

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				
0.46	0.50	0.80	1.51	0.90	0.1	0.73	0.24	0.06	0.46

Al analizar estos datos con R obtenemos lo siguiente:

Normales = c(0.46, 0.50, 0.80, 1.51, 0.90)
 Mutantes = c(0.1, 0.73, 0.24, 0.06, 0.46)

```
mean(Normales)
[1] 0.834 =  $\bar{x}$ 
mean(Mutantes)
[1] 0.318 =  $\bar{y}$ 
var(Normales)
[1] 0.17848 =  $s_1^2$ 
var(Mutantes)
[1] 0.07752 =  $s_2^2$ 
```

$\bar{x}$ = 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 un varón adulto diagnosticado de TDAH (Trastorno por Deficit de Atención e Hiperactividad) en su juventud tenga hijos con TDAH es de un tercio. Se planea tomar una muestra aleatoria de 100 padres con TDAH en esa población.

- a) (0.1 puntos)** ¿Cuál es la distribución de probabilidad de la proporción de ellos que tendrá hijos con ese trastorno?
- b) (0.2 puntos)** ¿Con qué probabilidad habrá más de un 40 % de esos padres que tengan progenie con TDAH?

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

$$s_p^2 = \frac{4 \cdot 0.17848 + 4 \cdot 0.07752}{8} = 0.128$$

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

$$IC_{95\%}(\mu_1 - \mu_2) = (0.834 - 0.318 \pm 2.306 \cdot \sqrt{0.128} \cdot \sqrt{\frac{1}{5} + \frac{1}{5}}) = (0.516 \pm 0.522)$$

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:

$$1.860 \sqrt{0.128} \sqrt{\frac{1}{5} + \frac{1}{5}} = 0.421$$

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

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

$$IC_{95\%}(\sigma^2) = \left(\frac{4 \cdot 0.17848}{11.14}, \frac{4 \cdot 0.17848}{0.484} \right) = (0.0641, 1.4750)$$

② $X = \begin{cases} 1 & \text{si un varón adulto TDAH tiene algún hijo TDAH} \\ 0 & \text{m " " " " no " ningún " "} \end{cases}$

$\sim \text{Bernoulli}(p)$ con $p = \frac{1}{3} \approx 0.3333$

$n = 100 \rightarrow X_1, \dots, X_{100}$

a) Proporción muestral de padres TDAH que tendrán hijos TDAH
 $= \hat{p} = \bar{X} = \frac{\sum_{i=1}^{100} X_i}{100} \sim N(p = 0.3333, \sqrt{\frac{0.3333 \cdot 0.6667}{100}} = 0.0471)$
 aprox por el TCL
 ($n = 100$ grande)

$$b) P\{\bar{X} > 0.4\} = P\left\{ \frac{\bar{X} - 0.3333}{0.0471} > \frac{0.4 - 0.3333}{0.0471} \right\} \approx P\{Z > 1.42\} = 0.0778$$

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