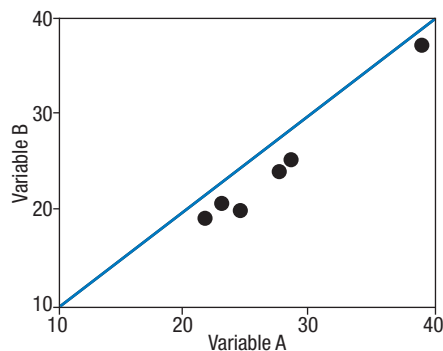


Figura 8-3 En datos apareados es posible estudiar la similitud entre ambas determinaciones.

Nota técnica



La **covarianza poblacional** entre las variables X e Y se define como:

$$\text{Cov}(X, Y) = E[X - E(X)][Y - E(Y)] = E(XY) - E(X)E(Y)$$

Nota técnica



Y la **covarianza muestral** se calcula como

$$\Sigma_{AB} = [\Sigma_{AB} - \Sigma_A \Sigma_B / n] / (n - 1) = [4229,543 - 164,440 \cdot 146,750 / 6] / 5 = 41,523$$

De donde la correlación vale:

$$r_{AB} = S_{AB} / S_A S_B = 41,523 / (6,279 \cdot 6,680) = 0,99$$

Esta **correlación** tan próxima a 1 está indicando que ambas variables van extremadamente juntas: hay un efecto paciente que origina que si el tratamiento A obtiene, por ejemplo, valores más elevados en un paciente, también se obtendrán valores más elevados en ese paciente con el tratamiento B. Puede demostrarse que la covarianza entre ambas variables coincide con la varianza entre unidades.

Ejemplo 8.7



Se desea averiguar si la frecuencia de relaciones sexuales modifica la probabilidad de reinfección por el virus del sida. Se dispone de una base de 120 pacientes de sida de transmisión sexual, de los que 6 se han reinfectado. A cada uno de ellos se le ha buscado en la base de datos un control lo más similar posible de acuerdo con los siguientes criterios: edad, género, hábitos sexuales y características clínicas. Finalmente, se ha escogido aquel control que tuviera un tiempo de seguimiento igual al tiempo de reinfección del caso.

Ejercicio 8.8



En la tabla siguiente figura la frecuencia de relaciones mensuales declaradas por cada caso y cada pareja del ejemplo anterior. Se pide realizar la prueba de datos apareados y la de datos independientes. Concluya en ambas y comente sobre el grado de apareamiento de los casos.

							\bar{y}_i	S_i^2	S^2
Y_{IA}	7,0	4,0	5,0	4,0	11,0	14,0	7,50	17,10	17,48
Y_{IB}	7,0	0,0	9,0	7,0	13,0	8,0	7,33	17,87	
$D_i = Y_{IA} - Y_{IB}$	0,0	4,0	-4,0	-3,0	-2,0	6,0	0,17	16,17	

Nota técnica



Podría darse la situación en la que unas observaciones proporcionen información sobre ambos tratamientos pero otras sólo tengan valores de uno (únicamente cierta porción de los datos está apareada). ¿Por qué se ha perdido esta información? ¿Son comparables los casos que no completan el proceso? Recuerde que es muy peligroso eliminar, sin más, estos casos.

Comentario



Consulte con un estadístico profesional si tiene datos parcialmente apareados.

Ejercicio 8.9



La estimación mediante datos apareados es más eficiente que la de datos independientes porque:

- a) aumenta la diferencia entre las medias;
- b) disminuye la diferencia entre las medias;
- c) aumenta el error típico de estimación;
- d) disminuye el error típico de estimación.

Ejercicio 8.10

La prueba definitiva para saber si unos datos son apareados es:

- a) que los tamaños muestrales sean iguales;
- b) que los datos vengan por parejas;
- c) que las parejas se hayan recogido muy juntas;
- d) que exista cierta similitud o correlación entre los valores de las parejas.

Comparación de dos varianzas con muestras independientes

Comentario



*Se ha visto la comparación de las medias de dos variables con distribución normal. De la misma forma podría verse la comparación de sus varianzas. Pero no lo haremos, especialmente porque creemos que la respuesta fundamental viene de las comparaciones de medias. Se puede argumentar que la igualdad de varianzas es una **premisa** necesaria para la comparación de medias. Y es cierto. Pero la posición mantenida en este texto es que las premisas son previas al establecimiento de la hipótesis y, por tanto, a la obtención de los datos. Si los datos presentan muchas novedades o sorpresas sobre las premisas, da la sensación de que el investigador se ha lanzado a un experimento sin saber muy bien de qué iba. Por otro lado, no se pretende afirmar que las premisas son una verdad absoluta sino razonable: el planteamiento estadístico presentado es **robusto** o sólido ante pequeñas desviaciones de las premisas.*

Soluciones a los ejercicios

8.1 Las respuestas idóneas deben ser SÍ, NO Y NO.

8.2 Cualquier ejemplo similar al del profesor. En los capítulos 11 y 12 se desarrollan con profundidad los aspectos relativos al diseño de experimentos y de observaciones.

8.3 Las respuestas correctas son la *a*) y la *b*), ya que la inferencia estadística hace referencia a los parámetros poblacionales o sus diferencias.

8.4 La respuesta correcta es la *c*). La amplitud del intervalo informa de la magnitud de la ignorancia sobre el efecto concreto, no de una variabilidad de este efecto a lo largo de los casos.

8.5 La respuesta correcta es la *a*).

8.6 La respuesta correcta es la *d*).

8.7 $S^2 = (155^2 \cdot 55 + 223^2 \cdot 74) / 129 \approx 38769,9$

$\hat{t} = (125 - 329) / \sqrt{[(38769,9/56) + (38769,9/75)]} \approx -204 / 34,77 \approx -5,87$

Se rechaza la hipótesis de igualdad de medias: el grupo SI tiene valores menores de CD4.

$IC_{95\%} = (125 - 329) \pm t_{129, 0,90} \cdot \sqrt{(S^2/50 + S^2/100)} = -204 \pm 1,657 \cdot 34,1 = -204 \pm 56,5$

$IC_{95\%} = (-260,5, -147,5)$

8.8 Independientes: $S^2 = (17,10 \cdot 5 + 17,87 \cdot 5) / 10 \approx 17,48$

$\hat{t} = (7,5 - 7,33) / \sqrt{[(17,48/6) + (17,48/6)]} \approx 0,17 / 2,41 \approx 0,07$

Apareados $\hat{t} = (0,17) / \sqrt{(16,17/6)} \approx 0,17 / 1,64 \approx 0,10$

Dado que el cociente señal/ruido está muy próximo a cero, no hace falta mirar en tablas.

Ninguna permite rechazar la nula de igualdad: no se han detectado diferencias.

El ruido en la prueba de datos apareados es sólo ligeramente inferior.

Parece que, para esta variable, aparear los datos no aporta grandes beneficios.

8.9 La respuesta correcta es la *d*).

8.10 La respuesta correcta es la *d*).

Capítulo 9:

Cálculo del tamaño muestral

Presentación

En este capítulo se introducen los argumentos estadísticos para poder determinar, de acuerdo con los objetivos de la investigación, el tamaño muestral necesario en un estudio comparativo. Mediante el método de comparación de medias, se ilustra como emplean los conceptos de potencia y efecto en estudio para determinar el tamaño muestral.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Sabrá calcular el tamaño muestral necesario para un diseño con muestras independientes.
- Interpretará el valor Δ como la diferencia entre los tratamientos para la que se desea cierta potencia determinada.
- Percibirá las consecuencias de cometer un riesgo α y un riesgo β .
- Percibirá la relación entre tamaño muestral y potencia.
- Percibirá la relación entre tamaño del efecto y potencia.
- Percibirá la relación entre el tamaño muestral y el tamaño del efecto.
- Sabrá calcular el tamaño muestral necesario para un diseño con muestras apareadas.
- Percibirá las ventajas en eficiencia de aparear los casos.
- Sabrá calcular el tamaño muestral necesario para un objetivo de equivalencia.
- Sabrá marcar un margen de seguridad al calcular el tamaño muestral en estudios de equivalencia.
- Distinguirá entre un tamaño muestral para establecer una hipótesis y un tamaño muestral para estimar, con cierta precisión deseada, un parámetro.
- Percibirá la determinación del tamaño muestral como un proceso iterativo y multidisciplinar.

Tamaño para comparar medias con datos independientes

Supóngase que, en la comparación de dos medias, se está interesado en tomar una decisión entre dos valores concretos, por ejemplo 0 y Δ , que situamos en las hipótesis nula y alternativa:

$$\begin{aligned} H_0 &: \mu_A - \mu_B = 0 \\ H_1 &: \mu_A - \mu_B = \Delta \end{aligned}$$

Ejemplo 9.1



Puede imaginarse, por ejemplo, que cierto tratamiento A tenga interés sanitario y comercial si, respecto a la versión clásica B, Δ representa aquella diferencia que hace rentable el desarrollo y la sustitución de B por A. Como es habitual en la prueba de diferencias, el valor de la hipótesis nula indica la absoluta igualdad entre ambos.

Supóngase que se conoce, en las unidades experimentales en las que se van a comparar, el grado de dispersión (σ) existente entre los resultados en varios pacientes sometidos al mismo tratamiento. Supóngase también que se ha decidido que los riesgos de adoptar decisiones erróneas sean exactamente α (bilateral) y β (unilateral).

Para determinar el número «n» de pacientes necesario en cada grupo, por simplicidad, se considera la situación (de máxima eficiencia) en la que se dispone exactamente del mismo número de casos en ambos grupos: $n_A = n_B = n$. Este número será:

$$n = [2 \sigma^2 (Z_{\alpha/2} + Z_{\beta})^2] / \Delta^2$$

Nota técnica



En esta situación, dado que la varianza de la diferencia de las medias en muestras independientes (asumiendo iguales las n y las s bajo cada tratamiento) es:

$$V(\bar{y}_1 - \bar{y}_2) = \sigma^2/n_1 + \sigma^2/n_2 = 2\sigma^2/n$$

Si la distribución de y es normal o el número de casos es razonable, la distribución de esta diferencia de medias será normal con esta varianza y centrada en 0, bajo H_0 , o en Δ , bajo H_1 .

En la figura 9-1 puede verse que la distancia entre los centros de ambas distribuciones es: $\Delta = Z_{\alpha/2} \sigma\sqrt{(2/n)} + Z_{\beta} \sigma\sqrt{(2/n)}$.

Si ahora se especifican los riesgos α (probabilidad de actuar A_1 siendo cierta H_0) y β (probabilidad de actuar A_0 siendo cierta H_1) que se está dispuestos a tolerar, ya se puede conocer el tamaño muestral n que se debe tener en cada muestra: $n = [2 \sigma^2 (Z_{\alpha/2} + Z_{\beta})^2] / \Delta^2$

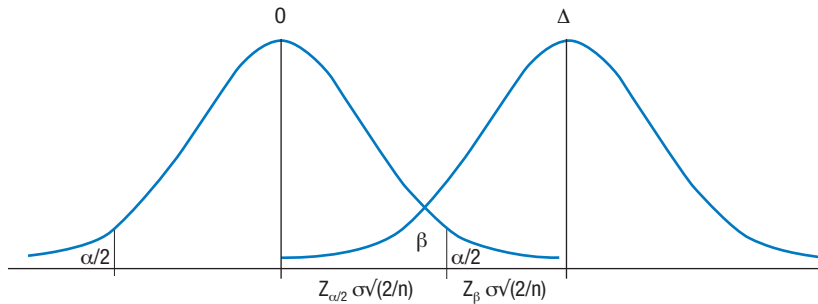


Figura 9-1 El tamaño muestral n requerido es aquel que permite que el solapamiento de las distribuciones del estadístico bajo H_0 y bajo H_1 proporcione los valores α y β especificados. Nótese que, si n aumenta, las distribuciones se hacen más «puntiagudas», disminuye el solapamiento y, por tanto, los riesgos α y β .

Recuerde



El cálculo del tamaño muestral depende de:

- los **riesgos α y β** que esté dispuesto a aceptar: cuanto menores, mayor tamaño.
- la **dispersión σ** del fenómeno estudiado: cuanto mayor, mayor tamaño.
- la **magnitud Δ del efecto** o diferencia que se desea demostrar: cuanto menor, mayor tamaño.

Comentario



Los autores creemos que, si se deben acotar los riesgos de actuaciones erróneas, el riesgo α debería ser unilateral; pero la posición de las Agencias del Medicamento, desde un punto de vista de inferencia de Fisher, pide que sea unilateral diciendo a los promotores: «si ustedes desean que, si el tratamiento no añade mejoras, yo acepté que es mejor en un $\alpha\%$ de ocasiones; en correspondencia, ustedes deben aceptar creer que es peor en otro $\alpha\%$ de ocasiones».

Ejercicio 9.1



¿Qué significa Δ ? ¿La diferencia «ideal» que se quiere demostrar? ¿La diferencia «real» que se cree que se puede demostrar?

Ejercicio 9.2



En un ensayo clínico de cuyo éxito depende la autorización comercial de un fármaco, ¿qué consecuencias se derivan de los riesgos α y β de cometer errores de 1.^a y 2.^a especie? ¿Qué implicaciones tienen para el usuario y el patrocinador?

Recuerde

La *potencia* de un estudio para establecer una alternativa de interés es el complementario del riesgo β .

Ejemplo 9.2

Un riesgo $\beta = 0,2$ implica una potencia de $0,8 = 80\%$.

Ejemplo 9.3

¿Qué tamaño muestral sería necesario para detectar una diferencia en la altura media de hombres y mujeres de 10 cm? Sea $\sigma = 8$ cm y los riesgos habituales.
 $n = [2 \cdot 8^2 (1,96 + 0,84)^2] / 10^2 \approx 10$
 Se necesitan 10 casos por grupo.

Nomograma para el cálculo del tamaño muestral

Se puede definir la diferencia tipificada o **efecto estandarizado** Δ_s como la razón entre la diferencia que se quiere detectar y la desviación típica (diferencia esperada entre dos observaciones): $\Delta_s = \Delta / \sigma$

Así, esta diferencia tipificada representa el efecto de manera relativa a la dispersión natural de los casos.

Ejemplo 9.4

Si se deseara aumentar la altura en 4 centímetros y la desviación típica se ha dicho que son 8 cm, el efecto tipificado sería del 50%.

Para un cálculo orientativo, se puede usar el gráfico de la figura 9-2 debido a Altman (46), donde N representa el tamaño total considerando ambos grupos ($N = 2 \cdot n$).

Ejercicio 9.3

¿Cuántos casos se necesitan si $\Delta = 5$ u, $\sigma = 8$ u, $\alpha = 0,05$ bilateral y $\beta = 0,20$?

Ejercicio 9.4

Calcule el tamaño necesario para un caso real propio.

Ejercicio 9.5



Sea $\sigma = 10$ u, $\alpha = 0,05$ bilateral y $\Delta = 5$ u. Ponga en una tabla los casos que necesita en total ($N = 2$ veces n por grupo) para potencias de 10, 20... 90, 95 y 99%. Dibuje un gráfico en el que se relacione la potencia con el tamaño. Interprete los resultados. Hágalo, de forma aproximada, utilizando el nomograma de Altman.

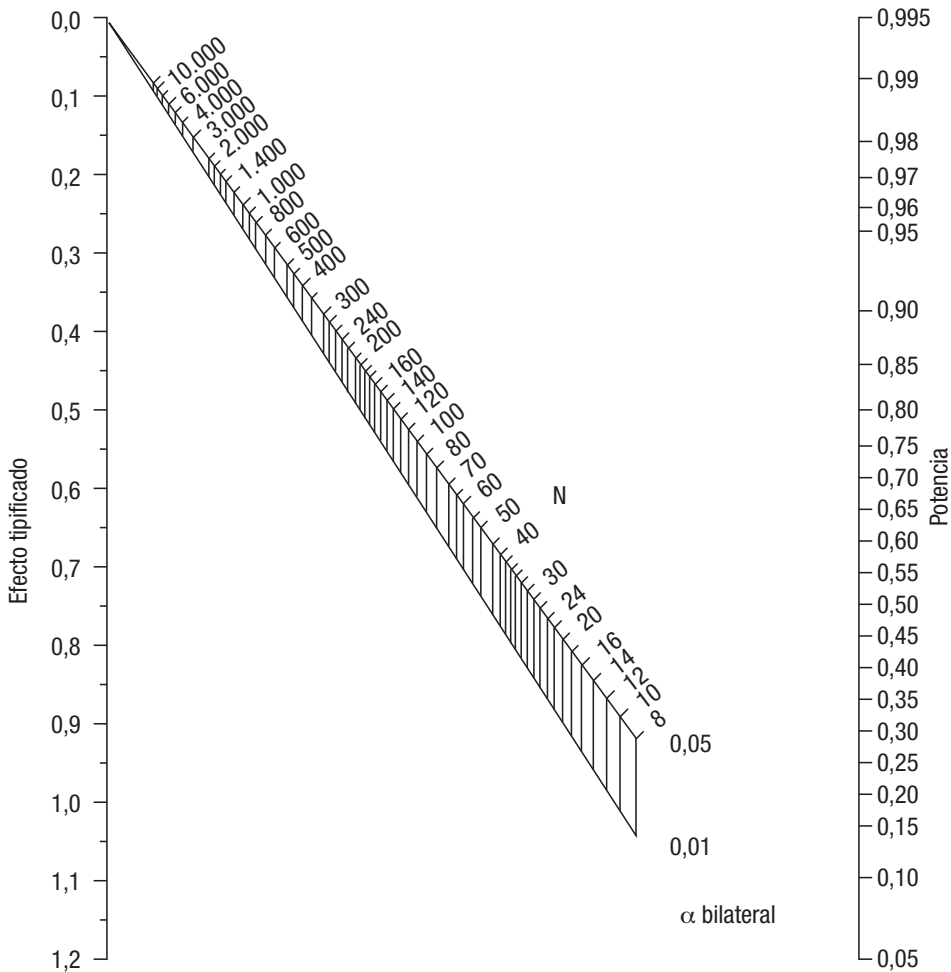


Figura 9-2 Nanograma para el cálculo del tamaño muestral. Debe trazarse una línea ante la columna de la izquierda (efecto tipificado Δ_s) y la de la derecha (Potencia $1 - \beta$). Al cruzar el riesgo α bilateral deseado (0,05 o 0,01), se obtiene el número total N de casos necesarios para comparar dos medias con datos independientes (Altman, 46)

Recuerde

Mayor *tamaño muestral* implica mayor *potencia*.

Ejercicio 9.6

Sea $\sigma = 10$ u, $\sigma = 0,05$ bilateral y $N=100$. Ponga en una tabla la potencia resultante para Δ desde 1 hasta 9 u. Dibuje un gráfico en el que se relacione la potencia con el efecto deseado.

Recuerde

Mayor *efecto Δ* en estudio implica *mayor potencia*.

Ejercicio 9.7

Sea $\sigma = 10$ u, $\alpha = 0,05$ bilateral y $\beta = 0,80$. Ponga en una tabla los casos que necesita para Δ desde 1 hasta 9u. Dibuje un gráfico en el que se relacione los casos necesarios con el efecto tipificado.

Recuerde

Mayor *efecto Δ* en estudio implica menor *tamaño muestral*.

Tamaño para comparar 2 medias con datos apareados

Los diseños con datos apareados consideran diferentes fuentes de variación. Recuerdese que, como cada unidad proporciona información sobre la diferencia del efecto de ambos tratamientos en comparación, se definía una nueva variable (D), diferencia entre la respuesta observada en ambas alternativas: $D_i = Y_{iA} - Y_{iB}$

¿Qué tiene que ver la varianza de esta nueva variable, σ_D^2 , con la varianza σ^2 que se utiliza en los datos independientes? Se vio que se podía descomponer σ^2 (la varianza total) en dos componentes: entre individuos σ_E^2 o verdaderas diferencias entre los casos e intraindividuos σ_I^2 o discordancias entre dos medidas en idénticas condiciones: $\sigma^2 = \sigma_E^2 + \sigma_I^2$

En datos apareados, se puede utilizar la **misma fórmula** teniendo en cuenta que la dispersión σ se refiere a la variabilidad **intrasujeto** σ_I^2 y la n resultante de la fórmula es la **n total**, ya que cada observación aporta los dos valores. La misma consideración puede hacerse para la tabla de Altman.

Recuerde

La varianza ahora es intrasujetos y la n resultante de cada grupo coincide con la N total.

Ejercicio 9.8

¿Cuántos casos totales se necesitan en un diseño paralelo y en uno apareado si $\Delta = 5$ u, $\sigma_E^2 = (9 \text{ u})^2$, $\sigma_I^2 = (4 \text{ u})^2$, $\alpha = 0,05$ bilateral y $\beta = 0,20$?

Ejercicio 9.9

¿Cuántos casos totales se necesitan en un diseño paralelo y en uno apareado si $\Delta = 5$ u, $\sigma_E^2 = 50 \text{ u}^2$, $\sigma_I^2 = 50 \text{ u}^2$, $\alpha = 0,05$ bilateral y $\beta = 0,20$?

Ejercicio 9.10

Invente valores de σ_E^2 y σ_I^2 para una variable que le sea familiar.

Tamaño para mostrar equivalencia

Se ha visto que el contraste de hipótesis para establecer equivalencia es:

$$\begin{cases} H_0: \mu_A - \mu_B = \Delta \\ H_1: \mu_A - \mu_B = 0 \end{cases}$$

La solución es idéntica a la anterior, por lo que el cálculo puede hacerse igual que antes, con la salvedad de que el planteamiento debe ser unilateral:

$$n = [2 \sigma^2 (Z_\alpha + Z_\beta)^2] / \Delta^2$$

Recuerde

En equivalencia, el planteamiento es *unilateral*.

Ejemplo 9.5

Se desea estudiar un cierto hipotensor, más seguro y barato. Se considera suficiente con demostrar que el clásico no le supera en 5 mmHg. ¿Cuántos casos se necesitan si $\sigma = 15$ mmHg, $\alpha = 0,025$ unilateral, $\beta = 0,8$?

$$H_0: \mu_A - \mu_B = 5$$

$$H_1: \mu_A - \mu_B = 0$$

$$n \approx [2 \cdot 15^2 (1,96 + 0,84)^2] / 5^2 \approx 141,28$$

→ 142 casos por grupo.

Ejercicio 9.11

¿Cuántos casos se necesitan si $\Delta = 5$ mmHg, $\sigma = 15$ mmHg, $\alpha = 0,05$ unilateral, $\beta = 0,8$?

Ejercicio 9.12

En los estudios de diferencias, Δ representaba cierta diferencia de interés, con relevancia clínica. ¿Qué significa Δ en los estudios de equivalencia?

Ejercicio 9.13

En equivalencia, ¿qué riesgo(s) α y/o β pueden perjudicar al usuario y al patrocinador?

Recuerde

Δ representa, en estudios de diferencias, al efecto *relevante*, mientras que en estudios de equivalencia es el *irrelevante*.

Ahora bien, se puede desear establecer equivalencia a pesar de que se crea que los dos tratamientos no son absolutamente idénticos: aunque tengan pequeñas diferencias entre ellos —que no alcancen relevancia—, podrían ser alternativas terapéuticas.

Lectura

Cuando en un ensayo clínico de inferioridad la potencia estadística se calcula para una diferencia cero, entonces el tamaño de la muestra necesario para alcanzar esta potencia va a ser insuficiente si el efecto del producto o la intervención evaluados es ligeramente inferior al del control activo (ICH-E9, 47).

Recuerde

*Es conveniente disponer de cierto **margen de seguridad**, por si los productos no fueran absolutamente idénticos.*

Para ello, en las fórmulas habituales se pone, en lugar del límite de no equivalencia Δ , la diferencia δ entre este límite de no equivalencia y el margen de seguridad MS que se desea cubrir:

$$\delta = \Delta - MS$$

donde δ : valor a ser utilizado en la fórmula

Δ : el límite de no equivalencia

MS: el margen de seguridad.

$$n = [2 \sigma^2 (Z_{\alpha} + Z_{\beta})^2] / \delta^2$$

Ejemplo 9.6



(Continuación del **ejemplo 9.5**) Ahora bien, se sospecha que este nuevo hipotensor más seguro y barato puede no ser absolutamente idéntico al clásico y se desea seguir teniendo la misma probabilidad de demostrar su eficacia incluso en el caso de que el clásico le superara en 1 mmHg. En resumen, ¿cuántos casos se necesitan si $\Delta = 5$ mmHg, MS = 1 mmHg, $\sigma = 15$ mmHg, $\alpha = 0,05$ unilateral, $\beta = 0,8$?
 $\delta = \Delta - MS = 5 - 1 = 4$
 $n \approx [2 \cdot 15^2 (1,645 + 0,84)^2] / 4^2 \approx 173,88 \rightarrow 174$ casos por grupo

Ejercicio 9.14



¿Cuántos casos se necesitan si aumentamos el margen hasta 2 mmHg [MS = 2 mmHg, $\Delta = 5$ mmHg, $\sigma = 15$ mmHg, $\alpha = 0,05$ unilateral, $\beta = 0,8$?

Tamaño muestral e intervalos de confianza

Es usual calcular el tamaño muestral de acuerdo con la metodología del contraste de hipótesis ya que permite considerar simultáneamente los riesgos α , β y la magnitud Δ que se desea establecer. En ocasiones, el objetivo del estudio puede ser más exploratorio y desear simplemente tener estimaciones de un parámetro con una cierta precisión, que se podría definir como la amplitud A del intervalo de confianza.

Comentario



Se vio que la amplitud del intervalo de confianza depende del error típico del estimador y del nivel de confianza. Si se desea limitar el grado de incertidumbre o amplitud de este intervalo sin disminuir el nivel de confianza $1 - \alpha$, la solución pasa por disminuir el error típico de estimación del parámetro. En el caso de la estimación de la media poblacional o esperanza matemática, el intervalo de confianza, asumiendo que conocemos σ , es: $IC_{1-\alpha}(\mu) = \bar{x} \pm Z_{\alpha/2} \sigma / \sqrt{n}$

Recuerde



La fórmula que hay que emplear es:

$$n = (2 \cdot Z_{\alpha/2} \sigma / A)^2$$

Ejemplo 9.7

Si, por ejemplo, se desea que la amplitud del intervalo de confianza valga A:

$$\begin{aligned} A &= \text{Límite superior IC} - \text{Límite inferior IC} = \\ &= \bar{x} + Z_{\alpha/2} \sigma / \sqrt{n} - (\bar{x} - Z_{\alpha/2} \sigma / \sqrt{n}) = \\ &= 2 \cdot Z_{\alpha/2} \sigma / \sqrt{n} \end{aligned}$$

De donde se obtiene la fórmula anterior que proporciona el tamaño n necesario para obtener la amplitud A deseada.

Ejemplo 9.8

Para estudiar la imagen de los diferentes políticos, se pide a los encuestados que los evalúen en una escala (continua) de 0 a 10 puntos. Si se acepta que la desviación típica de esta variable es de 1,5 puntos, ¿cuántos casos se necesitan para que la amplitud del intervalo de confianza al 95% de la media poblacional sea de 0,1 punto?

$$\begin{aligned} n &= (2 \cdot Z_{\alpha/2} \sigma / A)^2 = (2 \cdot 1,96 \cdot 1,5 / 0,1)^2 \approx 3457,44 \\ &\rightarrow 3458 \text{ casos} \end{aligned}$$

Ejercicio 9.15

Dado que 3458 casos son demasiados, se rebaja la ambición de conocimiento desde una amplitud de 0,1 puntos hasta 0,5 puntos.

En el caso de la **estimación de una probabilidad**, la fórmula es $n = (Z_{\alpha/2} / A)^2$

Nota técnica

El intervalo de confianza, en la situación de máxima incertidumbre, es:

$$\begin{aligned} \text{IC}_{95\%}(p) &= P \pm Z_{\alpha/2} \sqrt{[0,5 \cdot 0,5/n]} \\ \text{De donde: } A &= 2 \cdot Z_{\alpha/2} \sqrt{[0,5 \cdot 0,5/n]} \\ n &= (Z_{\alpha/2} / A)^2 \end{aligned}$$

Ejemplo 9.9

Para conocer el porcentaje de votos de un partido político, con una amplitud total del $\text{IC}_{95\%}$ igual a 1%, ¿cuántos casos se necesitan?

$$n = (Z_{\alpha/2} / A)^2 = (1,96 / 0,01)^2 \approx 38.416 \text{ casos}$$

Ejercicio 9.16

Decididamente 38.416 son demasiados casos, por lo que una vez más se rebaja la ambición de conocimiento desde una amplitud de 1 punto ($1\% = 0,01$) hasta 5 puntos ($5\% = 0,05$). ¿Cuántos casos se necesitan?

Comentario

En algunos textos se define la precisión de la estimación, con una confianza $1 - \alpha$, como la mitad de la amplitud A .

Consejos prácticos

Como pocas veces se tiene la oportunidad de desarrollar una investigación absolutamente original, el **primer paso** será siempre estudiar la literatura y mirar qué han hecho otros investigadores: ¿cuál era su objetivo, su tipo de diseño, su variable principal, el análisis estadístico, el tamaño final empleado, etc.?

El **segundo paso** es diseñar el borrador del propio estudio con la ayuda del nomograma de Altman y con valores aproximados de los parámetros necesarios (Δ , σ^2): ¿es razonable el número resultante? ¿Podrá disponer de este número de casos en un plazo y a un coste razonables? ¿Tendrán capacidad de convicción los resultados?

El **tercer paso** consiste en comparar su diseño con otros alternativos y repetir el segundo paso, hasta optar por un diseño concreto.

Una vez decidido el diseño, el **cuarto paso** consiste en realizar y justificar formalmente el cálculo del tamaño muestral, para lo que se requieren referencias para los valores de σ (y quizás de Δ) y utilizar tablas publicadas y programas validados para el cálculo definitivo.

Para este cálculo pueden usarse: Machin et al. (48), Badiella et al. (49), Dupont et al. (50) o Shuster (51). Además, los grandes paquetes comerciales de estadística también contienen ayudas al cálculo del tamaño muestral.

Comentario

En la pestaña «aplicaciones» de la página web del curso «Bioestadística para no estadísticos: principios para interpretar un estudio científico» encontrará hojas de cálculo que facilitan la determinación del tamaño muestral.

<http://www.fme.upc.edu/bioestadistica>

Lectura



En el desarrollo de una nueva alternativa terapéutica, además de sus efectos deseados, conviene estimar los no deseados, para conocer su seguridad. Aquí, el problema es muy diferente porque no se dispone de una variable respuesta única, sino de toda variable que represente un efecto secundario. Además, algunos de estos efectos, posiblemente los más severos, pueden presentarse con muy baja frecuencia, por lo que necesitarían tamaños muestrales desproporcionados para un razonable desarrollo del producto. Finalmente, el estudio de estos efectos se beneficia también del análisis globalizador de los diferentes ensayos disponibles.

La ICH-E1A (52) da recomendaciones sobre el número de casos necesarios para estudiar la seguridad de tratamientos crónicos. Pretende detectar aquellos fenómenos cuya incidencia supera el 1% a los 3 meses, pero no pretende caracterizar acontecimientos adversos por debajo del 1 por mil. Sugiere entre 300 y 600 casos seguidos y tratados durante 6 meses y 100 casos durante un año. También alerta sobre el necesario rigor científico de la comparación con los no tratados.

Soluciones a los ejercicios

9.1 Δ es el valor de la diferencia entre los tratamientos para el que se desea tener una probabilidad $1-\beta$ de demostrar que los tratamientos son diferentes. Conviene que coincida con la eficacia real y también con la ideal.

9.2 El riesgo α es la probabilidad de que un tratamiento no eficaz (H_0) se declare eficaz y se ponga en el mercado (A_1). El riesgo β es la probabilidad de que un tratamiento eficaz (H_1) se declare no eficaz y no se lleve al mercado (A_0). Ambos repercuten negativamente en el usuario, a quien representa la administración, y en el patrocinador. Al usuario porque puede estar pagando por un producto que no es eficaz (α) o porque no se puede beneficiar de uno que lo es (β). Al patrocinador, porque no comercializa un producto eficaz (β) o porque pierde energías en uno que no lo es (α).

9.3 $n = [2 \cdot 8^2 (1,96 + 0,84)^2] / 5^2 \approx 40,14$
 \rightarrow 41 casos por grupo (debe redondearse al alza).

El nomograma ofrece un resultado similar (doble, ya que su N es la total = $2n$)

9.4 Compruebe que coincide su cálculo con el nomograma.

9.5 El nomograma debe aproximarse a los resultados de la tabla 9-1:

Potencia	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%	99%
Casos total	8	22	34	48	62	80	100	126	170	208	294

Tabla 9-1 Casos N totales para potencias del 10 al 99% ($\alpha = 0,05$ bilateral, $\sigma = 10$ u y $\Delta = 5$ u)

Puede verse como, para un efecto Δ que representa el 50% de la desviación típica σ , se necesitan, para la potencia usual del 80%, algo más de 120 casos.

Nótese en la representación gráfica (figura 9-3) cómo va aumentando la potencia a medida que aumenta el número de casos. Obsérvese que el incremento de potencia es muy acusado desde la decena al centenar de casos por grupo, pero que, a partir del centenar de casos, el incremento en potencia es más reducido a pesar de aumentar el tamaño en centenares de casos.

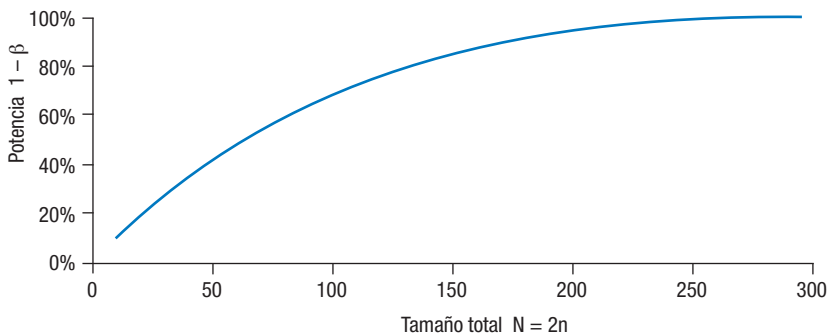


Figura 9-3 A nivel fijo de α , σ y Δ , cuando aumenta el número de casos n , aumenta la potencia.

9.6 El nomograma debe aproximarse a los resultados de la tabla 9-2:

Δ/σ	10%	20%	30%	40%	50%	60%	70%	80%	90%
Potencia	7%	17%	32%	52%	71%	85%	94%	98%	99%

Tabla 9-2 Para $N = 100$ ($\alpha = 0,05$ bilateral), potencia correspondiente para efectos tipificados del 10 al 90%

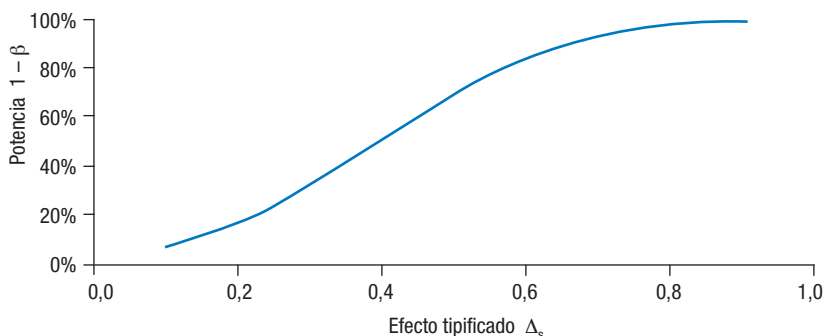


Figura 9-4 Para valores fijos de α , σ y n , cuando aumenta el efecto tipificado Δ_s , aumenta la potencia.

Puede verse que con 50 casos por grupo se tiene una potencia algo superior al 80% para un efecto que represente el 60% de la dispersión entre los casos estudiados. Nótese también que si el efecto se acerca al 100% de la dispersión entre los casos, con un diseño de 100 casos por grupo se tiene una potencia que se acerca al 100%, es decir, que, de ser cierto este efecto, la probabilidad de que el resultado del estudio sea significativo se acerca al 100%.

9.7 El nomograma debe aproximarse a los resultados de la tabla 9-3:

Δ/σ	10%	20%	30%	40%	50%	60%	70%	80%	90%
Casos	3140	786	350	198	126	88	66	50	40

Tabla 9-3 Casos N totales necesarios para efectos tipificados del 10 al 90% ($\alpha = 0,05$ bilateral y $\beta = 0,80$)

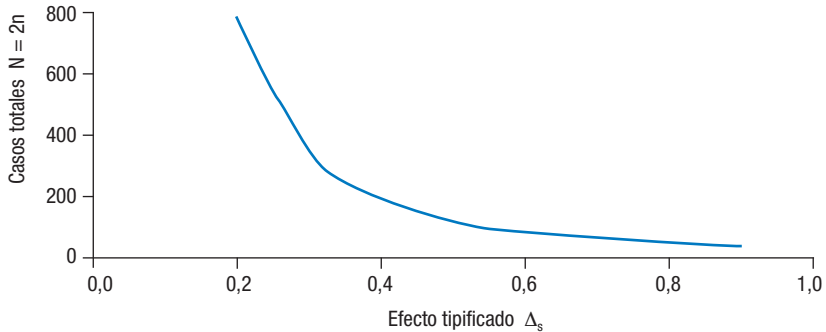


Figura 9-5 Para valores fijos de α , β y σ , cuando aumenta el efecto tipificado Δ_s , disminuye el número de casos.

Igual que antes, puede verse como, para un efecto Δ que representa el 50% de la desviación típica σ , se necesitan aproximadamente 120 casos por grupo. Nótese como va disminuyendo el número de casos necesarios a medida que aumenta la magnitud del efecto que se desea establecer. Obsérvese que el decremento de casos es progresivamente menos acusado.

9.8 $n = [2 \cdot (9^2 + 4^2) (1,96 + 0,84)^2] / 5^2 \approx 60,84 \rightarrow 61$ casos por grupo.

$n = [2 \cdot (4^2) (1,96 + 0,84)^2] / 5^2 \approx 10,04 \rightarrow 11$ casos apareados.

9.9 $n = [2 \cdot (50 + 50) (1,96 + 0,84)^2] / 5^2 \approx 62,72 \rightarrow 63$ casos por grupo.

$n = [2 \cdot (50) (1,96 + 0,84)^2] / 5^2 \approx 31,364 \rightarrow 32$ casos apareados.

9.10 Compruebe los valores con las referencias de la literatura.

9.11 $n \approx [2 \cdot 15^2 (1,645 + 0,84)^2] / 5^2 \approx 111,29 \rightarrow 112$ casos por grupo.

9.12 Mientras que en los estudios de diferencias, Δ representa la diferencia a partir de la cual se empieza a considerar relevantes a las diferencias entre los tratamientos, en los de equivalencia, deber representar un valor lo suficientemente pequeño como para que la diferencia entre los dos tratamientos sea irrelevante.

9.13 El riesgo α es la probabilidad de que dos tratamientos no equivalentes (H_0) se declaren equivalentes (A_1). El riesgo β es la probabilidad de que se declare que un tratamiento equivalente (H_1) no lo es (A_0). Igual que antes, los dos deben preocupar a ambos, pero las razones se invierten. El error tipo I del riesgo α implica sustituir un fármaco por otro cuando no son equivalentes, las consecuencias dependen de la di-

rección de la no equivalencia y del objetivo del estudio (eficacia o seguridad). Por su parte, el error tipo II del riesgo β implica no sustituirlo cuando en realidad sí que son equivalentes. Las consecuencias, como antes, dependen de la situación, aunque la habitual es de tipo económico, ya que no se autoriza un genérico más barato.

9.14 $n \approx [2 \cdot 15^2 (1,645 + 0,84)^2] / 3^2 \approx 309,13 \rightarrow 310$ casos por grupo.

9.15 $n \approx (2 \cdot Z_{\alpha/2} \sigma / A)^2 = (2 \cdot 1,96 \cdot 1,5 / 0,5)^2 \approx 138,2976 \rightarrow 139$ casos.

9.16 $n \approx (Z_{\alpha/2} / A)^2 = (1,96 / 0,05)^2 \approx 1.536,64 \rightarrow 1.537$ casos.

Capítulo 10:

Análisis de supervivencia

Presentación

El análisis de la supervivencia estudia una variable respuesta definida como el lapso de tiempo transcurrido entre dos sucesos. En general, cualquier tiempo de interés se denomina *tiempo de supervivencia* aunque no represente el tiempo hasta la muerte.

Dos aspectos del «tiempo entre dos eventos» caracterizan al análisis de supervivencia: la asimetría y la censura. La primera impide utilizar el modelo simétrico de la distribución normal. La censura proviene principalmente del hecho de que estos tiempos sólo se observan por completo cuando el suceso final ya se ha producido, mientras que en los restantes casos sólo se sabe que «por lo menos» superan un cierto valor.

En este capítulo se presenta su aplicación a las tres situaciones clásicas de la inferencia estadística correspondientes a los estudios univariante, bivariante y multivariante, para lo que se explican las funciones de supervivencia y de riesgo, y se explica cómo interpretar la comparación de curvas y el modelo de riesgos proporcionales de Cox.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Sabrá qué son datos censurados, cómo surgen y qué tipos de censura hay.
- Distinguirá entre función de supervivencia y función de riesgo.
- Sabrá interpretar la función de supervivencia y la función de riesgo.
- Entenderá que una función de riesgo constante permite estimar una tasa de riesgo común para todo el período de seguimiento.
- Sabrá interpretar curvas de la función de supervivencia.
- Pedirá estimaciones por intervalo de la supervivencia.
- Sabrá que las pruebas de Log-rank y de Gehan permiten comparar la supervivencia de dos grupos.
- Interpretará la razón de riesgo o *hazard ratio* como el cociente del riesgo (por unidad de tiempo) de presentar el evento entre dos grupos en comparación.
- Sabrá que la regresión de Cox permite estimar por intervalo la razón de riesgo ajustando por terceras variables.

Objetivos del análisis de supervivencia

El análisis de supervivencia (Gómez, 53; Gómez et al., 54; Gómez et al., 55) estudia el lapso de tiempo transcurrido entre dos eventos de interés, como podrían ser la aparición de un acontecimiento adverso tras una intervención terapéutica, o el tiempo transcurrido entre el inicio de una infección y su diagnóstico.

Comentario



El origen del nombre se debe a la construcción de tablas de vida en el siglo XVII.

Estudio univariante

Para describir y resumir los tiempos de vida, se emplean las funciones de supervivencia y de riesgo, que permiten predecir el comportamiento futuro de pacientes de características similares.

Ejemplo 10.1



Habiendo observado que 9 de 10 pacientes intervenidos de una neoplasia gástrica avanzada superaban los 6 meses de vida, ¿qué se sabe sobre lo que sucederá en futuros pacientes?

Estudio bivariante

Para comparar el patrón de supervivencia de dos poblaciones se emplean los métodos de Log-rank y de Gehan.

Ejemplo 10.2



En un ensayo clínico que compara un nuevo tratamiento en estudio con otro convencional, ¿es cierto que los miembros del primer grupo viven más tiempo y tienen mayor esperanza de vida o es simplemente un resultado casual observado en esta muestra, pero que no se repetirá en el futuro?

Estudio multivariante

El modelo de riesgos proporcionales de Cox selecciona aquellos factores de riesgo que más contribuyen a predecir el tiempo de vida teniendo en cuenta la influencia de los otros factores.

Ejemplo 10.3

A partir de la supervivencia de un grupo de pacientes con cáncer de pulmón y de variables como el género, la edad, los hábitos tabáquicos, etc., se desea estudiar cuál de estos factores ayuda a predecir la supervivencia.

O bien, para estimar el efecto de una intervención ajustando por las condiciones de los pacientes y del entorno en el que se aplica la intervención.

Ejemplo 10.4

En dos pacientes de características similares (edad, género, antecedentes tóxicos y fisiopatológicos, etc.) y en idénticas condiciones de administración (mismo centro, protocolo, tiempo previo de evolución, etc.), ¿cuánto mayor es la supervivencia en el paciente tratado con la nueva intervención?

El análisis basado en medias no es adecuado

Como el tiempo es una variable cuantitativa, una primera opción sería estudiar el tiempo de supervivencia mediante la media y la desviación típica. A continuación se exponen las dos razones que lo desaconsejan.

Asimetría de la variable tiempo

A diferencia de otras variables, el tiempo sólo se desplaza en una dirección: el colesterol puede subir o bajar, pero el tiempo nunca retrocede y siempre es positivo, con una distribución generalmente asimétrica.

Ejemplo 10.5

La figura 10-1 muestra la forma típica de esta variable para un ejemplo hipotético de meses de supervivencia tras el diagnóstico de una neoplasia de mal pronóstico: muchos casos tienen valores de unos pocos meses, pero algunos pacientes viven mucho, incluso uno de ellos puede llegar a superar los 8 años. Nótese los siguientes puntos:

- La media de esta variable, como resultado de estas observaciones extremas, se sitúa alrededor de 15 meses, valor que no representa en absoluto la supervivencia habitual de esta muestra. La mediana, en cambio, tiene un valor entre 10 y 11 meses, que resume mejor el tiempo de vida habitual o típico de estos pacientes.

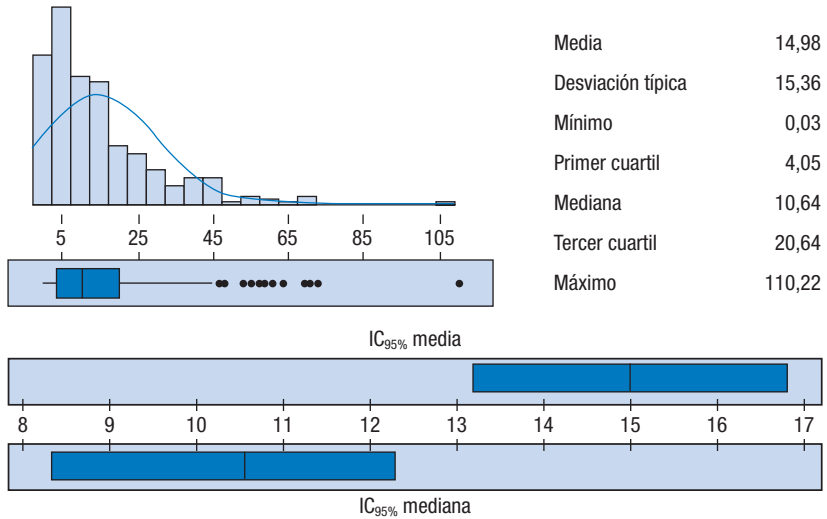


Figura 10-1 Histograma, con curva normal superpuesta, de los datos completos de supervivencia.

Ejemplo 10.5 (Cont.)



- Lo mismo sucede con la desviación típica, cuyo valor, algo mayor de 15 meses, es demasiado alto para representar la distancia típica de los casos que no alcanzan la supervivencia media y es demasiado bajo para representar esta distancia con el centro para los casos que sí superan el tiempo de vida más habitual.
- La superposición de una distribución normal (que tenga esta media y esta desviación típica) muestra que el modelo de Gauss-Laplace no puede ser utilizado para representar estos casos: ¡el 95% de los pacientes vivirían entre -15 y +45 meses!

Censura

La variable **tiempo hasta un cierto evento** requiere un largo plazo para ser observada en su totalidad.

Ejemplo 10.5 (Cont.)



En el ejemplo anterior, a los 24 meses de seguimiento ya se conocía el tiempo de vida exacto del 80% de los pacientes, pero se debería esperar aún más de 6 años para conocer el 20% restante, lo que restaría actualidad y valor científico a los resultados.

Los estudios científicos deben y suelen marcar un tiempo de seguimiento máximo.

Definición



Tiempo de seguimiento es el lapso de tiempo transcurrido entre la inclusión de un caso y el cierre del estudio.

Ejemplo 10.6



Un diseño habitual consistiría en incorporar pacientes durante 12 meses y parar el estudio tras otros 12 meses de seguimiento. De este modo, los pacientes incluidos el primer día del estudio habrán sido seguidos hasta un máximo de 24 meses. Y los incluidos al final del primer año hasta 12 meses. La figura 10-2 muestra los tiempos que se habrían observado si los pacientes de la figura 10-1 hubieran sido seguidos entre 12 y 24 meses.

Únicamente en los casos que hayan tenido una supervivencia menor a su tiempo de seguimiento se conocerá su tiempo exacto de vida. Para el resto, se sabrá que han vivido más que el tiempo de seguimiento observado, por lo que diremos que sus valores están censurados (por la derecha).

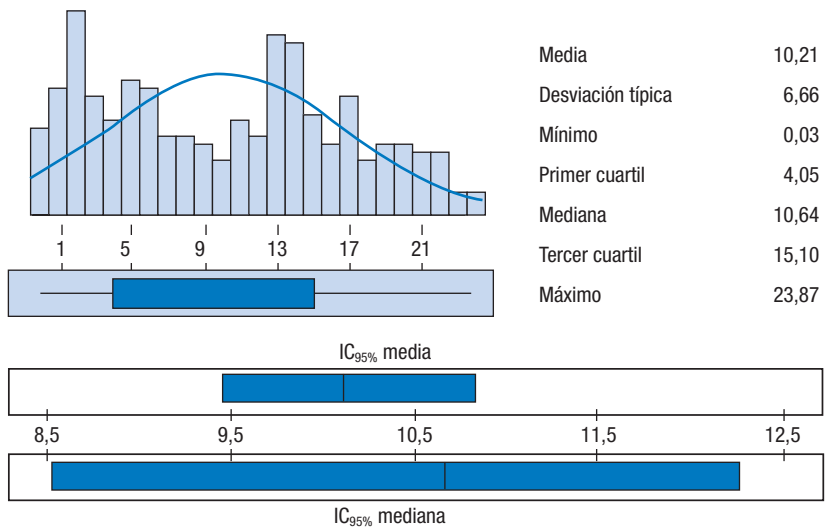


Figura 10-2 Histograma, con curva normal superpuesta, de los datos censurados de supervivencia.

Al haberse eliminado los mayores tiempos, tanto la media como la desviación típica ofrecen resultados (10 y 7 en cifras redondas) muy diferentes de los reales de la figura 10-1 (15 y 15). Ya se dijo que estos indicadores no resumían bien las variables asimétricas, y ahora se muestra que, en el caso de datos censurados, proporcionan resultados sesgados, que no informan ni de la auténtica media ni de la auténtica desviación típica.

En cambio, la mediana o percentil 50 puede ser estimado en este ejemplo, ya que hay un 69% de datos no censurados. Nótese que el valor estimado de la mediana (10,64) coincide con el anterior.

Recuerde



Los percentiles pueden estimarse siempre que el porcentaje de datos no censurados supere el percentil que se desea estimar.

Un problema adicional es que la censura sea el resultado de una pérdida (*lost to follow-up*) previa al final del período de seguimiento, especialmente si se desconoce la razón.

Ejemplo 10.7



Si, por ejemplo, al ver próximo su final los pacientes deciden *desaparecer* —quizá volver al pueblo o cambiar de centro asistencial—, se desconocería su defunción y se obtendrían cifras optimistas sobre la supervivencia.

La mayoría de los métodos usados en análisis de supervivencia presuponen la condición de no informatividad: los individuos censurados están sujetos a la misma probabilidad de muerte que aquellos que permanecen en el estudio.

Recuerde



Es preciso asumir que los casos censurados son idénticos a los restantes.

Consecuencias de la asimetría y de la censura

El hecho de que la variable de interés sea el tiempo, que se mide secuencialmente, tiene como consecuencia una distribución asimétrica y la presencia de censura. Estas circunstancias desaconsejan el uso de la distribución normal, que tan bien caracterizan sus parámetros media y desviación típica.

Si se renuncia a basar la comparación en un parámetro, como la media, se puede recurrir a los procedimientos no paramétricos, que no requieren tampoco una distribución concreta y permiten, además, considerar secuencialmente los datos, de for-

ma que cada individuo sólo contribuye al estudio mientras está bajo observación; es decir, mientras no aparece la censura. Por ello, son la alternativa más usada en estudios de supervivencia.

Funciones específicas

En este punto se explican las funciones que permiten especificar las preguntas de interés clínico: ¿cuál es la probabilidad de que un caso sobreviva cierto tiempo? Y ¿cuánto vale el riesgo en un instante determinado?

Función de supervivencia

Definición



La **función de supervivencia** proporciona la probabilidad de que un paciente sobreviva un determinado tiempo.

Ejemplo 10.8



La figura 10-3 representa cómo van variando, a lo largo del tiempo, la proporción de unidades que están libres del evento de interés. Puede verse, por ejemplo, que en el mes 10, el 60% de los pacientes del estudio aún no lo habían presentado. Y que el intervalo de confianza de lo que sucede en la población origen de la muestra es muy estrecho, lo que indica que se dispone de buena información (de hecho, este cálculo se basa en 1.000 pacientes).

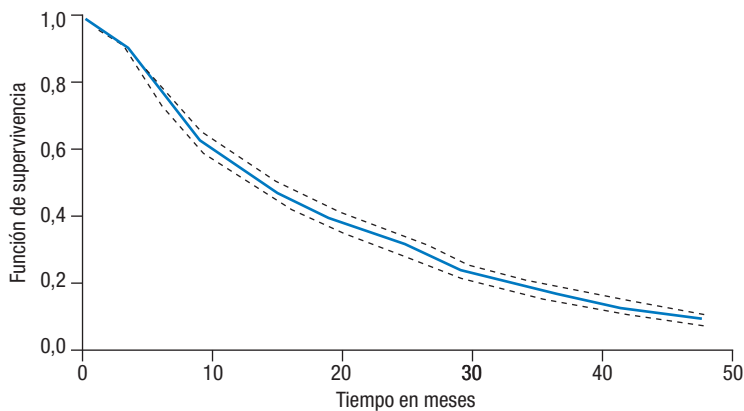


Figura 10-3 Función de supervivencia con su correspondiente intervalo de confianza al 95%.

Aun en el caso de un riesgo constante, la función de supervivencia mostrará un descenso más marcado al inicio por el simple hecho de que hay más casos expuestos al riesgo.

Ejemplo 10.9

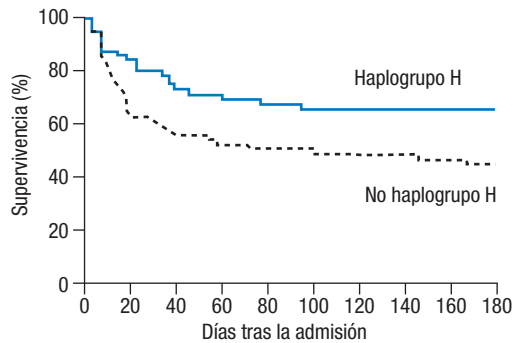


Si el evento de interés es la muerte, por ejemplo, aquellos casos ya fallecidos ¡no son susceptibles de volver a morir!

Ejercicio 10.1



Según la figura 10-4 (de Baudouin et al., 56):
 a) ¿Qué proporción de pacientes seguían vivos a los 100 días de seguimiento para cada grupo de tratamiento?
 b) ¿Qué indican los números 65, 55... al pie de la figura? ¿Por qué van haciéndose menores?



N.º en riesgo	
Haplogrupo H	65 55 48 45 44 42 42 42 42 41
No haplogrupo H	83 52 45 43 41 41 39 39 38 36

Figura 10-4 Función de supervivencia según haplogrupo.

Función de riesgo (hazard rate)

Para hacer comparables los descensos del inicio (que se obtienen de una población más grande) con los descensos finales (provenientes de una población restante más pequeña) se calcula la función de riesgo.

Definición



La **función de riesgo** o fuerza de mortalidad o tasa condicional de fallo (*hazard rate*) es la proporción de casos que presentan el evento en un momento determinado sobre el número de casos que llegan a ese momento.

Nota técnica

Otra forma de interpretar la función de riesgo en el instante t es entenderla como la distribución (densidad) residual del tiempo de vida después de t .

Definición

Una **función de riesgo** constante permite proporcionar una **tasa de riesgo** común para todo el período de seguimiento.

Ejemplo 10.10

En la figura 10-5 puede verse la función de riesgo correspondiente a los datos presentados en la figura 10-3. Nótese que es aproximadamente constante a lo largo del período de seguimiento, con una tasa de riesgo cercana al 0,05 mensual.

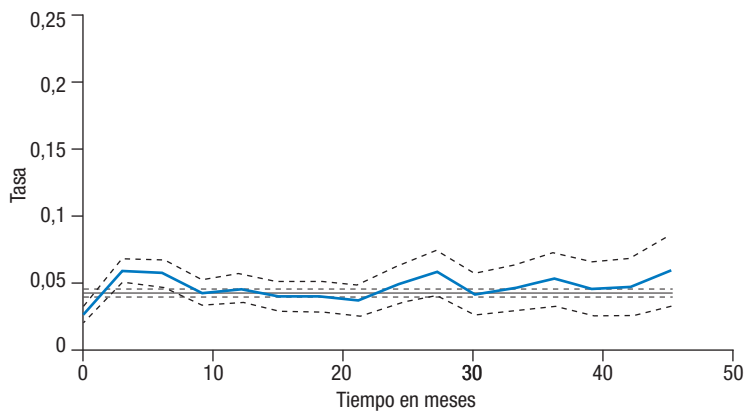


Figura 10-5 Función de riesgo (línea quebrada, azul y más gruesa) y tasa de riesgo (línea recta, negra y más fina) con sus correspondientes intervalos de confianza al 95%.

Ejemplo 10.11

Un ejemplo de tasa de riesgo constante sería la de las hemorragias digestivas tras antiinflamatorios, de los que se afirma que hay el mismo riesgo sea la primera o la enésima vez que se toman.

La forma de la función de riesgo dependerá del fenómeno estudiado.

Ejemplo 10.12



En el análisis del tiempo de vida de pacientes con leucemia que no responden al tratamiento (fig.10-6 A), la función de riesgo es creciente.

Recuerde



Un *riesgo creciente* corresponde a una población que envejece.

Ejemplo 10.13



Por ejemplo, después de una operación quirúrgica (fig. 10-6 B), la función de riesgo es decreciente.

Recuerde



Un *riesgo decreciente* corresponde a poblaciones en las que los individuos se fortalecen con el paso del tiempo.

Ejemplo 10.14



El tiempo de vida de un ser vivo o de un producto industrial suele tener una función de riesgo con forma de «bañera» (fig. 10-6 C).

Recuerde



El *riesgo* tiene forma de *bañera* en muchos seres vivos.

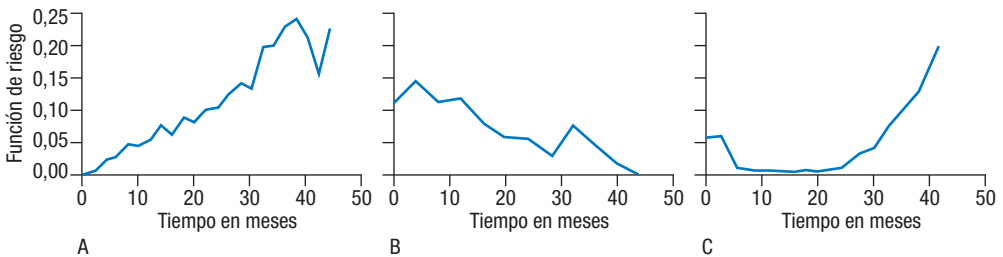


Figura 10-6 Funciones de riesgo creciente, decreciente y en «bañera».

Inferencia

En este punto se estudian, sucesivamente, la estimación, la comparación y la comparación ajustada.

Estimación instantánea (Kaplan-Meier) y por intervalo (actuarial)

De la misma manera que se pueden calcular o bien la velocidad instantánea o bien la velocidad «media» durante un período de tiempo, la función de riesgo también puede corresponder a un instante o a un intervalo.

Definición



El método actuarial (*life tables*) calcula, en un intervalo, la proporción de casos que lo superan respecto al total de casos que lo inician.

Ejemplo 10.15



Ha sido el método empleado en la figura 10-5.

Comentario



Una dificultad de este método la presentan los casos perdidos durante cada intervalo (están vivos al inicio pero no se sabe si lo están al final). La solución habitual considera que los individuos perdidos o que abandonan han sido seguidos hasta la mitad del intervalo y, por tanto, estuvieron a riesgo durante la mitad del mismo.

Definición



El **método instantáneo** de Kaplan-Meier «actualiza» la estimación de la función de supervivencia en cada momento en que aparece un evento.

Nota técnica



En estas definiciones, el método actuarial estima la función de riesgo; mientras que el instantáneo, la de supervivencia.

Nota técnica



Aun en el caso de que la tasa de fallo fuera constante, las fluctuaciones del muestreo originarán que las estimaciones difieran de un intervalo a otro.

Recuerde

Para visualizar correctamente la información contenida en la muestra conviene representar la banda de confianza que informa sobre los auténticos valores poblacionales que son compatibles con los resultados muestrales observados.

Comentario

En la figura 10-5 el intervalo de confianza de la función de riesgo se va ampliando con el tiempo, mostrando la mayor incertidumbre originada por el menor número de casos que alcanzan las mayores cifras del seguimiento.

Comentario

Si se decide asumir que la tasa es constante a lo largo de todo el período de seguimiento, se podrán combinar todas las tasas mensuales en una sola, con el doble beneficio de parsimonia científica y eficiencia estadística: parsimonia porque se explica el mismo fenómeno de forma más simple; y eficiencia porque se obtienen estimadores más precisos, con un intervalo de confianza más estrecho.

Ejercicio 10.2

¿Cuáles son ciertas?

- a) El intervalo de confianza no aporta información en el análisis de supervivencia.
- b) Si la función de riesgo es constante, se puede emplear una tasa de riesgo común para todo el período.
- c) Si la función de riesgo es constante, la función de supervivencia también decrece de forma constante.
- d) El método de Kaplan-Meier actualiza la función de supervivencia ante cada nuevo evento, lo que le confiere una forma escalonada típica.
- e) El método actuarial calcula la función de riesgo en unos intervalos predefinidos.
- f) Aun siendo constante la función de riesgo poblacional, la estimación del método actuarial tendría las fluctuaciones derivadas del proceso de muestreo.

Comparación de curvas**Comentario**

Para comparar dos curvas de supervivencia, una primera aproximación estudiaría un único tiempo, por ejemplo, los 3 años. Este procedimiento desprecia el resto de información, por lo que necesita muestras grandes para encontrar diferencias. Además, la elección de este punto en el tiempo puede ser controvertida.

Es conveniente utilizar la información de todos los tiempos para decidir si las curvas en comparación provienen de la misma población.

Definición

La **prueba del *log-rank*** compara las dos curvas otorgando la misma ponderación a todos los tiempos de seguimiento.

Nota técnica

Es equivalente a la prueba de Mantel y Haenszel para combinar tablas 2×2 .

Definición

La **prueba de Gehan** otorga una mayor ponderación a los tiempos iniciales, que tienen más observaciones.

Nota técnica

Es la generalización de la prueba de Wilcoxon y de Mann-Withney.


Ejercicio 10.3


Observe la figura 10-4. El valor de P proporcionado se basa en la prueba de *log-rank*, que da la misma ponderación a todos los tiempos. Si se hubiera realizado la prueba de Gehan, ¿se habría obtenido una significación mayor o menor?


Estos dos procedimientos proporcionan un nivel de significación (valor de P) que permite detectar y establecer la supremacía de una población respecto a la otra. Sin embargo, la ausencia de un parámetro que resuma las diferencias entre ambas poblaciones dificulta la evaluación de la relevancia clínica, que se basará en la comparación de la supervivencia en diferentes tiempos que, como se ha dicho, carece de objetividad y eficiencia estadística.

Comparación ajustada: modelo de riesgos proporcionales (regresión de Cox)

En este apartado se expone cómo el modelo de Cox proporciona un parámetro que permite resumir las diferencias entre las poblaciones en comparación.


<p>Definición</p> 	<p>El modelo semiparamétrico de Cox no exige ninguna forma para la función de riesgo (puede ser creciente, decreciente, etc.), pero en cambio proporciona un parámetro que es la razón entre las funciones de riesgo (hazard rate ratio, HRR) de los grupos en comparación.</p>
------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p>Nota técnica</p> 	<p>Es el equivalente en análisis de la supervivencia al modelo de regresión lineal.</p>
--------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------

<p>Comentario</p>	
	<p><i>Este planteamiento responde de forma natural a la pregunta del clínico, que no ambiciona conocer cuál es el tiempo exacto de vida de un paciente determinado (parte del modelo no parametrizada), pero centra sus esfuerzos en identificar aquellas intervenciones que puedan aumentarlo (parte del modelo parametrizada).</i></p>

El modelo de Cox no impone como premisa una forma para la función de riesgo, pero sí que asume que el cociente de su valor entre ambos grupos es constante, es decir, que la razón de riesgo toma el mismo valor a lo largo de todo el período de seguimiento.

Ejemplo 10.16

	<p>Como ejemplo, se puede pensar que si se desarrolla un nuevo tipo de antiinflamatorio, es deseable que la reducción del riesgo de hemorragia digestiva se mantenga mientras se toma la medicación. En la figura 10-7 puede verse la forma que adopta la función de supervivencia y la función de riesgo cuando este efecto es constante a lo largo del tiempo. Es decir, cuando el beneficio del tratamiento (reducción en el riesgo) es el mismo en todo momento.</p>
-------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Las funciones de supervivencia de ambos grupos se van separando progresivamente como resultado del mayor riesgo en uno de ellos. Por su parte, las funciones de riesgo son paralelas, con una cierta oscilación por las fluctuaciones del muestreo. La razón de riesgo de este ejemplo es 0,45 (IC_{95%}: 0,52 a 0,38), lo que indica que uno de los grupos tiene un riesgo que es algo menos de la mitad que el correspondiente al otro grupo (la reducción es del 55%, con un IC_{95%} desde el 48 al 62%).

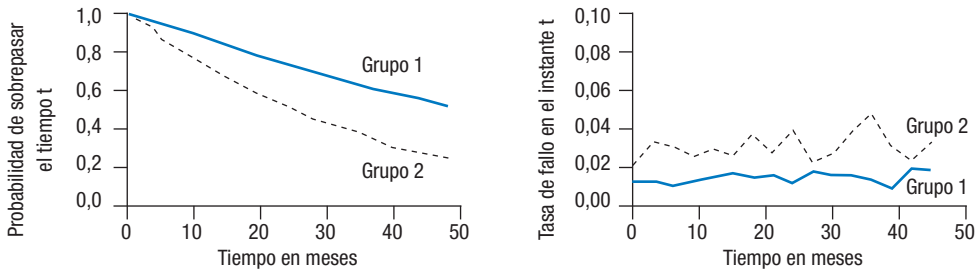


Figura 10-7 Forma de las funciones de supervivencia y de riesgo cuando la razón de riesgo es constante.

Nota técnica



El intervalo de confianza es razonablemente preciso. Una vez más, adoptar la **premisa de riesgos proporcionales** hará el modelo más parsimonioso y simple de comunicar, al mismo tiempo que aumenta su eficiencia estadística.

El médico está en condiciones de informar a su paciente: «Mire, yo no puedo decirle cuántos cumpleaños le quedan, pero puedo garantizarle que, si realiza la intervención que le propongo, su riesgo (instantáneo) queda reducido entre un 48 y un 62%».

Ejercicio 10.4



Véase en la siguiente tabla (Valery et al., 57), el HRR de la comparación de diferentes tratamientos

a) La tabla proporciona un HRR = 0,76 (para el análisis de Intención de tratar) con un IC_{95%} de 0,7-0,9. Interprete estos resultados.

b) Si se divide el tanto por ciento de la primera columna (riesgo de presentar el evento: 40,2%) por el de la segunda (50,9%) se obtiene 0,79, que es algo diferente de la razón entre las funciones de riesgo mostrada en la tabla (0,76). ¿A qué se debe?

		n autóctona (%)	n no autóctona (%)	HRR (IC _{95%})
Quimioterapia	No	515 (64,4)	435 (54,4)	1,0
	Sí	285 (35,6)	365 (45,6)	0,80 (0,7-0,9)
Radioterapia	No	437 (54,8)	390 (49,1)	1,0
	Sí	361 (45,2)	405 (50,9)	0,91 (0,8-1,0)
Cirugía (cualquiera)	No	427 (52,4)	341 (42,1)	1,0
	Sí	388 (47,6)	469 (57,9)	0,76 (0,7-0,9)
«Intención de tratar» con cirugía	No	457 (59,8)	362 (49,1)	1,0
	Sí	307 (40,2)	375 (50,9)	0,76 (0,7-0,9)

Tabla 10-1 Incidencia y riesgo de los criterios principal y secundario de valoración en la población por intención de tratar

La **premisa de proporcionalidad** de los riesgos no tiene por qué ser siempre correcta, y es obligación del investigador y del estadístico analizar hasta qué punto los datos permiten seguir empleándola.

Ejemplo 10.17

La figura 10-8 muestra una situación en la que el evento tiene una tendencia diferente a presentarse en ambos grupos: mientras que en uno de ellos se presenta con más frecuencia en la primera mitad del seguimiento; en el otro grupo, el evento acontece más a menudo en la segunda mitad.

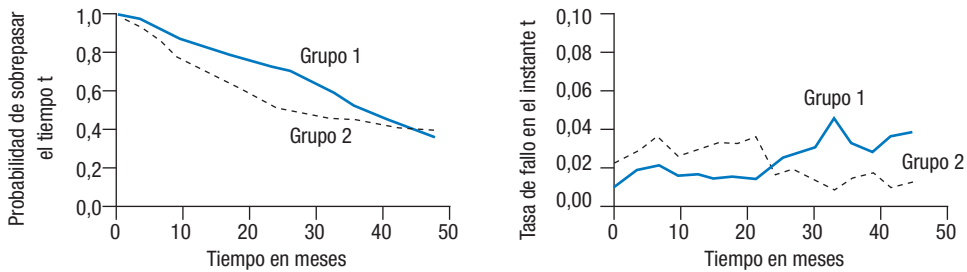


Figura 10-8 Forma de las funciones de supervivencia y de riesgo cuando la razón de riesgo cambia de una parte a otra del seguimiento.

Esta situación, sin duda más rebuscada, podría darse en un tratamiento con más de un mecanismo de acción, de forma que en una parte del seguimiento se observaran unos resultados contrarios a los de la otra parte.

Por supuesto, en esta situación no sería correcto asumir un efecto constante del tratamiento, entre otras cosas porque ambos efectos se anularían mutuamente y no se observaría nada (razón de riesgo = 1,095, $IC_{95\%}$: 0,940 a 1,227).

Nota técnica

La expresión «en cien años, todos calvos» nos recuerda que a largo plazo las curvas de supervivencia siempre se encontrarán en el mismo punto: ¡0%! Como en todos los modelos estadísticos, la premisa de proporcionalidad sólo aplica al período considerado.

El modelo de Cox puede considerar simultáneamente la hipotética relación de varias variables con la supervivencia y estudiar de esta manera si la relación de algunas persiste una vez se ha tenido en cuenta al resto. Para hacerlo, extiende la premisa de riesgos proporcionales a cada variable considerada, tanto cuantitativas como categóricas.

Comentario

La construcción del modelo es una tarea muy delicada, ya que se debe especificar aquella relación que cumple con esta premisa de proporcionalidad al mismo tiempo que se evita introducir variables que tengan entre ellas una elevada colinealidad.

Si el modelo está bien elaborado, su interpretación es más simple, ya que la razón de riesgo establecida para una variable puede interpretarse «independientemente», es decir, a nivel fijo, de las restantes variables.

Nota técnica

Recuerde que la decisión final sobre si una relación puede interpretarse como causal debe basarse en la existencia de la correspondiente hipótesis antes de la recogida de los datos; de la procedencia, observacional o experimental, de los mismos, y de otras pruebas empíricas que sustenten dicha relación.

En resumen, el análisis de supervivencia debe basarse en el estudio de las funciones de supervivencia y de riesgo. El modelo de riesgos proporcionales de Cox, sin imponer condiciones sobre la forma de la supervivencia, permite al investigador disponer de un estimador del efecto de un tratamiento, junto con su intervalo de confianza, a partir del cual puede evaluar tanto la significación estadística como la relevancia clínica.

Nota técnica

El modelo de Cox es una metodología sofisticada, con premisas que deben ser evaluadas por un estadístico.

Soluciones a los ejercicios

10.1 *a)* Alrededor de un 65% para el haplogrupo H y sobre un 50% para el no haplogrupo H.

b) Los números al pie de la figura indican la cantidad de casos que están informando sobre la supervivencia en cada momento del tiempo de seguimiento. Van haciéndose menores porque excluyen tanto a los casos que han abandonado el estudio antes de dicha fecha como a los que ya han fallecido.

10.2 Sólo son falsas la *a)*, ya que hay que hacer IC si se quieren usar los valores muestrales observados para estimar los poblacionales; y la *c)*, pues si la proporción de casos que presentan el evento, por intervalo de tiempo, es constante, la frecuencia de casos que lo presente será mayor al inicio, ya que hay más casos en observación.

10.3 Menor, ya que la prueba de Gehan otorga menor ponderación a los tiempos finales de seguimiento, que son los que, en este ejemplo, muestran una mayor diferencia entre los grupos.

10.4 *a)* El tratamiento reduce el riesgo de presentar alguno de los eventos que definen la variable principal en un 24% ($IC_{95\%}$ entre un 10 y un 30%), estimación puntual que indica el valor más verosímil del descenso del riesgo, mientras que los límites del IC indican el grado de incertidumbre debido al proceso de muestreo y de inferencia.

b) Varias razones pueden explicar esta pequeña discrepancia. La primera es que el cálculo realizado proporciona el riesgo relativo sin ajustar, es decir, sin «equilibrar» por el resto de variables que sí que ha tenido en cuenta el HRR estimado por el modelo de Cox. La segunda es que 0,79 se ha obtenido sin tener en cuenta el tiempo de seguimiento de los pacientes que el modelo de Cox sí que considera. Al tratarse de un EC aleatorizado, no cabe esperar grandes diferencias entre los grupos en estas variables, por lo que los resultados sin ajustar (0,79) y ajustados (0,76) no deberían diferir mucho.

Capítulo 11:

Objetivos de un estudio clínico

Introducción

Este capítulo aporta claves para identificar las preguntas que un determinado estudio permite contestar. Al terminar este capítulo, aquel lector que haya realizado los ejercicios podrá especificar los objetivos de un estudio concreto así como interpretar la utilidad de los originales de investigación más relevantes.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Distinguirá entre predicción e intervención.
- Percibirá las ventajas de anticipar el valor de una variable.
- Distinguirá entre una pregunta sobre causas y una pregunta sobre efectos.
- Clasificará como experimental un estudio si el investigador asigna los valores de la causa.
- Exigirá una hipótesis y un protocolo para poder considerar confirmatorio un estudio.
- Valorará las posibilidades de un análisis exploratorio.
- Sabrá que la confusión es una dificultad para la inferencia causal.
- Sabrá que para que exista confusión es necesario que la tercera variable prediga la respuesta.
- Sabrá que para que exista confusión es necesaria colinealidad.
- Identificará interacción con modificación del efecto.
- Pedirá a un diseño ausencia de sesgo.
- Pedirá a un diseño eficiencia.
- Sabrá que la interacción puede aparecer sin colinealidad.
- Sabrá que la interacción puede depender de la escala.
- Identificará el efecto causal con la diferencia en la respuesta entre dos acciones posibles.
- Identificará el efecto causal con la pregunta prospectiva, sobre efectos.
- Percibirá que el efecto causal requiere asignación de la causa.
- Percibirá que el efecto causal requiere una causa alternativa.
- Sabrá que las dos respuestas necesarias para la inferencia causal no son observables.
- Percibirá que la inferencia causal precisa premisas sobre la distribución del efecto.
- Conocerá cómo la inferencia estadística permite la inferencia causal.
- Interpretará el efecto causal ajustado como la pregunta natural de un investigador.

Cómo establecer objetivos

A continuación se interpretan, desde el punto de vista de sus implicaciones sanitarias, algunos de los resultados obtenidos en los capítulos anteriores. En este primer punto se estudia cómo establecer el objetivo de un estudio de investigación y las implicaciones que tiene para la interpretación de sus resultados.

El primer tipo de preguntas que abordan los estudios científicos hacen referencia a la distribución de una sola variable y dan respuesta al quién, cómo, dónde y cuándo. Las herramientas estadísticas disponibles se engloban bajo el concepto de «estadística univariante», y abordan el estudio gráfico, descriptivo y la inferencia mediante intervalos de confianza y pruebas de hipótesis.

Dos posibles usos de una relación: predicción e intervención

El segundo tipo de preguntas que abordan los estudios científicos incluyen el estudio simultáneo de dos variables: ¿qué relación existe entre el valor de la presión arterial y el tratamiento administrado? El establecimiento de esta relación puede tener dos utilidades prácticas bien diferentes: ¿se desea **predecir** (adivinar) acontecimientos futuros (el pronóstico), o se pretende **intervenir** sobre el curso de los mismos (el tratamiento y la prevención)? Así, mediante la administración de un tratamiento se pretende modificar la futura evolución de los pacientes. Nótese que un clínico puede decidir el tratamiento para un paciente; pero, en cambio, no puede ni parar ni acelerar el paso del tiempo, por lo que el tiempo previo de evolución sólo puede ser útil para predecir.

Ejemplo 11.1



Los partes meteorológicos permiten anticipar que pasará tal cosa con tal probabilidad, pero no permiten evitar que pase.

Cuando el objetivo es realizar esta predicción, por ejemplo para clasificar o para seleccionar, lo más importante es cuantificar la capacidad predictiva de la relación estudiada. Es decir, ¿hasta qué punto el conocimiento de una variable permite reducir la incertidumbre sobre la otra?

Ejemplo 11.2



Debemos tomar la presión arterial a Jaime, un nuevo paciente. ¿Por qué presión hay que empezar? ¿Hasta qué presión conviene subir el manómetro? Quizá haya que subir hasta 200 mmHg o quizá sea suficiente con 160 mmHg. Ahora bien, si Jaime nos dice «hace unos meses estaba a 120 de máxima». Ahora la «duda» se reduce: ¿subo hasta 160 o sólo hasta 140? Conocer la presión anterior ha reducido la incertidumbre de 40 (entre 200 y 160) a 20 (entre 160 y 140) mmHg.

Ejemplo 11.3

Mi amiga Julia tendrá un descendiente. ¿Niño o niña? Mi incertidumbre es absoluta. Si apuesto por niño, tendré un 50% de posibilidades de error. Y lo mismo si apuesto por niña. Ahora bien, si el ecógrafo, con un margen de error del 1%, predice niña, la incertidumbre se ha reducido del 50 al 1%.

Recuerde

El objetivo de la predicción es anticipar acontecimientos.

Ejemplo 11.4

Un 80% de los pacientes con un determinado nivel de CD4 no desarrollan sida durante los próximos 12 meses.

Recuerde

El objetivo de la intervención es dirigir los acontecimientos futuros.

Ejemplo 11.5

La protección con un preservativo prácticamente anula las probabilidades de contraer el sida.

Cuando el ambicioso objetivo es intervenir sobre una variable para cambiar el valor de otra, es necesario que exista relación de causa-efecto.

Recuerde

Para predecir el valor de una variable a partir del valor de otra es suficiente con establecer una simple relación numérica entre ambas variables. Sin embargo, para poder intervenir se requiere una relación más íntima, llamada de causa-efecto.

El primer punto necesario para estudiar una relación de causa-efecto es dejar claro qué representa la causa y qué el efecto. Demos el símbolo X a la causa potencial (o intervención, o tratamiento, o factor en estudio) y el símbolo Y al efecto en estudio (o consecuencia, o respuesta, o *end-point*).

Ejemplo 11.6

Al modificar la posición del interruptor (X), conseguimos cambiar el estado de la luz (Y). Si el sistema funciona, a botón apretado corresponde luz encendida; y a botón alzado, luz apagada. Dos características de este fenómeno merecen ser remarcadas. La primera es que, con la premisa de que el sistema funciona, existe una relación perfecta, al 100%, entre la posición del interruptor y el estado *encendido/apagado* de la luz. La relación es tan perfecta, que es determinista: conocido el valor de la variable X, es posible «determinar» el valor de la variable Y. La segunda característica es que, modificando la primera variable, se puede provocar un cambio en la segunda. Nótese que, de estas dos características, la segunda es la que nos induce a pensar en el concepto *causa*.

Ejemplo 11.7

Al aumentar el número de horas (X) dedicadas al estudio de una cierta materia se pretende influir en la puntuación en su examen (Y). Igual que en el caso anterior, modificando las horas de estudio, se desea influir en el dominio de esa materia y, asumiendo que el examen es válido y fiable, en la puntuación. Se cree, por tanto, que existe relación causal. Pero en este ejemplo se pueden imaginar otros fenómenos que también influyan en la respuesta: conocimientos previos, capacidad retentiva, habilidad mental, autoconfianza, suerte, etc. Así, a diferencia del ejemplo anterior, la relación no será *perfecta* o determinista. Pero sí que seguirá siendo causal.

Como el universo no se reduce a estas dos variables en estudio, X e Y, se deben considerar aquellas terceras variables Z que puedan estar relacionadas tanto con la causa X como con el efecto Y.

Ejemplo 11.8

El interruptor (X) permite encender la luz (Y) siempre que el resto del sistema (Z) esté en orden. En cambio, un ejemplo de relación numérica pero no causal sería el volumen (X) de la caja de un ordenador y su capacidad de cálculo (Y). Posiblemente, y muy en general, a mayor volumen (X), mayor capacidad (Y). Lo que permite incluso cierta predicción y seleccionar un buen ordenador. Pero no se pretenderá hacer un ordenador más voluminoso (X) para incrementar sus prestaciones (Y) sin modificar sus componentes (Z).

Ejemplo 11.9

Se podría reducir la incertidumbre sobre la mortalidad por cáncer de un país (Y) conociendo el nivel de uso de ordenadores (X), ya que ambas variables van asociadas al desarrollo socioeconómico (Z). Sin embargo, no se propondrá intervenir sobre el uso de la informática (X) para disminuir la mortalidad por cáncer (Y).

Ejemplo 11.10

En niños de 5 a 15 años existe una fuerte correlación entre su estatura y el número de preguntas correctamente contestadas a una prueba de inteligencia. Dado que ambas variables reflejan la maduración física y psíquica que los niños van desarrollando a esa edad, cabe esperar que a mayor estatura, mayor número de respuestas. Por lo tanto, se puede utilizar esta correlación para, conocida la estatura, determinar o, por lo menos, reducir la incertidumbre sobre el número de preguntas acertadas. Pero ¿se puede intervenir?, ¿se consigue aumentar el número de respuestas correctas (Y) si se aumenta la estatura de un niño (X)? El importante papel del paso del tiempo que representa la edad (Z) obliga a abandonar la idea de «manipular» la inteligencia mediante variaciones en la estatura.

Recuerde

La relación entre dos variables puede ser explicada por la relación de ambas con una tercera variable.

Más adelante se estudia con más detalle el papel que pueden desempeñar las terceras variables.

Perspectiva temporal: preguntas prospectivas y retrospectivas

En la relación de causa-efecto hay dos preguntas muy diferentes en función de su perspectiva temporal: la *prospectiva*, que mira hacia delante, y la *retrospectiva*, que mira hacia atrás.

Recuerde

La pregunta prospectiva es: «¿cuál es el efecto de esta causa?».

Ejemplo 11.11

Si me tomo esta medicina, ¿se me pasará el dolor?

Recuerde

La pregunta retrospectiva es: «¿cuál o cuáles son las causas de este efecto?».

Ejemplo 11.12

Se me ha pasado el dolor de cabeza, ¿es porque me tomé una medicina?

Los dos planteamientos difieren en la perspectiva desde la que se mira: el punto de partida de la pregunta puede ser la causa (la medicina) o bien el efecto (el dolor). Pero tienen en común que el efecto (Y) siempre sucede con posterioridad a la causa (X), lo que es un principio inmutable de las leyes físicas de causalidad.

Ejemplo 11.13

Primero, el buey tira, y luego, el carro se mueve.

Contraejemplo 11.14

Imagine que, para estudiar el efecto de la oración en la evolución de los pacientes, en el año 2000 se asignan a dos grupos distintos 400 historias clínicas de pacientes que enfermaron en el año 1990, uno en el que un grupo de personas rezan por ellos (¡en el año 2000!) y otro en el que no. Posteriormente se compara su nivel de salud en el año 1995. Nótese que el efecto antecede en, por lo menos, 5 años a la hipotética causa. Para justificar este diseño es necesario creer que el efecto de la oración no obedece a las leyes de la física y puede, por tanto, modificar actos pasados.

A lo largo de una investigación médica, las preguntas retrospectivas suelen anteceder a las preguntas prospectivas.

Ejemplo 11.15

En su investigación sobre las epidemias de asma en Barcelona (Antó et al., 58), el primer planteamiento (retrospectivo) de los doctores Antó y Sunyer fue: «Hay una epidemia de asma (Y), ¿cuál o cuáles pueden ser sus causas (X)?». Y observaron que las epidemias ocurrían cuando había descarga de soja en el puerto y cierto viento favorable. Entonces se hicieron la pregunta (prospectiva): «Si cubrimos con una lona (X) la descarga de soja, ¿se acabará la epidemia (Y) incluso en los días de viento (Z)?».

Obsérvese la necesaria interacción entre viento y soja. En la pregunta retrospectiva, ambas aparecen como causas (X); mientras que en la pregunta prospectiva, el viento se trata como una condición (Z) que no depende del investigador y se interviene sobre las condiciones de descarga de soja (X), que sí dependen de él.

Recuerde

Las preguntas retrospectivas suelen preceder, en la evolución del conocimiento, a las preguntas prospectivas.

Nota técnica

En el caso de fármacos, este orden se invierte. La farmacología clínica aborda preguntas prospectivas, más concretas («¿Cuál es el efecto de...?»); mientras que la farmacoepidemiología se concentra en preguntas retrospectivas, cuya formulación es más tentativa (¿cuáles son las causas de...?).

La tabla 11-1 resume los diferentes tipos de preguntas clínicas y sanitarias.

Preguntas típicas	Ejemplos	Pregunta prototipo
Descripción	¿Cuál es la proporción de hipertensos? ¿Cuánto tarda?	¿Cuál es la distribución de la variable Y? ¿Cuáles son sus valores centrales? ¿Qué grado de dispersión tiene?
Predicción	Relación entre filtrado glomerular y tiempo de evolución	¿Qué filtrado glomerular (Y) cabe esperar en un caso de tal tiempo de evolución (X)?
Intervención	Causa de... Efecto de...	Retrospectiva: ¿Cuál o cuáles son las causas (X) de este efecto (Y)? Prospectiva: ¿Cuál es el efecto (Y) de esta causa (X)?
	Se me ha pasado el dolor de cabeza, ¿es porque me he tomado una aspirina?	
	Si me tomo la aspirina, ¿se me pasará el dolor de cabeza?	

Tabla 11-1 Resumen de las respuestas que ofrece la inferencia estadística

Estudios experimentales y estudios observacionales

La diferencia primordial entre los estudios observacionales y los experimentales es que en los primeros el investigador «observa» el valor de la causa en estudio, mientras que en los segundos, «asigna» su valor a las unidades.

Por ello, en las investigaciones con seres vivos, la primera consideración para ir hacia un diseño observacional o hacia uno experimental es si la causa es asignable.

Ejemplo 11.16



La esperanza de vida es mucho mayor en las mujeres, pero ningún clínico aconsejará a un paciente masculino que cambie de género. Ni puede plantearse un diseño experimental en que el género se asigne al azar.

Contraejemplo 11.17



En un estudio sobre la discriminación laboral es posible asignar al azar nombres masculinos y femeninos a una serie de currículums y preguntar a un empresario qué salario cree que merecen.

Conviene distinguir si el efecto de esta causa X podría ser beneficioso o no. Si la expectativa es de un efecto positivo, es ético plantear un experimento en seres humanos.

Ejemplo 11.18



Un ensayo clínico sobre un nuevo fármaco. Existen indicios previos tanto de un posible efecto beneficioso como de niveles bajos de efectos tóxicos. La consideración simultánea de ambos efectos, positiva y negativa, permite equilibrar la expectativa en los dos grupos de tratamiento.

En cambio, en aquellas causas potenciales cuyos efectos cabe esperar negativos, no sería ético asignar un voluntario a este tóxico, por lo que sólo se pueden plantear estudios observacionales.

Ejemplo 11.19



Un estudio epidemiológico sobre el efecto del tabaco. No se puede asignar el tabaco a un individuo y pedirle que fume o no fume durante los próximos 20 años para ver si le aparece...

Recuerde



La *asignación* de la posible causa X en estudio divide a los estudios en experimentales y observacionales.

Definición

Ensayo clínico: toda investigación efectuada en seres humanos para determinar o confirmar los efectos clínicos, farmacológicos y/o demás efectos farmacodinámicos, y/o de detectar las reacciones adversas, y/o de estudiar la absorción, distribución, metabolismo y excreción de uno o varios medicamentos en investigación con el fin de determinar su seguridad y/o su eficacia (6).

Definición

Estudio observacional: estudio en el que los medicamentos se prescriben de la manera habitual, de acuerdo con las condiciones normales de la práctica clínica (aquellas establecidas en la autorización de comercialización). La asignación de un paciente a una estrategia terapéutica concreta no estará decidida de antemano por un protocolo de ensayo, sino que estará determinada por la práctica habitual de la medicina, y la decisión de prescribir un medicamento determinado estará claramente dissociada de la decisión de incluir al paciente en el estudio. No se aplicará a los pacientes ninguna intervención, ya sea diagnóstica o de seguimiento, que no sea la habitual de la práctica clínica, y se utilizarán métodos epidemiológicos para el análisis de los datos recogidos (6).

Además de la **causa X** y la **respuesta Y**, hay las terceras variables Z. Usualmente, son condiciones o atributos de los casos, como por ejemplo, el centro o la edad. Dado que su valor no es determinable o asignable por el investigador, sino que las unidades ya se presentan con dicho valor, su única utilidad práctica es predictiva.

Ejemplo 11.20

Ya se comentó que aunque la esperanza de vida es mayor en las mujeres, no tiene sentido aconsejar a un paciente varón que cambie de género.

Recuerde

Las terceras variables Z representan las condiciones o atributos con las que se muestran los casos.

Análisis exploratorio y análisis confirmatorio

El análisis de unos datos se puede realizar con una finalidad exploratoria o con una finalidad confirmatoria.

Definición

El análisis exploratorio pretende generar ideas e hipótesis.

Durante un análisis exploratorio se busca el modelo, más simple o parsimonioso, que mejor se ajuste a los datos. Para ello, todo vale y se torturan los datos en busca de hipótesis.

Recuerde



Un estudio exploratorio debe terminar con expresiones del tipo: «nuestros resultados sugieren...».

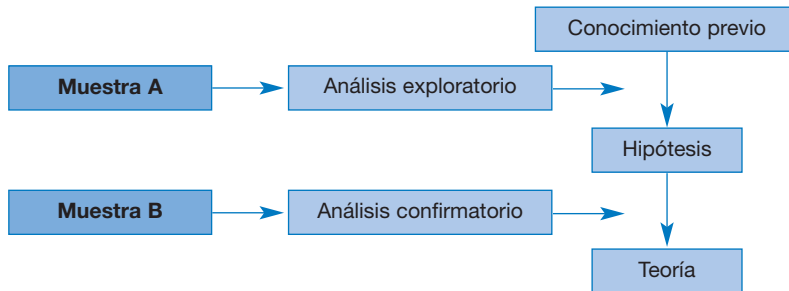


Figura 11-1 Momento del conocimiento científico en el que se emplea cada perspectiva de análisis.

En cambio, las técnicas estudiadas en los capítulos anteriores permitían poner a prueba una hipótesis previamente especificada.

Definición



El análisis confirmatorio pretende elevar a teoría una hipótesis previa.

Aquí, en el análisis confirmatorio ya no se busca la hipótesis que mejor se ajuste a los datos. Ahora se pregunta a los datos si permiten ratificar o rechazar el modelo previo. Dado que el contraste de hipótesis asume unos ciertos riesgos (α y β), este contraste sólo puede realizarse una vez. De lo contrario, estos riesgos no se mantendrían en los niveles deseados.

Ejemplo 11.21



Si se juega un número a la lotería, la probabilidad de que toque es pequeña, pero esta probabilidad sube cuantos más números se compran.

Recuerde



En un estudio confirmatorio sólo se puede hacer un contraste: el de aquella hipótesis especificada en el protocolo.

Historieta

Se dice que un bioestadístico es aquel señor que rechaza que Cristóbal Colón descubriera América porque no estaba en el protocolo de su viaje. Pero es más simple: un bioestadístico le pediría un segundo viaje a Colón para confirmar su hipótesis, tal y como hicieron los Reyes Católicos.

Contraejemplo 11.22

Se dispone de una base de datos que ha ido recogiendo a lo largo de cierto tiempo toda la información sobre partos de un hospital. El investigador ha ido observando que aquellos que hacían una determinada preparación al parto tenían mejor evolución, por lo que decide poner a prueba formalmente esta hipótesis. Si se basa en los mismos datos que la han generado, será un estudio exploratorio; pero si decide recoger nuevos datos (futuros, otros centros, etc.), será confirmatorio.

Definición

Un **estudio prolectivo** tiene una **hipótesis documentada previa** a la existencia de los datos; mientras que en el **retrolectivo** es posterior.

Lectura

Feinstein (59), punto 12.2.1. «Las palabras “prospectivo” y “retrospectivo” se usan a menudo para marcar la diferencia entre planificar la obtención de datos y desarrollar análisis a posteriori, pero su uso es ambiguo debido a que los mismos términos se utilizan para un tipo diferente de distinción temporal, en cuanto a la dirección de las observaciones. Por lo tanto, “prospectivo” también puede referirse a la dirección, hacia delante y longitudinal, en que se siguen las personas en un estudio de cohorte; y “retrospectivo” también puede referirse a la dirección hacia atrás en el seguimiento de un estudio etiológico de casos y controles [...] Para eliminar la ambigüedad, los términos “prolectivo” y “retrolectivo” se pueden utilizar para indicar la existencia de una hipótesis documentada previa a la recolección de datos.»

Lectura

«Las revistas pertenecientes al ICMJE exigen, como condición para la publicación, la inscripción previa en un registro público de ensayos clínicos. [...] Con este objetivo, el ICMJ define un ensayo clínico como cualquier proyecto de investigación que asigne de manera prospectiva a una serie de personas a un grupo de intervención o a un grupo de comparación, con objeto de evaluar la relación causa-efecto entre una intervención médica y un parámetro de salud.» (60)

En resumen, la información y las pruebas (la *evidencia*) aportados por unos datos concretos pueden servir o bien para construir una teoría o bien para ratificarla. Pero unos mismos datos no pueden dar soporte, a la vez, a la construcción y a la ratificación de una teoría.

Recuerde



Para poder finalizar un estudio diciendo «hemos demostrado que...» es necesario un protocolo y una hipótesis previos.

Ejercicio 11.1



Llano et al. (45) terminan su artículo con la siguiente frase: «En su conjunto, nuestros resultados sugieren un papel para IL-7 en el mantenimiento de la regeneración de las células T y la depleción por VIH en individuos infectados, y una posible relación entre IL-7 y la emergencia de variantes SI».

Estas afirmaciones, ¿son teorías o hipótesis? En otras palabras, ¿cree que se trata de resultados de un estudio exploratorio o de uno confirmatorio?

Ejercicio 11.2

¿Qué tipo de utilidad (predicción o modificación) cree que estos autores sugerirán en el texto para la relación entre IL-7 y SI?

¿Y entre IL-7 y la regeneración de células T?

Ejercicio 11.3

La probabilidad de tener un hijo con el síndrome de Down es aproximadamente de uno cada 600 embarazos en la población general, pero asciende hasta uno cada 60 en las gestantes mayores de 36 años.

¿Cuál es el objetivo de establecer relación entre la edad de cada progenitor y el riesgo de síndrome de Down: intervenir o predecir?

Digamos, para terminar este apartado, que las preguntas retrospectivas suelen ser más preliminares, abiertas y exploratorias, mientras que las preguntas prospectivas acostumbran a ser más concretas y confirmatorias.

Ejemplo 11.23

Recuerde que Barcelona estuvo a punto de quedarse sin sus Juegos Olímpicos por las epidemias de asma que aparecían por zonas. Obsérvese el carácter abierto de la primera pregunta retrospectiva que se formuló: «Hay una epidemia de asma, ¿cuál o cuáles pueden ser las causas?». Una vez se atribuyó el asma a la acción conjunta de la descarga de soja en el puerto y el viento que transportaba el alérgeno, se formuló la pregunta prospectiva, que fue concreta y cerrada: «Si reparamos los silos y protegemos con una lona la descarga de soja, ¿se acabará la epidemia?».

Ejercicio cooperativo

A partir de un artículo de investigación, proponga preguntas para que sus compañeros analicen las características del estudio siguiendo todos los puntos vistos en este primer apartado. Idealmente, convendría seleccionar frases en las que los autores sean ambiguos. Preste especial atención a la concordancia entre el objetivo del original y las conclusiones.

Papel de las terceras variables

Se ha dicho que en el estudio entre la causa (X) y el efecto (Y) deben considerarse las terceras variables (Z). En este punto se exponen, con la ayuda de ejemplos numéricos, las formas en las que estas terceras variables pueden interferir en el estudio de la relación entre el efecto y la causa.

El reto de la confusión o cómo evitar sesgos

Si una causa en estudio X está relacionada con una tercera variable Z, los posibles efectos de ambas no podrán diferenciarse y se dice que están confundidos. Por lo tanto, para que exista confusión es preciso que las variables X y Z estén relacionadas, lo que recibe el nombre de colinealidad.

Ejemplo 11.24

Dos tratamientos se utilizan en dos centros diferentes, de atención primaria y terciaria, si bien en el primero (CA1.^a) se tiene preferencia por el tratamiento A, de forma que A se administra 5 veces por cada vez que se ofrece B. En cambio, en el segundo centro (CA3.^a) es justo al revés, ya que A sólo se administra 1 vez por cada 5 de B. Todo ello conduce a un desequilibrio global que la *odds ratio* cuantifica en 25 (tabla 11-2). Existe, por tanto, una fuerte relación entre el tratamiento en estudio (X) y la tercera variable, género (Z).

	x = A	x = B	Razón
CA1. ^a	150	30	5 a 1
CA3. ^a	30	150	1 a 5

$$\text{odds ratio} = \text{OR} = 25$$

Tabla 11-2 El CA1.^a tiene tendencia a aconsejar la intervención A, mientras que el CA3.^a prefiere la B. Centro e intervención están relacionados (OR = 25)

Se dice, en esta situación, que X y Z son colineales y que sus posibles efectos estarán confundidos, como se muestra en la continuación del ejemplo.

Ejemplo 11.25



Pongamos que, por la razón que sea, los pacientes del CA1.^a que evolucionan favorablemente doblan a los que no lo hacen: la razón u *odd* está 2 a 1 (tabla 11-3). Mientras que en el CA3.^a, los que evolucionan favorablemente son la mitad de los que no lo hacen: la *odd* está 1 a 2 (es decir, $1/2$). El cociente de estas dos razones (u *odds ratio*) vale 4 ($2/1/2 = 4$), lo que indica que la razón entre evolución favorable y evolución no favorable es 4 veces superior en el centro 1 que en el centro 2. Dicho de otra manera, en el CA1.^a se curan, por el motivo que sea, 4 veces más. No es preciso ni se pretende afirmar que la relación entre el centro y la evolución de los pacientes sea causal: por ejemplo, los del CA1.^a podrían llegar antes o en mejores condiciones, con mayores posibilidades de mejorar. Imaginemos que los dos tratamientos A y B de este ejemplo no tienen ninguna diferencia en su respuesta, tal y como muestran las tablas 11-4, una para cada centro. Se observa en cada centro una *odds ratio* de 1, indicativo de que no existe relación entre tratamiento y evolución. Pero si, por brevedad o descuido, este investigador no tiene en cuenta el centro y estudia únicamente la tabla 11-5, obtendrá una *odds ratio* de 2,5 con un intervalo de confianza de 1,6 a 3,8, que indica asociación entre tratamiento y evolución. Al estudiar ambos centros en conjunto, el tratamiento parece tener un efecto (OR = 2,5) que en realidad no tiene. ¿Qué está sucediendo? Se están confundiendo los efectos de la tercera variable «centro» con los del tratamiento en estudio. Aquí no engaña la intuición: si el tratamiento A se estudia más en el CA1.^a, que tiene mejores resultados, y el tratamiento B en el CA3.^a, que tiene peores resultados, esta comparación estará sesgada a favor del tratamiento A. (Nótese que, en la tabla 11-2, 150 de los 180 pacientes tratados con A provienen del CA1.^a, y viceversa.) La comparación entre A y B en la tabla 11-5 no es «justa» porque los pacientes no tienen la misma procedencia.

	Y+	Y-	Razón
CA1. ^a	120	60	2 a 1
CA3. ^a	60	120	1 a 2

$$OR = 4$$

Tabla 11-3 La evolución es 4 veces (OR) mejor en el CA1.^a

CA1. ^a	Y+	Y-	Razón
X = A	100	50	2 a 1
X = B	20	10	2 a 1

$$OR = 1 \quad IC_{95\%} = 0,4 \div 2,3$$

CA3. ^a	Y+	Y-	Razón
X = A	10	20	1 a 2
X = B	50	100	1 a 2

$$OR = 1 \quad IC_{95\%} = 0,4 \div 2,3$$

Tabla 11-4 En ninguno de los dos centros, la intervención cambia la evolución

	Y+	Y-	Razón
X = A	110	70	11 a 7
X = B	70	110	7 a 11

$$OR \approx 2,5 \quad IC_{95\%} = 1,6 \div 3,8$$

Tabla 11-5 Estudiando ambos centros a la vez, existe relación entre el tipo de intervención y evolución

Recuerde



La relación entre los tratamientos (X) en comparación y la tercera variable (Z) provoca que se confundan sus efectos: un investigador poco atento podría atribuir al tratamiento una diferencia en el efecto que es explicable por la variable Z.

Un premio al buen diseño: ausencia de sesgo

Por supuesto, un buen investigador no cometería el error anterior, ya que habría considerado la variable Z, o bien en el diseño del estudio o bien en el análisis, y habría así evitado el fuerte desequilibrio del tratamiento en ambos centros. Veamos a continuación las ventajas de un buen diseño.

Ejemplo 11.26



Un buen diseño implicaría idéntica razón entre tratados y controles ambos centros, lo que se traduce por una $OR = 1$ (tabla 11-6). Es decir, ahora el tratamiento está balanceado entre centros, ya que la razón tratados con A frente a tratados con B es idéntica en ambos.

Ahora, si igual que antes no existieran diferencias entre los tratamientos, se podría observar, en cada centro, una situación idéntica a la anterior, como la que muestran las tablas 11-7.

En los dos centros, estudiados por separado, el tratamiento no tiene efecto: en el CA1.^a, ambos tratamientos ($X = A$ y $X = B$) tienen la misma razón «2 a 1» entre evolución positiva ($Y+$) y negativa ($Y-$), lo que conduce a una *odds ratio* de 1, indicador de que el tratamiento no tiene efecto. Y lo mismo en el CA3.^a: como la tendencia hacia una evolución negativa se observa por igual en ambos grupos (1 a 2), tampoco existe relación ($OR = 1$) entre tratamiento y evolución. Si, por parsimonia o por claridad, este investigador quisiera presentar los resultados comparativos de los tratamientos sin distinguir por centro, obtendría los siguientes datos (tabla 11-8), por lo que concluiría también que no existe relación entre tratamiento y evolución. Es decir, estudiando ambos centros a la vez, el tratamiento tampoco muestra efecto alguno, como era deseable.

Bueno	X = A	X = B	Razón
CA1. ^a	90	90	1 a 1
CA3. ^a	90	90	1 a 1

$OR = 1$

Tabla 11-6 Diseño balanceado: intervención y centro son independientes ($OR = 1$)

CA1. ^a	Y+	Y-	Razón
X = A	60	30	2 a 1
X = B	60	30	2 a 1

$OR = 1 \quad IC_{95\%} = 0,5 \div 1,9$

CA3. ^a	Y+	Y-	Razón
X = A	30	60	1 a 2
X = B	30	60	1 a 2

$OR = 1 \quad IC_{95\%} = 0,5 \div 1,9$

Tabla 11-7 En cada centro por separado, no hay relación entre intervención y evaluación ($OR = 1$)

	Y+	Y-	Razón
X = A	90	90	1 a 1
X = B	90	90	1 a 1

$$OR = 1 \quad IC_{95\%} = 0,7 \div 1,5$$

Tabla 11-8 Estudiando ambos centros a la vez, intervención y evolución también son independientes (OR = 1)

En resumen, un investigador cuidadoso obtiene los mismos resultados cuando realiza el análisis teniendo en cuenta la tercera variable Z que cuando no la considera. Y ello a pesar de que la evolución continúa siendo mejor en el CA1.^a que en el cen-CA3.^a, exactamente igual que antes.

Recuerde



Aunque una tercera variable Z esté relacionada con la respuesta Y, si se mantiene controlada mediante un buen diseño, no podrá sesgar la estimación del efecto del tratamiento.

Comentario



La tercera variable Z se puede relacionar con la respuesta Y por cualquier razón. Se podrían poner muchos ejemplos para explicar por qué los pacientes del CA1.^a se curan más que los del CA3.^a. Pero lo importante es que esta relación, sea del tipo que sea, causal o no, está ahí y no puede anularse. Ni tampoco ignorarse. La solución no pasa por «romper» la relación entre Z y Y que no puede depender del investigador: son «las cosas de la vida» que hay que considerar. La solución consiste en anular la relación entre Z y X, como se vio en el ejemplo anterior, mediante un diseño balanceado.

Ejercicio 11.4



En la tabla 11-9 figura un ejemplo de Bishop et al. (40) –analizado también por Freeman et al. (61)– sobre la evolución de un recién nacido (vive, muere) en función de la duración de la preparación materna al parto (cuidados > < 1 mes) y del hospital (A, B). ¿Qué opina sobre la influencia de la variable clínica en la evolución de recién nacido?

Ejercicio 11.5

Supongamos que en la situación anterior, los resultados fuesen los presentados en la tabla 11-10. ¿Qué opina en este caso sobre la influencia de la variable clínica en la evolución de recién nacido?

	Clínica A		Clínica B		Todos	
	Muere	Vive	Muere	Vive	Muere	Vive
Cuidados < 1	3	176	17	197	20	373
Cuidados > 1	4	293	2	23	6	316
OR IC_{95%} OR	1,25 0,28, 5,64		0,99 0,22, 4,57		2,82 1,12, 7,12	

Tabla 11-9 Evolución Y del recién nacido (vive/muere) según el nivel de cuidados prenatales (X) y la clínica (Z). Resultados discordantes

	Clínica A		Clínica B		Todos	
	Muere	Vive	Muere	Vive	Muere	Vive
Cuidados < 1	7	137	10	204	17	341
Cuidados > 1	5	92	7	137	12	229
OR IC_{95%} OR	0,94 0,29, 3,05		0,96 0,36, 2,58		0,95 0,45, 2,03	

Tabla 11-10 Evolución Y del recién nacido (vive/muere) según el nivel de cuidados prenatales (X) y la clínica (Z). Mismos resultados a nivel de clínica y global

Ejercicio 11.6



En el ejercicio anterior no había confusión al no predecir Z la respuesta Y: en la clínica A mueren 12 de 241 (5%) y en la B, 17 de 358 (5%). Por qué otra razón puede no haber confusión?

Se puede recurrir a la distribución marginal cuando no exista confusión, es decir, cuando Z sea independiente, bien de Y o bien de X.

Recuerde



Si una tercera variable Z está relacionada con las dos variables en estudio (es colineal con X y es predictora de la respuesta Y), deben estudiarse los resultados basados en las tablas definidas por la variable Z, lo que equivale a controlar o ajustar por la variable Z.

Ejercicio 11.7



Si quisiera inventarse unos datos en los que haya confusión, ¿qué tendría que tener en cuenta que pasara?

Recuerde

Para que Z sea confusora debe estar relacionada con las dos variables en estudio.

Nota técnica

La afirmación anterior sólo es estrictamente cierta en el modelo lineal (t-test, regresión, ANOVA...), que considera como respuesta Y a una variable cuantitativa. Lamentablemente, el modelo lineal generalizado (que engloba las tablas de frecuencia) carece de esta propiedad (Greenland, 62), por lo que, en raras situaciones, puede conducir a resultados paradójicos.

Otro premio al buen diseño: mayor eficiencia de las estimaciones

Además de variar el valor de la estimación puntual de la relación entre X e Y , el diseño y el ajuste también repercuten en el valor del error típico de estimación y en la consiguiente amplitud del intervalo de confianza.

Ejemplo 11.27

En los datos de los dos centros de los **ejemplos 11.25 y 11.26**, el intervalo más estrecho correspondía al buen diseño en la tabla global, sin ajustar $[0,7, 1,5]$; le seguía el intervalo ajustado obtenido en las dos subtablas, también del buen diseño $[0,5, 1,9]$ y el más amplio era el ajustado en las subtablas del mal diseño $[0,4, 2,3]$.

Comentario

No tiene secreto que el intervalo obtenido en la tabla global, sin ajustar, sea más estrecho que el obtenido en las tablas parciales, ya que dispone de mayor número de casos. Tampoco debe sorprender que el intervalo obtenido al comparar 90 casos de un tratamiento con 90 casos del otro sea más estrecho que el obtenido al comparar 150 casos de un tratamiento con 30 del otro, pues ya se vio que la comparación será más eficiente cuando los dos grupos estén más equilibrados.

Un ajuste planificado en el momento del diseño permite optimizar el rendimiento estadístico de las comparaciones, por lo que un mismo número de casos aporta mayor información y se traduce en un intervalo de confianza más estrecho.

Recuerde

Un buen diseño tiene dos objetivos: aportar **validez** (o controlar sesgos) y aumentar la **eficiencia**, con lo que permite estimaciones más precisas.

El reto de la modificación del efecto

La situación que se comenta a continuación no debe suponer una sorpresa para el lector, ya habituado a que ciertas condiciones (variables Z) modifiquen el efecto del tratamiento X en la respuesta Y .

Ejemplo 11.28



La ingesta de cierta cantidad de alcohol (X) tiene un efecto negativo en las habilidades motoras (Y), que es mayor en las mujeres que en los hombres (Z).

Esta modificación del efecto se conoce, en términos estadísticos, como interacción entre las variables Z y X en la respuesta Y . Esta situación es muy diferente de la provocada por la confusión. La primera diferencia es que un diseño equilibrado no puede proteger contra esta interacción.

Ejemplo 11.29



Supóngase que, para comparar dos tratamientos A y B , tanto en hombres como en mujeres, se ha procedido a realizar la siguiente asignación equilibrada (misma razón entre tratados con A y con B en ambos géneros), tabla 11-11. Si la variable género es modificadora del efecto, podrían observarse, por ejemplo, los resultados de las tablas 11-12, en las que se observa que mientras en los hombres es el tratamiento B el que obtiene mejores resultados, en las mujeres, los mejores resultados corresponden al tratamiento A . La tabla 11-13 muestra la información agregada de todos los casos. Nótese que no tiene ningún sentido, ya que no informa ni del efecto en los hombres ni del efecto en las mujeres. Una vez más, si un investigador no considera la tercera variable género, los resultados carecen de validez, ya que hablarían de dos tratamientos que tienen la misma eficacia, cuando se ha visto que su eficacia depende de la tercera variable.

	$X = A$	$X = B$	Razón
Hombres	60	60	1 a 1
Mujeres	60	60	1 a 1

$$OR = 1$$

Tabla 11-11 Asignación de la intervención X balanceada por género

Hombres	Y+	Y-	Razón
X = A	20	40	1 a 2
X = B	40	20	2 a 1

$$OR = 1/4$$

Mujeres	Y+	Y-	Razón
X = A	40	20	2 a 1
X = B	20	40	1 a 2

$$OR = 4$$

Tabla 11-12 En los hombres, B obtiene mejores resultados ($OR = 1/4$), pero en las mujeres lo hace A ($OR = 4$)

	Y+	Y-	Razón
X = A	60	60	1 a 1
X = B	60	60	1 a 1

$$OR = 1$$

Tabla 11-13 Al estudiar conjuntamente a hombres y mujeres, las dos intervenciones obtienen idénticos resultados

Recuerde



La *interacción* puede aparecer aunque no exista colinealidad.

Si hay interacción, para explicar el efecto del tratamiento se debe distinguir por el nivel de la tercera variable. En el caso de la confusión, era necesario considerar la tercera variable en el análisis y analizar los resultados condicionando o ajustando por ella, pero se podía presentar un efecto único del tratamiento, ya que era el mismo en los diferentes niveles de la variable Z.

Ejemplo 11.30



En este último ejemplo, la relación entre el tratamiento y la respuesta es diferente para cada género, y para explicar el efecto es necesario aclarar de qué género se está hablando. En el ejemplo 11.25, el efecto (nulo) era el mismo en ambos centros y se podía hablar de un único efecto (nulo) del tratamiento: todas las OR, tanto las parciales como la global, valían 1.

En el ejemplo 11.29 de modificación del efecto, la OR para los hombres vale $1/4$, mientras que la de las mujeres es de 4.

Recuerde

Si Z es una variable confusora, la relación parcial (condicionando por Z) difiere de la relación marginal (sin ajustar por Z).

Si Z es una variable modificadora del efecto, las relaciones parciales (condicionando por los diferentes niveles de Z) difieren entre sí.

Recuerde

La interacción, a diferencia de la confusión, requiere aportar resultados que varían según la tercera variable.

Ejercicio 11.8 (Cont. 6.20)

En un Centro de Atención Primaria han realizado un experimento para poner a prueba un nuevo tratamiento contra las molestias gástricas pospandriales. Para ello han asignado al azar o bien el tratamiento en estudio, o bien un placebo idéntico. El paciente, cuando le aparecía el dolor, en su domicilio, tomaba el tratamiento asignado y anotaba en un cuadernillo si había curado o no.

A pesar de que la asignación había sido realizada al azar, la «mala suerte» (o una selección diferente de los casos facilitada por un diseño no enmascarado) quiso que los pacientes quedaran desequilibrados en cuanto a una variable muy importante: el nivel de dolor que sentían al inicio en el momento en que decidían tomar la medicación. Los pacientes con dolor suave fueron mayoritariamente al grupo placebo, mientras que los pacientes con dolor fuerte pertenecían al grupo tratado. En la tabla 11-14 se exponen los resultados obtenidos. ¿Qué papel desempeña el dolor basal?

	Dolor suave		Dolor fuerte		Todos	
	Cura	No cura	Cura	No cura	Cura	No cura
Tratados	100	5	650	650	750	655
Placebos	550	80	10	200	560	280
OR	2,91		20		0,57	
IC_{95%} OR	1,15, 7,36		10,5, 38,1		0,48, 0,68	

Tabla 11-14 Todas las OR son diferentes (y sus IC_{95%} no se solapan). ¿Qué está pasando?

Comentario



Que exista interacción complica el análisis. Por tanto, la vida será más fácil si no existe. Pero demostrar o negar su existencia no es un objetivo tan importante como estudiar la relación entre X e Y. Si no puede asumirse que no haya interacción, posiblemente el primer paso puede ser establecer el efecto en un subgrupo homogéneo.

La interacción o modificación del efecto es un concepto íntimamente ligado a la medida del efecto empleada. Puede aparecer con una medida y no aparecer con otra.

Ejemplo 11.31



En la tabla 11-15 puede verse que mientras el riesgo relativo es idéntico, el riesgo atribuible es marcadamente diferente. Usando el riesgo relativo podrá decirse que el tratamiento multiplica por 1,5 la probabilidad de curarse, mientras que si se usa el riesgo atribuible habrá que especificar que, cuando el dolor es suave, la proporción de curados aumenta un 20% mientras que cuando es fuerte, sólo lo hace un 5%.

	Dolor suave		Dolor fuerte	
	Cura	No cura	Cura	No cura
Tratados	60	40	15	85
Placebos	40	60	10	90
OR	2,25		1,59	
RR	1,50		1,50	
RA	0,20		0,05	

Tabla 11-15 Al presentar interacción, un único valor del RA no sirve para resumir la relación en ambos grupos de pacientes

Dado que el objetivo científico de parsimonia aconseja utilizar el número mínimo de parámetros para explicar un fenómeno, siempre será preferible utilizar aquella medida que no precisa interacción, ya que entonces puede aplicarse a todos los grupos la misma medida del efecto.

Comentario



El objetivo principal de un estudio suele ser si interviniendo sobre X se consigue modificar la respuesta Y en las condiciones Z. La cuestión sobre si Z es o no confusora (o modificadora) es, en el fondo, secundaria y sirve para decidir el tipo de análisis, de presentación de resultados y la población objetivo. Por ello, en un estudio confirmatorio debe ser algo ya conocido previamente, que formaría parte de las premisas del estudio actual.

Ejercicio cooperativo

Busque ejemplos de la literatura en los que las terceras variables puedan actuar como confusoras o modificadoras del efecto. Proponga preguntas concretas que permitan comprobar que su compañero distingue ambos roles.

Definición de efecto causal

Llevamos un rato hablando de efectos. Vamos ya a definir el efecto causal, primero en una unidad y luego en la población de unidades. Posteriormente, abordaremos el reto de considerar a las terceras variables Z .

Definición del efecto causal en una unidad

Para definir el efecto causal necesitamos establecer un marco. Escojamos primero a la respuesta Y en la que se desea estudiar dicho tratamiento. En segundo lugar, por simplicidad, estudiemos el efecto de cambiar el control c por el tratamiento t . Representemos ahora por $Y_{t,i}$ el valor de la respuesta que se observaría en el individuo i si se le asignara t . Sea $Y_{c,i}$ la respuesta que se observaría si se le asignara c .

Ejemplo 11.32

En el control de la presión arterial, si se desea cambiar en el paciente José López el tratamiento «dieta baja en sal» (c) por el tratamiento «dieta baja en sal más bloqueadores β » (t) y se decide analizar como respuesta Y a la presión arterial diastólica; $Y_{t,i}$ representa la PAD que tendrá el Sr. José López cuando se le administren bloqueadores β más dieta, mientras que $Y_{c,i}$ representa la PAD del mismo paciente si se le hubiera aconsejado dieta sola.

Definición

El efecto causal (Rubin, 63; Holland, 64) es la diferencia entre la respuesta al asignar t y la respuesta al asignar c : $Y_{t,i} - Y_{c,i}$

Ejemplo 11.33

El efecto, en la PAD del Sr. López, de añadir bloqueadores β a la dieta es la PAD de este paciente cuando se le administran bloqueadores β y dieta menos la PAD del mismo señor cuando sólo se le aconseja dieta. Esta definición, aunque pretende ser simple, tiene una serie de connotaciones nada triviales.

La definición de efecto causal es prospectiva

En primer lugar, nótese que se ha definido el *efecto de una causa* y no al revés. Es decir, se está contestando la pregunta prospectiva.

Ejemplo 11.34



Al buscar la causa del tétanos (pregunta retrospectiva) podría encontrarse cierta competencia entre causas que se engloban (el *Clostridium* engloba la toxina), de forma que diversos investigadores, en diferentes momentos del conocimiento científico, podrían atribuir la causa a diferentes componentes del *Clostridium*. En cambio, no habría ambigüedad a la hora de estimar el efecto de cada una de estas posibles diferentes causas (pregunta prospectiva).

Comentario



Esta definición «prospectiva» es la que responde al objetivo final de la sanidad, que es, mediante intervenciones, mejorar el estado de salud de la población.

Nota técnica



Aickin (65) ha propuesto un planteamiento alternativo que pretende modelar el entorno multicausal que suele aparecer en las preguntas retrospectivas. En cambio, la aproximación prospectiva que se presenta aquí considera a las concausas como condiciones en las que se estima el efecto de la causa que se cree que se está en condiciones de intervenir.

La definición de efecto causal habla de asignación

Una segunda consideración en la definición realizada es que se ha hablado de asignación. Esta definición sólo estudia el efecto de aquellos fenómenos en los que el investigador puede o podrá decidir el valor que toman. En una palabra, sólo sirve para aspectos **intervenibles**.

Ejemplo 11.35



Ya se ha dicho que aunque la esperanza de vida de las mujeres es superior a la de los hombres, un clínico, al ver entrar a un paciente masculino no le pide que abandone ese mal hábito.

De acuerdo con esta definición, es fundamental distinguir entre tratamientos —causas susceptibles de una intervención— y los atributos de las unidades, que son condiciones o características poseídas por los individuos. La capacidad de intervención se reduce a los primeros.

Ejemplo 11.36

Mientras no se pueda asignar la edad a un ser vivo, esta definición no permitirá estudiar si la relación entre edad y riesgo de morir es causal. Pero ¿es una limitación? A fin de cuentas, si no se puede decidir el valor de la variable *edad*, ¿para qué importa saber si es una relación causal? Desde una perspectiva práctica, es irrelevante si la relación es causal o no: dado que la edad no depende del investigador, se ha perdido toda posibilidad de influir en la esperanza de vida por medio de la edad. Tanto si la relación es causal como si no lo es, se podrán utilizar las propiedades predictivas de una relación y así predecir el tiempo de vida en función de la edad. También se pueden definir subgrupos homogéneos para un estudio determinado. Pero no tiene utilidad práctica, mientras no se pueda *asignar* la edad, preguntarse si esta relación es o no es causal.

Contraejemplo 11.37

Ciertos hábitos de vida pueden acelerar el envejecimiento celular. Tiene pleno sentido preguntarse si cambiarlos tiene un efecto causal sobre la cantidad y la calidad de vida.

Lectura

El concepto de lo que es intervenible no es inmutable. Como apunta Cox (66), aunque no se puede cambiar la edad de una persona, quizá sí se pueden ralentizar ciertos cambios bioquímicos o moleculares ligados al envejecimiento. Así, la edad no sería assignable, pero sí que lo sería la decisión de ralentizar o no esos cambios.

Ejemplo 11.38

Cabe esperar que un paciente hospitalizado cumpla al 100% con el tratamiento establecido, pero eso no es así para un paciente ambulatorio. El profesional sanitario puede tomar la decisión de asignar un tratamiento u otro a un paciente, pero el grado de cumplimiento posterior depende de este último. Si se quiere valorar el impacto de una intervención sanitaria, se debe estudiar aquello que se desea cambiar en el futuro: la decisión del clínico.

Contraejemplo 11.39

Por supuesto, pueden diseñarse intervenciones para mejorar el grado de cumplimiento. Por ejemplo, podría compararse el nivel de cumplimiento de la medicación entre un blister clásico y un expendedor electrónico con voz que recordara el momento de la toma de la medicación. En ese caso, el clínico podría decidir entre asignar el paciente a un método u otro para dispensar el tratamiento.

Ejercicio 11.9

Cuando se habla de que cierta variable es un factor de riesgo, ¿quiere decir que la relación es causal y por tanto, que mediante una intervención sobre esa variable se lograría disminuir el riesgo?

La definición de efecto causal es relativa a otra causa

Una tercera consideración relevante de la definición presentada es que se ha definido el efecto de una causa de forma relativa a otra causa alternativa. Ello no tiene por qué ser una limitación, ya que emula el proceso de decisión de la vida real, en la que, por lo menos, aparece la alternativa de no hacer nada. ¿Para qué se podría desear conocer las consecuencias o efectos de nuestros actos, si no hay alternativa posible?

Ejemplo 11.40

Supóngase que acude a urgencias un paciente con una fractura de hueso largo para la que el mínimo tratamiento éticamente aceptable es la reducción e inmovilización. Por tanto, no puede observarse la evolución de una alternativa que consista en no hacer nada, en no aplicar el tratamiento.

Existe una situación en la que es posible sustituir la observación de la evolución por el valor hipotético de lo que habría pasado si no se hubiera hecho nada. Se trata de la invarianza o no variabilidad de los datos. Exige poder predecir sin error los valores futuros en caso de que no se interviniera.

Ejemplo 11.41

¿Aceptaría un oftalmólogo la premisa de invarianza en un paciente con cataratas? Si los miembros más críticos de este colectivo científico estuvieran dispuestos a asumir que, en ausencia de intervención, la evolución de las cataratas es perfectamente conocida, creerían en esta invarianza.

Ejemplo 11.42

En los primeros años de la aparición del sida, ningún paciente de cierto estadio evolutivo lograba sobrevivir más de, pongamos, seis meses desde el diagnóstico: invariablemente morían antes. Si un científico aplica un tratamiento nuevo a un paciente que, sin duda alguna, tiene el sida en dicho estadio evolutivo, y este caso único permanece vivo a los doce meses, el resto de la comunidad científica estará muy interesada en dicho tratamiento. A pesar de disponer de una sola observación.

La premisa de invarianza es raramente sostenible. La situación habitual es justamente la contraria: existe una variación notable entre e intra individuos. Si la evolución de los pacientes es ciertamente impredecible, la obtención de datos sobre el tratamiento control es obligada.

Además de esta incertidumbre sobre la evolución de los pacientes, si existe variabilidad intrasujeto y si se seleccionan a los individuos que estén en peores condiciones, aparecerá un fenómeno denominado regresión a la media que originará que, en una nueva observación, estos casos estén en mejores condiciones.

Ejemplo 11.43

La artritis o «reuma» es una afectación con grandes variabilidades dentro de un mismo paciente. Si un paciente decide acudir al médico cuando «ya no aguanta más», posiblemente esté acudiendo en el peor momento de su enfermedad, por lo que tiene más posibilidades de encontrarse mejor en los próximos días, aun en ausencia de tratamiento.

Ejemplo 11.44

En el rendimiento académico suele existir una cierta variabilidad, ya que puede influir la suerte o la inspiración del momento, por ejemplo. El alumno que ha sacado el mejor puesto ya sabe lo difícil que es mantenerse ahí. Es frecuente dar un estímulo negativo (castigo o bronca) después de los peores rendimientos, y uno positivo (premio), después de los mejores. Pero si existe variabilidad intraindividual en el rendimiento académico, cabe esperar que después de la peor nota (que se acompaña de bronca) vengan mejores resultados, y después de la mejor nota (que se acompaña de premio), peores resultados. Dado que los estímulos positivos y negativos se dan en condiciones diferentes, un observador poco atento podría confundir los efectos de la regresión a la media con los hipotéticos efectos de los estímulos.

Recuerde

La presencia de variabilidad obliga a recoger información sobre el control.

Comentario

Puede argumentarse que si existe un único tratamiento para una determinada enfermedad, no tiene ningún interés estimar su efecto ya que no hay alternativa y dicho tratamiento deberá aplicarse forzosamente. Sin embargo, el concepto de éticamente aceptable depende del entorno y de los recursos disponibles: el responsable de la administración de los recursos sanitarios decide el patrón de mínimo tratamiento éticamente aceptable en función del beneficio y el coste de las diferentes intervenciones sanitarias. Ahora, la pregunta no es asignar un tratamiento u otro a un mismo paciente, sino asignar el recurso sanitario a una u otra intervención. A diferencia del profesional asistencial, que cuida personas, el administrador debe velar por la salud de toda una población. Su paciente es la comunidad, y debe decidir la prioridad de las diferentes intervenciones, exactamente de la misma forma en que debe hacerlo un clínico delante de un paciente politraumatizado.

Problema fundamental de la inferencia causal

Para estimar el efecto causal en una unidad determinada, se ha propuesto hacer la resta entre los dos valores de la respuesta bajo los dos tratamientos alternativos, $Y_{t,i}$ y $Y_{c,i}$. Ahora bien, nótese que estas dos respuestas no son observables a la vez y en las mismas condiciones.

Ejemplo 11.45

O bien se puede observar la respuesta al tratamiento bloqueadores β más dieta (PAD_{BB+D}) en unas unidades y la respuesta a la dieta aislada (PAD_D) en otras; o bien, en el mismo individuo, se puede observar la respuesta PAD_{BB+D} , López en unas condiciones en un cierto momento y la respuesta PAD_D , López en otras condiciones de otro momento. Pero nunca en idénticas circunstancias.

Al exponer la confusión, se vio que estas diferencias —sea en unidades, sea en condiciones, sea en ambas— abren la posibilidad de que terceras variables Z fueran la auténtica explicación de las diferencias observadas.

Ejemplo 11.46

Podría ser que se estudiaran pacientes de diferente edad (Z_1) en cada grupo. O que ambos grupos sean de diferentes centros (Z_2). O que en un segundo período ya no exista alérgeno en el ambiente. O que el primer tratamiento haya conseguido curar la enfermedad antes de dar el segundo.

¿Puede una tercera variable Z ser una explicación de la relación observada entre X e Y ? O, dicho de otra manera, ¿cómo se podrá afirmar que las diferencias en la respuesta Y son debidas a X y no a diferencias entre unidades, o a diferentes condiciones o variables Z ?

Recuerde

Se ha definido el efecto causal en una unidad, pero no es directamente observable. Este problema es tan delicado que recibe el nombre de problema fundamental de la inferencia causal, y constituye el reto de la inferencia estadística y el diseño de experimentos.

Definición de efecto causal en la población

El primer paso consiste en extender al conjunto de todas las unidades el efecto definido en una única unidad. Para generalizar el efecto causal es preciso realizar alguna premisa sobre su distribución a lo largo de las unidades. Al estudiar la comparación de dos medias mediante la t de Student se ha usado la premisa de un efecto constante y aditivo, es decir: idéntico en cada individuo, y que simplemente se suma o resta a su tendencia natural.

Comentario

Este modelo se conoce técnicamente con el nombre de «efectos fijos». En él, los diferentes resultados en las distintas unidades se interpretan por divergencias iniciales entre los individuos, no porque el efecto del tratamiento vaya cambiando de una unidad a otra.

Nota técnica

Ya se vio en el capítulo 8 que esta premisa tan simple está detrás del análisis estadístico habitual: el efecto del tratamiento se manifiesta en las medias, pero no en las dispersiones.

Recuerde

El efecto promedio, o diferencia entre medias, se corresponde con el efecto del tratamiento en cada uno de los casos.

Bajo la premisa de efecto constante, todas las unidades proporcionan información del mismo efecto y pueden combinarse para estimarlo. Ello permite extender a toda la población aquella definición que sólo servía para el Sr. José López.

Definición



El efecto causal es la diferencia entre la respuesta media al administrar t y la respuesta media al asignar c : $E[Y_t] - E[Y_c]$.

Nota técnica



$E[Y_t]$ se lee como «esperanza de Y_t », y es la media poblacional, μ_t , que representa el centro de todos los posibles casos tratados con t .

Establecer una premisa sobre el comportamiento del efecto a lo largo de las unidades permite, no sólo observar qué ha sucedido en el pasado en la unidad estudiada, el Sr. José López, sino también anticipar qué sucederá en las próximas unidades en las que se administre el tratamiento.

Nota técnica



Si la premisa de efecto constante se considera simplista, puede relajarse de varias formas. Una de ellas, conocida como modelo de efectos aleatorios, asume que el efecto del tratamiento tiene una distribución de valores que puede resumirse mediante su media y su desviación típica. Ahora se deben estimar, pues, dos parámetros: el del efecto promedio y el de la distancia usual del efecto en cada unidad a este efecto promedio. Además de requerir diseños sofisticados, que implican determinaciones repetidas en cada individuo, este modelo tiene el inconveniente de su aplicabilidad posterior, ya que, cuanto mayor sea esta dispersión, mayor será la incertidumbre de cuál será el efecto en una unidad determinada, lo que puede llegar a requerir estudios « n igual a 1» para determinar el efecto en cada caso particular.

Podría ser que esta variación del efecto entre las unidades siguiera cierto patrón, de forma que dependiera de una característica observable, por ejemplo, el género del paciente. Así, el tratamiento podría tener un efecto en los hombres y otro en las mujeres, que empezaríamos por considerar como constantes o fijos. Se trata de la interacción ya vista anteriormente entre género y tratamiento (o de modificación del efecto del tratamiento por el género). Ahora, se deben estimar dos parámetros, el del efecto en cada una de las poblaciones, lo que requiere duplicar el esfuerzo experimental.

Ejemplo 11.47

La ingestión de alimentos puede modificar (o interferir o interactuar) la absorción y, por tanto, el efecto de un tratamiento. Por lo general, las agencias reguladoras no exigen estudios en los que se recoja simultáneamente la información en las dos situaciones, antes y después de la ingesta, sino que aceptan estudios complementarios que aporten la información adicional no contemplada en los estudios principales.

Definición de efecto causal ajustado por terceras variables

El siguiente paso para solucionar el problema fundamental de la inferencia causal es considerar las terceras variables.

Todo estudio se realiza en ciertas condiciones. Un investigador debe explicar estas condiciones, no sólo para describir la situación en la que se realiza el estudio sino también para permitir su réplica por futuros investigadores.

Se debe, por tanto, matizar la definición de efecto causal, incluyendo estas características Z.

Definición

El efecto causal ajustado por Z es: $E[Y_t|Z] - E[Y_c|Z]$.

Nota técnica

De la misma forma que $E[Y_t]$ representa la media μ en los casos tratados con t, ahora $E[Y_t|Z]$ representa la media μ en los casos tratados con t a nivel fijo de la variable Z.

En esta expresión, utilizar el mismo símbolo Z en los dos términos de la resta significa que se está condicionando en ambos por los mismos valores de Z. Es decir, se habla de valores de Y_t «dada Z»: se ha dejado «fijo» o constante el valor de Z.

Ejemplo 11.48

Si en el estudio de la PAD, las variables Z que se desean controlar son la edad y el centro, el efecto causal ajustado será el incremento al cambiar el tratamiento c por el tratamiento t en la PAD ajustada por edad y centro, es decir, para un paciente de una edad y de un centro determinados.

Esta definición contesta la pregunta de interés y, además, convierte en constantes las variables Z que, por tanto, ya no pueden ser una explicación alternativa a la relación causal en estudio: ¡no puede existir confusión porque las variables Z son fijas!

Comentario



El problema fundamental de la inferencia causal aparecía porque el efecto causal definido en las unidades no podía observarse, ya que la misma unidad no podía ser estudiada en las mismas condiciones bajo ambos tratamientos.

En cambio, una población sí que puede ser estudiada en las mismas condiciones, bajo tratamientos diferentes. Esta estimación del parámetro central de una población es el reto habitual de la estadística, que ya se ha solucionado en los capítulos anteriores.

Si se conoce la influencia de las terceras variables y, además, éstas son observables y medibles sin error, es posible tenerlas en cuenta mediante las técnicas de ajuste estadístico.

Comentario



Un investigador serio, precavido y buen conocedor de su profesión tendrá en cuenta estas variables al diseñar el estudio y dejará constancia de ello en el protocolo del mismo. Un estudio confirmatorio que pretenda convertir una hipótesis en teoría debe especificar de forma concreta, en el protocolo, las condiciones del estudio y la metodología para tenerlas en cuenta. Las desviaciones del protocolo restarán credibilidad a la inferencia causal.

Ejercicio 11.10



(Continuación del ejercicio 11.3) Si el objetivo de conocer la relación entre la edad materna y el riesgo de Down es intervenir sobre esta edad aconsejando a las mujeres que adelanten su gestación, ¿qué papel juega la edad paterna?

(Nótese la elevada correlación entre la edad de los padres: por un lado, se sabe que a mayor edad de uno, cabe esperar mayor edad del otro; por otro, si se deja fija la edad de uno de los padres, disminuye considerablemente la variabilidad de la edad del otro. Por tanto, también es razonable que exista relación entre la edad del padre y el riesgo de síndrome de Down.)

Soluciones a los ejercicios

11.1 La expresión *sugieren* («*suggest*» en el original) es claramente indicativa de un estudio exploratorio. Estos autores están aportando al conocimiento nuevas ideas que deberán ser contrastadas posteriormente.

11.2 No queda claro. Estas expresiones son habituales en la jerga científica: podrían estar redactadas así para fomentar la lectura del texto y la discusión, donde se aclara que IL-7 puede ser utilizado como «indicador» de la evolución (utilidad predictiva). No se afirma nada sobre si, además, puede ser utilizado para intervenir.

11.3 Si lo que se pretende es cambiar la edad de la madre o del padre para disminuir el riesgo, se trata de una intervención. Pero si se desea seleccionar aquellos embarazos con elevada probabilidad de síndrome de Down para justificar una prueba diagnóstica no exenta de riesgo, entonces el objetivo es predecir.

11.4 Mientras que el estudio marginal, juntando todos los casos, muestra la existencia de relación, la asociación parcial no detecta esta relación. Por tanto, la variable clínica es confusora: condicionando por ella, no es necesario considerar la relación entre evolución y cuidados. Nótese la mayor administración de cuidados en la clínica A (relación ZX) y la mayor mortalidad en la B (relación ZY).

11.5 Se puede observar que la OR marginal y las asociaciones parciales coinciden, lo que indica que la variable clínica no es confusora y podemos recurrir a la distribución marginal, donde concluimos que la variable clínica no tiene relación con la evolución del recién nacido. Utilizar la tabla marginal cumple con el principio científico de parsimonia (si es posible, utilice el modelo más simple) y con la necesaria claridad en toda comunicación. Pero además, tiene el premio de una mayor eficiencia estadística: los intervalos de confianza son más estrechos, más precisos.

11.6 Porque las variables Z y X, por el diseño experimental, sean independientes.

11.7 Para que exista confusión, las variables Z y X deben ser colineales (estar relacionadas) y Z debe ser predictora de la respuesta Y.

11.8 Utilizando los intervalos de confianza al 95%, la interpretación sería la siguiente. En los que tienen dolor suave, el fármaco tiene un efecto positivo moderado: aumenta la disparidad «cura/no cura» entre 1,15 y 7,36 veces. En los que tienen un dolor fuerte, el efecto es más marcado: aumenta la razón «cura/no cura» entre 10,5 y 38,1 veces. En cambio, al realizar el estudio global con todos los casos, se llega a la conclusión contraria: el fármaco disminuye las posibilidades de curarse entre 0,48 y 0,68 veces.

Ajustando por dolor basal, ambos intervalos no se solapan, lo que hace pensar en un efecto diferente del tratamiento según la intensidad del dolor basal: esta variable actúa como modificadora del efecto (dolor y tratamiento interactúan). No tendría sentido ofrecer una estimación conjunta o global del efecto del tratamiento, ya que éste no es el mismo en los dos grupos de dolor basal. Pero, además, mientras que los dos intervalos ajustados se sitúan por encima de 1, lo que indica que los curados se asocian con los tratados, el intervalo sin ajustar se sitúa por debajo del 1, indicador

de la situación contraria: los curados se asocian con los placebos. Es decir, el dolor basal, además de modificador del efecto, es confusor. Esta indeseable confusión ha podido aparecer por el sesgo de selección (los más graves han ido al grupo tratado), resultado de una asignación no enmascarada.

11.9 No hay uniformidad en su uso. Por un lado, el término riesgo hace referencia tan sólo a una cierta probabilidad de un evento negativo, por lo que no especifica si detrás hay una relación causal. Por otro, algunos autores aconsejan, quizá acertadamente, reservar el término «factor de riesgo» para relaciones causales y hablar de factor pronóstico cuando no es así. Dada esta diversidad en el uso del término, cuando un autor se refiera a una variable como factor de riesgo, es muy aconsejable no dar por sentado que la relación es causal, mientras no se diga explícitamente.

11.10 Si el objetivo fuera intervenir, para poder establecer la necesaria relación causal, la mejor respuesta vendría de un diseño experimental en el que se estudiara la edad de cada uno de los padres dejando equilibrada la del otro. Por ejemplo, se aparearían tanto las madres de 20 como las de 40 años, por igual, con padres de 20 y de 40 años. Este diseño no sería ético, por supuesto, pero es que además no tendría sentido práctico en nuestro entorno sociocultural. ¿Se plantea una madre o un padre buscar una pareja «independientemente» de su propia edad? Si una madre potencial decidiera adelantar su edad de maternidad, posiblemente también estaría adelantando la edad de su pareja.

Capítulo 12:

Tipos de diseños

Presentación

Este capítulo describe las herramientas de las que dispone el investigador para controlar las terceras variables: primero los métodos para aquellas que son observables, y luego cómo el azar ayuda a controlar las terceras variables no conocidas o no observables.

Finalmente, se describen los diseños más habituales en Ciencias de la Salud.

Objetivos

Al terminar este capítulo, un lector que haya realizado los ejercicios:

- Distinguirá entre variables de las unidades (condiciones Z) y variables del investigador (intervenciones X).
- Se preocupará por observar si el criterio de ajuste se ha decidido independientemente de los resultados.
- Valorará más un ajuste ya preespecificado en el protocolo.
- Sabrá que los principales métodos de ajuste son la restricción, el estudio de subgrupos y el modelado estadístico.
- Se preocupará porque el estudio de subgrupos persiga que la intervención en estudio esté balanceada en todos los grupos.
- Identificará el apareamiento con el caso extremo de subgrupos en el que cada subgrupo está formado por una pareja.
- Sabrá que el modelado también conviene especificarse en el protocolo.
- Interpretará a nivel exploratorio los ajustes no especificados en el protocolo.
- Centrará el interés de un estudio de intervención en el efecto de la causa X en la respuesta Y dadas las condiciones Z.
- Al diseñar estudios confirmatorios de intervención, las relaciones de Z con X y de Z con Y formarán parte de las premisas previas.
- En los estudios exploratorios de intervención, antes de ajustar por Z se preguntará por la viabilidad de intervenir sobre X a nivel fijo de Z.
- En presencia de modificación del efecto, preferirá empezar por el estudio del efecto en un único subgrupo.
- Distinguirá entre las condiciones observables Z y las no observables V.

- Previamente a realizar inferencia causal, se preguntará por el posible papel de las condiciones no observables V.
- Sabrá que la asignación aleatoria permite equilibrar a nivel poblacional las condiciones no observables V.
- Sabrá que la asignación aleatoria no garantiza un equilibrio absoluto entre las muestras en comparación.
- Recurrirá a los métodos de ajuste para equilibrar las condiciones Z relevantes.
- Recurrirá a la asignación al azar para equilibrar las condiciones V no observables y las Z poco relevantes.
- Definirá como transversal el estudio en que la recogida de los valores de X e Y sean simultáneos.
- Definirá como longitudinal el estudio en el que los valores de X e Y están separados por un lapso de tiempo suficiente para que los hipotéticos efectos de X actúen en Y.
- Definirá como prospectivo el estudio longitudinal en el que la recogida de la información sigue su orden natural (se observa X y se espera en el tiempo a que acontezca Y).
- Definirá como retrospectivo el estudio longitudinal en el que la recogida de la información invierte el orden natural (tras observar Y, se averigua X en el pasado).
- Distinguirá si el término control lo es de una causa (como en el ensayo clínico) o de un efecto (como en diseño de casos y controles).
- Se preocupará por la eficiencia de un diseño y de un análisis.
- Sabrá que los diseños apareados o los del estudio del cambio persiguen una mayor eficiencia eliminando la variabilidad interunidad.
- En un estudio que relacione dos variables se preocupará por recoger la información de cada una de ellas sin que la segunda variable pueda afectar los valores de la primera.
- Identificará como experimental el estudio en el que el investigador decida el valor de la causa en estudio (asigne el tratamiento).
- Percibirá las ventajas del diseño experimental en cuanto a equilibrar las variables desconocidas V y a optimizar la eficiencia de la estimación.
- Se preocupará porque el investigador que incluya los pacientes en el estudio desconozca la secuencia de aleatorización.
- Identificará los ensayos comunitarios como aquellos en los que la asignación del grupo de intervención se decide en conjunto para un grupo de unidades.
- Percibirá la grave amenaza de los desvíos del protocolo.
- Sabrá que sólo el análisis por intención de tratar (ITT) respeta el carácter experimental del estudio.

Métodos de control de las terceras variables observables

Hay que tener muy claro qué papel desempeñan las diferentes variables en un estudio de intervención.

Recuerde



En este texto, se representa por Y a la variable respuesta; por X, a la intervención y por Z, a las condiciones de las unidades.

Ejemplo 12.1



Un tratamiento farmacológico o un consejo dietético pueden ser considerados como variables potencialmente intervinientes (X).

Ejercicio 12.1

Un estudio pretende evaluar el efecto que, sobre el nivel de cumplimiento de la medicación, tiene el número de visitas de seguimiento.

Suponga:

- a) un estudio experimental, que asigna a los pacientes un determinado número de visitas, y
- b) un estudio observacional, que simplemente observa el número de visitas espontáneamente solicitadas por el paciente.

¿El número de visitas es X o Z en cada estudio?

Recuerde

Una variable susceptible de ser intervenida debe ser «asignable».

El capítulo anterior se terminó definiendo el efecto causal ajustado por los valores de las condiciones Z. Antes de comentar las herramientas de las que dispone el investigador para dejar fijas estas variables Z, conviene recordar que un ajuste bien planificado suele conllevar un beneficio en la precisión de la estimación.

Ejemplo

En el ejemplo de la PAD, cabe esperar menos diferencias entre dos pacientes de 40 años tratados en el mismo centro que entre dos pacientes de diferentes edades y diferentes centros. La comparación directa de estos dos pacientes, al estar menos influida por fluctuaciones aleatorias, aportará más información.

Recuerde

El ajuste permite controlar la variabilidad de la variable Z que se deja fija. Eso implica, en un buen diseño, mayor eficiencia estadística, por lo que un mismo número de casos permite un menor error de estimación y un intervalo de confianza más estrecho.

Un ajuste no planificado en el protocolo puede tener consecuencias imprevisibles sobre la eficiencia; puede conllevar un beneficio o, por el contrario, aumentar considerablemente el error típico de estimación de los parámetros.

Asimismo, sólo el ajuste previamente planificado (y, por tanto, enmascarado con respecto a los resultados) permite garantizar la ausencia de sesgos (fig. 12-1).



Figura 12-1 Ajuste por variables conocidas. Un profesional que desee tener controlada una variable conocida y medida sin error, utilizará un sistema que le garantice que queda igualmente distribuida en los grupos en comparación, de la misma forma que al freír un bistec nos aseguramos que cada lado se fría justo en la cantidad que deseamos.

Recuerde



Sólo el ajuste bien planificado puede garantizar dos beneficios: ausencia de sesgo y mayor eficiencia de las estimaciones.

La tabla 12-1 resume las diferentes posibilidades para el ajuste distinguiendo entre aquellas planificadas en el momento del diseño y aquellas que se introducen durante el análisis. En los siguientes puntos se comentan sucesivamente.

Restricción

La forma más sencilla de ajuste es la restricción de la población en estudio, bien sea mediante la definición de **criterios de inclusión** en el protocolo, bien sea mediante el **estudio de un subgrupo** en el momento del análisis.

Ejemplo 12.2



Se limita el estudio a casos de una cierta edad y de un hospital concreto. Ahora, estas dos variables ya no pueden ser una explicación alternativa a la relación observada entre el tratamiento X y la respuesta Y.

Su gran ventaja es la sencillez de aplicación, mientras que su inconveniente es la reducción de la población objetivo, tanto para obtener casos para el estudio como para aplicar después sus resultados.

Opción	Fase	Nombre	Ventajas	Inconvenientes
Restricción	Diseño	Criterios de inclusión	Control completo Barato Simple de diseñar Simple de analizar	Reduce la población Número de variables limitado Posible confusión residual (si las restricciones son amplias)
	Análisis	Análisis de un subgrupo		
Estudio por subgrupos	Diseño	Bloques (apareamiento)	Potencia Eficiencia	Pierde flexibilidad Coste Dispersión de casos en estratos Diferentes estratificaciones Difícil «sumarización»
	Análisis	Estratificación (apareamiento)	Sin premisas Directa Cálculo simple	
Modelado Estadístico	Diseño	Modelado	Factible con pocos casos Redondea efectos menores Permite predicciones Permite variables continuas Permite varias variables	Muchas premisas Elección del modelo Elecciones variables Interpretación Parametrización del software
	Análisis	Covarianza, regresión, otros		
Ajuste global	Diseño	Minimización	Permite considerar varias Z No reduce la población objetivo	Logística sofisticada
	Análisis	Pareja óptima		

Tabla 12-1 Opciones para el ajuste. Adaptada de Kleinbaum et al. (34)

Ejemplo 12.3



En el ejemplo anterior, si un profesional de otro centro desea utilizar los resultados de ese estudio deberá analizar con detalle si existen características diferenciales entre ambos centros que puedan comprometer los efectos observados.

Ejercicio 12.2



En un estudio experimental muy controlado, se ha logrado establecer el efecto de un tratamiento que disminuye el área necrosada tras un infarto. Los criterios de inclusión de dicho estudio exigían que el paciente fuera tratado en un plazo de tiempo muy limitado. Por esta razón, el protocolo del centro en el que usted trabaja no incluye este tratamiento para aquellos casos que superan este límite de tiempo. Usted desea tratar a este paciente. ¿Qué argumentos deben utilizarse?

Definición

Los **criterios de selección** son la aplicación de una restricción decidida en la fase de diseño del estudio.

Lectura

La declaración CONSORT (9) recomienda no distinguir entre criterios de inclusión y exclusión, y llamarlos simplemente criterios de selección o de elegibilidad.

Si una variable Z dispone de un amplio rango de variación y los criterios de inclusión son generosos, puede persistir una cierta confusión residual.

Ejemplo 12.4

Si la edad tiene un efecto claro en la respuesta, que hace muy diferente un caso con 18 años de uno con 70, limitar el estudio a pacientes de 18 a 70 es correcto, pero puede no ser suficiente.

Definición

El **estudio de un subgrupo** aplica la restricción en el momento del análisis.

Dada la ausencia de hipótesis previa, el análisis de un subgrupo debe siempre ser interpretado de forma exploratoria.

Estudio de subgrupos

La siguiente herramienta de ajuste es la definición de subgrupos, que reciben el nombre de **bloques** si se especificaron en el diseño, y **estratos** si se decidieron en el momento del análisis.

Ejemplo 12.5

En un ensayo clínico multicéntrico se especifica en el protocolo que se desea tener controlada la variable centro, y la asignación aleatoria se planifica de forma que en cada centro exista el mismo número de casos en cada grupo de tratamiento. Se dice que cada centro es un bloque y que está balanceado.

Ejemplo 12.6

En un estudio sobre el efecto de los programas de atención a domicilio en la supervivencia de las personas mayores que se quedan solas, se decide estratificar en el análisis por el nivel de actividades que realiza cada persona.

Si la variable que define los subgrupos no es modificadora del efecto, el principio de parsimonia aconseja proporcionar una **estimación única** y global del efecto que represente al efecto común de cada subgrupo.

Recuerde

Sea parsimonioso al exponer sus resultados. Si es razonable, proporcione una única medida del efecto que sirva para todos los subgrupos, sean bloques o estratos.

Comentario

No hay nada más tedioso que un informe que va presentando resultados por subgrupos cuando el efecto es muy similar en todos ellos.

Nota técnica

Una cuestión más delicada es ¿cuándo es razonable asumir que no hay interacción y que, por tanto, es lícito proporcionar una única medida? Desgraciadamente, la estadística es de poca ayuda en esta cuestión, ya que el tamaño muestral se define para el objetivo principal (p. ej., estudiar el efecto de una intervención), pero el estudio de la interacción (¿difiere este efecto por subgrupos?) requiere un mayor número de casos. Una prueba formal sólo detectaría interacción si ésta es de gran magnitud. Por ello, el consejo habitual del estadístico suele ser que las premisas del estudio conviene sustentarlas en datos previos, no sea que una sorpresa deje sin potencia el estudio.

Ejemplo 12.7

Si la presencia de alimentos en el tubo digestivo puede modificar el efecto del tratamiento, no tiene sentido combinar los resultados obtenidos en pacientes en ayunas con los de pacientes en plena digestión. Es más simple y directo definir al estado pandrial como criterio de inclusión. Si conviene, más adelante se estudiará la posible interacción o se complementará el estudio ya realizado con otro de la condición pandrial no considerada.

Ejercicio 12.3

Imagine dos estudios multicéntricos, ambos de $n = 200$, para comparar un tratamiento t con un control c . El estudio A ha sido diseñado para recoger información de cinco centros. Cada uno aporta 40 casos, 20 en cada brazo de tratamiento. El estudio B se inicia en veinte centros, de los que uno aporta 160 casos; y los otros 19 centros, los 40 restantes. Destacan 10 centros que sólo aportan un caso cada uno. ¿Cuáles de las siguientes afirmaciones son ciertas?

- El estudio A seguro que evita la posible confusión por la variable centro, ya que está equilibrado o balanceado.
- La razón entre el número de casos tratados con t y con c es constante para todos los centros del estudio A. Esto implica que la variable centro y la variable tratamiento son independientes (no son colineales), por lo que no puede haber confusión.
- En el estudio B, hay 10 centros que no pueden aportar, por sí solos, información del efecto del tratamiento y, por tanto, no podrán constituir cada uno de ellos un bloque separado.
- El estudio A, por su equilibrio en número de casos, parece que será más eficiente, proporcionando intervalos de confianza más estrechos.
- El estudio B al abarcar más centros, es más extrapolable a otros centros nuevos.
- El estudio B no es un estudio multicéntrico sólido.

El **apareamiento** es un caso extremo en el que cada bloque o estrato está únicamente compuesto por dos observaciones que comparten características que les diferencian de las otras parejas. El grado de conexión entre estas parejas, naturales o artificiales, puede variar.

Ejemplo 12.8

Los diseños con intercambio del tratamiento (*cross-over*) consisten en aplicar un tratamiento en un período y otro tratamiento en otro período, cambiando el orden de aplicación en diferentes pacientes. Cada paciente define una pareja de observaciones y constituye un bloque.

Ejemplo 12.9

En la búsqueda de causas X , a cada paciente que ha desarrollado cierta enfermedad ($Y+$) se le busca una pareja sana ($Y-$) que tenga los mismos valores en las terceras variables Z . Cada pareja constituye un estrato.

Modelado

La siguiente opción de ajuste es el modelado estadístico, del que las diferentes técnicas de regresión son el procedimiento más habitual. El protocolo puede especificar en mayor o menor grado el modelo que se utilizará para realizar el ajuste.

Ejemplo 12.10



Al comparar dos fármacos antihipertensivos (X), se puede especificar que el análisis de la respuesta (p. ej., $Y = \text{PAD}$ a las 12 semanas) se ajustará mediante un modelo de regresión lineal que incluye el valor inicial o basal (Z) de dicha presión.

Nótese que la relación entre la presión basal (Z) y final (Y) no es, ni mucho menos, el objetivo principal del estudio (un investigador experto en el tema debería conocer la magnitud de esta relación con anterioridad). El auténtico objetivo es conocer el efecto del tratamiento X en la respuesta Y ajustando por la condición Z .

Cuanto más **concreto** sea el protocolo sobre el modelo que se utilizará para realizar el ajuste, mayor será el carácter **confirmatorio** del estudio.

Ejemplo 12.11



Un ejemplo muy sencillo de modelo que se especifica completamente en el protocolo es el llamado estudio del cambio o diferencia entre los valores finales y los iniciales o basales.

Nota técnica



Este cambio, evolución o diferencia final-basal se corresponde (Senn, 67) con un modelo de regresión lineal en que el coeficiente («pendiente») de la relación entre la PAD a las 12 semanas y la PAD inicial vale 1.

Los paquetes de programas estadísticos dan una falsa sensación de facilidad en el uso de estos modelos de regresión. La selección de variables a incluir no es un tema trivial, especialmente en presencia de colinealidades o correlaciones entre las variables a introducir. Dejar el ajuste para el final es muy desaconsejable, ya que el margen de maniobra es siempre muy inferior.

Recuerde

Si no quiere encontrarse con sorpresas al final del estudio, considere al inicio el método que usará para controlar las terceras variables.

Comentario

La práctica habitual de introducir en el modelo las variables que han resultado previamente significativas es muy peligrosa si se desea interpretar causalmente los resultados. Quizá podría ser correcta en un estudio exploratorio, que terminara proponiendo hipótesis que deberían ser ratificadas en estudios posteriores. Pero aun en esta situación, la colinealidad entre las variables introducidas generará problemas en la calidad de la estimación y en la interpretación de los resultados, como se comenta en el apartado «Papel del azar» (p. 302).

Ejercicio 12.4

En el estudio del efecto de un tratamiento X en la PAD de las 12 semanas, se ha realizado un modelo de regresión que incluía la PAD en la semana inicial. Los estudios que se comentan a continuación han llegado al modelo por diferentes caminos.

Ordénelos de mayor a menor credibilidad que le ofrezcan las conclusiones.

Y remarque cuáles representan un salto importante en credibilidad.

- a) Tal y como especificaba el protocolo, la respuesta estudiada ha sido el cambio o diferencia entre la PAD final y la inicial.
- b) Con la PAD12 como respuesta, se ha introducido la PAD0 en el modelo con un coeficiente que ya venía especificado en el protocolo.
- c) Con la PAD12 como respuesta y tal como especificaba el protocolo, se ha introducido la PAD0 en el modelo y se ha dejado al programa que escoja el coeficiente por el procedimiento habitual (que minimiza el error de predicción de la respuesta).
- d) Como *a* pero sin que figurara en el protocolo.
- e) Como *b* pero sin que figurara en el protocolo.
- f) Como *c* pero sin que figurara en el protocolo.
- g) Tal como decía el protocolo, entre un montón de modelos posibles se ha seleccionado aquel en el que era más significativa la variable X .

Ajuste global (minimización)

Finalmente, el cuarto método de ajuste considera simultáneamente varias variables en el ajuste. En un diseño experimental con aleatorización, persigue mejorar el equilibrio proporcionado por esta asignación al azar.

Ejemplo 12.12



En el estudio de la PAD se quiere tener controladas la edad, el centro, el estrés y el sedentarismo. En el momento de introducir un nuevo paciente en el estudio, el investigador proporciona los datos de estas variables a un centro de aleatorización que, sin romper el enmascaramiento del investigador, asigna el nuevo caso al grupo que consigue un mayor equilibrio en estas variables.

La minimización permite tener equilibrados los grupos sin necesidad de recurrir a criterios de inclusión más restrictivos ni a la definición de muchos subgrupos, lo que podría comprometer el reclutamiento de casos para un estudio.

Nota técnica



Los algoritmos habituales de **minimización** suelen ponderar por separado las variables consideradas en el proceso y persiguen un equilibrio en cada una de ellas.

La minimización puede combinarse también con la asignación aleatoria, manteniendo el enmascaramiento de los investigadores y de los pacientes.

Ejemplo 12.13



Un investigador interesado en ajustar por edad, género, estrés y sedentarismo (estas dos últimas medidas en escalas de 0 a 10) ha determinado en estudios previos que la relación de estas 4 variables con la PAD es la siguiente: cada año supone 1 mmHg más, los hombres tienen 8 mmHg más; cada punto de sedentarismo son 2 mmHg más y cada punto de estrés son 3 mmHg más. Así, un índice global combina estos valores en cada paciente y el próximo se asigna al grupo que hace más similares las medias de este índice en los dos grupos. Es óptimo, no para cada variable Z por separado, sino para el conjunto de todas ellas. Es decir, pretende compensar los desequilibrios de unas variables con los de otras.

También puede usarse en los estudios observacionales.

Ejemplo 12.14

En la misma búsqueda anterior de causas X , y de la misma forma que antes, a cada paciente que ha desarrollado cierta enfermedad ($Y+$) se le busca una pareja sana ($Y-$). La diferencia ahora es que la pareja debe tener un valor próximo a los valores de las terceras variables Z , sea en cada una de ellas, como en el ejemplo 12.12 o en su combinación en un índice como en el ejemplo 12.13.

Técnicas de ajuste frente a la confusión y a la modificación del efecto

Las variables Z se incluyen en el modelo porque son condiciones previas sobre las que se estudia el efecto de la hipotética causa. Pero, por cuestiones logísticas, no se pueden considerar todos los atributos. ¿Cuáles seleccionar? La respuesta es: aquellos que puedan tener sus efectos confundidos con el tratamiento o bien aquellos que lo modifiquen.

La experiencia y conocimiento del investigador deben permitirle saber, a priori, qué variables Z predicen la evolución Y (y podrían, por tanto, acabar teniendo sus efectos confundidos con la intervención X) y qué variables Z interactúan con el tratamiento X (y, por tanto, modifican el efecto de X en Y).

Recuerde

Concentre la atención y los riesgos estadísticos en el auténtico objetivo del estudio: la relación entre X e Y dadas las condiciones Z . (Las relaciones de Z con X e Y son una premisa, no un objetivo.)

Ejercicio 12.5

Sean Z el centro, X el tratamiento e Y la evolución, ¿qué confunde a qué? Es decir, ¿qué tablas de las siguientes (tabla 12-2) tiene interés estudiar y presentar?

Si una variable Z es **modificadora del efecto**, es decir, si existe interacción, es complicado explicar en primer lugar un efecto común del tratamiento que debe matizarse en segundo tiempo según el valor de la variable Z . Es más sencillo explicar cuál es el efecto del tratamiento en cada nivel de la variable Z .

Z+	Y+	Y-	Z-	Y+	Y-		Y+	Y-
X+	6	5	X+	180	6	X+	186	11
X-	45	150	X-	150	20	X-	195	170
OR = 4			OR = 4			OR = 14,74		
Y+	Z+	Z-	Y-	Z+	Z-		Z+	Z-
X+	6	180	X+	5	6	X+	11	186
X-	45	150	X-	150	20	X-	195	170
OR = 0,11			OR = 0,11			OR = 0,05		
X+	Z+	Z-	X-	Z+	Z-		Z+	Z-
Y+	6	180	Y+	45	150	Y+	51	330
Y-	5	6	Y-	150	20	Y-	155	26
OR = 0,04			OR = 0,04			OR = 0,03		

Tabla 12-2 ¿Qué tabla tiene más sentido estudiar? ¿Qué información aporta cada tabla?

Ejercicio 12.6



Suponga los siguientes resultados sobre el efecto del alcohol, modificado por el género, en las habilidades motoras. Aunque los números (imaginarios) que se presentan son idénticos, ¿de qué forma es más fácil interpretarlos?

- Dos vasos de vino reducen un 40% la rapidez de reflejos, siendo esta reducción un 25% superior en las mujeres y un 25% inferior en los hombres.
- Dos vasos de vino reducen la rapidez de reflejos un 50% en las mujeres y un 30% en los hombres.

Si Z es modificadora del efecto, es preciso estimar más parámetros, lo que dificulta la interpretación y requiere más casos. Si una de las categorías de Z es especialmente relevante tiene sentido concentrar el estudio en esa categoría y dejar para más adelante el resto.

Ejemplo 12.15



Se sospecha que cierto fármaco, que se usa muy poco, puede modificar el efecto de nuestro producto X en estudio. Será más práctico poner como criterio de exclusión la toma de ese fármaco y dejar para más adelante el estudio de la posible interacción.

Interpretación de resultados ajustados

Se acaba de proponer el ajuste para estimar el efecto causal de una variable X habiendo eliminado el posible efecto de terceras variables Z . Veamos ahora cómo se deben interpretar estos resultados ajustados si la **variable Z es posterior** a la causa X .

Ejercicio 12.7



Se sospecha que cierto gen ($X_1 = \text{sí}$, $X_2 = \text{no}$) provoca la aparición de enfermedades cardiocirculatorias ($Y_1 = \text{sí}$, $Y_2 = \text{no}$). Se sabe que las hiperlipemias ($Z_1 = \text{sí}$, $Z_2 = \text{no}$) originan también las enfermedades (Y).

Además, se sospecha que el gen X provoca las hiperlipemias Z . El orden natural de aparición de estas variables es X, Z, Y . Suponga que se ha decidido estudiar la relación entre el gen X y las enfermedades cardiocirculatorias Y a partir de los pacientes de un servicio (Z_1) especializado en el tratamiento de pacientes hiperlipídicos. ¿A qué conclusión llegarían investigadores de este servicio que dispusieran únicamente de los datos de la subtabla Z_1 (tabla 12-3)?

Z_1	Y_1	Y_2	Z_2	Y_1	Y_2		Y_1	Y_2
X_1	503	185	X_1	37	151	X_1	540	336
X_2	151	56	X_2	83	337	X_2	234	393
OR = 1,01 IC _{95%} 0,7, 1,4			OR = 0,99 IC _{95%} 0,6, 1,5			OR = 2,7 IC _{95%} 2,2, 3,3		

Tabla 12-3 Ejemplo 12-7, en el que la variable Z es posterior a la variable X

Ejercicio 12.8



Suponga que otros investigadores hacen el mismo estudio en la población general representada en la subtabla de la derecha. ¿A qué conclusión llegarían con estos datos?

Ejercicio 12.9

Interprete el conjunto de los ejercicios 12.7 y 12.8.

Recuerde



Si las variables Z son posteriores a la causa X en estudio, la interpretación del ajuste por ellas es muy delicada.

Conviene recordar que, en ocasiones, el objetivo es realizar una **predicción**, no una intervención.

Ejemplo 12.16

Recuperemos el ejercicio 11.3: la probabilidad de tener un hijo con el síndrome de Down es aproximadamente de 1 cada 600 embarazos en la población general, pero sube hasta alrededor de 1 cada 60 en las gestantes mayores de 40 años. Ahora bien, ¿y la edad del padre? La elevada correlación entre la edad de los padres puede originar que se estén confundiendo sus efectos y si, por ejemplo, se deseara conocer el efecto causal de la edad del padre, se debería ajustar por la edad de la madre. Ya se comentó en el ejercicio 11.10 que la única utilidad práctica (a nivel inmediato) de conocer el efecto de la edad del padre ajustado por la de la madre, sería poder decir a las futuras madres que se busquen un padre joven pero sin adelantar su propia edad de maternidad. En cambio, si la finalidad de establecer la relación entre la edad de los padres y el riesgo de síndrome de Down es poder seleccionar aquellos embarazos con elevada probabilidad de que el feto presente esta enfermedad, entonces el objetivo es predecir, y no es necesario buscar una relación causal por medio del ajuste.

Si el objetivo es realizar una **predicción**, se debe seleccionar, para hacer la predicción, aquella variable que sea más temprana en el tiempo, más fiable (en el sentido de obtener valores más similares en sucesivas determinaciones) e incluso más fácil de interpretar.

Comentario

*Si el objetivo **no** es intervenir, y desea seleccionar un conjunto pequeño de variables que permitan realizar una predicción o pronóstico, los métodos habituales de selección de variables disponibles en los programas de regresión multivariante pueden ser de gran ayuda para complementar los criterios clínicos.*

Recuerde

Para realizar una predicción puede escoger aquella variable «más cómoda».

Ejercicio 12.10

En el ejemplo del síndrome de Down, ¿qué ventajas puede tener la edad de la madre respecto a la del padre?

Volviendo al efecto causal, el atractivo del ajuste es que permite estudiar el efecto de la variable X, a la vez que se dejan igual, sin modificar, las variables por las que se ha ajustado. Esto es lo que pretende significar la expresión «el efecto de X **independientemente de Z**». Por ello, la decisión de ajustar o no por una tercera variable Z, que puede estar relacionada con el tratamiento X en estudio, depende de la capacidad futura de intervención: ¿se puede intervenir sobre X sin alterar simultáneamente Z? O por el contrario, ¿es más factible conseguir cambios simultáneamente en X y en Z que sólo en X?

Comentario



Recuerde el diseño experimental para distinguir el efecto de la edad de ambos progenitores: dos grupos de madres (jóvenes y adultas) se cruzan, al azar, con dos grupos de padres (jóvenes y adultos). Ahora, la edad de ambos progenitores es independiente, pero éste es un resultado experimental que no reproduce ninguna realidad concreta. La pregunta que contesta este diseño no tiene interés práctico en nuestro entorno cultural.

Ejemplo 12.17



Al estudiar el efecto del tabaco, se puede desear ajustar por el nivel de alcohol. Lo que será muy interesante si se dispone de una intervención sanitaria que consigue disminuir el hábito tabáquico sin alterar el hábito alcohólico. Pero si la intervención sanitaria consigue, al mismo coste, disminuir ambos hábitos tóxicos, tendrá más interés práctico conocer el efecto simultáneo de ambos.

Por supuesto, este principio de «caja negra» paga el precio de no satisfacer la curiosidad racional: «Al final de todo, ¿cuál es la auténtica causa de...?».

Ejercicio 12.11



Suponga que los expertos afirmen que dejar el tabaco a la vez que se modera el consumo de alcohol es prácticamente imposible y, por tanto, que no puede intervenir sobre ambas variables a la vez. ¿Tiene sentido estimar los efectos de una de estas variables a nivel fijo de la otra? (Es decir, ¿ajustará en su estimación?)

Ejercicio 12.12

Suponga que cierta teoría psicológica establece que es más fácil cambiar simultáneamente los hábitos sobre dieta y ejercicio que cambiar cada hábito por separado. ¿Conviene estudiar el efecto de la dieta a nivel fijo de ejercicio?

Ejercicio 12.13

¿Qué variables hay que tener controladas? Conteste cierto o falso a las siguientes afirmaciones.

- a) Hay que tener presente todo lo dicho a propósito del ajuste, de la confusión y de la modificación del efecto.
- b) No hay que ajustar por otras consecuencias Y' de la causa en estudio X .
- c) Sí que puede ajustarse por los atributos Z (características poseídas por los pacientes).
- d) Debe ajustarse por aquellos atributos Z que, si no quedan controlados, pueden comprometer, por su conocida relación con la respuesta, la inferencia causal.
- e) Se puede ajustar por otras concausas X' si se cree que la intervención sobre X podrá hacerse «independientemente» del valor de la concausa X' .
- f) Si es más fácil intervenir simultáneamente sobre X y sobre la con-causa X' que sobre X independientemente de X' , tiene más sentido no ajustar por X' .
- g) Si se puede intervenir sobre X sin alterar X' entonces puede tener sentido ajustar por X' .

Ejercicio 12.14

El servicio de infecciones de cierto centro dispone de un centenar de casos de sida (Y) y desean averiguar el posible efecto del nivel de promiscuidad (X) en la aparición de la enfermedad. Para ello, desean obtener unos controles sin sida ajustando por las siguientes características Z : preferencias sexuales, uso de drogas por vía parenteral, empleo de preservativo, edad y nivel socioeconómico. De todas ellas, la única de la que se sospecha que pueda ser modificadora del efecto de la promiscuidad es el uso del preservativo. El servicio de traumatología del mismo centro tiene recogidas, con el permiso de los pacientes, todas estas variables en un archivo informatizado. Diga, de cada frase, si le parece correcta o no.

- a) Tiene sentido hacer dos grupos según el uso del preservativo en los que se estimará, por separado, el efecto de la promiscuidad.

Ejercicio 12.14 (Cont.)

- b) Dentro de cada grupo de uso de preservativo, se busca un control de forma que sus diferencias sean mínimas en el resto de variables.
- c) Para ello, se recurre a la minimización: con la ayuda de un algoritmo informático, se forman las parejas de manera que las diferencias entre los valores de las restantes variables Z sea mínima.
- d) El criterio de apareamiento se ha realizado sin considerar (incluso, «a ciegas») el valor de la causa en estudio (el nivel de promiscuidad).

Ejercicio adicional

Busque o proponga un ejemplo de una intervención de salud para modificar hábitos de vida (tabaquismo, consumo de alcohol, dieta, sedentarismo, estrés, etc.). Defina el papel de cada variable (X , Y , Z). Modifique el diseño de forma que cambie el papel de las variables en el modelo.

Ejercicio 12.15

Una discusión habitual es si conviene luchar contra la presión arterial sistólica o contra la diastólica. Unos trabajos incluyen una variable y otros a la otra. ¿Usted qué opina?

Recuerde

*Si el objetivo es intervenir mediante cambios en X , antes de ajustar por Z pregúntese si puede **modificar X a nivel fijo de Z** .*

Ejercicio adicional

A partir de un EC real (o simulado), invente preguntas sobre los criterios de inclusión para que sus compañeros respondan si la razón para figurar como tales es por razones estadísticas (posibles confusoras o posibles modificadoras del efecto), éticas o logísticas.

Las terceras variables no observables

El ajuste puede aplicarse sobre las variables Z que influyen en la respuesta siempre que su relación sea conocida y puedan ser medidas sin error. Sin embargo, en la respuesta pueden influir también otras variables no conocidas o con grandes dificultades en su medición. Llamémoslas V . Si estas variables V también estuvieran relacionadas con el tratamiento en estudio, podrían confundirse sus efectos.

Ejemplo 12.18



Un ejemplo podría ser el estrés, que aunque se sospecha su influencia, es muy difícil de medir y cuantificar.

Como estas variables V , por definición, no son observables, no se puede recurrir al ajuste o control. La única solución es asumir alguna **premisa**: alguna suposición sobre su existencia o sobre su comportamiento.

Lectura



Stone (68) menciona cuatro premisas (tabla 12-4) para poder inferir causalidad:

1. La primera y más simple es la **suficiencia del modelo**, que consiste en suponer que todas las variables importantes ya han sido consideradas en el ajuste. Y, por tanto, que las variables V no existen. Para ser verosímil, exige una gran predictibilidad de la respuesta, por lo que su uso está limitado a ciencias muy deterministas como, por ejemplo, la Física.
2. En Farmacología, y en aquellos entornos en los que es éticamente aceptable la comparación experimental entre los grupos, suele recurrirse a la **asignación aleatoria** de las causas en comparación. Esta randomización garantiza que ambos grupos vienen de la misma población y tienen, por tanto, idéntica distribución poblacional de todas las variables V . Se trata de la única «premisa» que un buen diseño puede establecer.
3. La tercera premisa, usual en Epidemiología, asume que las variables V están igualmente distribuidas o repartidas entre los grupos que se comparan (V y X son independientes, no hay colinealidad). Por tanto, los efectos de V y de X **no** están **confundidos**. Es decir: se asume que las variables que influyen en la respuesta o bien son observables (Z) (y ajustaremos por ellas); o bien, si no son observables (V), son independientes del tratamiento.
4. Finalmente, dado que puede darse aquella situación en la que la influencia conjunta en la respuesta de todas las variables V resulte equilibrada a pesar de que existan desequilibrios particulares en algunas de ellas, se ha propuesto basar la interpretación de los estudios observacionales en la premisa de **asignación ignorante** del tratamiento, que supone que la respuesta potencial, previamente a la administración del tratamiento, es la misma en ambos grupos.

Nombre	Premisa sobre las variables V	Entorno usual
Suficiencia del modelo	No hay variables V	Ciencias deterministas
Asignación aleatoria	Las V están equilibradas en los grupos	Farmacología
No confusión	Las que hay son independientes del tratamiento	Epidemiología
Asignación ignorante	Su influencia conjunta está equilibrada	Ciencias sociales

Tabla 12-4 Premisas habituales sobre las variables no observables (Stone, 68)

Recuerde



Entre todas las condiciones que equilibran las variables V no observables destaca la **asignación aleatoria**, pues es la única que está en manos del investigador.

Un estudio que no controle las variables desconocidas V es *caduco* en el sentido de que un nuevo conocimiento científico que convierta una variable desconocida, V, en conocida, Z, obliga a repetirlo.

Ejercicio adicional



Proponga ejemplos basados en su experiencia para que sus compañeros discutan qué terceras variables podrían influir en los resultados y si éstas son observables o no. A su vez, deben indicar qué técnica de ajuste proponen para controlarla.

Papel del azar

La inferencia estadística puede recurrir al azar en dos ocasiones, para seleccionar a las unidades o para asignar los tratamientos. Como se explica a continuación, en el primer caso se persigue representatividad, y en el segundo, comparabilidad (fig. 12-2).

Obtención al azar: representatividad

Ya se dijo que, si los elementos de la muestra eran seleccionados al azar, la teoría de probabilidad permite cuantificar (error estándar, intervalos de confianza) la incertidumbre al inferir al conjunto de toda la población los resultados obtenidos en la muestra.

Ejemplo 12.19



Al realizar la planificación sanitaria de una comunidad se recurre a una muestra aleatoria para conocer el estado de salud de sus ciudadanos.

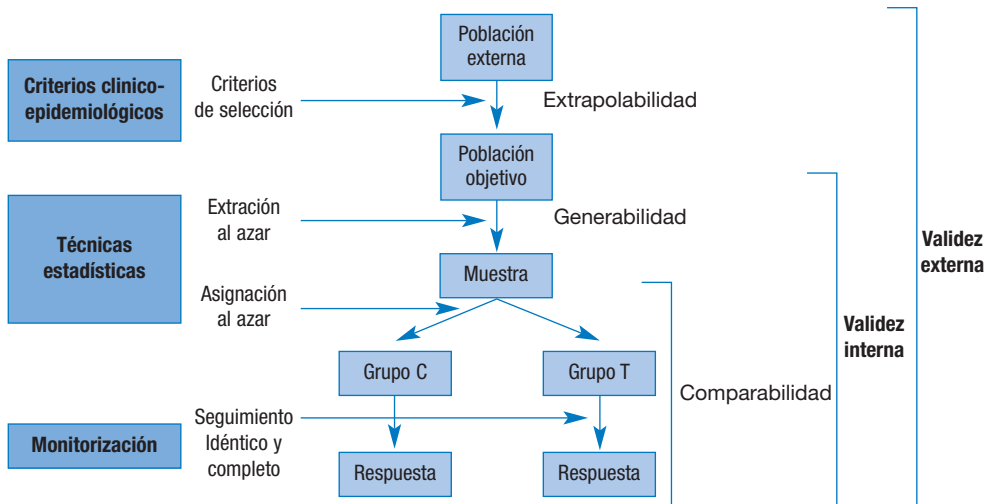


Figura 12-2 Dos aportaciones del azar a la validez: generabilidad y comparabilidad.

Recuerde



La teoría de probabilidad permite cuantificar las oscilaciones de los resultados debidas al azar del muestreo.

Ejercicio 12.16



Para realizar el estudio del ejemplo 12.19, a partir de un listado de todos los elementos de la población, se seleccionan al azar una serie de individuos en los que se evalúa su nivel de salud. Dado que es un derecho no participar, ¿qué consecuencias puede tener que algunos casos rehúsen?

Recuerde



La obtención al azar de la muestra persigue la representatividad o generabilidad de los resultados.

Asignación al azar (randomización): comparabilidad

Una vez se han seleccionado de la población aquellos individuos que formarán parte del estudio, el proceso de asignación aleatoria del tratamiento genera dos muestras de esta misma población. Dado que ambas muestras provienen, de manera aleatoria, de la misma población, la distribución de todas aquellas variables no intervenidas será la misma a nivel poblacional.

Dado que, a nivel poblacional, todas las variables están igualmente distribuidas en los dos grupos, no existe ninguna variable con la que se puedan confundir los efectos del tratamiento, por lo que la asignación aleatoria del tratamiento ofrece una estimación insesgada del efecto causal.

Recuerde



La asignación al azar del tratamiento persigue la *comparabilidad* de los grupos (fig. 12-3).



Figura 12-3 Papel del azar. Un profesional que desee tener *controlado* un elevado número de variables con pequeños efectos en la respuesta puede recurrir al azar para conseguir un reparto suficientemente similar del conjunto de todas ellas, de la misma forma que al freír unos champiñones introducimos entropía, azar, ruido, para que se hagan homogéneamente. ¡Y así resulta!

Ejemplo 12.20



Imagínese una población con dos unidades: Montserrat Martí y José López (Rubin, 63). En el ejemplo de la PAD y los bloqueadores β , el efecto causal, en la unidad Montserrat Martí, de los bloqueadores β más dieta frente a la dieta sola, sería la PAD cuando se le asigna BB+D menos la PAD cuando se le aconseja D. En símbolos:

efecto causal en Montserrat Martí

$$e_{\text{Martí}} = \text{PAD}_{\text{BB+D, Martí}} - \text{PAD}_{\text{D, Martí}}$$

efecto causal en José López

$$e_{\text{López}} = \text{PAD}_{\text{BB+D, López}} - \text{PAD}_{\text{D, López}}$$

Y el efecto causal promedio en esta población de 2 unidades sería:

$$\begin{aligned} E[\text{PAD}_{\text{BB+D}} - \text{PAD}_{\text{D}}] &= 1/2 (e_{\text{Martí}} + e_{\text{López}}) = \\ &= 1/2 ([\text{PAD}_{\text{BB+D, Martí}} - \text{PAD}_{\text{D, Martí}}] + [\text{PAD}_{\text{BB+D, López}} - \text{PAD}_{\text{D, López}}]) \end{aligned}$$

Ejemplo 12.20 (Cont.)

Ahora bien, ya se vio que $e_{\text{Martí}}$ y $e_{\text{López}}$ no son observables a la vez y en las mismas condiciones. En un diseño experimental típico sólo se observaría o bien la muestra en la que se asigna el tratamiento a la unidad «Martí» y el control al individuo «López»:

$$PAD_{BB+D, \text{Martí}} - PAD_{D, \text{López}}$$

o bien la muestra con la situación contraria:

$$PAD_{BB+D, \text{López}} - PAD_{D, \text{Martí}}$$

Como estas dos muestras constituyen todas las posibles realizaciones muestrales, la esperanza del conjunto del experimento será:

$$\frac{1}{2} ([PAD_{BB+D, \text{Martí}} - PAD_{D, \text{López}}] + [PAD_{BB+D, \text{López}} - PAD_{D, \text{Martí}}])$$

lo que por simple reordenación de sus términos

$$= \frac{1}{2} ([PAD_{BB+D, \text{Martí}} - PAD_{D, \text{Martí}}] + [PAD_{BB+D, \text{López}} - PAD_{D, \text{López}}]) =$$

$$= \frac{1}{2} (e_{\text{Martí}} + e_{\text{López}}) = E[PAD_{BB+D} - PAD_D]$$

conduce al efecto causal promedio.

El conjunto de todas las posibles realizaciones muestrales se encuentra centrado alrededor del auténtico efecto causal, que es el valor que se desea conocer. Por ello, la asignación aleatoria permite una estimación insesgada.

Ahora bien, un buen investigador no tiene suficiente con el hecho de que la ciencia avance porque el conjunto de los estudios apuntan en la dirección correcta. Él quiere que su estudio concreto acierte en sus estimaciones (fig. 12-4).



Figura 12-4 Límites del azar. Si una variable importante debe quedar equilibrada entre los grupos, el azar no lo garantiza, especialmente si la muestra es pequeña. El azar que le sirve al cocinero para *equilibrar* los champiñones, no funciona con los bistecs. Puede argumentarse que, si se dispusiera de muchos bistecs, la mitad de ellos se harían por su cara proximal y la otra mitad por la distal, llegando a un equilibrio a «largo plazo». Ciertamente, habría *equilibrio* para el conjunto de los comensales. Pero cada uno de ellos se comería un bistec que tendría una cara cruda y la otra hecha. De la misma manera, si un investigador quiere garantizar que en su estudio concreto una variable importante quede equilibrada, no debe descansar en el azar.

Ejemplo

En el **ejemplo 12.20**, un investigador concreto asignaría cada observación a un tratamiento, y en función del azar sólo observaría o bien el tratamiento en estudio en la primera unidad y el control en la segunda ($PAD_{BB+D, Martí}$ y $PAD_{D, López}$), o bien la situación contraria ($PAD_{BB+D, López}$ y $PAD_{D, Martí}$). En ambas posibilidades, el efecto del tratamiento estaría «confundido» con cualquier característica que diferencie los casos tratados de los controles. En este ejemplo, el género podría ser una explicación alternativa a posibles diferencias entre los casos.

Recuerde

*Aunque la asignación aleatoria permite estimaciones insesgadas a lo largo de todos los experimentos, **no garantiza** que cada posible experimento esté equilibrado para toda tercera variable.*

Ejemplo 12.21

(Prestado de S. Senn): A nivel global puede ser razonable aceptar que en 1 de cada 10.000.000.000 vuelos se averíen todos los motores, pero a los pasajeros de ese vuelo...

¿Qué puede hacerse? Muy sencillo: no dejar en manos del azar aquellas terceras variables que, si resultaran desequilibradas, pudieran comprometer las conclusiones del estudio. Nótese que las variables no observables V nunca comprometerán la credibilidad del estudio (porque no puede observarse su desequilibrio). Pero las variables Z sí que pueden observarse. Si una de ellas puede comprometer la credibilidad de los resultados, no conviene dejarla en manos del azar.

Ejemplo 12.22

Si el estadio inicial de la enfermedad es una variable crucial, que determina claramente la evolución de los pacientes, conviene controlarla por los métodos vistos en el ajuste de variables conocidas Z .

Contraejemplo 12.23

Si la situación personal, familiar y profesional es un conjunto de variables que influyen (un poco cada una de ellas) en la presión arterial, no conviene ajustar por todas ellas, ya que se haría inviable el estudio, al menos mientras no se defina una nueva variable que resuma todos sus efectos (fig. 12-5).



Figura 12-5 Límites del ajuste. Pretender controlar cada pequeña variable haría inviable el estudio. Si se definen bloques de edad, género y color de los ojos en cada centro, siempre faltarán casos en alguno de los bloques y nunca se podrá cerrar el estudio.

Recuerde



Las terceras variables *observables* Z se controlan con las técnicas de ajuste, y las *no observables* V con la asignación al azar.

Ejercicio 12.17



De las siguientes expresiones, remarque aquellas que son ciertas.

- Si realizo al azar la asignación del tratamiento, las dos muestras, que vienen de la misma población, tienen una idéntica distribución poblacional.
- Si realizo al azar la asignación del tratamiento, las dos muestras, que vienen de la misma población, toman idénticos valores en ambas muestras.
- Si realizo al azar la asignación del tratamiento, la estimación es insesgada.
- Si realizo al azar la asignación del tratamiento, no debe preocuparme ningún posible desequilibrio entre las muestras, ya que la estimación es insesgada.
- Si mediante las técnicas de ajuste controlo todas las terceras variables Z , la asignación al azar ya no aporta nada al diseño.
- Si se acepta que la ciencia empieza por descubrir y conocer las variables más importantes, las técnicas de ajuste (que controlan las variables conocidas Z) tienen un papel más relevante que la asignación al azar (que controlan a las desconocidas V).

Consideraciones éticas y papel del científico

Los principios éticos que rigen el diseño de los ensayos clínicos (autonomía, beneficencia y no maleficencia, equidad o justicia distributiva y confidencialidad), ¿cómo afectan al método de asignación al azar?

Nótese que de la «asignación al azar», el término que tiene implicaciones éticas es «**asignación**», pero **no** «**azar**». Si ambas pautas de tratamiento son éticamente asignables a los pacientes y ellos renuncian mediante el consentimiento informado a sus derechos de autonomía, ¿qué principio ético impide que sea el azar quien decida? Puede ser estéticamente feo, pero no es éticamente discutible.

Recuerde



«Asignar al azar» tiene implicaciones que pueden ser estéticamente feas, pero no éticamente reprobables.

Ejemplo 12.24



El famoso ensayo clínico sobre la estreptomicina, realizado en los años cuarenta por el Medical Research Council (69), asignaba los pacientes al azar al grupo con antibiótico más reposo o bien al grupo tratado únicamente con reposo. Con los conocimientos de hoy en día, sobrecoge pensar qué procedimiento aleatorio pudo decidir sobre una evolución que implicaba la vida de esas personas. Dos consideraciones apoyan, una a la asignación, y la otra, a su aleatoriedad. La primera, común a casi todos los estudios, es que, con los conocimientos de aquel momento, los (posibles) efectos positivos se compensaban con los (posibles) efectos negativos y, en ausencia de conocimiento sólido, ambos tratamientos son éticamente asignables («equipoise»). La segunda consideración es un aspecto poco conocido de este ensayo: el Medical Research Council sólo disponía de estreptomicina para tratar a unos 150 enfermos, ante 300 candidatos. Y propusieron asignarlo al azar para garantizar el principio ético de equidad.

Comentario



El compromiso ético del ensayo clínico no viene impuesto tanto por la aleatoriedad como por la elección del tratamiento control (asignabilidad) y de las medicaciones de rescate.

Ejercicio 12.18



¿Es ético el uso de placebo?

Lectura



¿Hasta qué punto es lícito disminuir el procedimiento terapéutico habitual en un estudio? Por supuesto, ésta es una pregunta que debe ser contestada por los comités de ética y por los profesionales que consensuaron el protocolo que define dicho «procedimiento terapéutico habitual». En cualquier caso, conviene recordar que no sólo están en juego los derechos de los pacientes actuales con la enfermedad ya desarrollada, sino también los de los pacientes futuros. Por ello, suele decirse que, en situación de duda, la actitud más ética es empezar inmediatamente un estudio aleatorizado bien diseñado. Rawls (70) sugiere considerar la «situación de partida»: antes de saber el resultado del reparto de papeles que la vida nos depara, es decir, independientemente de si somos el paciente actual, el futuro o el investigador, ¿en qué sociedad preferimos vivir: en una que haga ensayos clínicos aleatorizados y aprenda rápidamente o en una que no los haga?

Nuestra postura es que la actitud más ética consiste en empezar lo antes posible un estudio bien diseñado («asignar al azar el primer paciente») que permita que el máximo número de pacientes se beneficien de sus resultados (Cobo, 71).

Recuerde



Las obligaciones éticas de los investigadores y de los estadísticos son:

- 1. Garantizar que el diseño permitirá contestar el objetivo del estudio, con el mínimo «coste» para los pacientes actuales.*
- 2. Iniciar, lo antes posible, dicho estudio.*
- 3. Ofrecer credibilidad en las conclusiones alcanzadas.*

Criterios para clasificar un estudio

En este punto se definen dos claves para clasificar los estudios: la secuencia temporal y el plan de muestreo. A continuación se explican las implicaciones que tiene en el diseño considerar la eficiencia estadística y la calidad de la recogida de variables.

Según la secuencia temporal

En el capítulo anterior se distinguió entre preguntas prospectivas (¿efecto de una causa?) y retrospectivas (¿causas de un efecto?). Ahora se verá que el diseño de recogida de datos puede también plantearse de forma prospectiva o retrospectiva. E incluso simultánea o transversal.

Definición

En un diseño **transversal**, la respuesta y la causa en estudio se miden en el mismo momento de tiempo.

Ejemplo 12.25

Se investiga en una serie de unidades, en el día de hoy, si tienen el sida (Y) y si son promiscuos (X). La limitación de este diseño proviene de que el sida de hoy puede ser una consecuencia de la promiscuidad de ayer, pero la (¿menor?) promiscuidad de hoy puede ser una consecuencia del sida de ayer.

En un estudio transversal, al realizar las mediciones en el mismo instante no se puede saber qué es lo que mueve qué: ¿qué hace de carro y qué hace de buey?

Recuerde

*Un estudio **transversal** no permite dilucidar el orden **causal**. En los estudios transversales debe poderse asumir que la causa X en estudio ha permanecido inmutable desde antes de que se desarrollara la consecuencia Y.*

Comentario

Dado que esta invariabilidad sólo suele ser razonable para atributos inmutables, como el género, la repercusión científica de estos diseños es limitada.

Ejemplo 12.26

Puede ser razonable asumir que el género de una persona ha permanecido invariable desde antes de desarrollar la enfermedad. Pero este carácter de invariabilidad es el que impide luego intervenir sobre el género («hágase usted mujer y disminuirá sus probabilidades de...»).

Recuerde

La mayor limitación de los estudios transversales es que no contemplan el lapso de tiempo necesario para que la causa en estudio X ejerza su influencia.

En un estudio longitudinal, la hipotética causa en estudio X está situada en el tiempo con anterioridad a la respuesta Y. Ahora bien, esta información puede recogerse en diferente orden.

Definición

En un diseño **retrospectivo**, tras observar la respuesta Y, se investiga en el pasado sobre los valores de la hipotética causa X.

Ejemplo 12.27

La pregunta sobre la promiscuidad (X) puede hacer referencia, no a la situación actual, sino a la situación en un pasado concreto, definido con un margen de tiempo suficiente.

Definición

En un diseño **prospectivo**, el registro de los datos sigue el orden natural de aparición de los acontecimientos: se observa primero la hipotética causa X y se espera en el tiempo a que acaezca la respuesta Y.

Ejemplo 12.28

Se selecciona a todos los pacientes actuales con sida de un cierto servicio que tengan normalizadas sus cifras de CD4 y se les mide el grado de promiscuidad actual y el uso del preservativo. Al cabo de un cierto tiempo, se observa cuántos de ellos se han reinfestado por una nueva cepa del virus. El proceso del diagnóstico debe ser idéntico en ambos grupos y, a ser posible, el evaluador de la respuesta en estudio desconocerá los valores de promiscuidad y uso de preservativos.

Ejercicio 12.19

En el ejemplo anterior, no hay dudas de que la reinfección es la respuesta Y en estudio. Promiscuidad y preservativo, ¿qué papel desempeñan: son X o Z?

Ejercicio 12.20

En el hospital WW disponen de una lista de 100 pacientes que han padecido infarto. A cada uno de ellos se le busca un control de su misma edad, centro, obesidad y hábitos dietéticos y tóxicos. Toman, en todos ellos, una muestra de sangre y determinan la concentración del metabolito X. La comparación mediante la *t* de Student de datos apareados resulta significativa y concluyen que este metabolito es un factor que provoca el infarto. ¿Qué opina de esta conclusión? (Pista: piense en términos de causa (X) y efecto (Y) y en la secuencia temporal entre ambos): ¿cuándo se obtiene el metabolito X?

Un ensayo clínico es un diseño prospectivo, ya que se asigna la causa en estudio (X) y se espera en el tiempo para observar la respuesta (Y).

Recuerde



Los ensayos clínicos son estudios prospectivos.

Además, el ensayo clínico es un estudio confirmatorio, ya que parte de una hipótesis previa especificada en el protocolo.

Recuerde



La especificación de la hipótesis en el protocolo permite el carácter confirmatorio de los ensayos clínicos.

Ejercicio 12.21



Se han realizado 4 ensayos clínicos idénticos para valorar el efecto de un fármaco en el descenso de la presión arterial, habiendo resultado todos ellos estadísticamente significativos en esta respuesta, pero no en el descenso de los accidentes vasculares. Se sugiere que la potencia extra que proporcionaría un análisis conjunto de los cuatro estudios permitiría alcanzar la significación estadística. ¿Se trata de un estudio prospectivo o retrospectivo? ¿Es confirmatorio o exploratorio?

Comentario



Ciertos estudios **retrospectivos** utilizan bases de datos elaboradas de forma **prospectiva**. Tienen la ventaja de que las causas X están más libres de sesgos en su recolección, pero no garantizan que estén recogidas todas las variables relevantes. Además, para poder ser confirmatorios, deben especificar la hipótesis en el protocolo y justificar que ésta no ha venido sugerida por los resultados observados en esos pacientes, ya que entonces, la misma información habría sido utilizada para sugerir y confirmar la hipótesis, lo que se conoce como «capitalización del azar».

Lectura

Los términos “prospectivo” y “retrolectivo” suelen usarse también para distinguir entre estudios planificados antes o después de la ocurrencia de los datos. Pero este uso introduce ambigüedad, ya que estos términos suelen usarse para distinguir la secuencia temporal de la recogida de datos [...]. Para eliminar esta ambigüedad, los términos “prolectivo” y “retrolectivo” pueden emplearse y así evitar expresiones como [...] “estudio prospectivo retrospectivo”, [...] “prospectivo histórico”, “cohorte histórica” o “cohorte retrospectiva” para lo que aquí es denominado “cohorte retrolectiva” (Feinstein, 59).

Según el plan de muestreo

La recogida de información sobre las variables Y, X, y Z puede realizarse de varias formas.

Definición

En un estudio de cohortes, se recoge, en una población determinada, información en todas las unidades sobre las variables Y, X y Z.

Ejemplo 12.29

A partir de un censo de prostitutas masculinos de una cierta comunidad, se recoge información sobre sus hábitos y, en su momento, se valora su respuesta inmunitaria a los virus del sida y de la hepatitis.

Dado que ninguna de estas variables se ha fijado en el diseño, todas ellas ofrecen información sobre su distribución poblacional.

Definición

En un ensayo clínico se deja fija en el diseño la intervención X.

Ejemplo 12.30

Para estudiar el efecto de usar dos preservativos, se aconseja a la mitad de los prostitutas masculinos emplear uno; y, a la otra mitad, emplear dos.

Ejercicio 12.22

En este estudio, ¿informa la proporción de prostitutas que emplean dos preservativos de la correspondiente proporción poblacional?

Definición

En un diseño de caso-control se deja fija en el diseño a la respuesta Y.

De la misma forma que un ensayo clínico prospectivo necesita una muestra con la intervención control, un diseño retrospectivo, que parte de una serie de casos con la enfermedad (Y+), necesita una serie de controles sin la enfermedad (Y-).

Nota técnica

En un estudio prospectivo, se requiere un control para la intervención X, mientras que en uno retrospectivo se requiere el control para la evolución Y.

Ejemplo 12.31

No basta con observar que los pacientes con sida tienen un cierto grado de promiscuidad. Es necesario constatar que los que no tienen sida tienen menor promiscuidad.

Recuerde

En un ensayo clínico, el término «control» indica la opción terapéutica o causa de referencia; pero en un estudio retrospectivo de casos y controles indica la ausencia de enfermedades o del efecto en estudio.

Ejemplo 12.32

Se dispone de una muestra de 20 prostitutas que han desarrollado el sida (casos). Se selecciona otra muestra de 20 prostitutas que no lo han desarrollado (controles). En ambas muestras se les pregunta si han usado preservativos.

Ejercicio 12.23

Diga, de las siguientes proporciones, ¿cuáles pueden ser estimadas en el estudio del ejemplo 12.32 y cuáles no?

- La proporción de prostitutas que desarrollan el sida.
- La proporción de prostitutas con sida que usan preservativos.
- La proporción de prostitutas sin sida que usan preservativos.
- Entre los que usan preservativos, la proporción que desarrollan el sida.

En un diseño que fija alguna variable no pueden estimarse todos los parámetros.

Comentario

Esta limitación influye en el análisis estadístico, ya que ciertos cálculos habituales, como por ejemplo los riesgos relativo y atribuible, no podrán hacerse.

La *odds ratio* tiene la ventaja de que puede ser usada en cualquier tipo de diseño.

Recuerde

*En los diseños en los que se deja una **variable fija** hay algunas estimaciones que no pueden realizarse. La **odds ratio** no tiene esta limitación.*

En los estudios de casos y controles deben seleccionarse controles que sean «comparables» con los casos en estudio.

Comentario

Por comparables se debe entender que puedan proporcionar una información insesgada sobre el efecto del tratamiento, es decir, que tengan los mismos valores en aquellas terceras variables que puedan originar confusión. En otras palabras, se desea que la única diferencia entre casos y controles sea, además de la presencia o ausencia de la enfermedad que los define, su grado de exposición al factor. Así, para poder hacer inferencia causal, se controlan las condiciones observables Z mediante las herramientas disponibles, y se debe asumir, por ejemplo, que las variables V no observables están igualmente repartidas entre los grupos (no confusión).

Ejemplo 12.33

Para cada caso con sida se busca un control sin sida que tenga la misma edad, hábitos tóxicos, antecedentes patológicos, etc.

La tabla 12-5 muestra diferentes estrategias para obtener los controles, así como sus ventajas e inconvenientes.

Origen del control	Ventaja	Inconveniente
Población (p. ej., censo)	Seguro que todos los controles pertenecen a la población	Mala colaboración
Hospital: historia clínica, registros de enfermedades, etc.	Fácilmente accesibles	Mala representación
Vecindario		Mala colaboración Representación no asegurada
Amigos, familiares	Buena participación, fácilmente localizables	Representación no asegurada

Tabla 12-5 Tipos de controles

Recuerde

Los estudios de cohortes y los de casos y controles difieren en la perspectiva temporal y en el plan de muestreo (tabla 12-6).

	≥ 2 muestras (fijan una variable)	1 muestra
Retrospectivos	Caso-control	
Prospectivos	Ensayo clínico	Cohortes

Tabla 12-6 Perspectiva temporal y plan de muestreo en los ejemplos prototípicos de ensayos clínicos y estudios de cohortes y caso-control

La tabla 12-7 resume las propiedades principales de los estudios de casos y controles y de los de cohortes.

	Ventajas	Inconvenientes
Casos y controles	Rápidos de realizar Permiten obtener más casos	Menor fiabilidad de la información pasada El análisis estadístico puede sofisticarse
Cohortes	Son más confirmatorios Mejor información sobre X Mejores opciones de diseño	Más lentos Más caros Pueden requerir una cohorte grande

Tabla 12-7 Propiedades de los estudios de cohortes y de casos y controles

Eficiencia estadística

En los estudios de cohortes podría ser que sólo unos pocos individuos desarrollaran la enfermedad en estudio. Esta situación, muy deseable desde un punto de vista humano y sanitario, es fatal para la eficiencia estadística, que viene determinada por el tamaño del grupo menor.

Nota técnica



Si, por ejemplo, se desea comparar medias, la **eficiencia** estadística para una N total fijada se hace máxima cuando el tamaño n de los dos grupos es el mismo ($n = N/2$).

Para optimizar la eficiencia estadística de la comparación o para evitar el coste de recoger datos que no mejoran la información disponible se puede recurrir a diferentes estrategias estadísticas de muestreo. Así, en caso de enfermedades raras, puede recurrirse, por ejemplo, a estudiar sólo una pequeña proporción de las unidades que no desarrollan la enfermedad.

Ejemplo 12.34



Se dispone de muestras de sangre congeladas para todos los individuos de la cohorte. Se estudian todos los casos que desarrollan la enfermedad, pero de los que no, se analizan únicamente los de una pequeña proporción («**casos y controles anidados en una cohorte**»).

Comentario



El uso de las técnicas de optimización permite toda una serie de diseños alternativos que mejoran el coste y el seguimiento: «cohortes de casos y controles», «casos y controles híbridos», cohortes y casos, etc.

Recuerde



*Para una misma « n » total es más **eficiente** aquel diseño y análisis que conduce a un menor error típico de estimación.*

El error típico de estimación y los correspondientes intervalos de confianza dependen del tamaño muestral y de la magnitud de la varianza de la variable respuesta. Por tanto, una forma de obtener un error típico de menor magnitud para un mismo número de observaciones consiste en disminuir la magnitud de esta varianza, eliminando, por ejemplo, la variabilidad entre individuos.

Recuerde



Un estudio *apareado* pretende mejorar la eficiencia eliminando la variabilidad entre unidades.

Ejemplo 12.35



Dos antifúngicos tópicos son administrados simultáneamente en dos zonas afectadas del mismo individuo y de tamaño similar. La respuesta es la reducción de la zona afectada al cabo de un tiempo determinado. La comparación directa de zonas de un mismo individuo permite eliminar la variabilidad de la respuesta entre diferentes pacientes.

Ejemplo 12.36



En un estudio sobre el nivel de cumplimiento de la medicación, los pacientes utilizan alternativamente durante dos períodos de 6 meses dos modernos dispensadores de la medicación que permiten cuantificar el número de tomas dentro de un margen de tiempo preespecificado.

Para eliminar la variabilidad interunidades puede recurrirse a estudiar el cambio desde una medida inicial o basal de la respuesta; si bien, al valorar dos veces esta variable, la variabilidad intraindividual (que incluye el error de medida) afecta en dos ocasiones.

Ejemplo 12.37



En la comparación de los dos antifúngicos tópicos anteriores, sólo se dispone de una lesión por paciente, pero la respuesta es la disminución del tamaño de la zona afectada desde los valores iniciales. Dado que el tamaño se valora al inicio y al final, el error de medida se comete en estas 2 ocasiones, a diferencia de un estudio que sólo mida el valor final que tendrá sólo el error de medida (y la variabilidad intraindividual) debida a la determinación final.

Recuerde

Para que un **estudio del cambio** (o evolución desde el nivel basal) sea más **eficiente**, la variabilidad entre unidad que se elimina debe ser superior a la intraunidad que se introduce (Senn, 67).

Ejercicio 12.24

En estos estudios, ¿qué afirmaciones son correctas?

- a) El beneficio del estudio del cambio es mayor si todas las lesiones tienen la misma magnitud, ya que de esta manera se hace menor la variabilidad entre individuos.
- b) El estudio del cambio suprime la variabilidad debida a diferentes condiciones iniciales en la lesión.
- c) El diseño apareado, respecto al estudio del cambio, presenta la ventaja de que también suprime la variabilidad debida a diferencias entre individuos durante el período de seguimiento.
- d) En el diseño del cambio, a diferencia del diseño apareado, no es necesario controlar aquellas variables (hábitos higiénico-dietéticos y tóxicos, por ejemplo) que podrían introducir diferencias durante la evolución.
- e) El diseño apareado, respecto al estudio del cambio presenta la ventaja de que no introduce la variabilidad interunidad.

Calidad de la información

Toda imprecisión en la recogida de datos implica pérdida de información y, en consecuencia, limita la eficiencia estadística, aumentando el error de las estimaciones.

Pero, además, si el error que se comente está ligado a las variables en estudio (X e Y), la estimación de la relación entre ambas estará sesgada.

La calidad de la recogida de la información sobre la hipotética causa X es menor en los diseños retrospectivos. Si además viene afectada por la consecuencia Y , se estará produciendo un sesgo o diferencia sistemática entre los grupos en comparación.

Ejemplo 12.38

Podría ser que la capacidad de recordar los hábitos sexuales pasados dependa de tener actualmente el sida. Un paciente que tiene 5 parejas por año puede no acordarse de ningún *partenaire* si no ha desarrollado el sida; pero, en caso de contagio, se «acordará» muy bien de cada uno de ellos.

En los estudios **prospectivos**, la calidad de la recogida de información es mayor. Los datos sobre todas las variables se recogen en su momento de aparición. La causa X no tendrá sesgos al valorarse antes de la aparición de la consecuencia Y. Es esta última la que puede recoger algún sesgo, especialmente si se trata de variables «blandas», que deban ser evaluadas con alguna carga subjetiva.

Ejemplo 12.39



Un ejemplo sería la evolución de la enfermedad, a criterio del investigador o del propio paciente, valorada en escala ordinal (curación, mejoría, estable, etc.).

Contraejemplo



Un contraejemplo sería la mortalidad. (Aunque sería discutible si hubiera que atribuirle el origen.)

Para minimizar estos sesgos, en los ensayos clínicos puede recurrirse al enmascaramiento del tratamiento, de forma que se pueda administrar y evaluar a ciegas, sin que sea posible identificar los dos tratamientos en comparación.

Comentario



Se suele distinguir entre estudios **a ciegas** si cuando, o el investigador, o el paciente, desconocen el tratamiento administrado, y **con doble cegado** cuando ni uno ni otro lo conocen. Si ello no es posible y la evaluación no está libre de subjetividades puede recurrirse a la evaluación a ciegas. Para permitir este enmascaramiento debe recurrirse a la administración de un placebo que tiene idéntica apariencia exterior que el vehículo del tratamiento en estudio. Si se están comparando dos productos que se administran por vías diferentes puede recurrirse al doble placebo o **doble simulador** (double dummy), que consiste en administrar a cada grupo su tratamiento más el placebo-simulador del otro. Conviene mantener el enmascaramiento hasta que finalice la limpieza de la base de datos, que se **blinda** para no permitir modificaciones posteriores.

Recuerde



Diga «a ciegas», «enmascarado» o «cegado» y evite «ciego».

Historieta



Recuerde que blind \neq blinded.
¿Apoya la gran cantidad de estudios realizados con investigadores ciegos y evaluadores ciegos las tesis de Ernesto Sábato en su «Informe sobre ciegos»?

Recuerde

En todo estudio que relaciona dos variables debe evitar que el error de medida de una de ellas guarde relación con la otra variable.

*El **enmascaramiento** también es deseable en los estudios observacionales.*

Comentario

Parte del llamado efecto placebo podría indicar la evolución debida a la autosugestión del paciente y del clínico (además de los efectos de la regresión a la media y de los otros tratamientos recibidos).

En las intervenciones que implican cambios de hábitos por el paciente, es imposible que éste desconozca el grupo de tratamiento al que pertenece. Si debe hacer ejercicio o dieta, tiene que saberlo. Y también el profesional que le asesora durante el período del estudio. En estos casos, conviene que el evaluador desconozca el grupo de tratamiento.

Ejercicio 12.25

Imagine un estudio en el que un grupo recibe consejo dietético y el otro no, con evaluación a ciegas. ¿Cuáles de las siguientes expresiones son ciertas?

- a) El efecto directo de la intervención y los posibles efectos indirectos originados por el hecho de conocer el paciente la intervención que se le aplica están confundidos, en el sentido de que no pueden descomponerse: el estudio proporciona una estimación de la suma de ambos efectos.
- b) Desde el punto de vista de adquisición de conocimiento, sería interesante poder distinguir entre los efectos directos e indirectos del ejercicio.
- c) El sesgo originado por el hecho de conocer el paciente su tratamiento elimina la utilidad posterior de los resultados.
- d) Dado que un paciente siempre sabrá si está haciendo ejercicio o no, preocuparse por este posible sesgo es irrelevante a efectos prácticos.

Lectura

La tabla 12-8 reproduce la traducción realizada por el doctor Esteve Fernández del tercer borrador de la guía de publicación STROBE sobre estudios epidemiológicos (72). A finales de 2007 se publicará la versión internacional definitiva.

	Punto n.º	Estudios de cohortes	Estudios de casos de controles	Estudios transversales
Título y resumen	1	(a) Identifique el artículo como un estudio de cohortes en el título o en el resumen (b) El resumen debería ser una síntesis estructurada y muy informativa del artículo, teniendo en cuenta los puntos desarrollados en la lista de comprobación que sigue	(a) Identifique el artículo como un estudio de casos y controles en el título o en el resumen	(a) Identifique el artículo como un estudio transversal en el título o en el resumen
Introducción				
Antecedentes/fundamentos	2	Explique los antecedentes científicos y los fundamentos del estudio		
Objetivos	3	Consigne los objetivos específicos y cualquier hipótesis preespecificada		
Métodos				
Diseño del estudio	4	Presente los elementos clave del diseño del estudio. Declare el objetivo general del estudio original, si el artículo es uno de varios de un estudio en curso		
Marco	5	Describa el marco, lugares y fechas, definiendo los períodos de recogida de datos		
Participantes	6	(a) Explique los criterios de inclusión y exclusión, la fuente y los métodos de selección de los participantes (b) Describa los períodos y los métodos de seguimiento	(a) Para los casos y los controles por separado, explique los criterios de inclusión y exclusión, la fuente y los métodos de selección (b) Dé criterios diagnósticos precisos de los casos y motivos para escoger a los controles	Explique los criterios de inclusión y exclusión, la fuente y los métodos de selección de los participantes
Variables de interés	7	Proporcione una lista y defina claramente todas las variables de interés, indicando cuáles son tratadas como medidas de efecto o de exposición y cuáles son potenciales predictoras, variables de confusión o modificadoras		
Medidas	8*	(a) Proporcione detalles de los métodos de determinación para cada variable de interés (b) Si procede, describa la comparabilidad de los procedimientos en los diferentes grupos		
Sesgos	9	Describa cualquier tipo de medida utilizada para afrontar las potenciales fuentes de sesgo		
Tamaño de la muestra	10	Describa la justificación del tamaño del estudio, incluidas tanto las razones prácticas como las estadísticas		
Métodos estadísticos	11	(a) Describa todos los métodos estadísticos, incluyendo aquellos para el control de la confusión (b) Describa cómo se rataron las pérdidas en el seguimiento y los datos faltantes o perdidos (c) Si procede, describa los métodos para los análisis de subgrupos y los análisis de sensibilidad	(b) Describa cómo se trataron el emparejamiento y los datos faltantes o perdidos	(b) Describa cómo se trataron los efectos del diseño y los faltantes o perdidos
Exposiciones cuantitativas	12	(a) Explique cómo se analizaron las exposiciones cuantitativas; por ejemplo, qué categorizaciones o agrupaciones se escogieron y por qué (b) Si está indicado, presente los resultados de los análisis conjuntos o continuos y de los análisis agrupados		
Financiación	13	Explique la fuente de financiación y el papel de los financiadores en el presente estudio y, si fuera necesario, del estudio original en el que se basa		

Resultados				
Participantes	14*	(a) Para cada fase de estudio, informe del número de individuos: potencialmente elegibles, examinados para ser elegidos (si se conoce), incluidos en el estudio, que han acabado el seguimiento y analizados (b) Exlique las razones de no participación en cada etapa del proceso (c) Se recomienda la inclusión de un diagrama de flujo (d) Proporcione las fechas que definen el período de reclutamiento (e) En estudios apareados proporcione el número de controles por caso		
Datos descriptivos	15*	(a) Describa las características de los participantes en el estudio (p. ej., datos demográficos, clínicos o sociales) e informe sobre las exposiciones y los potenciales confusores (b) Indique, para cada variable de interés del estudio, la exhaustividad de los datos (c) Proporcione la media, la duración total y las fechas del seguimiento		
Datos de resultados	16*	Proporcione el número de sucesos del evento o tasas temporales	Informe del número de casos y controles en cada categoría de exposición	Informe del número de sucesos del evento o de medidas resumen para cada grupo en comparación
Principales resultados	17	(a) Proporcione medidas de asociación directas («crudas») o sin ajustar, así como ajustadas por las variables de confusión y su precisión (p. ej., intervalos de confianza del 95%). Deje claro por qué variables de confusión se ajustó y la justificación para incluir esas variables de ajuste y otras no (b) Para comparaciones que usen categorías derivadas de variables cuantitativas, indique el rango de valores o el valor mediano para cada grupo (c) Transforme las medidas relativas en diferencias absolutas de riesgo para periodos de tiempo significativos que no se extiendan más allá del rango de los datos (d) Proporcione los resultados ajustados por las variables de confusión para poblaciones objetivo realistas		
Otros análisis	18	Informe de otros análisis realizados, por ejemplo, análisis de subgrupos y análisis de sensibilidad		
Discusión				
Hallazgos clave	19	Resuma los hallazgos más importantes en relación con las hipótesis del estudio		
Limitaciones	20	(a) Discuta las limitaciones del estudio, teniendo en cuenta las fuentes potenciales de sesgo o de falta de precisión, y los problemas que pudieran derivarse por la multiplicidad de análisis, exposiciones y resultados estudiados. Discuta tanto la dirección como la magnitud de los potenciales sesgos (b) La discusión de las limitaciones no puede ser un sustituto de análisis cuantitativos de sensibilidad		
Generalización	21	Discuta la generalización (validez externa) de los resultados del estudio		
Interpretación	22	Aporte una interpretación global cauta de los resultados en el contexto de la evidencia disponible y las limitaciones del estudio, con especial atención a las interpretaciones alternativas		

*Proporcione esta información por separado para casos y controles en estos estudios y, si es aplicable, para los grupos expuestos y no expuestos en los estudios de cohortes y transversales.

Fuente: STROBE statement. Checklist of essential items Version 3 (Sept 2005) [citado 15 Nov 2005]. Disponible en: [www.strobe-statement.org/PDF/STROBE-Checklist-Versión 3.pdf](http://www.strobe-statement.org/PDF/STROBE-Checklist-Versión%203.pdf)

Tabla 12-8 Declaración STROBE: Lista de comprobación de puntos esenciales (Fernández, 72)

Ejercicio 12.26

De las siguientes características del ensayo clínico, ¿cuáles pertenecen también a los estudios epidemiológicos según STROBE?

- Hipótesis previa y, por tanto, protocolo.
- Diagrama de flujo de participantes.
- Controlar la multiplicidad de análisis.
- Enmascaramiento.

Estudios experimentales**Definición de ensayo clínico**

La característica fundamental de los estudios experimentales es que el investigador determina el grado de exposición a la causa en estudio. En una palabra: puede *asignar* el tratamiento. El diseño experimental aplicado en Medicina recibe el nombre de ensayo clínico.

Lectura

«Ensayo clínico es el método experimental aplicado a establecer los efectos de un procedimiento terapéutico. Se entiende por procedimiento terapéutico cualquier maniobra, intervención o producto químico o biológico administrado con cualquier intención, preventiva o curativa» (RD, 6).

Ejemplo 12.40

Ya se ha comentado el ejemplo de la estreptomicina en el tratamiento de la tuberculosis (69), en el que se comparó la Estreptomicina más reposo en $n = 55$ pacientes frente al reposo en $n = 52$. La respuesta era la evaluación enmascarada por dos radiólogos y un clínico de una placa radiográfica realizada a los seis meses de tratamiento.

Lectura

La ventaja más importante de asignar al azar el tratamiento es que equilibra las terceras variables V no observadas. (Los estudios observacionales deben recurrir a asumir alguna premisa sobre estas variables V .) Además, esta asignación permite optimizar la eficiencia estadística, lo que en general implica asignar igual número de casos a los grupos. Finalmente, permite optimizar el ajuste de las terceras variables, de forma que mejoren también la eficiencia estadística. En cambio, los estudios observacionales, al ajustar, pueden resultar en grupos desequilibrados, con pérdida de eficiencia.

Recuerde

Los diseños experimentales: 1) permiten controlar mediante la asignación al azar a las terceras variables V no observables; 2) permiten optimizar la eficiencia, y 3) disponen de mayor margen de maniobra para controlar a las terceras variables Z .
Todo ello repercute en un diseño más válido y más eficiente.

El carácter insesgado de estos estudios se expone de forma muy didáctica mediante los dos términos siguientes.

Definición

El **principio de comparabilidad** establece que dos o más grupos son comparables si difieren tan sólo en aquello que se desea comparar.

Definición

El **principio de intercambiabilidad** (*exchangeability*) dice que un diseño es insesgado o válido si se hubiera obtenido la misma estimación en el caso de que los dos tratamientos se hubieran intercambiado.

La tabla 12-9 resume algunas estrategias de asignación del tratamiento.

Asignación	Elección del criterio de asignación de tratamientos
Según criterio	El médico decide el tratamiento asignado a cada paciente
Aleatorización simple	Cada paciente es asignado a un grupo con la misma probabilidad π
Aleatorización balanceada por bloques	Cada «n» casos, el número de pacientes en cada grupo se equilibra. Los bloques se definen por terceras variables (orden, centro, edad, etc.)
Aleatorización por minimización	La probabilidad de asignar al paciente a un grupo se define de forma que, a la larga, se haga mínima la diferencia entre los grupos
Aleatorización según severidad (<i>cut-off</i>)	Se procede en función de los valores de severidad. Por ejemplo, los casos severos se comparan con un activo, y los leves, con placebo
Aleatorización y libre elección (<i>comprehensive cohort</i>)	Se dispone de datos de pacientes que han aceptado la asignación aleatoria y de datos de pacientes que han elegido libremente el tratamiento
Apostar por el mejor (<i>Play the winner</i>)	Se da mayor probabilidad de asignar un paciente al grupo con mejores resultados hasta el momento

Tabla 12-9 Posibilidades de asignación del tratamiento (Deeks et al., 73)

Para garantizar que las muestras generadas por la aleatorización son las que definen realmente los grupos, se debe evitar que el investigador que incluye los casos en el estudio conozca a qué grupo irán asignados antes de decidir incluirlos, ya que podría originar «sesgo de selección».

Ejemplo 12.41



Durante las hambrunas de inicios del siglo xx, un estudio pretendía evaluar el efecto en el peso de los escolares irlandeses de un vaso de leche extra en la merienda. Al finalizar el estudio pesaban más los que no tomaron leche, ya que se dejó en manos del maestro la selección de los alumnos que sí tomaban y éste lo administró a los alumnos de menor peso.

Recuerde



Para evitar el **sesgo de selección**, el grupo de asignación se enmascara o camufla (concealment) al investigador que incluye los pacientes.

Lectura



Las recomendaciones CONSORT (9) dedican los apartados 8 a 10 a la «aleatorización»: «8) **generación de secuencia aleatoria**: método que se usa para generar la secuencia de asignación aleatoria, incluyendo detalles de cualquier restricción (p. ej., bloqueo, estratificación, etc.); 9) **ocultamiento de la secuencia aleatoria**: método que se usa para implementar la generación de secuencia aleatoria (p. ej., contenedores numerados o teléfono central), aclarando si la secuencia se oculta hasta que las intervenciones se han asignado; y 10) **implementación de la secuencia aleatoria**, quién genera la secuencia de asignación, quién selecciona a los participantes y quién asigna los participantes a cada grupo».

Ejercicio 12.27



Compruebe si el trabajo de Chiasson et al. (74) cumple las directrices CONSORT sobre el informe de los métodos de aleatorización.

«De manera aleatoria, se asignó a pacientes idóneos, placebo o 100 mg de acarbosa 3 veces al día, tomada al principio de las comidas. Usamos un programa de ordenador para generar la secuencia de asignación aleatoria, estratificada por centro. La asignación aleatoria se llevó a cabo en bloques de 4 o 6. Se usaron envases de fármacos numerados para llevar a cabo la asignación aleatoria. Los pacientes se aleatorizaron secuencialmente en cada centro, pues el código de aleatorización se estratificó por centro. Las claves de aleatorización se ocultaron en un

Ejercicio 12.27 (Cont.)

envase etiquetado de tres compartimentos que se separó de los envases y se guardó en previsión de que los investigadores necesiten conocer el estatus de aleatorización del paciente. Un estadístico independiente, miembro del comité de revisión de la seguridad y calidad de datos, generó la secuencia de asignación. El reclutamiento y la aleatorización fueron gestionados en los centros.»

Tipos de ensayos clínicos

Se pueden realizar diferentes clasificaciones de los ensayos clínicos. La mayoría escapan de los objetivos de este curso.

Comentario

Una clasificación muy utilizada tiene en cuenta la fase de desarrollo de un fármaco. Una vez se dispone de un producto químico o biológico ya «inventado», el desarrollo empieza con el estudio de los niveles tolerados en el organismo (fase I), prosigue con el estudio exploratorio (fase II) y confirmatorio (fase III) de eficacia y termina con el estudio de efectividad (fase IV). Todos aportan información sobre la tolerabilidad. La demostración de eficacia de la fase III pretende conseguir el registro del producto en el mercado (tabla 12-10).

Lectura

*Lo que tiene **fase** es el **desarrollo** de un fármaco, no los ensayos clínicos que se utilizan en cada fase del desarrollo (ICH-E8, 75). Por ejemplo, en un fármaco ya comercializado puede desearse ampliar su ficha técnica con nuevas indicaciones o con nuevos estudios de tolerabilidad, o con estudios de interacción con otros fármacos, etc.*

Fase	Objetivo	Voluntarios	Tamaño	Diseño	Duración
I	Tolerabilidad y cinética	Sanos	Unidades	Dosis crecientes	Semanas
II	Exploración de eficacia	Pacientes	Decenas	Placebo, Aleatorizado, enmascarado	Meses
III	Confirmación de eficacia	Pacientes	Centenas	Placebo/Activo Aleatorizado, enmascarado	Años
IV	Efectividad y seguridad en grandes poblaciones	Pacientes	Millares	Activo	

Tabla 12-10 Características de los ensayos clínicos según la fase de desarrollo del fármaco

Comentario



Se suelen utilizar tres términos para determinar los efectos positivos, deseables del fármaco: la **eficacia** se establece en condiciones ideales (pacientes atendidos en un corto período de tiempo desde el inicio de los síntomas, con las mejores exploraciones complementarias...); la **efectividad**, en cambio, se estudia en condiciones reales, de práctica clínica habitual; finalmente, la **eficiencia** tiene en cuenta el coste de la intervención para la toma de decisiones. Su correcta evaluación se debe hacer considerando los incrementos de efectividad y coste respecto al procedimiento estándar actual, lo que se denomina **eficiencia diferencial**. Por otro lado, las técnicas psicométricas han permitido disponer de medidas validadas (insesgadas) y fiables (repetibles) de la **calidad de vida** de los pacientes, lo que facilita complementar la perspectiva clínica de eficacia con el punto de vista de los usuarios, beneficiarios últimos de las intervenciones sanitarias.

Ensayo clínico comunitario

En el ensayo clínico habitual la asignación del tratamiento se realiza a los pacientes, que constituyen las mismas unidades en las que se obtendrá la evaluación de los resultados. Pero estas unidades podrían estar agrupadas en macrounidades en las que se asigna el tratamiento. Este tipo de diseños ha recibido los nombres de ensayos comunitarios o aleatorizados en grupo (*cluster*) o multinivel.

Ejemplo 12.42



En un estudio sobre dos alternativas terapéuticas que precisan un elevado grado de entrenamiento de los profesionales, los centros sanitarios se asignan al azar y todos los pacientes de cada centro reciben el mismo tratamiento.

Ejemplo 12.43



En un estudio de sensibilización de escolares al sida, se asignan los centros escolares a una u otra intervención y todos los escolares de un centro reciben la misma información.

Ejemplo 12.44



En un estudio de prevención del tabaquismo, se desea comparar dos campañas publicitarias locales. Las comunidades se asignan al azar y luego se valora la respuesta en los individuos.

Nótese que en este tipo de diseños pueden establecerse mecanismos de «contagio» que rompan la independencia de las unidades y hagan que dos casos del mismo colectivo sean más similares entre sí que dos de diferentes colectivos, lo que debe considerarse en el análisis.

Recuerde

En un **ensayo comunitario**, la **asignación** del tratamiento se decide simultáneamente para un **grupo** de unidades.

Ejercicio 12.28

¿Son los estudios muticéntricos un caso particular de ensayos multinivel?

Desvíos del plan del ensayo clínico

Algunos casos no cumplen estrictamente el protocolo establecido (Cobo, 76). Esta situación, que sería incomprensible en un estudio sobre la calidad de fabricación de un producto, debe aceptarse en un ensayo clínico, dados los derechos de los pacientes incluidos en el mismo. Las desviaciones usuales consisten en pacientes que dejan de acudir a las visitas concertadas o no se toman la medicación prescrita.

Recuerde

Una buena intervención, un buen diseño, un buen investigador y un buen seguimiento conseguirán mantener al mínimo las desviaciones del protocolo. El crédito de los investigadores y la intervención se reduce a medida que aumentan los desvíos del protocolo.

Comentario

Algunos autores hablan de **quebrantamiento** (violation) del protocolo cuando el investigador incumple conscientemente el compromiso adquirido en el protocolo.

Las desviaciones del protocolo presentan dos grandes dificultades metodológicas. La primera es que pueden romper el equilibrio entre los grupos. La segunda, y definitivamente irremediable, es que impiden conocer el efecto real de la intervención. La potestad del clínico es aconsejar uno u otro tratamiento (X), pero el nivel de cumplimiento de las instrucciones pertenece al paciente (Z), y quizá sea una respuesta al tratamiento (Y'), pero no puede considerarse la causa X en estudio.

Recuerde

La variable experimental (X) es la asignación a un grupo de tratamiento. El cumplimiento del protocolo es una variable observacional (Z) que incluye atributos del paciente.

Por ello, el análisis principal de un ensayo clínico debe comparar a los pacientes a quienes se les ha recomendado un tratamiento con los que han recibido el otro consejo, lo que recibe el nombre de análisis por **intención de tratar** (ITT) o según la asignación del tratamiento (*as randomized*). El análisis alternativo, estudiando únicamente los pacientes que han cumplido el protocolo recibe el nombre de **análisis por protocolo** (PP). La medida del efecto futuro al asignar el tratamiento la proporciona el análisis por intención de tratar.

Recuerde



Los desvíos del protocolo amenazan la validez que proporciona el diseño experimental y, por tanto, lo acercan al diseño observacional.

Comentario



Un farmacólogo puro puede argumentar que él desea conocer el efecto del tratamiento en aquellos casos que cumplieron el protocolo, pero un farmacólogo clínico se preocupará por el efecto en aquellos casos a los que se les asignó el tratamiento.

Ejemplo 12.45



Se asignan al azar 200 casos a dos tratamientos alternativos A y B. En el grupo A, todos los pacientes cumplen y acaban el protocolo, mientras que en el grupo B, 50 de ellos dejan de tomar la medicación. En el grupo A se curan 60 y en el B, 40. (Tabla 12-11).

Así, entre los que acaban (PP), el B tiene mayor eficacia, ya que son un 80% (40 de 50) respecto al 60% (60 de 100) del A. En cambio, respecto a los que empiezan (ITT), considerando que no tomar la medicación es un fracaso del tratamiento, es el grupo A el que tiene mayor eficacia, 60% respecto al 40% del grupo B.

	N	Pérdidas	Acaban	Curan	P(Cura acaba)	P(cura empieza)
A	100	0	100	60	60%	60%
B	100	50	50	40	80%	40%

Tabla 12-11 El grupo B tiene un 50% de pérdidas y dificulta la interpretación de resultados

Comentario



El análisis por cumplimiento del protocolo es más inmediato, pero no tiene interpretación posterior. En cambio, la gran dificultad del análisis por intención de tratar reside en que no se dispone de valores para los casos que no han cumplido el protocolo. Si hay que incluirlos en el análisis, ¿qué valor se les puede asignar? ¿Cómo tenerlos en cuenta? La solución más aceptada y popular consiste en copiar el último valor disponible (last observation carried forward o LOCF).

Ejemplo 12.46



Si el paciente sólo participó en el estudio hasta el tercer mes, se asigna, a la evaluación de los 12 meses, el valor observado en el tercer mes.

Comentario



El problema de los datos faltantes es muy serio y de imposible solución ya que, por definición, falta información que no puede «inventarse». Una elegante clasificación distingue entre datos ausentes completamente aleatorios (no dependen de nada); aleatorios (sólo dependen de otras variables observadas); o no ignorables (dependen de variables observadas y no observadas). Los primeros pueden generarse al azar, los segundos, en función de las variables observadas y los terceros son los únicos que no pueden generarse y pueden originar problemas. Pero la clasificación de un dato ausente en una de estas categorías debe hacerse por creencias del investigador, ya que no se dispone de las variables no observadas. Por tanto, esta clasificación descansa en premisas no contrastables, lo que le resta valor.

Las conclusiones serán firmes si se pueden generar los datos ausentes bajo diferentes premisas y los resultados del análisis permanecen estables, independientemente de cómo fueron generados los datos. Ello requiere que el estudio tenga un número razonablemente reducido de pérdidas.

Comentario



Dos argumentos aconsejan usar únicamente los casos que han cumplido el protocolo. El primero son los estudios de seguridad del fármaco, donde no tiene sentido incluir casos que, por ejemplo, no han tomado el producto, ya que seguro que no tienen acontecimientos adversos. El segundo argumento aparece en los estudios de equivalencia, donde añadir

Comentario (Cont.)



datos NO fiables tiende a diluir las diferencias entre los grupos. Incluir estos casos se convertiría en un premio para un mal diseño o un mal seguimiento, lo que de ninguna manera parece razonable.

Recuerde



*La comparación más importante es **por intención de tratar**, si bien para el establecimiento de equivalencia o en el estudio de tolerabilidad, la comparación **por cumplimiento del protocolo** puede ser más relevante.*

Ya se ha dicho que los desvíos implican pérdida de información. Cuanto menor sea, por tanto, mejor. El estudio sólo recuperará toda la credibilidad perdida si cualquier consideración sobre estos desvíos no conduce a conclusiones diferentes.

Recuerde



Conviene que ambos análisis (ITT y PP) lleven a la misma conclusión (Porta et al., 77).

Son especialmente preocupantes las pérdidas relacionadas con la evolución, ya que podrían afectar de forma diferente a los grupos. Ello sucederá en los casos de eficacia nula o absoluta y de falta de tolerabilidad.

Comentario



Si el paciente murió durante el período de seguimiento, por la sana costumbre de los médicos de no pasar visita a los muertos, no se tendrá un valor introducido en la respuesta al final del estudio; pero ello no implica que a ese paciente no le corresponda, en esa variable, un valor que indique fracaso.

Recuerde



Si la pérdida está relacionada con la evolución, se sabe mucho sobre el dato ausente y se debe intentar asignarle, a ciegas del tratamiento, un valor que considere esta información.

Si las pérdidas son previas a la administración de un tratamiento enmascarado o se trata de pacientes erróneamente incluidos, podrían no afectar a la comparabilidad.

Finalmente, ¿qué se puede pensar de una situación en la que el paciente o no aparece o retira su consentimiento a participar en el estudio o simplemente no ha tomado toda la medicación? Todas estas situaciones pueden afectar a la comparabilidad.

Recuerde



Las pérdidas prealeatorización atentan contra la generabilidad de los resultados, y las pérdidas postaleatorización, contra la comparabilidad.

Comentario



*El uso de las tecnologías de la información ha permitido mejorar la calidad de los datos de forma muy notable. Quizá más por la posibilidad de establecer **filtros automáticos de coherencia** que permitan contrastar la información en el mismo momento en que se genera, que no por facilitar la transmisión o el almacenamiento.*

Historieta



Una anécdota del bioestadístico Dr. Alberto Cobos ilustra la importancia de estos filtros. El sistema avisó de que un paciente ya fallecido no podía tener presión arterial. Pero no era un error, estaba en espera de donación para trasplante. Al haberse ejecutado el filtro en el mismo momento en que se genera la información, la validación fue inmediata y no generó ninguna molestia.

Ejercicio 12.29



Para estudiar la protección para el cirujano del uso de un doble par de guantes (Cobo, 76), se ha realizado un ensayo (tabla 12-12) en el que los cirujanos son asignados al azar al grupo de utilizar uno o dos pares de guantes. Debido a ciertas creencias individuales (menor sensibilidad con dos pares, menor protección con un par, etc.), algunos cirujanos no han seguido el protocolo y han utilizado un número de pares diferente al asignado. En la tabla figuran los casos que han sido asignados a cada grupo, los pares de guantes que finalmente han usado y el número de casos en los que ha habido rotura y posible contaminación. Dadas estas transgresiones del protocolo se pueden plantear tres diferentes estrategias de análisis.

284 casos seleccionados				
Asignados a	Uno		Dos	
Casos	142		142	
Pares realmente usados	Uno	Dos	Uno	Dos
Casos totales	108	34	46	96
Se rompen	46	3	35	6
Resisten	62	31	11	90
% rotos	42,6%	8,8%	76,1%	6,2%

Tabla 12-12 Datos del ensayo sobre el efecto de usar doble par de guantes

Ejercicio 12.29 (Cont.)



1) Se incluyen todos los casos y se comparan los asignados a un par (49/142: 34,5%) frente a los asignados a dos pares (41/142: 28,9%), lo que se traduce en un $IC_{95\%}$ de la diferencia de proporciones de $-5,2$ a $+16,4\%$ que incluye el 0 y es, por tanto, no significativo.

2) Se consideran evaluables sólo aquellos casos que han cumplido el protocolo y se comparan los asignados a uno que usan uno (46/108: 42,6%) con los asignados a dos que usan dos (6/96: 6,2%), lo que se traduce en un $IC_{95\%}$ de 25,8 a 46,9%.

3) Se comparan todos los casos según los pares de guantes que realmente han usado: uno (81/154: 52,6%) o dos (9/130: 6,9%), lo que se traduce en un $IC_{95\%}$ de 36,7 a 54,7%.

¿Cuáles de las siguientes son ciertas?

- El análisis por intención de tratar es el 1.
- El análisis por cumplimiento de protocolo es el 2.
- En el análisis 3 se está estudiando un atributo de los pacientes, no una variable experimental. Por lo tanto, la conclusión pierde la fuerza causal de que estaba dotada por ser ensayo clínico y debe ser considerado con la prudencia de los estudios no experimentales.
- En el análisis 1 se concluye que no se ha podido demostrar que aconsejar el uso de dos guantes mejore la seguridad de los cirujanos.
- En los análisis 2 y 3 los resultados son estadísticamente significativos.
- En el análisis 2, la selección realizada a los casos no garantiza que los grupos sean comparables. Se están eliminando la segunda y la tercera columna, que incluyen los casos más particulares. No es de extrañar que los «super precavidos» de la segunda columna (que aunque les toca usar uno, usan dos) tengan un valor muy bajo de rotu-

Ejercicio 12.29 (Cont.)

ras (8,8%); mientras que los «despreocupados» de la tercera columna (que deberían usar dos pero utilizan uno) tengan el porcentaje más alto de roturas (76,1%). Al eliminar estos casos tan diferentes, los resultados serán distintos.

g) En el análisis 3 se demuestra que si se aconseja en el futuro el uso de doble guante descenderá la proporción de roturas.

Recapitulación: valor científico de los diferentes diseños

Ciertas características de los estudios científicos hacen que sus resultados sean más reproducibles.

La asignación al azar es la base del proceso de estimación

Ejercicio 12.30

Una crítica habitual de los ensayos clínicos es que están realizados en unas condiciones alejadas de las reales de aplicación. Por ello, se sugiere realizar estudios más «pragmáticos» en los que evaluar también la eficiencia para permitir la posterior toma de decisiones. ¿Cuáles de las siguientes expresiones son ciertas?

- Algunos ensayos clínicos se hacen en condiciones que garantizan su validez interna (comparabilidad) pero que limitan su validez externa (representatividad o generabilidad).
- Puede mejorarse la validez externa sacrificando la validez interna.
- Para que la validez externa esté comprometida se requiere que las características diferenciales del ensayo sean una variable modificadora del efecto de la intervención.
- La asignación al azar es la principal responsable de la pérdida de validez externa, ya que las restricciones éticas que implica el azar limitan mucho la inclusión de casos.

Recuerde

No se puede sacrificar la validez interna en nombre de la validez externa.

Lectura



Se suele olvidar que el «error típico» de estimación de la diferencia de las medias de dos muestras se calcula, bajo la hipótesis nula, asumiendo que ambas vienen de la misma población, por lo que el error típico se limita a cuantificar la incertidumbre generada por el azar. Si no hay asignación al azar, ello implica que las dos muestras provienen de poblaciones que pueden diferir en algo más, lo que origina una incertidumbre adicional que no cuantifica el error típico ni el intervalo de confianza. Se ha demostrado que esta incertidumbre adicional genera, además de resultados sesgados, oscilaciones mayores de las cuantificadas por las medidas estadísticas habituales, fenómeno conocido como **sesgo impredecible** (73).

Recuerde



Eliminar la **asignación al azar** puede provocar estimaciones sesgadas tanto del efecto como de la variabilidad de su estimador.

Valor del diseño para imputar causalidad

Olkin (78) propone ordenar de menor a mayor según el nivel de evidencia científica para **imputar relación causal** los diferentes diseños según figura en la tabla 12-13.

Esta clasificación es útil, pero orientativa: conviene estudiar detalladamente cada estudio para poder valorar el nivel de credibilidad de su inferencia causal.

Comentario



Además de sus ventajas intrínsecas por asignación de la causa en estudio, los diseños experimentales suelen acompañarse también de: 1) un protocolo con una hipótesis que garantizan su carácter confirmatorio; 2) idéntico seguimiento y evaluación (enmascarada) de los grupos en com-

1	Informe de caso aislados
2	Series de casos sin controles
3	Series de casos con controles de la bibliografía o históricos
4	Análisis de bases informatizadas
5	Estudios retrospectivos de casos-controles
6	Series prospectivas
7	Ensayo clínico
8	EC confirmado incluyendo metaanálisis
9	Metaanálisis con datos originales

Tabla 12-13 Jerarquía de los niveles de evidencia para establecer causalidad según Olkin (78)

Comentario (Cont.)



paración; 3) un sistema de gestión de datos que mejora su validez y fiabilidad, a la vez que agiliza el proceso de la información; 4) una auditoría de calidad; 5) un registro que permite localizar los estudios para las evaluaciones globales; y 6) un permiso legal que regula la responsabilidad ética de los investigadores.

De la misma forma que los ensayos clínicos están beneficiándose de procedimientos de análisis clásicos en el entorno observacional, este último se beneficia cada día más de la metodología del entorno experimental.

Ejercicio 12.31



¿Qué diferencia fundamental existe entre los métodos globales de epidemiología y el ensayo clínico controlado?

Ejercicio 12.32

Explique brevemente la razón de ser del ensayo clínico controlado. Es decir: ¿con qué argumentos técnicos y éticos se justifica la experimentación en seres humanos?

Ejercicio 12.33

Los grandes avances médicos en esperanza de vida están provocando un nuevo interés por la calidad de vida. Unos investigadores han desarrollado una escala que mide la calidad de vida de los ancianos con fractura de cadera. Han medido esta calidad de vida antes y después de la intervención quirúrgica mediante una t de Student apareada. Estiman, con una confianza del 95%, que la prótesis de cadera mejora la calidad de vida entre un 40 y un 50%. ¿Qué premisas son necesarias para poder afirmar que la cirugía es la causa de esta mejora?

Recuerde



Al valorar la **calidad de un estudio** de inferencia causal conviene saber si:

- 1) es confirmatorio, con hipótesis previa expresada de forma unívoca;
- 2) es longitudinal y respeta el lapso de tiempo para que la causa actúe;
- 3) recoge sin sesgo las variables en estudio;
- 4) es experimental, ya que asigna la causa a las unidades;
- 5) controla mediante el ajuste las condiciones Z observables;
- 6) controla mediante el azar las condiciones V no observables;

Recuerde (Cont.)



7) tiene un diseño eficiente y un tamaño muestral adecuado que conduce a un error aleatorio reducido; y
8) ha seguido con fidelidad el protocolo previsto.

Lectura



La tabla 12-14 (73) ordena los diseños para evaluar una intervención en función de la calidad o credibilidad de sus conclusiones.

Ejercicio adicional



Busque en internet la última versión, en inglés o en castellano, del Manual de la Cochrane Collaboration y conteste las siguientes preguntas:
¿Cuáles son los objetivos de esta Colaboración?
¿Qué apartados contiene el Manual?
¿Qué sesgos resalta en su capítulo sobre «Evaluación crítica de los estudios»?
Explique cada uno de estos sesgos a un compañero, poniendo ejemplos de situaciones en las que se podrían estar cometiendo.

Diseños experimentales

Un estudio en el que el investigador controla al menos algunas condiciones del propio estudio, especialmente las decisiones relativas a la asignación de los pacientes a los distintos grupos de intervención.

1. Ensayo clínico con asignación aleatoria

Los participantes son asignados aleatoriamente a los grupos de intervención o control, con seguimiento a lo largo del tiempo para determinar las posibles diferencias en los parámetros de evolución. El proceso de asignación aleatoria y el enmascaramiento de la asignación de los pacientes a los distintos grupos de intervención garantizan que, en promedio, los determinantes conocidos y desconocidos de la evolución se distribuyen de manera homogénea entre los distintos grupos.

2. Ensayo clínico con asignación casi aleatoria

Los participantes son asignados por el investigador a los grupos de intervención o de control, pero el método de asignación no es exactamente una asignación aleatoria genuina con enmascaramiento del proceso (p. ej., asignación en función de la fecha de nacimiento, el número de historia clínica hospitalaria, etc.)

3. Ensayo clínico semiexperimental sin asignación aleatoria

El investigador controla el proceso de asignación de los participantes a los distintos grupos, pero no lleva a cabo un procedimiento de asignación aleatoria (p. ej., la asignación se realiza en función de las preferencias del paciente o el médico). Difiere del «estudio de cohorte» por el hecho de que la intención es experimental más que de observación.

Diseños de observación

Estudio en el que se investiga la variación natural de las intervenciones (o la exposición) en los participantes en el estudio, con objeto de evaluar el efecto de las intervenciones (o la exposición) sobre distintos parámetros de salud.

4. Estudio con determinaciones iniciales y finales

Estudio de seguimiento sobre pacientes en los que se ha realizado una forma de intervención y sobre pacientes en los que no se ha efectuado dicha intervención, con cuantificación de la variable de evolución tanto en el momento de inicio del estudio como después del período de intervención, y con comparación de los cambios en las puntuaciones o, si los grupos son comparables, de los valores finales. También se puede considerar un diseño experimental si el investigador posee el control sobre la introducción de la intervención o bien puede realizar manipulaciones deliberadas en la misma.

5. Estudio de cohorte concurrente

Estudio de seguimiento en el que se compara la evolución entre los participantes que han recibido una intervención y los participantes que no la han recibido. Los participantes son estudiados durante el mismo período (de manera concurrente), bien de manera prospectiva o, con mayor frecuencia, retrospectiva.

6. Estudio de cohorte histórica

Variación sobre el estudio de cohorte tradicional en el que se establece el resultado de una nueva intervención aplicada sobre los participantes estudiados a lo largo de un período de tiempo concreto, y se compara dicho resultado con el obtenido en los pacientes que no recibieron la intervención en un período previo; es decir, los participantes no son estudiados de manera concurrente.

7. Estudio de casos y controles

Se identifican los pacientes con y sin una forma de evolución dada (casos y controles, respectivamente) y se comparan entre ambos grupos los resultados de la exposición a una intervención concreta.

8. Estudio con control histórico previo

Se realiza la comparación de los resultados obtenidos en los participantes del estudio antes y después de una intervención. Las mediciones previas y posteriores pueden realizarse en los mismos participantes o en muestras distintas. También se puede considerar un diseño experimental si el investigador tiene el control de la aplicación de la intervención o bien puede manipularla deliberadamente.

9. Estudio con diseño transversal

Evalúa en un momento concreto la relación existente entre la enfermedad y otras variables de interés existentes en una población definida.

10. Serie de casos

Descripción de un número de casos respecto a la forma de intervención y a los resultados (no se realiza la comparación con un grupo control).

Adaptada de CDR Report 4175.

Tabla 12-14 Taxonomía de los diseños de los estudios para evaluar la efectividad de una intervención (Deeks, 73)

Soluciones a los ejercicios

12.1 En el estudio experimental, el número de visitas es la intervención X asignada a los pacientes. En el observacional, en cambio, se trata de un atributo o condición Z de los pacientes. Este atributo podemos utilizarlo para predecir el nivel posterior de cumplimiento (objetivo para el que usualmente empleamos las Z). Pero si queremos que desempeñe el papel de una X y hacer inferencia causal, el tema es más delicado, ya que ahora es posible sospechar que el número de visitas es consecuencia de cierta característica personal que influye también sobre el nivel de cumplimiento de la medicación. Además, dado que este número de visitas no ha sido asignado, es preciso asumir que es posible hacerlo y que, cuando se haga, el paciente cumplirá con ellas. Nótese la importancia del término *asignación*. Incluso cuando no se acompaña del calificativo «aleatoria».

12.2 ¡Atención! Esta pregunta tiene trampa (mis sinceras disculpas). Este capítulo discute un tema científico, de adquisición de conocimiento: ¿qué sabemos sobre un cierto tema? En cambio, la pregunta hace referencia a una decisión concreta: tratar o no tratar a ese paciente. Y la decisión requiere otra metodología. Por supuesto que el conocimiento sobre la eficacia influye en esta decisión, pero no es el único factor. A las consideraciones técnicas sobre hasta qué punto el lapso de tiempo extra que ha transcurrido puede comprometer esta eficacia, hay que añadir consideraciones sobre la seguridad (más fácilmente extrapolable) y, en especial, sobre el coste.

12.3 Sólo es falsa la e), ya que los resultados de este estudio se basan, en el fondo, en un único centro. Nótese que las respuestas a) y b) son dos formas correctas de expresar lo mismo.

12.4 Ya están bien ordenadas. a) y b) ofrecen la misma credibilidad. Igual que d) y e). Los dos saltos más importantes son de c) a d) porque deja de estar especificado en el protocolo, y el último, ya que ese criterio es una auténtica barbaridad, algo así como si dijéramos: «No pares de analizar los datos hasta que den el resultado que yo quiero».

12.5 Como el objetivo es estudiar la relación entre X e Y ajustando por Z , nos centraremos en la primera fila de tablas. La segunda fila de tablas aclara que X y Z están relacionadas: son colineales. Y la tercera fila que Z e Y están también relacionadas: la condición Z es predictora de la respuesta Y .

La estimación de la relación parece ser diferente al condicionar por centro ($OR = 4$ en las dos primeras subtablas) que al considerar globalmente los casos ($OR = 14,74$), por lo que tiene sentido condicionar por centro ya que es una característica previa de los casos: cada paciente entrará y saldrá del estudio condicionado a que es de su centro. Así, los resultados de interés son los de las dos primeras subtablas. Como el centro no es modificadora del efecto, se presenta un único valor del efecto ($OR = 4$) mediante la combinación de las dos observaciones en ambos grupos de centro. (¡Y ojalá esto coincida con lo especificado en el protocolo!)

12.6 Es más sencillo y legible el segundo sistema.

12.7 En los pacientes del servicio de lípidos (Z_1) no se observa relación entre la presencia del gen y el desarrollo de las enfermedades. Dado que el intervalo de confianza es razonablemente pequeño, se pueden descartar valores elevados de relación y no parece que una muestra mayor pudiera cambiar mucho las conclusiones.

12.8 En la población general, estudiando conjuntamente los que tienen alterados los lípidos y los que no los tienen, se detecta una relación significativa entre el gen y las enfermedades cardiovasculares.

12.9 Al condicionar por lípidos, no se observa la relación global establecida entre gen y evolución. Por lo tanto, parece que todo el posible efecto de los genes en la evolución viene mediado por los lípidos. Si se pudiera intervenir sobre los lípidos puede esperarse que se consiga anular el efecto de los genes. La moraleja de estos 3 ejercicios es que el ajuste por una variable posterior es mucho más delicado de interpretar.

12.10 Si se acepta que el objetivo es identificar aquellos embarazos con una probabilidad de que el feto presente síndrome de Down lo suficientemente elevada para compensar los riesgos de exploraciones complementarias cruentas, se puede escoger para hacer la predicción, aquella variable que sea más barata, más temprana en el tiempo, más fiable (en el sentido de obtener valores más similares en sucesivas determinaciones) e, incluso, más fácil de interpretar. En todas estas propiedades, la edad de la madre antecede a la del padre.

12.11 Sí. Dado que no puede intervenir a la vez sobre ambas, puede tener sentido estudiar sus efectos independientes de cara a priorizar las intervenciones.

12.12 A nivel práctico, no. Quizá para satisfacer la curiosidad racional y contestar la pregunta retrospectiva sobre «¿cuáles son las causas de...?», pero no de cara a la intervención sanitaria.

12.13 Todas son ciertas.

12.14 Todas tienen sentido.

12.15 Ahora el objetivo pretende ir más allá de la predicción, ya que se quiere modificar el riesgo vascular interviniendo sobre la presión arterial. La pregunta crucial es ¿se puede bajar una de estas presiones dejando fija la otra? Si las intervenciones disponibles van a actuar disminuyendo ambas, no tiene sentido estudiar el efecto de bajar una de ellas dejando fija la otra. Una posibilidad sería considerarlas simultáneamente, por ejemplo, mediante la presión arterial media.

12.16 Evidentemente, comprometen la representatividad de la muestra. Convendría estudiar si tienen características diferenciales con los que sí han aceptado participar.

12.17 Son falsas la *b*), la *d*) y la *e*). Para saber si es cierta la *f*) haría falta confirmar la hipótesis de que el conocimiento de las variables predictoras avanza de las más importantes a las menos importantes.

12.18 Otra pregunta con trampa (mis disculpas). El placebo es tan sólo un simulador para enmascarar el tratamiento. La pregunta correcta es si el tratamiento que recibe el grupo control está de acuerdo con los protocolos profesionales. Recuérdese que al grupo «placebo» no se le suprimen los tratamientos de conocida eficacia. Por eso, muchos autores prefieren llamarlos estudios «*add-on*» para recordar que lo que se hace es añadir el tratamiento en estudio.

12.19 Esta pregunta es más de ética que de estadística. ¿Cuál de las dos variables es más fácilmente intervenible sin atentar contra los derechos humanos que establece el principio de autonomía? Recuérdese la estrategia de los investigadores del asma y la soja en Barcelona: «Ya que no podemos cambiar los vientos ni interrumpir la descarga de soja, repararemos los silos y pondremos una lona para aislar esta descarga».

12.20 Es una barbaridad. El metabolito, para poder provocar el infarto, debería ser anterior en el tiempo, no posterior. Para poder aceptar estas conclusiones es necesario asumir que el valor de este metabolito en este paciente era el mismo antes del infarto. Además, como el estudio no es aleatorizado, debe aceptarse que no hay confusión (es decir, que las variables no controladas por el diseño y el análisis están igualmente repartidas entre los grupos). Si se deciden a publicar estos resultados, se deberían discutir estas premisas para que los lectores puedan opinar sobre su verosimilitud.

12.21 Aunque los datos han sido recogidos de forma prospectiva, la hipótesis no figuraba en los protocolos respectivos, por lo que se trata de un análisis exploratorio, que puede generar hipótesis («nuestros resultados sugieren...») pero no teorías («hemos demostrado que...»). Si la hipótesis hubiera figurado en los protocolos originales, debería haberse repartido el riesgo α entre todos los contrastes de hipótesis que llevan a la misma decisión.

12.22 Evidentemente que no, la variable X ha quedado fija por diseño. No es una variable aleatoria que estime nada.

12.23 Desarrollar el sida es una variable fija, por lo que no se pueden estimar ni a) ni d), pero sí b) y c). Nótese que para ofrecer una estimación global de los prostitutos que usan preservativos deberían ponderarse las estimaciones b) y c) por la prevalencia de la enfermedad.

12.24 Son falsas la a), ya que si los individuos son muy homogéneos, el cambio eliminaría una variabilidad pequeña; y la d), pues siempre es necesario que los grupos en comparación tengan el mismo seguimiento y evaluación.

12.25 Todas figuran explícitamente, menos el enmascaramiento, que estaría implícito en medidas para evitar sesgos.

12.26 La única falsa es la c). El argumento d) es correcto en el caso de un hábito. Nótese que, por el contrario, en el caso de un fármaco, es imprescindible poder distinguir ambos efectos.

12.27 Los cumple.

12.28 No, siempre que la asignación se haga a nivel de los pacientes.

12.29 La única falsa es la última.

12.30 *a)* y *c)* son correctas; *b)* es falsa, ya que para poder generalizar algo a un grupo más extenso, primero debe poderse aplicar al grupo más reducido. Y también es falsa la *d)*, ya que los principios éticos afectan al término asignación. Ya se dijo que una vez se ha decidido que una intervención es éticamente asignable a un paciente, añadir el calificativo «aleatoria» tiene connotaciones estéticas pero no éticas.

12.31 La asignación del tratamiento.

12.32 Ante la duda de asignar un nuevo tratamiento, la mejor forma de evaluar su efecto es precisamente asignándolo. Todos los tratamientos que se asignan a los pacientes deben estar de acuerdo con los protocolos estándar del entorno. Además, el ensayo debe estar bien diseñado para poder aplicar de la forma más inmediata sus conclusiones.

12.33 Es necesario poder asumir que la calidad de vida de estos pacientes no hubiera evolucionado nada en ausencia de intervención. Esta premisa debe ser explícita en la discusión final del artículo. También es necesario justificar que no hay errores de medida ni de diagnóstico.

- 1 Stevens SS. On the theory of scales of measurement. *Science* 1946;103:677-80.
- 2 Stevens SS. Mathematics, measurement, and psychophysics. En: Stevens SS (ed). *Handbook of experimental psychology*. Nueva York: John Wiley; 1951.
- 3 Bollen KA. *Structural equation with latent variables*. Nueva York: Wiley; 1989.
- 4 Hand D. Statistics and the theory of Measurement. *J R Stat Soc Ser A*. 1996;159(3):445-92.
- 5 Velleman P, Wilkinson L. Nominal, Ordinal, Interval, and Ratio Typologies are Misleading. *Am Stat*. 1993;47(1):65-72.
- 6 Real Decreto 223/2004, de 6 de febrero. BOE núm. 33, 5429-43.
- 7 Página oficial de la ICH. Ginebra [consultada 18-6-2007]. Accesible en: <http://www.ich.org/>
- 8 ICH Topic E6 (R1). Guideline for Good Clinical Practice. Harmonised Tripartite Guideline. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. 1996.
- 9 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. *Ann Intern Med*. 2001;134:663-94.
- 10 Listas de comprobación para autores, revisores y editores de revistas médicas. *Med Clin (Barc)*. 2005; 125(Supl 1).
- 11 Cobos A. Ensayos clínicos aleatorizados (CONSORT). *Med Clin (Barc)*. 2005; 125(Supl 1):21-7.
- 12 Martín M, Llombart A, Lluch A, Alba G, Munáriz B, Tusquets I, et al. Estudio epidemiológico del grupo GEICAM sobre el cáncer de mama en España (1990-1993): proyecto «El Álamo». *Med Clin (Barc)*. 2004;122(1):12-7.
- 13 Grupo para la Asistencia Médica Integrada y Continua de Cádiz (GAMIC). La consultoría de medicina interna en atención primaria mejora la eficacia de la atención. *Med Clin (Barc)*. 2004;122(2):46-52.
- 14 Soriano A, García S, Ortega M, Almeda M, Gallart X, Vila J, et al. Tratamiento de la infección aguda de la artroplastia total o parcial de cadera con desbridamiento y régimen antibiótico oral. *Med Clin (Barc)*. 2003;121(3):81-5.
- 15 Bobes J, Bulbena A, Luque A, Dal-Ré R, Ballesteros J, Ibarra N, et al. Evaluación psicométrica comparativa de las versiones en español de 6, 17 y 21 ítems de la Escala de valoración de Hamilton para la evaluación de la depresión. *Med Clin (Barc)*. 2003;120(18):693-700.
- 16 González JA, Jover L. Los gráficos en la comunicación y el razonamiento científicos: ¿instrumento u ornamento? *Med Clin (Barc)*. 2004;122(Supl 1):3-10.

- 17 Ramos R, Solanas P, Cordón F, Rohlfis I, Elosua R, Sala J, et al. Comparación de la función de Framingham original y la calibrada del REGICOR en la predicción del riesgo coronario poblacional. *Med Clin (Barc)*. 2003;121(14):521-6.
- 18 Gisbert JP, González L, Cara C, Pajares JM, Moreno R. Actividad de la tiopurina metiltransferasa en pacientes con hepatitis autoinmune. *Med Clin (Barc)*. 2003; 121(13):481-4.
- 19 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration. *Clinical Chemistry* 2003;49:7-18
- 20 Gisbert JP, Luna M, Legido J, Hermida C, Maté J, Pajares JM. Anticuerpos anticitoplasma de los neutrófilos en el diagnóstico de la colitis ulcerosa y la enfermedad de Crohn. *Med Clin (Barc)*. 2004;122(4):134-5.
- 21 Baños JE, Brotons C, Farré M. Glosario de investigación clínica y epidemiológica. Monografías Dr. Antonio Esteve. Fundación Dr. Antonio Esteve. 1998; núm. 23.
- 22 Fumaz CR, Tuldrà A, Ferrer MJ, Paredes R, Bonjoch A, Jou T, et al. Quality of Life, emotional status, and adherence of HIV-1 - infected patients treated with efavirenz versus protease inhibitor - containing regimens. *J Acquir Immune Defic Syndr*. 2002;29(3):244-53.
- 23 Regidor E, Gutiérrez JL, Calle ME, Otero A. Patrón de mortalidad en España, 1998. *Med Clin (Barc)*. 2002;118(1):13-5.
- 24 Guyatt G, Drummond R. Guías para usuarios de literatura médica. Manual para la práctica clínica basada en la evidencia. The Evidence-Based Medicine Working Group. Barcelona: Ars Médica; 2004.
- 25 Cohn JN, Tognoni G. A randomized trial of the angiotensin-receptor bloker valsartan in chronic heart failure. *N Engl J Med*. 2001;345(23):1667-75.
- 26 Martín P, Leín A, Benítez E, Gómez JM, Mangas A. Comparación de dos modelos de educación para pacientes asmáticos. *Med Clin (Barc)*. 2001;116(11):408-12.
- 27 Ricart W, Bach C, Fernández JM, Biernés J, Sabrià J. Impacto de un cribado selectivo de la diabetes gestacional en una población española. *Med Clin (Barc)*. 1999;113(9):331-33.
- 28 Banegas JR, Rodríguez F, de la Cruz JJ, de Andrés B. Mortalidad relacionada con la hipertensión y la presión arterial en España. *Med Clin (Barc)*. 1999;112(13): 489-95.
- 29 Everitt BS. *Medical Statistics from A to Z. A guide for clinicians and medical students*, Cambridge: Cambridge University Press; 2003.
- 30 Chalmers AF. *What is this thing called science?* Open University Press; 1999.
- 31 Bunge M. *A la caza de la realidad*. Barcelona: Gedisa; 2006.
- 32 Costa A, Yust A, Solà J, Conget I, et al. Detección de la diabetes mellitus en consultas externas hospitalarias. Utilidad de un cuestionario de cribado. *Med Clin (Barc)*. 2003;120(8):287-91.
- 33 Cilla G, Pérez E, Montes M, Darío L, Beristain X. Seroconversión y serorreversión frente a *Helicobacter pylori* en mujeres que acuden al hospital por parto. *Med Clin (Barc)*. 2003;121(3):86-8.
- 34 Kleinbaum DG, Kupper L, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. Nueva York: Wiley; 1982.
- 35 Dal-Ré R, Ortega R, Espada J. Eficiencia de los investigadores en la selección de los enfermos para ensayos clínicos: a propósito de un estudio multinacional. *Med Clin (Barc)*. 1998;110(14):521-3.

- 36 Mediavilla JD, Sabio JM, Carrillo PL, Fernández C, Aliaga L, Jiménez J. Factores predictivos de hipertensión arterial en pacientes con dudas diagnósticas de hipertensión mantenida. *Med Clin (Barc)*. 2002; 119(11):401-4.
- 37 Gosset WS [artículo de la Wikipedia en inglés]. Accesible en: http://en.wikipedia.org/wiki/William_Sealy_Gosset
- 38 Agresti A. *Categorical Data Analysis*. Nueva York: Wiley; 2002.
- 39 Serra M, Díaz E, Verde Y, Gost J, Serra E, Puig M. Prevalencia del déficit de yodo y factores asociados en escolares de 4 años. *Med Clin (Barc)*. 2003;120(7):246-9.
- 40 Bishop YM; Fienberg SE, Holland PW. *Discrete Multivariate Analysis*. Cambridge MA: MIT Press; 1975.
- 41 Fisher RA. *Smoking and the cancer controversy*. Oliver and Boyd; 1959.
- 42 Hill AB. The environment and disease: association and causation? *Proc R Soc Med*. 1965;58:295-303.
- 42b Greenland S. Science versus public health action: those who were wrong are still wrong. *Am J Epidemiol*. 1995;133:435-6.
- 43 Popper KR. *El desarrollo del conocimiento científico*. Buenos Aires: Paidós; 1962.
- 44 Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ*. 2003;327:1459-61.
- 45 Llano A, Barretina J, Gutiérrez A, Blanco J, Cabrera C, Clotet B, et al. Interleukin-7 in plasma correlates with CD4 t-cell depletion and may be associated with emergence of syncytium-inducing variants in human immunodeficiency virus type 1-positive individuals. *J Virol*. 2001;103:19-25.
- 46 Altman DG. How large a sample? *Statistics in Practice*. Gore SM & Altman DG, eds. Londres: British Medical Association; 1982.
- 47 ICH Topic E9. Statistical principles for clinical trials. Harmonised Tripartite Guideline. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. 1998.
- 48 Machin D, Campbell M. *Statistical tables for the design of clinical trials*. 2.ª edición. Oxford: Blackwell Science; 1997.
- 49 Badiella LL, Marino AP. Cálculo del tamaño muestral (TM) con el programa Ene 2.0. *Ene-CTM*. 2005.
- 50 Dupont W, Plummer W. Power and sample size calculations: a review and computer program. *Control Clin Trials*. 1990 Apr;11(2):116-28.
- 51 Shuster JJ. *Practical Handbook of sample size guidelines for clinical trials*. Boca Raton: CRC Press; 1993.
- 52 ICH Topic E1. The extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life-threatening conditions. Harmonised Tripartite Guideline. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. 1994.
- 53 Gómez G. Técnicas estadísticas en análisis de supervivencia. *ICB digital*. 2002; 5:2-8.
- 54 Gómez G, Cobo E. Hablemos de... Análisis de supervivencia. *GH Continuada* 2004;3(4):185-91.
- 55 Gómez G, Calle ML, Langohr K. Análisis de supervivencia (En Cobo E, Videla S, Clotet B Editores. *Investigación clínica y estadística*. Fundación de Lluita contra la Sida) 2003;tema5:281-348.

- 56 Baudouin SV, Saunders D, Tiangyou W, Elson JL, Poynter J, Pyle A, et al. Mitochondrial DNA and survival after sepsis: a prospective study. *Lancet*. 2005;366:2118-21.
- 57 Valery PC, Coory M, Stirling J, Green AC. Cancer diagnosis, treatment, and survival in Indigenous and non-Indigenous Australians: a matched cohort study. *Lancet*. 2006;367:1842-48.
- 58 Anto JM, Sunyer J, Rodriguez R, Suarez M, Vazquez L. The toxicoepidemiological committee, community outbreaks of asthma associated with inhalation of soybean dust. *N Eng J Med*. 1989;320:1097-102.
- 59 Feinstein AR. *Clinical Epidemiology. The Architecture of Clinical Research*. Philadelphia: WB Saunders; 1985.
- 60 De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *JAMA* 2005;172(13):1700-2.
- 61 Freeman Jr, Daniel H. *Applied Categorical Data Analysis. Statistics, a series of textbooks and monographs*. Nueva York: Marcel Dekker; 1987.
- 62 Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*. 1996;7:498-501.
- 63 Rubin DB. Estimating causal effects of treatment in randomized and non randomized studies. *J Educ Psychol*. 1974;66:688-701.
- 64 Holland P. Statistics and causal inference. *J Am Stat Assoc* 1986;81:945-70.
- 65 Aickin M. *Causal Analysis in Biomedicine and Epidemiology; Based on Minimal Sufficient Causation*. Londres: Chapman & Hall; 2001.
- 66 Cox DR. Causality: some statistical aspects. *J R Stat Soc Ser A*. 1992; 155(2):291-301.
- 67 Senn S. *Statistical issues in drug development*. Chichester: Wiley; 1997.
- 68 Stone R. The assumptions on which causal inference rest. *J R Statist Soc (B)* 1993;55:455-66.
- 69 Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ*. 1948;2:769-82.
- 70 Rawls, J. (http://en.wikipedia.org/wiki/John_Rawls).
- 71 Cobo E. Papel ético del estadístico en la experimentación humana. *Qüestió*. 1999;23(1):155-65.
- 72 Fernández E. Estudios epidemiológicos (STROBE). *Med Clin (Barc)*. 2005;25 (Supl 1):43-8.
- 73 Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomized intervention studies. *Health Technol Assess*. 2003;7(27).
- 74 Chiasson JL, et al. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *Lancet*. 2002;359(9323):2072-7.
- 75 ICH Topic E8. General Considerations for clinical trials. Harmonised Tripartite Guideline. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use. 1997.
- 76 Cobo E. Diseño y análisis de un ensayo clínico: el aspecto más crítico. *Med Clin (Barc)*. 2004;122(5):184-9.
- 77 Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol*. 2007;60:663-9.
- 78 Olkin I. Keynote addresses. Meta-analysis: reconciling the results of independent studies. *Stat Med*. 1995;14(5):457-72.

Bibliografía recomendada

Introducción a la estadística médica

- Campbell MJ, Machin D. Medical Statistics, a commonsense approach. Chichester: Willey; 1999.
- Gardner MJ, Altman DG. Statistics with Confidence. Confidence Intervals and Statistical Guidelines. Londres: BMJ Books; 1989.
- Wonnacott TH, Wonnacott RJ. Introducción a la Estadística. México: Limusa Noriega; 1997.
- Campbell MJ. Statistics at Square Two, understanding modern statistical application in medicine. Londres: BMJ; 2001.
- Guyatt G, Rennie D. Guías para usuarios de literatura médica. Manual para la práctica clínica basada en la evidencia. Barcelona: Ars Médica; 2004.
- Everitt BS. Statistical Methods for Medical Investigations. Oxford: Oxford University Press; 1989.

Aplicación de los conceptos de medicina (epidemiología)

- Feinstein AR. Clinical Epidemiology: The Architecture of Clinical Research. Filadelfia: WB Saunders; 1985.

Libros teóricos

- Peña D. Estadística: Modelos y Métodos. Madrid: Alianza Universidad Textos; 1986. Vol. 1 y 2.

Introducción general a la investigación

- Greenfield T (ed.). Research Methods for Postgraduates. Londres: Arnold; 2002.

Referencia terminológica (inglés)

- Everitt BS. Medical Statistics from A to Z, A guide for clinicians and medical students. Cambridge: Cambridge University Press; 2003.

Determinación del cálculo muestral

- Badiella LL, Marino AP. Cálculo del tamaño muestral (TM) con el programa Ene 2.0. Ene-CTM. 2005.
- Machin D, Campbell M. Statistical tables for the design of clinical trials. 2.^a ed. Oxford: Blackwell Science; 1997.
- Shuster JJ. Practical Handbook of sample size guidelines for clinical trials. Boca Raton: CRC Press; 1993.
- Lemeshow S, Hosmer D, Klar J, Lwanga S. Adequacy of sample size in Health Studies. Chichester: Wiley; 1989.
- Dupont W, Plummer W. Power and sample size calculations: a review and computer program. Control Clin Trials; 1990 Apr;11(2):116-28.

- A**
- Análisis
 - confirmatorio, 255
 - exploratorio, 255
 - Auditoría de calidad, 337
- C**
- Causa o intervención (X), 259
 - Contraste de hipótesis, CH, 158-171
 - decisión, 171
 - riesgos de error, 171
 - riesgo α , 174
 - riesgo β , 174
 - potencia, 175
 - el error
 - de primera especie o tipo I, 173
 - de segunda especie o tipo II, 174
 - intervalos de confianza, pruebas de significación y contraste de hipótesis, 175
 - Correlación, 208-209
 - covarianza
 - muestral, 209
 - poblacional, 208
 - en datos pareados, 210
 - Criterios para clasificar un estudio, 309
 - orden casual, 310
 - según la secuencia temporal, 309
 - prospectivo(s), 309, 311, 312
 - retrospectivo(s), 309, 311, 312
 - transversal, 309-310
- D**
- Datos faltantes, 331
 - aleatorios, 331
 - análisis
 - as *randomized*, 330
 - por cumplimiento del protocolo, 330, 332
 - por intención de tratar, 330, 332
 - completamente aleatorios, 331
 - desvío del protocolo del ensayo clínico, 329
 - filtros, 333
 - de coherencia, 333
 - no ignorables, 331
 - pérdidas, 333
 - previas, 333
 - relacionadas con la evolución, 332
 - sesgo impredecible, 336
 - Descripción de los participantes, 31
 - Directrices
 - ICH, 9
 - directrices relevantes para realizar y documentar EC, 10
 - guías de las revistas biomédicas, 11
 - CONSORT, 12
- Strobe* 321
- Diagnóstico
 - especificidad, 45-47
 - falsos
 - negativos, 51
 - positivos, 51
 - información
 - a posteriori, 52
 - a priori, 52
 - sensibilidad, 45-47
 - valor(es) predictivo(s), 46-47
 - negativo, 45
 - positivo, 45
 - Distribución
 - normal, 71-72
 - aplicaciones al diagnóstico, 83
 - curvas ROC (*Receiver Operating Characteristic curves*), 83
 - desvío tipificado, 76
 - parámetros, 72
 - desviación típica, 72
 - media, 72
 - varianza, 72
 - uso
 - con aplicaciones informáticas, 82
 - de las tablas, 74
 - de las tablas, ir de los porcentajes a los valores, 74
 - de las tablas, ir de los valores a los porcentajes, 74
 - t* de Student, 163
 - condiciones de aplicación, 141
 - intervalo de confianza
 - de μ usando *S*, 137
 - premisas para estimar μ sin conocer σ , 138
 - distribución normal, 138
 - tamaños muestrales, 138
- E**
- Efecto causal
 - en la población, 276
 - en una unidad, 270, 276
 - ajustado, 278
 - asignación, 271
 - intervenibles, 271
 - premisa de invarianza, 273
 - regresión a la media, 274
 - efectos fijos, 276
 - estudios «n igual a 1», 277
 - modelo, 276
 - problema fundamental de la inferencia causal, 275
 - definición prospectiva, 271
 - Eficiencia, 317, 319
 - apareado, 318
 - casos y controles, 317
 - anidados en una cohorte, 317

- estudio del cambio, 319
 - Ensayo clínico, 327
 - asignación al azar, 336
 - ensayo clínico comunitario, 328
 - en grupo, 329
 - ensayo clínico con asignación aleatoria, 339
 - ensayo clínico con asignación casi aleatoria, 339
 - ensayo clínico semiexperimental sin asignación aleatoria, 339
 - permiso legal, 337
 - protocolo, 336
 - registro, 337
 - efectividad, 328
 - eficacia, 328
 - eficiencia, 328
 - fases, 327
 - I Tolerabilidad cinética, 327
 - II Exploración de eficacia, 327
 - III Confirmación de eficacia, 327
 - IV Efectividad y seguridad, 327
 - Equivalencia
 - intercambiabilidad, 189
 - planteamientos de equivalencia, «no inferioridad» y «no superioridad»
 - equivalencia
 - individual, 189
 - poblacional, 189
 - no inferioridad», 187
 - no superioridad», 185, 187
 - prescribibilidad, 189
 - Escalas de medida, 2, 5
 - Escala
 - de intervalo, 3
 - de razón, 4
 - nominal, 2
 - ordinal, 3
 - propiedades, 5
 - Especificidad, 45-47
 - Estadísticos, 20, 22
 - apropiados según la escala de medida, 26
 - dispersión, 22
 - desviación típica, 23
 - varianza, 23
 - escala
 - de intervalo, 21, 25
 - media, 21
 - promedio, 21
 - nominal, 20
 - moda, 20
 - ordinal, 20
 - cuartiles, 20
 - deciles, 20
 - mediana, 20
 - probabilidades acumuladas, 20
 - quintiles, 20
 - tendencia central, 20
 - media, 21
 - mediana, 20
 - moda, 20
 - Estudio(s)
 - enmascarado, 320
 - experimental, 254, 324, 339
 - aleatorización
 - balanceada por bloques, 325
 - por minimización, 325
 - según severidad, 325
 - simple, 325
 - y libre elección, 325
 - apostar por el mejor, 325
 - asignación según criterio, 325
 - CONSORT, 326
 - generación de secuencia aleatoria, 326
 - implementación de secuencia aleatoria, 326
 - ocultación de la secuencia aleatoria, 326
 - ensayo clínico, 324 (v. *Ensayo clínico*)
 - principio
 - de comparabilidad, 325
 - de intercambiabilidad, 325
 - sesgo de selección, 326
 - longitudinal, 251, 309, 337
 - metaanálisis con datos originales, 336
 - observacional, 254, 321, 339
 - guía de publicación STROBE, 321
 - análisis de bases informatizadas, 336
 - estudio con determinaciones iniciales y finales, 339
 - estudio con control histórico previo, 339
 - estudio de cohorte concurrente, 339
 - estudio de cohorte histórica, 339
 - informe de casos aislados, 336
 - series
 - de casos con controles de la bibliografía o históricos, 336
 - de casos sin controles, 336
 - retrospectivas, 336
 - prolectivo, 257
 - hipótesis documentada previa, 257, 336
 - estudio confirmatorio, 337
 - prospectivos 251, 309, 311, 312
 - retrolectivo, 257
 - según el plan de muestreo, 313
 - calidad de la información, 319
 - eficiencia, 319
 - sesgo, 319
 - caso-control o de casos y controles, 314, 315-316, 339
 - odds ratio*, 315
 - variable, 315
 - comparables, 315
 - cohortes, 313, 316
 - y casos, 317
 - retrospectivos, 251, 309, 311, 312 319
 - sesgo, 319
 - según la perspectiva temporal, 316
 - transversal, 309-310
- F**
- Falsos
 - negativos, 51
 - positivos, 51
 - Función
 - de riesgo, 236
 - de supervivencia, 235
- G**
- Gráficos, 34
 - escala de medida, 34
 - variables
 - continuas, 36
 - box-plot* o diagrama de caja, 37
 - histograma, 36
 - discretas, 34
 - diagrama de barras, 34, 35
 - gráfico de sectores, 34
 - Gestión de datos, 337
- I**
- Importancia
 - de estudiar las medias, 197
 - de la amplitud del intervalo de confianza
 - grado de ignorancia, 199
 - no implica variabilidad del efecto, 199
 - «respondedores», 199
 - Diferencia entre las medias
 - muestrales, 200
 - estima el efecto en cada unidad, 197

- esperanza, 200
 varianza, 200
 poblacionales, 200
 estimador único ponderado, 201
 homocedasticidad, 198
 de tener un control, 194
 Inferencia, 91
 distribución del estadístico \bar{X} , 119
 error(es)
 aleatorio(s), 103, 105
 sesgos, 103
 sistemáticos, 103
 típico
 ¿desviación típica o error típico? 114
 de la media de X , 113
 de \bar{X} , 112, 113
 estadística, 92, 98, 106
 error aleatorio, 106
 información empírica o «evidencia», 98
 método científico-técnico, 101
 pasos del método científico, 101
 estimación puntual, 119
 estimador(es), 97, 117
 insesgado, no sesgo, 108
 convergente, 117
 eficiente, 117
 individuo (caso o unidad), 93-94
 muestra, 93-94
 aleatoria, 98
 simple (MAS), 98
 parámetro(s), 96
 esperanza, 96
 media, 96
 población(es), 93-94
 muestreada, 103
 externa, 103
 objetivo o diana, 103
 preciso, 105
 principios, 91
 sesgo, 105
 unidad (individuo o caso), 93-94
 validez, 105
 externa, 105
 interna, 105
 Intervalo(s) de confianza
 de la probabilidad π , 140
 error típico del estimador p , 140
 de μ usando S , 137
 de las medidas de asociación, 143
 Odds Ratio (OR), 147
 error típico del $\ln(\text{OR})$, 147
 $\text{IC}_{95\%} \ln(\text{OR})$, 147
 requisito: tamaño muestral grande, 147
 riesgo atribuible (RA), 143
 error típico de RA, 144
 $\text{IC}_{95\%} \text{RA}$, 144
 requisito: tamaño muestral grande, 144
 riesgo relativo (RR), 145
 error típico del $\ln(\text{RR})$, 144
 $\text{IC}_{95\%} \ln(\text{RR})$, 145
 requisito, tamaño muestral grande, 145
 del 95%, 131
 cobertura, 131
 Intervención, 248
- M**
- Media 96, 106
 Medida(s), 6
 de asociación, 61
 número necesario de casos tratados, 65
 number needed to treat, 65
 odds ratio (OR), 64, 65
 riesgo atribuible (RA), 62
 relativo (RR), 63, 64
 de posición relativa de los individuos, 28
 desvío tipificado, 29
 caso extremo, 29
 fuera de la banda (*outlier*), 30
 tipificar o estandarizar, 29
 propiedades
 error
 aleatorio, 6, 7
 sistemático, 6, 7
 fiabilidad, 6-7
 validez, 6-7
 variable(s)
 latente, 6
 observables, 6
 Monitorización, 303
 seguimiento, 303
 Métodos de control de las terceras variables
 observables, 284
 ajuste global (minimización), 293
 efecto causal ajustado, 285
 ausencia de sesgo, 286
 control de la variabilidad, 285
 eficiencia, 285-286
 estudio de subgrupos, 288
 apareamiento, 290
 bloques, 288
 estimación única, 289
 estratos, 288
 interpretación de resultados ajustados, 296
 efecto de X independientemente de Z , 298
 variable posterior, 296
 variable Z , 296
 modelado, 291
 parsimonia, 289
 restricción, 286
 criterios de selección, 288
 reducción de la población objetivo, 286
 Muestras
 dependientes o apareadas, 205
 descomponer la varianza, 206
 efecto unidad, 207
 variabilidad intraunidad, 206
 independientes, 204
- N**
- Normas de buena práctica en estudios clínicos, 8
- O**
- Objetivos clínicos
 ¿cuál o cuáles son las causas?, 252
 descripción, 253
 farmacoepidemiología, 253
 farmacología clínica, 253
 intervención, 249, 253
 predicción, 248, 249, 253
 preguntas
 prospectivas, 251
 retrospectivas, 251
 prevención, 60
 primaria, 60
 secundaria, 60
 terciaria, 60
 pronóstico, 248
 relación causa-efecto, 249, 336
Odd, 54-55
 Odds diagnóstica, 58
 función soporte, 59
 información a posteriori, 59
 información a priori, 59
 razón de verosimilitud, 58
 Odds ratio, 64, 65, 147, 315

- P**
- Papel del azar, 302
 - asignación al azar, 303
 - comparabilidad, 303, 304
 - consideraciones éticas, 308
 - asignación, 308
 - asignar al azar el primer paciente, 309
 - medicaciones de rescate, 308
 - obtención o extracción al azar, 302, 303
 - critérios de selección, 303
 - extrapolabilidad, 303
 - generabilidad, 303
 - muestra, 303
 - población externa, 303
 - población objetivo, 303
 - representatividad, 302
 - validez externa, 303
 - validez interna, 303
 - Papel ético del científico, 308
 - Probabilidad, 41-42
 - condicionada, 43, 45
 - incidencia, 59
 - prevalencia, 59-60
 - riesgo, 52
 - tasa, 53
 - teorema de Bayes, 49-50
 - Prueba de significación, PS
 - «evidencia» empírica, 160
 - estadísticamente significativo, 160
 - valor de P , 158, 160
 - condiciones de aplicación, 161
 - muestra grande, 161
 - probabilidad π no extrema, 161
 - hipótesis H , 159, 162
 - información, 160
 - mecánica de la prueba de significación, 161
 - distribución del estadístico bajo H , 161
 - hipótesis (H), 161
 - premisas, 161
 - poner a prueba, 158
 - prueba de significación de una media, 164
 - cociente señal/ruido, 167
 - pruebas de significación e intervalos de confianza, 167
 - pruebas de significación unilaterales y bilaterales, 169
 - bilaterales, 169
 - unilateral por la derecha, 170
 - unilateral por la izquierda, 169
 - R**
 - Razón t , 202
 - cantidad de información, 204
 - diseño óptimo, 204
 - premisas, 202
 - distribución normal, 202
 - muestras independientes, 202
 - varianzas iguales, 202
 - ruido, 202
 - señal, 202
 - tipificada, 204
 - Riesgo
 - atribuible 62, 144
 - relativo 63, 145
 - S**
 - Sensibilidad, 45-47
 - Supervivencia 230
 - asimetría de la variable tiempo, 231
 - casos censurados, 234
 - censura, 232
 - cierre, 233
 - comparación de curvas, 240
 - prueba de *log-rank*, 241
 - prueba de Gehan, 241
 - consecuencias de la asimetría y la censura, 234
 - el análisis basado en medias no es adecuado, 231
 - estimación
 - instantánea (Kaplan-Meier), 239
 - por intervalo (actuarial), 239
 - estimador del efecto, 245
 - modelo de riesgos proporcionales, 242
 - modelo de Cox, 245
 - regresión a Cox, 242
 - premisa
 - de riesgos proporcionales, 243
 - estudio
 - bivariante, 230
 - multivariante, 230
 - univariante, 230
 - inclusión, 233
 - tiempo de seguimiento, 233
 - T**
 - Tamaño muestral necesario
 - para comparar medias con datos independientes, 214
 - dispersión σ , 215, 218
 - magnitud del efecto δ , 215
 - efecto estandarizado, 216
 - para estimar una probabilidad, 222
 - margen de seguridad, 220
 - planteamiento, unilateral, 219
 - potencia, 216, 218
 - precisión, 221
 - amplitud, 221
 - intervalo de confianza, 221
 - riesgos α y β , 215
 - variabilidad intrasujeto, 218
 - varianza de la diferencia de las medias, 214
 - Tendencia central, moda, 20
 - Teoría representativa de la medida, 5
 - Terceras variables (Z , observables), 259
 - buen diseño, 261
 - ausencia de riesgo, 261
 - confusión, 259
 - validez, 265
 - eficiencia, 265
 - modificación del efecto, 266
 - colinealidad, 267
 - interacción, 267, 269
 - parsimonia, 269
 - relación marginal, 268
 - relación parcial, 268
 - Terceras variables (V , no observables), 301
 - asignación
 - aleatoria, 301
 - ignorante de tratamiento, 301
 - no confusión, 301
 - premisa, 301
 - suficiencia del modelo, 301
 - V**
 - Valor(es) predictivo(s), 45-47
 - negativo, 45
 - positivo, 45
 - Variable respuesta o efecto (Y) en estudio, 259
 - Varianzas diferentes, 204
 - aplicar logaritmos, 205
 - estadístico señal/ruido, 205
 - heterocedasticidad, 205
 - información de Y , 205