

Nombre: _____

Control 2 de ESTADÍSTICA APLICADA (25 de abril de 2019)
Primer curso del grado en Bioquímica

EXPLICAR DEBIDAMENTE LA RESOLUCIÓN DE CADA PROBLEMA

1. Consideramos los niveles de expresión de un cierto gen en 5 ratones normales y en otros 5 ratones mutantes:

| Normales | | | | | Mutantes | | | | |
|----------|------|------|------|------|----------|------|------|------|------|
| 0.40 | 0.52 | 0.83 | 1.45 | 0.85 | 0.05 | 0.76 | 0.31 | 0.15 | 0.53 |

Al analizar estos datos con R obtenemos lo siguiente:

Normales = c(0.40, 0.52, 0.83, 1.45, 0.85)

Mutantes = c(0.05, 0.76, 0.31, 0.15, 0.53)

mean(Normales)

[1] 0.810 = \bar{x}

mean(Mutantes)

[1] 0.360 = \bar{y}

var(Normales)

[1] 0.16595 = s_1^2

var(Mutantes)

[1] 0.08290 = s_2^2

$$m = n = 5$$

\bar{x} = nivel de expresión del gen en un ratón normal

\bar{y} = nivel de expresión del gen en un ratón mutante

a) (0.3 puntos) Calcula un intervalo de confianza al 95 % para la diferencia de los niveles esperados de expresión del gen entre los ratones normales y los mutantes. Especifica las suposiciones previas que garantizan la validez del procedimiento empleado.

b) (0.2 puntos) En base al intervalo de (a), con una confianza del 95 % ¿hay diferencia entre los dos grupos de ratones en cuanto al gen en cuestión? ¿Qué sucede si rebajamos la confianza al 90 %?

c) (0.2 puntos) Calcula un intervalo de confianza al 95 % para la varianza del nivel de expresión del gen en el grupo de ratones normales. Especifica las suposiciones previas que garantizan la validez del procedimiento empleado.

2. En una cierta población la probabilidad de que una mujer celíaca adulta tenga hijos celíacos es de un 25 %. Se planea tomar una muestra aleatoria de 90 mujeres celíacas en esa población.

a) (0.1 puntos) ¿Cuál es la distribución de probabilidad de la proporción de ellas que tendrá hijos celíacos?

b) (0.2 puntos) ¿Con qué probabilidad habrá más de un 30 % de esas madres muestreadas que tengan progeñe celíaca?

B

① a) Suponemos que $X \sim N(\mu_1, \sigma)$ e $Y \sim N(\mu_2, \sigma)$ son independientes.

$$s_p^2 = \frac{4 \cdot 0.16595 + 4 \cdot 0.08290}{8} = 0.124425$$

$$t_{8;0.025} = 2.306$$

$$IC_{95\%}(\mu_1 - \mu_2) = (0.810 - 0.360 \pm 2.306 \cdot \sqrt{0.124425} \sqrt{\frac{1}{5} + \frac{1}{5}}) = (0.45 \pm 0.514)$$

b) Como $0 \in IC_{95\%}(\mu_1 - \mu_2)$, no hay diferencias significativas (al 95% de confianza) entre μ_1 y μ_2 .

Para un 90% de confianza la semiamplitud del intervalo es:

$$1.860 \sqrt{0.124425} \sqrt{\frac{1}{5} + \frac{1}{5}} = 0.415$$

luego $0 \notin IC_{90\%}(\mu_1 - \mu_2)$. Al 90% de confianza sí habría diferencia entre los niveles de expresión esperados del gen en ratones normales y mutantes.

c) Suponemos $X \sim N(\mu_1, \sigma)$. Usamos $\chi_{4;0.025}^2 = 11.14$ y $\chi_{4;0.975}^2 = 0.484$.

$$IC_{95\%}(\sigma^2) = \left(\frac{4 \cdot 0.16595}{11.14}, \frac{4 \cdot 0.16595}{0.484} \right) = (0.0596, 1.371)$$

② $X = \begin{cases} 1 & \text{si una mujer celíaca adulta tiene hijos celíacos} \\ 0 & \text{si no} \end{cases} \sim \text{Bernoulli}(p)$
con $p = 0.25$

$$n = 90 \rightarrow X_1, \dots, X_{90}$$

a) $\bar{X} = \frac{\sum_{i=1}^{90} X_i}{90} = \hat{p}$ = proporción muestral de mujeres con hijos celíacos

$$\sim N(p = 0.25, \sqrt{\frac{0.25 \cdot 0.75}{90}} = 0.0456)$$

aprox por el TCL ($n = 90$ grande)

$$b) P\{\bar{X} > 0.3\} = P\left\{ \frac{\bar{X} - 0.25}{0.0456} > \frac{0.3 - 0.25}{0.0456} \right\} \approx$$

$$\approx P\{Z > 1.09\} = 0.1379$$

esto es aprox $N(0,1)$ por el TCL

$$Z \sim N(0,1)$$