

Solución

Estamos ante una situación que el Texto Base identifica como “contraste de la bondad del ajuste (primer caso)”. La forma de proceder es calcular el valor del estadístico

$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - e_i)^2}{e_i}$$

y comparar con el dato que nos proporciona la distribución χ_9^2 (9 grados de libertad) para el nivel de significación seleccionado.

Respecto al cálculo del estadístico, el enunciado nos proporciona la frecuencia observada (O_i) en una muestra aleatoria de tamaño 200, mientras que la frecuencia esperada (e_i) bajo la hipótesis de que los 10 dígitos fuesen igualmente probables es, en cada caso, 20. Entonces,

$$(O_i - e_i)^2 : 100, 0, 1, 1, 1, 25, 1, 4, 25, 36$$

y

$$\sum_{i=1}^{10} \frac{(O_i - e_i)^2}{e_i} = \frac{194}{20} = 9.7$$

Entre los datos auxiliares encontramos $\chi_{9;0.05}^2 = 16.92$; entonces, dado que

$$9.7 < 16.92$$

las desviaciones observadas no son significativas, al nivel $\alpha = 0.05$, y por tanto, aceptamos la hipótesis de que *en la población los 10 dígitos son equiprobables*. \square

Problema 4. Primera semana. De una población normal se extrae una muestra aleatoria de tamaño 10, obteniéndose las siguientes observaciones:

$$1, 2, 5, 1, 3, 7, 4, 2, 3, 2$$

- (a) Obtener una estimación insesgada de la media poblacional.
- (b) Contrastar la hipótesis de que la media poblacional es 3.5, frente a la alternativa de que es menor. Tómese un nivel de significación de $\alpha = 0.01$.

Solución

- (a) Se sabe que la media muestral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es un estimador insesgado

de la media poblacional μ , ya que $E\{\bar{X}\} = \frac{1}{n} \sum_{i=1}^n E\{X_i\} = \mu$.

Para la muestra extraída la estimación viene dada por

$$\bar{x} = \frac{1 + 2 + 5 + 1 + 3 + 7 + 4 + 2 + 3 + 2}{10} = 3.$$

- (b) La muestra obtenida ha proporcionado un valor para la media muestral de $\bar{x} = 3$. Se trata de comprobar si ese valor proporciona evidencia estadística en favor de la hipótesis de que la media poblacional es $\mu = 3.5$ o por el contrario apoya la alternativa de que es inferior.

El problema se plantea en términos de un contraste de hipótesis en el que

$$H_0 : \mu = 3.5 \qquad H_1 : \mu < 3.5.$$

Dado que la variable aleatoria poblacional sigue una distribución normal de media y varianza desconocidas, el contraste se lleva a cabo utilizando la cantidad pivotal

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1},$$

donde

- μ_0 : la media poblacional bajo H_0 . En este caso $\mu_0 = 3.5$.
- \bar{X} : media muestral.
- S^2 : cuasivarianza muestral.
- n : tamaño muestral.

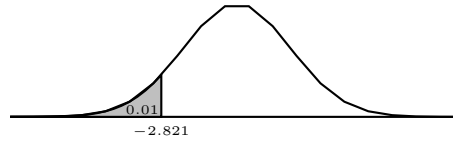
La región crítica del test —conjunto de todas las muestras para las que se rechaza H_0 — viene dada por

$$\mathcal{R} = \left\{ (X_1, X_2, \dots, X_n) : \frac{\bar{X} - \mu_0}{S/\sqrt{n}} < c \right\}$$

donde el punto crítico c se determina con la condición del nivel de significación, la cual viene dada por:

$$P_{H_0}((X_1, X_2, \dots, X_n) \in \mathcal{R}) = P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < c\right) = P(t_{n-1} < c) = 0.01.$$

En este caso, para un tamaño muestral de $n = 10$, el punto crítico vale $c = -2.821$, ya que para una distribución t con 9 grados de libertad se tiene que $P(t_9 < -2.821) = 0.01$ (véase la figura 1, en la que el área de la región sombreada coincide con el nivel de significación $\alpha = 0.01$). Por tanto, la región crítica \mathcal{R} está definida por $\mathcal{R} = \left\{ \frac{\bar{X} - \mu_0}{S/\sqrt{n}} < -2.821 \right\}$.

Figura 1: Punto crítico para una t_9 ($\alpha = 0.01$)

Para la muestra extraída, se sabe que $\bar{x} = 3$ y que

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(1-3)^2 + (2-3)^2 + \dots + (2-3)^2}{9} = \frac{32}{9},$$

de donde se sigue que

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{3 - 3.5}{\sqrt{32/90}} = -1.5 \times \sqrt{\frac{10}{32}} \notin \mathcal{R}.$$

Consecuentemente, la muestra extraída no proporciona evidencia estadística como para rechazar H_0 y mantendremos la afirmación de que la media de la población es 3.5.

□

Problema 1. Segunda Semana. Con el fin de describir un conjunto de 10 observaciones del par de variables (X, Y) , se calcularon la recta de regresión de Y sobre X y la correspondiente varianza residual, resultando:

$$Y = 23.5 - 10.2X \quad ; \quad \text{Varianza residual} = 9,$$

respectivamente. Se pide:

- ¿Cuál es el signo de la covarianza? (Justifique la respuesta sin calcular dicho dato)
- Si la varianza residual representa el 19 % de la varianza de Y , ¿cuánto vale el coeficiente de correlación muestral?

Solución

- La relación entre la covarianza y la pendiente (*pend*) de la recta de regresión de Y sobre X se pone de manifiesto en la siguiente igualdad

$$\text{cov}_{x,y} = \text{pend} \times v_x$$