Soluciones de Ejercicios de Inferencia Estadística

Solución Ejercicio 1

a) Tamaño muestral

Partimos del intervalo de confianza $(1-\alpha)$ para una proporción:

$$\left(p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}},\ p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right)$$

El error máximo de estimación es:

$$E = z_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}$$

Como no disponemos de una estimación previa de π, utilizamos el valor máximo:

$$\pi(1-\pi) \leq 0{,}25$$

Imponemos que el error sea como máximo 0,03:

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{0{,}25}{n}} \leq 0{,}03$$

Para un nivel de confianza del 99 %:

$$z_{1-\frac{\alpha}{2}} = 2{,}575$$

Entonces:

$$2{,}575\sqrt{\frac{0{,}25}{n}} \leq 0{,}03 \quad \Longrightarrow \quad n \geq 1841{,}84$$

Por tanto, el tamaño mínimo de muestra es:

$$\boxed{n = 1842}$$

b) Contraste de hipótesis para dos proporciones

Las proporciones muestrales son:

$$p_1 = \frac{63}{500} = 0{,}126 \qquad p_2 = \frac{79}{700} = 0{,}113$$

Planteamos el contraste:

$$\begin{cases} H_0 : \pi_1 - \pi_2 = 0 \\ H_A : \pi_1 - \pi_2 > 0 \end{cases}$$

El estadístico de contraste es:

$$Z_{\text{obs}} = \frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Sustituyendo valores:

$$Z_{\text{obs}} = \frac{0{,}126 - 0{,}113}{\sqrt{\frac{0{,}126 \cdot 0{,}874}{500} + \frac{0{,}113 \cdot 0{,}887}{700}}} = 0{,}68$$

El p-valor es:

$$p\text{-valor} = P(Z > 0{,}68) = 0{,}2483$$

Como p-valor $> \alpha$ tanto para $\alpha = 0{,}05$ como para $\alpha = 0{,}01$, se concluye que:

$$\boxed{\text{No se rechaza } H_0}$$

No hay evidencia suficiente para afirmar que la aceptación sea mayor en Madrid que en Barcelona.

c) Intervalo de confianza para la diferencia de proporciones

El intervalo de confianza $(1 - \alpha)$ para $\pi_1 - \pi_2$ es:

$$\left(p_1 - p_2 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}},\ p_1 - p_2 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \right)$$

Para $\alpha = 0{,}05$, $z_{1-\frac{\alpha}{2}} = 1{,}96$:

$$(0{,}013 - 1{,}96 \cdot 0{,}019,\ 0{,}013 + 1{,}96 \cdot 0{,}019) = (0{,}013 \pm 0{,}037)$$

Por tanto:

$$\boxed{(-0{,}024,\ 0{,}05)}$$

- Con un 95 % de confianza, la diferencia de proporciones está entre un $-2{,}4$ % (mayor en Barcelona) y un 5 % (mayor en Madrid).
- Como el valor 0 pertenece al intervalo, no se detectan diferencias significativas entre las proporciones.

$$\boxed{\text{Se acepta la igualdad de proporciones al nivel } \alpha = 0{,}05}$$

Solución Ejercicio 2

Disponemos de las proporciones muestrales:

$$p_2 = \frac{33}{200\,745} = 0{,}000164 \quad \text{(proporción muestral de niños vacunados que enfermaron de polio)}$$

$$p_1 = \frac{110}{201\,229} = 0{,}00055 \quad \text{(proporción muestral de niños con placebo que enfermaron de polio)}$$

Queremos hacer inferencias sobre el parámetro $\pi_1 - \pi_2$, la diferencia entre las proporciones poblacionales de niños que enfermaron (placebo frente a vacuna), que mide la efectividad de la vacuna.

a) Intervalo de confianza del 99 % para $\pi_1 - \pi_2$

Construimos el intervalo de confianza para la diferencia de proporciones:

$$P\left(a < \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} < b \right) = 0{,}99$$

En la tabla $N(0,1)$, para 0,99:

$$a = z_{0{,}005} = -2{,}575, \qquad b = z_{0{,}995} = 2{,}575$$

Luego:

$$P\left(-2{,}575 < \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} < 2{,}575 \right) = 0{,}99$$

Equivalentemente,

$$(p_1 - p_2) - 2{,}575\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < \pi_1 - \pi_2 < (p_1 - p_2) + 2{,}575\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Sustituyendo:

$$\pi_1 - \pi_2 \in (0{,}000386 \pm 2{,}575 \cdot 0{,}0000286) = (0{,}000386 \pm 0{,}0000736)$$

Por tanto, el intervalo es:

$$\boxed{IC_{0{,}99}(\pi_1 - \pi_2) = (0{,}00031,\ 0{,}00046)}$$

Interpretación:

- Con un 99 % de confianza, la proporción de niños **no vacunados** que contrajeron la polio estuvo entre 31 y 46 por cada 100 000 por encima de la proporción de niños **vacunados** que enfermaron.

- Es decir, la vacuna consiguió reducir entre 31 y 46 casos por cada 100 000.

b) Conclusión sobre la efectividad: contraste para π_2

Planteamos un contraste para comprobar si la vacuna logra que la proporción de niños vacunados que se contagian esté en 15 por cada 100 000, es decir, $\pi_2 = 0{,}00015$.

$$\begin{cases} H_0: \ \pi_2 = 0{,}00015 & \text{(la vacuna reduce la incidencia a la mitad)} \\ H_A: \ \pi_2 > 0{,}00015 & \text{(la vacuna no logra el objetivo)} \end{cases}$$

El estadístico es:

$$Z_{\text{obs}} = \frac{p_2 - \pi_2}{\sqrt{\frac{\pi_2(1-\pi_2)}{n_2}}} = \frac{0{,}000164 - 0{,}00015}{\sqrt{\frac{0{,}00015 \cdot 0{,}99985}{200\,745}}} = 0{,}0000273$$

$$p\text{-valor} = P(Z > 0{,}0000273) = 0{,}5 > \alpha \quad \Longrightarrow \quad \boxed{\text{Aceptamos } H_0}$$

Conclusión:

- Con un 1 % de margen de error, podemos admitir que la incidencia de polio en los niños vacunados fue de 15 por cada 100 000,
- por tanto, se redujo el valor inicial a la mitad y la vacuna se considera **efectiva**.

Solución Ejercicio 3

Denotamos:

$$\mu_1 = \text{importe medio de todos los préstamos de la entidad A}$$

$$\mu_2 = \text{importe medio de todos los préstamos de la entidad B.}$$

Datos muestrales (en miles de euros):

$$n_1 = 41, \quad \bar{x} = 15, \quad s_1 = 9{,}8; \qquad n_2 = 49, \quad \bar{y} = 13, \quad s_2 = 9{,}3.$$

a) Contraste para la diferencia de medias

Planteamos el contraste:

$$\begin{cases} H_0: \ \mu_1 - \mu_2 = 0, \\ H_A: \ \mu_1 - \mu_2 > 0. \end{cases}$$

Es un contraste sobre diferencia de medias con varianzas poblacionales desconocidas y tamaños muestrales grandes (> 40), por lo que usamos la aproximación normal:

$$Z_{\text{obs}} = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(15 - 13) - 0}{\sqrt{\frac{96{,}04}{41} + \frac{86{,}49}{49}}} = 0{,}987.$$

El p-valor (cola derecha) es:

$$p\text{-valor} = P(Z > 0{,}987) = 0{,}1611 > 0{,}01.$$

Por tanto,

$$\boxed{\text{Aceptamos } H_0.}$$

Conclusión: al 1 % no existen diferencias significativas entre los importes medios de los créditos concedidos por las dos entidades.

b) Intervalo de confianza del 99 % para μ_2 (muestra grande, σ desconocida)

Para la entidad B, el intervalo $(1-\alpha)$ para la media es:

$$\left(\bar{y} - z_{1-\frac{\alpha}{2}}\frac{s_2}{\sqrt{n_2}},\ \bar{y} + z_{1-\frac{\alpha}{2}}\frac{s_2}{\sqrt{n_2}}\right).$$

Con $1-\alpha = 0{,}99$, $z_{1-\frac{\alpha}{2}} = z_{0{,}995} = 2{,}575$:

$$\left(13 \pm 2{,}575\frac{9{,}3}{\sqrt{49}}\right) = (13 \pm 3{,}42) = (9{,}58,\ 16{,}42).$$

Por tanto,

$$\boxed{IC_{0{,}99}(\mu_2) = (9{,}58,\ 16{,}42) \text{ (miles de euros).}}$$

Interpretación:

- Con una confianza del 99 %, el importe medio de los créditos de la entidad B está entre 9 580 y 16 420 euros.
- Si tomamos muchas muestras de 49 créditos y construimos el intervalo anterior, aproximadamente el 99 % de esos intervalos contendrán el valor real de μ_2.

c) Tamaño muestral para estimar μ_1 con error máximo 300 euros

Queremos que el error máximo sea $E = 0{,}3$ (miles de euros). Exigimos:

$$z_{1-\frac{\alpha}{2}}\frac{s_1}{\sqrt{n}} \leq 0{,}3.$$

Con $z_{0{,}995} = 2{,}575$ y $s_1 = 9{,}8$:

$$2{,}575\frac{9{,}8}{\sqrt{n}} \leq 0{,}3 \quad \Longrightarrow \quad n \geq 7075{,}61.$$

Luego, el tamaño mínimo es:

$$\boxed{n = 7076.}$$

Es decir, hay que analizar al menos 7076 créditos de la primera entidad para lograr ese margen de error con confianza 0,99.