Estadística

Soluciones ejercicios: Inferencia

Versión 8

Emilio Letón

1. Nivel 1

1. Decir si es verdadera o falsa la siguiente afirmación: "Para querer probar que una media poblacional es superior a un valor teórico μ_0 hay que plantear el contraste $H_0 : \bar{x} \leq \mu_0$ frente a $H_1 : \bar{x} > \mu_0$". En el supuesto de que sea falsa, plantear el contraste correcto.

   SOLUCIÓN:

   Falso. La forma de plantear el contraste es $H_0 : \mu \leq \mu_0$ frente a $H_1 : \mu > \mu_0$.

2. Se ha realizado un estudio para ver si la variable SEXO puede explicar el IMC. Los resultados experimentales resumidos por un programa estadístico han sido los siguientes:

	SEXO=0 (H)	SEXO=1 (M)
Count	50	50
Average	24,2197	21,9712
Variance	2,97796	4,22
Range	1,72568	2,05426

   Se asume que la variable IMC es Normal en los dos grupos y que hay homocedasticidad (igualdad de varianzas en los dos grupos).

   Teniendo en cuenta que

$z_{0,025} = 1{,}9600$	$z_{0,050} = 1{,}6450$

$t_{48,0,025} = 2{,}0106$	$t_{48,0,050} = 1{,}6772$
$t_{49,0,025} = 2{,}0096$	$t_{49,0,050} = 1{,}6766$
$t_{50,0,025} = 2{,}0086$	$t_{50,0,050} = 1,6759$

$t_{98,0,025} = 1{,}9845$	$t_{98,0,050} = 1{,}6606$
$t_{99,0,025} = 1{,}9842$	$t_{99,0,050} = 1{,}6604$
$t_{100,0,025} = 1{,}9840$	$t_{100,0,050} = 1{,}6602$

$\chi^2_{48,0,025} = 69{,}0226$	$\chi^2_{48,0,050} = 65{,}1708$	$\chi^2_{48,0,975} = 30{,}7545$	$\chi^2_{48,0,950} = 33{,}0981$
$\chi^2_{49,0,025} = 70{,}2224$	$\chi^2_{49,0,050} = 66{,}3386$	$\chi^2_{49,0,975} = 31{,}5549$	$\chi^2_{49,0,950} = 33{,}9303$
$\chi^2_{50,0,025} = 71{,}4202$	$\chi^2_{50,0,050} = 67{,}5048$	$\chi^2_{50,0,975} = 32{,}3574$	$\chi^2_{50,0,950} = 34{,}7643$

   Se pide:

   a) Calcular el $IC95\,\%\,(\mu_H)$.

b) Calcular el $IC95\,\%\left(\sigma_H^2\right)$ y el $IC95\,\%\left(\sigma_H\right)$

c) Realizar el contraste

$$\begin{aligned} H_0 &: \mu_H - \mu_M \leq 0 \\ H_1 &: \mu_H - \mu_M > 0 \end{aligned}$$

con un nivel de significación de 0.05.

d) Explicar con palabras el resultado del apartado anterior.

SOLUCIÓN:

a) El $IC95\,\%\left(\mu_H\right)$ viene dado por

$$\left(\overline{x} \mp t_{n_H-1,0{,}05/2} \cdot \sqrt{\frac{\widehat{s}_H^2}{n_H}}\right)$$

$$= \left(24{,}2197 - 2{,}0096 \cdot \sqrt{\frac{2,97796}{50}}, 24{,}2197 + 2{,}0096 \cdot \sqrt{\frac{2,97796}{50}}\right)$$

$$= (23{,}7294, 24{,}7100)$$

b) El $IC95\,\%\left(\sigma_H^2\right)$ viene dado por

$$\left(\frac{(n_H-1)\,\widehat{s}_H^2}{\chi^2_{n_H-1,0{,}025}}, \frac{(n_H-1)\,\widehat{s}_H^2}{\chi^2_{n_H-1,0{,}975}}\right)$$

$$= \left(\frac{49 \cdot 2{,}97796}{70{,}2224}, \frac{49 \cdot 2{,}97796}{31{,}5549}\right) = (2{,}0780, 4{,}6243)$$

Por tanto el $IC95\,\%\left(\sigma_H\right)$ viene dado por

$$\left(\sqrt{2{,}0780}, \sqrt{4{,}6243}\right) = (1{,}4415, 2{,}1504)$$

c) La regla de decisión es:

$$\text{Si } \frac{\overline{x}_{H-}\overline{x}_M - 0}{\sqrt{\frac{\widehat{s}_T^2}{n_H} + \frac{\widehat{s}_T^2}{n_M}}} > t_{n_H+n_M-2,0{,}05} \text{ se rechaza } H_0$$

Dado que

$$\widehat{s}_T^2 = \frac{(n_H-1)\,\widehat{s}_H^2 + (n_M-1)\,\widehat{s}_M^2}{n_H - n_M - 2} = \frac{49 \cdot 2{,}97796 + 49 \cdot 4{,}22}{98} = 3{,}59898$$

y que de las tablas

$$t_{n_H+n_M-2,0{,}05} \simeq 1{,}6606$$

se tiene que

$$\frac{\overline{x}_{H-}\overline{x}_M}{\sqrt{\frac{\widehat{s}_T^2}{n_H} + \frac{\widehat{s}^2}{n_M}}} = \frac{2{,}2485}{0{,}3794} = 5{,}9261 > 1{,}6606$$

con lo que se rechaza H_0 con un nivel de significación de $\alpha = 0{,}05$, con lo que el p-valor<0.05. De las tablas completas se obtiene que el p-valor $< 0{,}0005$.

d) Hay evidencia suficiente para afirmar que $\mu_H - \mu_M > 0$.

Sí influye el hecho de estar en un grupo o en otro a la hora de explicar IMC.

Los del grupo H tienen un IMC superior al grupo M de forma significativa (p-valor$< 0{,}05$).

El azar no puede explicar las diferencias observadas: el grupo influye.

3. Se ha realizado un estudio para ver si la variable MOTOR puede explicar el consumo de LITROS de gasolina a los 100 km. Los resultados experimentales resumidos por un programa estadístico han sido los siguientes:

	MOTOR=1	MOTOR=2
Count	41	31
Average	7.3154	6.98581
Median	7.25977	6.929
Variance	0.255072	0.179241
Standard deviation	0.505046	0.423369
Standard error	0.078875	0.0760392
Range	2.25145	1.65585

Se asume que la variable LITROS es Normal en los dos grupos y que hay homocedasticidad (igualdad de varianzas en los dos grupos).

Teniendo en cuenta que

$z_{0,025} = 1{,}9600$	$z_{0,050} = 1{,}6450$

$t_{70,0,025} = 1{,}9940$	$t_{70,0,050} = 1{,}6669$
$t_{71,0,025} = 1{,}9939$	$t_{71,0,050} = 1{,}6666$
$t_{72,0,025} = 1{,}9944$	$t_{72,0,050} = 1,6663$

$t_{70,0,0025} = 2{,}8987$	$t_{70,0,0050} = 2,6479$
$t_{71,0,0025} = 2{,}8974$	$t_{71,0,0050} = 2,6469$
$t_{72,0,0025} = 2{,}8961$	$t_{72,0,0050} = 2,6459$

Se pide:

a) Calcular el $IC95\,\%\,(\mu_1 - \mu_2)$.

b) Realizar el contraste

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$

con un nivel de significación de 0.05.

c) ¿Son coherentes los apartados a) y b)?

d) Explicar con palabras el resultado del apartado anterior.

SOLUCIÓN:

a) El $IC95\,\%\,(\mu_1 - \mu_2)$ viene dado por

$$\left(\overline{x}_1 - \overline{x}_2 \mp t_{n_1+n_2-2,0,05/2} \cdot \sqrt{\frac{\widehat{s}_T^2}{n_1} + \frac{\widehat{s}_T^2}{n_2}} \right)$$

siendo

$$\widehat{s}_T^2 = \frac{(n_1-1)\,\widehat{s}_1^2 + (n_2-1)\,\widehat{s}_2^2}{n_1 - n_2 - 2} = \frac{40 \cdot 0{,}255072 + 30 \cdot 0{,}179241}{70} = 0{,}226$$

$$t_{n_1+n_2-2,0,025} = 1{,}994$$

Por tanto el $IC95\,\%\,(\mu_1 - \mu_2)$ es

$$\left(7{,}3154 - 6{,}98581 \mp 1{,}994 \cdot \sqrt{\frac{0{,}226}{41} + \frac{0{,}226}{31}}\right)$$

$$= (0{,}106, 0{,}553)$$

b) La regla de decisión es:

$$\text{Si } \left| \frac{\overline{x}_1 - \overline{x}_2 - 0}{\sqrt{\frac{\widehat{s}_T^2}{n_1} + \frac{\widehat{s}_T^2}{n_2}}} \right| > t_{n_1+n_2-2,0{,}025} \text{ se rechaza } H_0$$

Dado que

$$\left| \frac{\overline{x}_1 - \overline{x}_2 - 0}{\sqrt{\frac{\widehat{s}_T^2}{n_1} + \frac{\widehat{s}_T^2}{n_2}}} \right| = \left| \frac{7{,}3154 - 6{,}98581}{\sqrt{\frac{0{,}226}{41} + \frac{0{,}226}{31}}} \right| = \left| \frac{0{,}3296}{0{,}1131} \right| = 2{,}9352 > 1{,}994$$

con lo que se rechaza H_0 con un nivel de significación de $\alpha = 0{,}05$. De los valores dados se obtiene que el p-valor es aproximadamente

$$\text{p-valor} = 2 \cdot 0{,}0025 = 0{,}0050$$

c) Son coherentes ya que el contraste es bilateral y se observa que el valor de diferencias de medias de cero no pertenece al intervalo de confianza y que se rechaza la hipótesis nula.

d) Hay evidencia suficiente para afirmar que $\mu_1 - \mu_2 = 0$.

Sí influye el hecho de estar en un grupo o en otro a la hora de explicar LITROS.

Los del grupo 1 tienen un LITROS distinto al grupo 2 de forma significativa (p-valor$<$ 0,05).

El azar no puede explicar las diferencias observadas: el grupo influye.

No es irrelevante el hecho de tener un motor 1 o tener un motor 2.

2. Nivel 2

1. Se ha realizado un estudio para ver si la variable MOTOR puede explicar el consumo de LITROS de gasolina a los 100 km. Los resultados experimentales resumidos por un programa estadístico han sido los siguientes:

	MOTOR=1	MOTOR=2
Count	41	31
Average	7.3154	6.98581
Median	7.25977	6.929
Variance	0.255072	0.179241
Standard deviation	0.505046	0.423369
Standard error	0.078875	0.0760392
Range	2.25145	1.65585

Teniendo en cuenta que

$F_{40,30,0,99} = 0{,}4538$	$F_{41,31,0,99} = 0{,}4589$	$F_{30,40,0,99} = 0{,}4349$	$F_{31,41,0,99} = 0,4408$
$F_{40,30,0,975} = 0{,}5147$	$F_{41,31,0,975} = 0{,}5194$	$F_{30,40,0,975} = 0{,}4978$	$F_{31,41,0,975} = 0{,}5034$
$F_{40,30,0,95} = 0{,}5732$	$F_{41,31,0,95} = 0{,}5777$	$F_{30,40,0,95} = 0{,}5581$	$F_{31,41,0,95} = 0{,}5633$

y asumiendo Normalidad, se pide:

a) Calcular el $IC95\,\%\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$.

b) Realizar el contraste

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$
$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

con un nivel de significación de 0.05.

c) ¿Son coherentes los apartados a) y b)?

d) ¿Se puede asumir homocedasticidad?

SOLUCIÓN:

a) El $IC95\,\%\left(\frac{\sigma_1^2}{\sigma_2^2}\right)$ viene dado por

$$\left(\frac{\widehat{s}_1^2}{\widehat{s}_2^2}\frac{1}{F_{n_1-1,n_2-1,0{,}025}}, \frac{\widehat{s}_1^2}{\widehat{s}_2^2}\frac{1}{F_{n_1-1,n_2-1,0{,}975}}\right)$$

$$= \left(\frac{0{,}255072}{0{,}179241}\frac{1}{2{,}00887}, \frac{0{,}255072}{0{,}179241}\frac{1}{0{,}5147}\right) = (0{,}70839, 2{,}76490)$$

ya que

$$F_{40,30,0{,}025} = \frac{1}{F_{30,40,0{,}975}} = \frac{1}{0{,}4978} = 2{,}00887$$

b) La regla de decisión es:

$$\text{Si } \frac{\widehat{s}_1^2}{\widehat{s}_2^2} > F_{n_1-1,n_2-1,0{,}025} \text{ ó } \frac{\widehat{s}_1^2}{\widehat{s}_2^2} < F_{n_1-1,n_2-1,0{,}975} \text{ se rechaza } H_0$$

Dado que

$$\frac{\widehat{s}_1^2}{\widehat{s}_2^2} = \frac{0{,}255072}{0{,}179241} = 1{,}423$$

no se rechaza H_0 con un nivel de significación de $\alpha = 0{,}05$, con lo que el p-valor $> 0{,}05$.

c) Son coherentes ya que el contraste es bilateral y se observa que el valor de uno pertence al intervalo de confianza y que no se rechaza la hipótesis nula.

d) Se pude asumir homocedasticidad. No hay evidencia suficiente para concluir heterocedasticidad.

3. Nivel 3

1. Se ha realizado un estudio para ver si la variable MOTOR puede explicar el consumo de LITROS de gasolina a los 100 km. Los resultados experimentales resumidos por un programa estadístico han sido los siguientes en cuanto al ajuste de normalidad por el test de Chi-Cuadrado:

	MOTOR=1	MOTOR=2
Chi-square	2.68272	8.99982
d.f.	6	5
p-value	0.847484	0.109071

¿Se puede asumir que la variable LITROS en ambos grupos se distribuye según una Normal con un nivel de significación de 0.05?

SOLUCIÓN:

En ambos casos se pude asumir normalidad, ya que el p-valor>0.05.

2. Se desea realizar un estudio para saber si los alumnos de la UC3M están satisfechos con los medios informáticos que la universidad pone a su disposición. Se desea una confianza del 99 % y una imprecisión del 6 % (amplitud del 12 %). Calcular el tamaño muestral necesario para hacer dicho estudio en los dos supuestos siguientes:

   a) No hay información previa acerca del posible porcentaje de satisfacción.

   b) Se cree que el porcentaje de satisfacción es de un 85 %. ¿El tamaño muestral será mayor o menor que el calculado en el apartado a)? ¿Por qué?

   SOLUCIÓN:

   a) La fórmula para el tamaño muestral para estimar una proporción es

   $$n = \left(\frac{z_{\alpha/2}}{L}\right)^2 pq$$

   Si no se tiene información a priori sobre p se toma $p = 0,50$ (situación de máxima incertidumbre). Con los demás datos, la fórmula se simplifica a

   $$n = \left(\frac{2,58}{0,06}\right)^2 \cdot 0,50 \cdot 0,50 = 463$$

   b) Si se tiene información a priori sobre p cercano a $0,85$, la fórmula se simplifica a

   $$n = \left(\frac{2,58}{0,06}\right)^2 \cdot 0,85 \cdot 0,15 = 236$$

   El tamaño muestral es menor que en el supuesto del apartado a), ya que se dispone de información adicional. La falta de información a priori "se paga" con un mayor tamaño muestral.