

## 7. INFERENCIA PARA MEDIAS Y DESVIACIONES TÍPICAS

### WILLIAM S. GOSSET

¿Qué podría explicar que el jefe de producción de la famosa fábrica de cerveza Guinness de Dublín, Irlanda, no sólo utilizara la estadística sino que además inventara nuevos métodos estadísticos? El anhelo de mejorar la calidad de la cerveza, por supuesto.

William S. Gosset (1876-1937) empezó a trabajar en 1899 como técnico en la fábrica de cerveza Guinness, justo después de licenciarse en la Oxford University. Muy pronto empezó a realizar experimentos y se dio cuenta de la necesidad de utilizar la estadística para comprender los resultados de éstos. ¿Cuáles son las mejores variedades de cebada y de lúpulo para producir cerveza? ¿Cómo se tienen que cultivar? ¿Cómo se deben secar y almacenar? Los resultados de los experimentos de campo, como puedes adivinar, variaban. La inferencia estadística permite descubrir la pauta que esta variación deja oculta. A principios de siglo, los métodos de inferencia se reducían a una versión de las pruebas  $z$  para las medias—incluso los intervalos de confianza eran desconocidos—.

En su trabajo, Gosset se enfrentó con el problema que hemos señalado al utilizar el estadístico  $z$ : no conocía la desviación típica poblacional  $\sigma$ . Es más, en los experimentos de campo se obtenían pocas observaciones, por lo que la simple substitución de  $\sigma$  por  $s$  en el estadístico  $z$  y la suposición de que el resultado era aproximadamente normal, no daban unas conclusiones suficientemente precisas. En consecuencia, Gosset se planteó la pregunta clave, ¿cuál es la distribución exacta del estadístico  $(\bar{x} - \mu)/s$ ?

En 1907, Gosset ya era el responsable de la investigación que se desarrollaba en la cervecera Guinness. Además, Gosset también había encontrado la respuesta a la pregunta anterior y había calculado una tabla de números críticos de su nueva distribución, a la que llamamos distribución  $t$ . La nueva prueba  $t$  identificó la mejor variedad de cebada y Guinness, rápidamente, adquirió toda la semilla disponible. Guinness permitió que publicara sus descubrimientos, pero no con su propio nombre. Gosset utilizó el nombre “Student”, y, en su honor, la prueba  $t$  es llamada a veces “ $t$  de Student”.

## 7.1 Introducción

Una vez vistos los principios de la inferencia estadística, podemos pasar a la práctica. Este capítulo describe los intervalos de confianza y las pruebas de significación para la media de una población y para la comparación de las medias de dos poblaciones. Una sección optativa discute una prueba aplicada a la comparación de las desviaciones típicas de dos poblaciones. En capítulos posteriores se describirán procedimientos de inferencia aplicados a las proporciones poblacionales, a la comparación de las medias de más de dos poblacionales y al estudio de la relación entre variables.

## 7.2 Inferencia para la media de una población

Los intervalos de confianza y las pruebas de significación para la media  $\mu$  de una población normal se basan en la media muestral  $\bar{x}$ . La media de la distribución de  $\bar{x}$  es  $\mu$ . Es decir,  $\bar{x}$  es un estimador insesgado de la  $\mu$  desconocida. La dispersión de  $\bar{x}$  depende del tamaño de la muestra y de la desviación típica poblacional  $\sigma$ . En el capítulo anterior hicimos el supuesto, poco realista, de que conocíamos el valor de  $\sigma$ . En la práctica,  $\sigma$  es desconocida. Por tanto, tenemos que estimar  $\sigma$  a partir de los datos, incluso si nuestro principal interés es  $\mu$ . La necesidad de estimar  $\sigma$  cambia algunos detalles de las pruebas de significación y de los intervalos de confianza para  $\mu$ , pero no su interpretación.

He aquí los supuestos de los que partimos al hacer inferencia para la media poblacional.

### SUPUESTOS DE LA INFERENCIA PARA LA MEDIA

- Nuestros datos son una **muestra aleatoria simple** de tamaño  $n$  de una población. Este supuesto es muy importante.
- Las observaciones proceden de una población que tiene una **distribución normal** con media  $\mu$  y desviación típica  $\sigma$ . En la práctica, a no ser que la muestra sea muy pequeña, es suficiente con que la población sea simétrica y con un solo pico. Tanto  $\mu$  como  $\sigma$  son parámetros desconocidos.

En esta situación, la media muestral  $\bar{x}$  tiene una distribución normal con media  $\mu$  y desviación típica  $\frac{\sigma}{\sqrt{n}}$ . Debido a que no conocemos  $\sigma$ , la estimaremos a partir de la desviación típica muestral  $s$ . A continuación estimaremos la desviación típica de  $\bar{x}$  a partir de  $\frac{s}{\sqrt{n}}$ . Este valor se llama *error típico* de la media muestral  $\bar{x}$ .

### ERROR TÍPICO

Cuando la desviación típica del estadístico se estima a partir de los datos, el resultado se llama **error típico** del estadístico. El error típico de la media muestral  $\bar{x}$  es  $\frac{s}{\sqrt{n}}$ .

#### 7.2.1 Distribuciones $t$

Cuando conocemos el valor de  $\sigma$ , basamos los intervalos de confianza y las pruebas para  $\mu$  en el estadístico  $z$  de una muestra

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Este estadístico  $z$  tiene una distribución normal estandarizada  $N(0, 1)$ . Cuando no conocemos  $\sigma$ , sustituimos la desviación típica de  $\bar{x}$ ,  $\frac{\sigma}{\sqrt{n}}$ , por su error típico  $\frac{s}{\sqrt{n}}$ . El estadístico resultante no tiene una distribución normal. Su distribución, que es nueva para nosotros, se llama *distribución  $t$* .

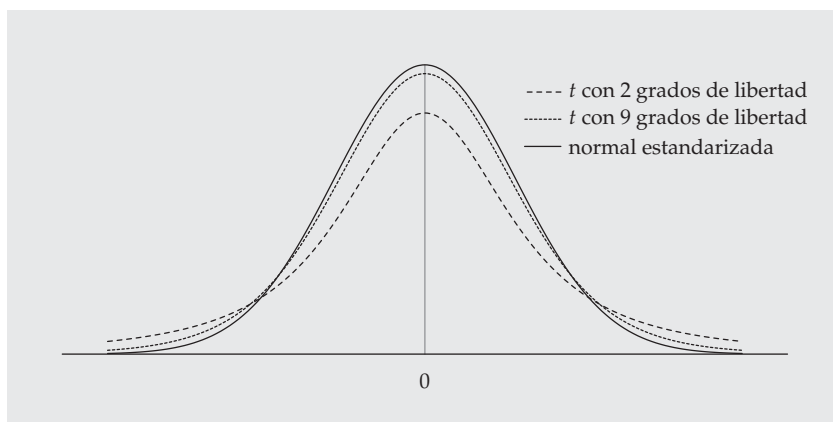
### EL ESTADÍSTICO $t$ DE UNA MUESTRA Y LAS DISTRIBUCIONES $t$

Obtén una muestra aleatoria simple de tamaño  $n$  de una población que tenga una distribución normal con media  $\mu$  y desviación típica  $\sigma$ . El **estadístico  $t$  de una muestra**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

tiene una **distribución  $t$**  con  $n - 1$  grados de libertad.

El estadístico  $t$  tiene la misma interpretación que cualquier estadístico estandarizado: indica a qué distancia se encuentra  $\bar{x}$  de la media  $\mu$ , expresada en desviaciones típicas. Existe una distribución  $t$  distinta para cada tamaño de muestra. Concretamos una distribución  $t$  determinada, dando sus **grados de libertad**. Los grados de libertad del estadístico  $t$  de una muestra se obtienen a partir de la desviación típica muestral  $s$  en el denominador de  $t$ . En el capítulo 1 vimos que  $s$  tenía  $n - 1$  grados de libertad. Existen otros estadísticos  $t$  con diferentes grados de libertad, algunos de los cuales describiremos más adelante en este capítulo. Indicaremos, de forma abreviada, una distribución  $t$  con  $k$  grados de libertad como  $t(k)$ .



**Figura 7.1.** Curvas de densidad de dos distribuciones  $t$  con 2 y 9 grados de libertad, respectivamente, y la distribución normal estandarizada. Todas son simétricas con centro 0. Las distribuciones  $t$  tienen más probabilidad en las colas que la distribución normal estandarizada.

La figura 7.1 compara la curva de densidad de la distribución normal estandarizada con las curvas de densidad de dos distribuciones  $t$  con 2 y 9 grados de libertad, respectivamente. La figura ilustra las siguientes características de las distribuciones  $t$ :

- La forma de las curvas de densidad de las distribuciones  $t$  es similar a la forma de la curva normal estandarizada. Todas ellas son simétricas, con centro en cero, con un solo pico y con forma de campana.
- La dispersión de las distribuciones  $t$  es algo mayor que la dispersión de la distribución normal estandarizada. Las distribuciones  $t$  de la figura 7.1 tienen más probabilidad en las colas y menos en el centro que la normal estandarizada. Esto es debido a que la sustitución del parámetro fijo  $\sigma$  por el estadístico  $s$  introduce más variación en el estadístico  $t$ .
- A medida que aumentan los grados de libertad  $k$  de  $t$ , la curva de densidad de  $t(k)$  se parece más a la curva de densidad de una normal estandarizada  $N(0, 1)$ . Esto es así porque, a medida que aumenta el tamaño de la muestra, la estimación de  $\sigma$  a partir de  $s$  se va haciendo más precisa. Por tanto, la utilización de  $s$  en lugar de  $\sigma$  causa poca variación adicional cuando la muestra es grande.

La tabla C, que se encuentra en la parte final de este libro, da algunos valores críticos para algunas distribuciones  $t$ . Cada fila de la tabla contiene los valores críticos de una de las distribuciones  $t$ ; los grados de libertad aparecen a la izquierda de cada fila. Para mayor comodidad, los valores de la tabla se denominan según  $p$ , la probabilidad de la cola de la derecha que se utiliza en las pruebas de significación, y según los niveles de confianza  $C$  (en porcentaje) necesarios para los intervalos de confianza. Ya has utilizado los valores críticos de la distribución normal estandarizada  $z^*$  situados en la última fila de la tabla C. Puedes comprobar, examinando cualquier columna de arriba a bajo, como a medida que aumentan los grados de libertad de  $t$ , los valores críticos de  $t$  se aproximan cada vez más a los valores críticos de una distribución normal estandarizada. Como en el caso de la tabla normal, los programas estadísticos a menudo hacen innecesaria la utilización de la tabla C.

## APLICA TUS CONOCIMIENTOS

**7.1.** A menudo, los investigadores resumen sus datos dando  $\bar{x}$  y su desviación típica, en vez de dar  $\bar{x}$  y  $s$ . El error típico de la media  $\bar{x}$  a menudo se expresa como ETM.

(a) Un estudio médico halla que  $\bar{x} = 114,9$  y  $s = 9,3$  de un conjunto de datos sobre la presión sistólica de un grupo experimental de 27 sujetos. Halla el error típico de la media.

(b) Unos biólogos que estudian la concentración de un determinado producto en embriones de gambas, presentan sus resultados como, “valores de las medias  $\pm$  ETM de tres muestras independientes”. El valor del ATP era  $0,84 \pm 0,01$ . Los investigadores llevaron a cabo tres determinaciones de ATP, con media  $\bar{x} = 0,84$ . ¿Cuál era la desviación típica muestral  $s$  de estas determinaciones?

7.2. ¿Qué valor crítico  $t^*$  de la tabla C cumple cada una de las siguientes condiciones?

(a) La distribución  $t$  con 5 grados de libertad tiene una probabilidad 0,05 a la derecha de  $t^*$ .

(b) La distribución  $t$  con 21 grados de libertad tiene una probabilidad 0,99 a la izquierda de  $t^*$ .

7.3. ¿Qué valor crítico  $t^*$  de la tabla C cumple cada una de las siguientes condiciones?

(a) El estadístico  $t$  de una muestra de 15 observaciones tiene una probabilidad 0,025 a la derecha de  $t^*$ .

(b) El estadístico  $t$  de una muestra aleatoria simple de una muestra de 20 observaciones tiene una probabilidad 0,75 a la izquierda de  $t^*$ .

### 7.2.2 Intervalos y pruebas $t$

Para analizar muestras de poblaciones normales con  $\sigma$  desconocida, basta con sustituir la desviación típica de  $\bar{x}$ ,  $\frac{\sigma}{\sqrt{n}}$  por su error típico,  $\frac{s}{\sqrt{n}}$ , en los procedimientos  $z$  descritos en el capítulo 6. Los procedimientos  $z$  se convierten, entonces, en los *procedimientos  $t$  de una muestra*. Utiliza los valores  $P$  o los valores críticos de la distribución  $t$  con  $n - 1$  grados de libertad en lugar de los valores normales. La justificación y los cálculos de los procedimientos  $t$  de una muestra son similares a los procedimientos  $z$  del capítulo 6. Por tanto, nos centraremos en la utilización práctica de los procedimientos  $t$ .

### PROCEDIMIENTOS $t$ DE UNA MUESTRA

Obtén una muestra aleatoria simple de tamaño  $n$  de una población de media  $\mu$  desconocida. Un intervalo de confianza de nivel  $C$  para  $\mu$  es

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

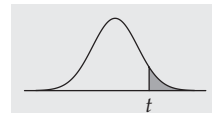
donde  $t^*$  es el valor crítico superior  $\frac{(1-C)}{2}$  de la distribución  $t(n-1)$ . Este intervalo es exacto cuando la distribución de la población es normal y aproximadamente correcto para muestras grandes en los demás casos.

Para contrastar la hipótesis  $H_0 : \mu = \mu_0$  a partir de una muestra aleatoria simple de tamaño  $n$ , calcula el estadístico  $t$  de una muestra

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

En términos de la variable  $T$  que tiene una distribución  $t(n-1)$ , el valor  $P$  para contrastar  $H_0$  en contra de

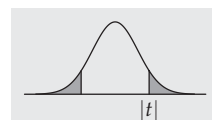
$$H_a : \mu > \mu_0 \quad \text{es} \quad P(T \geq t)$$



$$H_a : \mu < \mu_0 \quad \text{es} \quad P(T \leq t)$$



$$H_a : \mu \neq \mu_0 \quad \text{es} \quad 2P(T \geq |t|)$$



Estos valores  $P$  son exactos si la distribución de la población es normal y son aproximadamente correctos para muestras grandes en los demás casos.

*EJEMPLO 7.1. Metabolismo de las cucarachas*

Para estudiar el metabolismo de los insectos, unos investigadores alimentaron unas cucarachas con soluciones azucaradas. Después de 2, 5 y 10 horas, los investigadores diseccionaron algunas de las cucarachas y analizaron el contenido de azúcar en varios de sus tejidos.<sup>1</sup> Después de 10 horas, los contenidos de D-glucosa (en microgramos) en los intestinos de cinco cucarachas que se alimentaron con una solución que contenía D-glucosa, eran los siguientes:

55,95   68,24   52,73   21,50   23,78

Los investigadores calcularon un intervalo de confianza del 95% para el contenido medio de D-glucosa en los intestinos de las cucarachas en las condiciones anteriores.

Primero calcularon

$$\bar{x} = 44,44 \quad y \quad s = 20,741$$

Los grados de libertad son  $n - 1 = 4$ . En la tabla C encontramos que para un intervalo de confianza del 95%,  $t^* = 2,776$ . El intervalo de confianza es

$$\begin{aligned} \bar{x} \pm t^* \frac{s}{\sqrt{n}} &= 44,44 \pm 2,776 \frac{20,741}{\sqrt{5}} \\ &= 44,44 \pm 25,75 \\ &= 18,69 \text{ a } 70,19 \end{aligned}$$

La comparación de esta estimación con las estimaciones para otros tejidos y para diferentes momentos de disección permitió saber más sobre el metabolismo de las cucarachas y sobre nuevos métodos de eliminación de cucarachas en casas y restaurantes. El hecho de que el error de estimación sea grande se debe a que la muestra es pequeña y a que la dispersión es relativamente grande, lo que se refleja en la magnitud de  $s$ . ■

<sup>1</sup>D. L. Shankland *et al.*, "The effect of 5-thio-D-glucose on insect development and its absorption by insects", *Journal of Insect Physiology*, 14, 1968, págs. 63-72.



El intervalo de confianza  $t$  de una muestra tiene la forma

$$\text{estimación} \pm t^* \text{ET}_{\text{de la estimación}}$$

donde “ET” significa “error típico”. Encontraremos diversos intervalos de confianza que tienen esta misma forma. Al igual que con los intervalos de confianza, las pruebas  $t$  son muy parecidas a las pruebas  $z$  que vimos anteriormente. He aquí un ejemplo. En el capítulo 6 utilizamos la prueba  $z$  para estos datos. Para ello, tuvimos que suponer, de forma poco realista, que conocíamos la desviación típica  $\sigma$  de la población. Ahora podemos hacer un análisis más realista.

### EJEMPLO 7.2. Refrescos light

Los fabricantes de refrescos prueban nuevas fórmulas para evitar la pérdida de dulzor de los refrescos *light* durante el almacenamiento. Unos catadores experimentados evalúan el dulzor de los refrescos antes y después del almacenamiento. He aquí las pérdidas de dulzor (dulzor antes del almacenamiento menos dulzor después del almacenamiento) halladas por 10 catadores para una nueva fórmula de refresco.

2,0 0,4 0,7 2,0 -0,4 2,2 -1,3 1,2 1,1 2,3

Estos datos, ¿constituyen una buena evidencia de que el refresco perdió dulzor durante el almacenamiento?

**Paso 1: Hipótesis.** Existen diferencias sobre la percepción de la pérdida de dulzor por parte de los catadores. Por este motivo, planteamos las hipótesis como la pérdida media de dulzor  $\mu$  de una gran población de catadores. La hipótesis nula establece que la “pérdida es nula” y la hipótesis alternativa que “existe una pérdida”.

$$H_0 : \mu = 0$$

$$H_a : \mu > 0$$

**Paso 2:** Estadístico de contraste. Los estadísticos básicos son

$$\bar{x} = 1,02 \quad \text{y} \quad s = 1,196$$

El estadístico  $t$  de una muestra es

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1,02 - 0}{1,196/\sqrt{10}} \\ &= 2,70 \end{aligned}$$

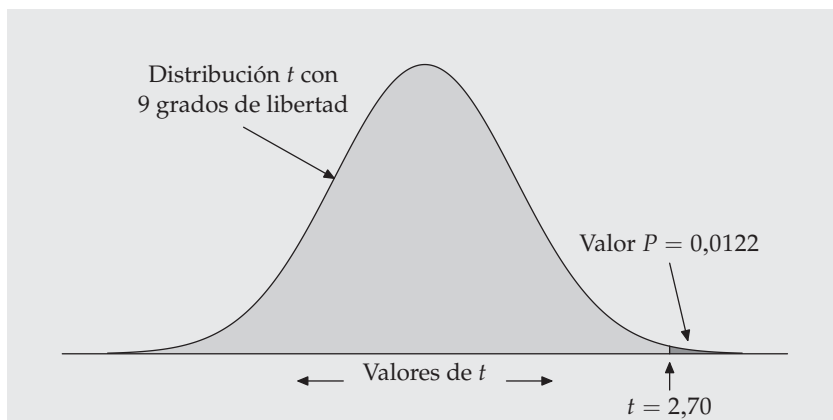


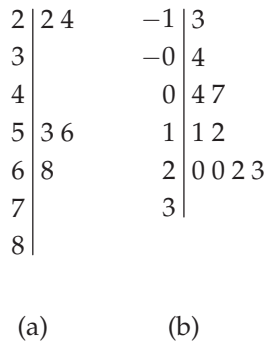
Figura 7.2. El valor  $P$  de la prueba  $t$  de una cola del ejemplo 7.2.

**Paso 3: Valor  $P$ .** El valor  $P$  para  $t = 2,70$  es el área situada a la derecha de 2,70 por debajo de la curva de la distribución  $t$  con  $n - 1 = 9$  grados de libertad (gl). La figura 7.2 muestra este área. Sin la ayuda de un ordenador no podemos hallar el valor exacto de  $P$ . De todas formas, utilizando la tabla C podemos situar  $P$  entre dos valores. Busca en la fila correspondiente a 9 grados de libertad entre qué valores se encuentra  $t = 2,70$ . Debido a que el valor observado de  $t$  se encuentra entre los valores críticos de 0,02 y 0,01, el valor  $P$  se halla entre 0,01 y 0,02. Un programa estadístico da el valor más exacto,  $P = 0,0122$ . Existe una evidencia bastante fuerte a favor de que se produce una pérdida de dulzor. ■

### 7.2.3 Utilización de procedimientos $t$ de una muestra

¿Son fiables las conclusiones que obtuvimos en los ejemplos 7.1 y 7.2? Los dos ejemplos son experimentos. En ambos casos los investigadores tomaron especiales precauciones para evitar sesgos. Las cucarachas fueron asignadas al azar a las distintas soluciones azucaradas y a los distintos tiempos antes de la diseción. Desde cualquier otro punto de vista, todas las cucarachas fueron tratadas exactamente igual. Los catadores trabajaron en cabinas aisladas para evitar que se influyeran mutuamente. En consecuencia, podemos fiarnos de los resultados obtenidos en estos dos experimentos concretos.

Tenemos que tratar a las cucarachas y a los catadores como muestras aleatorias simples de poblaciones grandes si queremos sacar conclusiones, en general, sobre las cucarachas o sobre los catadores. Las cucarachas fueron escogidas al azar de una población de cucarachas criadas en laboratorio con fines experimentales. Todos los catadores tenían una formación similar. Aunque, en realidad, no tengamos muestras aleatorias simples de las poblaciones en las que estamos interesados, estamos dispuestos a actuar como si lo fueran. Ésta es una cuestión a juzgar en cada caso.



**Figura 7.3.** Diagramas de tallos con los datos del ejemplo 7.1 en (a) y del ejemplo 7.2 en (b).

Los procedimientos  $t$  dan por supuesto que la población de todas las cucarachas y la de todos los catadores tienen distribuciones normales. El supuesto de que la distribución de la población es normal no se puede comprobar eficazmente con sólo 5 o 10 observaciones. En parte, los investigadores confían en la experiencia con variables similares. También examinan los datos. La figura 7.3 representa los diagramas de tallos de las dos distribuciones (los datos de las cucarachas se han redondeado). La distribución de las diferencias de dulzor de los 10 catadores no tiene una forma regular, pero no se observan espacios vacíos ni observaciones atípicas ni otros signos de falta de normalidad. Los datos de las cucarachas, por otro lado, presentan un gran espacio vacío entre las dos observaciones menores y las tres mayores. Con datos observacionales, esto podría indicar la existencia de dos tipos de cucarachas. En este caso sabemos que todas las cucarachas proceden de una misma población que se ha criado en laboratorio. El espacio vacío se debe a la variación del azar en muestras muy pequeñas. Afortunadamente, veremos que los niveles de confianza y los valores  $P$  de los procedimientos  $t$  no son muy sensibles a la falta de normalidad.

**APLICA TUS CONOCIMIENTOS**

7.4. ¿Qué valor crítico  $t^*$  de la tabla C utilizarías en un intervalo de confianza para la media poblacional en cada una de las siguientes situaciones?

- (a) Un intervalo de confianza del 95% basado en  $n = 10$  observaciones.
- (b) Un intervalo de confianza del 99% de una muestra aleatoria simple con 20 observaciones.
- (c) Un intervalo de confianza del 80% de una muestra de tamaño 7.

7.5. El estadístico  $t$  de una muestra para contrastar

$$H_0 : \mu = 0$$

$$H_a : \mu > 0$$

de una muestra de  $n = 15$  observaciones tiene el valor  $t = 1,82$ .

- (a) ¿Cuántos grados de libertad tiene este estadístico?
- (b) Halla los dos valores críticos  $t^*$  de la tabla C entre los que se encuentra  $t$ . ¿Cuáles son las probabilidades  $p$  de la cola de la derecha de estos dos valores?
- (c) ¿Entre qué dos valores se encuentra el valor  $P$  de la prueba?
- (d) ¿Es significativo el valor  $t = 1,82$  a un nivel del 5%? ¿Es significativo a un nivel del 1%?

7.6. El estadístico  $t$  de una muestra de  $n = 25$  observaciones para contrastar la prueba de dos colas de

$$H_0 : \mu = 64$$

$$H_a : \mu \neq 64$$

tiene un valor  $t = 1,12$ .

- (a) ¿Cuántos grados de libertad tiene  $t$ ?
- (b) Sitúa los dos valores críticos  $t^*$  de la tabla C entre los que se encuentra  $t$ . ¿Cuáles son las probabilidades  $p$  de la cola de la derecha de estos dos valores?
- (c) ¿Entre qué dos valores se encuentra el valor  $P$  de la prueba? (Ten en cuenta que  $H_a$  tiene dos colas.)
- (d) El valor  $t = 1,12$ , ¿es estadísticamente significativo a un nivel del 10%? ¿Y a un nivel del 5%?

7.7. **Envenenamiento con DDT.** El envenenamiento con DDT causa temblores y convulsiones. En un estudio sobre envenenamiento con DDT, unos investigadores suministraron una determinada cantidad de este producto a un grupo de

ratas. Más tarde se tomaron datos sobre sus sistemas nerviosos para averiguar cómo el DDT causa temblores. Una variable importante era el “periodo absolutamente refractario”, es decir, el tiempo que necesita un nervio para recuperarse después de un estímulo. La variación de este periodo es normal. Las mediciones hechas en cuatro ratas dieron los siguientes datos (en milisegundos):<sup>2</sup>

1,6 1,7 1,8 1,9

(a) Halla la media del periodo absolutamente refractario  $\bar{x}$  y su error típico.

(b) Calcula un intervalo de confianza del 90% para la media del periodo absolutamente refractario para todas las ratas de este tipo que fueron sometidas al mismo tratamiento.

(c) Supón que la media del periodo absolutamente refractario para ratas no envenenadas es de 1,3 milisegundos. El envenenamiento con DDT debería retrasar la recuperación del nervio y, por tanto, aumentar dicho periodo. Los datos, ¿proporcionan buena evidencia a favor de esta afirmación? Plantea  $H_0$  y  $H_a$ , y lleva a cabo la prueba  $t$ . ¿Entre qué valores de la tabla C se halla el valor  $P$ ? ¿Qué conclusiones obtienes de esta prueba?

#### 7.2.4 Procedimientos $t$ para diseños por pares

El estudio del ejemplo 7.1 estimó el contenido medio de azúcar en los intestinos de unas cucarachas, aunque, para tener una visión más general, los investigadores compararon después los resultados en varios tejidos y en distintos tiempos antes de la disección. La prueba de cata del ejemplo 7.2 corresponde a un estudio por pares, en el cual los mismos 10 catadores valoraron el dulzor de los refrescos antes y después de su almacenamiento. Los estudios comparativos son más convincentes que las investigaciones basadas en una sola muestra. Por este motivo, la inferencia basada en una muestra es menos habitual que la inferencia comparativa. Un diseño experimental bastante frecuente utiliza los procedimientos  $t$  de una muestra para comparar dos tratamientos. En un **diseño por pares**, los sujetos se agrupan por pares y cada sujeto del par recibe uno de los dos tratamientos. El investigador puede lanzar una moneda al aire para asignar los tratamientos a los sujetos de cada par. Otra situación a la que conviene aplicar un diseño por pares es cuando hay observaciones del antes y el después con los mismos sujetos, como en la prueba de cata del ejemplo 7.2.

*Diseño por pares*

<sup>2</sup>D. L. Shankland, “Involvement of spinal cord and peripheral nerves in DDT-poisoning syndrome in albino rats”, *Toxicology and Applied Pharmacology*, 6, 1964, págs. 197-213.

### PROCEDIMIENTOS $t$ EN DISEÑOS POR PARES

Para comparar las respuestas de dos tratamientos en un diseño por pares, aplica los procedimientos  $t$  de una muestra a las diferencias observadas.

El parámetro  $\mu$  en un procedimiento  $t$  de un diseño por pares es la media de las diferencias entre las respuestas a los dos tratamientos de los sujetos de cada par en toda la población.

Tabla 7.1. Tiempo medio necesario para hallar la salida de un laberinto.

Sujeto	Sin esencia	Con esencia	Diferencia	Sujeto	Sin esencia	Con esencia	Diferencia
1	30,60	37,97	-7,37	12	58,93	83,50	-24,57
2	48,43	51,57	-3,14	13	54,47	38,30	16,17
3	60,77	56,67	4,10	14	43,53	51,37	-7,84
4	36,07	40,47	-4,40	15	37,93	29,33	8,60
5	68,47	49,00	19,47	16	43,50	54,27	-10,77
6	32,43	43,23	-10,80	17	87,70	62,73	24,97
7	43,70	44,57	-0,87	18	53,53	58,00	-4,47
8	37,10	28,40	8,70	19	64,30	52,40	11,90
9	31,17	28,23	2,94	20	47,37	53,63	-6,26
10	51,23	68,47	-17,24	21	53,67	47,00	6,67
11	65,40	51,10	14,30				

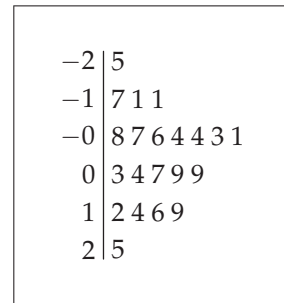
#### EJEMPLO 7.3. Olores florales y aprendizaje

Se dice que escuchar a Mozart mejora los resultados de las pruebas realizadas por estudiantes. Es posible que los buenos olores tengan el mismo efecto. Para probar esta idea, 21 sujetos tuvieron que pasar una prueba que consistía en marcar con un lápiz la trayectoria a seguir para salir de un laberinto mientras llevaban puesta una mascarilla. La mascarilla era inodora o bien contenía una esencia floral. La variable respuesta es el tiempo medio en tres ensayos. Cada sujeto tenía que pasar la prueba con cada una de las mascarillas. El orden de realización de las pruebas era aleatorio. La aleatorización es importante ya que los sujetos tienden

a mejorar sus tiempos al repetir la prueba varias veces. La tabla 7.1 da los tiempos medios con ambas mascarillas.<sup>3</sup>

Para analizar estos datos, resta los tiempos “con esencia” de los tiempos “sin esencia”. Las 21 diferencias forman una sola muestra. Aparecen en la columna de “diferencias” de la tabla 7.1. El primer sujeto, por ejemplo, tardó 7,37 segundos menos cuando llevaba la mascarilla con esencia, por tanto la diferencia es negativa. Debido a que tiempos menores indican un mejor resultado, las diferencias positivas indican que el sujeto lo hizo mejor cuando llevaba una mascarilla con esencia. Un diagrama de tallos de las diferencias (figura 7.4) muestra que su distribución es simétrica con una forma razonablemente normal.

**Figura 7.4.** Diagrama de tallos de las diferencias de tiempo para hallar la salida de un laberinto. Datos correspondientes a los 21 sujetos del ejemplo 7.3. Los datos se han redondeado al valor entero más próximo. Fíjate en que el tallo 0 debe aparecer dos veces, para poder mostrar las diferencias entre  $-9$  y  $0$  y entre  $0$  y  $9$ .



**Paso 1. Hipótesis.** Para valorar si el perfume floral mejoró los resultados de forma significativa, contrastamos

$$H_0 : \mu = 0$$

$$H_a : \mu > 0$$

Aquí  $\mu$  es media de las diferencias en la población de la cual se obtuvieron los sujetos. La hipótesis nula dice que no tiene lugar ninguna mejora,  $H_a$  indica que como media, los tiempos con mascarilla sin esencia son mayores que los tiempos con esencia floral.

<sup>3</sup>A. R. Hirsch y L. H. Johnston, “Odors and learning”, *Journal of Neurological and Orthopedic Medicine and Surgery*, 17, 1996, págs. 119-126.

**Paso 2. Estadístico de contraste.** Las 21 diferencias tienen

$$\bar{x} = 0,9567 \quad y \quad s = 12,5479$$

En consecuencia, el estadístico  $t$  de una muestra es

$$\begin{aligned} t &= \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{0,9567 - 0}{12,5479/\sqrt{21}} \\ &= 0,349 \end{aligned}$$

**Paso 3. Valor  $P$ .** Halla el valor  $P$  de la distribución  $t(20)$ . (Recuerda que los grados de libertad son el tamaño de la muestra menos 1.) La tabla C muestra que 0,349 es menor que el valor crítico de  $t(20)$  que le corresponde un área a su derecha de 0,25. En consecuencia, el valor  $P$  es mayor que 0,25. Un programa estadístico da el valor  $P = 0,3652$ .

**Conclusión:** Los datos no apoyan la afirmación de que los perfumes florales mejoran los resultados. Como media, la mejora es pequeña, ya que los sujetos que llevaban la mascarilla sin esencia tardaron 0,96 segundos más. Este valor es relativamente pequeño con relación a los 50 segundos de media que tarda en terminar el laberinto un sujeto. Esta pequeña mejora no es estadísticamente significativa incluso a un nivel del 25%. ■

El ejemplo 7.3 ilustra cómo expresar los datos de una prueba por pares como si fueran datos de una sola muestra. Para ello basta con calcular las diferencias entre cada par. En realidad, estamos haciendo inferencia sobre una sola población, la población de todas las diferencias entre cada par. Es incorrecto ignorar los pares y analizar los datos como si tuviéramos dos muestras, una de los sujetos que llevaban mascarillas sin esencia y otra de los sujetos que llevaban mascarillas perfumadas. Los procedimientos inferenciales para comparar dos muestras se basan en el supuesto de que las dos muestras se obtienen de forma independiente. Este supuesto no se cumple cuando los mismos sujetos son considerados dos veces. El análisis adecuado depende del diseño utilizado para obtener los datos.

Debido a que los procedimientos  $t$  son tan comunes, todos los programas estadísticos pueden hacerte los cálculos. Si conoces los procedimientos  $t$ , puedes comprender los resultados de cualquier programa. La figura 7.5 muestra los resultados del ejemplo 7.3 obtenidos a partir de tres programas estadísticos: Minitab, Data Desk y Excel. En cada caso, hemos introducido los datos y pedimos al programa que haga la prueba  $t$  para datos por pares de una cola. Los tres resultados contienen información ligeramente distinta, sin embargo todos incluyen los cálculos básicos:  $t = 0,349$ ,  $P = 0,365$ .



**Paired T-Test and Confidence Level**

Paired T for sin\_esencia - con\_esencia

	N	Mean	StDev	SE Mean
sin_esencia	21	50.01	14.36	3.13
con_esencia	21	49.06	13.39	2.92
diferencia	21	0.96	12.55	2.74

95% CI for mean difference: (-4.76, 6.67)

T-Test of mean difference = 0 (vs > 0): T-Value = 0.35 P-Value = 0.365

(a) Minitab

sin\_esencia - con\_esencia:

Test Ho:  $\mu(\text{sin\_esencia-con\_esencia}) = 0$  vs Ha:  $\mu(\text{sin\_esencia-con\_esencia}) > 0$

Mean of Paired Differences = 0.956667 t-Statistic = 0.349 w/20 df

Fail to reject Ho at Alpha = 0.0500

p = 0.3652

(b) Data Desk

t-Test: Paired Two Sample for Means

	Variable 1	Variable 2
Mean	50.01429	49.05762
Variance	206.3097	179.1748
Observations	21	21
Pearson Correlation	0.593026	
Hypothesized Mean Difference	0	
df	20	
t Stat	0.349381	
P(T<=t) one-tail	0.365227	
t Critical one-tail	1.724718	
P(T<=t) two-tail	0.730455	
t Critical two-tail	2.085962	

(c) Excel

**Figura 7.5.** Resultados correspondientes a tres programas estadísticos, de la prueba *t* del diseño por pares del ejemplo 7.3. Es fácil hallar los resultados básicos en los datos de cualquiera de los tres programas.

## APLICA TUS CONOCIMIENTOS

A partir de ahora, muchos ejercicios te piden que halles el valor  $P$  de una prueba  $t$ . Si tienes una calculadora o un ordenador con un programa estadístico adecuado, da el valor  $P$  exacto. Si no lo tienes, utiliza la tabla C para dar los dos valores entre los que se halle  $P$ .

**7.8. Cultivo de tomates.** Un experimento agrícola compara el rendimiento de dos variedades comerciales de tomates. Unos investigadores dividen por la mitad 10 parcelas situadas en distintas localidades y plantan cada variedad de tomate en cada una de las mitades de las parcelas. Después de la cosecha, los investigadores comparan los rendimientos, en kilos por planta, en cada localidad. Las 10 diferencias (Variedad A – Variedad B) dan  $\bar{x} = 0,75$  y  $s = 1,83$ . ¿Existe suficiente evidencia de que como media la Variedad A tiene un rendimiento mayor?

- (a) Describe con palabras qué es el parámetro  $\mu$  en este contexto.
- (b) Plantea  $H_0$  y  $H_a$ .
- (c) Halla el estadístico  $t$  y da el valor  $P$ . ¿Cuáles son tus conclusiones?

**7.9. Derecha versus izquierda.** El diseño de mandos e instrumentos influye en la facilidad con que la gente puede utilizarlos. El proyecto de un alumno consistió en investigar este efecto pidiendo a 25 compañeros diestros que accionaran un mando giratorio (con la mano derecha) que por la acción del giro desplazaba un indicador. Había dos instrumentos idénticos, uno con el mando que giraba hacia la derecha y otro con el mando que giraba hacia la izquierda. La tabla 7.2 da los tiempos en segundos que tardó cada sujeto en desplazar el indicador una determinada distancia.<sup>4</sup>

(a) Cada uno de los 25 estudiantes que participaron en el experimento utilizó ambos instrumentos. Comenta brevemente cómo utilizarías la aleatorización para preparar el experimento.

(b) El proyecto esperaba demostrar que las personas diestras utilizan más fácilmente los mandos que giran hacia la derecha. ¿Cuál es el parámetro  $\mu$  de una prueba  $t$  por pares? Plantea  $H_0$  y  $H_a$  en términos de  $\mu$ .

(c) Lleva a cabo una prueba con tus hipótesis. Da el valor  $P$  e informa de tus conclusiones.

<sup>4</sup>Datos de Timothy Sturm.

Tabla 7.2. Tiempos empleados utilizando mandos de giro hacia la derecha y mandos de giro hacia la izquierda.

Sujeto	Giro hacia la derecha	Giro hacia la izquierda	Sujeto	Giro hacia la derecha	Giro hacia la izquierda
1	113	137	14	107	87
2	105	105	15	118	166
3	130	133	16	103	146
4	101	108	17	111	123
5	138	115	18	104	135
6	118	170	19	111	112
7	87	103	20	89	93
8	116	145	21	78	76
9	75	78	22	100	116
10	96	107	23	89	78
11	122	84	24	85	101
12	103	148	25	88	123
13	116	147			

**7.10.** Calcula un intervalo de confianza del 90% para la media de la ganancia del tiempo empleado con los mandos que giran hacia la derecha con relación al tiempo empleado con los mandos que giran hacia la izquierda, en el contexto del ejercicio 7.9. ¿Crees que el tiempo ahorrado tendría una importancia práctica si la tarea se efectuara muchas veces, por ejemplo, por parte de un trabajador de una cadena de montaje? Para ayudarte a contestar a esta pregunta, halla el tiempo medio empleado con los mandos que giran hacia la derecha como porcentaje del tiempo medio empleado con los mandos que giran hacia la izquierda.

### 7.2.5 Robustez de los procedimientos $t$

Los procedimientos  $t$  de una muestra sólo son completamente exactos cuando la población es normal. Pero las poblaciones reales nunca son exactamente normales. Por tanto, la utilidad de los procedimientos  $t$ , en la práctica, depende de cómo se vean afectados por la falta de normalidad.

#### PROCEDIMIENTOS ROBUSTOS

Un intervalo de confianza o una prueba de significación son considerados **robustos** si el nivel de confianza o el valor  $P$  no cambian mucho cuando se violan los supuestos en los que se basa el procedimiento.

Debido a que las colas de las curvas normales descienden muy rápidamente, las muestras de poblaciones normales deben tener muy pocas observaciones atípicas. Las observaciones atípicas sugieren que los datos no constituyen una muestra de una población normal. **De forma similar a lo que ocurre con  $\bar{x}$  y con  $s$ , los procedimientos  $t$  se ven muy influidos por las observaciones atípicas.**

#### EJEMPLO 7.4. Efecto de las observaciones atípicas

Los datos de las cucarachas del ejemplo 7.1 eran

55,95   68,24   52,73   21,50   23,78

Si las dos observaciones de valores muy pequeños se incrementaran en 20 microgramos, a 41,5 y 43,78, la media aumentaría de  $\bar{x} = 44,44$  a  $\bar{x} = 52,44$ , y la desviación típica bajaría de  $s = 20,74$  a  $s = 10,69$ . El intervalo de confianza para  $\mu$  sería la mitad de ancho que el del ejemplo 7.1. ■

Afortunadamente, los procedimientos  $t$  son bastante robustos respecto a la falta de normalidad de la población cuando no hay observaciones atípicas, especialmente cuando la distribución de los datos es aproximadamente simétrica. Las muestras grandes mejoran la precisión de los valores  $P$  y de los valores críticos de las distribuciones  $t$  cuando la población no es normal. La principal razón para este comportamiento es el teorema del límite central. El estadístico  $t$  utiliza la media muestral  $\bar{x}$ , la cual es más normal a medida que aumenta el tamaño de la muestra, incluso cuando la población no tiene una distribución normal.

Siempre que tengas muestras pequeñas, antes de utilizar los procedimientos  $t$ , dibuja un gráfico para detectar asimetrías o la presencia de observaciones atípicas. Si dispones de 15 observaciones o más, los procedimientos  $t$  que hemos visto se pueden aplicar de forma segura cuando  $n \geq 15$  a no ser que entre los datos existan observaciones atípicas o que la distribución sea muy asimétrica. He aquí unas reglas prácticas para hacer inferencia para una sola media.<sup>5</sup>

<sup>5</sup>Harry O. Posten, "The robustness of the one-sample  $t$ -test over the Pearson system", *Journal of Statistical Computation and Simulation*, 9, 1979, págs. 133-149 y E. S. Pearson y N. W. Please, "Relation between the shape of population distribution and the robustness of four simple test statistics", *Biometrika*, 62, 1975, págs. 223-241.

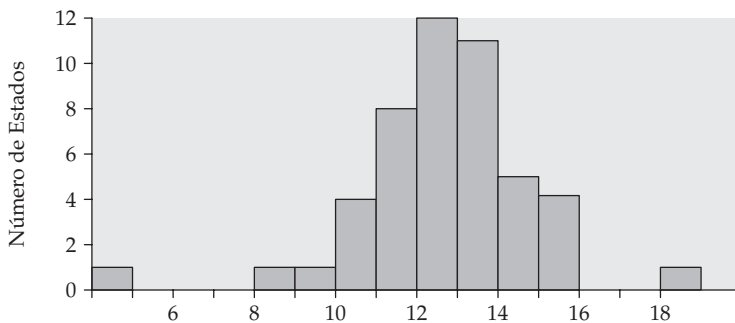
### UTILIZACIÓN DE LOS PROCEDIMIENTOS $t$

- Excepto en el caso de muestras pequeñas, el supuesto de que los datos sean una muestra aleatoria simple de la población de interés es más importante que el supuesto de que la distribución de la población sea normal.
- *Tamaño de muestra menor que 15.* Utiliza los procedimientos  $t$  si los datos son aproximadamente normales. Si los datos no son claramente normales o si existen observaciones atípicas, no utilices los procedimientos  $t$ .
- *Tamaño de la muestra mayor o igual a 15.* Los procedimientos  $t$  se pueden utilizar a no ser que existan observaciones atípicas o que la distribución sea muy asimétrica.
- *Muestras grandes.* Los procedimientos  $t$  se pueden utilizar incluso para distribuciones muy asimétricas cuando la muestra sea grande, aproximadamente cuando  $n \geq 40$ .

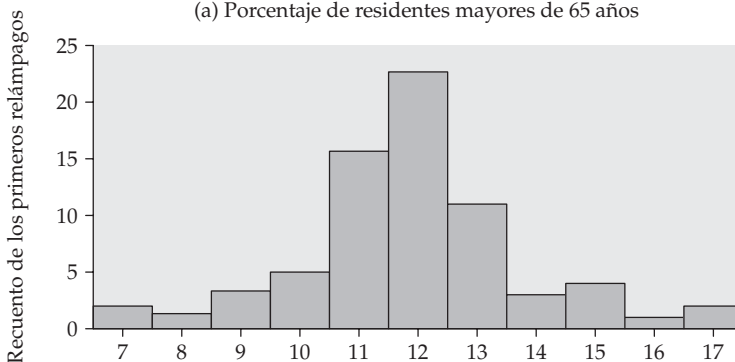
#### EJEMPLO 7.5. ¿Podemos utilizar una $t$ ?

Considera algunos de los conjuntos de datos que representamos gráficamente en el capítulo 1. La figura 7.6 muestra los histogramas.

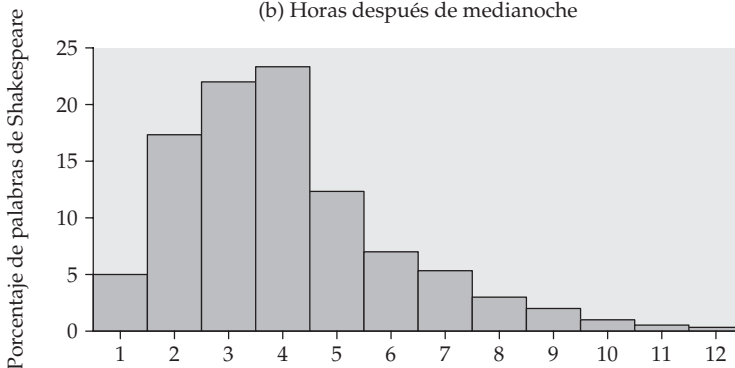
- La figura 7.6(a) es un histograma de los porcentajes de residentes de 65 o más años en cada uno de los Estados de EE UU. *Tenemos datos de toda la población de los 50 Estados; por tanto, no tiene sentido hacer inferencia.* Podemos calcular de manera exacta la media de la población. No tenemos la incertidumbre que se presenta cuando sólo disponemos de una muestra de la población, por lo que no es necesario ningún intervalo de confianza ni ninguna prueba de significación.



(a) Porcentaje de residentes mayores de 65 años



(b) Horas después de medianoche



(c) Número de letras en las palabras

**Figura 7.6.** ¿Podemos utilizar los procedimientos *t* con estos datos? (a) Porcentajes de residentes con al menos 65 años en los distintos Estados de EE UU. No: se trata de toda la población, no de una muestra. (b) Hora del día en la que se produce el primer relámpago en una localidad de Colorado. Sí, hay más de 70 observaciones con una distribución simétrica. (c) Longitud de las palabras en las obras de Shakespeare. Sí, si la muestra es suficientemente grande como para contrarrestar la asimetría hacia la derecha.

- La figura 7.6(b) muestra la distribución de la hora del día en que se produce el primer relámpago en una región montañosa de Colorado, EE UU. Los datos contienen más de 70 observaciones que tienen una distribución simétrica. Puedes utilizar los procedimientos  $t$  para sacar conclusiones sobre la media de la hora del día en la que se produce el primer relámpago con toda fiabilidad.
- La figura 7.6(c) muestra que la distribución de la longitud de las palabras de las obras de Shakespeare es muy asimétrica hacia la derecha. No sabemos el número de observaciones de que disponemos. Se pueden utilizar los procedimientos  $t$  con una distribución de este tipo si el número de observaciones es igual o mayor que 40. ■

## APLICA TUS CONOCIMIENTOS

**7.11. La dependencia de la cafeína, ¿es real?** Nuestros sujetos son 11 personas diagnosticadas como dependientes de la cafeína. A cada una de ellas se le impidió tomar café, cola o cualquier otra sustancia que contuviera cafeína. Sin embargo, se les suministró unos comprimidos que contenían su dosis habitual de cafeína. Asimismo, durante diferentes periodos de tiempo, los sujetos estudiados tomaron comprimidos placebo. El orden en el cual los sujetos tomaron la cafeína o el placebo fue al azar. La tabla 7.3 contiene datos sobre dos pruebas que pasaron los sujetos. “Depresión” es el resultado de la prueba Beck. Una puntuación mayor en esta prueba indica más síntomas de depresión. “Pulso” es el número de pulsaciones por minuto de cada sujeto después de pedirle que accionara un mando 200 veces de la forma más rápida posible. Queremos saber si la falta de cafeína afecta a estos resultados.<sup>6</sup>

(a) El estudio se llevó a cabo doblemente a ciegas. ¿Qué significa esto?

(b) Este experimento por pares, ¿proporciona evidencia de que la falta de cafeína aumenta los valores de los indicadores de depresión? Dibuja un diagrama de tallos para comprobar que no existe una falta de normalidad exagerada de la distribución de las diferencias. Plantea las hipótesis, y lleva a cabo la prueba. ¿Cuáles son tus conclusiones?

<sup>6</sup>E. C. Strain, G. K. Mumford, K. Silverman y R. R. Griffiths, “Caffeine dependence syndrome: evidence from case histories and experimental evaluation”, *Journal of the American Medical Association*, 272, 1994, págs. 1.604-1.607.

(c) Ahora dibuja un diagrama de tallos correspondiente a las diferencias del número de pulsaciones por minuto con cafeína o sin ella. ¿Se te plantea alguna duda a la hora de utilizar los procedimientos  $t$  con estos datos? ¿Por qué?

Tabla 7.3. Resultados del estudio sobre la dependencia de la cafeína.

Sujeto	Depresión (cafeína)	Depresión (placebo)	Pulso (cafeína)	Pulso (placebo)
1	5	16	281	201
2	5	23	284	262
3	4	5	300	283
4	3	7	421	290
5	8	14	240	259
6	5	24	294	291
7	0	6	377	354
8	0	3	345	346
9	2	15	303	283
10	11	12	340	391
11	1	0	408	411

**7.12. ¿Cargarán más?** Un banco se pregunta si la eliminación de la cuota anual de la tarjeta de crédito de los clientes que carguen en ella un mínimo de 2.400 € al año haría aumentar los cargos en sus tarjetas. El banco hace esta oferta a una muestra aleatoria simple de 200 clientes con tarjeta de crédito. Posteriormente, compara lo que han cargado esos clientes este año con lo que cargaron el año anterior. El aumento medio es de 332 € y la desviación típica es de 108 €.

(a) ¿Existe evidencia significativa a un nivel del 1% de que la cantidad media cargada aumenta con la oferta de eliminación de la cuota? Plantea  $H_0$  y  $H_a$ , y lleva a cabo una prueba  $t$ .

(b) Calcula un intervalo de confianza del 99% para la media del aumento.

(c) La distribución de las cantidades cargadas en las tarjetas de crédito es asimétrica hacia la derecha, pero no existen observaciones atípicas debido al límite de crédito que el banco impone a cada tarjeta. La utilización de los procedimientos  $t$  está justificada en este caso, a pesar de que la distribución de la población no es normal. Explica por qué.

(d) Un observador crítico puntualiza que los clientes posiblemente habrían cargado en sus tarjetas más este año que el año pasado incluso sin la oferta del banco, debido a que la situación económica actual es mejor que la del año pasado y los tipos de interés son más bajos. Describe brevemente el diseño de un experimento para estudiar el efecto de la eliminación de la cuota que evitaría esta crítica.



**7.13. Cigüeñales de auto.** He aquí las mediciones (en milímetros) de una dimensión crítica de 16 cigüeñales de un motor para automóviles:

224,120	224,001	224,017	223,982	223,989	223,961
223,960	224,089	223,987	223,976	223,902	223,980
224,098	224,057	223,913	223,999		

Se supone que la dimensión crítica de los cigüeñales es de 224 mm y que la variabilidad del proceso de fabricación es desconocida. ¿Existe suficiente evidencia de que la dimensión media no es de 224 mm?

(a) Comprueba gráficamente que no haya observaciones atípicas o una fuerte asimetría que pudieran amenazar la validez de los procedimientos  $t$ . ¿A qué conclusión llegas?

(b) Plantea  $H_0$  y  $H_a$ , y lleva a cabo una prueba  $t$ . Halla el valor  $P$  (de la tabla C o de un programa estadístico). ¿A qué conclusión llegas?

## RESUMEN DE LA SECCIÓN 7.2

Las pruebas y los intervalos de confianza para la media  $\mu$  de una población normal se basan en la media muestral  $\bar{x}$  de una muestra aleatoria simple. El teorema del límite central garantiza que estos procedimientos son aproximadamente correctos con otro tipo de distribuciones poblacionales cuando las muestras son grandes.

La media muestral estandarizada es el **estadístico  $z$  de una muestra**,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Cuando conocemos  $\sigma$ , utilizamos el estadístico  $z$  y la distribución normal estandarizada.

En la práctica, no conocemos  $\sigma$ . Sustituye la desviación típica  $\frac{\sigma}{\sqrt{n}}$  de  $\bar{x}$  por el **error típico**  $\frac{s}{\sqrt{n}}$  para obtener el **estadístico  $t$  de una muestra**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

El estadístico  $t$  tiene una **distribución  $t$**  con  $n - 1$  grados de libertad.

Existe una distribución  $t$  distinta para cada valor positivo  $k$  de **grados de libertad**. Todas las distribuciones  $t$  son simétricas con forma similar a la distribución normal estandarizada. La distribución  $t(k)$  se aproxima a la distribución  $N(0, 1)$  a medida que  $k$  aumenta.

Un **intervalo de confianza** de nivel  $C$  exacto para la media  $\mu$  de una población normal es

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

donde  $t^*$  es el valor crítico superior  $\frac{(1-C)}{2}$  de una distribución  $t(n-1)$ .

Las **pruebas de significación**  $H_0 : \mu = \mu_0$ , se basan en el estadístico  $t$ . Utiliza los valores  $P$  o los niveles de significación predeterminados de la distribución  $t(n-1)$  para contrastar las hipótesis.

Utiliza estos procedimientos  $t$  de una muestra para analizar los datos de los **diseños por pares**. Primero tienes que calcular la diferencia dentro de cada par para obtener una sola muestra.

Los procedimientos  $t$  son relativamente **robustos** cuando la población no es normal, especialmente para tamaños de muestra grandes. Los procedimientos  $t$  son útiles con datos no normales cuando  $n \geq 15$ , a no ser que los datos contengan observaciones atípicas o muestren una fuerte asimetría.

## EJERCICIOS DE LA SECCIÓN 7.2

Cuando un ejercicio pida un valor  $P$ , dalo de forma exacta si tienes una calculadora adecuada o un programa estadístico. Si no es así, utiliza la tabla C para dar dos valores entre los que se encuentra  $P$ .

**7.14.** El estadístico  $t$  de una muestra para contrastar

$$H_0 : \mu = 10$$

$$H_a : \mu < 10$$

basado en  $n = 10$  observaciones tiene un valor  $t = -2,25$ .

(a) ¿Cuántos grados de libertad tiene este estadístico?

(b) ¿Entre qué dos probabilidades  $p$  de la tabla C se encuentra el valor  $P$  de la prueba?

**7.15. Investigación de mercado.** Un fabricante de pequeños electrodomésticos contrata una empresa de investigación de mercados para estimar las ventas de sus productos al por menor. Dicha empresa obtiene la información a partir de una muestra de tiendas minoristas. Este mes, una muestra aleatoria simple de 75 tiendas pone de manifiesto que este tipo de establecimientos vendieron como media 24 batidoras de este fabricante, con una desviación típica de 11.

(a) Calcula un intervalo de confianza del 95% para la media de las batidoras vendidas en todas las tiendas.

(b) La distribución de las ventas es muy asimétrica hacia la derecha, ya que hay muchas tiendas pequeñas y unas pocas muy grandes. La utilización de  $t$  en (a) es razonablemente segura a pesar de esta violación del supuesto de normalidad. ¿Por qué?

**7.16. Tiburones.** Los tiburones blancos son grandes y muy feroces. He aquí la longitud en metros de 44 ejemplares:<sup>7</sup>

5,61	3,69	5,58	4,92	4,71	5,49	4,38	4,74	4,47	5,28	3,63
4,92	5,01	5,28	4,86	3,78	5,34	4,14	3,66	4,56	4,41	3,72
3,96	4,34	4,29	4,98	2,82	5,46	3,96	4,08	4,59	4,83	4,05
5,73	4,86	6,84	5,04	4,08	3,96	4,71	5,91	5,61	3,96	5,04

(a) Examina la distribución de estos datos en relación con su forma, centro, dispersión y presencia de observaciones atípicas. La distribución es razonablemente normal excepto por la presencia de dos observaciones atípicas, una en cada dirección. Debido a que estas observaciones no son extremas y además se conserva la simetría de la distribución, la utilización de los procedimientos  $t$  es segura con 44 observaciones.

(b) Calcula un intervalo de confianza del 95% para la media de la longitud de los tiburones blancos. Basándote en este intervalo, ¿existe evidencia significativa al 5% para rechazar la afirmación: “como media los tiburones blancos miden 6 metros”?

(c) No queda del todo claro qué parámetro  $\mu$  estimaste en (b). Para decir exactamente qué es  $\mu$ , ¿qué información necesitas?

**7.17. Calcio y presión sanguínea.** En un experimento comparativo aleatorizado sobre el efecto del calcio de la dieta sobre la presión sanguínea, unos investigadores dividieron al azar a 54 hombres sanos en dos grupos. Uno recibió calcio; el otro, un placebo. Al comienzo del estudio, los investigadores midieron una serie de variables en los sujetos. El informe del estudio da  $\bar{x} = 114,9$  y  $s = 9,3$  para la presión sanguínea sistólica de los 27 sujetos del grupo placebo.

(a) Calcula un intervalo de confianza del 95% para la media de la presión sanguínea de la población de la que proceden estos sujetos.

<sup>7</sup>Datos de Chris Olsen.

(b) ¿Qué supuestos en relación con la población y el diseño del estudio exige el procedimiento que utilizaste en (a)? ¿Cuáles de estos supuestos son, en este caso, importantes para la validez del procedimiento?

**7.18. Calibración de instrumentos.** La cromatografía de gases es una técnica utilizada para analizar pequeñas concentraciones de determinados compuestos. Los cromatógrafos de gases se calibran haciendo lecturas repetidas de un patrón con una concentración conocida. Un estudio sobre la calibración de estos aparatos utiliza un patrón que contiene 1 nanogramo por litro (es decir,  $10^{-9}$  gramos) de un determinado compuesto. Las lecturas que se obtienen son las siguientes.<sup>8</sup>

21,6   20,0   25,0   21,9

Se sabe por experiencia que las lecturas varían de acuerdo con una distribución normal, a no ser que una observación atípica indique un error en el análisis. Estima la media de las lecturas obtenidas y halla el error de estimación para un nivel de confianza de tu elección. Luego explica a un químico que no sepa estadística lo que significa tu error de estimación.

**7.19. Embriones de artemia.** Los embriones de artemia pueden entrar en una fase de latencia en la cual la actividad metabólica desciende a un nivel bajo. Unos investigadores que estudiaron esta fase de latencia determinaron los niveles de varias sustancias importantes para el metabolismo. Los investigadores presentaron sus resultados en forma de tabla, con la nota, “los valores son medias  $\pm$  ETM de tres muestras independientes”. El valor del nivel de ATP en la tabla era  $0,84 \pm 0,01$ . Los biólogos que leen el artículo tienen que ser capaces de descifrar qué significa esto.<sup>9</sup>

(a) ¿Qué significa la abreviatura ETM?

(b) Los investigadores realizaron tres determinaciones del nivel de ATP, que dieron  $\bar{x} = 0,84$ . Calcula un intervalo de confianza del 90% para la media del nivel de ATP en los embriones de artemia en latencia.

**7.20. Adopción de la cultura anglosajona.** La prueba ARSMA (Acculturation Rating Scale for Mexican Americans) mide el grado de adopción de la cultura anglosajona por parte de los estadounidenses de origen mexicano. Durante la etapa de

<sup>8</sup>D. A. Kurtz (ed.), *Trace Residue Analysis*, American Chemical Society Symposium Series, n° 284, 1985.

<sup>9</sup>S. C. Hand y E. Gnaiger, “Anaerobic dormancy quantified in *Artemia* embryos”, *Science*, 239, 1988, págs. 1.425-1.427.

elaboración del ARSMA, se sometió a la prueba a un grupo de 17 mexicanos. Sus puntuaciones, en un intervalo posible de 1,00 a 5,00, mostraron una distribución simétrica con  $\bar{x} = 1,67$  y  $s = 0,25$ . Debido a que resultados bajos indicarían una fuerte presencia de la cultura mexicana, estos resultados ayudaron a validar la prueba.<sup>10</sup>

(a) Calcula un intervalo de confianza del 95% para la media de los resultados de los mexicanos en la prueba ARSMA.

(b) ¿Qué supuestos exige tu intervalo de confianza? ¿Cuál de estos supuestos es el más importante en este caso.

**7.21. Cicatrización de heridas.** Entre distintos puntos de la piel existen de forma natural diferencias de potencial. El campo eléctrico creado espontáneamente por este hecho, ¿facilita la cicatrización de las heridas? Si fuese así, un cambio en el campo eléctrico podría ralentizar la cicatrización. Los sujetos del estudio son tritones anestesiados. Haz un corte, con una navaja, en la piel de cada una de las patas traseras de un tritón. Deja que una de las heridas cicatrice de forma natural (el grupo de control). Utiliza un electrodo para bajar a la mitad el campo eléctrico en la piel de la otra pata. Después de dos horas, mide la velocidad de cicatrización. La tabla 7.4 proporciona las velocidades de cicatrización (en micrómetros por hora) de 14 tritones.<sup>11</sup>

(a) Como es habitual, el artículo donde se publicó este experimento no proporcionaba los resultados obtenidos en el mismo. Se supone que los lectores son capaces de interpretar los resúmenes que aparecen en dicho artículo. Éste resumió las diferencias de la tabla como  $-5,71 \pm 2,82$  y decía: “Todos los valores se expresan como medias  $\pm$  el error típico de la media”. Indica de dónde vienen los números  $-5,71$  y  $2,82$ .

(b) Los investigadores querían saber si un cambio en el campo eléctrico reduciría la velocidad media de cicatrización de las heridas. Plantea las hipótesis, lleva a cabo la prueba e indica tus conclusiones. El resultado de la prueba, ¿es estadísticamente significativo a un nivel de significación del 5%? ¿Y a un nivel del 1%? (Los investigadores compararon diversos campos eléctricos y llegaron a la conclusión de que el campo eléctrico natural es el más adecuado para acelerar la cicatrización.)

<sup>10</sup>I. Cuellar, L. C. Harris y R. Jasso, “An acculturation scale for Mexican American normal and clinical populations”, *Hispanic Journal of Behavioral Sciences*, 2, 1980, págs. 199-217.

<sup>11</sup>D. D. S. Iglesia, E. J. Cragoe, Jr. y J. W. Venable, “Electric field strength and epithelization in the newt (*Notophthalmus viridescens*)”, *Journal of Experimental Zoology*, 274, 1996, págs. 56-62.

Tabla 7.4. Velocidades de cicatrización de heridas en la piel de tritones.

Tritón	Pata experimental	Pata control	Diferencia
13	24	25	-1
14	23	13	10
15	47	44	3
16	42	45	-3
17	26	57	-31
18	46	42	4
19	38	50	-12
20	33	36	-3
21	28	35	-7
22	28	38	-10
23	21	43	-22
24	27	31	-4
25	25	26	-1
26	45	48	-3

(c) Calcula un intervalo de confianza del 90% para el cambio medio de la velocidad de cicatrización como consecuencia de la modificación del campo eléctrico. A continuación, explica qué significa que tenemos “una confianza del 90%” en el resultado.

**7.22. ARSMA versus BI.** La prueba ARSMA (ejercicio 7.20) se comparó con una prueba similar, la prueba BI (Bicultural Inventory), haciéndolas pasar a 22 ciudadanos de EE UU de origen mexicano. Las dos pruebas tienen el mismo intervalo de resultados (de 1,00 a 5,00) y se ajustaron de manera que las medias de las dos pruebas al aplicarse a los grupos experimentales fueran similares. Se observó una alta correlación entre los resultados de las dos pruebas, lo que es una evidencia a favor de que ambas miden las mismas características. Los investigadores querían saber si las medias de los resultados poblacionales eran iguales en las dos pruebas. Las diferencias de los resultados (ARSMA – BI) de los 22 sujetos fueron  $\bar{x} = 0,2519$  y  $s = 0,2767$ .

(a) Describe brevemente cómo organizarías la aplicación de las dos pruebas a los 22 sujetos. Incluye la aleatorización.

(b) Lleva a cabo una prueba de significación para la hipótesis de que las dos pruebas tienen la misma media poblacional. Halla el valor  $P$  y justifica tus conclusiones.

(c) Calcula un intervalo de confianza del 95% para la diferencia entre los resultados medios de las dos pruebas.

**7.23. Salarios de altos ejecutivos.** Un estudio sobre el salario de los altos ejecutivos examinó los ingresos, después de tener en cuenta la inflación, de esta clase de ejecutivos en 104 empresas durante el periodo 1977-1988. Entre los datos había las medias de los aumentos salariales anuales de cada uno de los 104 ejecutivos. La media de los aumentos porcentuales era del 6,9%. Los datos mostraron una gran variación, con una desviación típica del 17,4%. La distribución era claramente asimétrica hacia la derecha.<sup>12</sup>

(a) A pesar de la asimetría de la distribución, no había observaciones atípicas extremas. Explica por qué podemos utilizar los procedimientos  $t$  con estos datos.

(b) ¿Cuántos grados de libertad hay? Cuando los grados de libertad exactos no están en la tabla C, utiliza los grados de libertad con el valor menor inmediato de la tabla.

(c) Calcula un intervalo de confianza del 99% para el aumento medio de los salarios de los altos ejecutivos. ¿Qué condición esencial deben cumplir los datos para que los resultados sean fiables?

**7.24. Comparación de dos medicamentos.** Los fabricantes de medicamentos genéricos deben demostrar que éstos no son significativamente distintos de los medicamentos de “referencia” con la misma materia activa. Un aspecto en el que pueden diferir ambos medicamentos es en su absorción en la sangre. La tabla 7.5 proporciona datos de absorción de los dos tipos de medicamentos obtenidos en 20 hombres sanos no fumadores.<sup>13</sup> Estamos ante un diseño por pares. Los sujetos del 1 al 10 tomaron primero el medicamento genérico y los sujetos del 11 al 20 tomaron en primer lugar el medicamento de referencia. Entre la toma de ambos medicamentos se dejó pasar suficiente tiempo para asegurar que los restos del medicamento tomado en primer lugar desaparecieran de la sangre. Los números de referencia de cada sujeto se escogieron al azar para asegurar la aleatorización en el orden de toma de los medicamentos.

(a) Analiza los datos de las diferencias de absorción entre los medicamentos de referencia y los genéricos. ¿ Existe algún motivo para no aplicar los procedimientos  $t$ ?

(b) Utiliza una prueba  $t$  para responder a la pregunta clave: ¿los medicamentos difieren significativamente en su absorción?

<sup>12</sup>Charles W. L. Hill y Phillip Phan, “CEO tenure as a determinant of CEO pay”, *Academy of Management Journal*, 34, 1991, págs. 707-717.

<sup>13</sup>Liannng Yuh, “A biopharmaceutical example for undergraduate students”, manuscrito no publicado.

Tabla 7.5. Absorción de medicamentos de referencia y genéricos.

Sujeto	Medicamento de referencia	Medicamento genérico	Sujeto	Medicamento de referencia	Medicamento genérico
15	4.108	1.755	4	2.344	2.738
3	2.526	1.138	16	1.864	2.302
9	2.779	1.613	6	1.022	1.284
13	3.852	2.254	10	2.256	3.052
12	1.833	1.310	5	938	1.287
8	2.463	2.120	7	1.339	1.930
18	2.059	1.851	14	1.262	1.964
20	1.709	1.878	11	1.438	2.549
17	1.829	1.682	1	1.735	3.340
2	2.594	2.613	19	1.020	3.050

**7.25. Edades de los presidentes.** La tabla 1.7 ofrece las edades de los presidentes de EE UU cuando tomaron posesión del cargo. No tiene sentido utilizar los procedimientos  $t$  (o cualquier otro procedimiento estadístico) para calcular un intervalo de confianza del 95% para la media de la edad de los presidentes. Explica por qué.

**7.26 (Optativo). Potencia de la prueba  $t$ .** El banco del ejercicio 7.12 contrastó una nueva idea con una muestra de 200 clientes. A un nivel de significación  $\alpha = 0,01$ , el banco quiere detectar un incremento medio de  $\mu = 100$  € en la cantidad cargada. Quizás esto se podría conseguir con una muestra de sólo  $n = 50$  clientes. Halla de forma aproximada la potencia de la prueba con  $n = 50$ , con una alternativa  $\mu = 100$ , de la siguiente manera:

(a) ¿Cuál es el valor crítico  $t^*$  de la prueba de una cola con  $\alpha = 0,01$  y  $n = 50$ ?

(b) Describe el procedimiento para rechazar  $H_0 : \mu = 0$  en términos del estadístico  $t$ . Luego, utiliza  $s = 108$  (una estimación basada en los datos en los datos del ejercicio 7.12). Describe el criterio de rechazo de  $H_0$  en términos de  $\bar{x}$ .

(c) Supón que  $\mu = 100$  (la alternativa considerada) y que  $\sigma = 108$  (una estimación basada en los datos del ejercicio 7.12). La potencia aproximada es la probabilidad del suceso que hallaste en (b) calculada, ahora, bajo estos nuevos supuestos. Halla la potencia. ¿Recomendarías que el banco contrastara la nueva idea con sólo 50 clientes? ¿Deberían incluirse más clientes?

**7.27 (Optativo). Potencia de la prueba  $t$ .** El ejercicio 7.22 habla de un pequeño estudio que compara la prueba ARSMA con la prueba BI, dos pruebas que determinan la orientación cultural de los ciudadanos de EE UU de origen mexicano.



Este estudio, ¿detectaría una diferencia de 0,2 en los resultados medios? Para contestar a esta pregunta, calcula la potencia aproximada del contraste (con  $n = 22$  sujetos y  $\alpha = 0,05$ ) de

$$H_0 : \mu = 0$$

$$H_a : \mu \neq 0$$

Con la alternativa  $\mu = 0,2$ . Suponemos que  $\sigma$  es conocido.

(a) ¿Cuál es el valor crítico de la tabla C para  $\alpha = 0,05$ ?

(b) Describe el procedimiento para rechazar  $H_0$  a un nivel  $\alpha = 0,05$ . Luego, toma  $s = 0,3$ , el valor aproximado observado en el ejercicio 7.22, y expresa el criterio de rechazo en términos de  $\bar{x}$ . Fíjate en que es una prueba de dos colas.

(c) Halla la probabilidad de este suceso cuando  $\mu = 0,2$  (la alternativa considerada) y  $\sigma = 0,3$  (estimada con los datos del ejercicio 7.22) mediante un cálculo de probabilidad normal. Esta probabilidad es la potencia aproximada.

### 7.3 Comparación de dos medias

La comparación de dos poblaciones o de dos tratamientos es una de las situaciones más comunes que hay que afrontar en estadística aplicada. A estas situaciones las llamaremos *problemas de dos muestras*.

#### PROBLEMAS DE DOS MUESTRAS

- El objetivo de la inferencia es la comparación de las respuestas de dos tratamientos o la comparación de las características de dos poblaciones.
- Tenemos una muestra distinta de cada población o de cada tratamiento.

#### 7.3.1 Problemas de dos muestras

Un problema de dos muestras puede surgir de un experimento comparativo aleatorizado que divida al azar a los sujetos en dos grupos y exponga a cada uno a un

tratamiento distinto. La comparación de dos muestras aleatorias seleccionadas independientemente de dos poblaciones también es un problema de dos muestras. A diferencia de los diseños por pares que hemos estudiado anteriormente, las unidades experimentales no se agrupan por pares. Las dos muestras pueden ser de distinto tamaño. Los procedimientos inferenciales para los datos de dos muestras son distintos de los procedimientos inferenciales de los datos por pares. He aquí algunos problemas típicos de dos muestras.

#### *EJEMPLO 7.6. Problemas de dos muestras*

(a) Un investigador médico está interesado en el efecto sobre la presión sanguínea de un incremento de calcio en la dieta. Por este motivo, el investigador lleva a cabo un experimento comparativo aleatorizado en el cual un grupo de sujetos recibe un suplemento de calcio y un grupo de control recibe un placebo.

(b) Un psicólogo desarrolla una prueba que determina la capacidad de un individuo para captar la personalidad de la gente que le rodea. El psicólogo quiere comparar la capacidad de estudiantes universitarios de sexo masculino con la de los de sexo femenino. Con este fin, el psicólogo, hace pasar la prueba a un grupo numeroso de estudiantes de cada sexo.

(c) Un banco quiere saber cuál de sus dos planes para incentivar la utilización de las tarjetas de crédito es mejor. El banco aplica cada plan a una muestra aleatoria simple de sus clientes y después compara la cantidad cargada durante seis meses en las tarjetas de crédito de los sujetos de los dos grupos. ■

### **APLICA TUS CONOCIMIENTOS**

**7.28. ¿Qué diseño experimental?** Las siguientes situaciones exigen hacer inferencia sobre una o dos medias. Identifica cada situación como (1) de una muestra, (2) de un diseño por pares o (3) de dos muestras. Los procedimientos de la sección 7.2 se aplican a los casos (1) y (2). Estamos a punto de iniciar el estudio de los procedimientos que se aplican al caso (3).

(a) Un pedagogo quiere saber si es más efectivo plantear preguntas antes o después de introducir un nuevo concepto en un texto de matemáticas destinado a la enseñanza primaria. El pedagogo prepara dos extractos del texto que introducen el concepto; uno con preguntas para suscitar el interés de los niños antes de introducir el concepto y el otro con preguntas de repaso después de introducir

el concepto. El pedagogo utiliza cada extracto del texto con un grupo distinto de niños y compara los resultados de los dos grupos mediante un examen sobre este tema.

(b) Una pedagoga enfoca el mismo problema de forma diferente. Esta pedagoga prepara extractos de un texto sobre dos temas sin relación entre sí. Cada extracto tiene dos versiones, una con preguntas antes de introducirlos y otra con preguntas después de introducir los conceptos. Los sujetos son un solo grupo de niños. Cada niño estudia los dos temas, uno (escogido al azar) con preguntas antes y otro con preguntas después de los nuevos conceptos. La investigadora compara los resultados de los exámenes de cada alumno en los dos temas para ver qué tema aprendió mejor.

**7.29. ¿Qué diseño experimental?** Las siguientes situaciones exigen hacer inferencia para una o dos medias. Identifica cada situación como (1) de una muestra, (2) de un diseño por pares o (3) de dos muestras. Los procedimientos de la sección 7.2 se aplican a los casos (1) y (2). Estamos a punto de iniciar el estudio de los procedimientos que se aplican al caso (3).

(a) Para comprobar la fiabilidad de un nuevo método de análisis, una química tiene un patrón con una concentración conocida de una determinada sustancia. La química lleva a cabo con el nuevo método 20 análisis de la concentración de la sustancia en este patrón y comprueba si existe un sesgo comparando la media de los resultados con la concentración conocida.

(b) Otra química comprueba la fiabilidad del mismo método de análisis utilizando otro procedimiento. No tiene ninguna muestra de referencia, pero dispone de un método analítico conocido. La química quiere saber si los resultados de los dos métodos concuerdan. Con este objetivo, prepara un patrón de concentración desconocida y analiza su concentración 10 veces con el nuevo método y 10 veces con el método conocido.

### 7.3.2 Comparación de dos medias poblacionales

Podemos examinar gráficamente los datos de dos muestras comparando sus diagramas de tallos (para muestras pequeñas) o bien sus histogramas o diagramas de caja (para muestras grandes). Ahora aplicaremos las ideas de la inferencia formal a esta situación. Cuando las dos distribuciones poblacionales son simétricas y, especialmente, cuando al menos son aproximadamente normales, la comparación de las respuestas medias en las dos poblaciones es el objetivo más común de la inferencia. He aquí los supuestos que haremos.

### SUPUESTOS PARA LA COMPARACIÓN DE DOS MEDIAS

- Tenemos **dos muestras aleatorias simples** de dos poblaciones distintas. Las muestras son **independientes**. Es decir, una muestra no tiene ninguna influencia sobre la otra. Así, por ejemplo, la agrupación por pares viola la independencia. Medimos la misma variable en las dos muestras.
- Las dos poblaciones tienen **distribuciones normales**. Las medias y las desviaciones típicas de las dos poblaciones son desconocidas.

Llamaremos  $x_1$  a la variable que medimos en la primera población y  $x_2$  a la variable que medimos en la segunda, ya que la variable puede tener distribuciones distintas en las dos poblaciones. La notación que utilizaremos para describir las dos poblaciones es

Población	Variable	Media	Desviación típica
1	$x_1$	$\mu_1$	$\sigma_1$
2	$x_2$	$\mu_2$	$\sigma_2$

Existe un total de cuatro parámetros desconocidos, las dos medias y las dos desviaciones típicas. Los subíndices nos recuerdan qué población describe cada parámetro. Queremos comparar las dos medias poblacionales, calculando un intervalo de confianza para su diferencia  $\mu_1 - \mu_2$  o contrastando la hipótesis de que no existen diferencias,  $H_0 : \mu_1 = \mu_2$ .

Utilizamos las medias y las desviaciones típicas muestrales para estimar los parámetros desconocidos. De nuevo, los subíndices nos recuerdan de qué muestra procede cada estadístico. La notación que describen las muestras es

Población	Tamaño muestral	Media muestral	Desviación típica muestral
1	$n_1$	$\bar{x}_1$	$s_1$
2	$n_2$	$\bar{x}_2$	$s_2$

Para inferir sobre la diferencia  $\mu_1 - \mu_2$  entre las dos medias poblacionales, partimos de la diferencia entre las dos medias muestrales  $\bar{x}_1 - \bar{x}_2$ .

## EJEMPLO 7.7. ¿Se degrada el poliéster?

¿Con qué rapidez se degradan en el suelo productos sintéticos como el poliéster? Un investigador enterró tiras de poliéster en el suelo durante diferente número de semanas, luego desenterró las tiras y midió la fuerza necesaria para romperlas. Esta fuerza es fácil de medir y es un buen indicador de su estado de degradación. Fuerzas de rotura pequeñas son un indicio de que el producto se ha degradado.

Parte del estudio consistió en enterrar 10 tiras en un suelo bien drenado durante el verano. Cinco tiras, escogidas al azar, se desenterraron al cabo de 2 semanas; las otras 5 al cabo de 16. He aquí las fuerzas de rotura en kilogramos:<sup>14</sup>

2 semanas	260	278	278	265	284
16 semanas	273	216	243	309	243

A partir de los datos, calcula los resúmenes estadísticos:

Grupo	Tratamiento	$n$	$\bar{x}$	$s$
1	2 semanas	5	272,93	10,15
2	16 semanas	5	256,61	35,46

La tira que estuvo enterrada más tiempo tiene una fuerza de rotura media algo menor; asimismo, presenta más variabilidad. La diferencia observada en las fuerzas medias es

$$\bar{x}_1 - \bar{x}_2 = 272,93 - 256,61 = 16,32 \text{ kilogramos}$$

¿Existe evidencia clara de que el poliéster se degrada más en 16 que en dos semanas? ■

El ejemplo 7.7 corresponde a la comparación de dos muestras. Escribimos las hipótesis en términos de la fuerza media de rotura que observaríamos en toda la población de tiras de poliéster,  $\mu_1$  para las tiras enterradas durante 2 semanas y  $\mu_2$  para las tiras enterradas durante 16 semanas. Las hipótesis son

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 > \mu_2$$

<sup>14</sup>Sapna Aneja, "Biodeterioration of textile fibers in soil", M.S. thesis, Purdue University, 1994.

Queremos contrastar estas hipótesis y también estimar la disminución media de la fuerza de rotura,  $\mu_1 - \mu_2$ .

¿Se satisfacen los supuestos? Debido a la aleatorización, podemos considerar los dos grupos experimentales como dos muestras aleatorias simples independientes de tiras de poliéster. Aunque las muestras son pequeñas, examinando los datos, comprobamos que no existen signos de falta de normalidad importantes. He aquí un diagrama de tallos doble de las respuestas.

2 semanas		16 semanas
	21	6
	22	
	23	
	24	33
	25	
50	26	
88	27	3
4	28	
	29	
	30	9

El grupo de las 16 semanas presenta una mayor dispersión, tal como sugiere su desviación típica. Aunque tenemos pocas observaciones, no podemos decir que observemos ningún síntoma de falta de normalidad que impida la utilización de los procedimientos  $t$ .

### 7.3.3 Procedimientos $t$ de dos muestras

Para valorar la significación de la diferencia observada entre las medias de nuestras dos muestras, seguimos un procedimiento parecido. Que la diferencia observada entre dos muestras sea sorprendente depende tanto de la dispersión como de las dos medias. Diferencias muy grandes entre las medias pueden surgir simplemente por azar si las observaciones individuales varían mucho. Para tener en cuenta la variación, estandarizamos la diferencia observada dividiendo por su desviación típica. Esta desviación típica es

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Esta desviación típica se hace mayor cuando aumenta la variabilidad de la población, es decir, cuando  $\sigma_1$  o  $\sigma_2$  aumentan. Se hace menor cuando el tamaño de las muestras  $n_1$  y  $n_2$  aumenta.

Debido a que no conocemos las desviaciones típicas poblacionales, las estimamos a partir de las desviaciones típicas muestrales de las dos muestras. El resultado es el **error típico**, o desviación típica estimada, de la diferencia de las medias muestrales:

*Error típico*

$$ET = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

El resultado de la estandarización de la diferencia entre medias dividiendo por su error típico es el **estadístico  $t$  de dos muestras**:

*Estadístico  
 $t$  de dos  
muestras*

$$t = \frac{\bar{x}_1 - \bar{x}_2}{ET}$$

La interpretación del estadístico  $t$  es la misma que la de cualquier estadístico  $z$  o  $t$ : nos indica a qué distancia se encuentra  $\bar{x}_1 - \bar{x}_2$  de 0 tomando como unidad de medida la desviación típica.

Desgraciadamente, el estadístico  $t$  de dos muestras *no* tiene una distribución  $t$ . Sin embargo, en los problemas de inferencia sobre dos muestras, este estadístico se utiliza con los valores críticos  $t$ . Hay dos formas de hacerlo.

**Opción 1.** Utiliza los procedimientos basados en el estadístico  $t$  con los valores críticos de una distribución  $t$  con grados de libertad calculados a partir de los datos. Los grados de libertad generalmente no son números enteros. De esta forma se obtiene una aproximación muy exacta de la distribución  $t$ .

**Opción 2.** Utiliza los procedimientos basados en el estadístico  $t$  con los valores críticos de una distribución  $t$  con grados de libertad iguales al menor de los valores  $n_1 - 1$  y  $n_2 - 1$ . Estos procedimientos son siempre conservadores para dos poblaciones normales.

La mayoría de programas estadísticos utilizan, en los problemas de dos muestras, el estadístico  $t$  de dos muestras con la opción 1, a no ser que el usuario exija otro método. La utilización de esta opción sin la ayuda de un ordenador es algo complicada. Es por este motivo que presentamos en primer lugar la opción 2, que es más sencilla. Te recomendamos que utilices la opción 2 cuando hagas los cálculos sin la ayuda de un ordenador. Si utilizas un programa estadístico, el programa hará los cálculos, por defecto, con la opción 1. Más adelante, en una sección operativa, veremos algunos detalles sobre la opción 1. He aquí una descripción de la opción 2 que incluye una indicación de por qué se trata de un procedimiento "conservador".

### PROCEDIMIENTOS $t$ DE DOS MUESTRAS

Obtén una muestra aleatoria simple de tamaño  $n_1$  de la población normal de media  $\mu_1$  desconocida y una muestra aleatoria simple independiente de tamaño  $n_2$  de otra población normal de media  $\mu_2$  desconocida. El intervalo de confianza para  $\mu_1 - \mu_2$  dado por

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

tiene un nivel de confianza de *al menos*  $C$ , independientemente de cuáles sean las desviaciones típicas poblacionales. Aquí  $t^*$  es el valor crítico superior de  $\frac{(1-C)}{2}$  de la distribución  $t(k)$ , donde  $k$  es el menor de los valores  $n_1 - 1$  y  $n_2 - 1$ .

Para contrastar la hipótesis  $H_0 : \mu_1 - \mu_2$ , calcula el estadístico  $t$  de dos muestras

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

y utiliza los valores  $P$  o los valores críticos de la distribución  $t(k)$ . El verdadero valor  $P$  o el nivel de significación predeterminado siempre será igual o menor que el valor calculado a partir de  $t(k)$ , independientemente de cuáles sean los valores que tengan desviaciones poblacionales desconocidas.

Fíjate que el intervalo de confianza de la  $t$  de dos muestras, de nuevo tiene la forma

$$\text{estimación} \pm t^* \text{ET}_{\text{de la estimación}}$$

Este procedimiento  $t$  de dos muestras siempre es seguro, puesto que da valores  $P$  mayores y niveles de confianza menores que los verdaderos valores. La diferencia entre los valores obtenidos y los verdaderos suele ser bastante pequeña, a no ser que el tamaño de las dos muestras sea pequeño y desigual. A medida que aumenta el tamaño de las muestras, los valores probabilísticos basados en



la distribución  $t$  con grados de libertad iguales al menor de los valores  $n_1 - 1$  y  $n_2 - 1$  son cada vez más exactos.<sup>15</sup>

## APLICA TUS CONOCIMIENTOS

**7.30. Tratamiento de la encefalopatía espongiiforme.** La encefalopatía espongiiforme es una enfermedad degenerativa que afecta al sistema nervioso central. En un estudio sobre el tratamiento de esta enfermedad, se probó el IDX. Como sujetos experimentales se utilizaron 20 hámsteres infectados. Se inyectó IDX a 10 hámsteres escogidos al azar. Los restantes 10 no recibieron ningún tratamiento. Los investigadores registraron la supervivencia de cada hámster. Dice el informe: “De esta manera, todos los hámsteres que actuaron como control murieron 94 días después de la infección (media  $\pm$  ETM =  $88,5 \pm 1,9$  días). Mientras que los hámsteres que fueron tratados con IDX sobrevivieron más de 128 días (media  $\pm$  ETM =  $116 \pm 5,6$  días)”.<sup>16</sup>

(a) Completa la tabla siguiente:

Grupo	Tratamiento	$n$	$\bar{x}$	$s$
1	IDX	?	?	?
2	Control	?	?	?

(b) Si utilizas el procedimiento  $t$  de dos muestras conservador, ¿cuántos grados de libertad tendrá  $t$ ?

**7.31. Empresas en quiebra.** Un estudio compara una muestra de empresas griegas que quebraron con una muestra de empresas griegas sin problemas. Un indicador de la salud financiera de una empresa es el cociente entre el activo y el pasivo de la empresa,  $A/P$ . El año anterior al de la quiebra, el estudio halló un

<sup>15</sup>Información detallada sobre los procedimientos  $t$  conservadores se puede hallar en las siguientes publicaciones. Paul Leaverton y John J. Birch, “Small sample power curves for the two sample location problem”, *Technometrics*, 11, 1969, págs. 299-307. En Henry Scheffé, “Practical solutions of the Behrens-Fisher problem”, *Journal of the American Statistical Association*, 65, 1970, págs. 1.501-1.508. D. J. Best y J. C. W. Rayner, “Welch’s approximate solution for the Behrens-Fisher problem”, *Technometrics*, 29, 1987, págs. 205-210.

<sup>16</sup>F. Tagliavini *et al.*, “Effectiveness of anthracycline against experimental prion disease in Syrian hamsters”, *Science*, 276, 1997, págs. 1.119-1.121.

$A/P$  medio de 1,72565 en el grupo de empresas sin problemas y de 0,78640 en el grupo de empresas que quebró. El estudio afirma que  $t = 7,36$ .<sup>17</sup>

(a) Puedes sacar conclusiones a partir de esta  $t$  sin utilizar una tabla e incluso sin conocer los tamaños de las muestras (siempre que éstas no sean muy pequeñas). ¿Cuál es tu conclusión? ¿Por qué no necesitas conocer el tamaño de la muestra y la tabla de valores críticos de  $t$ ?

(b) En realidad, el estudio investigó a 33 empresas que quebraron y a 68 empresas sin problemas. ¿Cuántos grados de libertad utilizarías para la prueba  $t$  si consideras la aproximación conservadora que se recomienda utilizar cuando no se dispone de un programa estadístico?

### 7.3.4 Ejemplos de procedimientos $t$ de dos muestras

*EJEMPLO 7.8. ¿Se degrada el poliéster?*

Utilizaremos los procedimientos  $t$  de dos muestras para comparar las fuerzas de rotura de las tiras del ejemplo 7.7. El estadístico de contraste para la hipótesis nula  $H_0 : \mu_1 - \mu_2$  es

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{272,93 - 256,61}{\sqrt{\frac{10,15^2}{5} + \frac{35,46^2}{5}}} \\ &= \frac{16,32}{16,50} = 0,989 \end{aligned}$$

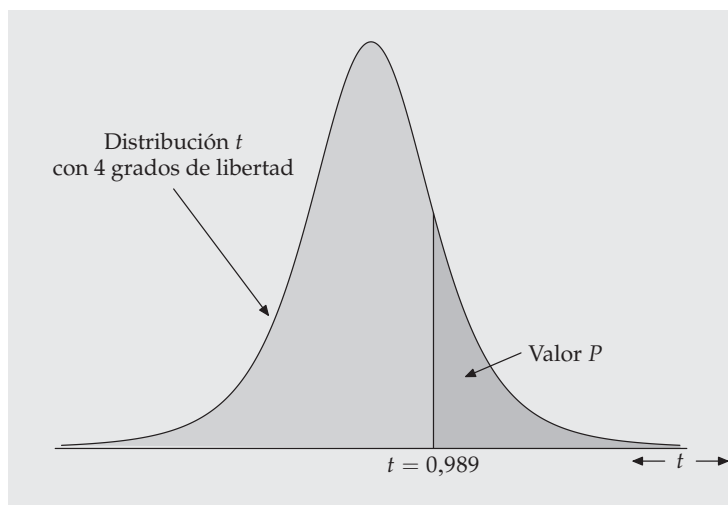
Tenemos 4 grados de libertad, el menor de  $n_1 - 1 = 4$  y  $n_2 - 1 = 4$ . Como  $H_a$  es de una cola, la cola de la derecha, el valor  $P$  es el área situada a la derecha de  $t = 0,989$  por debajo de la curva  $t(4)$ . La figura 7.7 ilustra este valor  $P$ . La tabla C muestra que se halla entre 0,15 y 0,20. Un programa estadístico nos dice que el verdadero valor es  $P = 0,189$ . El experimento no halló evidencia de que el poliéster se degradara más después de 16 semanas que después de dos semanas.

<sup>17</sup>Costas Papoulias y Panayiotis Theodossiou, "Analysis and modeling of recent business failures in Greece", *Managerial and Decision Economics*, 13, 1992, págs. 163-169.

Para un intervalo de confianza del 90%, la tabla C indica que el valor crítico de  $t(4)$ , es  $t^* = 2,132$ . Tenemos una confianza del 90% de que la diferencia entre las fuerzas medias de rotura después de 2 y de 16 semanas,  $\mu_1 - \mu_2$ , se halle en el intervalo

$$\begin{aligned} (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} &= (272,93 - 256,61) \pm 2,132 \sqrt{\frac{10,15^2}{5} + \frac{35,46^2}{5}} \\ &= 16,32 \pm 35,17 \\ &= -18,85 \text{ a } 51,49 \end{aligned}$$

Que el cero esté incluido en el intervalo de confianza del 90%, nos lleva a no poder rechazar la  $H_0 : \mu_1 = \mu_2$ , de una prueba de dos colas, para un nivel de significación  $\alpha = 0,10$ . ■



**Figura 7.7.** Valor  $P$  del ejemplo 7.8. En este ejemplo se ha utilizado el procedimiento conservador, según el cual la distribución  $t$  tiene 4 grados de libertad.

El tamaño de la muestra tiene una gran influencia en el valor  $P$  de una prueba. Un efecto que no sea significativo para un determinado nivel  $\alpha$  en una muestra pequeña será significativo en una muestra mayor. Como las muestras del ejemplo 7.7 son más bien pequeñas, sospechamos que más datos podrían indicar que una mayor duración del tiempo de enterramiento, reduce la fuerza de forma significativa. Incluso si fuera significativo, la reducción puede ser bastante pequeña.

Los datos sugieren que el poliéster enterrado se degrada muy lentamente, ya que la media de la fuerza disminuyó sólo de 272,93 a 256,61 kilos entre 2 y 16 semanas.

*EJEMPLO 7.9. Prueba Chapin con hombres y mujeres*

La prueba Chapin (*Chapin Social Insight Test*) es una prueba psicológica diseñada con el objeto de determinar la precisión con la que un individuo capta la personalidad de los que le rodean. Los posibles resultados de la prueba van de 0 a 41. Durante el desarrollo de la prueba Chapin, ésta se aplicó a diferentes grupos de personas. He aquí los resultados obtenidos con un grupo de estudiantes universitarios de Arte de ambos sexos:<sup>18</sup>

Grupo	Sexo	<i>n</i>	$\bar{x}$	<i>s</i>
1	Hombres	133	25,34	5,05
2	Mujeres	162	24,94	5,44

¿Se puede afirmar que como media la capacidad de mujeres y de hombres para captar la personalidad de las personas que les rodean es distinta?

**Paso 1: Hipótesis.** Debido a que antes de analizar los datos no teníamos pensado que la diferencia entre hombres y mujeres pudiese ir en una u otra dirección, escogemos una hipótesis alternativa de dos colas. Las hipótesis son

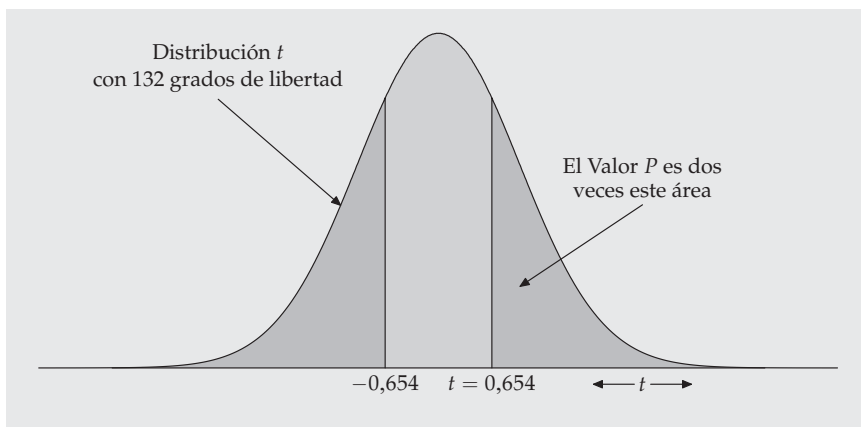
$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

**Paso 2: Estadístico de contraste.** El estadístico *t* de dos muestras es

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{25,34 - 24,94}{\sqrt{\frac{5,05^2}{133} + \frac{5,44^2}{162}}} \\ &= 0,654 \end{aligned}$$

<sup>18</sup>H. G. Gough, *The Chapin Social Insight Test*, Consulting Psychologists Press, Palo Alto, Calif., 1968.



**Figura 7.8.** Valor  $P$  del ejemplo 7.9. Para hallar el valor  $P$ , determina el área por encima de  $t = 0,654$  y dobla su valor, ya que la alternativa es de dos colas.

**Paso 3: Valor  $P$ .** Hay 132 grados de libertad, el menor de

$$n_1 - 1 = 133 - 1 = 132 \quad \text{y} \quad n_2 - 1 = 162 - 1 = 161$$

La figura 7.8 ilustra el valor  $P$ . Hállalo comparando 0,654 con valores críticos de la distribución  $t(132)$  y luego doblando  $p$ , ya que la alternativa es de dos colas. En la tabla C no aparece el valor correspondiente a 132 grados de libertad. En consecuencia, utilizaremos el valor de la tabla menor más próximo, 100 grados de libertad.

La tabla C muestra que 0,654 no alcanza el valor crítico 0,25, que es la mayor probabilidad de la cola de la derecha de la tabla C. El valor  $P$  es, por tanto, mayor que 0,50. Los datos no proporcionan suficiente evidencia de que existan diferencias entre hombres y mujeres en las medias de los resultados de la prueba Chapin ( $t = 0,654$ ,  $gl = 132$ ,  $P > 0,5$ ). ■

El investigador del ejemplo 7.9 no llevó a cabo ningún experimento, sino que comparó muestras de dos poblaciones. Muestras grandes permiten que el supuesto de que las poblaciones tengan distribuciones normales pierda importancia. Las medias muestrales serán, en cualquier caso, aproximadamente normales. El mayor problema que se plantea es saber a qué población se pueden aplicar las conclusiones. Estos estudiantes no son una muestra aleatoria simple de todos los estudiantes de Arte del país. Si son voluntarios de una sola universidad, los resultados muestrales no se pueden extender a una población más amplia.

## APLICA TUS CONOCIMIENTOS

**7.32. Explotación de la selva tropical.** La explotación forestal, ¿perjudica la selva tropical? Un estudio comparó parcelas forestales en Borneo que nunca se explotaron forestalmente con parcelas vecinas similares que se explotaron forestalmente 8 años antes. El estudio halló que los efectos de la explotación forestal eran menos severos de lo que se esperaba. He aquí datos sobre el número de especies arbóreas en 12 parcelas que nunca se explotaron y en 9 parcelas que se explotaron 8 años antes:<sup>19</sup>

Nunca explotadas	22	18	22	20	15	21	13	13	19	13	19	15
Explotadas	17	4	18	14	18	15	15	10	12			

(a) Dice el estudio, “los cultivadores no sabían que se valorarían los efectos de la explotación forestal”. ¿Por qué es importante que no lo supieran? En el estudio también se puede leer que las parcelas se podían considerar como escogidas al azar.

(b) Después de 8 años, ¿la explotación forestal reduce el número medio de especies? Plantea las hipótesis y lleva a cabo la prueba  $t$ . El resultado de la prueba, ¿es significativo a un nivel del 5%? ¿Y a un nivel del 1%?

(c) Calcula un intervalo de confianza del 90% para la diferencia entre la media del número de especies en las parcelas que no se explotaron nunca y las que se explotaron desde hacía 8 años.

**7.33. Cada día soy mejor en matemáticas.** Un mensaje subliminal es un mensaje que no podemos percibir de forma consciente, pero que sin embargo nos influye. Los mensajes subliminales, ¿pueden ayudar a los estudiantes a aprender matemáticas? Un grupo de estudiantes que suspendieron matemáticas, estuvieron de acuerdo en participar en un estudio para averiguarlo.

Todos ellos recibieron mensajes subliminales, que aparecían en la pantalla de sus televisores durante lapsos de tiempo tan breves que era imposible leerlos de forma consciente. El grupo experimental estaba formado por 10 estudiantes escogidos al azar que recibieron el mensaje: “Cada día soy mejor en matemáticas”. El grupo de control estaba formado por 8 estudiantes que recibieron un

<sup>19</sup>C. H. Cannon, D. R. Peart, y M. Leighton, “Tree species diversity in commercially logged Bornean rainforest”, *Science*, 281, 1998, págs. 1.366-1.367.

mensaje neutro: “La gente pasea por la calle”. Todos los estudiantes participaron en un curso de verano que tenía como objetivo subir su nivel de matemáticas. Al final del curso, todos los estudiantes pasaron una prueba para valorar su nivel. La tabla 7.6 proporciona datos sobre las puntuaciones recibidas por los estudiantes, antes y después del curso.<sup>20</sup>

Tabla 7.6. Notas de matemáticas antes y después de un mensaje subliminal.

Grupo experimental		Grupo de control	
Calificación antes del mensaje	Calificación después del mensaje	Calificación antes del mensaje	Calificación después del mensaje
18	24	18	29
18	25	24	29
21	33	20	24
18	29	18	26
18	33	24	38
20	36	22	27
23	34	15	22
23	36	19	31
21	34		
17	27		

(a) ¿Existe evidencia de que el mensaje subliminal del grupo experimental contribuyó más a aumentar el nivel de matemáticas que el mensaje neutro del grupo de control? Plantea las hipótesis y lleva a cabo la prueba. ¿Cuáles son tus conclusiones? Tu resultado, ¿es significativo a un nivel del 5%? ¿Y a un nivel del 10%?

(b) Calcula un intervalo de confianza para la diferencia entre la media del grupo experimental y la del grupo de control.

**7.34. Escarabajos de la avena.** En un estudio sobre los daños causados por escarabajos en cultivos de avena, unos investigadores contaron el número de larvas de escarabajo por tallo en pequeñas parcelas sembradas después de aplicar aleatoriamente uno de los dos tratamientos siguientes: ningún insecticida o malathion (un insecticida) a una dosis de 275 gramos por hectárea. Los datos aparecen como aproximadamente normales. He aquí los estadísticos de resumen.<sup>21</sup>

<sup>20</sup>Datos proporcionados por Warren Page, New York City Technical College. Son datos de un estudio de John Hudesman.

<sup>21</sup>M. C. Wilson et al., “Impact of cereal leaf beetle larvae on yields of oats”, *Journal of Economic Entomology*, 62, 1969, págs. 699-702.

Grupo	Tratamiento	$n$	$\bar{x}$	$s$
1	Control	13	3,47	1,21
2	Malathion	14	1,36	0,52

A un nivel de significación del 1%, ¿los datos proporcionan suficiente evidencia de que el malathion reduce el número medio de larvas por tallo? Asegúrate de plantear  $H_0$  y  $H_a$ .

### 7.3.5 Otra vez la robustez

Los procedimientos  $t$  de dos muestras son más robustos que los procedimientos  $t$  de una muestra, especialmente cuando las distribuciones no son simétricas. Cuando los tamaños de las dos muestras son iguales y las dos poblaciones que se comparan tienen distribuciones similares, las probabilidades de la tabla  $t$  son bastante exactas para un amplio espectro de distribuciones, incluso cuando el tamaño de las muestras es tan pequeño como  $n_1 = n_2 = 5$ .<sup>22</sup> Cuando la forma de las distribuciones de las dos poblaciones es distinta, se necesitan muestras mayores.

Como guía práctica, adapta las indicaciones de la sección 7.2 para la utilización de los procedimientos  $t$  de una muestra a los procedimientos  $t$  de dos muestras, mediante la sustitución de “tamaño de la muestra” por “suma de los tamaños de las muestras”,  $n_1 + n_2$ . Estas indicaciones son seguras, especialmente cuando las dos muestras son de igual tamaño. Cuando prepares un estudio de dos muestras, procura elegir, siempre que puedas, muestras de igual tamaño. Los procedimientos  $t$  de dos muestras son más robustos ante una posible falta de normalidad en este caso y los valores de las probabilidades conservadoras son más exactos.

## APLICA TUS CONOCIMIENTOS

**7.35. Un maíz más nutritivo.** El maíz común no tiene la cantidad de lisina que necesitan los animales en su pienso. Unos científicos han desarrollado ciertas

<sup>22</sup>Ver los siguientes estudios. Harry O. Posten, “The robustness of the two-sample  $t$ -test over the Pearson system”, *Journal of the Statistical Computation and Simulation*, 6, 1978, págs. 295-311. Harry O. Posten, H. Yeh y Donald B. Owen, “Robustness of the two-sample  $t$ -test under violations of the homogeneity assumption”, *Communications in Statistics*, 11, 1982, págs. 109-126.



variedades de maíz que contienen una mayor cantidad de lisina. En una prueba sobre la calidad del maíz con alto contenido en lisina destinado a pienso animal, un grupo experimental de 20 pollos de un día de edad empezó a recibir una ración que contenía el nuevo maíz. Un grupo de control de otros 20 pollos recibió una ración que era idéntica a la anterior, con la excepción de que contenía maíz normal. He aquí las ganancias de peso (en gramos) de los pollos a los 21 días.<sup>23</sup>

Grupo de control				Grupo experimental			
380	321	366	356	361	447	401	375
283	349	402	462	434	403	393	426
356	410	329	399	406	318	467	407
350	384	316	272	427	420	477	392
345	455	360	431	430	339	410	326

(a) Representa gráficamente los datos. ¿Hay observaciones atípicas o asimetrías claras que pudieran impedir la utilización de los procedimientos *t*?

(b) ¿Existe suficiente evidencia de que los pollos alimentados con el maíz con un alto contenido en lisina ganan peso más deprisa? Lleva a cabo una prueba. ¿Cuáles son tus conclusiones?

(c) Calcula un intervalo de confianza del 95% para la diferencia entre la media de los pollos alimentados con maíz con un alto contenido en lisina y la media de los pollos alimentados con maíz común.

**7.36. Actitud de los estudiantes.** La encuesta SSHA (*Survey of Study Habits and Attitudes*) es una prueba psicológica que mide la motivación, la actitud hacia la universidad y los hábitos de estudio de los estudiantes. Los resultados van de 0 a 200. Una selecta universidad privada pasa la encuesta SSHA a una muestra aleatoria simple de estudiantes de primer curso de ambos sexos. Los resultados de las mujeres son los siguientes:

154 109 137 115 152 140 154 178 101  
 103 126 126 137 165 165 129 200 148

Y los de los hombres:

108 140 114 91 180 115 126 92 169 146  
 109 132 75 88 113 151 70 115 187 104

<sup>23</sup>G. L. Cromwell *et al.*, "A comparison of the nutritive value of opaque-2, floury-2 and normal corn for the chick", *Poultry Science*, 47, 1968, págs. 840-847.

(a) Examina cada muestra gráficamente, prestando especial atención a las observaciones atípicas y a las asimetrías. ¿Es aceptable la utilización de un procedimiento  $t$  con estos datos?

(b) La mayoría de los estudios han hallado que la media de los resultados de la prueba SSHA de los hombres es menor que la media de los resultados de un grupo comparable de mujeres. ¿Es esto cierto para los estudiantes de primer curso de esta universidad? Lleva a cabo una prueba y resume tus conclusiones.

(c) Calcula un intervalo de confianza del 90% para la diferencia entre la media de los resultados en la prueba SSHA de los hombres y la de las mujeres estudiantes de primer curso de esta universidad.

### 7.3.6 Procedimientos $t$ de dos muestras más precisos\*

El estadístico  $t$  de dos muestras no tiene una distribución  $t$ . Es más, la distribución exacta cambia a medida que las desviaciones típicas poblacionales desconocidas,  $\sigma_1$  y  $\sigma_2$ , cambian. De todas formas, se dispone de una excelente aproximación.

#### DISTRIBUCIÓN APROXIMADA DEL ESTADÍSTICO $t$ DE DOS MUESTRAS

La distribución del estadístico  $t$  de dos muestras es aproximadamente una distribución  $t$  con los grados de libertad  $gl$  dados por

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2}$$

Esta aproximación es bastante precisa cuando ambos tamaños muestrales  $n_1$  y  $n_2$  son iguales o mayores que 5.

Los procedimientos  $t$  de dos muestras son exactamente iguales a los procedimientos  $t$  que hemos visto hasta ahora, la única diferencia es que utilizamos

\*La lectura de esta sección se puede omitir, a no ser que quieras entender cómo los programas estadísticos calculan las probabilidades.

la distribución  $t$  con  $gl$  grados de libertad para obtener los valores críticos y los valores  $P$ .

**EJEMPLO 7.10.** *¿Se degrada el poliéster?*

En el experimento de los ejemplos 7.7 y 7.8, los datos sobre el poliéster enterrado dieron

Grupo	Tratamiento	$n$	$\bar{x}$	$s$
1	2 semanas	5	272,93	10,15
2	16 semanas	5	256,61	35,46

Para mejorar la precisión, podemos utilizar los valores críticos de una distribución  $t$  con los grados de libertad dados por

$$\begin{aligned}
 gl &= \frac{\left(\frac{10,15^2}{5} + \frac{35,46^2}{5}\right)^2}{\frac{1}{4}\left(\frac{10,15^2}{5}\right)^2 + \frac{1}{4}\left(\frac{35,46^2}{5}\right)^2} \\
 &= \frac{74.031,24}{15.916,98} = 4,65
 \end{aligned}$$

Fíjate en que los grados de libertad  $gl$  no son un número entero. En este ejemplo, existe poca diferencia entre los 4 grados de libertad del procedimiento conservador y los 4,65 resultantes del procedimiento más elaborado. ■

Los procedimientos  $t$  de dos muestras son exactamente iguales a los de antes, siendo la única diferencia la utilización de una distribución  $t$  con más grados de libertad. El número  $gl$  es siempre al menos tan grande como el menor de los valores  $n_1 - 1$  y  $n_2 - 1$ . Por otro lado,  $gl$  nunca es mayor que la suma de los dos grados de libertad individuales  $n_1 + n_2 - 2$ . El número de grados de libertad  $gl$  no es, generalmente, un número entero. Existe una distribución  $t$  para cualquier valor positivo de los grados de libertad, a pesar de que la tabla C contiene sólo datos correspondientes a valores enteros de grados de libertad. Algunos programas estadísticos hallan  $gl$  y luego utilizan la distribución  $t$  que tiene el número entero positivo menor más próximo. Otros programas estadísticos utilizan  $t(gl)$  incluso cuando  $gl$  no es un número entero positivo. No te aconsejamos la utilización habitual de este método a no ser que el ordenador haga los cálculos. Con un ordenador, en cambio, el procedimiento más preciso no cuesta ningún esfuerzo.

He aquí los resultados del programa estadístico SAS para los datos de las fuerzas de rotura de la tiras de poliéster.<sup>24</sup>

## TTEST PROCEDURE

Variable: FUERZAS

SEMANAS	N	Mean	Std Dev	Std Error
2	5	273.0	10.04988	4.49444
16	5	256.8	35.47112	15.86316

Variances	T	DF	Prob> T
Unequal	0.983	4.64	0.3743
Equal	0.983	8	0.3546

El programa SAS proporciona los resultados de dos procedimientos *t*: el procedimiento de dos muestras usual (suponiendo que las dos varianzas poblacionales son distintas, *unequal variances*) y un procedimiento especial que supone que las dos varianzas poblacionales son iguales. Estamos interesados en el primero de estos dos procedimientos. El estadístico *t* de dos muestras toma el valor  $t = 0,9889$ , lo que está en consonancia con nuestros resultados del ejemplo 7.8. Los grados de libertad son  $gl = 4,7$ , un valor redondeado concordante con nuestro resultado del ejemplo 7.10. El valor *P* de dos colas de la distribución  $t(4,7)$  es 0,3718. Debido a que el SAS siempre da valores *P* de pruebas de dos colas, tenemos que dividir por 2 para hallar el valor *P* correspondiente a la prueba de una cola  $P = 0,1859$ .

La diferencia entre los procedimientos *t* que utilizan el método conservador y los que utilizan el método aproximado tiene poca importancia práctica. Es por este motivo que te recomendamos que utilices el método conservador, que es más simple, siempre que no dispongas de un ordenador.

## APLICA TUS CONOCIMIENTOS

**7.37. Envenenamiento con DDT.** En un experimento comparativo aleatorizado, unos investigadores compararon 6 ratas envenenadas con DDT con un grupo de control de 6 ratas no envenenadas. La medida de la actividad eléctrica de

<sup>24</sup>En el ejemplo 7.10 no utilizamos ni Minitab ni Data Desk. Estos programas calculan los grados de libertad de la prueba *t* de dos muestras utilizando el procedimiento más preciso; sin embargo, redondean hasta el número entero menor para hallar el valor *P*.

los nervios es la clave para conocer la naturaleza del envenenamiento por DDT. Cuando un nervio es estimulado, se produce en él una respuesta eléctrica pronunciada seguida por una segunda respuesta menor. El experimento halló que la segunda respuesta eléctrica era mayor en las ratas envenenadas con DDT que en las ratas del grupo de control.<sup>25</sup>

Los investigadores midieron la intensidad de la segunda respuesta al estimular un nervio de una pata de rata, como un porcentaje de la intensidad de la primera respuesta. En las ratas envenenadas los resultados fueron

12,207 16,869 25,050 22,429 8,456 20,589

Los datos del grupo de control fueron

11,074 9,686 12,064 9,351 8,182 6,642

Las dos poblaciones son razonablemente normales, en la medida en que se puede juzgar a partir de seis observaciones. He aquí los resultados del programa estadístico SAS con estos datos:

TTEST PROCEDURE

Variable: RESPUESTA

GROUP	N	Mean	Std Dev	Std Error
DDT	6	17.60000000	6.34014839	2.58835474
CONTROL	6	9.49983333	1.95005932	0.79610839

Variances	T	DF	Prob> T
Unequal	2.9912	5.9	0.0247
Equal	2.9912	10.0	0.0135

(a) Los investigadores querían saber si ratas envenenadas con DDT eran distintas que ratas no envenenadas. Plantea  $H_0$  y  $H_a$ .

(b) ¿Qué valor toma del estadístico  $t$  de dos muestras? ¿Y su valor  $P$ ? (Fíjate en que SAS proporciona el valor  $P$  de dos colas. Si necesitas el valor  $P$  de una cola, divide el valor  $P$  de dos colas por 2.) ¿Cuáles son tus conclusiones?

(c) Calcula un intervalo de confianza del 90% para la diferencia entre las medias de las ratas envenenadas y la de las ratas no envenenadas. (Cuando un programa proporciona los grados de libertad como un número no entero, utiliza como grados de libertad el valor entero menor más próximo de la tabla C.)

<sup>25</sup>D. L. Shankland, "Involvement of spinal cord and peripheral nerves in DDT-poisoning syndrome in albino rats", *Toxicology and Applied Pharmacology*, 6, 1964, págs. 197-213.

**7.38.** El ejercicio 7.37 muestra el análisis de datos sobre los efectos del DDT. El programa estadístico utiliza la prueba  $t$  de dos muestras con los grados de libertad calculados con el procedimiento más elaborado. A partir de los resultados de  $\bar{x}_i$  y  $s_i$  proporcionados por el ordenador, comprueba que el valor del estadístico de contraste  $t = 2,99$  y el de los grados de libertad  $gl = 5,9$ , proporcionados por el programa estadístico, sean correctos.

**7.39. Autoestima de estudiantes.** He aquí los resultados del SAS correspondientes a un estudio sobre la autoestima de estudiantes de séptimo. La variable  $SC$  es el resultado de una prueba (*Piers-Harris Self Concept Scale*). Se hizo el análisis para ver si había diferencias entre la media de autoestima de chicos y chicas.<sup>26</sup>

## TTEST PROCEDURE

Variable: SC

SEXO	N	Mean	Std Dev	Std Error
M	31	55.51612903	12.69611743	2.28029001
H	47	57.91489362	12.26488410	1.78901722

Variances	T	DF	Prob> T
Unequal	-0.8276	62.8	0.4110
Equal	-0.8336	76.0	0.4071

Resume el resultado del análisis muy brevemente, en una o dos líneas, como si tuvieras que preparar un informe.

### 7.3.7 Procedimientos $t$ de dos muestras con varianza común\*

En el ejemplo 7.10, el programa estadístico ofreció dos posibilidades de ejecución de la prueba  $t$ . Una llevaba el nombre en inglés de *unequal variances* (varianzas distintas) y otra llevaba el nombre de *equal variances* (varianzas iguales). El procedimiento para varianzas distintas es nuestro procedimiento  $t$  de dos muestras. Esta prueba es válida tanto si las varianzas poblacionales son iguales como si son distintas. La otra posibilidad es una versión especial del estadístico  $t$  de dos muestras que supone que las dos poblaciones tienen la misma varianza. Este procedimiento calcula la media ponderada, de las dos varianzas muestrales para estimar

<sup>26</sup>Datos proporcionados por Darlene Gordon, School of Education, Purdue University.

\*Este apartado es un tema especial optativo.

la varianza poblacional común. El estadístico resultante se llama el *estadístico t de dos muestras* con varianza común (en los programas de ordenador aparece a veces con el nombre inglés *pooled t statistic*). Es igual a nuestro estadístico *t* si el tamaño de las dos muestras es igual, pero no en caso contrario. Podríamos utilizar el estadístico *t* de dos muestras con varianza común en las pruebas de significación y en los intervalos de confianza.

El estadístico *t* de dos muestras con varianza común tiene la ventaja de que tiene exactamente una distribución *t* con  $n_1 + n_2 - 2$  grados de libertad si las dos varianzas poblacionales son realmente iguales. Obviamente, las varianzas poblacionales son, a menudo, distintas. Es más, el supuesto de igualdad de varianzas es difícil de comprobar a partir de los datos. La utilización del estadístico *t* de dos muestras con varianza común era frecuente antes de que el uso de los ordenadores facilitara el empleo de la aproximación exacta a la distribución de nuestro estadístico *t* de dos muestras. Actualmente sólo es útil en situaciones especiales. No podemos utilizar la *t* de dos muestras con varianza común en el ejemplo 7.10, ya que está claro que la varianza es mucho mayor entre las tiras que se enterraron durante 16 semanas.

### RESUMEN DE LA SECCIÓN 7.3

Los datos en un **problema de dos muestras** son dos muestras aleatorias simples independientes, cada una de ellas obtenida de una población distinta distribuida normalmente.

Las pruebas de significación y los intervalos de confianza para la diferencia entre las medias  $\mu_1$  y  $\mu_2$  de dos poblaciones parten de la diferencia  $\bar{x}_1 - \bar{x}_2$  entre las dos medias muestrales. Con distribuciones no normales, el teorema del límite central garantiza que los procedimientos de cálculo son aproximadamente correctos cuando las muestras son grandes.

Obtén muestras aleatorias simples independientes de tamaños  $n_1$  y  $n_2$  de dos poblaciones normales de parámetros  $\mu_1, \sigma_1$  y  $\mu_2, \sigma_2$ , respectivamente. **El estadístico *t* de dos muestras** es

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

El estadístico *t* no tiene exactamente una distribución *t*.

Los procedimientos de inferencia conservadores para comparar  $\mu_1$  y  $\mu_2$  utilizan el estadístico *t* de dos muestras con la distribución  $t(k)$ . El valor de los grados

de libertad  $k$  es el menor de  $n_1 - 1$  y  $n_2 - 1$ . Para valores de probabilidad más exactos, utiliza la distribución  $t$ (gl) con los grados de libertad gl estimados a partir de los datos. Este procedimiento es el que se utiliza normalmente en los programas estadísticos.

El **intervalo de confianza** para  $\mu_1 - \mu_2$  dado por

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

tiene un nivel de confianza de al menos  $C$  si  $t^*$  es el valor crítico superior  $\frac{(1-C)}{2}$  de  $t(k)$ , donde  $k$  es el menor de  $n_1 - 1$  y  $n_2 - 1$ .

Las **pruebas de significación** para  $H_0 : \mu_1 = \mu_2$  basadas en

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

tienen un verdadero valor  $P$  que no es mayor que el calculado para  $t(k)$ .

Los consejos prácticos sobre el uso de los procedimientos  $t$  de dos muestras son similares a los consejos prácticos sobre el uso del estadístico  $t$  de una sola muestra. Se recomienda que el tamaño de las muestras sea igual.

### EJERCICIOS DE LA SECCIÓN 7.3

En los ejercicios en los que deban utilizarse procedimientos  $t$  de dos muestras, puedes emplear como grados de libertad el más pequeño de  $n_1 - 1$  y  $n_2 - 1$  o el valor gl más exacto. Te recomendamos la primera opción, a no ser que utilices un ordenador. Muchos de estos ejercicios te piden que reflexiones sobre la aplicación práctica de la estadística, además de llevar a cabo los procedimientos  $t$ .

**7.40. Tratamiento de la encefalopatía esponjiforme.** El ejercicio 7.30 contiene los resultados de un estudio que determina si el IDX es un tratamiento efectivo para la encefalopatía esponjiforme.

(a) ¿Existe evidencia de que los hámsteres tratados con IDX viven como media más tiempo?

(b) Calcula un intervalo de confianza del 95% para la diferencia entre la media del grupo experimental y la del grupo de control.



**7.41. Talentos de 13 años.** Unos “buscadores de talentos” sometieron a la prueba SAT (*Scholastic Assessment Test*), pensada para jóvenes que han terminado sus estudios de secundaria, a muchachos de 13 años. Entre 1980 y 1982, participaron en las pruebas 19.883 muchachos y 19.937 muchachas. Los resultados medios de los dos sexos en la prueba de Lengua son casi iguales, pero hay una clara diferencia entre ambos sexos en la prueba de Matemáticas. No se conoce cuál es la razón de esta diferencia. He aquí los datos.<sup>27</sup>

Grupo	$\bar{x}$	$s$
Chicos	416	87
Chicas	386	74

Calcula un intervalo de confianza del 99% para la diferencia entre la media de los resultados de los muchachos y la media de los resultados de las muchachas de la población. Los resultados de la prueba SAT, ¿tienen que tener una distribución normal para que tu intervalo de confianza sea válido? ¿Por qué?

**7.42. ¿Formas extraterrestres?** Algunas moléculas presentan formas dextrorrotatorias y formas levorrotatorias. En la Tierra, la forma habitual de algunas moléculas que se hallan en la materia orgánica es la levorrotatoria. La existencia de formas levorrotatorias, ¿es anterior a la aparición de la vida? Para averiguarlo, unos científicos analizaron meteoritos procedentes del espacio exterior y material estándar procedente de la Tierra. He aquí los resultados que muestran los porcentajes de formas levorrotatorias de una determinada molécula en dos análisis distintos.<sup>28</sup>

Análisis	Meteorito			Estándar		
	$n$	$\bar{x}$	$s$	$n$	$\bar{x}$	$s$
1	5	52,6	0,5	14	48,8	1,9
2	10	51,7	0,4	13	49,0	1,3

Los investigadores utilizaron la prueba  $t$  para ver si el meteorito presentaba un porcentaje mayor de formas levorrotatorias. Lleva a cabo la prueba. Resume

<sup>27</sup>De un anuncio que apareció en *Science*, 224, 1983, págs. 1.029-1.031.

<sup>28</sup>John R. Cronin y Sandra Pizzarello, “Enantiometric excesses in meteoritic amino acids”, *Science*, 275, 1997, págs. 951-955.

tus resultados. Las conclusiones de los investigadores fueron: “Las observaciones sugieren que la materia orgánica de origen extraterrestre puede haber jugado un papel esencial en el origen de la vida terrestre”.

**7.43. Enseñanza activa versus enseñanza pasiva.** Un estudio sobre métodos de enseñanza asistidos por ordenador, utilizó pictogramas (piensa, por ejemplo, en los jeroglíficos egipcios) para la formación de niños con dificultades de comunicación. El investigador diseñó dos lecciones de ordenador para enseñar la misma materia, los mismos ejemplos. En una de las lecciones era necesario que los niños interaccionaran con el ordenador; en la otra, los niños simplemente podían controlar la velocidad de desarrollo de la lección. Vamos a llamar a estos dos tipos de enseñanza: “Activa” y “Pasiva”, respectivamente. Después de las lecciones, el ordenador presentaba una prueba en la que los niños tenían que identificar 56 pictogramas. El número de identificaciones correctas de los 24 niños del grupo Activo fueron las siguientes:<sup>29</sup>

29	28	24	31	15	24	27	23	20	22	23	21
24	35	21	24	44	28	17	21	21	20	28	16

En el grupo Pasivo, el número de identificaciones correctas fue:

16	14	17	15	26	17	12	25	21	20	18	21
20	16	18	15	26	15	13	17	21	19	15	12

(a) ¿Existe evidencia de que el método de enseñanza activo es mejor que el método pasivo? Plantea las hipótesis, lleva a cabo la prueba y halla el valor  $P$ . ¿Cuáles son tus conclusiones?

(b) Calcula un intervalo de confianza del 90% para el número medio de pictogramas identificados correctamente por una gran población de niños que han pasado la lección activa de ordenador.

(c) ¿En qué supuestos se basan los procedimientos utilizados en (a) y (b)? ¿En cuál de estos procedimientos puedes utilizar los datos para comprobar si se cumplen los supuestos? Para los procedimientos que sea oportuno, utiliza los datos para comprobar si se cumplen los supuestos que permiten su utilización. ¿Cuáles son tus conclusiones?

<sup>29</sup>Orit E. Hetzroni, “The effects of active versus passive computer-assisted instruction on the acquisition, retention, and generalization of Blissymbols while using elements for teaching compounds”, Ph.D. thesis, Purdue University, 1995.

**7.44. Coeficientes de inteligencia de chicos y chicas.** A continuación damos los resultados de una prueba de inteligencia de 31 chicas de secundaria de una determinada zona rural.<sup>30</sup>

114	100	104	89	102	91	114	114	103	105
108	130	120	132	111	128	118	119	86	72
111	103	74	112	107	103	98	96	112	93

Los coeficientes de inteligencia de 47 chicos de secundaria de la misma zona rural son

111	107	100	107	115	111	97	112	104	106	113
109	113	128	128	118	113	124	127	136	106	123
124	126	116	127	119	97	102	110	120	103	115
93	123	79	119	110	110	107	105	105	110	77
90	114	106								

(a) Halla la media de los coeficientes de las chicas y también la de los chicos. En general, en pruebas estándar, los resultados de los chicos son ligeramente superiores a los de las chicas. ¿Ocurre con nuestros datos?

(b) Dibuja diagramas de tallos o histogramas con los dos conjuntos de datos. Debido a que la distribución de los datos es razonablemente simétrica, sin la presencia de observaciones atípicas, se pueden utilizar los procedimientos  $t$ .

(c) Trata estos datos como si fueran una muestra aleatoria simple de los estudiantes de secundaria de una determinada zona rural. ¿Existe evidencia de que las medias de los coeficientes de inteligencia de chicas y chicos son distintas?

(d) Calcula un intervalo de confianza del 90% para la diferencia entre las medias de los coeficientes de inteligencia de todos los chicos y chicas de esta zona rural.

(e) Antes de aceptar los resultados como representativos de todos los chicos y chicas de secundaria de esta zona rural, ¿qué información adicional necesitas?

**7.45. Forma física y personalidad.** Estar en buena forma física está relacionado con ciertas características de la personalidad. En un estudio sobre esta relación, el profesorado de mediana edad de una universidad que había participado voluntariamente en un programa atlético fue dividido, mediante un reconocimiento médico, en dos grupos. En un grupo, los que estaban en buena forma y en el otro grupo los que estaban en baja forma. Posteriormente, los sujetos pasaron la prueba CSPFQ (*Cattell Sixteen Personality Factor Questionnaire*) para determinar

<sup>30</sup>Véase la nota 26.

su personalidad. He aquí datos sobre la “fuerza de personalidad” de cada uno de estos sujetos.<sup>31</sup>

Grupo	Forma física	$n$	$\bar{x}$	$s$
1	Buena	14	4,64	0,69
2	Mala	14	6,43	0,43

(a) La diferencia entre las medias de “fuerza de personalidad” de los dos grupos, ¿es significativa a un nivel del 5%? ¿Y a un nivel del 1%? Asegúrate de plantear  $H_0$  y  $H_a$ .

(b) ¿Puedes extender estos resultados a la población de todos los hombres de mediana edad? Explica por qué.

**7.46. Investigación de mercado.** Una empresa de investigación de mercados proporciona a unos fabricantes estimaciones sobre las ventas de sus productos al por menor a partir de muestras de tiendas minoristas. Los directores de marketing tienden a fijarse en la estimación y a ignorar el error de ésta. Este mes, una muestra aleatoria simple de 75 tiendas da una media de ventas de 52 unidades de un pequeño electrodoméstico, con una desviación típica de 13 unidades. Durante el mismo mes del año anterior, otra muestra aleatoria simple de 53 tiendas dio unas ventas medias de 49 unidades, con una desviación típica de 11. Un aumento de 49 a 52 unidades es un incremento del 6%. El director de marketing está contento porque las ventas han subido un 6%.

(a) Utiliza el procedimiento  $t$  de dos muestras para calcular un intervalo de confianza del 95% para la diferencia entre el número medio de unidades vendidas este año y el año pasado en todas las tiendas minoristas.

(b) Explica con un lenguaje que el directivo pueda entender por qué no podemos estar seguros de que las ventas subieran un 6%, y que incluso es posible que hayan bajado.

**7.47. ¿Utilizarán más la tarjeta de crédito?** Un banco compara dos propuestas para fomentar la utilización de las tarjetas de crédito entre sus clientes. (El banco gana un porcentaje sobre la cantidad cargada en la tarjeta de crédito, que pagan las tiendas que la aceptan.) La propuesta A ofrece eliminar la cuota anual para los clientes que carguen al menos 2.400 € durante el año. La propuesta B ofrece

<sup>31</sup>A. H. Ismail y R. J. Young, “The effect of chronic exercise on the personality of middle-aged men”, *Journal of Human Ergology*, 2, 1973, págs. 47-57.

devolver en metálico a los clientes un pequeño porcentaje de la cantidad total cargada al final del año. El banco ofrece cada propuesta a una muestra aleatoria simple de 150 clientes con tarjeta de crédito. Al final del año el banco registra la cantidad total cargada por cada cliente en su tarjeta de crédito. He aquí los estadísticos resumen de estas cantidades.

Grupo	$n$	$\bar{x}$ (€)	$s$ (€)
A	150	1.987	392
B	150	2.056	413

(a) ¿Existen diferencias significativas entre las cantidades medias cargadas por los clientes de las dos opciones? Plantea las hipótesis nula y alternativa, y calcula el estadístico  $t$  de dos muestras. Obtén el valor  $P$ . Expón tus conclusiones prácticas.

(b) Las distribuciones de las cantidades cargadas son asimétricas hacia la derecha, pero no hay observaciones atípicas debido a los límites que impone el banco en las cantidades que se pueden cargar en las tarjetas. ¿Crees que la asimetría amenaza la validez de la prueba que utilizaste en (a)? Justifica tu respuesta.

(c) El estudio del banco, ¿es un experimento? ¿Por qué? ¿Cómo afecta esto a las conclusiones que el banco pueda extraer de este estudio?

**7.48. Hallar la salida de un laberinto.** La tabla 7.1 contiene datos sobre el tiempo medio necesario para hallar la salida de un laberinto de 21 sujetos que llevaban mascarillas con esencia o sin ella. En el ejemplo 7.3 se utiliza la prueba  $t$  por pares para mostrar hasta que punto no hay diferencias en los tiempos. Ahora queremos saber si hay un efecto de aprendizaje de manera que los tiempos sean menores en el segundo intento. Todos los sujetos de la tabla 7.1 identificados con un número impar, llevaron a cabo en primer lugar la prueba con la máscara inodora. Los sujetos con números de identificación par, hicieron primero la prueba con la mascarilla que contenía una esencia. La asignación de las etiquetas identificadoras se hizo al azar.

(a) Queremos comparar los tiempos “sin esencia” de los sujetos que en primer lugar hicieron la prueba con las mascarillas sin esencia y los tiempos “sin esencia” de los sujetos que en primer lugar pasaron la prueba con esencia.

(b) Creemos que como media, los sujetos tardan más cuando la prueba sin esencia se lleva a cabo en primer lugar. Dibuja un diagrama de tallos doble con los datos de los tiempos “sin esencia” de los sujetos que hicieron en segundo lugar la prueba sin esencia y de los sujetos que la hicieron en primer lugar. Halla las

medias de los tiempos de los dos grupos. Estos datos, ¿apoyan nuestra sospecha? ¿Existe algo en los datos que impida la utilización de los procedimientos  $t$ ?

(c) Estos datos, ¿aportan evidencia estadísticamente significativa a favor de nuestra sospecha? Plantea las hipótesis y lleva a cabo la prueba. ¿Cuáles son tus conclusiones?

**7.49. Rapidez de aprendizaje.** Los investigadores que estudian el aprendizaje del habla suelen comparar mediciones hechas sobre grabaciones del habla de adultos y de niños. Una variable de interés es el momento del inicio de la voz (MIV). He aquí los resultados de niños de 6 años y de adultos a los que se les pidió que pronunciasen la palabra “bees”. El MIV se mide en milisegundos y puede ser positivo o negativo.<sup>32</sup>

Grupo	$n$	$\bar{x}$	$s$
Niños	10	-3,67	33,89
Adultos	20	-23,17	50,74

(a) Los investigadores querían saber si el MIV diferencia a los adultos de los niños. Plantea  $H_0$  y  $H_a$ , y lleva a cabo una prueba  $t$  de dos muestras. Halla el valor  $P$  ¿Cuáles son tus conclusiones?

(b) Calcula un intervalo de confianza del 95% para la diferencia entre las medias del MIV de niños y adultos cuando se pronuncia la palabra “bees”. Explica por qué sabías a partir de tu resultado en (a) que este intervalo contendría el 0 (ninguna diferencia).

**7.50.** Los investigadores del estudio comentado en el ejercicio 7.49 analizaron los MIV de adultos y niños al pronunciar distintas palabras. Explica por qué no deberían hacer una prueba  $t$  de dos muestras distinta para cada palabra y concluir que aquellas palabras con una diferencia significativa ( $P < 0,05$ ) distinguen a los niños de los adultos. (Los investigadores no cometieron este error.)

**7.51 (Optativo). Potencia de la prueba  $t$  de dos muestras.** Un banco te pide que compares dos proyectos para fomentar el uso de sus tarjetas de crédito. El Proyecto A ofrecería a los clientes una devolución de dinero según la cantidad total cargada en las tarjetas. El Proyecto B reduciría la tasa de interés que se cobra sobre

<sup>32</sup>M. A. Zlatin y R. A. Koenigsknecht, “Development of the voicing contrast: a comparison of voice onset time in stop perception and production”, *Journal of Speech and Hearing Research*, 19, 1976, págs. 93-111.

los saldos negativos de las tarjetas. El banco cree que el proyecto B será mejor. La variable respuesta es la cantidad total que carga un cliente en su tarjeta durante el periodo de prueba. Decides ofrecer los Proyectos A y B a distintas muestras aleatorias simples de clientes con tarjeta de crédito. En el pasado, la cantidad media cargada en las tarjetas durante un periodo de 6 meses fue aproximadamente de 1.100 €, con una desviación típica de 400 €. Una prueba  $t$  de dos muestras basada en muestras aleatorias simples de 350 clientes en cada grupo, ¿detectaría una diferencia de 100 € entre las cantidades medias cargadas en los dos proyectos?

Calcularemos la potencia aproximada de la prueba  $t$  de dos muestras

$$H_0 : \mu_B = \mu_A$$

$$H_a : \mu_B > \mu_A$$

en contra de la alternativa  $\mu_B - \mu_A = 100$ . Utilizaremos el valor 400 como una estimación aproximada de las desviaciones típicas de las poblaciones y de las muestras

(a) ¿Cuál es el valor aproximado del valor crítico del estadístico  $t$  de dos muestras  $t^*$  para  $\alpha = 0,05$ , cuando  $n_1 = n_2 = 350$ ?

(b) **Paso 1.** Escribe la regla para rechazar  $H_0$  en términos de  $\bar{x}_B - \bar{x}_A$ . La prueba rechaza  $H_0$  cuando

$$\frac{\bar{x}_B - \bar{x}_A}{\sqrt{\frac{s_B^2}{n_B} + \frac{s_A^2}{n_A}}} \geq t^*$$

Considera que tanto  $s_A$  como  $s_B$  son iguales a 400, y que  $n_A$  y  $n_B$  son iguales a 350. Halla el número  $c$  tal que la prueba rechace  $H_0$  cuando  $\bar{x}_B - \bar{x}_A \geq c$ .

(c) **Paso 2.** La potencia es la probabilidad de rechazar  $H_0$  cuando la alternativa es cierta. Supón que  $\mu_B - \mu_A = 100$  y que tanto  $\sigma_A$  como  $\sigma_B$  son iguales a 400. La potencia que buscamos es la probabilidad de que  $\bar{x}_B - \bar{x}_A \geq c$  bajo estos supuestos. Calcula dicha potencia.

## 7.4 Inferencia para la dispersión poblacional\*

Las dos características más básicas para describir una distribución son su centro y su dispersión. En una población normal medimos el centro y la dispersión mediante la media y la desviación típica. Utilizamos los procedimientos  $t$  para hacer inferencia sobre la media poblacional de las poblaciones normales y sabemos que

\*Esta sección no es necesaria para comprender los contenidos que aparecen más adelante excepto para el capítulo 10.