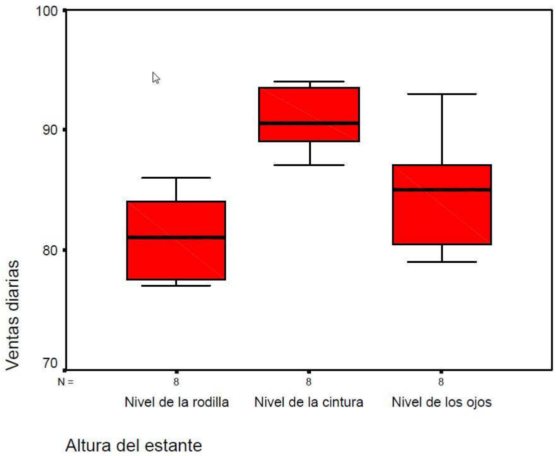
Prueba de Estadística

Apelidos: **Nome:** **DNI:**

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas[1].

Pregunta 1	a	b	c	d
Pregunta 2	a	b	c	d
Pregunta 3	a	b	c	d

1. *(1 punto)* En el siguiente diagrama de cajas se representa los diagramas de caja del número de productos vendidos diariamente en función de altura de colocación en los expositores en un determinado supermercado. ¿Qué afirmaciones son correctas?:



   a) Las ventas diarias a nivel de la cintura siempre son mayores que las ventas diarias a nivel de la rodilla y a nivel de los ojos.

   b) Al menos el 50 % de las ventas diarias a nivel de los ojos superan el 75 % de las ventas diarias a nivel de la rodilla.

   c) La venta media diaria a nivel de la cintura es superior a la venta media diaria a nivel de la rodilla.

   d) Ninguna de las anteriores

   solución correcta: b,c)

2. *(1 punto)* Un sistema de computación en-línea tiene 4 líneas de entrada. Cada línea cubre un porcentaje de tráfico de entrada y cada línea tiene un % de mensajes con errores. La tabla siguiente describe estos porcentajes.

Línea	% mens. por línea	% mens. sin error
1	40	99.8
2	30	99.9
3	10	99.7
4	20	99.2

   Si un mensaje se recibió erróneamente, ¿cuál es la probabilidad de que entrara por la línea 1?

   a) $\frac{4}{15}$ prob. de bayes = (2*0.4)/(2*0.4+1*0.3+3*.1+8*0.2)=8/3

   b) 0.40

   c) $\frac{1}{4}$

   d) Ninguna de las anteriores.

3. *(1 puntos)* Sea X una variable con distribución $N(0,2)$. Calcular la mediana de la variable $Y = |X|$ (redondear a 3 decimales).

   a) 0

   b) 0.674

   c) 1.349

   d) Ninguno de los anteriores

   solución correcta c): por simetría a mediana é o 3 cuartil da N(0,2): qnorm(0.75, mean=0,sd=3)

4. *(1 punto)* Un servicio telefónico de urgencias recibe por término medio 10 llamadas cada minuto según una distribución de Poisson, ¿cuál es la probabilidad de recibir más de 550 llamadas en una hora?(Se supone que las llamadas son independientes)

   a) 0.979.

   b) 0.417.

   c) 1.0

   d) Ninguna de las anteriores.

   solución correcta: a)

Solución: Sea X_t = número de mensajes en el intervalo $(0,t)minutos \sim Pois(10t)$, se pide que:

$$P(X_{60} > 550) = 1 - P(X_{60} \leq 550) =$$
$$1 - ppois(550, lambda = 600) = 0.9794596$$

□

[1] 1 respuesta incorrecta penaliza de tal forma que 3 respuestas incorrectas equivale a una respuesta correcta. Las preguntas en blanco no penalizan. Es necesario justificar la opción marcada, usar la parte trasera del enunciado.

1. *(2.5 puntos)* El archivo de datos en formato R *HardWater.RData*[2] contiene información sobre mortalidad en 61 ciudades de Inglaterra y Gales durante los años 1958-1964 y la concentración de Calcio, concretamente las variables contenidas en el data.frame *HardWater* son:

   - *location*: variable indicadora si la ciudad está tan al norte como la ciudad de Derby,
   - *town*: nombre de la ciudad,
   - *mortality*: mortalidad anual promedio por en tantos por cien mil,
   - *hardness*: concentración de calcio (en partes por millón).

   a) ¿Qué tipo de variables estadísticas son?. Justifica la respuesta.

   b) Describe completamente la variable *location*. Da una representación gráfica adecuada. Interpreta los resultados.

   c) Agrupar la variable *hardness* en los subintervalos $((0, 20], (20, 40], (40, 60], (60, +\infty])$. Da su distribución de frecuencias completa. Dar la representación gráfica más adecuada manteniendo esos intervalos.

   d) Resume numéricamente la variable *hardness* y la variable *hardness* agrupada. Interpreta todos los resultados que se indiquen.

2. (*3.5 puntos*) Si la probabilidad de que un virus infecte un ordenador es 0.0001, ¿cuál es la probabilidad de infectar 4 o más ordenadores en 50000 ataques? Da el resultado exacto y usando la correspondiente aproximación. Se supone independencia. Una mejora de la infección del virus da una probabilidad de 0.01, ¿cuantos ataques, como mínimo, tienen que realizarse para que la probabilidad de infectar a 4 o más ordeandores sea inferior al 40 %? En este último apartado suponer aproximación por distribución normal.

   Solución: Sea X = número de ordenadores infectados $\sim Bi(n = 50000, p = 0.0001)$ Aproximación teórica:

$$P(X \geq 4) = 1 - P(X < 4) = 1 - P(X <= 3)$$
$$1 - pbinom(3, size = 50000, p = 0.0001) = 0.7349881$$

   Aproximación por poisson dado que $n > 20$ e $p < 0.05$. Sea $\hat{X} \sim Pois(np = 5)$

$$P(X \geq 4) \approx 1 - P(\hat{X} <= 3) = 1 - ppois(3, lambda = 5) = 0.7349741$$

   La aproximación por la distribución normal no es procedente, dado que el npq<5

   Ahora la variable es $X \sim Bi(n, 0.01)$ y la condición indica que n debe verificar que

$$P(X \geq 4) < 0.4$$
$$P(X < 4) \geq 0.6$$

   Usando la aproximación por la normal obtenemos que puedo aproximar X por $\tilde{X} \sim N(0.01 * n, \sqrt{0.0099 * n})$. Tipificando la ecuación anterior obtenemos que:

$$P\left(\tilde{X} < 4\right) = P\left(\frac{\tilde{X} - 0.01n}{\sqrt{0.0099 * n}} < \frac{4 - 0.5 - 0.01n}{\sqrt{0.0099 * n}}\right) \geq 0.6$$

   donde ahora la varible tipificada es N(0,1) y por lo tanto el valor pedido es su cuantil 0.4, es decir, *qnorm(0.4, mean=0,sd=1)=-0.2533471*. Entonces

$$\frac{3.5 - 0.01n}{\sqrt{0.0099 * n}} \geq -0.2533471$$
$$3.5 - 0.01n \geq -0.2533471 * \sqrt{0.0099 * n}$$

[2] El archivo de datos *HardWater.RData* se puede descargar desde la url *https://dl.dropboxusercontent.com/u/29008031/HardWater.RData*

Sólo resta resolver la inecuación:

$$
\begin{aligned}
3.5 - 0.01n &= -0.2533471 * \sqrt{0.0099 * n} \\
(3.5 - 0.01n)^2 &= 0.2533471^2 * 0.0099 * n \\
12.25 - 2 * 3.5 * 0.01n + 10^{-4}n^2 &= 0.2533471^2 * 0.0099 * n \\
10^{-4}n^2 - \left(0.07 + 0.2533471^2 * 0.0099\right) * n + 12.25 &= 0 \\
10^{-4}n^2 - 0.07063543n + 12.25 &= 0 \\
n^2 - 706.3543n + 122500 &= 0
\end{aligned}
$$

Soluciones de la ecuación de 2º grado dadas por: n= 305.35 (raíz negativa) y n = 400.44 (raíz positiva). Realizando las comprobaciones se obtiene que para que se verifique la probabilidad de inferior a 0.4 debe ser $n \geq 400.44$.

$pnorm(3.5, mean = 400.44 * 0.01, sd = sqrt(400.44 * 0.01 * 0.99))$

$\square$