Soluciones - Inferencia estadística

Curso: Universidad 1

Solución del ejercicio 1

Planteamiento. Si la población es normal, la media de una muestra aleatoria también lo es, con la misma media y desviación típica $\frac{\sigma}{\sqrt{n}}$. Tipificamos con ese error estándar y usamos la tabla de la normal.

a. $\bar{X} \sim N\left(170, \frac{10}{\sqrt{25}}\right) = N(170, 2)$.
b. $P\left(Z > \frac{174-170}{2}\right) = P(Z > 2) \approx 0{,}0228$.
c. $P(-2$.
d. Para un estudiante: $P\left(Z > \frac{4}{10}\right) = P(Z > 0{,}4) \approx 0{,}3446$. La media de 25 estaturas varía mucho menos que una estatura individual.

Solución del ejercicio 2

Planteamiento. Cuando np y $n(1-p)$ superan 5, la proporción muestral es aproximadamente normal con media p y desviación $\sqrt{\frac{p(1-p)}{n}}$. Tipificamos y calculamos las probabilidades con la tabla.

a. $np = 40$ y $n(1-p) = 60$, ambos mayores que 5: $\hat{p} \approx N\left(0{,}4,\ \sqrt{\frac{0{,}4 \cdot 0{,}6}{100}}\right)$.
b. $\sqrt{0{,}0024} \approx 0{,}0490$.
c. $P\left(Z > \frac{0{,}1}{0{,}0490}\right) = P(Z > 2{,}04) \approx 0{,}0206$.
d. $P(-1{,}02$.

Solución del ejercicio 3

Planteamiento. El intervalo de confianza para la media con σ conocida es $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Calculamos el error estándar, lo multiplicamos por el valor crítico del nivel pedido e interpretamos el resultado.

a. $\frac{\sigma}{\sqrt{n}} = \frac{6}{6} = 1$ cl.
b. $52 \pm 1{,}96 \cdot 1 = [50{,}04;\ 53{,}96]$ cl.
c. $52 \pm 2{,}576 = [49{,}42;\ 54{,}58]$: más confianza exige un intervalo más ancho.
d. El método produce intervalos que contienen la media real en el 95 % de las muestras; no significa que la media «tenga» un 95 % de probabilidad de estar en este intervalo concreto, porque la media es fija.

Solución del ejercicio 4

Planteamiento. Despejamos el tamaño muestral de la fórmula del error máximo y redondeamos siempre hacia arriba, porque un tamaño menor no garantizaría el error pedido. Comparando dos casos vemos cómo crece n.

a. $n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$, redondeando siempre hacia arriba.
b. $n = (1{,}96 \cdot 6)^2 = 138{,}30$: hacen falta 139 envases.
c. $n = \left(\frac{1{,}96 \cdot 6}{0{,}5}\right)^2 = 553{,}19$: 554 envases.
d. Se multiplica por 4, porque el error es inversamente proporcional a $\sqrt{n}$.

Solución del ejercicio 5

Planteamiento. Para una proporción, el intervalo es $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Calculamos la proporción muestral, el error estándar y el intervalo, y lo usamos para responder a la pregunta.

a. $\hat{p} = \frac{240}{600} = 0{,}4$.
b. $\sqrt{\frac{0{,}4 \cdot 0{,}6}{600}} = 0{,}02$.
c. $0{,}4 \pm 1{,}96 \cdot 0{,}02 = [0{,}361;\ 0{,}439]$.
d. Sí: todo el intervalo está por debajo de 0,5.

Solución del ejercicio 6

Planteamiento. Un contraste sigue siempre los mismos pasos: hipótesis, estadístico, región crítica o p-valor y decisión. Aquí la alternativa es bilateral, así que la región crítica tiene dos colas y el p-valor se duplica.

a. $H_0 : \mu = 500$ frente a $H_1 : \mu \neq 500$.
b. $z = \frac{496 - 500}{12/\sqrt{36}} = \frac{-4}{2} = -2$.
c. Al 5 %: región crítica $|z| > 1{,}96$; como $|-2| > 1{,}96$, se rechaza H_0. Al 1 %: $|z| > 2{,}576$; no se rechaza.
d. p-valor $= 2P(Z$. Es la probabilidad de obtener una media tan alejada si la máquina estuviera bien ajustada; al ser menor que 0,05 pero mayor que 0,01, la evidencia es moderada.

Solución del ejercicio 7

Planteamiento. Estimamos la proporción, construimos el intervalo de confianza y usamos la fórmula del tamaño muestral con el peor caso $p = 0{,}5$. Por último, valoramos qué información debe acompañar a una estimación.

a. $\hat{p} = \frac{456}{1200} = 0{,}38$, es decir, un 38 %.
b. $0{,}38 \pm 1{,}96\sqrt{\frac{0{,}38 \cdot 0{,}62}{1200}} = 0{,}38 \pm 0{,}027$: entre el 35,3 % y el 40,7 %.
c. $n = \frac{1{,}96^2 \cdot 0{,}25}{0{,}03^2} = 1067{,}1$: 1068 personas.
d. El margen de error y el nivel de confianza. Además, la intención de voto declarada no es el voto real.

Solución del ejercicio 8

Planteamiento. La sospecha de que los paquetes pesan menos da un contraste unilateral a la izquierda. Calculamos el estadístico, el p-valor como área de la cola izquierda y lo comparamos con cada nivel de significación.

a. $H_0 : \mu \geq 250$ (o $\mu = 250$) frente a $H_1 : \mu$.
b. $z = \frac{248{,}5 - 250}{5/\sqrt{50}} \approx -2{,}12$.
c. p-valor $= P(Z$.
d. Al 5 %: se rechaza H_0 (p-valor menor que 0,05): hay evidencia de que los paquetes pesan menos. Al 1 %: no se rechaza. La asociación puede denunciarlo, pero la evidencia no es abrumadora.

Solución del ejercicio 9

Planteamiento. Con una muestra grande, el teorema central del límite permite usar la normal y estimar σ con s. Construimos el intervalo al 99 % y lo comparamos con el objetivo; después calculamos el tamaño para un error menor.

a. Con $n = 64$ el teorema central del límite garantiza que la media muestral es aproximadamente normal, y s es una buena estimación de σ.
b. $85 \pm 2{,}576 \cdot \frac{15}{8} = 85 \pm 4{,}83$: $[80{,}17;\ 89{,}83]$ min.
c. Sí, aunque por poco: todo el intervalo queda por debajo de 90 minutos, ya que su extremo superior es 89,83 minutos.
d. $n = \left(\frac{2{,}576 \cdot 15}{3}\right)^2 \approx 165{,}9$: 166 pacientes.

Solución del ejercicio 10

Planteamiento. Planteamos un contraste unilateral para la proporción, usando en el error estándar el valor de p de la hipótesis nula. Decidimos con la región crítica y el p-valor y reflexionamos sobre el diseño del estudio.

a. $H_0 : p = 0{,}6$ frente a $H_1 : p > 0{,}6$.
b. $\hat{p} = 0{,}675$ y $z = \frac{0{,}675 - 0{,}6}{\sqrt{0{,}6 \cdot 0{,}4/200}} \approx 2{,}17$.
c. Región crítica $z > 1{,}645$: se rechaza H_0. p-valor $\approx 0{,}0152$.
d. Aporta evidencia estadística, pero no prueba la causa: puede que el grupo fuera

diferente o el examen más fácil. Haría falta un grupo de control comparable en el mismo curso.

Solución del ejercicio 11

Planteamiento. Calculamos la región de aceptación bajo H_0 y, suponiendo cierta una alternativa concreta, la probabilidad de caer en ella, que es el error de tipo II. Así vemos la relación entre α, β y el tamaño de la muestra.

a. Tipo I: parar la máquina cuando está bien ajustada (probabilidad $\alpha = 0{,}05$). Tipo II: no pararla cuando está desajustada.
b. $10 \pm 1{,}96 \cdot \frac{0{,}2}{4} = [9{,}902;\ 10{,}098]$ mm.
c. $\beta = P(9{,}902$: casi la mitad de las veces no se detecta.
d. Aumentando el tamaño de la muestra: el error estándar disminuye y las distribuciones bajo H_0 y bajo H_1 se solapan menos.

Solución del ejercicio 12

Planteamiento. Para comparar dos medias con muestras independientes, el error estándar de la diferencia es la raíz de la suma de varianzas divididas por los tamaños. Con él hacemos el contraste y el intervalo de confianza.

a. $\sqrt{\frac{16}{40} + \frac{25}{50}} = \sqrt{0{,}9} \approx 0{,}949$ kg.
b. $z = \frac{32-30}{0{,}949} \approx 2{,}11 > 1{,}96$: se rechaza la igualdad de medias.
c. $2 \pm 1{,}96 \cdot 0{,}949 = [0{,}14;\ 3{,}86]$ kg.
d. El abono A parece producir algo más, entre 0,14 y 3,86 kg por parcela; antes de decidir conviene comparar también los precios de los abonos.

Solución del ejercicio 13

Planteamiento. Con una muestra pequeña y σ desconocida, la media tipificada con s sigue una t de Student con $n-1$ grados de libertad. Usamos su valor crítico, mayor que el de la normal, para construir el intervalo.

a. Porque σ es desconocida y la muestra es pequeña: al estimarla con s aparece una incertidumbre adicional que la t tiene en cuenta con colas más pesadas.
b. $n - 1 = 8$ grados de libertad.
c. $12{,}2 \pm 2{,}306 \cdot \frac{0{,}6}{3} = 12{,}2 \pm 0{,}461$: $[11{,}739;\ 12{,}661]$ h.
d. Con 1,96 saldría $12{,}2 \pm 0{,}392$, un intervalo más estrecho que no alcanzaría la confianza declarada.

Solución del ejercicio 14

Planteamiento. En la prueba de bondad de ajuste comparamos las frecuencias

observadas con las esperadas si el dado fuera equilibrado. El estadístico χ^2 mide la discrepancia y se compara con el valor crítico de la tabla.

a. H_0: el dado es equilibrado (cada cara con probabilidad $\frac{1}{6}$). H_1: no lo es.
b. Esperadas: $120 \cdot \frac{1}{6} = 20$ cada cara. $\chi^2 = \frac{25+4+4+25+9+9}{20} = 3{,}80$.
c. $3{,}80$: no se rechaza H_0. Las diferencias son compatibles con el azar.
d. Las 6 frecuencias deben sumar 120, de modo que solo 5 son libres.

Solución del ejercicio 15

Planteamiento. Hacemos un estudio completo: estimación puntual, intervalo de confianza y contraste unilateral de una afirmación concreta. El intervalo permite valorar las dos afirmaciones y la fórmula del tamaño, planificar un nuevo estudio.

a. $\hat{p} = \frac{280}{800} = 0{,}35$; error estándar $\sqrt{\frac{0{,}35 \cdot 0{,}65}{800}} \approx 0{,}0169$.
b. $0{,}35 \pm 0{,}033 = [0{,}317;\ 0{,}383]$.
c. La patronal queda refutada (todo el intervalo supera 0,30). La del sindicato no se puede confirmar: el intervalo incluye valores por debajo de $\frac{1}{3}$.
d. $z = \frac{0{,}35 - 0{,}3333}{\sqrt{\frac{1}{3} \cdot \frac{2}{3} / 800}} \approx 1{,}00$: no se rechaza H_0; los datos no bastan para afirmar que son más de un tercio.
e. $n = \frac{1{,}96^2 \cdot 0{,}35 \cdot 0{,}65}{0{,}02^2} \approx 2185$: unas 2185 personas.

Solución del ejercicio 16

Planteamiento. Seguimos los pasos del contraste unilateral para la media con muestra grande, estimando σ con s. Completamos la conclusión con un intervalo de confianza, que informa de la magnitud de la diferencia.

a. $H_0 : \mu = 5$ frente a $H_1 : \mu > 5$.
b. $z = \frac{5{,}3 - 5}{0{,}8/\sqrt{40}} \approx 2{,}37$.
c. $P(Z > 2{,}37) \approx 0{,}0089$.
d. p-valor menor que 0,05 y que 0,01: se rechaza H_0 en ambos niveles.
e. Hay evidencia sólida de que el consumo real supera el anunciado: con un 95 % de confianza está entre 5,05 y 5,55 L/100 km.

Solución del ejercicio 17

Planteamiento. Un intervalo de confianza se construye como media más o menos margen de error. Recuperamos ambos números y usamos la equivalencia entre intervalos y contrastes bilaterales para decidir sin cálculos.

a. $\bar{x} = \frac{12{,}4 + 15{,}8}{2} = 14{,}1$ y $E = 1{,}7$.
b. $\mu = 12$ está fuera del intervalo: se rechaza. $\mu = 15$ está dentro: no se rechaza.

c. $\frac{1{,}7}{1{,}96} \approx 0{,}867$.
d. $z = \frac{14{,}1-12}{0{,}867} \approx 2{,}42$; p-valor $= 2P(Z > 2{,}42) \approx 0{,}0155$.
e. Un contraste bilateral al nivel α rechaza $H_0 : \mu = \mu_0$ exactamente cuando μ_0 queda fuera del intervalo de confianza al $1 - \alpha$.

Solución del ejercicio 18

Planteamiento. En la prueba de independencia, las frecuencias esperadas se obtienen multiplicando los totales de fila y columna y dividiendo entre el total. El estadístico χ^2 compara observadas y esperadas con un grado de libertad.

a. H_0: el uso es independiente de la rama. H_1: depende.
b. $E_{ij} = \frac{\text{fila} \cdot \text{columna}}{200}$: Ciencias $52{,}5$ y $47{,}5$; Letras $52{,}5$ y $47{,}5$.
c. $\chi^2 = \sum \frac{(O-E)^2}{E} \approx 4{,}511$.
d. $4{,}511 > 3{,}841$: se rechaza la independencia al 5 %.
e. Ciencias: 60 %; Letras: 45 %. Los estudiantes de Ciencias usan más la biblioteca semanal, y la diferencia no parece debida al azar.

Solución del ejercicio 19

Planteamiento. Un estimador es insesgado si su esperanza es el parámetro; entre dos insesgados es mejor el de menor varianza. Calculamos esperanzas y varianzas con las propiedades de las combinaciones lineales de variables independientes.

a. $E(T_1) = \frac{3\mu}{3} = \mu$ y $E(T_2) = \frac{\mu+2\mu+\mu}{4} = \mu$.
b. $\operatorname{Var}(T_1) = \frac{3\sigma^2}{9} = \frac{\sigma^2}{3}$ y $\operatorname{Var}(T_2) = \frac{\sigma^2+4\sigma^2+\sigma^2}{16} = \frac{3\sigma^2}{8}$.
c. T_1, porque $\frac{1}{3}$: con el mismo centro, sus valores se concentran más alrededor de μ.
d. Porque es insesgada: $E(s^2) = \sigma^2$, mientras que la varianza muestral con denominador n subestima σ^2 en un factor $\frac{n-1}{n}$.

Solución del ejercicio 20

Planteamiento. Despejamos el tamaño de la fórmula del error y acotamos $p(1-p)$ completando cuadrados para obtener una fórmula válida en cualquier caso. Después aplicamos la corrección para poblaciones finitas.

a. $n = \frac{z_{\alpha/2}^2\, p(1-p)}{E^2}$.
b. $p(1-p) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \le \frac{1}{4}$, con máximo en $p = \frac{1}{2}$. Entonces $n \le \frac{z_{\alpha/2}^2}{4E^2}$, que vale para cualquier p.
c. $n = \frac{1{,}96^2}{4 \cdot 0{,}0001} = 9604$.

d. $n' = \frac{9604}{1+9603/2000} \approx 1655{,}4$: unas 1656 personas.