

Nombre: _____

Examen final de ESTADÍSTICA APLICADA (20 de mayo de 2019)
Primer curso del grado en Bioquímica

EXPLICAR DEBIDAMENTE LA RESOLUCIÓN DE CADA PROBLEMA

1. Pérez-Moreno *et al.* (2019)¹ estudiaron la adulteración y contaminación de resina de hachís obtenida en las calles de Madrid. Entre otros contaminantes, se examinó la presencia de *Escherichia coli* en 60 muestras, 43 de ellas con forma de bellota y 17 con forma de tableta. De dichas muestras, 40 bellotas y 5 tabletas estaban contaminadas con *E. coli*.

a) (0.75 puntos) Construye un intervalo de confianza al 95 % para la proporción p_1 de bellotas de hachís vendidas en la calle y contaminadas con *E. coli*.

b) (0.75 puntos) A un nivel de significación del 5 %, ¿hay suficiente evidencia muestral para afirmar que p_1 es superior al 75 %?

c) (0.5 puntos) Explica qué análisis realiza el siguiente código R y qué conclusiones se obtienen.

```
prop.test(40,43,p=0.75,conf.level=0.9)
```

```
1-sample proportions test with continuity correction
```

```
data: 40 out of 43, null probability 0.75
X-squared = 6.5194, df = 1, p-value = 0.01067
alternative hypothesis: true p is not equal to 0.75
90 percent confidence interval:
 0.8228321 0.9783592
sample estimates:
      p
0.9302326
```

d) (0.5 puntos) Una de las afirmaciones del estudio era la siguiente:

There was a strong association between sample shape and contamination by E. coli. Thus, 88.9 % of acorn samples were contaminated with E. coli while this percentage was only 11.1 % in ingot samples ($p < 0.0001$).

Esta afirmación se puede entender de la siguiente manera (olvidando la parte del comentario sobre los porcentajes):

(*) Dependiendo de la forma de la resina, hay mayor o menor contaminación con *E. coli*. El contraste para ver si hay suficiente evidencia muestral a favor de la hipótesis de que p_1 es mayor que p_2 (la proporción de tabletas de hachís vendidas en la calle y contaminadas con *E. coli*) tiene un p-valor inferior a 0.0001.

¿Te parece correcta la afirmación (*) o hay alguna objeción para llevar a cabo ese análisis? Razona tu respuesta.

e) (0.5 puntos) Especifica qué variables aleatorias estás empleando en este problema, su distribución de probabilidad y la estimación de máxima verosimilitud de sus parámetros.

¹Pérez-Moreno, Pérez-Lloret, González-Soriano y Santos-Álvarez (2019). Cannabis resin in the region of Madrid: Adulteration and contamination. *Forensic Science International*, 298, 34–38. Ver también https://elpais.com/elpais/2019/03/29/ciencia/1553875974_641867.html

2. Las focas de Weddell viven en un entorno de la Antártida. Se alimentan de pescado durante inmersiones largas y profundas en agua a muy bajas temperaturas. Aunque estas inmersiones benefician a las focas a nivel nutricional, tienen un coste metabólico. Unos investigadores intentaron averiguar si el coste metabólico de una inmersión (medido, en $\text{ml O}_2 \text{ kg}^{-1}$, como el consumo de oxígeno de la foca cuando emerge para respirar tras el buceo) es diferente en una inmersión para alimentarse y en una inmersión no alimentaria. Escogieron 10 focas Weddell en cada una de las cuales midieron el coste metabólico en dos inmersiones, una alimentaria y otra no. Los datos resultantes aparecen en la Tabla 1.

Foca	No alimentaria	Alimentaria	Foca	No alimentaria	Alimentaria
1	42.2	71.0	6	82.0	112.8
2	51.7	77.3	7	81.3	121.2
3	59.8	82.6	8	81.3	126.4
4	66.5	96.1	9	96.0	127.5
5	81.9	106.6	10	104.1	143.1

Tabla 1

Al analizar los datos con R obtenemos:

```
Datos = read.table("FocasWeddell.txt")
Datos
  V1    V2    V3
1 42.2  71.0
2 51.7  77.3
3 59.8  82.6
4 66.5  96.1
5 81.9 106.6
6 82.0 112.8
7 81.3 121.2
8 81.3 126.4
9 96.0 127.5
10 104.1 143.1

X = Datos$V2
Y = Datos$V3
mean(X)
[1] 74.68
mean(Y)
[1] 106.46
var(X)
[1] 376.0884
var(Y)
[1] 580.5116
cov(X,Y)
[1] 451.6824
```

a) (1 punto) Determina un intervalo de confianza al 99% para la diferencia de los costes metabólicos medios entre una inmersión no alimentaria y otra alimentaria. Especificar las suposiciones previas que garantizan la validez del procedimiento empleado.

b) (1 punto) A un nivel de significación del 10%, ¿hay evidencia a favor de la hipótesis de que la inmersión alimentaria causa en promedio un mayor coste metabólico que una no alimentaria? ¿Y a niveles de significación del 5% y del 1%? ¿Qué puedes decir del p-valor del contraste?

c) (0.5 puntos) Utilizando las salidas de R del enunciado y especificando claramente las fórmulas empleadas, calcula el coeficiente de correlación entre los costes metabólicos en una inmersión alimentaria y una no alimentaria.

d) (0.5 puntos) Utilizando las salidas de R del enunciado y especificando claramente las fórmulas empleadas, determina la recta de regresión del coste metabólico en una inmersión alimentaria sobre aquél de una no alimentaria.

e) (0.5 puntos) En el recuadro de la Figura 1, representa un diagrama de dispersión de la muestra y superpón la recta de regresión calculada en **d)**. Utilizando el gráfico y la correlación calculada en **c)**, evalúa el grado de relación lineal entre las dos variables.

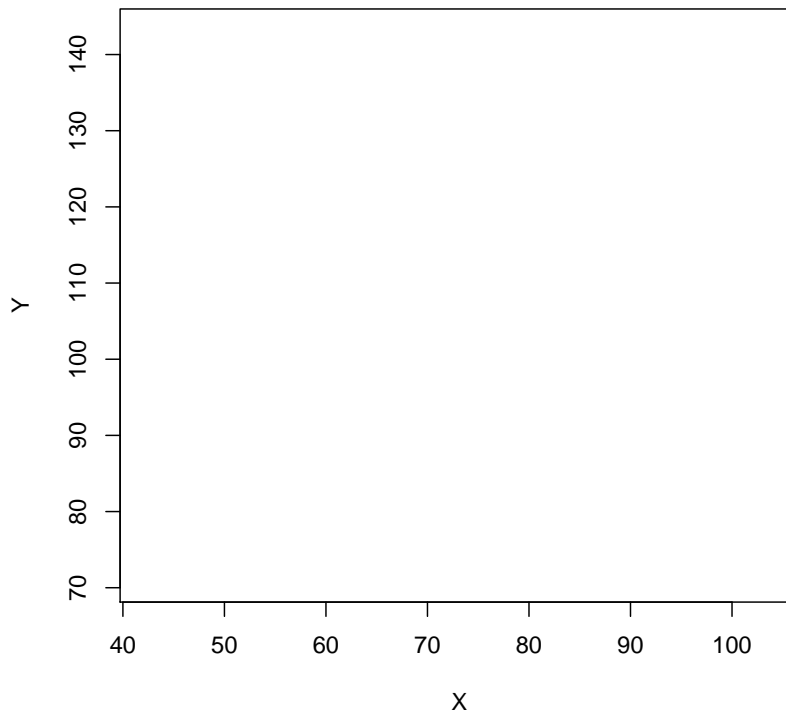


Figura 1

f) (0.5 puntos) Al ejecutar el código R

```
boxplot(X,Y,names=c("No alimentaria","Alimentaria"))
```

se obtiene la Figura 2. Interpreta este gráfico y di si es coherente con lo obtenido en el apartado **b)**.

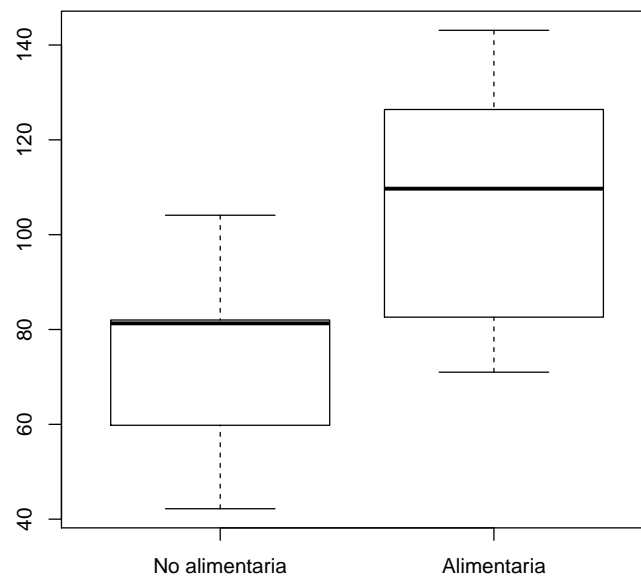


Figura 2

3. Cuando se recoge una muestra de agua de un cierto pozo, la medida X de la cantidad de oxígeno en disolución (en mg/L) sigue una distribución normal de media 5.6 y desviación típica 2.3.

a) (0.5 puntos) Calcula la probabilidad de que X esté entre 3 y 8.

b) (1 punto) Se toman tres medidas, X_1 , X_2 y X_3 , del oxígeno en disolución en el pozo. Calcula la probabilidad de que la media $\bar{X} = (X_1 + X_2 + X_3)/3$ de las tres medidas esté entre 3 y 8.

c) (0.5 puntos) Calcula el valor del primer cuartil de la distribución de X , es decir, aquel valor Q_1 tal que el 25 % de las medidas del oxígeno en disolución son menores que Q_1 .

d) (1 punto) Se toman diez medidas de la concentración de oxígeno. Calcula la probabilidad de que como mucho siete de ellas estén entre 3 y 8 mg/L.