Capítulo 13

PRUEBA DE HIPÓTESIS

Sumario

Objetivos específicos

Al finalizar el estudio del capítulo, el estudiante será capaz de:

1. Explicar en qué consiste la prueba de hipótesis y en los principios en los cuales se fundamenta.
2. Indicar los pasos que se siguen al realizar una prueba de hipótesis.
3. Seleccionar la distribución que debe usarse, dependiendo del tamaño de la muestra y del conocimiento existente sobre la variancia poblacional.
4. Señalar los dos tipos de error que pueden darse en una prueba de hipótesis e indicar su importancia práctica.
5. Realizar pruebas de hipótesis para la media poblacional μ y para P, e interpretar estadística y sustantivamente los resultados de la prueba.
6. Elaborar e interpretar pruebas de igualdad de medias y proporciones.

Resumen

En este capítulo se explica el concepto de *prueba de hipótesis estadística* y los principios en los cuales se fundamenta, se comentan los tipos de error y su importancia, se discuten e ilustran las técnicas empleadas para realizar pruebas para μ y P, y para las diferencias de medias y proporciones.

La inferencia estadística, como ya se mencionó, comprende dos problemas básicos: la estimación y la prueba de hipótesis. El primero fue presentado en el capítulo 12, ahora la atención será dirigida a la prueba de hipótesis. Tomando en cuenta el alcance y enfoque de este libro, se introducirá el tema y se concentrará en el caso de la media poblacional y de las proporciones, considerando el caso de hipótesis para μ y para P, así como para hipótesis de igualdad de medias y de proporciones.

13.1. UN PROBLEMA DE DECISIÓN

Considere un problema de decisión simple: se tiene una moneda y se quiere saber si es perfecta o con sesgo a favor de una de las caras, es decir, si al lanzarla tiende a salir con más frecuencia escudo o corona.

Suponga la inexistencia de un sistema físico o matemático que, por simple inspección de las características de la moneda: peso, diámetro, etc., permita descubrir si es perfecta. ¿Cómo puede resolverse este problema?

Sin mucho esfuerzo, se llega a la conclusión de que la única forma de proceder es lanzando la moneda un cierto número de veces y luego, con base en los resultados obtenidos, llegar a una decisión. Es decir, el problema planteado debe resolverse con un cierto monto de evidencia empírica, necesariamente incompleta.

La forma de actuar puede describirse más apropiadamente en la siguiente forma:

a) Se supone que la moneda es perfecta y tiene igual oportunidad de mostrar escudo o corona. Es decir, se establece la hipótesis de que la probabilidad de obtener escudo al lanzar la moneda una vez es igual a $\frac{1}{2}$.

b) Se lanza la moneda un cierto número de veces, por ejemplo 10; luego, de acuerdo con el número de escudos obtenido, se mantiene la hipótesis planteada o se rechaza. Si el número de escudos es muy pequeño o muy alto, se concluirá que la moneda no es perfecta, si es cercano a cinco, se concluirá que sí lo es.

Es decir, se parte de la hipótesis de que la moneda está buena, es perfecta, y de acuerdo con el resultado, se mantiene esa hipótesis o se desecha. En realidad, se tienen dos hipótesis:

Hipótesis nula $H_0 : P = \frac{1}{2}$.

Hipótesis alternativa $H_1 : P \neq \frac{1}{2}$.

La que se somete a prueba es la hipótesis nula H_0, pero al tomarse una decisión acerca de ella, automáticamente se decide lo contrario respecto a H_1.

Antes de profundizar más en este punto de la prueba de hipótesis, es conveniente señalar dos puntos de especial importancia:

La decisión que se tome con respecto a la hipótesis nula, rechazarla o mantenerla, debe basarse en un cierto número de observaciones, 10 en el presente caso.

Se está ante un problema típico de decisión ante la incertidumbre. Se decide con base en información incompleta (los 10 lanzamientos de la moneda) y se debe procurar, entonces, tomarla en la forma más eficiente, es decir, haciendo lo más pequeño posible el margen de error asociado con ella. La regla de decisión y el procedimiento de prueba deben buscar ese objetivo.

Evidentemente, en este ejemplo, la muestra puede aumentarse, ya que es fácil lanzar la moneda un número mayor de veces: 20, 50, 100, 500 o más.

Sin embargo, hay situaciones prácticas en las cuales el número de observaciones o pruebas está limitado (por razones de costo, tiempo, u otro tipo, como sería el caso en que los elementos se destruyen o modifican al observarse) y la decisión, entonces, debe tomarse con solo el número de ellas disponible.

Ejemplo 1

Un oficial recibe un contenedor lleno de granadas y está interesado en evaluar el estado de ese cargamento. Obviamente, no puede probar una gran cantidad de ellas o la totalidad, ya que probarlas implica destruirlas. Su decisión debe basarse en un número reducido de ellas.

Ejemplo 2

En un laboratorio se ha desarrollado una droga A para un padecimiento muy grave, se tiene la hipótesis de que produce mareo y otros efectos colaterales muy negativos en un tercio o más de las personas tratadas, proporción que es considerada inaceptable para efectos de uso y comercialización. Determinar si esta hipótesis es cierta es esencial para decidir si el desarrollo de la droga se continúa o se suspende. El investigador a cargo del estudio dispone de una muestra representativa de 15 voluntarios que tienen el padecimiento. La prueba toma dos semanas y debe hacerse en un centro médico bajos condiciones muy controladas. El investigador debe tomar su decisión acerca de los efectos colaterales de la droga, a partir de los resultados dados por esa muestra de 15 pacientes, ya que le es prácticamente imposible aumentarla y menos esperar a que la use gran cantidad de los pacientes actuales para llegar a una decisión.[1]

13.2. LA REGLA DE DECISIÓN

Si una moneda es perfecta, por instinto se sabe que mostrará escudo en aproximadamente la mitad de las tiradas, pero también es posible que surjan variaciones aleatorias, es decir, alejamientos de esa proporción debido al azar, los cuales serán más marcados cuanto más pequeño sea el número de lanzamientos.

Se espera entonces, si la moneda es perfecta (H_0 es cierta), que en 10 lanzamientos aparezcan cinco escudos. Sin embargo, el que se produzcan desviaciones de uno o dos escudos del valor esperado cinco no debe sorprender, pues se pueden atribuir perfectamente a variaciones aleatorias. Por lo tanto, resultados como 3, 4, 5, 6, 7 harán pensar que la moneda está buena; en cambio, valores bastante alejados de cinco, como 0 y 10 escudos, harán dudar que la moneda sea perfecta.

Parece lógico que la hipótesis H_0 (la moneda es perfecta) deba mantenerse cuando se obtiene un resultado cercano a cinco escudos y rechazarse cuando el valor se aleje mucho de cinco.

El problema que se plantea es fijar por cuáles resultados se rechazará la hipótesis y cuáles permitirán mantenerla. Es decir, determinar qué es "cercano a cinco" y "muy alejado de cinco".

1. Note que se trata de una droga que se encuentra aún en una etapa experimental. Si estuviera en la etapa de prueba definitiva, el tamaño de la muestra y las condiciones para realizarla son fijados por las disposiciones establecidas por la oficina nacional encargada de autorizar el uso de las nuevas drogas.

Se tiene de nuevo el problema fundamental de decidir bajo condiciones de incertidumbre. Con ningún número de tiradas, por supuesto, se estará completamente seguro de que la moneda es perfecta o no. Lo que sí puede hacerse es seguir un procedimiento que mida la probabilidad de cometer un error, ya sea al decidir que la moneda es perfecta, cuando en realidad no lo es, o al asegurar que tiene sesgo cuando es perfecta.

Más adelante se discutirá, con más detalle, cómo puede fijarse ese criterio para decidir cuándo debe rechazarse o mantenerse H_0. Ahora, para efectos de la explicación, se establece, arbitrariamente, pero con cierta lógica, que si salen 0 o 1 escudos, se considerará que la moneda tiene sesgo a favor de corona $P < \frac{1}{2}$; si salen 9 o 10, se valorará que la moneda tiene tendencia a caer escudo $P > \frac{1}{2}$, y si el número de escudos está entre 2 y 8, ambos incluidos, la moneda está "buena" $P = \frac{1}{2}$.

De acuerdo con esto, la regla de decisión será:

Rechazar H_0 si el número de escudos es 0, 1, 9 o 10.

Aceptar H_0 si el número de escudos es 2, 3, 4, 5, 6, 7, 8.

Rechazar H_0 implica que la moneda no es perfecta, que tiene sesgo. Mantener o aceptar H_0 es porque debe considerarse perfecta y no hay base para dudar de ella.

El paso siguiente es calcular las probabilidades de error que involucra esta regla de decisión.

13.3. DEFINICIÓN DE LA ZONA DE RECHAZO Y DE ACEPTACIÓN

La probabilidad de que al lanzar una moneda perfecta 10 veces aparezcan 0, 1, 2, 3, ..., 10 escudos puede obtenerse utilizando las reglas de cálculo de las probabilidades vistas en el capítulo 10, notando que P(Escudo) = P(Corona) $= \frac{1}{2}$.

Por ejemplo, la probabilidad de 0 escudos en los diez lanzamientos es equivalente a P(10 coronas): $P(0) = \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) \cdot \ldots \cdot \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right) \cdot 10 = \frac{1}{1024}$.

En igual forma se obtiene la probabilidad de 1 escudo, $P(1) = 10 \cdot \left(\frac{1}{2}\right)^{10} = \frac{10}{1024}$. El número 10 obedece al hecho de que el escudo puede aparecer en cualquiera de los 10 lanzamientos.[2]

2. Una forma más simple y directa de obtener la distribución de probabilidades es usando la distribución binomial con $n = 10$ y $P = 1/2$. Sus valores pueden calcularse con el procedimiento DISTR. BINOM que aparece en las funciones estadísticas de la hoja de cálculo Excel.

Si se representa con x el número de escudos que se obtiene al tirar 10 veces la moneda y con $f(x)$ la probabilidad de obtener x escudos, la distribución de probabilidad de x es la siguiente:

Cuadro 13.1

PROBABILIDADES SIMPLES Y ACUMULADAS DEL NÚMERO DE ESCUDOS QUE PUEDEN APARECER AL LANZAR 10 VECES UNA MONEDA PERFECTA

Número de escudos x	**Probabilidad** $f(x)$	**Probabilidad acumulada** $F(x)$
0	1/1024 = 0,0010	0,0010
1	10/1024 = 0,0098	0,0108
2	45/1024 = 0,0439	0,0547
3	120/1024 = 0,1172	0,1719
4	210/1024 = 0,2051	0,3770
5	252/1024 = 0,2461	0,6231
6	210/1024 = 0,2051	0,8282
7	120/1024 = 0,1172	0,9454
8	45/1024 = 0,0439	0,9893
9	10/1024 = 0,0098	0,9991
10	1/1024 = 0,0010	1,0001*

$F(x)$ se obtuvo acumulando $f(x)$. El valor final no es exactamente igual a 1,00 debido a los redondeos.

Puede apreciarse que la probabilidad de obtener 0 escudos (o 10) es muy baja, uno en mil, y que la mayor probabilidad corresponde a 5 escudos, con alrededor de 25 por cada 100. Se nota, además, que la distribución es simétrica con centro precisamente en $x = 5$.

La **zona de rechazo o zona crítica**, de acuerdo con la regla de decisión antes establecida, está formada por los valores 0, 1, 9 y 10 de x y la **zona de aceptación,** por los valores de 2 a 8.

De acuerdo con ellas, si al arrojarse la moneda 10 veces se obtienen 9 escudos, la hipótesis nula H_0 será rechazada. Se procede así porque 9 está dentro de la zona de rechazo y, en consecuencia, se concluirá que la moneda no es perfecta, o sea, tiene sesgo.

¿Qué pasaría si se obtienen 7 escudos? Se mantiene H_0 porque el valor $x = 7$ está dentro de la zona de aceptación?

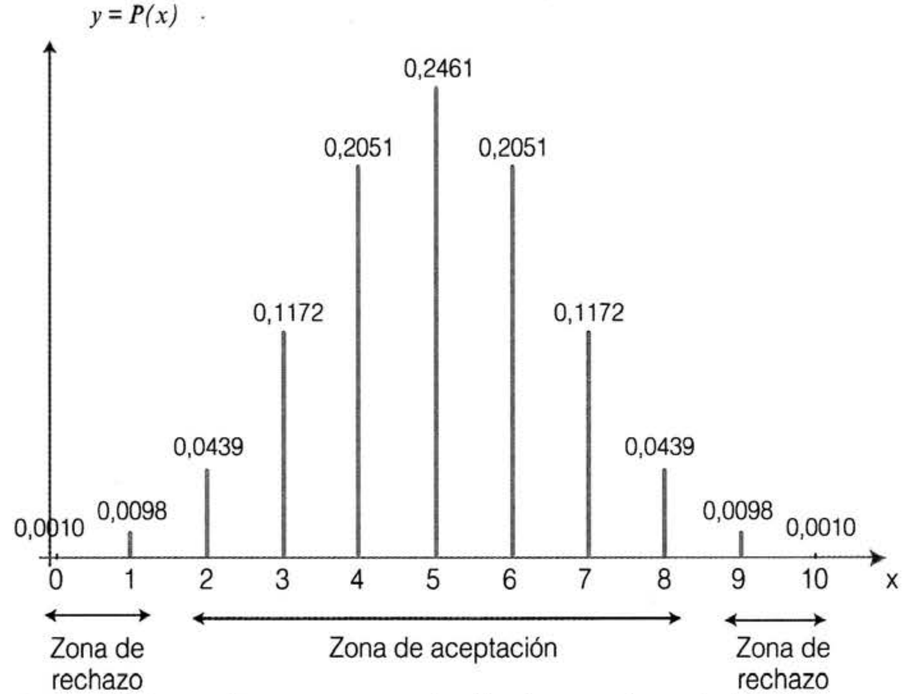


Figura 13.1. Gráfico que muestra la distribución de probabilidad para el número de escudos al tirar 10 veces una moneda "perfecta"

En el presente caso, en el cual la zona de rechazo está ubicada en ambos extremos (colas) de la distribución, se dice que la prueba es de dos colas. En otras situaciones, la naturaleza del problema y los elementos de juicio disponibles hacen que la zona de rechazo se concentre únicamente en una de las dos colas de la distribución, entonces se habla de pruebas de **una cola**.

El que la prueba de hipótesis sea de una cola o de dos depende de la hipótesis alternativa H_1, la cual, a su vez, es fijada tomando en cuenta la información disponible acerca del problema y los intereses y propósitos del investigador.

En el ejemplo de la moneda que se ha venido considerando, la prueba es de dos colas; es así porque la alternativa es $H_1: P \neq \frac{1}{2}$, lo cual significa que si P no es igual a $\frac{1}{2}$ podría ser mayor o menor que ese valor, por ello la zona de rechazo debe ubicarse en las dos colas de la distribución. Si se tienen sospechas bien fundadas de que la moneda tiende a mostrar escudo (o corona), la hipótesis alternativa debe reflejar ese hecho, se indicaría de la siguiente manera: $H_1: P > \frac{1}{2}$ o $H_1: P < \frac{1}{2}$, según corresponda y la prueba sería de una cola.

13.4. TIPOS DE ERROR

En el proceso de someter a prueba y rechazar o mantener una hipótesis estadística, se pueden cometer dos tipos de errores.

Error de tipo I: consiste en rechazar la hipótesis nula H_0 siendo realmente cierta. La magnitud de este error, es decir, la probabilidad de cometerlo, se representa usualmente con la letra α. Simbólicamente:

$$\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ cierta}).$$

Error de tipo II: consiste en aceptar la hipótesis nula H_0 cuando es falsa. La probabilidad de este error se representa con la letra β. Simbólicamente:

$$\beta = P(\text{Aceptar } H_0 \mid H_0 \text{ falsa}).$$

Estas definiciones se pueden presentar más cómodamente con el cuadro 13.2:

Cuadro 13.2

TIPOS DE ERROR QUE SE PUEDEN COMETER UN UNA PRUEBA DE HIPÓTESIS ESTADÍSTICA

DECISIÓN	SITUACIÓN REAL	
	H_0 **cierta**	H_0 **falsa**
Rechazar H_0	**Error de tipo I**	Decisión correcta
Aceptar H_0	Decisión correcta	**Error de tipo II**

La importancia de estos tipos de error y sus implicaciones prácticas varían según las circunstancias. Sobre esto se volverá con mayor detalle. Ahora considere el cálculo de la magnitud del error de tipo I para el ejemplo que se ha venido discutiendo.

En el caso de la moneda, el error de tipo I consiste en decir que no es perfecta (rechazar H_0) cuando realmente lo es (H_0 es cierta). Esto sucederá cuando, por azar, se obtengan 0, 1, 9 o 10 escudos, valores que, de acuerdo con la regla de decisión adoptada, llevan al rechazo de la hipótesis H_0.

La probabilidad α de cometer un error de tipo I, por lo tanto, es la de obtener un valor de x dentro de la zona de rechazo siendo H_0 cierta.

En símbolos:

$$\alpha = P(\text{Rechazar } H_0 \mid H_0 \text{ cierta})$$
$$= P\left(x = 0, 1, 9 \text{ o } 10 / P = \tfrac{1}{2}\right) = f(0) + f(1) + f(9) + f(10).$$

Esta probabilidad puede calcularse con la distribución que aparece en el cuadro 13.1 de la sección 13.3.

$$\alpha = 0,010 + 0,0098 + 0,0010 = 0,0216.$$

El valor $\alpha = 0,0216$ representa la probabilidad del error de tipo I en esta experiencia, con la regla de decisión adoptada. Significa que si se repite lanzar una moneda buena 10 veces, y se aplica la regla de decisión adoptada, la hipótesis H_0 de que la moneda es perfecta se rechazaría un 2,16% de las veces, a pesar de que lo es. Se estarían tomando decisiones equivocadas, con respecto a una moneda perfecta, con una frecuencia de 2,16%.

El complemento $1 - \alpha = 0,9784$ representa la probabilidad de tomar una decisión correcta, aceptando H_0 cuando realmente es cierta.

Si se tomara como zona de rechazo x igual 0 y 10, es decir, los valores extremos, y como zona de aceptación el resto de los valores de x (1 a 9), evidentemente la probabilidad de rechazar H_0, siendo cierta, se reduce y α será más pequeña:

$$\alpha = P\left(x = 0 \text{ o } x = 10 / P = \tfrac{1}{2}\right) = f(0) + f(10)$$
$$= 0,0010 + 0,0010 = 0,002.$$

Con esta regla de decisión, solo se rechazaría H_0 siendo cierta en 0,2% de las veces, o sea, en 2 de cada 1000 repeticiones de la experiencia de lanzar una moneda buena 10 veces. Esto sugiere que una correcta estrategia es reducir al máximo la zona de rechazo. Sin embargo, antes de seguir por este camino, debe tomarse en cuenta el comportamiento del error de tipo II y su relación con el de tipo I.

Cuanto más pequeña se haga la zona de rechazo, menor será la probabilidad de objetar H_0 siendo cierta, y de cometer un error de tipo I; pero, correlativamente, más amplia será la zona de aceptación y mayor la probabilidad de un error de tipo II, o sea, de aceptar H_0 siendo falsa.

El problema está en que, dado un tamaño de muestra n, no se pueden reducir simultáneamente α y β. La reducción en uno de ellos trae un aumento en el otro.[3]

3. La única forma de reducir ambos errores simultáneamente es aumentando el tamaño de la muestra. Si fuera muy grande, la magnitud de los dos tipos de error sería insignificante. En la práctica, sin embargo, en la mayoría de las situaciones, las muestras no pueden ser muy grandes y la probabilidad de cometer alguno de los dos tipos de error es de cierta magnitud.

Lo que se hace en la práctica, entonces, es fijar α a un nivel convencional y luego establecer la regla de decisión que haga mínimo el valor β. Los valores más usados de α son 0,05 y 0,01, pero pueden utilizarse otros niveles como 0,10.

El error de tipo II consiste, como ya se indicó, en aceptar, mantener o rechazar H_0 cuando es falsa, es decir, cuando la que es cierta es H_1. Por definición, el cálculo de la probabilidad de cometer este tipo de error requiere el conocimiento del valor alternativo H_1.

Simbólicamente, β se puede expresar así:

$$\beta = P(x \text{ en zona de aceptación } \mid H_1 \text{ cierta}).$$

En el caso de la moneda que se ha venido comentando si H_0 no fuera cierta y se conociera el verdadero valor de P, sería posible calcular β, evaluando la probabilidad de obtener un valor dentro de la zona de aceptación para ese P correcto.

13.5. IMPORTANCIA DE LOS TIPOS DE ERROR

La importancia de cada uno de estos tipos de error depende, en gran medida, de las circunstancias y consecuencias que pueda tener el cometer uno de ellos. Seguidamente, se incluyen dos ejemplos para ilustrar este punto.

Ejemplo 3

En una universidad, para decidir sobre la admisión de un candidato, se le somete a una prueba de habilidad académica, se le acepta si obtiene 80 puntos o más y se le rechaza si el puntaje es menor.

La hipótesis que se somete a prueba, en el caso de un candidato, puede plantearse de la siguiente forma:

H_0: El candidato X merece ser admitido (es un buen candidato).

La hipótesis alternativa H_1 es: X es un mal candidato (no merece ser admitido).

La regla de decisión será: rechazar H_0 si X obtiene menos de 80 puntos,(nota obtenida <80) y mantener H_0 si saca 80 o más (nota obtenida ≥80).

El error de tipo I consistiría en rechazar un buen candidato y el tipo II en aceptar uno malo. ¿Cuál es el más importante? Como ya se dijo, es cuestión de criterio y de circunstancias. En este caso, los dos tipos de error tienen parecida significación, aunque posiblemente sea más grave el error tipo I, porque implica cerrarle las puertas de la universidad a un candidato y quizás acabar para siempre con sus inquietudes y aspiraciones. En cambio, un error tipo II, que sería aceptar un mal candidato, no parece tan grave, ya que al seguir los estudios debe hacer exámenes y pasar otro tipo de pruebas en las cuales, posiblemente, por ser un mal candidato, quedaría eliminado.

Una situación diferente, en cuanto a la importancia de los dos tipos de error, surge al considerarse el caso de las pruebas finales de graduación. En este caso, parece de mayor importancia el error de tipo II, el cual debería hacerse lo más pequeño posible para evitar que se "filtre" algún mal aspirante a graduado (un futuro mal profesional). Rechazar un buen candidato no es tan grave, ya que normalmente tiene una nueva oportunidad de hacer la prueba y obtener el título, y es casi seguro que la aprovechará. En cambio, si se acepta un mal candidato y se le permite graduarse, las consecuencias pueden ser graves y es muy difícil corregir ese error.

Ejemplo 4

Una enfermedad grave del ganado vacuno se cura, usualmente, con medicamento estándar D_e. Suponga que se desea estudiar el efecto de una nueva droga D_n y compararlo con el de la estándar.

En este caso, las hipótesis nula y alternativa serán:

$H_0 : D_n = D_e$ (droga nueva es igual a la estándar).

$H_1 : D_n > D_e$ (droga nueva es mejor que la estándar).

El error tipo I sería rechazar H_0 siendo cierta, o sea, aceptar que la droga nueva es superior a la estándar cuando realmente son iguales.

El error tipo II consistiría en aceptar H_0 siendo falsa, es decir, no recomendar el uso de la nueva droga cuando en realidad es superior.

Analice los efectos de ambos errores suponiendo diferentes circunstancias:

a) La nueva droga es barata (vale solo un poco más que la estándar): en este caso, aceptar que D_n es superior a D_e, cuando en realidad son iguales (error tipo I), implica una pérdida económica reducida, de manera que no es necesario utilizar un α pequeño.

   El error tipo II, que consiste en no recomendar D_n cuando en realidad es superior, sí es un error substancial: de esto puede depender la vida de animales (vacunos) que, por lo general, tienen un alto valor, y el uso de D_n significar una importante ganancia, β debe ser pequeño.

b) La nueva droga es cara (cuesta mucho más que la estándar): en esta situación, aceptar que $D_n > D_e$, siendo en realidad iguales (error tipo I), implica una fuerte pérdida económica al pagarse mucho más por una droga que no aporta ninguna mejoría, α debe ser pequeño.

No recomendar D_n, cuando en realidad es superior, es también un error importante y parece lógico exigir que β deba ser pequeño. En esta situación, lo apropiado es aumentar el tamaño de la muestra en la cual se basará la decisión.

En general, que α sea más o menos importante que β depende de las circunstancias; en este caso en particular, del balance entre el costo de la droga y el valor de los animales.

13.6. RESUMEN DE CONCEPTOS BÁSICOS EN RELACIÓN CON LA PRUEBA DE HIPÓTESIS

Antes de discutir algunas ilustraciones del uso de los procedimientos de prueba de hipótesis, se hará una recapitulación de los conceptos más importantes en relación con este tema.

Hipótesis estadística: es una suposición acerca de un valor de la población o parámetro, o sobre la relación entre varios valores de la población.

Prueba de hipótesis: es un procedimiento para decidir la aceptación o rechazo de una hipótesis estadística.

Hipótesis nula (H_0): es la hipótesis que se somete a prueba, o sea, aquella sobre la cual se toma la decisión de mantenerla o rechazarla.[4]

Hipótesis alternativa (H_1): especifica la decisión a la que se llegará si la hipótesis nula se rechaza. Se plantea como una oposición o alternativa a la nula.

Regla de decisión: es un criterio para decidir, de acuerdo con los resultados muestrales obtenidos, si la hipótesis H_0 debe rechazarse o mantenerse.

Zona de rechazo o zona crítica: está formada por el conjunto de valores que, de presentarse en la muestra, determinarán el rechazo de la hipótesis nula H_0.

Zona de aceptación: está formada por los valores no incluidos en la zona de rechazo, es decir, aquellos que, de aparecer en la muestra, harán que la hipótesis H_0 se mantenga.

Error de tipo I: consiste en rechazar la hipótesis H_0 siendo correcta. La probabilidad de que ocurra, es decir, su magnitud, se indica con α.

Error de tipo II: consiste en aceptar la hipótesis H_1 siendo falsa. La magnitud de este error, es decir, la probabilidad de que ocurra, se indica por el símbolo β.

Significancia: este término se utiliza frecuentemente cuando se hace referencia a las pruebas de hipótesis, a las cuales también se les conoce como *pruebas de significancia.* En ocasiones se dice que la prueba resultó significativa, lo que indica que la H_0 se rechazó porque se dio una diferencia significativa entre el resultado esperado, bajo el supuesto de H_0 cierta, y el realmente observado en la muestra. En resumen, el término significancia implica el rechazo de la hipótesis nula H_0.

4. El término hipótesis nula se originó en las aplicaciones iniciales de la prueba de hipótesis en la agricultura y la medicina. Cuando se deseaba probar una nueva medicina o fertilizante, se tomaba como referencia el producto usado o existente y se suponía, como hipótesis básica, que el nuevo tenía una efecto nulo.

También se emplea el término *nivel de significancia* para indicar el valor de α tomado como referencia para someter a prueba la hipótesis nula. Como ya se indicó, los niveles de significancia más usuales son 5% y 1%.

Cuando una hipótesis H_0 no es rechazada, no significa que sea verdadera y se acepta como tal, sino que la información disponible no ofrece evidencia suficiente para rechazarla y, por lo tanto, se decide mantenerla y admitirla como si fuera cierta, al menos mientras no la ponga en duda la aparición de nueva evidencia.

13.7. PRUEBA DE HIPÓTESIS PARA LA MEDIA POBLACIONAL μ

En las secciones precedentes, se discutió el problema de la prueba de hipótesis, los principios que la guían y una serie de conceptos relacionados con este tema. Ahora se abordará un problema muy común en la investigación: el de la prueba de hipótesis para la media poblacional. Dado que los principios básicos ya se han visto, se entrará de inmediato a considerar un ejemplo.

Ejemplo 5

Suponga que, hace unos 10 años, la sección de salud de una universidad hizo un estudio de una serie de características de la población universitaria y obtuvo, respecto al peso, los siguientes resultados:

- Peso promedio (μ) 130 libras
- Desviación estándar (σ) 14 libras

Ahora, para ciertos fines, desean conocer si la situación es la misma o ha cambiado. Respecto a la desviación estándar hay acuerdo en que, posiblemente, no ha variado y, por lo tanto, puede considerarse que $\sigma = 14$ es una buena aproximación. En cuanto al peso promedio sí hay diferencias de criterio. En cuanto al peso promedio (μ), para el médico jefe no ha cambiado; para otros, ha aumentado debido a mejoras en la alimentación y los demás funcionarios consideran que ha disminuido porque se incrementó el número de estudiantes jóvenes y de mujeres.

Ante esta situación, el médico jefe decide tomar una muestra para determinar si su opinión es correcta o, por el contrario, el peso promedio en realidad se ha modificado (aumentado o disminuido).

Se trata, entonces, de un típico problema de prueba de hipótesis:

- $H_0 : \mu = 130$
- $H_0 : \mu \neq 130$

Ahora bien, si se toma una muestra al azar de 49 alumnos, se pesan y se obtiene el promedio muestral, ¿cómo se va a proceder ante un determinado valor de $\overline{X}$? En otras palabras, ¿cuándo se va a rechazar H_0 y cuándo se mantendrá?

Evidentemente es necesario fijar una regla de decisión, para esto deben conocerse:

a) El nivel de error tipo I o nivel de significancia (α) con el que se desea trabajar.

b) Si la prueba es de una cola o de dos.

c) La distribución de probabilidades que sigue la variable de interés $\overline{X}$, la cual es una variable aleatoria.

En el presente caso, se sabe que la prueba es de dos colas porque la hipótesis alternativa es $\mu \neq 130$. Además, por el teorema de Límite Central, se sabe que tiene distribución normal con media $\mu = 130$, y $\sigma = \frac{14}{\sqrt{49}} = 2$, bajo el supuesto de que la hipótesis H_0 es cierta. Suponga, además, que se escoge un nivel de significancia $\alpha = 0,05$.

Con estos elementos pueden definirse inmediatamente las zonas de rechazo y de aceptación que requiere la regla de decisión; todo se reduce a escoger los valores de $\overline{x}$ equidistantes de $\mu = 130$, los cuales abarquen, entre ellos, el 95% del área bajo la curva y dejen por fuera un 5%.

Para calcularlos, debe recordarse que en cualquier distribución normal, entre $\mu - 1,96\sigma$ y $\mu + 1,96\sigma$, se tiene el 95% del área bajo la curva y en la distribución de $\overline{X}$ el 95% de los valores estarán entre $\mu - 1,96 \cdot \frac{\sigma}{\sqrt{n}}$ y $\mu + 1,96 \cdot \frac{\sigma}{\sqrt{n}}$, si la hipótesis es cierta. Por lo tanto, los valores que determinan las zonas de rechazo y aceptación son:

$$\mu \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}} = 130 \pm 1,96 \cdot 2 = 130 \pm 3,92 = \begin{cases} 126,08 \\ 133,92 \end{cases}$$

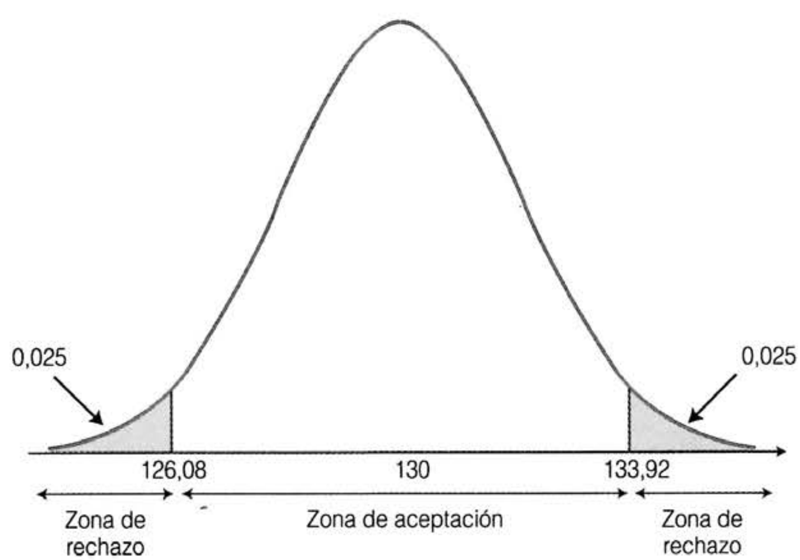


Figura 13.2. Gráfico que muestra las zonas de rechazo y aceptación

La regla de decisión es:

Rechazar $H_0: \mu = 130$ si $\overline{X}$ es mayor que 133,92 o menor que 126,08

$$\overline{X} > 133,92 \text{ o } \overline{X} < 126,08.$$

Aceptar $H_0: \mu = 130$ si $\overline{X}$ es un valor entre 126,08 y 133,92

$$126,8 \leq \overline{X} \leq 133,92.$$

La *zona crítica o de rechazo* la forman los valores de $\overline{X}$ situados a la derecha de 133,92 y a la izquierda de 126,08. Cuando el valor observado de $\overline{X}$ queda dentro de esa zona, la H_0 se rechaza, pues la probabilidad de que, siendo la correcta $(\mu = 130)$, ocurra un valor tan alejado de 130, es menor que el nivel de significancia fijado de 0,05, mínimo que se está dispuesto a aceptar.

◇◇

Ejemplo 6. Uso de la regla de decisión

Suponga dos posibles resultados de $\overline{X}$ en muestras al azar de 49 estudiantes de la población universitaria.

Si $\overline{X} = 128$: valor dentro de la zona de aceptación. Se mantiene H_0.

Conclusión: no hay evidencia para afirmar que el peso promedio de 130 libras haya cambiado.

Si $\overline{X} = 135$: valor dentro de la zona de rechazo. Se rechaza H_0 al 5%.

Conclusión: el resultado da suficiente evidencia para afirmar que el peso promedio ha variado y es mayor ahora.

Siempre para el caso del peso de los estudiantes universitarios suponga, ahora, que se quiere someter a prueba la hipótesis $H_0: \mu = 130$ pero usando un $\alpha = 001$. Como primer paso, debe calcularse la zona crítica correspondiente. A continuación se indica la forma de proceder:

1. Se buscan, en una normal estándar, los valores de z que contengan el 99% central del área bajo la curva:

   $\alpha = 0,01$

   $\frac{\alpha}{2} = 0,005$

   $1 - \frac{\alpha}{2} = 0,995$

   $z_{0,005} = 2,576$

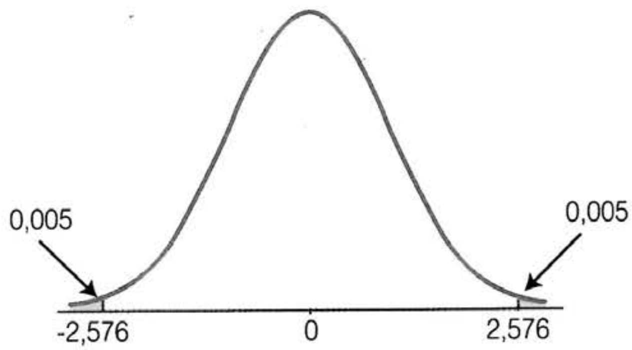


Figura 13.3. Gráfico que muestra la curva normal estándar para el peso de estudiantes universitarios

2. Se calculan los valores que definen la zona crítica:

$$\mu \pm 2,576 \cdot \frac{\sigma}{\sqrt{n}} = 130 \pm 2,576 \cdot 2 = 130 \pm 5,15 = \begin{cases} 135,5 \\ 124,85 \end{cases}$$

3. Se establece la regla de decisión.

   En este nivel de significancia, solo se rechazaría H_0 si $\overline{X}$ es mayor que 135,15 o menor que 124,85. El valor $\overline{X} = 135$, el cual permitió rechazar H_0 al 5%, no admite hacerlo si se trabaja al 1%, al quedar 135 dentro de la zona de aceptación para este nivel de significancia.

◇◇

13.7.1. Procedimiento clásico para la prueba de hipótesis

En la práctica, las pruebas de hipótesis no se realizan definiendo la regla de decisión en términos de $\overline{X}$, sino usando directamente la distribución normal estándar. Para ello se procede en la siguiente forma:

a) Se plantean las hipótesis nula y alternativa, se fija el nivel de significancia α. Se obtiene la z calculada (z_c), definida como:

$$z_c = \frac{\overline{X} + \mu}{\frac{\sigma}{\sqrt{n}}}.$$

b) Con base en el α fijado y tomando en cuenta si la prueba es de una cola o de dos, se define la zona crítica en la curva normal. Es decir, se establece una z tabular (z_t), la cual determina la zona crítica. La z_t se busca directamente en la tabla de la normal estándar; si es de dos colas, se busca el valor para $1 - \frac{\alpha}{2}$, y si es de una, se busca el correspondiente a $1-\alpha$, luego se le da el signo, positivo o negativo, de acuerdo con la dirección de la hipótesis alternativa.[5]

c) Se compara z_c con z_t para ver si z_c está dentro de la zona de rechazo o de la de aceptación. Para ello se usa la siguiente regla:

- Si $|z_c| > |z_t|$ se rechaza H_0
- Si $|z_c| \leq |z_t|$ se mantiene H_0[6]

Ejemplo 7

Para ilustrar el procedimiento suponga, para el ejemplo del peso que se viene considerando, se obtiene $\overline{X} = 135$. Seguidamente, se calcula la z_c y se presentan los valores de z_t para los dos niveles de significancia usuales, los cuales se usan para marcar las zonas críticas:[7]

$$z_c = \frac{135-130}{14/\sqrt{49}} = \frac{5}{2} = 2,5.$$

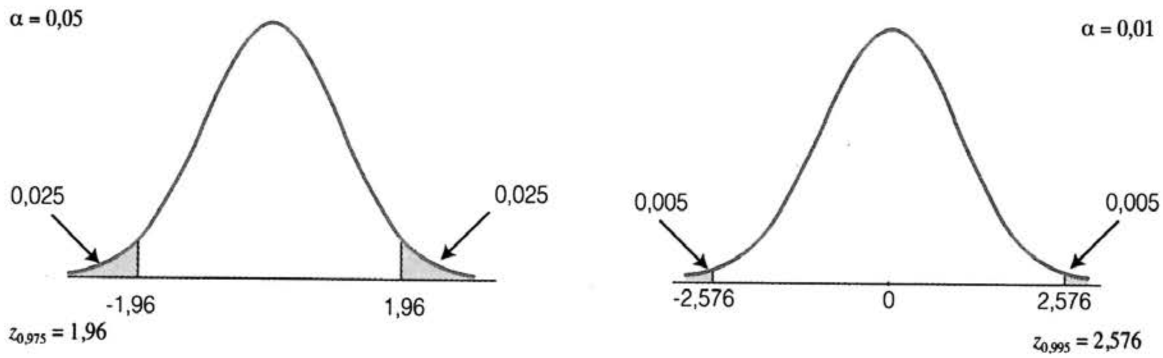


Figura 13.4. Gráfico de las zonas críticas para el peso de estudiantes universitarios

La comparación de los valores de $z_c = 2,5$ con los de z_t correspondientes para los niveles de significancia de 5% y 1% muestran:

$|2,5| > |1,96|$ se rechaza H_0 al 5%

$|2,5| \leq |1,96|$ se mantiene H_0 al 1%

El valor 2,5 está dentro de la zona crítica para 5%, pero al 1% queda dentro de la zona de aceptación.

5. El uso de los valores tabulares para $1-\frac{\alpha}{2}$ o $1-\alpha$ suponen que se tiene una tabla normal estándar N(0,1) acumulada "menos de", que parte de z = 0. Si se tuviera una tabla completa podría buscarse directamente $\frac{\alpha}{2}$ o α en la cola negativa de la tabla. Una forma más simple de obtener la z_t, por supuesto, es usando la función DISTR.NORM.ESTAND.INV de la hoja de cálculo Excel.
6. El símbolo $|\ |$ indica valor absoluto; así, $|-z| = |z| = z$ si z es positivo.
7. En términos generales, el nivel de significancia debe fijarse por anticipado pero cuando no es así, la prueba se realiza a los dos niveles convencionales: 5% y 1%.

13.7.2. Prueba de hipótesis usando el cálculo directo de la probabilidad

Una forma un poco diferente, pero igualmente válida de someter a prueba las hipótesis estadísticas, es calculando directamente la probabilidad de que se produzca un $\overline{X}$ como el observado ($\overline{X}_0$) o mayor, si H_0 es cierta, y tomando la decisión de rechazarla, si esa probabilidad es baja, o, de mantenerla, si sucede lo contrario. Simbólicamente, esa probabilidad se expresa en la siguiente forma:

$$\mathrm{P}(\overline{X} \geq \overline{X}_0 / H_0 \text{ es cierta}).$$

Considere el ejemplo 8.

Ejemplo 8

La experiencia indica que los niños de prekinder (3 a 4 años) de un sistema escolar público pesan en promedio 18 kilos y una $\sigma = 2$. Se tiene la impresión de que, en los últimos años, ese peso promedio ha aumentado. Se toma una muestra al azar de 14 niños y el peso promedio resulta $\overline{X} = 19,4$ kilos.

¿Confirma este resultado la impresión de que el peso promedio ha aumentado?

$$H_0 : \mu = 18$$
$$H_1 : \mu > 18$$

Calcule la probabilidad de que, siendo H_0 cierta, pueda ocurrir un promedio muestral de 19,4 o mayor.

$$z_c = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{19,4 - 18}{\frac{2}{\sqrt{14}}} = \frac{1,4}{0,5345} = 2,619.$$

$$P(\bar{X} \geq 19,4 \,/\, H_0 \text{ es cierta}) = P(z \geq 2,62) = 1 - 0,9956 = 0,0044.$$ [8]

El valor 0,44% indica una probabilidad muy baja de que, siendo cierta la hipótesis nula H_0, se observe un promedio muestral de 19,4 o más. Como el valor es muy pequeño, inferior al 1%, se concluye que la hipótesis H_0 no es cierta y debe rechazarse. Por lo tanto, se acepta que el peso promedio ha aumentado. El nivel del error de tipo I, en esta decisión, es de 0,0044%.

Es obvio, si se utiliza el método clásico, se llega a la misma conclusión, ya que el valor $z_c = 2,62$ es mayor que los valores de z tabular para 0,05 y 0,01; 1,645 y 2,325, respectivamente.

8. $P(z \geq 2,62)$ puede obtenerse usando la función DISTR.NORM.ESTAND de Excel, la cual, como trabaja con la función acumulada menos de la N(0,1), al suministrársele el valor $z = 2,62$ devuelve el valor $1 - P(z \geq 2,62) = 0,9956$, del cual, por diferencia con 1, se obtiene 0,0044.

13.8. APLICACIÓN DE LA PRUEBA DE HIPÓTESIS

Ejemplo 9

Un cierto material viene en cajas cuyo peso promedio es de 100 kg y $\sigma = 1,40$ kg. Se recibe un cargamento grande y el importador sospecha que el peso promedio de las cajas es inferior al usual. Para verificar su sospecha, toma una muestra al azar de 16 cajas y las pesa, obtiene $\overline{X} = 99,50$ kilogramos. ¿Indica este resultado que el peso promedio de las cajas es inferior al acostumbrado?

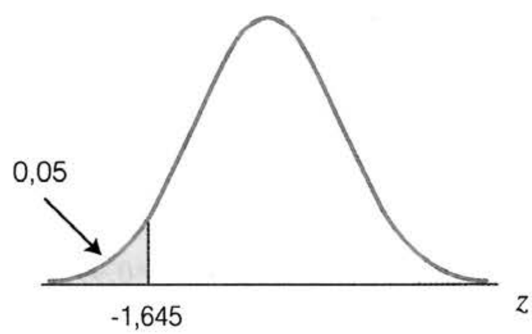
Esta pregunta puede ser respondida mediante una prueba de hipótesis. Se usará un $\alpha = 0,05$.

Las hipótesis nula y alternativa son:

$$H_0 : \mu = 100$$
$$H_1 : \mu < 100$$

Observación: $H_0 : \mu = 100$ significa suponer que el nuevo cargamento tiene el peso acostumbrado, es decir, un promedio de 100 kg. También, la prueba es de una cola, ya que $H_1 : \mu < 100$ porque hay una base, la sospecha del importador, para pensar que si el cargamento no tiene el peso acostumbrado, posee uno menor.

$$z_c = \frac{x - \mu}{\sigma / \sqrt{n}} = \frac{99,5 - 100}{1,40 / \sqrt{16}} = \frac{0,50}{1,40 / 4} = \frac{0,50}{0,35} = 1,43$$



$z_{0,05} = -1,645$

$|-1,43| < |-1,645|$ se mantiene H_0

Figura 13.5. Gráfico que muestra el peso promedio del cargamento del ejemplo 9

Conclusión: no hay evidencia suficiente para afirmar que el cargamento tiene un peso promedio inferior al usual.

Ejemplo 10

Un índice que mide la asimilación al medio urbano ha sido aplicado a un número grande de adultos que llegaron a un área metropolitana procedentes de las zonas costeras. Los resultados indican $\mu = 70$ y $\sigma = 10$. Se toma una muestra al azar de 49 inmigrantes provenientes de las zonas altas del interior y se obtiene $\overline{X} = 67$.

¿Se puede concluir que el grado de asimilación al medio urbano, del inmigrante proveniente de las zonas altas, es diferente al que se observa en quienes provienen de las costeras o puede pensarse que es igual y la diferencia obtenida se debe al azar, es decir, a fluctuaciones del muestreo?

La hipótesis nula es $H_0: \mu = 70$ y la alternativa, $H_1: \mu \neq 70$.

$$z_c = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{67 - 70}{\frac{10}{\sqrt{49}}} = \frac{-3}{\frac{10}{7}} = \frac{-3}{1,43} = -2,10.$$

Como la prueba es de dos colas y se usa un nivel de significancia de 0,05, el valor de z_t que corresponde es 1,96.

$$|-2,10| > |1,96| \text{ se rechaza } H_0 \text{ al } 5\%.$$

Conclusión: a un nivel de $\alpha = 0,05$ no puede considerarse que el resultado sea debido al azar. Se concluye que, en promedio, el grado de asimilación del inmigrante venido de las zonas altas del interior es inferior al del proveniente de las costas.

Ejemplo 11

Suponga que se tiene:

$$H_0: \mu = 60 \qquad H_1: \mu = 56$$

Para someter a prueba H_0, se decide tomar una muestra al azar de tamaño 36 y usar $\alpha = 0,05$. Se sabe que $\sigma = 9$.

a) *Defina la zona de aceptación y la de rechazo en términos de $\overline{X}$.*

b) Calcule el tamaño de β, probabilidad de error tipo II.

c) Dibuje las dos curvas normales y señale la magnitud de β y α.

Solución

a) Se sabe que $z_c = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$. Despejando el valor crítico $x_0 = \mu - z_\alpha \dfrac{\sigma}{\sqrt{n}}$.

En este caso, se usa $z_\alpha = 1,645$ porque la prueba es de una cola y todo el α debe acumularse en la cola de la izquierda.

Valor crítico $x_0 = 60 - 1,645 \cdot \dfrac{9}{\sqrt{36}} = 60 - 2,47 = 57,53$.

Zona de rechazo de $H_0 : \overline{\mathrm{X}} < 57,53$.

Zona de aceptación de $H_1 : \overline{\mathrm{X}} \geq 57,53$.

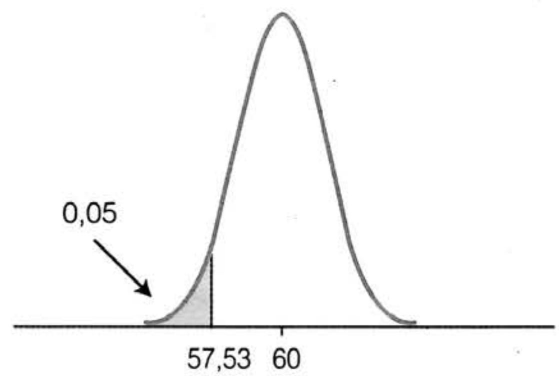


Figura 13.6. Gráfico que muestra la zona de aceptación y de rechazo en términos de $\overline{X}$

b) β es la probabilidad de mantener H_0 siendo falsa, es decir, la probabilidad de que $\overline{X}$ sea mayor o igual a 57,53 cuando realmente $\mu = 56$.

$$\beta = P(\overline{X} \geq 57,53 / \mu = 56).$$

$$z_c = \frac{57,53 - 56}{\frac{9}{\sqrt{36}}} = \frac{1,53}{1,50} = 1,02.$$

$$\beta = P(z \geq 1,02) = 1 - 0,846 = 0,154 = 15,4\,\%.$$

c)

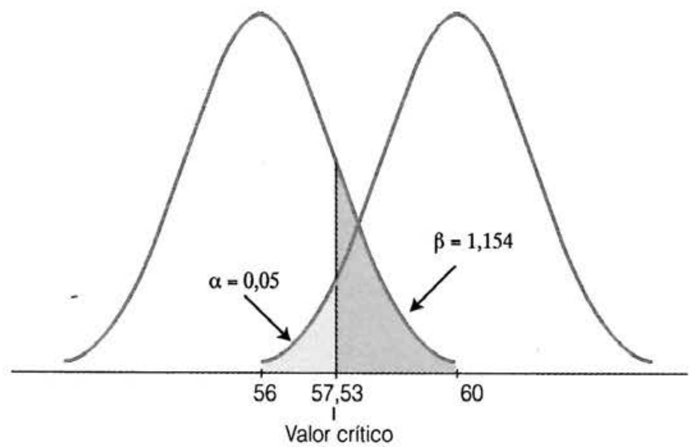


Figura 13.7. Gráfico que muestra dos curvas normales

Ejemplo 12

Se sabe, para una cierta lámpara eléctrica, que su duración media es de 1250 horas. Se introducen ciertas variaciones en el proceso de producción orientadas a mejorar su duración, luego se somete a prueba una muestra al azar de 60 lámparas fabricadas con el nuevo procedimiento, con lo cual se obtiene una duración media de $\overline{X} = 1271$ horas y una variancia muestral $s^2 = 3600$. ¿Existe base para afirmar que el nuevo procedimiento aumenta la duración de los bombillos?

En este caso, la hipótesis nula y alternativa son las siguientes:

$$H_0 : \mu = 1250 \qquad H_1 : \mu = 1250$$

La prueba es de una cola y se usa un $\alpha = 0,05$.

$$z_c = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1271 - 1250}{\frac{60}{\sqrt{50}}} = \frac{21}{8,485} = 2,47.$$

Como el valor de z calculado es mayor que el valor tabular $z_t = 1,645$, se rechaza la hipótesis nula y se concluye que el nuevo procedimiento aumenta la duración de los bombillos.

Es importante señalar que, en este caso, no se conoce σ, la desviación estándar de la población, y se ha estimado con la s desviación estándar de la muestra. Esto puede hacerse sin mucho riesgo porque la muestra es relativamente grande. En términos generales, se considera que una muestra es grande para estos fines cuando es de 30 casos o mayor. Un tamaño como ese garantiza que la distribución de $\overline{X}$ es aproximadamente normal y el valor de s^2 es muy cercano a σ^2 como para que el uso de la desviación estándar muestral no introduzca un error de consideración en los cálculos y procedimientos, por ello se usó la curva normal.

13.9. MUESTRAS PEQUEÑAS: la *t* de *Student*

Los procedimientos de prueba de hipótesis discutidos en la sección precedente, y los cálculos de intervalos de confianza para μ, presentados en el capítulo 12, requieren que la variancia σ^2 de la población sea conocida y que $\overline{X}$ sea normal o aproximadamente normal. De acuerdo con el teorema del Límite Central, esta última condición se logra si la muestra –sin importar la distribución de la población– es bastante grande ($n \geq 30$); también, un tamaño de muestra amplio garantiza que el valor de s es cercano a σ.

Sin embargo, en muchas situaciones prácticas, especialmente de tipo experimental, resulta muy difícil o costoso obtener muestras grandes y es necesario trabajar con valores inferiores o menores a 30 casos.

Ahora bien, ¿qué sucede cuando la muestra es pequeña y no se conoce σ? En estos casos, se emplea s en lugar de σ, pero no se puede garantizar que $\overline{X}$ tenga distribución normal y menos que la expresión $t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$, la cual depende tanto de la variabilidad de $\overline{X}$ como de la de s, tenga distribución normal. Debido a esto, sabiendo que la normal no constituye un modelo apropiado para abordar las inferencias acerca de μ, se recurre a otra distribución denominada t de *Student*, utilizada cuando se cumplen tres condiciones: a) la muestra es pequeña, b) σ no es conocida y c) la población de donde se ha tomado, población madre, tiene distribución norma1 o muy cercana a la normal.[9]

La distribución t de *Student*[10] es una distribución simétrica con respecto a su media, semejante a la normal, pero un poco más extendida, esto hace que el área bajo la curva, con respecto a la normal, sea mayor en las colas y menor en la parte central. Su forma depende del tamaño de la muestra o, más concretamente, de un parámetro denominado grados de libertad asociados a s^2 que es igual a $(n - 1)$. Por ello, no hay una sola distribución t, sino una familia de distribuciones, una para cada número de grados de libertad (tamaño de la muestra). Cuanto mayor sean los grados de libertad asociados a s^2, más se acerca la distribución t a la distribución normal estándar, y cuando la muestra es muy grande, ambas son prácticamente iguales.

El término grados de libertad implica la capacidad de escoger libremente valores. Su sentido puede verse con un ejemplo simple: suponga que una familia tiene cuatro hijos y la edad promedio de ellos es 7 años, ¿cuáles son las edades de los niños?

Se pueden plantear diversas combinaciones de edades, pero con la condición de que el promedio sea 7 o su suma 28, ya que $4 \cdot 7 = 28$.

9. Cuando la población madre no es normal o se desconoce su forma, no es válido utilizar la t de *Student*, entonces se abren dos caminos: a) aumentar el tamaño de la muestra para poder utilizar la curva normal; (o normal b) hacer las inferencias recurriendo a las técnicas denominadas no paramétricas o de distribución libre. Estos procedimientos, sin embargo, están fuera del alcance de este texto y deben ser buscadas en la literatura pertinente.

10. Recibe este nombre porque el estadístico inglés William S. Gosset, quien la descubrió, publicaba sus artículos bajo el seudónimo de Student, debido a que laboraba en una destilería y sus patrones prohibían publicar artículos que pudieran poner en riesgo los secretos industriales de la empresa.

Así, por ejemplo, si se supone que tienen uno de 2 años, uno de 5 y otro de 8, el cuarto debe tener 13 años necesariamente, porque la suma debe ser 28 o su promedio 7. Se pueden escoger libremente las edades de tres, pero la del cuarto queda determinada por la restricción de que el promedio sea 7; en otras palabras, hay 3 grados de libertad para escoger las edades.[11]

Para el uso práctico de la distribución *t* se han construido tablas en las cuales aparecen los valores de *t* para diferentes niveles de significancia y un número amplio de grados de libertad. Su uso es muy similar al de la normal, salvo el hecho de que, para cada caso, debe escogerse la correspondiente según el número de grados de libertad involucrados. A continuación se presenta una tabla donde aparecen los valores de *t* para niveles de probabilidad (α) seleccionados y un cierto número de grados de libertad, suponiendo que la prueba es de una o de dos colas.

11. Los grados de libertad ayudan, en este caso, a seleccionar la distribución *t* de *Student* que corresponde de acuerdo con el tamaño de la muestra. También, en cierta forma indican la confianza que merece s como estimador de σ. En este caso, los grados de libertad son iguales a $n-1$, pero en otras situaciones pueden no ser $n-1$; así, por ejemplo, en el caso de la regresión lineal (capítulo 14), los grados de libertad para pruebas sobre el coeficiente de regresión, y sobre cualquier otra propiedad de la recta de regresión, son $n-2$.

Cuadro 13.3

VALORES DE LA t PARA DIFERENTES NÚMEROS DE GRADOS DE LIBERTAD Y NIVELES DE SIGNIFICANCIA SELECCIONADOS

GRADOS DE LIBERTAD	PROBABILIDAD DE UN VALOR DE t MAYOR QUE EL INDICADO (SIGNO IGNORADO)				
	0,1	0,05	0,02	0,01	DOS COLAS
	0,05	0,025	0,01	0,005	UNA COLA
1	6,314	12,706	31,821	63,656	
2	2,92	4,303	6,965	9,925	
3	2,353	3,182	4,514	5,841	
4	2,132	2,776	3,747	4,604	
5	2,015	2,571	3,365	4,032	
6	1,943	2,447	3,147	3,707	
7	1,895	2,365	2,998	3,499	
8	1,86	2,306	2,896	3,335	
9	1,833	2,262	2,821	3,25	
10	1,812	2,228	2,764	3,169	
11	1,796	2,201	2,718	3,106	
12	1,782	2,179	2,681	3,055	
13	1,771	2,16	2,65	3,012	
14	1,761	2,145	2,624	2,977	
15	1,753	2,131	2,602	2,947	
16	1,746	2,12	2,583	2,921	
17	1,74	2,11	2,567	2,898	
18	1,734	2,101	2,552	2,878	
19	1,729	2,093	2,539	2,861	
20	1,725	2,086	2,528	2,845	
21	1,721	2,08	2,518	2,831	
22	1,717	2,074	2,508	2,819	
23	1,714	2,069	2,5	2,807	
24	1,711	2,064	2,492	2,797	
25	1,708	2,06	2,485	2,787	
26	1,706	2,056	2,479	2,779	
27	1,703	2,052	2,473	2,771	
28	1,701	2,048	2,465	2,763	
29	1,699	2,045	2,462	2,756	
30	1,697	2,042	2,457	2,75	
40	1,684	2,021	2,423	2,704	
50	1,676	2,009	2,403	2,678	
60	1,671	2	2,39	2,66	
∞	1,645	1,96	2,326	2,576	

La interpretación de los valores en la tabla es simple, por ejemplo: 1,812, que corresponde a 10 grados de libertad y a 0,10 de probabilidad, en el caso de las dos colas, significa que, trabajando con 10 grados de libertad, la probabilidad de obtener un valor numérico t (sin considerar el signo) mayor que 1,812 es 0,10. El valor 0,10 incluye, por lo tanto, ambas colas de la distribución. En símbolos:

$$P(-1{,}812 < t < 1{,}812) = 0{,}90.$$

Gráficamente:

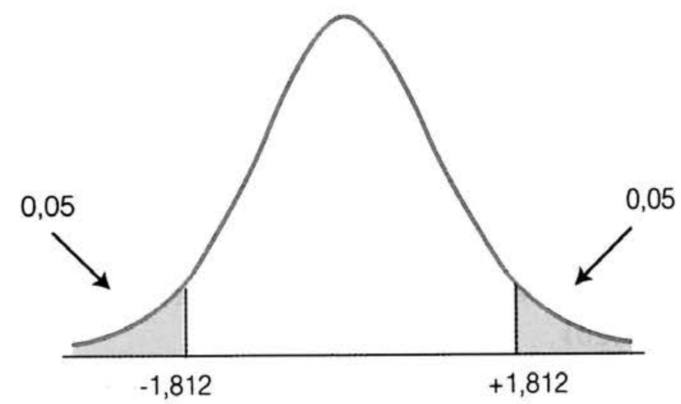


Figura 13.8. Gráfico que muestra la *t* de *Student* para muestras pequeñas

De igual forma, al valor 1,812 para una cola le corresponde una probabilidad de 0,05 y significa la probabilidad de obtener un valor de t calculado mayor que 1,812, y también la de conseguir uno menor que –1,812. Al usar una tabla, es necesario fijarse si es de dos colas o de una, ya que la no consideración de este hecho puede conducir a errores graves.[12]

Finalmente, la teoría estadística muestra cómo la distribución de la variable $t = \dfrac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$ es exactamente la t de *Student*, siempre que la variable básica (x), o sea, la población madre, tenga distribución normal. Esto implica que la t de *Student* deberá usarse solamente cuando se sepa que la variable analizada se basa en una muestra obtenida de una población normal o muy cercana a la normal.

12. Las probabilidades para valores de la t de *Student* pueden obtenerse recurriendo a DISTR.T y DISTR.T.INV que aparecen en las funciones estadísticas de la hoja de cálculo Excel.

Ejemplo 13

Se cree que el contenido de nicotina de los cigarrillos de cierta marca se distribuye normalmente con media 19 miligramos. Se toma una muestra al azar de 12 cigarrillos y se mide el contenido de nicotina, obteniéndose:

$$\overline{X} = 22 \text{ y } s_2 = 16.$$

a) Someta a prueba la hipótesis $\mu = 19$, usando un $\alpha = 0,05$.

b) Calcule un intervalo de confianza del 95% para la media poblacional.

Solución

a) $H_0: \mu = 19 \qquad H_1: \mu \neq 19 \qquad n = 12 \cdot \overline{X} = 22 \qquad s = 4$

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{22 - 19}{\frac{4}{\sqrt{12}}} = \frac{3\sqrt{12}}{4} = 0,75 \cdot 3,464 = 2,598.$$

Consultando la tabla anterior para 11 grados de libertad, $\alpha = 0,05$ y dos colas, se obtiene: $t_{0,05(11)} = 2,201$.

Se tiene $|2,598| > |2,201|$, entonces se rechaza H_0 al 5% y se concluye que el contenido de nicotina es superior a 19 miligramos.[13]

b) Cálculo del intervalo de confianza para μ

$$L_i = \overline{X} \pm t_{\alpha(n-1)} \frac{s}{\sqrt{n}}$$

$$L_i = 22 \pm 2,201 \cdot \frac{4}{\sqrt{12}} = 22 \pm 2,201 \cdot 1,155$$

$$= 22 \pm 2,54 = \begin{cases} 19,46 \\ 24,54 \end{cases}$$

$$P(19,46 \leq \mu \leq 24,54) = 0,95.$$

Se tiene una confianza del 95% de que el contenido promedio de nicotina está entre 19,5 y 24,5 miligramos.

13. Si se suministra a la función DIST.T de la hoja de cálculo Excel, el valor $t_c - 2,598$, se le indica que los grados de libertad son 11 y la prueba es de dos colas, se obtiene una probabilidad de 0,02478, la cual señala que, siendo la hipótesis H_0 cierta, la probabilidad de observar un valor como el indicado o mayor –signo ignorado– es de 2,48%, la cual, por ser tan pequeña, lleva a la conclusión de que la H_0 es falsa y a su consecuente rechazo. Se infiere que el contenido de nicotina es superior a los 19 miligramos.

Por otra parte, si se le suministra a la función DISTR.T.INV, los valores $\alpha = 0,05$ y $gl = 11$, la función arroja el valor 2,201, el cual corresponde al obtenido de la tabla: $t_{0,05(11)} = 2,201$.

13.10. LA DIFERENCIA ENTRE DOS MEDIAS

Hasta el momento, se han tratado únicamente pruebas de hipótesis que se refieren al promedio de una población. Un problema mucho más corriente es el de la comparación de las medias de dos muestras, con el fin de determinar si los promedios de las poblaciones de proveniencia son o no iguales. El procedimiento de prueba tiene algunas diferencias dependiendo de si se conocen o no las variancias de las poblaciones.

13.10.1. Comparación cuando las variancias de las poblaciones son conocidas

Suponga que se tienen dos poblaciones con medias μ_1 y μ_2 y con variancias σ_1^2 y σ_2^2, y se desea determinar si sus medias son iguales o no, es decir, someter a prueba la hipótesis de igualdad de medias. Simbólicamente:

$$H_0: \mu_1 = \mu_2 \quad (\text{o sea, } \mu_1 - \mu_2 = 0)$$

Contra la hipótesis alternativa:

$$H_0: \mu_1 \neq \mu_2 \quad (\text{o sea, } \mu_1 - \mu_2 \neq 0)$$

Para realizar la prueba de hipótesis, se toma una muestra de n_1 elementos de la primera población y una de n_2 elementos de la segunda población, y se calculan los promedios $\overline{X}_1$ y $\overline{X}_2$.

Si las muestras son suficientemente grandes o si las poblaciones de donde fueron tomadas son normales, se sabe que los promedios $\overline{X}_1$ y $\overline{X}_2$ tienen distribución normal y la diferencia de las medias muestrales $d = \overline{X}_1 - \overline{X}_2$ poseen también distribución normal con la media y la variancia siguientes:

$$\mu_d = \mu_1 - \mu_2.$$

$$\sigma_d^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Por lo anterior, la hipótesis H_0 puede ser probada calculando el valor de la expresión z_c, que luego se define, y comparándolo con el valor de z_t para el nivel de significancia establecido, tal como se procedió en el caso de las pruebas de hipótesis para una sola media.

$$z_c = \frac{d - \mu_d}{\sigma_d} = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_d} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Note que la diferencia $\mu_d = \mu_1 - \mu_2$ desaparece, porque, de acuerdo con la hipótesis H_0 planteada, $\mu_1 - \mu_2 = 0$.

Ejemplo 14

Suponga que 64 estudiantes de último año, de colegios privados, y 81 de colegios públicos son medidos y se encuentran estaturas promedio de 170 y 168 cm, respectivamente. Por estudios previos, se conoce que las desviaciones estándar de esas poblaciones son 6 y 9 cm. ¿Puede concluirse que, en promedio, las estaturas de los estudiantes de colegios privados y públicos difieren? Use $\alpha = 0,05$.

Solución

En este caso, las hipótesis nula y alternativa son las siguientes:

$$H_0 : \mu_1 = \mu_2 \text{ y } H_1 : \mu_1 \neq \mu_2,$$

donde el 1 representa a los colegios privados y el 2 los públicos.

Se tiene una prueba de dos colas, como el nivel de significancia es de 0,05, el valor de z tabular es $z_t = 1,96$. Por otra parte,

$$z_c = \frac{170 - 168}{\sqrt{\frac{6^2}{64} + \frac{9^2}{81}}} = \frac{2}{\sqrt{\frac{36}{64} + \frac{81}{81}}} = \frac{2}{\sqrt{0,5625 + 1}} = \frac{2}{\sqrt{1,5625}} = 1,60.$$

Como el z calculado (1,60) es menor que 1,96, la H_0 se mantiene y se concluye que las estaturas promedio de los estudiantes de último año de los colegios privados y públicos no difieren.

Este mismo procedimiento se utiliza también cuando se desea someter a prueba la hipótesis de igualdad de dos promedios y no se conocen las variancias poblacionales, pero las muestras son grandes. Simplemente se usan las variancias muestrales en lugar de las σ^2 poblacionales. Al ser las muestras grandes, se garantiza que las medias muestrales tienen distribución normal y el uso de s^2 en lugar de σ^2, no introduce un error de significación.

13.10.2. Comparación cuando las variancias poblacionales no son conocidas y las muestras son pequeñas

Cuando las muestras son pequeñas, las poblaciones madre son normales o aproximadamente normales, y las variancias poblacionales son desconocidas, pero puede suponerse que son iguales, es decir $\sigma_1^2 = \sigma_2^2 = \sigma^2$. La prueba de hipótesis de igualdad de medias puede ser realizada utilizando la t de *Student*.[14]

14. Cuando las variancias de las poblaciones difieren significativamente, este procedimiento no es válido y debe recurrirse a otras técnicas que están fuera del alcance de este libro.

Lo que se hace es calcular primero s_w^2, o sea, la variancia combinada o promedio de las dos muestras, empleando la siguiente expresión:

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

El objetivo perseguido al combinar las variancias muestrales es tener una mejor estimación de la variancia poblacional σ^2, la cual se supone que es la misma en ambas poblaciones. La variancia combinada es realmente un promedio ponderado de las dos variancias muestrales, y la ponderación se realiza con los grados de libertad. El denominador $n_1 + n_2 - 2$ de la expresión representa el número de grados de libertad asociados a s_w^2.

Luego, con la s_w^2 y la otra información muestral se procede a calcular la expresión:

$$t_c = \frac{\overline{X}_1 - \overline{X}_2}{s_d} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s_w^2}{n_1} + \dfrac{s_w^2}{n_2}}},$$

la cual se compara con un valor t de *Student* para el nivel de significancia a escogido con $n_1 + n_2 - 2$ grados de libertad.

Ejemplo 15

En una investigación, suponga que se probó el efecto de dos abonos A y B sobre el crecimiento de las plantas de café. El estudio se realizó en un invernadero, usando 14 macetas, bajo condiciones experimentales muy homogéneas en cuanto a luz, tipo de suelo, semillas, etcétera.

Se escogieron al azar 7 macetas y se les aplicó el abono A y a las restantes 7, el tratamiento B. Dos plantas abonadas con B se perdieron.

Los resultados del crecimiento experimentado, en centímetros, fueron los siguientes:

A: 5, 6, 5, 7, 4, 8, 7.

B: 6, 6, 9, 8, 10.

Se somete a prueba $H_0: \mu_A = \mu_B$ contra $H_1: \mu_A \neq \mu_B$.

Para realizar la prueba hay que calcular $\overline{X}_A$, $\overline{X}_B$, s_A^2 y s_B^2, al igual que s_w^2. Seguidamente, se indican los cálculos básicos necesarios para obtener esos valores:

		$\sum x_i$	n	$\overline{X}$	$\frac{(\sum x)^2}{n}$	$\sum x_i^2$	x^2
ABONO A	5, 6, 5, 7, 4, 8, 7	42	7	6,0	252	264	2,0
ABONO B	6, 6, 9, 8, 10	39	5	7,8	304,2	317	3,2

$$s_A^2 = \frac{1}{7-1}\cdot(264 - 252) = \frac{12}{6} = 2 \qquad s_B^2 = \frac{1}{5-1}\cdot(317 - 304{,}2) = \frac{12{,}8}{4} = 3{,}2.$$

Se sabe que:

$$s_w^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$

$$s_w^2 = \frac{(7-1)\cdot 2 + (5-1)\cdot 3,2}{7+5-2} = \frac{12+12,8}{10} = \frac{24,8}{10} = 2,48$$

$$s_{(\bar{X}_A - \bar{X}_B)} = \sqrt{\frac{s_w^2}{n_A} + \frac{s_w^2}{n_B}} = \sqrt{\frac{2,48}{7} + \frac{2,48}{5}} = \sqrt{0,8503} = 0,922$$

$$t_c = \frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{s(\overline{X}_A - \overline{X}_B)} = \frac{\overline{X}_A - \overline{X}_B - 0}{s(\overline{X}_A - \overline{X}_B)} = \frac{6,0 - 7,8}{0,922} = \frac{-1,8}{0,922} = -1,952.$$

Los grados de libertad son $n_1 + n_2 - 2 = 7 + 5 - 2 = 10$ y como la prueba es de dos colas, se tiene:

$$t_{0,05}(10) = 2,228 \qquad |-1,953| < |2,228|$$
$$t_{0,01}(10) = 3,169 \qquad |-1,952| < |3,169|$$

Como la t calculada resultó menor que la t_t al 0,05, se concluye que la hipótesis nula (igualdad de efectos de los abonos) no puede rechazarse a ese nivel de significancia y menos al 0,01. No hay razón, por lo tanto, para afirmar que los abonos difieran en su efecto sobre el crecimiento de las plantas.

Esta prueba se puede realizar también usando la función DISTR.T del Excel; si se le suministra el valor 1,952, se le dan los grados de libertad (10) y se le indica que es de dos colas, la función suministra una probabilidad de 0,0795, la cual indica que la probabilidad de que se produzca una diferencia de promedios como la observada, siendo la hipótesis cierta, es de 7,95%. Al ser un valor que supera el 5%, se concluye que el resultado es coherente con la hipótesis de igualdad y esta debe mantenerse. También puede usarse la DISTR.T.INV dándole una probabilidad de 0,05 y 10 grados de libertad, dará el valor $t_c = 2,228$.

13.11. MUESTRAS DEPENDIENTES O PAREADAS

El método presentado en la sección anterior para probar la igualdad entre dos medias supone que las muestras son independientes, es decir, los elementos observados en la primera muestra no tienen ninguna relación –son independientes– con las vistas en la segunda muestra.

En la práctica, sin embargo, en algunas oportunidades se tienen muestras dependientes o "pareadas", ya sea porque se realizaron dos observaciones sobre el mismo elemento o individuo o porque se hizo un pareo de elementos de acuerdo con ciertas características. En estos casos se tiene, realmente, es una muestra de "pares" de observaciones, y lo que interesa es determinar si, en promedio, las diferencias entre esos pares difieren significativamente o no de cero.

Para una mejor comprensión del problema, considere el siguiente ejemplo: una muestra de 25 personas mayores de 30 años que desean ingresar a una universidad son sometidos a una prueba de matemática básica y luego a un curso de refrescamiento de 6 semanas, al final se aplica, de nuevo, otra prueba de matemática básica de nivel muy similar a la primera. Los resultados de las pruebas aparecen en el cuadro 13.4, junto con los cálculos básicos requeridos para hacerla.

El sentido común sugiere que el promedio de la prueba inicial y el de la final se comparen directamente mediante una prueba de promedios como la vista en la sección 13.10. Sin embargo, eso no sería correcto; por tratarse de muestras pareadas, no independientes, el procedimiento debe ser diferente.

Aquí el interés reside en determinar si el curso de refrescamiento mejoró las habilidades matemáticas básicas de los estudiantes, y corresponde analizar las diferencias que se dan entre la primera prueba y la segunda para cada aspirante. Lo apropiado es calcular, para cada estudiante, la diferencia entre la prueba final y la inicial: $d_i = x_D - x_A$, y usar esa diferencia como la variable de análisis para realizar la prueba de hipótesis. Lo que se tiene, entonces, es una sola muestra de 25 diferencias y los procedimientos de prueba de hipótesis para una sola media son totalmente válidos.

Las hipótesis de interés son:

$H_0: \mu_d = 0$ (o sea, $\mu_D - \mu_A = 0$, la diferencia es = 0)

$H_1: \mu_A > 0$ (o sea, $\mu_D > \mu_A$, la diferencia es positiva).

Y como la muestra es pequeña, se utilizará la t de *Student*.

La expresión básica es

$$t_c = \frac{\overline{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{\overline{d}}{\frac{s_d}{\sqrt{n}}}.$$

Note que $\overline{d}$ es el promedio de las diferencias y que s_d corresponde a la desviación estándar de estas. Su cálculo se hace con la información sobre las diferencias, tal como aparece al pie del cuadro 13.4.

$$t_c = \frac{\overline{d}}{\frac{s_d}{\sqrt{n}}} = \frac{3}{\frac{5,2}{\sqrt{25}}} = \frac{3}{1,04} = 2,88.$$

Como el valor $t_c = 2,88$ es superior al valor tabular 2,064 para $\alpha = 0,05$ con 24 grados de libertad, la hipótesis nula se rechaza y se concluye que el curso de refrescamiento fue beneficioso para los estudiantes.[15]

Cuadro 13.4

PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA ENTRE DOS MEDIAS, MUESTRAS DEPENDIENTES O PAREADAS

Estudiante	Nota en prueba inicial (x_A)	Nota en prueba final (x_D)	$d = x_D - x_A$	d^2
1	65	67	+2	4
2	72	70	–2	4
3	64	72	+8	16
4	43	50	+7	49
5	55	54	–1	1
6	84	86	+2	4
7	72	80	+8	64
8	52	50	–2	4
9	49	62	+13	169
10	80	81	+1	1
11	38	56	+18	324
12	93	90	–3	9
13	77	78	+1	1
14	62	64	+2	4
15	69	72	+3	9
16	58	57	–1	1
17	45	55	+10	100
18	90	88	–2	4
19	60	62	+2	4
20	54	52	–2	4
21	72	70	–2	4
22	49	53	+4	16
23	53	56	+3	9
24	82	84	+2	4
25	66	70	+4	16
			75	873

15. El cálculo directo utilizando la DISTR.T de Excel indica $P(t > 2,88) = 0,0041$, lo que señala un valor observado muy poco probable si H_0 es cierta; por ello se rechaza.

$$\overline{d} = \frac{\sum d_i}{n} = \frac{75}{25} = 3$$
$$s_d^2 = \frac{1}{n-1}\left[\sum d_i^2 - \frac{(\sum d_i)^2}{n}\right] = \frac{1}{24}\left[873 - \frac{(75)^2}{25}\right] = 27.$$
$$s_d = \sqrt{27} = 5,20.$$

Las muestras pareadas son muy útiles cuando se hacen comparaciones, porque al realizar las mediciones y comparaciones en los mismos sujetos se mantienen muchos factores constantes, se reduce el error experimental y se facilita la evaluación de los efectos que se quieren analizar. Aparecen con cierta frecuencia en la investigación educativa, cuando se desea determinar los efectos de un método de enseñanza en un mismo grupo de alumnos o comparar pruebas iniciales de diagnóstico con las pruebas finales. También son comunes en el campo médico, cuando se busca evaluar los efectos de tratamientos y medicamentos. Es importante señalar que los investigadores deben estar muy atentos para evitar el análisis de los resultados de muestra dependientes como si fueran independientes, ya que al hacerlo se pierde precisión en el análisis y es menos probable que se detecte el factor detectar evaluar.[16]

13.12. PRUEBA DE HIPÓTESIS PARA PROPORCIONES

Las pruebas de hipótesis para P, salvo algunas excepciones, se realizan utilizando la distribución normal y siguiendo el mismo procedimiento visto para la media poblacional μ; la única condición importante es que la muestra sea lo suficientemente grande para que sea satisfactoria la aproximación normal a la binomial.[17] Considere el siguiente ejemplo.

Ejemplo 16

El presidente de la Federación de Estudiantes de una universidad afirma que 70% de los alumnos está de acuerdo con su propuesta para modificar el sistema actual de elección del Consejo Directivo de la Federación; con el fin de verificar su afirmación, un representante estudiantil selecciona al azar una muestra de 200 alumnos y les pide indicar su opinión sobre la propuesta del presidente, 129 (64,5%) expresan estar de acuerdo con la propuesta.

16. En rigor, si se sigue el método visto en la sección 13.10, no se comete un error sino que se aplica un procedimiento menos eficaz, porque se pierde la ganancia en precisión lograda al hacer las dos mediciones en cada uno de los sujetos. Debido a esto, la prueba puede indicar que el refrescamiento no tuvo efecto cuando en realidad sí lo tuvo.
17. Revisar el capítulo 12, sobre estimación.

¿Cuál sería su conclusión a partir de ese resultado muestral? Utilice un nivel de significancia de 5%.

En este caso,

$$H_0: P = 0,70 \quad H_1: P \neq 0,70$$

$$z = \frac{p-P}{\sigma_p} = \frac{0,645-0,700}{\sqrt{\frac{0,70 \cdot 0,30}{200}}} = \frac{-0,055}{\sqrt{\frac{0,21}{200}}} = \frac{-0,055}{\sqrt{0,00105}} = \frac{-0,055}{0,0324} = -1,70.$$

El valor tabular para $\alpha = 0,05$, con dos colas, es $z_{0,025} = 1,96$.

Como $|-1,7| < |1,96|$, se mantiene H_0.

La hipótesis nula se mantiene. Debe aceptarse la afirmación del presidente de la Federación, no hay base para rechazarla.

◇◇

13.13. DIFERENCIA ENTRE DOS PROPORCIONES

En muchas situaciones prácticas, interesa comparar proporciones (o porcentajes muestrales) para determinar si la diferencia entre ellas es real o debe atribuirse al azar. Así, por ejemplo, al analizar por sexo los resultados de una encuesta de teleaudiencia podría interesar saber si la mayor preferencia observada dentro de la muestra de hombres por los programas deportivos, en relación con las mujeres, debe ser atribuido al azar o si puede considerarse "real" y reflejar una mayor preferencia por esos programas dentro de la población total de hombres.

Igualmente, puede interesar conocer si el hábito del fumado difiere entre los hombres menores de 35 años y los mayores de esa edad o si el abstenerse de votar varía de acuerdo con la residencia urbana o rural. También, en el área industrial sería importante determinar si el porcentaje de piezas defectuosas producidas difiere entre la jornada normal y la jornada nocturna; y en el campo médico, si los infartos son más comunes dentro de los trabajadores de "cuello blanco" o los trabajadores manuales.

Estos problemas de prueba de hipótesis respecto a la diferencia entre dos proporciones se realiza en forma muy similar a como se hace la prueba entre medias poblacionales, con algunos ajustes menores.

Si las muestras son suficientemente grandes, se garantiza que la diferencia entre las proporciones muestrales $p_1 - p_2$ tiene distribución normal y el error estándar de la diferencia viene dado por:

$$\sigma_{p_1-p_2} = \sqrt{\frac{P_1 \cdot Q_1}{n_1} + \frac{P_2 \cdot Q_2}{n_2}}.$$

Sin embargo, como P_1 y P_2 no se conocen –si se conocieran no tendría sentido hacer la prueba de hipótesis– se calcula un valor promedio $\overline{P}$ a partir de los datos de las muestras y se usa en lugar de P_1 y P_2 para calcular el error estándar de la diferencia:

$$\overline{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}.$$

$$\sigma_{p_1-p_2} = \sqrt{\frac{\overline{p} \cdot \overline{q}}{n_1} + \frac{\overline{p} \cdot \overline{q}}{n_2}}.$$

Ejemplo 17

En una encuesta telefónica realizada en agosto del 2009, se preguntó a los entrevistados si les parecía bien o mal que el presidente Óscar Arias hubiera aceptado servir de mediador en la crisis surgida en Honduras con motivo de la destitución del presidente Manuel Zelaya. La muestra fue aleatoria e incluyó 130 hombres y 170 mujeres. Respondieron afirmativamente 91 hombres y 97 mujeres. ¿Podría concluirse, con esta información, que las opiniones positivas respecto a la medición difieren por sexo dentro de la población con teléfono?[18]

En este caso, las hipótesis son: $H_0 : P_1 = P_2$ y $H_1 : P_1 \neq P_2$.

Los valores de p_1 y p_2 son: $p_1 = \frac{91}{130} = 0,70$ y $p_2 = \frac{97}{170} = 0,57$.

En este caso, las hipótesis son: $H_0 : P_1 = P_2$ y $H_1 : P_1 \neq P_2$.

Los valores de p_1 y p_2 son: $p_1 = \frac{40}{80} = 0,50$ y $p_2 = \frac{42}{70} = 0,60$.

$$\overline{p} = \frac{130 \cdot 0,70 + 170 \cdot 0,97}{130 + 170} = \frac{91 + 97}{300} = 0,627 \approx 0,63.$$

$$\begin{aligned}\sigma_{p_1-p_2} &= \sqrt{0,627 \cdot 0,373\left(\frac{1}{130} + \frac{1}{170}\right)} \\ &= \sqrt{0,2331(0,00769 + 0,00588)} \\ &= \sqrt{0,2331(0,01357)} = 0,0562 \\ z_c &= \frac{0,70 - 0,57}{0,0562} = \frac{0,13}{0,0562} = 2,135.\end{aligned}$$

Dado que el valor calculado de z es mayor que el de z tabular para 5% (curva normal): $|\,2,135\,|1,96$, se rechaza la hipótesis H_0 y se concluye que la proporción a la cual le parece bien la mediación del presidente Arias, en el conflicto hondureño, es mayor dentro de los hombres.

18. La encuesta fue de alcance nacional y abarcó a personas de 18 años y más, costarricenses por nacimiento o naturalización, residentes en viviendas particulares que disponen de teléfono de línea fija.

EJERCICIOS DE AUTOEVALUACIÓN

1. Un amigo suyo afirma tener un sistema que le permite acertar los caballos ganadores en las carreras. Él dice que, en un número grande de carreras, con su sistema acierta el caballo ganador en al menos un 50% de ellas. Además, le hace notar que, atinando los ganadores en un 50% de las carreras, se pueden obtener grandes ganancias a mediano plazo. Finalmente, su amigo le dice que, como él no tiene dinero, lo acepta a usted como socio para que dé el dinero y apuesten en un número grande de carreras, dividiendo la utilidad en partes iguales. Como usted no está seguro de que su amigo realmente tenga un "sistema" para ganar el 50% de las carreras, decide hacer una prueba antes de resolver si le entrega o no el dinero para el negocio. Para ello escoge 10 carreras al azar, de las que se realizarán el próximo fin de semana, y le pide a su amigo indicar cuáles serán los caballos ganadores, a fin de que usted pueda verificar en qué proporción de carreras acertó (adaptado de C. Sprowls, *Elementary Statistics*, pp. 1-3).
   a) ¿Cuáles son las hipótesis nula y alternativa en este caso? Escríbalas.
   b) Si su decisión es aceptar H_0, ¿cuál sería su curso de acción respecto al negocio con el amigo? Y si su decisión es rechazar H_0, ¿cuál sería su curso de acción?

DECISIÓN	ACCIÓN
Aceptar H_0	
Rechazar H_0 (Aceptar H_1)	

   c) ¿Qué tipos de error puede usted cometer al tomar una decisión sobre H_0? Indíquelos. Exprese también esos errores en términos del problema considerado.
   d) ¿Cuántas carreras debería acertar su amigo para que usted decida invertir su dinero en el negocio? Es decir, ¿cuál regla de decisión utilizaría usted? Escríbala.
   e) ¿Cuál es el valor de α que corresponde a la regla de decisión adoptada en d)?
   f) ¿Cuál tipo de error considera usted que es el más importante en este problema? ¿Sería adecuado un α alto? Razone su respuesta.

2. Experiencias de laboratorio han mostrado que la longitud promedio de las hembras de cierta especie de chinches criadas en el laboratorio es de 29 mm. Un microbiólogo recoge 15 hembras de chinches silvestres y encuentra una longitud promedio de 29,8 mm, y una variancia de 0,96. El microbiólogo, con base en los resultados obtenidos, concluye que la crianza en laboratorios tiende a reducir el tamaño promedio de los chinches. ¿Considera usted válido ese razonamiento?
   a) Para resolver el problema realice una prueba de hipótesis con un $\alpha = 0{,}01$.
   b) Calcule un intervalo de confianza del 95% para la media de la población.
   c) ¿A qué otro factor, diferente de la cría en el laboratorio, podría atribuirse la mayor longitud de los chinches silvestres?

3. En una prueba de hipótesis, la H_0 es $\mu = 80$ y la alternativa $H_1{:}\mu > 80$. Se sabe que $\sigma = 18$ y la muestra por utilizar es de 36 elementos. Se usará un $\alpha = 0{,}10$.

   Con base en esa información, diga cuál será la zona de rechazo en términos de z y $\overline{X}$.

4. Un instituto técnico está interesado en desarrollar un seminario de preparación para el examen de admisión universitario. De acuerdo con sus cálculos, para que ese proyecto sea rentable y pueda ser establecido, se necesita que más del 20% de los bachilleres que se graduarán este año y el siguiente decidan tomar el seminario. Para llegar a una decisión, se decide realizar una encuesta entre 200 estudiantes de 4° y 5° años de los colegios del país.
   a) En este caso particular, ¿cuál sería la hipótesis nula y cuál la alternativa?
   b) ¿Cuál sería el error de tipo I y cuál sería el de tipo II?
   c) ¿Cuál de estos errores considera usted que es más importante en este caso?
   d) Si se encuentra en la muestra que un 24% de los entrevistados está dispuesto a matricularse en el seminario, ¿cuál sería su decisión acerca de la H_0 y cuál su posición acerca de desarrollar el proyecto? Use $\alpha = 0{,}05$.

5. ¿Cambiaría usted su respuesta en 4d) si supiera que la población que sigue estudios de secundaria está creciendo muy rápidamente? ¿Por qué?

6. Una universidad ubicada en la zona central urbana del país tiene una Prueba de Conocimientos Generales (PCG) que se aplica rutinariamente a los aspirantes que desean ingresar. Los resultados de los últimos 10 años señalan que la PCG tiene $\mu = 200$ y $\sigma = 10$.

   La universidad decide abrir una sede en la zona costera y le aplica la PCG a los 90 aspirantes, obteniendo $\overline{X} = 197$.

¿Puede afirmarse que los resultados confirman la idea de los funcionarios de la universidad, de que los aspirantes de la zona costera tienen, en promedio, un nivel más bajo de conocimientos generales que los de la zona central urbana? Use $\alpha = 0{,}05$.

a) Escriba la hipótesis nula y la alternativa.

b) Realice la prueba e indique cuál es la decisión acerca de la H_0.

c) Escriba la conclusión a la cual se llega. ¿Confirman los resultados la idea de los funcionarios universitarios respecto a los aspirantes de la zona costera?

7. Se sabe, en relación con el tiempo necesario para fabricar cierto artículo por parte de obreros especializados, que $\mu = 16$ min y $\sigma = 4{,}2$ min.

Con el fin de reducir el tiempo que se consume en la operación, se diseña una nueva herramienta y, para probarla, se fabrican 49 unidades y se gastan 710,5 min en la labor.

¿Constituye el resultado observado suficiente evidencia de que la nueva herramienta reduce el tiempo necesario para producir una unidad del artículo? Use $\alpha = 0{,}05$.

a) Escriba las hipótesis nula y alternativa.

b) Someta a prueba la hipótesis y llegue a una decisión acerca de H_0.

c) ¿Qué decisión debe tomar la fábrica: adoptar o no la nueva herramienta?

d) ¿Qué tipo de error puede haber cometido la fábrica al tomar esa decisión? ¿Qué implicaciones prácticas tiene?

8. Al preguntarse a una muestra de 10 mujeres casadas, con hijos menores, aproximadamente cuántas veces al mes visitaba la familia un establecimiento de comida rápida, se obtuvieron las respuestas siguientes:

$$8, 15, 11, 9, 10, 8, 11, 18, 17, 13$$

Pruebe la hipótesis de que $\mu = 15$ contra alternativa $\mu \neq 15$, use $\alpha = 0{,}05$.

9. Un total de 200 alumnos tomó un examen, 133 respondieron correctamente a la pregunta 2 y 148 a la 3. Someta a prueba la hipótesis de que las dos preguntas tienen el mismo nivel de dificultad contra la alternativa de que la 3 es más fácil. Utilice $\alpha = 0{,}10$.

10. Suponga que se tienen dos técnicas de enseñanza del inglés, la tradicional que denominaremos T y una nueva que denominaremos N; se desea determinar si la nueva es más eficaz que la tradicional. Para hacer la comparación, se toman los 50 participantes en un curso de verano de dos meses y se dividen aleatoriamente en dos

grupos: uno de 30 y otro de 20; al primero se le aplica el método tradicional (T) y al segundo el método nuevo (N). La diferencia en el tamaño de los grupos se debe a que el método nuevo requiere equipo especial que solo se tiene para 20 alumnos. Se toman medidas adecuadas para mantener los dos grupos lo más separados posible: distintos alojamientos y lugares de reunión y estudio separados.

De acuerdo con lo previsto, al finalizar el curso de dos meses se aplica una prueba estandarizada de dominio del inglés muy reconocida internacionalmente denominada ELIT, y los resultados arrojan los siguientes valores:

Método T: $\overline{X} = 80$ $s = 10$ $n = 30$

Método N: $\overline{X} = 85$ $s = 11$ $n = 20$

¿Podría afirmarse, con base en estos resultados, que el método nuevo es más eficaz para la enseñanza del inglés? Use $\alpha = 0,01$.

a) Plantee las hipótesis nula y alternativa.

b) Realice la prueba de hipótesis y llegue a una conclusión acerca de la hipótesis nula. Justifique el uso de la distribución que empleó para hacer la prueba.

c) Indique a qué conclusión se llega respecto al nuevo método de enseñanza.

11. En un Programa de Selección de Personal para Empresas Transnacionales, una prueba que consiste en armar un rompecabezas fue aplicada a 21 hombres y 17 mujeres seleccionados al azar dentro de un grupo numeroso de aspirantes. Los tiempos consumidos en la prueba, en minutos, mostraron $\overline{x} = 16,2$ y $s = 2$, para los hombres y $\overline{x} = 14,5$ y $s = 1,8$ para las mujeres.

    a) ¿Puede concluirse que la capacidad para completar el rompecabezas difiere entre hombres y mujeres?

    b) Si hubiera que realizar una tarea industrial similar a la del rompecabezas, ¿cuál sería un intervalo de confianza del 90% para el tiempo que se ahorraría con mujeres en lugar de hombres para realizar la tarea? Calcúlelo e interprételo.

12. Una muestra de distritos de un país fue tomada y en cada uno se indagó la disposición de la población a clasificar la basura y separarla en bolsas de diferentes colores. Luego se realizó una campaña educativa sobre el tema y dos meses después se repitió el estudio. Conviene señalar que, en algunos distritos, durante ese período, el servicio de recolección de basura sufrió un deterioro importante. La proporción que se mostró de acuerdo con clasificar la basura y separarla, en ambas oportunidades, aparece seguidamente:

	Estudio A	Estudio B		Estudio A	Estudio B
1	68	66	10	73	79
2	63	60	11	47	40
3	65	71	12	68	75
4	73	80	13	59	55
5	74	77	14	74	81
6	47	58	15	72	84
7	65	72	16	72	79
8	72	72	17	68	66
9	55	49	18	80	88

¿Puede concluirse de esta información que la disposición a clasificar la basura subió en el período? Haga la prueba con $\alpha = 0{,}05$ y $\alpha = 0{,}01$.

13. Una empresa que se dedica a realizar encuestas telefónicas está interesada en determinar si dos entrevistadoras A y B difieren en el tiempo que tardan en las entrevistas. Para investigar el punto, se toman las entrevistas efectuadas en el último estudio y se calcula, para cada entrevistadora, el promedio y la variancia de la duración. Es importante señalar que las personas que se consultan son seleccionadas al azar y todas las entrevistadoras se desempeñan bajo el mismo horario y condiciones:

Entrevistadora A	Entrevistadora B
$n = 30$	$n = 28$
$\overline{x} = 15{,}2$ minutos	$\overline{x} = 16{,}4$ minutos
$s^2 = 13$	$s^2 = 20$

¿Los datos dan base para concluir que la rapidez para realizar las entrevistas difiere entre las entrevistadoras?

14. En una encuesta realizada en el Área Metropolitana de San José, en la que se estudiaron los hábitos de audiencia de la televisión de las mujeres de 13 años y más, se dividió la población en dos estratos socioeconómicos: el Popular, formado por familias de ingresos bajos y el Medio-Alto que abarcó la clase media y los grupos de ingresos altos. Una de las preguntas indagó el hábito de ver telenovelas.

Dentro de la muestra, quedaron un total de 181 mujeres estudiantes, 80 pertenecientes al Popular y 101 pertenecientes al Medio-Alto; de ellas, 43 en el Popular y 41 en el Medio-Alto, indicaron que tenían la costumbre de ver telenovelas.

¿Los resultados permiten concluir que es correcta la creencia de que el hábito de ver telenovelas es mayor dentro de las jóvenes de sectores populares? Use un $\alpha = 0{,}02$.

Plantee las hipótesis e incluya los cálculos y razonamientos que lo llevan a la conclusión.

15. Usando la información dada en el problema anterior y sabiendo que el total de mujeres entrevistadas fue de 552 en el estrato Popular y de 537 en el Medio-Alto, ¿podría concluirse que la asistencia a la educación formal difiere entre los estratos socioeconómicos?

16. Sabiendo que, en general, las mujeres que estudian son menores de 35 años, y para el problema 15, que el total de mujeres estudiadas menores de 35 años fue de 320 en el estrato Popular y de 264 en el estrato Medio-Alto:

    a) ¿Modificaría usted las conclusiones a las cuáles llegó en el problema 15? ¿Por qué?

    b) ¿A qué atribuye usted la diferencia en los resultados obtenidos en este problema y en el problema 15?

17. En un periódico de circulación nacional se presentaron los resultados de una encuesta, elaborada por una conocida consultora que realiza estudios de opinión. Al comentar los resultados en el texto se dice que "los electores jóvenes (18 a 24 años) difieren de la muestra total en la mayor importancia que asignan a la Deuda Interna". Los resultados de la encuesta presentados para apoyar esta conclusión son:

    a) Entre los jóvenes, un 11,9% considera la Deuda Interna el problema más importante de Costa Rica, mientras que, dentro de la muestra total, el porcentaje es de 9,9%.

    b) Entre los jóvenes un 10,8% considera que solucionar la Deuda Interna es la tarea más importante del próximo gobierno mientras que, dentro de la muestra total, esa proporción es de 7,6%.

    En su opinión, estos resultados de la encuesta ¿permiten aceptar la conclusión difundida por el periódico? El número de jóvenes de 18-24 años entrevistados es de 232 y la muestra total es de 1205 personas.

18. Por experiencia se sabe que los días de estancia de quienes se someten a un cierto tipo de procedimiento quirúrgico en un gran hospital público tienen una media de 9,7 días y una desviación estándar de 3,6 días. Un nuevo director hace modificaciones para mejorar los procedimientos operatorios y el manejo postoperatorio de los pacientes. Al revisar las cifras del primer mes, después de la introducción de los cambios, el director encuentra que se han realizado 110 operaciones y el número de días de estancia promedio es de 9,1 días.

    ¿A partir de esta información puede concluirse que los cambios redujeron los días de estancia hospitalaria para el procedimiento quirúrgico en cuestión?

RESPUESTA A LOS EJERCICIOS DE AUTOEVALUACIÓN

1. a) $H_0{:}P = 0{,}50$ $\quad$ $H_1{:}P < 0{,}50$

b)

DECISIÓN	ACCIÓN
Aceptar H_0	Darle el dinero, entrar en el negocio.
Rechazar H_0	No darle el dinero, no entrar en el negocio.

c) Dos tipos de error:

- Rechazar H_0 siendo cierta, o sea, concluir que el amigo no tiene la capacidad de "adivinar" los resultados de las carreras en al menos el 50% de los casos, cuando en realidad sí la tiene.
- Mantener H_0 cuando en realidad es falsa, es decir, cuando el amigo no tiene la capacidad de adivinar los resultados de las carreras.

d) Obviamente se mantiene la hipótesis nula si el número de aciertos es razonablemente elevado y se rechaza si es bajo; por ello, usando el juicio, se fija como regla de decisión mantener H_0 cuando acierte 3 o más y rechazarla si acierta 2 o menos.

e) El valor de α que corresponde a la regla de decisión utilizada puede calcularse con la distribución de probabilidades presentada en el cuadro 13.1.

$$\begin{aligned}\alpha &= P(x = 0 \text{ o } x = 1 \text{ o } x = 2 \,/\, H_0 \text{ cierta, } P = 0{,}50)\\ &= f(0) + f(1) + f(2)\\ &= 0{,}0010 + 0{,}0098 + 0{,}0439 = 0{,}0547.\end{aligned}$$

f) Si usted está pensado en invertir en un proyecto como el comentado, se supone que es una persona a quien le interesa no dejar pasar una oportunidad de hacer un buen negocio. Como α mide la probabilidad de rechazar una hipótesis siendo cierta y en este caso significa concluir que el amigo no tiene la capacidad de adivinar los resultados de las carreras, cuando en realidad sí la tiene, α debe ser bajo porque al ser alto habría una alta probabilidad de rechazar la H_0 y de perder, en consecuencia, una buena oportunidad de inversión.

2. $\mu = 29$ mm $\quad n = 15 \quad \bar{x} = 29,8$ mm $\quad s^2 = 0,96$

   a) $H_0: \mu = 29 \quad H_1: \mu > 29 \quad t_t = 2,624$

   $$t = \frac{29,8 - 29}{\sqrt{\frac{0,96}{15}}} = \frac{0,8}{0,253} = 3,16 \quad 3,16 > 2,624 \Rightarrow \text{Se rechaza } H_0$$

   Se concluye que el tamaño promedio de los chinches silvestres es mayor que el de los criados en el laboratorio.

   b) $L = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}} = 29,8 \pm 2,624 \cdot \frac{0,98}{\sqrt{15}} = 29,8 \pm 0,66$

   $P(29,14 < \mu < 30,46) = 0,95$.

   c) Una posibilidad es que la mayor longitud de los chinches silvestres se deba a la selección natural que les permiten sobrevivir, principalmente, a los más fuertes y grandes.

3. $H_0: \mu = 80 \quad H_1: \mu > 80 \quad \sigma = 18 \quad n = 36 \quad \alpha = 0,10$

   Como la prueba es de una cola y $\alpha = 0,10$, la hipótesis nula se rechazará cuando $z > 1,28$.

   La zona de rechazo en términos de $\bar{x}$ es $\bar{x} > 83,84$.

   $$L = \mu \pm z \cdot \frac{\sigma}{\sqrt{n}} = 80 + 1,28 \cdot \frac{18}{\sqrt{36}} = 80 + 3,84 = 83,84.$$

4. a) $H_0: P = 0,20 \quad H_1 > 0,20$

   b) El error de tipo I, rechazar H_0 siendo cierta, implicaría concluir que el porcentaje que tomará el seminario es mayor de 20% cuando en realidad es 20% o menor.

   El error tipo II, aceptar H_0 siendo falsa, implicaría concluir que el porcentaje es de 20% o menor cuando en realidad es mayor.

   c) En este caso, el error de tipo I parece ser el más grave porque implicaría el desarrollo del proyecto cuando en realidad la proporción dispuesta a tomar el seminario es de 20% o menor, lo que causaría una pérdida económica. El error tipo II implica una pérdida de oportunidad, es decir, no hacer la ganancia que daría el proyecto si es viable y se lleva a cabo.

   d) $P = 0,24 \quad z = 1,65$

   $$z = \frac{0,24 - 0,20}{\sqrt{\frac{0,20 \cdot 0,80}{200}}} = \frac{0,04}{0,0283} = 1,41 < 1,65.$$

   Se mantiene H_0, no se desarrollaría el proyecto.

5. Posiblemente, porque aunque el porcentaje no sea superior a 20%, el rápido crecimiento de la población de secundaria garantiza que en unos pocos años la matrícula sea mucho más alta y el proyecto alcanzaría el punto de equilibrio y sería viable.

6. $\mu = 200$ $\quad\sigma = 10$ $\quad\overline{x = 197}$ $\quad n = 90$

   a) $H_0 : \mu = 200$ $\quad H_1 : \mu > 200$ $\quad\alpha = 0,05$ $\quad z_t = 1,645.$

   b) $z = \dfrac{197 - 200}{\dfrac{10}{\sqrt{90}}} = \dfrac{-3}{1,054} = -2,85.$

   Como $|-1,85|1,645$, la H_0 se rechaza.

   c) Los resultados confirman la opinión de que los estudiantes de la zona costera, en promedio, tienen un nivel más bajo de conocimientos generales.

7. $\mu = 16$ minutos $\quad\sigma = 4,2$ $\quad\sum x = 710,5$ $\quad n = 49$ $\quad\bar{x} = \dfrac{710,5}{49} = 14,5$

   a) $H_0 : \mu = 16$ $\quad H_1 : \mu < 16.$

   b) $z = \dfrac{14,5 - 16}{\dfrac{4,2}{\sqrt{49}}} = \dfrac{-1,5}{0,6} = -2,50.$ Como $|-2,50|1,645$, la H_0 se rechaza.

   c) Si el criterio básico para tomar la decisión es que la herramienta rebaje el tiempo promedio para realizar la tarea, la empresa debe adoptar la nueva herramienta.

   d) Al rechazar H_0 y tomar la decisión de adoptar la nueva herramienta, la fábrica puede haber cometido un error de tipo I: rechazar una hipótesis siendo cierta. En este caso, la nueva herramienta no reduce el tiempo y la fábrica sufre una pérdida derivada del costo de cambiar las viejas herramientas y adiestrar el personal para el uso de la nueva.

8. $n = 10$ $\quad H_0 : \mu = 15$ $\quad H_1 : \mu \neq 15$

   $\sum x = 120$ $\qquad \overline{x} = \dfrac{\sum x}{n} = \dfrac{120}{10} = 12$

   $\sum (x - \overline{x})^2 = 118$ $\qquad s^2 = \dfrac{118}{9} = 13,11$ $\qquad s = 3,62$ $\qquad t_t = 2,26$

   $t = \dfrac{\overline{x} - \mu}{\dfrac{s}{\sqrt{n}}} = \dfrac{-3}{1,145} = -2,62.$

   $|-2,62| > 2,26$ se rechaza la hipótesis de que $\mu = 15$, y se concluye que μ es menor que 15.

9. $H_0 : P_2 = P_3 \qquad H_1 : P_3 > P_2 \qquad \alpha = 0,10 \qquad z_t = 1,282$

$$p_2 = \frac{133}{200} = 0,665 \qquad p_3 = \frac{148}{200} = 0,74$$

$$p = \frac{0,665 + 0,74}{2} = 0,70$$

$$z = \frac{0,74 - 0,665}{\sqrt{0,70 \cdot 0,30\left(\frac{1}{200} + \frac{1}{200}\right)}} = \frac{0,075}{\sqrt{0,21 \cdot 0,01}} = \frac{0,075}{0,046} = 1,63.$$

Como el valor calculado 1,63 es mayor que 1,282, la hipótesis de igualdad de proporciones se rechaza y, por lo tanto, se concluye que la pregunta 3 es más fácil que la pregunta 2.

10. $H_0 : \mu_N = \mu_T \qquad H_1 : \mu_N > \mu_T \qquad \alpha = 0,10 \qquad t_t = 2,41$

$$\bar{s}^2 = \frac{(n_T - 1)s_T^2 + (n_N - 1)s_N^2}{n_T + n_N - 2} = \frac{29 \cdot 10^2 + 19 \cdot 11^2}{30 + 20 - 2} = \frac{2900 + 2299}{48} = 108,3125.$$

$$t = \frac{85 - 8}{\sqrt{108,31 \cdot \left(\frac{1}{30} + \frac{1}{30}\right)}} = \frac{5}{\sqrt{108,31 \cdot 0,0833}} = \frac{5}{\sqrt{9,02}} = \frac{5}{3} = 1,67.$$

La hipótesis de igualdad de medias se mantiene porque $1,67 < 2,41$.

Se emplea la t de *Student*, porque no se conocen las variancias, pero es razonable pensar que las variables tienen una distribución cercana a la normal.

El valor $t_t = 2,41$ para 48 grados de libertad y un $\alpha = 0,01$, una cola se obtuvo de la tabla por interpolación.

c) El método nuevo no es más eficaz que el tradicional.[19]

11.

Hombres:	$n = 21$	$\bar{x}_H = 16,2$	$s = 2,0$
Mujeres:	$n = 17$	$\bar{x}_M = 14,5$	$s = 1,8$

a) $H_0 : \mu_H = \mu_M \qquad H_1 : \mu_H \neq \mu_M \qquad t_t = 2,029$

$$\bar{s}^2 = \frac{20 \cdot 4 + 16 \cdot 3,24}{21 + 17 - 2} = \frac{80 + 51,84}{36} = \frac{131,84}{36} = 3,66$$

$$t = \frac{16,2 - 14,5}{\sqrt{3,66 \cdot \left(\frac{1}{21} + \frac{1}{17}\right)}} = \frac{1,7}{\sqrt{3,66 \cdot 0,1064}} = \frac{1,7}{0,624} = 2,72.$$

$2,72 > 2,03$ H_0 se rechaza; la capacidad difiere, es mayor dentro de las mujeres.

19. Si se utiliza la distribución t de *Student* de la hoja de cálculo Excel, se obtiene $P(t > 1,67) = 0,0507$. Este valor es muy superior al $\alpha = 0,01$ establecido como criterio para evaluar la hipótesis nula, por lo que es razonable mantener dicha hipótesis.

b) Aquí lo que interesa es un intervalo de confianza del 90% para la diferencia en el tiempo que, en promedio, les toma a los estudiantes de diferente sexo realizar la tarea, o sea $\mu_H - \mu_M$. El cálculo se realiza con la expresión que seguidamente se indica, y usando el error estándar de la diferencia estimado en el punto anterior: $s_d = 0,624$ y un valor tabular de $t = 1,69$, que corresponde a 36 grados de libertad y un $\alpha = 0,10$, obtenido por interpolación de la tabla t de *Student*.

$L_i = (\mu_H - \mu_M) \pm t\bar{s}_d = 1,7 \pm 1,69 \cdot 0,624 = 1,7 \pm 1,05.$

$P(0,65 < \mu_H - \mu_M < 2,75) = 0,90.$

Con una confianza de 90% puede afirmarse que en una tarea industrial similar, con mujeres en lugar de hombres, se ahorraría entre 0,65 y 2,75 minutos por unidad producida.

12. En este caso, se trata de muestras pareadas, es decir, no independientes, ya que las observaciones se refieren al mismo distrito en dos momentos diferentes.

DIST	Est 1	Est 2	DIF	DIST	Est 1	Est 2	DIF
1	68	66	–2	10	73	79	6
2	63	60	–3	11	47	40	–7
3	65	71	6	12	68	75	7
4	73	80	7	13	59	55	–4
5	74	77	3	14	74	81	7
6	47	58	11	15	72	84	12
7	65	72	7	16	72	79	7
8	72	72	0	17	68	66	–2
9	55	49	–6	18	80	88	8

$H_0: \mu_d = 0 \qquad H_1: \mu_d > 0 \qquad t_t = 1,74\,(al\,5\%) \qquad t_t = 2,57\,(al\,1\%)$

$$d = \frac{\sum d}{n} = \frac{57}{18} = 3,17.$$

$$s_d^2 = \frac{1}{n-1}\left(\sum d_i^2 - \frac{(\sum d_i)^2}{n}\right) = \frac{1}{17}\left(773 - \frac{(57)^2}{18}\right) = \frac{1}{17} \cdot 592,5 = 34,85.$$

$s_d = 5,90.$

$$t = \frac{\bar{d} = \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{3,17}{\frac{5,90}{\sqrt{18}}} = \frac{3,17}{1,39} = 2,28.$$

Si se usa $\alpha = 0,05$, puede concluirse que la disposición a clasificar la basura subió en el período, ya que $2,28 > 1,74$; pero si se es más estricto y se emplea $\alpha = 0,01$, no se puede llegar a esa conclusión, pues $2,28 < 2,57$.

13.

Entrevistadora A: $n = 30$ $\bar{x} = 15,2$ $s^2 = 13$

Entrevistadora B: $n = 28$ $\bar{x} = 16,4$ $s^2 = 20$

$H_0: \mu_A = \mu_B$ $H_1: \mu_A \neq \mu_B$ $t_t = 2,00$

$$\bar{s}^2 = \frac{29 \cdot 13 + 27 \cdot 20}{30 + 28 - 2} = \frac{377 + 540}{56} = \frac{917}{56} = 16,38$$

$$t = \frac{15,2 - 16,4}{\sqrt{16,38 \cdot \left(\frac{1}{30} + \frac{1}{29}\right)}} = \frac{-1,2}{\sqrt{16,38 \cdot 0,06905}} = -1,13.$$

$|-1,13| < 2,00$ se mantiene H_0, las entrevistadoras no difieren en la rapidez con que realizan las entrevistas.

14.

Estrato social	Ven telenovelas	No ven	Total	% Ven
1. Popular	43	37	80	53,8
2. Medio-Alto	41	60	101	40,6
Total	84	97	181	46,4

$H_0: P_1 = P_2$ $H_1: P_1 > P_2$ $\alpha = 0,02$ $z_t = 2,05$

$$p = \frac{43 + 41}{181} = 0,46.$$

$$z = \frac{0,538 - 0,406}{\sqrt{0,46 \cdot 0,54 \cdot \left(\frac{1}{80} + \frac{1}{101}\right)}} = \frac{0,132}{\sqrt{0,2484 \cdot 0,0224}} = 1,77.$$

$1,75 < 2,05$ no hay base para concluir que el hábito de ver telenovelas es mayor dentro de las jóvenes del estrato Popular.

15.

Estrato social	Estudiantes	Total	% Estud.
1. Popular	80	552	14,5
2. Medio-Alto	101	537	18,8
Total	181	1 089	16,6

$H_0: P_1 = P_2$ $H_1: P_1 \neq P_2$ $\alpha = 0,05$ $z_t = 1,96$

$$p = \frac{80 + 101}{1089} = 0,116.$$

$$z = \frac{0,188 - 0,145}{\sqrt{0,166 \cdot 0,834 \cdot \left(\frac{1}{552} + \frac{1}{537}\right)}} = \frac{0,043}{\sqrt{0,1384 \cdot 0,02255}} = 1,91.$$

$1,91 < 1,96$ La hipótesis de igualdad de proporciones se mantiene; no puede concluirse que la asistencia a la educación formal difiere entre los estratos sociales. La diferencia observada debe atribuirse al azar.

16.

Estrato social	Estudiantes	Mujeres Menos 35 años	% Estud.
1. Popular	80	320	25,0
2. Medio-Alto	101	264	38,3
Total	181	584	31,0

$H_0: P_P = P_{MA}$ $\quad\quad H_1: P_P \neq P_{MA}$

a) $$z = \frac{p-P}{\sigma_p} = \frac{0,383 - 0,250}{\sqrt{0,31 \cdot 0,69 \cdot \left(\frac{1}{320} + \frac{1}{264}\right)}} = \frac{0,133}{\sqrt{(0,2139)(0,006913)}} = \frac{0,133}{0,03845} = 346.$$

$3,46 > 13,96$ Se rechaza H_0. Este análisis sí da base para concluir que la asistencia a la educación formal es mayor dentro de las mujeres de estrato Medio-Alto.

b) La diferencia con el resultado obtenido en el punto 15 se debe a que, en este caso, la medición es más precisa porque se relacionan las mujeres que estudian con un grupo de población que está más en edad de asistir a la educación formal: menos de 35 años. En el punto 15 se utilizaron todas las mujeres como denominador del porcentaje, y resulta que en el estrato Medio-Alto hay un ligero predominio de las mujeres de 35 años y más.

17.

	18-24 años	Total
Número de entrevistados	232	1 202
La Deuda Interna es el principal problema	11,9%	9,9%
Solucionar problema de D.I es la principal tarea del próximo gobierno	10,8	7,6

Dado que la muestra total es grande, pueden tomarse sus valores como si correspondieran a la población y someter a prueba la hipótesis nula $H_0: P = 0,099$ contra $H_1: P \neq 0,099$ y proceder en forma similar para la segunda opinión.

$$z = \frac{p-P}{\sigma_p} = \frac{0,119 - 0,099}{\sqrt{0,099 \cdot \frac{0,901}{232}}} = \frac{0,02}{0,0196} = 1,02.$$

$1,02 < 1,96$, se mantiene la hipótesis de que $P = 0,099$. No hay base para concluir que los jóvenes muestran más preocupación que la población general por la Deuda Interna.

$H_0: P = 0,076$ $\quad\quad H_1: P \neq 0,076$

$$z = \frac{p-P}{\sigma_p} = \frac{0,108 - 0,076}{\sqrt{0,076 \cdot \frac{0,924}{232}}} = \frac{0,032}{0,0174} = 1,84.$$

$1,84 < 1,96$, se mantiene la hipótesis H_0. La proporción de jóvenes con la opinión de que la Deuda Interna es la tarea más importante del próximo gobierno no difiere de la observada en la población general. No hay base para concluir que los jóvenes muestran más preocupación que la población general con la Deuda Interna.

En resumen, usando el nivel de significancia más usual, 5%, no hay evidencia de que las opiniones de los jóvenes y de la población general difieran en la importancia asignada a la Deuda Interna. La afirmación del periódico es incorrecta.

18. $\sigma = 3,6 \qquad \bar{x} = 9,1 \qquad n = 110$

$H_0 : \mu = 9,7 \qquad H_1 : \mu < 9,7$

$$z_c = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{9,1 - 9,7}{\frac{3,6}{\sqrt{110}}} = \frac{-0,6}{\frac{3,6}{10,49}} = \frac{-0,6}{0,343} = 1,749.$$

Como $|-1,749| > 1,645$ puede afirmarse, con un nivel de significancia de 5%, que los cambios redujeron los días promedio de estancia hospitalaria, para los pacientes que se someten al procedimiento quirúrgico.[20]

20. Si se usa el nivel de significancia de 1%, la prueba no resulta significativa porque $|-1,749| < 1,96$.