

Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica¹

Segunda edición

Richard K. Riegelman y Robert P. Hirsch

PARTE IX:

Capítulo 26. Principios básicos
Capítulo 27. Análisis univariantes

¹ Título original: *Studying a Study and Testing a Test How to Read the Medical Literature*. Second edition. © Richard K. Riegelman, Robert P. Hirsch. Publicado por Little, Brown and Company, Boston, Massachusetts 02108, Estados Unidos de América. Los pedidos del libro en inglés deben dirigirse a esta dirección

Versión en español autorizada por Little, Brown and Company; se publica simultáneamente en forma de libro (Publicación Científica 531) y como serie en el *Boletín de la Oficina Sanitaria Panamericana*. Traducción de José María Borrás, revisada por el Servicio Editorial de la Organización Panamericana de la Salud

© Little, Brown and Company, 1989. Todos los derechos reservados. Ninguna parte de esta publicación puede ser reproducida ni transmitida en ninguna forma ni por ningún medio de carácter mecánico o electrónico, incluidos fotocopia y grabación, ni tampoco mediante sistemas de almacenamiento y recuperación de información, a menos que se cuente con la autorización por escrito de Little, Brown and Company

PRINCIPIOS BÁSICOS

La estadística aplicada a la investigación médica persigue tres finalidades: 1) sintetizar numerosas mediciones en un número limitado de datos manejables, 2) realizar estimaciones e inferencias a partir de las muestras extraídas de poblaciones, teniendo en cuenta la influencia del azar, y 3) ajustar los datos según la influencia de las variables de confusión en esas estimaciones e inferencias. Nuestro objetivo en *La selección de una prueba estadística* es arrojar algo de luz sobre la forma en que la estadística puede ayudar a conseguir estos fines. No suponemos que la información brindada en estas pocas páginas pueda reemplazar la participación de un estadístico en las fases de planificación, ejecución e interpretación de la mayor parte de los proyectos de investigación médica; pero sí esperamos proporcionar las herramientas necesarias para que los lectores de la literatura de investigación sepan valorar la sección de los “métodos estadísticos” de tal forma que el análisis, la interpretación y la extrapolación de los resultados de la investigación se puedan comprender totalmente.

Para utilizar la estadística en la investigación médica, en primer lugar es preciso escoger un método estadístico apropiado. En segundo lugar, las mediciones de la investigación deben ser manipuladas de acuerdo con el método seleccionado. Por último, los resultados de estas manipulaciones han de interpretarse correctamente. La primera y la última de estas tareas están íntimamente relacionadas con el tema de la Parte 4, *La selección de una prueba estadística*. Sin embargo, no trataremos de discutir a fondo las manipulaciones de los datos que son necesarias para producir los resultados estadísticos. Sin lugar a dudas, el estudio de estas manipulaciones requiere una comprensión más profunda de los métodos estadísticos, pero, en nuestra opinión, no es preciso tener ese nivel de conocimientos para poder evaluar por qué se selecciona un método determinado y cómo podemos interpretar los resultados de su aplicación.

Empezaremos echando un vistazo a la forma de enfocar las primeras dos finalidades de la estadística. La tercera, ajustar los datos según el efecto de las variables de confusión, se realizará mediante el análisis multivariable, que se presentará en el capítulo 29.

SÍNTESIS DE LAS MEDIDAS

Como se ha afirmado anteriormente, una de las finalidades de los métodos estadísticos consiste en resumir grandes cantidades de datos en un número reducido y manejable de ellos. Para cumplir con esta tarea debemos darnos cuenta, en primer lugar, de que las mediciones realizadas en los sujetos de una investigación son una parte o una *muestra* de un grupo más numeroso de individuos que podrían haber sido incluidos en la misma. Este grupo más numeroso se denomina *población*.¹

¹ En medicina, habitualmente pensamos en mediciones realizadas en personas, en lugar de animales u objetos. Esto puede crear la falsa impresión de que el término estadístico *población* es el mismo que se utiliza para describir distintos conjuntos de personas en política o en geografía. Aunque una población estadística podría ser uno de esos grupos de personas, no se limita a ellos. Una *población estadística* se define como el conjunto de todas las mediciones posibles (no necesariamente realizadas en personas) de las cuales se selecciona una muestra.

FIGURA 26-1. Una distribución poblacional hipotética de las mediciones de la concentración de bilirrubina sérica

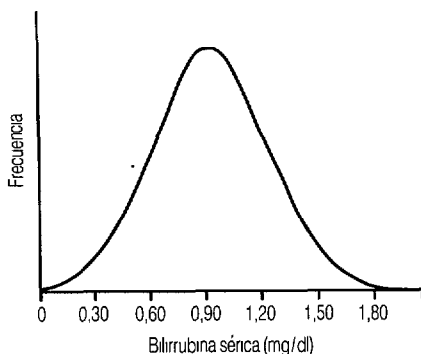
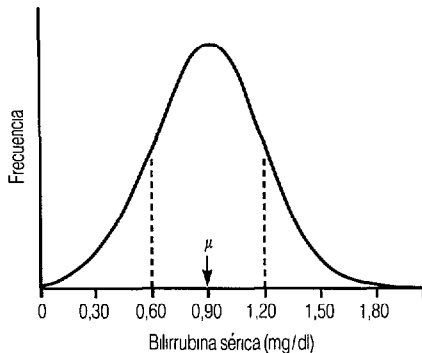


FIGURA 26-2. Una distribución gaussiana hipotética de la concentración de bilirrubina sérica con una media de 0,9 mg/dl y una desviación estándar de 0,3 mg/dl. Las líneas discontinuas indican los valores iguales a la media \pm la desviación estándar



Si marcamos en una gráfica la frecuencia con que aparecen los distintos valores de una variable en la población, obtendremos una representación gráfica de la *distribución poblacional*. La distribución poblacional describe la frecuencia con que aparecen los valores en la población de la que se extraen las muestras para observación (figura 26-1). No obstante, es difícil asimilar o transmitir la información contenida en los datos representados en esa gráfica.

Los métodos estadísticos ofrecen una medida sintética de la distribución poblacional, en lugar de su descripción gráfica. Cada tipo de distribución poblacional tiene un número limitado de valores sintéticos, denominados *parámetros*, que se utilizan para describir completamente la distribución concreta de las mediciones. Por ejemplo, para describir íntegramente una *distribución gaussiana*,² se necesitan dos parámetros: la *media*³ (la posición de la distribución en una escala continua o, más concretamente, su "centro de gravedad") y la *desviación estándar*⁴ (la *dispersión* de la distribución, dado que indica cuán alejados de la media se encuentran los valores individuales). La figura 26-2 muestra una distribución gaussiana con la media, como medida de posición de la distribución, y la desviación estándar, como medida de dispersión.

² La distribución gaussiana también se conoce como distribución *normal*. Evitaremos la utilización del último término, porque normal tiene otro sentido en medicina. La distribución gaussiana es la distribución poblacional que se supone la mayor parte de las veces en estadística.

³ Con frecuencia el término *promedio* se utiliza como sinónimo de media. En terminología estadística no son lo mismo. La media se calcula sumando todas las mediciones y dividiéndolas por el número de mediciones realizadas. Un promedio, por su lado, se calcula multiplicando cada una de las mediciones por unos valores concretos, denominados *pesos*, antes de sumarlos. Esta suma se divide entonces por la suma de los pesos. La media es un tipo especial de promedio en el cual el peso de cada medición es igual a 1.

⁴ La desviación estándar (σ) es la raíz cuadrada de la varianza (σ^2). La varianza es igual a la suma de las desviaciones de los datos (x_i) respecto de la media (μ) al cuadrado. Por lo tanto, la desviación estándar poblacional es

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Para demostrar lo que queremos decir con la posición de una distribución, supongamos que la media de la concentración sérica de bilirrubina en la población es de 1,2 mg/dl, en lugar de 0,9 mg/dl. La distribución gaussiana de la concentración sérica de la bilirrubina sería entonces como la que aparece en la figura 26-3.

Observe que la forma general de la distribución de la figura 26-3 no se modifica al cambiar la media, pero la posición de su centro de gravedad se mueve 0,3 mg/dl hacia la derecha. No obstante, si hubiésemos cambiado la dispersión de la distribución de la figura 26-2, su forma se habría modificado sin cambiar su posición. Por ejemplo, compare la distribución de la figura 26-2 con la de la figura 26-4, en la cual se ha cambiado la distribución estándar de 0,3 mg/dl a 0,4 mg/dl.

ESTIMACIÓN E INFERENCIA

En muy pocas ocasiones podemos realizar todas las mediciones posibles en una población. No obstante, podemos calcular valores numéricos para *estimar* el valor de los parámetros de la población mediante el empleo de las mediciones observadas en una muestra extraída de esa población. Estas estimaciones muestrales de los parámetros poblacionales son el fin que persiguen los métodos estadísticos. De hecho, ¡estas estimaciones se denominan *estadísticos*! Un estadístico individual utilizado para estimar el valor de un parámetro poblacional determinado se conoce como *estimación puntual*. Estas estimaciones puntuales son los estadísticos que usamos para resumir grandes cantidades de mediciones en unas pocas manejables.

Hasta el momento, solo hemos considerado la primera finalidad de los métodos estadísticos: sintetizar las observaciones. No obstante, es un paso importante para valorar la influencia del azar en esas observaciones. Como hemos afirmado anteriormente, una muestra es un subgrupo de todas las posibles mediciones de una población. En *todos* los métodos estadísticos se supone que la muestra es un subgrupo *aleatorio* de la población de la que se ha extraído. Aunque los subgrupos aleato-

FIGURA 26-3. Una distribución gaussiana hipotética de la concentración de bilirrubina sérica con una media de 1,2 mg/dl y una desviación estándar de 0,3 mg/dl. La comparación de esta distribución con la de la figura 26-2 ilustra lo que se pretende decir con posiciones diferentes de las distribuciones poblacionales

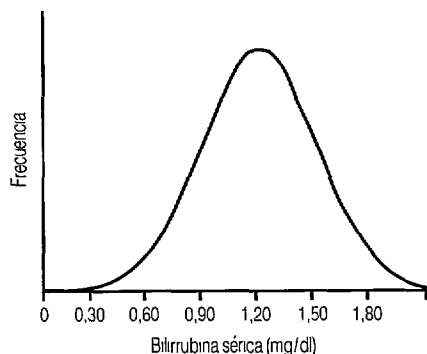
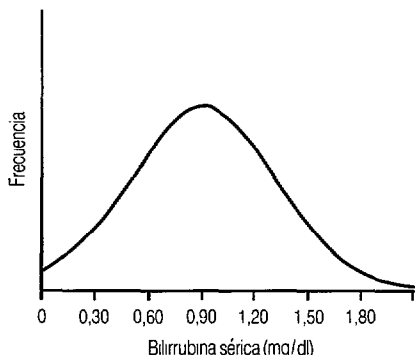


FIGURA 26-4. Una distribución gaussiana hipotética de la concentración de bilirrubina sérica con una media de 0,9 mg/dl. La comparación de esta distribución con la de la figura 26-2 ejemplifica lo que se pretende decir con dispersiones diferentes de las distribuciones poblacionales



rios se pueden obtener por distintos métodos, en *La selección de una prueba estadística* solo consideraremos el más simple de todos ellos (y el más habitual), denominado *muestra aleatoria simple*. En una muestra aleatoria simple, todas las mediciones de la población tienen la misma probabilidad de ser incluidas en la muestra.⁵ Por consiguiente, el azar dicta cuáles de esas mediciones se incluyen realmente en la muestra.

Cuando se estiman los parámetros poblacionales utilizando estadísticos muestrales, la selección aleatoria de las mediciones realmente incluidas en la muestra determina cuánto se aproxima el estadístico muestral al valor real del parámetro poblacional. Lamentablemente, nunca sabemos cuán correctamente un estadístico refleja el valor del parámetro poblacional correspondiente, porque tendríamos que efectuar mediciones en todos los integrantes de la población para conocer los parámetros poblacionales reales. No obstante, lo que podemos saber es cuánto se espera que varíe el estadístico en relación con el valor hipotético del parámetro poblacional sobre la base de la variabilidad del azar entre las muestras aleatorias. Este conocimiento constituye la base de la *inferencia estadística* o de las *pruebas de significación estadística*.

El marco de la inferencia estadística ha sido descrito en la Parte 1 (*Boletín* de julio, pp. 63–71). En ese apartado se señaló que las pruebas de significación estadística se realizan suponiendo que la hipótesis nula es cierta. La hipótesis nula nos proporciona el valor hipotético con el que podemos comparar nuestras estimaciones.

Como se ha comentado en la Parte 1, el “objetivo” en las pruebas de significación estadística es el cálculo del valor *P*.⁶ El valor *P* se calcula a partir de las observaciones de la investigación convirtiéndolas, en primer lugar, a una *distribución estándar*. Utilizamos una distribución estándar, porque los valores *P* se pueden obtener a partir de las tablas estadísticas en cualquier lugar de estas distribuciones. Buena parte de lo que se considera metodología de la estadística tiene que ver con la conversión de las observaciones a una distribución estándar.⁷

En la Parte 1 también comentamos que una alternativa al uso de las pruebas de significación estadística para investigar la influencia del azar en las estimaciones muestrales es el cálculo del *intervalo de confianza* o la *estimación por intervalo*.⁸ Dentro de un intervalo de confianza, tenemos un nivel de confianza determinado (con frecuencia de 95%) de que está incluido el parámetro poblacional.⁹ Generalmente, los intervalos de confianza se calculan modificando mediante el álgebra los cálculos realizados en las pruebas de significación estadística.

Cuando realizamos una prueba de significación estadística o calculamos un intervalo de confianza, podemos usar técnicas *unilaterales* o *bilaterales*. Una prueba de significación estadística *bilateral* o estimación por intervalo se emplea cuando

⁵ En un sentido general, una muestra aleatoria implica que cualquier individuo en la población tiene una probabilidad conocida de ser incluido en la muestra. Aquí limitamos esas probabilidades conocidas a la condición de que sean iguales.

⁶ Recuerde que el valor *P* es la probabilidad de obtener una muestra que sea como mínimo tan distinta de la indicada por la hipótesis nula como la muestra realmente obtenida si la hipótesis nula realmente describe la población. No es, como se supone frecuentemente, la probabilidad de que el azar haya influido sobre las observaciones muestrales. Esa probabilidad es igual a 1 (es decir, estamos seguros de que el azar ha influido en nuestras observaciones).

⁷ Ejemplos de distribuciones estándares son la normal, la de la *t* de Student, la de *ji al cuadrado* y la de la *F*. Estas distribuciones se presentarán en capítulos posteriores.

⁸ Algunas veces, este intervalo se denomina “límites de confianza”. En la terminología estadística, los límites de confianza son los valores numéricos que marcan los límites de un intervalo de confianza.

⁹ En la estadística clásica, una *estimación por intervalo* significa que, si examinamos un número infinito de muestras de un mismo tamaño, un porcentaje determinado (esto es, el 95%) de las estimaciones por intervalo incluirán el parámetro poblacional. Una visión más moderna entre los estadísticos es que esto equivale a suponer que existe una determinada posibilidad (de 95%) de que el valor del parámetro poblacional esté incluido en el intervalo. Esta última interpretación es la que habitualmente tiene interés para el investigador en medicina.

el investigador no está seguro en qué lado del valor del parámetro implicado en la hipótesis nula se encuentra realmente el parámetro poblacional. Esta es la situación habitual, pero en algunas circunstancias se pueden encontrar en la literatura médica pruebas de significación estadística o estimaciones por intervalo *unilaterales*. Una prueba o intervalo de confianza unilateral se aplica cuando el investigador está dispuesto a suponer que conoce la dirección del efecto estudiado y el análisis solo se centra en el examen de la magnitud o de la fuerza de tal efecto.

Para ilustrar la distinción entre las técnicas unilaterales o bilaterales, imaginaremos un ensayo clínico en el que se mide la tensión arterial diastólica en un grupo de individuos antes y después del tratamiento con un nuevo fármaco antihipertensivo. Antes de examinar los datos resultantes de este estudio, podríamos suponer en nuestra hipótesis de estudio que la tensión arterial diastólica disminuye cuando los pacientes toman el medicamento. En otras palabras, supondríamos que es *imposible* que el medicamento aumente la tensión arterial diastólica. Con este supuesto, la prueba de significación estadística o la estimación por intervalo puede ser unilateral y la potencia estadística de nuestro análisis aumentará. Por otro lado, si nuestra hipótesis de estudio es que la tensión arterial diastólica cambiará cuando los pacientes tomen el medicamento, las pruebas de significación o la estimación por intervalo deben ser bilaterales. Esto se debe a que consideramos *posible*, aunque improbable, que el nuevo medicamento antihipertensivo aumente la presión arterial diastólica.

LA SELECCIÓN DE LOS MÉTODOS ESTADÍSTICOS

Centremos ahora nuestra atención en la selección de los métodos estadísticos para analizar los datos de la investigación médica. Antes de seleccionar un método, debemos tomar dos decisiones: 1) cuál es la variable dependiente y cuál la independiente, y 2) qué tipo de datos constituyen cada una de esas variables. En primer lugar, veamos qué queremos decir con variables dependientes e independientes.

Una *variable* es una característica que se mide en un estudio. Por ejemplo, si medimos la edad, podemos hablar de la edad como una de las variables de nuestro estudio. La mayor parte de los métodos estadísticos distinguen entre variables *dependientes* e *independientes*. Así se indican las funciones o el propósito de una variable en un análisis determinado. Por lo general, una serie de variables diseñadas para investigar una hipótesis de estudio solo incluirá una variable dependiente. Esta variable dependiente puede identificarse como la de interés principal o el desenlace principal del estudio. Queremos contrastar hipótesis o hacer estimaciones, o efectuar ambos procedimientos, acerca de la variable dependiente. Por otro lado, en la serie de variables puede que no haya ninguna variable independiente o que se incluya una o más. Las variables independientes determinan las características que es necesario tener en cuenta o las condiciones en que se contrastan las hipótesis o se realizan las estimaciones.

Para ilustrar la distinción entre variables dependientes e independientes, considere un estudio de cohortes en el que se investiga la relación entre el consumo de tabaco y la enfermedad coronaria. Suponga que solo se miden dos variables en cada individuo: consumo de tabaco (frente a no consumo) y enfermedad coronaria (frente a no enfermedad). Para analizar estos datos, primero decidimos que estamos interesados principalmente en estimar o contrastar una hipótesis sobre el riesgo anual de enfermedad coronaria. Por consiguiente, la enfermedad coronaria es la variable dependiente. Además, deseamos comparar el riesgo de enfermedad coronaria entre los fumadores y los no fumadores. Por este motivo, el consumo de tabaco es la variable independiente.

El número de variables independientes determina el tipo de método estadístico que es apropiado para analizar los datos. Por ejemplo, si nos interesara estimar el riesgo anual de enfermedad coronaria en una comunidad sin tener en cuenta el consumo de tabaco o cualquier otra característica de los individuos, aplicaríamos los métodos estadísticos conocidos como *análisis univariantes*. Estas técnicas se aplican a una serie de observaciones que contienen una variable dependiente y ninguna independiente. Para examinar el riesgo de enfermedad coronaria en relación con el hecho de ser fumador, como en el ejemplo anterior, usaríamos los métodos de *análisis bivariante*. Estos métodos se aplican a grupos de observaciones con una variable dependiente y una independiente. Por último, si nos interesara el riesgo de enfermedad coronaria en los individuos de diversas edades, sexo y hábito de fumar, aplicaríamos los métodos de *análisis multivariante* (*multivariable* en inglés).¹⁰ Estos métodos se utilizan para grupos de observaciones que consisten en una variable dependiente y más de una independiente, como la edad, el sexo y el hábito tabáquico. Los métodos multivariantes se aplican con frecuencia para cumplir la tercera finalidad de los métodos estadísticos: ajustar según la influencia de las variables de confusión.

Las investigaciones médicas suelen incluir diversas series o grupos de variables. Por ejemplo, suponga que hemos realizado un ensayo clínico controlado en el cual los sujetos han recibido el fármaco X o un placebo para facilitar su recuperación de una enfermedad determinada. Dado que nos interesa conocer la influencia de la edad y el sexo en la recuperación (porque la edad y el sexo pueden ser variables de confusión), las incluimos en los registros de datos de la investigación. Por lo tanto, nuestro estudio contiene cuatro variables: tratamiento (fármaco X o placebo), recuperación (sí o no), edad y sexo. En el grupo de datos que incluye las cuatro variables, la recuperación sería la variable de interés, es decir, la variable dependiente. El tratamiento, la edad y el sexo serían las variables independientes, que reflejan nuestro interés en analizar la recuperación en relación con el tratamiento específico que ha recibido el sujeto, su edad y sexo. Sin embargo, incluso antes de contrastar hipótesis o de realizar estimaciones sobre la recuperación, probablemente nos interesaría saber si mediante la asignación al azar de los participantes se obtuvieron distribuciones de edad desiguales en los dos grupos de tratamiento. El grupo de variables que nos permitiría comparar las distribuciones de edad incluye la edad como variable dependiente y el tratamiento como variable independiente, ya que la edad es la variable de interés y el grupo de tratamiento, la condición en la que estamos valorando la edad. Por este motivo, la decisión sobre cuál es la variable dependiente y cuál la independiente *depende de la pregunta que se intenta responder*.

TIPOS DE DATOS

Además de caracterizar la función de las variables en el análisis, para seleccionar la prueba estadística debemos determinar el tipo de datos que constituyen las mediciones de cada variable. Con el fin de categorizar los tipos de datos, realizaremos una primera distinción entre datos *continuos* y *discretos*.

¹⁰ Un error habitual en el uso de la terminología estadística es referirse a las técnicas diseñadas para una variable dependiente y varias independientes como análisis *multivariado* (*multivariate* en inglés). Sin embargo, este término se refiere en rigor a las técnicas diseñadas para tratar con *más de una* variable dependiente. El uso de técnicas multivariadas es raro en la investigación médica. No hemos incluido estas técnicas en nuestro diagrama y mencionamos el término en su aplicación más habitual (variables dependientes nominales multivariantes).

Los datos continuos se definen como los que ofrecen la posibilidad de observar alguno de ellos entre un número infinito de valores regularmente espaciados entre dos puntos cualesquiera de su intervalo de medidas. Son ejemplos de datos continuos la tensión arterial, la concentración de colesterol sérico, la edad y el peso. Para cada una de estas variables podemos escoger dos valores cualesquiera e imaginar mediciones intermedias que sería posible observar, al menos, teóricamente, entre esos valores. Podemos considerar, por ejemplo, las edades de 35 y 36 años. Podríamos imaginar que las edades entre los 35 y 36 años se distinguen por el número de días transcurridos desde el 35o. cumpleaños de la persona. Además, podríamos imaginar el número de horas y de minutos que han transcurrido desde el cumpleaños. Teóricamente, no existe un límite de la precisión con que podríamos medir el tiempo. No obstante, observe que no es necesario que los datos continuos tengan un intervalo infinito de posibles valores, sino un número infinito de posibles valores dentro de su intervalo. Este intervalo puede tener, y de hecho lo tiene frecuentemente, un límite superior y uno inferior. La edad es un buen ejemplo. El límite inferior es cero y es difícil imaginar individuos que tengan edades por encima de los 120 años.

Los datos discretos, por otro lado, solo pueden tener un número finito de valores en su intervalo de medidas. Son ejemplos de datos discretos el número de hijos, el estadio de las enfermedades y el sexo. Para cada una de estas variables podemos seleccionar dos valores entre los cuales no es posible imaginar otros valores. Por ejemplo, no podemos imaginar que el número de hijos de una familia se encuentre entre 2 y 3.

En la práctica, a veces no se puede distinguir claramente entre datos continuos y discretos. Esto ocurre porque no existe ninguna variable en la que podamos medir realmente un número infinito de valores.¹¹ Este problema se soluciona al reconocer que, si se puede efectuar un elevado número de mediciones en el intervalo de medidas posibles y si los intervalos entre las mediciones son uniformes, esas mediciones son virtualmente continuas. Sin embargo, esto crea otra fuente de confusión, pues permite que se redefinan como continuos datos que son, incluso teóricamente, discretos. Por ejemplo, el número de cabellos en la cabeza es con certeza un dato discreto: no podemos imaginar un valor entre 99 999 y 100 000 cabellos. Con todo, el número de posibles valores dentro del intervalo del número de cabellos es muy elevado. ¿Podemos considerar esta variable como realmente continua? Sí; para casi todos los fines sería totalmente correcto.

Los datos pueden definirse además por su *escala* de medida. Los datos continuos se miden en escalas, denominadas *escala de razón* y *escala de intervalo*,¹² que se definen por estar constituidas por un intervalo constante o uniforme entre mediciones consecutivas. Algunas mediciones discretas se realizan en una *escala ordinal*. Los datos en una escala ordinal tienen una ordenación o posición específica, como en el caso de los datos continuos, pero no es preciso que el intervalo entre mediciones consecutivas sea constante. Un tipo de variable que se mide habitualmente con una escala ordinal es la clasificación conocida como el *estadio de la enfermedad*. Sabemos que el estadio 2 es más avanzado que el 1, pero no podemos afirmar que la diferencia entre los dos estadios sea la misma que la diferencia entre el estadio 2 y el 3.

¹¹ Por ejemplo, podríamos imaginar, aunque no medir, la tensión arterial en, digamos, picómetros de mercurio. Así que, en realidad, ¡todos los datos son discretos!

¹² La distinción entre la escala de razón y la de intervalo consiste en que la primera incluye el valor cero verdadero mientras que la segunda no. Ciertos tipos de datos discretos, como los recuentos, tienen intervalos uniformes entre las mediciones y, por lo tanto, también se miden mediante escalas de razón o de intervalo. Otros tipos de datos discretos se miden en escalas ordinales o en escalas nominales.

Cuando no se puede aplicar algún tipo de ordenamiento a los datos discretos, decimos que se midieron en una *escala nominal*. Son ejemplos de características medidas con datos discretos en escalas nominales el tratamiento, el sexo, la raza y el color de los ojos. Los datos que tratamos como nominales incluyen mediciones con dos categorías, aunque se pueda considerar que tienen un orden intrínseco, porque uno es claramente mejor que el otro (por ejemplo, vivo y muerto).

Es importante darse cuenta de que el término variable nominal puede causar confusión. En su uso común, una variable nominal es una característica como el sexo o la raza que tiene dos o más categorías potenciales. Sin embargo, desde un punto de vista estadístico, una variable nominal se limita solamente a dos categorías. De este modo, debemos referirnos a la raza o al color de los ojos como datos nominales que requieren más de una variable nominal. El número de variables nominales es igual al número de categorías potenciales menos uno.

Con el fin de seleccionar una técnica estadística o de interpretar el resultado de una técnica, es importante distinguir entre tres categorías de variables:

1. *Variables continuas* (comprenden datos continuos como la edad y datos discretos que contienen un número elevado de posibles valores como el número de cabellos).
2. *Variables ordinales* (comprenden los datos ordinales con un mínimo de tres valores posibles aunque con un número total limitado, como los estadios de los tipos de cáncer).
3. *Variables nominales* (comprenden los datos nominales que no tienen un orden como la raza, y los datos que solo pueden tomar dos valores posibles, como vivo o muerto).

El orden en el que se han enumerado estos tres tipos de variables indica la cantidad relativa de información que cada una contiene. Esto es, las variables continuas contienen más información que las ordinales y estas, más que las nominales. Por esta razón, las variables continuas se sitúan a un nivel más elevado que las ordinales y estas, a un nivel más elevado que las nominales.

Las mediciones de un nivel de información concreto pueden ser *reescaladas* a un nivel inferior. Por ejemplo, la edad (medida en años) es una variable continua. Podríamos reescalarla de forma legítima y transformarla en una variable ordinal al definir a las personas como niños (0–18 años), jóvenes (19–30 años), adultos (31–45 años), adultos maduros (45–65 años) y ancianos (>65 años). Podríamos reescalarla otra vez para convertirla en una variable nominal. Por ejemplo, podríamos dividir las personas en dos categorías: jóvenes y viejas. Sin embargo, no podemos reescalar las variables a un nivel superior al que se midieron realmente.

Cuando reescalamos una medida a un nivel inferior perdemos información. Es decir, tenemos menos detalles sobre una característica si la medimos en una escala nominal que si la medimos en escala ordinal o continua. Por ejemplo, sabemos menos acerca de una persona cuando la identificamos como de edad madura que si decimos que tiene 54 años. Si una persona tuviera 54 años de edad y midiéramos su edad en una escala continua, podríamos distinguir su edad de la de otra persona que tuviera 64 años. Sin embargo, si la edad se registrara en la escala ordinal antes indicada, no podríamos diferenciar la edad de estos individuos.

La pérdida de información que se produce al utilizar mediciones reescaladas en las técnicas estadísticas tiene el efecto de aumentar el error de tipo II, si todo lo demás se mantiene igual. Es decir, reescalar a un nivel inferior reduce la potencia estadística, lo que hace más difícil establecer el nivel de significación estadística y,

en consecuencia, rechazar una hipótesis nula falsa. Por otra parte, reescalando a un nivel inferior se evita la necesidad de aceptar ciertos supuestos, como la uniformidad de los intervalos, que puede ser un requisito para realizar determinadas pruebas estadísticas. En los siguientes capítulos se describirán con mayor detalle varios ejemplos concretos de determinadas pruebas que requieren estos supuestos y de las que permiten evitarlos.

Hasta aquí, hemos revisado los pasos iniciales que deben darse para seleccionar una prueba estadística. Estos pasos son:

1. Identificar una variable dependiente y todas las variables independientes a partir de la pregunta que se intenta responder con el estudio.
2. Determinar si cada variable es continua, ordinal o nominal.

Una vez completados estos pasos, estamos preparados para iniciar el proceso de selección de una prueba estadística.

EL ESQUEMA

Los capítulos restantes de este libro están organizados como ramas de un esquema diseñado para facilitar la selección e interpretación de los métodos estadísticos. Se han incluido la mayor parte de ellos, aunque no todos los que pueden encontrarse en la literatura médica.

Para usar este esquema (figura 26-5), primero se debe determinar cuál es la variable dependiente entre el grupo de variables. Si hay más de una variable dependiente que usted quiere considerar simultáneamente en el mismo análisis, quizá le interese un análisis multivariante para el cual debe consultar a un estadístico. Si su grupo de variables parece contener más de una variable dependiente, es muy probable que los datos planteen más de una hipótesis de estudio. En ese caso, se deben considerar las variables relevantes para una hipótesis de estudio específica.

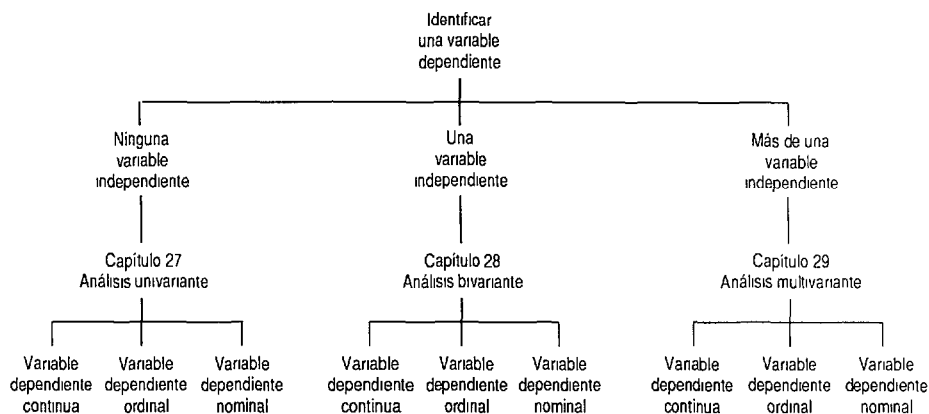
Una vez identificada una sola variable dependiente, el número de variables independientes en la investigación le orientará hacia el capítulo que trata de ese número de variables independientes. Cada capítulo contiene tres grandes divisiones. La primera hace referencia a los grupos de variables en los que la variable dependiente es continua. La segunda se centra en las variables dependientes ordinales y la tercera, en las variables dependientes nominales. Dentro de cada una de estas divisiones se describen las técnicas para variables independientes continuas, ordinales y nominales, cuando se dispone de ellas. El capítulo 30 reúne los esquemas presentados en los capítulos 27, 28 y 29.

RESUMEN

En este capítulo hemos aprendido que los métodos estadísticos utilizados para analizar los datos de una investigación médica tienen tres finalidades. La primera es la de resumir los datos. Las distribuciones de los datos en grandes poblaciones se resumen mediante valores numéricos denominados parámetros. Los valores de estos parámetros poblacionales se estiman a partir de muestras aleatorias mediante estimaciones puntuales denominadas estadísticos.

La segunda finalidad de la estadística es la de tener en cuenta la influencia del azar en las estimaciones puntuales calculadas a partir de las observaciones muestrales seleccionadas al azar de la población. Hay dos enfoques generales para considerar el azar. Uno está constituido por las pruebas de significación estadística. Bajo este enfoque, las observaciones muestrales se comparan con lo que sería de esperar si

FIGURA 26-5. Esquema para identificar el capítulo y la sección en los que se tratan las pruebas estadísticas referentes a un grupo de variables en particular



no existiese una asociación entre variables o una diferencia entre los grupos de la población. Si las observaciones son lo suficientemente inesperadas o no existe una verdadera asociación (o diferencia), rechazamos la hipótesis de que no existe una asociación (o diferencia) en la población. Un enfoque alternativo para considerar el azar es el cálculo de los intervalos de confianza de la estimación puntual. En este caso, podemos suponer con un grado de confianza determinado que el parámetro poblacional se halla incluido en el intervalo de confianza. Aunque las pruebas de significación estadística y la estimación por intervalo son procesos que aparentemente se interpretan de forma distinta, consisten sencillamente en expresiones matemáticas diferentes de un mismo principio.

La tercera finalidad de la estadística es la de ajustar los datos según el efecto de las variables de confusión en nuestras observaciones muestrales. Este objetivo se alcanza mediante el análisis multivariante, que será el tema que nos ocupará en el capítulo 29.

Para cumplir con estas finalidades, debemos seleccionar una técnica estadística apropiada para responder a la cuestión en estudio. Para realizar esta selección, procederemos de la siguiente forma:

1. Decidir cuál es la variable dependiente. Esta será la variable de interés principal en la hipótesis del estudio. Las variables restantes son las variables independientes.
2. Determinar cuántas variables independientes contiene el conjunto de observaciones. Si no existe ninguna, debemos realizar un análisis univariante. Con una variable independiente, el análisis bivalente es el apropiado. Si, por otro lado, la serie contiene más de una variable independiente, usaremos un método multivariante. Recuerde que para los datos nominales, el número de variables es igual al número de categorías potenciales menos una.
3. Definir qué tipo de variable dependiente es la de interés. Si la variable dependiente tiene un número ilimitado de valores uniformemente espaciados, se trata de una variable continua. Si la variable dependiente contiene un número de valores limitado que pueden seguir un orden, se trata de una variable ordinal. Una variable dependiente nominal simplemente identifica la presencia o la ausencia de una condición.

ANÁLISIS UNIVARIANTES

Para analizar un conjunto de mediciones que contiene una variable dependiente y ninguna independiente, los métodos estadísticos utilizados son un tipo de *análisis univariante*. En la literatura médica, el análisis univariante tiene tres usos comunes. El primero se encuentra en estudios descriptivos (por ejemplo, en las series de casos) en los que solo se ha examinado una muestra. Por ejemplo, un investigador puede presentar una serie de casos de una enfermedad determinada describiendo diversas mediciones demográficas y patofisiológicas de los pacientes. El propósito del análisis en ese estudio sería el de explicar la influencia del azar en las mediciones de cada característica. Dado que no existen grupos diferentes de personas para comparar, ni interés en comparar una característica con otra, cada característica de las personas enfermas se considera una variable dependiente en un análisis univariante individual.

El análisis univariante también se utiliza comúnmente cuando se extrae una muestra para incluirla en un estudio. Por ejemplo, antes de hacer la selección aleatoria en un ensayo clínico controlado, puede que sea conveniente realizar mediciones en toda la muestra objeto de estudio. Es decir, podríamos determinar el porcentaje y la media de edad de las mujeres en el grupo seleccionado para muestreo al azar antes de asignarlas al grupo de control o al de estudio. Como en el estudio descriptivo comentado antes, cada característica examinada en la muestra es una variable dependiente en un análisis univariante individual.

Por lo general, en los estudios descriptivos o cuando se examina una sola muestra, el interés se centra en la estimación puntual y por intervalo, en lugar de las pruebas de significación estadística. En el esquema univariante se pueden realizar pruebas de hipótesis, pero en la hipótesis nula debe especificarse un valor para el parámetro poblacional. Muchas veces esto no se puede hacer en el análisis univariante. Por ejemplo, es difícil imaginar qué valor se tomará como hipótesis de la prevalencia de hipertensión entre los individuos de una comunidad determinada.¹ Sin embargo, en la tercera aplicación del análisis univariante es fácil imaginar ese valor hipotético. Este es el caso en el que una medición, como la tensión arterial diastólica, se realiza dos veces en el mismo individuo o en uno muy semejante y el interés se centra en la diferencia entre las dos mediciones. En esta aplicación, es lógico imaginar una hipótesis nula que afirme que la diferencia entre las dos mediciones es igual a cero. De este modo, la diferencia entre las mediciones de la tensión arterial diastólica es la variable dependiente. Aunque la diferencia, por su misma naturaleza, es una comparación entre grupos, las diferencias en sí mismas *no* son comparadas con ningún grupo. Por lo tanto, no existe ninguna variable independiente. Cuando se comparan dos mediciones en un mismo individuo o en individuos muy semejantes, estamos tratando con un problema univariante. Por eso, en una investigación que emplea datos apareados y en la que cada par

¹ A primera vista puede parecer que la hipótesis nula sería que la prevalencia en una comunidad determinada es igual a la prevalencia en otra comunidad o a la prevalencia estimada en otro estudio. Sin embargo, es importante recordar que el valor propuesto como parámetro poblacional tiene que ser *conocido sin error*. Esto no sería cierto a no ser que todos los miembros de la comunidad que se compara se incluyeran en el cálculo de la prevalencia.

constituye una observación, los datos se analizan usando métodos univariantes. Los pares pueden consistir en datos de un individuo o de dos individuos que se aparean antes de analizar los datos.

VARIABLE DEPENDIENTE CONTINUA

Comenzaremos a analizar la figura 27-1 preguntando ¿cuál es el aspecto de la distribución poblacional que nos interesa, su posición o su dispersión?² A continuación es preciso considerar la estimación puntual que puede emplearse para representar ese aspecto de la distribución poblacional.

En el análisis univariante de una variable dependiente continua se acostumbra suponer que los datos provienen de una población con una distribución gaussiana. Por lo tanto, la media se usa habitualmente como medida de posición. La dispersión de las distribuciones gaussianas se mide mediante la desviación estándar o, alternativamente, por el cuadrado de la desviación estándar, denominado *varianza*. Para fines de análisis, tanto la varianza como el coeficiente de variación —descrito más adelante— se usan para medir la dispersión de los datos de la distribución poblacional. Por último, cada diagrama clasificará la categoría general de las pruebas estadísticas que se emplean más frecuentemente para calcular los intervalos de confianza o para contrastar las hipótesis estadísticas.

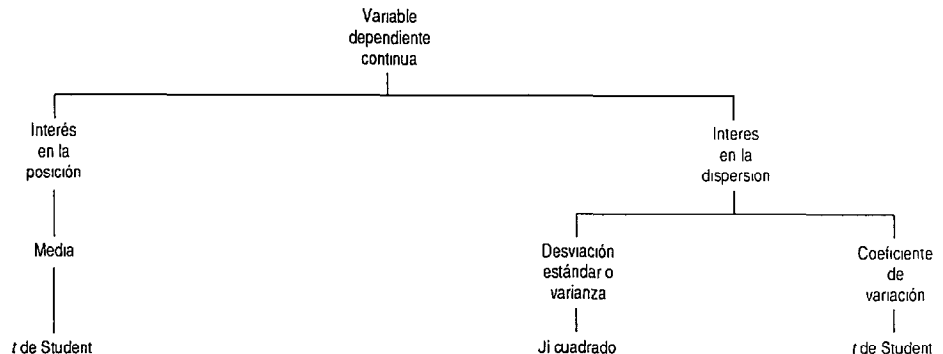
En el capítulo 26 aprendimos que los primeros pasos para escoger una técnica estadística son:

1. Decidir cuál es la variable dependiente.
2. Determinar cuántas variables independientes, si las hubiera, contiene el grupo de observaciones.
3. Definir el tipo de variable dependiente como continua, ordinal o nominal.

A estos pasos, ahora añadimos el siguiente:

4. Seleccionar el parámetro de la distribución poblacional sobre el que desearíamos contrastar hipótesis o efectuar estimaciones. En otras palabras, ¿nos interesa la posición o la dispersión?

FIGURA 27-1. Esquema para seleccionar un método estadístico univariante para variables dependientes continuas (continuación de la figura 26-5)



² En los siguientes capítulos centraremos nuestro interés en la posición.

Si seguimos estos pasos en la figura 27-1, observamos que nos conducen al nombre de un tipo general de pruebas estadísticas. Estas pruebas suelen ser apropiadas tanto para determinar la significación estadística como para calcular los intervalos de confianza.

Interés en la posición

Como se ha afirmado anteriormente, la media muestral es una estimación de la posición de la media poblacional. A menudo, la media poblacional es el parámetro que intentamos estimar. Para calcular el intervalo de confianza de la media de una muestra, la *distribución de la t de Student* es la más frecuentemente empleada. La distribución de la *t* de Student es una distribución estándar en la cual se transforman las medias de variables dependientes continuas para facilitar el análisis. Esta distribución es parecida a la gaussiana, pero requiere de un parámetro adicional conocido como *grados de libertad*. El propósito de los grados de libertad en la distribución de la *t* de Student es reflejar el papel del azar en la estimación de la desviación estándar.³

La distribución de la *t* de Student nos permite construir los intervalos de confianza a partir de la media observada y de su error estándar. En la Parte 3 se señaló que el error estándar de una media disminuye a medida que aumenta el tamaño de la muestra. De forma más precisa, el error estándar es igual a la desviación estándar dividida por la raíz cuadrada del tamaño de la muestra.

El error estándar se emplea en la distribución de la *t* de Student para calcular las estimaciones por intervalo de las medias de las variables continuas. El intervalo de confianza de una media es igual a la estimación muestral de la media + el valor de la *t* de Student para el nivel de confianza deseado y multiplicado por el error estándar. Para una estimación bilateral con un nivel de confianza de 95%, el valor de la *t* de Student es aproximadamente igual a 2 si las muestras contienen 20 casos o más. Sumando y restando a la estimación puntual de la media un valor igual al doble del error estándar, se puede obtener un intervalo de confianza *aproximado*. Es decir, la media poblacional se encuentra en el intervalo comprendido entre la media muestral \pm dos errores estándares, con un nivel de confianza de 95%.⁴ Por ejemplo, si leemos en un informe de investigación que la media \pm el error estándar de la concentración de colesterol sérico en una muestra es igual a 150 ± 30 mg/dl, podemos tener un nivel de confianza de 95% de que la media poblacional se encuentra dentro del intervalo aproximado comprendido entre 120 y 180 mg/dl.

Como se mencionó anteriormente, en el análisis univariante existe una situación especial en la que se pueden aplicar las pruebas de significación estadística. El caso más frecuente es el de un estudio en el que una variable dependiente continua se mide dos veces en el mismo individuo. Por ejemplo, podríamos medir la tensión arterial antes y después de que un paciente reciba un medicamento antihipertensivo. Si lo que realmente nos interesa no son las mediciones antes y después del tratamiento,

³ Al utilizar la distribución de la *t* de Student para realizar estimaciones por intervalo de las medias, se reconoce el hecho de que la desviación estándar se estima a partir de la muestra. Es decir, no se conoce con precisión la desviación estándar.

⁴ De forma similar, se pueden estimar otros intervalos de confianza considerando múltiplos del error estándar. Aproximadamente dos tercios de las medias muestrales posibles se encuentran dentro de un error estándar de la media poblacional. Más de 99% de las posibles medias muestrales se encuentran dentro del intervalo de la media poblacional \pm tres errores estándares. Sin embargo, es importante recordar que, cuando aplicamos estas interpretaciones a los intervalos de confianza o a sus aproximaciones, estamos suponiendo que la población de todas las medias posibles tiene una distribución gaussiana.

sino la diferencia entre las mediciones, nos encontramos frente a un *diseño apareado*.⁵ Este es un problema univariante, dado que la variable dependiente es la diferencia entre las mediciones y no existe una variable independiente. Mediante un diseño apareado, hemos tratado de eliminar la influencia de la variación entre los sujetos en la medición inicial o *de base*.

De la misma manera que se emplea en otros análisis univariantes, la distribución *t* de Student se emplea para contrastar hipótesis o para realizar estimaciones por intervalo para los datos continuos a partir de un diseño apareado. Aunque las pruebas estadísticas utilizadas para analizar los datos de un diseño apareado no son distintas de otras pruebas univariantes, en los textos introductorios de estadística frecuentemente se tratan por separado. En estos casos, la prueba utilizada para examinar la diferencia entre las medias de los datos de un diseño apareado se denomina *prueba de la t de Student para datos apareados*.

Más que la media de la muestra \pm el error estándar, con frecuencia vemos los datos univariantes presentados como la media de la muestra \pm la desviación estándar. La media muestral \pm el error estándar informa del nivel de confianza que podemos tener en nuestra estimación de la media poblacional. El error estándar es un indicador de la *dispersión de las medias muestrales* que podrían obtenerse extrayendo una muestra de la población. Sin embargo, la media de la muestra \pm la desviación estándar plantea una cuestión distinta. La desviación estándar de los datos de la muestra estima la *dispersión de las mediciones* en la población. Aproximadamente, el 95% de los valores de una población se encuentran dentro del intervalo de la media poblacional \pm dos desviaciones estándares.⁶ Por lo tanto, cuando aplicamos una prueba estadística univariante a una variable dependiente continua, podemos estar interesados tanto en la estimación de la posición de la media poblacional y, por ese motivo, en su error estándar, como en la descripción de la dispersión de los valores y, por consiguiente, en la desviación estándar.

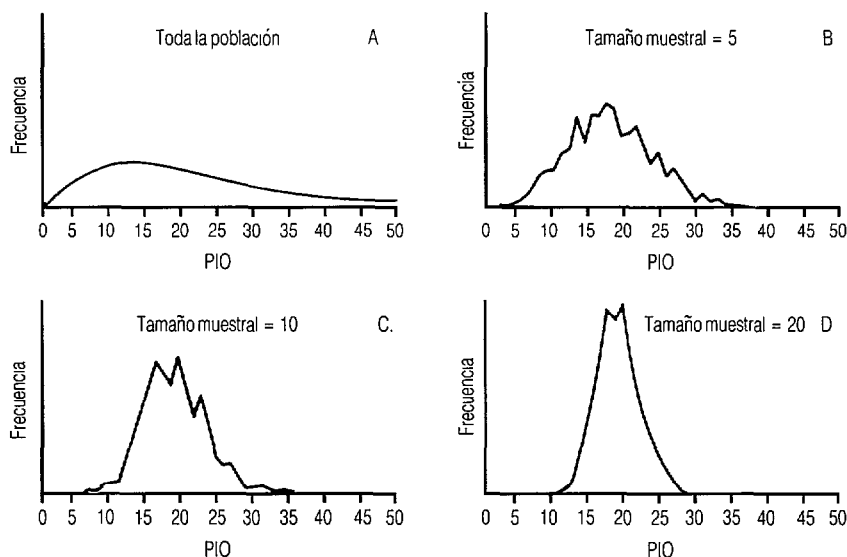
Para ilustrar cómo se escoge entre la presentación de la media \pm la desviación estándar y la media \pm el error estándar, imaginemos un estudio en el que se describe una serie de casos de una enfermedad determinada. Supongamos que una de las variables medidas en esos pacientes es la concentración del colesterol sérico. Si el objetivo del estudio es estimar los valores de la concentración del colesterol sérico que se podrían observar en los pacientes *individuales* con esa enfermedad, se debe presentar la desviación estándar, dado que estamos interesados en la dispersión de los datos poblacionales. Si, por otro lado, el propósito del estudio es estimar la media de la concentración del colesterol sérico de un *grupo* de pacientes con la enfermedad, se debe presentar el error estándar (o la estimación por intervalo), pues estamos interesados en la dispersión de las medias muestrales obtenidas al azar de la población.

Es importante entender la diferencia entre los supuestos que realizamos cuando interpretamos la media \pm el error estándar y la media \pm la desviación estándar. Cuando utilizamos el error estándar, suponemos que las medias de las muestras obtenidas al azar de la población siguen una distribución gaussiana. En el caso de la media \pm la desviación estándar, suponemos que los datos poblacionales por sí mismos

⁵ Otro diseño apareado sería el correspondiente a una variable dependiente continua medida en dos individuos apareados que sean similares en las características compartidas que se considera posible que influyan en la magnitud de la variable dependiente.

⁶ Asimismo, aproximadamente dos tercios de los datos poblacionales se encuentran dentro del intervalo formado por la media \pm una desviación estándar y más de 99%, dentro del intervalo de la media \pm tres desviaciones estándares. Para aplicar estas interpretaciones debemos suponer que los datos poblacionales siguen una distribución gaussiana.

FIGURA 27-2. Demostración del teorema central del límite. Cuando medimos la tensión intraocular en muchos individuos (A) observamos que la distribución de las mediciones individuales no es gaussiana. A pesar de ello, la distribución de la media de la presión intraocular tiende a seguir una distribución gaussiana (B-D). Esta tendencia aumenta a la par que el tamaño muestral



siguen una distribución gaussiana. A menudo este supuesto será cierto para la media \pm el error estándar, como veremos, si escogemos muestras suficientemente grandes. Sin embargo, el supuesto muchas veces no será cierto para la media \pm la desviación estándar.

Si los datos poblacionales siguen una distribución gaussiana, las medias de las muestras de esa población también seguirán una distribución gaussiana. Incluso cuando los datos poblacionales no siguen una distribución gaussiana, las medias de un elevado número de muestras obtenidas mediante muestreos aleatorios repetidos de la misma población a la larga seguirán una distribución gaussiana (figura 27-2). La probabilidad de que las medias sigan una distribución gaussiana aumenta a la par que el número de observaciones en cada muestra. Este importante fenómeno se conoce como el *teorema central del límite* y explica el interés de los estadísticos tanto en las medias como en la distribución gaussiana. También les permite a los investigadores médicos emplear los métodos estadísticos que suponen una distribución gaussiana para analizar los valores de las medias obtenidas de poblaciones en las que los datos no siguen una distribución gaussiana. Esto supone una gran ventaja, ya que muchas de las variables de interés en medicina provienen de poblaciones en las cuales las distribuciones de los datos no son gaussianas.

Interés en la dispersión

Con mucho, la media es el parámetro poblacional que se estima con mayor frecuencia en el análisis univariante de las variables continuas. Sin embargo, este no es el único parámetro que podemos estimar con ese tipo de datos y no es siempre el que mejor refleja nuestro interés por una serie de observaciones. Quizá nos in-

terese la dispersión de las mediciones en la población. En este caso, nuestro interés se centra en la varianza o, de forma equivalente, en su raíz cuadrada: la desviación estándar de la población.

Cuando queremos obtener una medida de posición de la población de la cual hemos extraído una serie de observaciones univariantes, generalmente estimamos esa posición con la media de la muestra. El error estándar refleja la dispersión de las medias de la muestra. Empleamos la distribución de la *t* de Student para contrastar hipótesis estadísticas o para realizar estimaciones por intervalo de la media poblacional. Por otro lado, cuando nos interesa la dispersión de los datos de la población por sí mismos, estimamos la desviación estándar o la varianza de la población a partir de nuestras observaciones muestrales. Si deseamos contrastar hipótesis estadísticas o construir intervalos de confianza de la varianza poblacional, empleamos la *distribución de ji al cuadrado*. Sin embargo, el uso de la varianza o de la desviación estándar puede inducir a error si deseamos comparar la dispersión entre grupos distintos. Examinaremos esta situación y una solución habitual.

Una de las propiedades teóricas de los datos que siguen una distribución gaussiana es que la desviación estándar y la media son independientes. Es decir, para una media determinada, cualquier desviación estándar es igualmente probable. En la práctica, esto no ocurre con frecuencia. Por ejemplo, considere los pesos corporales desde el nacimiento hasta los 5 años de edad (cuadro 27-1). Queda claro que la variación del peso aumenta con la edad, así como el propio peso. Sin embargo, la asociación entre la media y la desviación estándar hace difícil comparar medidas de dispersión correspondientes a diferentes pesos medios. Por ejemplo, las variaciones de un kilogramo entre lactantes representan una variabilidad mucho mayor para su tamaño que una variación de un kilogramo en niños de 5 años de edad.

Una solución sencilla para este problema consiste en dividir la desviación estándar por la media con el fin de "ajustar" los datos según las diferencias entre las medias. Si multiplicamos esta razón por 100, obtenemos lo que se conoce como el *coeficiente de variación*. En el cuadro 27-2 se presentan los coeficientes de variación de los pesos corporales de niños varones.

CUADRO 27-1. Medias y desviaciones estándares del peso corporal (niños)

| Edad (años) | Peso (kg) | |
|-------------|-----------|---------------------|
| | Media | Desviación estándar |
| Nacimiento | 3,50 | 0,53 |
| 1 | 10,20 | 1,01 |
| 5 | 18,50 | 2,17 |

(Fuente: Smith DS. *Growth and its disorders*. Philadelphia: Saunders; 1977.)

CUADRO 27-2. Medias y coeficientes de variación del peso corporal (niños)

| Edad (años) | Peso (kg) | |
|-------------|-----------|--------------------------|
| | Media | Coeficiente de variación |
| Nacimiento | 3,50 | 15,1% |
| 1 | 10,20 | 9,9% |
| 5 | 18,50 | 11,7% |

(Fuente: Smith DS. *Growth and its disorders*. Philadelphia: Saunders; 1977.)

El examen de las variaciones absolutas de los pesos, estimadas mediante la desviación estándar, sugiere que la menor variación se observa entre los recién nacidos (cuadro 27-1). Sin embargo, esta variación se da entre niños que, como promedio, pesan menos. La variación del peso *en relación con la media* del peso en cada grupo, tal como muestran los coeficientes de variación, sugiere precisamente lo contrario (cuadro 27-2). La variación del peso al nacer en relación con el peso total al nacer es mayor que en cualquier otra edad considerada.

Por este motivo, el coeficiente de variación es una medida útil para examinar la dispersión relativa de las variables dependientes continuas cuando se cree que la media y la desviación estándar no son independientes y queremos comparar estimaciones univariantes de dispersión. En los intervalos de confianza y las pruebas de hipótesis estadísticas del coeficiente de variación se utiliza la distribución de la *t* de Student.

VARIABLE DEPENDIENTE ORDINAL

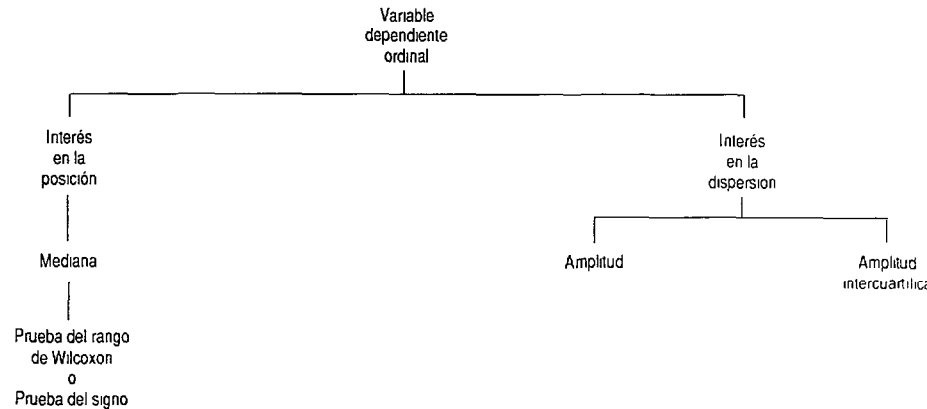
Los métodos estadísticos univariantes para las variables dependientes ordinales se presentan en la figura 27-3.

A diferencia de las variables continuas, con las variables ordinales no suponemos una distribución concreta de los datos poblacionales, tal como la distribución gaussiana. Los métodos utilizados para las variables ordinales se denominan por este motivo de *distribución libre o no paramétricos*. Es importante darse cuenta de que estos métodos no están libres de supuestos. Por ejemplo, seguimos suponiendo que nuestra muestra es representativa de alguna población de interés.

Interés en la posición

Dado que no suponemos una distribución determinada de los datos medidos en una escala ordinal, no podemos estimar parámetros poblacionales que sintetizen la distribución. No obstante, es posible que nos interese describir la posición de los datos ordinales en una escala continua. Eso lo podemos hacer mediante la *mediana*. La mediana es el punto medio de una serie de datos, seleccionada de forma tal que la mitad de los valores sean más altos y la otra mitad más bajos que la mediana.

FIGURA 27-3. Esquema para seleccionar un método estadístico univariante para una variable dependiente ordinal (continuación de la figura 26-5)

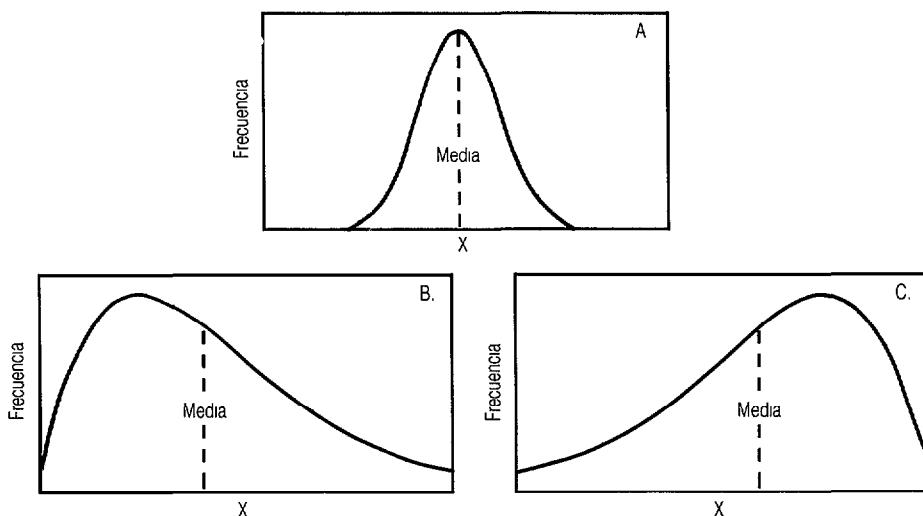


La mediana no tiene una distribución poblacional teórica como medida de su posición, pero puede utilizarse como una estimación *robusta*⁷ de la media de una distribución gaussiana. La mediana soslaya un supuesto que realizamos cuando calculamos la media: que los intervalos entre las mediciones de una distribución son uniformes y conocidos. Como la mediana se calcula empleando solamente el rango relativo u orden de las mediciones, la estimación de la mediana sería la misma independientemente de que los intervalos sean conocidos y uniformes o no. Por lo tanto, podemos usar la mediana para estimar la media de una población de datos continuos. Esto se lleva a cabo organizando las observaciones muestrales en orden relativo. De este modo, los datos continuos se convierten a una escala ordinal mediante la sustitución de los rangos por las observaciones reales.

En sentido estricto, la mediana puede emplearse como una estimación de la media poblacional solo cuando la distribución de la población es simétrica. Si esto es cierto, la media y la mediana poblacionales tienen el mismo valor (figura 27-4). No obstante, aunque la distribución poblacional sea simétrica, es posible que las observaciones obtenidas en una muestra de esa población sean, sin lugar a dudas, asimétricas. Un motivo habitual de esa asimetría es la posibilidad de incluir *valores extremos o aislados (outliers)* en la muestra. Estos valores extremos se producen en la población con muy poca frecuencia. En ocasiones, una muestra incluirá uno o más de estos valores extremos. Cuando esto sucede, las observaciones muestrales sugieren que esos valores extremos han aparecido con una frecuencia mayor de la que realmente tienen en la población.

Debido a que la media es el "centro de gravedad" de una distribución, su valor es influido más por los valores extremos que por los cercanos al centro

FIGURA 27-4. Posición de la media en una distribución simétrica (A) y en distribuciones asimétricas (B,C). X indica la posición de la mediana



⁷ En términos estadísticos, una estimación robusta es aquella que no se ve sustancialmente influida por desviaciones menores de los supuestos de la prueba.

de la distribución. Por consiguiente, en las muestras que incluyen valores extremos, la media muestral puede ser bastante distinta de la poblacional. La mediana muestral, por su lado, es *resistente* a aquellos valores extremos. Es decir, los valores extremos tienen el mismo impacto sobre la mediana que los valores que se encuentran cerca del centro de la distribución muestral. Por lo tanto, paradójicamente, cuando una muestra de una distribución poblacional simétrica incluye valores extremos, la mediana muestral es un estimador de la media poblacional mejor que la media muestral.

El uso de la mediana para estimar la media poblacional constituye, sin embargo, un inconveniente. Dado que la mediana se basa solamente en la clasificación relativa de las observaciones, contiene menos información que la media. Siempre que utilizamos menos información al aplicar métodos estadísticos corremos un riesgo más elevado de cometer un error de tipo II. En otras palabras, la probabilidad de no poder rechazar una hipótesis nula incorrecta es más alta. Solo vale la pena correr ese riesgo cuando hay razones para sospechar que la información excluida crearía otros errores más graves si se incluyera en el análisis de los datos.

Aunque la mediana se emplea como una estimación robusta y resistente de la media poblacional, es importante recordar que también es por derecho propio una medida legítima de la posición de una distribución. Por ejemplo, si una distribución poblacional es asimétrica, podría interesar menos su centro de gravedad o media que su punto medio o mediana.

Si nos interesa contrastar la hipótesis nula de que la mediana es igual a cero en un análisis univariante, podemos emplear tanto la *prueba del rango con signo de Wilcoxon* como la *prueba del signo*. Habida cuenta de que la mediana no es un parámetro de ninguna distribución determinada, en general no podemos construir un intervalo para ese parámetro. Sin embargo, cuando se emplea la mediana como estimación robusta y resistente de la media poblacional, es correcto realizar una estimación por intervalo de esa media. Para esta estimación se dispone de métodos basados en la prueba del rango con signo de Wilcoxon y en la prueba del signo.⁸

Interés en la dispersión

Como ocurre con la media muestral, el cálculo de la desviación estándar supone que los intervalos entre los valores son conocidos y uniformes. El cálculo de la desviación estándar está influido en gran medida por los valores extremos. Como alternativa, en los artículos de investigación frecuentemente se presenta como medida de dispersión el *recorrido (range)* (diferencia entre el valor más alto y el más bajo). Aunque el recorrido es útil para describir la dispersión de un conjunto de observaciones muestrales, no es una buena estimación de la dispersión de los datos poblacionales. Esto se debe al hecho de que los valores de los extremos de la mayor parte de las distribuciones poblacionales raramente se observan en las poblaciones y, por este motivo, tampoco en las muestras. El recorrido se calcula a partir de esos extremos, así que el recorrido calculado en una muestra subestima el recorrido poblacional casi con toda seguridad. Por eso, según se reduce el tamaño muestral, la probabilidad de observar valores extremos también decrece. El resultado es que las estimaciones muestrales del recorrido varían directamente con el tamaño de la muestra.

⁸ Del mismo modo, se podría calcular una estimación robusta y resistente de la desviación estándar (descrita más adelante) y emplear la distribución de la *t* de Student para construir un intervalo de confianza de la media poblacional.

Como alternativa, se puede utilizar el *recorrido intercuartílico* (*interquartile range*) para describir la dispersión de una muestra de observaciones, así como para estimar la dispersión en la población. Los cuartiles dividen una distribución en cuatro partes que contienen el mismo número de observaciones, de la misma forma que la mediana divide una distribución en dos partes iguales. El intervalo entre el valor de los datos que se encuentran un cuartil por debajo de la mediana y un cuartil por encima de la mediana se conoce como recorrido intercuartílico. Dentro de ese intervalo o recorrido se encuentran la mitad de los datos muestrales. Dado que el recorrido intercuartílico no depende de los valores extremos de una distribución, es mucho menos dependiente del tamaño de la muestra que el recorrido.

En una distribución gaussiana, dos tercios de los valores poblacionales se encuentran en el intervalo comprendido por la media \pm una desviación estándar. Por lo tanto, en una distribución gaussiana, la media poblacional \pm $\frac{2}{3}$ del recorrido intercuartílico se puede considerar una estimación robusta y resistente de la media \pm una desviación estándar. Si nos preocupa el supuesto de los intervalos conocidos y uniformes o si la muestra contiene valores extremos de validez cuestionable, podemos estimar la desviación estándar poblacional calculando los dos tercios del recorrido intercuartílico en lugar de usar la desviación estándar calculada a partir de los datos muestrales.

No se realizan pruebas de significación estadística ni cálculo de los intervalos de confianza del recorrido o del recorrido intercuartílico. Por otro lado, si el recorrido intercuartílico se emplea para estimar la desviación estándar poblacional, podemos contrastar una hipótesis estadística o calcular un intervalo de confianza. En ese caso, el método sugerido para la medida de la dispersión podría utilizarse para las variables dependientes continuas.

VARIABLE DEPENDIENTE NOMINAL

Como indica el término, una *variable dependiente nominal* consiste solamente en el nombre de una condición determinada. Además, recuerde que hemos limitado los datos nominales a indicadores de que la condición existe o, por defecto, no existe. Ejemplos de las variables dependientes nominales incluyen vivo o muerto, curado o no curado y enfermo o sano. La cantidad de información contenida en una variable dependiente aislada es bastante limitada, en comparación con la que contienen las variables dependientes continuas, como la edad, o las ordinales, como el estadio de la enfermedad.

Cuando utilizamos variables dependientes nominales solo es necesario referirnos a medidas de posición. Esto puede sorprender, dado que, cuando considerábamos las variables dependientes continuas u ordinales, discutimos la importancia de las estimaciones de la dispersión y de la posición. En las variables dependientes continuas, la dispersión constituye una cuestión importante, porque frecuentemente se supone que siguen una distribución poblacional gaussiana caracterizada, en parte, por la independencia entre la posición y la dispersión. Esto equivale a decir que, para una distribución gaussiana, el conocimiento de la media no nos dice nada acerca de cuál puede ser la varianza de la distribución. Para una media determinada, son posibles infinitas varianzas. Esto *no* es verdad para las distribuciones aplicables a las variables nominales. Antes bien, esas distribuciones tienen medidas de dispersión que dependen totalmente de las medidas de posición (lo cual significa que pueden calcularse a partir de las medidas de posición o son iguales a un valor constante). Por eso, una vez que conocemos la medida de posición, sabemos o podemos calcular la medida de dispersión.

El método estadístico univariante específico que utilizamos para analizar una variable dependiente nominal (figura 27-5) varía según se trate de una proporción como la prevalencia o de una tasa como la incidencia. Veamos en primer lugar, los métodos aplicables a las proporciones.

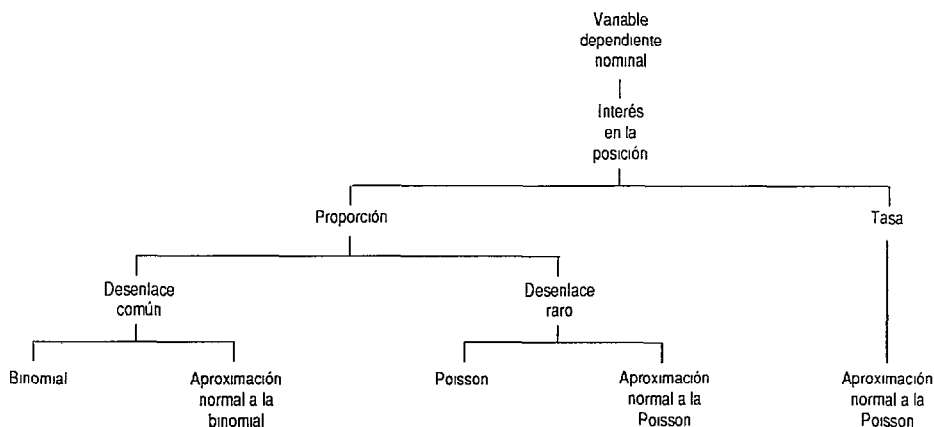
Interés en las proporciones

Para cada medición u observación de una variable compuesta por datos nominales, solo determinaremos la presencia o la ausencia de la condición en estudio. Por ejemplo, podemos determinar si un individuo de una muestra tiene o no una enfermedad concreta. En una muestra constituida por más de una observación podemos estimar la *frecuencia* o el número de veces que la condición ocurre en la población. Por ejemplo, podemos estimar el número de personas que tienen una enfermedad en la población. Más a menudo esa frecuencia nos interesa en relación con el número de observaciones en la muestra. Si dividimos el número de veces que se observa una condición en una muestra por el número de observaciones, estamos calculando la *proporción* de observaciones en la muestra que tienen esa condición. Una proporción calculada a partir de las observaciones muestrales es una estimación puntual de la proporción de la población con la condición. Una forma equivalente de interpretar la proporción de la muestra es estimar la *probabilidad* de la presencia de la condición en la población. Dos proporciones o probabilidades que se calculan habitualmente en la investigación médica son la prevalencia y el riesgo. Estas medidas se comentan en la Parte 1 y en la Parte 3.

Las probabilidades no siguen una distribución gausiana. Se supone que siguen una distribución *binomial* o una de *Poisson*. Se puede aplicar una distribución binomial a toda probabilidad calculada a partir de datos nominales que cumplan los siguientes criterios: 1) la probabilidad de que cualquier observación obtenida mediante un muestreo aleatorio pertenezca a una categoría determinada, denominada *condición nominal*, es la misma para cada observación y 2) las observaciones son independientes entre sí. *Independiente* quiere decir que el resultado de una observación no influye en el resultado de otra.

Una distribución de Poisson es un caso especial de la distribución nominal en la cual el suceso nominal observado, como la muerte o la enfermedad, es

FIGURA 27-5. Esquema para seleccionar un método estadístico univariante para una variable dependiente nominal (continuación de la figura 26-5)



muy infrecuente y el número de observaciones es elevado. El cálculo de la distribución de Poisson es más sencillo que el de la binomial. En general, constituye una buena aproximación a la distribución binomial cuando el número de individuos observado con la condición es 5 o menos y, además, el número de individuos en la muestra es 100 o más.

Las pruebas de significación estadística y el cálculo de los intervalos de confianza de las distribuciones binomial y de Poisson resultan difíciles si deseamos utilizar técnicas *exactas* que realmente usen las distribuciones de Poisson o binomial. Afortunadamente, muchas veces no nos vemos en la necesidad de usar esas técnicas.

Es mucho más sencillo calcular los intervalos de confianza o realizar las pruebas de significación estadística para variables dependientes nominales cuando, en ciertas condiciones, se puede realizar una aproximación a las distribuciones binomial o de Poisson mediante la distribución gaussiana. Podemos utilizar una aproximación gaussiana, casi siempre denominada *aproximación normal*, a las distribuciones binomial o de Poisson cuando el número de individuos con la condición es mayor de 5 y el número de observaciones es mayor de 10.⁹

Interés en las tasas

En la terminología estadística se reserva el término *tasa* para hacer referencia a una razón que incluya una medida del tiempo en el denominador, en contraposición con el término *proporción*, que solo incluye el número total de observaciones en el denominador. La medida de interés más habitual en la investigación médica que cumple la definición de tasa es la incidencia.

Para ilustrar esta distinción, imagine que hemos observado 100 personas que, al inicio de nuestro período de observación, no tenían cierta enfermedad. A los tres años, 30 de las 100 habían enfermado. Si estuviéramos interesados en conocer la probabilidad de que una persona seleccionada al azar de la población de la que se ha extraído la muestra desarrolle esa enfermedad en un período de tres años, calcularíamos la proporción trianual o el riesgo de padecer la enfermedad dividiendo 30 por 100 = 0,30. Sin embargo, si estuviéramos interesados en la *tasa* con la que aparecen nuevos casos de la enfermedad en la muestra de población, calcularíamos la incidencia de la enfermedad como $30/(100 \times 3) = 0,10$ por año. Observe que las probabilidades no tienen unidades y que las tasas se expresan en unidades de 1/tiempo o de sucesos por unidad de tiempo.

Dado que las enfermedades por lo común se producen de forma infrecuente por unidad de tiempo, en el análisis univariante muchas veces se supone que las tasas siguen una distribución de Poisson. Al igual que sucede con las proporciones, es posible aplicar técnicas exactas a las tasas, pero habitualmente las pruebas de significación estadística y la construcción del intervalo de confianza se basan en la aproximación normal. De este modo, se emplean las mismas técnicas para las tasas y las probabilidades, excepto cuando se realizan pruebas de significación estadística y estimaciones por intervalo, para las cuales se emplea la distribución de Poisson o su aproximación normal.

⁹ En la aproximación normal a la distribución de Poisson o a la binomial, solo necesitamos estimar la probabilidad de observar un suceso, dado que el error estándar se calcula a partir de esa probabilidad. Esto difiere de la distribución gaussiana para variables dependientes continuas, en la cual debemos realizar estimaciones separadas para la posición y para la dispersión. Como resultado, no es necesario o, de hecho, apropiado utilizar la distribución de la *t* de Student para tener en cuenta, mediante los grados de libertad, la precisión con que se haya estimado la dispersión. En su lugar, se emplea la distribución normal estándar.

RESUMEN

En este capítulo hemos presentado solamente las técnicas univariantes. Estos métodos se emplean cuando un grupo de observaciones contiene una variable dependiente y ninguna independiente. En su mayor parte, el análisis univariante se centra en el cálculo de los intervalos de confianza más que en las pruebas de hipótesis estadísticas. Una excepción a esta regla es la medición de los valores de una variable dependiente continua dos veces en los mismos individuos o en sujetos muy semejantes. En este caso, la variable dependiente es la diferencia entre dos mediciones. Para contrastar la hipótesis nula de que la diferencia es igual a cero, suele emplearse una prueba de significación estadística para datos apareados.

Durante la presentación del análisis univariante de las variables dependientes continuas, hemos examinado diversos principios importantes del análisis de datos continuos. Uno de ellos, el teorema central del límite, nos ayudó a entender por qué las pruebas estadísticas para las medias se basan tan frecuentemente en la distribución gaussiana. Este teorema afirma que las medias tienden a seguir una distribución gaussiana, aunque no la sigan en la población de la que proceden.

Otro principio importante es la distinción entre dos medidas de dispersión: la desviación estándar y el error estándar. La desviación estándar es una medida de la dispersión de los datos en la población. Utilizamos la media más y menos la desviación estándar cuando nos interesa comunicar la variabilidad estimada de las observaciones individuales. Por su lado, el error estándar es una medida de la dispersión de las medias de las muestras extraídas de una población. Utilizamos el error estándar cuando nos interesa mostrar la diferencia esperada entre las medias muestrales. Para contrastar hipótesis estadísticas y para construir los intervalos de confianza de las medias empleamos el error estándar.

Al realizar las pruebas de significación estadística y al construir intervalos de confianza para el análisis univariante de las variables dependientes continuas se supone que la población de la que se extrae la muestra sigue una distribución gaussiana. Cuando dudamos que sea así, podemos transformar la variable dependiente continua a una escala ordinal. Con una variable dependiente ordinal o con una variable dependiente continua transformada en una variable ordinal podemos realizar cálculos estadísticos paralelos a los comentados cuando tratamos el tema de las variables dependientes continuas, aunque no requieren suponer que la población siga una distribución determinada de los datos. Estos métodos estadísticos se denominan no paramétricos. De forma alternativa, podemos efectuar estimaciones de los parámetros de la distribución gaussiana transformando los datos continuos a una escala ordinal y empleando la mediana como una estimación robusta de la media y los dos tercios del intervalo intercuartílico como estimación robusta de la desviación estándar. Esta aproximación es útil cuando la muestra contiene valores extremos o aislados.

El análisis univariante de las variables dependientes nominales se distingue de otros porque en él no se realizan estimaciones independientes de la posición y la dispersión. Las estimaciones de la posición de las variables dependientes nominales pueden ser tasas o proporciones. Los tipos de distribuciones supuestas con más frecuencia para las variables dependientes nominales son la distribución de Poisson y la binomial. La distribución de Poisson se usa siempre que la condición estudiada sea muy poco frecuente. En el análisis se pueden utilizar estas distribuciones directamente o, para simplificar los cálculos, emplear la aproximación normal a las mismas.