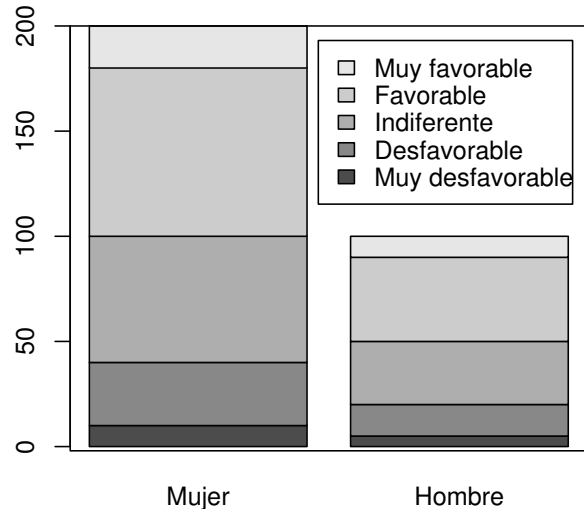


Atención: Contesta en la hoja de respuestas, pero entrega todos los cálculos. Razona siempre las respuestas

1. (0.5pt) Una encuesta sobre el grado de satisfacción de los clientes por un servicio (valorado desde "Muy desfavorable" hasta "Muy favorable") analiza las respuestas conjuntamente con el sexo (hombre o mujer) del cliente. Según el gráfico resultante:

¿Influye el sexo en la distribución del nivel de satisfacción de los clientes?



(Responde y justifica la respuesta)

2. (0.75pt) Compara la dispersión de las dos muestras presentadas (es decir, **di cuál es más homogénea**) teniendo en cuenta la diferencia de magnitud de los datos de cada una.

Muestra A	5.46	7.75	4.89	6.24	7.90	5.99	5.77	4.79
Muestra B	241.03	264.66	221.67	255.64	230.36	233.25	265.83	297.17

3. (1.25pt) Una muestra, recogida para un estudio sobre el descenso de la concentración de una sustancia en la sangre con el paso del tiempo, da lugar a los datos que se presentan en la tabla:

Tiempo (h)	1	2	3	4	5	6
Concentración (mg/l)	122.80	100.54	82.32	67.39	55.18	45.17

Realiza predicciones de la concentración que habrá a las 3 horas y media de haber consumido la sustancia, y a las 10 horas, **y razona sobre la calidad** de las predicciones realizadas.

4. (0.5pt) Para valorar la duración de un tipo de bombillas se conectaron de forma simultánea 7 bombillas a la corriente. A fecha de hoy se nos pide dar un valor global que resuma la duración de las bombillas de este tipo.

El hecho es que 6 de ellas ya se han fundido y **queda una todavía funcionando**. Las duraciones de las bombillas fundidas (en meses) han sido:

7.2 13.8 15.3 21.7 36.6 41.9

Da un valor **que tenga en cuenta la muestra completa** y **justifica** la conveniencia de usar ese valor frente a otros posibles.

5. (0.75pt) Una prueba tipo test consta de 10 preguntas con 4 alternativas de solución por pregunta. Una persona contesta al azar todas las preguntas. Calcula:
- a) (0.25pt) Número esperado de preguntas correctas.
 - b) (0.5pt) Probabilidad de aprobar la prueba (es decir, de obtener 5 ó más respuestas correctas).
-
6. (1pt) Un vendedor (que vende a través de llamadas telefónicas) estima que la probabilidad de que una llamada acabe siendo una venta es de 0.05. Su objetivo diario es realizar 3 ventas, y termina su jornada de trabajo, sea la hora que sea. Calcula:
- a) (0.5pt) Número medio (a la larga) de llamadas realizadas por día.
 - b) (0.5pt) Probabilidad de acabar la jornada laboral en exactamente 50 llamadas.
-
7. (0.75pt) Una oposición consta de un examen escrito en el que se desarrolla una pregunta, de una lista de 71 preguntas posibles que forman el temario.
- Al llegar al examen, el opositor debe escoger 3 bolas (numeradas del 1 al 71) de una urna, y su examen consistirá en una de las 3 preguntas indicadas por las bolas (la que quiera).
- Suponiendo que el opositor sabe perfectamente 30 de las 71 preguntas, y desconoce absolutamente las otras 41 preguntas ¿qué probabilidad exacta tiene de poder responder al examen?
-
8. (1pt) Una empresa lleva registro de las incidencias ocurridas diariamente. Atendiendo sólo al número de incidencias diarias se comprueba que ha habido un promedio de 3.71 incidencias diarias en la historia de la empresa. Calcula:
- a) (0.5pt) La probabilidad de que el día siguiente no haya ninguna incidencia.
 - b) (0.5pt) Si la fábrica lleva 2327 días en funcionamiento, ¿en cuántos de ellos (aprox.) no se produjo ningún incidente?
-
9. (1.5pt) Un proveedor de máquinas ofrece dos modelos de máquina (M_1 y M_2) a sus clientes, que las precisan para fabricar piezas de una longitud teórica de 150 mm.
- El cliente C_1 admite como buenas las piezas con longitud entre 148.5 y 151.5 mm, mientras que el cliente C_2 es más exigente y sólo admite como buenas las piezas cuya longitud esté entre 148.8 y 151.2 mm.
- Un estudio realizado sobre la calidad de las máquinas revela que la máquina M_1 hace piezas con longitud completamente arbitraria dentro del intervalo $[148.3, 151.7]$. La máquina M_2 realiza sus piezas con longitud normalmente distribuida de media 150 mm y desviación típica 1.0 mm.
- Haz los cálculos necesarios para que cada cliente escoja la máquina que más le convenga (puede ser la misma o diferente).
-
10. (0.5pt) Un partido político quiere encargar una encuesta para estimar la proporción de votantes entre el electorado en las próximas elecciones. Calcula el tamaño mínimo que debe tener la muestra para que la estimación (por intervalo de confianza al 95 %) tenga un error máximo de 0.01.



1. Sí / No. Razón: _____

2. La más homogénea es la muestra ____ porque _____

3. Predicción para 3 horas y media: _____
Predicción para 10 horas: _____
Razonamiento calidad predicciones: _____

4. Valor: _____
Razonamiento: _____

5. (a) _____
(b) _____
6. (a) _____
(b) _____
7. Resultado: _____
8. (a) _____
(b) _____
9. Cliente C_1 prefiere máquina ☐ porque _____

Cliente C_2 prefiere máquina ☐ porque _____

10. Resultado: _____
Razonamiento de por qué no hace falta conocer \hat{p} : _____

1. Práctica 1

Escribe el código R que:

(1) Calcula el valor $\frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{100000}}$

(2) Genera un vector sexo cuyas componentes son las cadenas 'Mujer' (repetida 200 veces) y 'Hombre' (repetida 100 veces).

(3) Genera dos vectores opinamujer y opinahombre cuyas componentes son las cadenas 'En contra', 'Indiferente' y 'A favor', de manera que se repiten: en opinamujer, 13, 103 y 85 veces cada una, y en opinahombre, 22, 73 y 5 veces cada una, respectivamente.

(4) Genera el vector opina uniendo los vectores opinamujer y opinahombre.

(5) Genera la hoja de datos muestra con dos variables sexo y opinion, de manera que tengan, como datos, los datos de los vectores sexo y opina de antes.

2. Práctica 2

(1) Genera el diagrama de barras de la muestra bivalente muestra del ejercicio anterior (Práctica 1-(5), en esta hoja).

(2) Si los vectores tiempo y concentracion tienen los datos de dichas variables (mirar el ejercicio del examen escrito que corresponde a estas variables), escribe los comandos R que permiten realizar la recta de regresión de la concentración en función del tiempo.

(3) Escribe los comandos R que dibujan en un mismo gráfico los datos y la recta de regresión (suponiendo que la ecuación de la recta de regresión del apartado anterior fuera, por ejemplo, concentracion = 140 - 20*tiempo).

3. Práctica 3

Escribe el código R que resuelve los problemas del examen escrito del 5 al 9

4. Práctica 4

Programa la función int.conf.media que calcula el intervalo de confianza sobre la media poblacional, con las siguientes condiciones:

- El valor que devuelva la función sea un vector con dos componentes (la primera el extremo inferior del intervalo y la segunda el extremo superior).
- Los argumentos que pida la función para hacer el cálculo del intervalo sean:
 - El nivel de confianza nconf, que tendrá un valor por defecto de 0.95.
 - La variable poblacional vpobl que quedará NULL si no es conocida por el usuario, o bien tendrá el valor que éste le dé.
 - El argumento esnormal con el que el usuario indique si la muestra viene de una población normal o no.
 - Un vector datos que tenga los datos de la muestra, si es que el usuario dispone de ellos (o que quede como NULL si no los tiene).
 - Argumentos mediamuestral, varianzamuestral y tamanyomuestral por si el usuario tiene ya calculados dichos estadísticos y no dispone de la muestra de datos (y que queden como NULL si el usuario no los tiene).
- La función debe mostrar por pantalla un mensaje de aviso, anunciando que el intervalo de confianza no es fiable en el caso de que los datos **no** vengan de una distribución normal ni sean más de 30.

1. **No . Razón:** La distribución del nivel de satisfacción de las mujeres es la misma que la de los hombres, puesto que las barras de hombres y mujeres son exactamente proporcionales (todo es el doble en las mujeres). Por tanto NO influye el sexo en la distribución del nivel de satisfacción.
2. **La más homogénea es la muestra B porque** tiene menor CV. Calculando se tiene que $CV_A = 0.1926198$ y $CV_B = 0.09821318$.
3. **Predicción para 3 horas y media:** 78.92 mg/l
Predicción para 10 horas: -21.18 mg/l
Razonamiento calidad predicciones: La recta de regresión es $C = 132.82 - 15.40T$ (C es concentración y T temperatura). Los coeficientes $r = -0.9896474$ y $R^2 = 0.979402$ indican un grado muy alto de correlación lineal. Por tanto la nube de puntos está muy cerca de la recta de regresión. La calidad para $T = 3.5$ es muy alta (muy buena), pero la calidad para $T = 10$ ya no se sabe pues $T = 10$ está lejos de los datos, y no se sabe qué tendencia pueden seguir los datos.
4. **Valor:** 21.7
Razonamiento: Para resumir la muestra se puede usar la media, mediana y/o moda. Dado que falta un dato (pero sabemos que es el mayor de todos) sólo podemos calcular la mediana, ordenandolos de menor a mayor, y como hay 7 datos, la mediana es el cuarto dato.
5. $X = \text{"Num. preguntas acertadas"} \sim \text{Bin}(n = 10, p = 1/4 = 0.25)$
(a) $E(X) = np = 2.5$
(b) $P(X \geq 5) = 1 - F(4) = 1 - 0.9219 = 0.0781$
6. $X = \text{"Num. llamadas hasta acabar jornada"} \sim \text{BinNeg}(r = 3, p = 0.05)$
(a) $E(X) = r/p = 60$
(b) $P(X = 50) = f(50) = \binom{50-1}{3-1} 0.05^3 (1 - 0.05)^{50-3} = 0.01319249$
7. $X = \text{"Num. bolas buenas que extrae"} \sim \text{Hyper}(n = 3, N = 71, K = 30)$
Resultado: $P(X \geq 1) = 1 - f(0) = 1 - \frac{\binom{30}{0} \binom{71-30}{3-0}}{\binom{71}{3}} = 1 - 0.18651 = 0.81349$
8. $X = \text{"Num. incidencias de cada día"} \sim \text{Po}(\lambda = 3.71)$
(a) $P(X = 0) = e^{-3.71} \frac{3.71^0}{0!} = 0.024477$
(b) $2327 \cdot 0.024477 = 56.9592$
9. **Cliente C_1 prefiere máquina M_1 porque** hace un 88.2353 % de piezas buenas para él (la otra máquina hace un 86.6385 % de piezas buenas para él, que es inferior).
Cliente C_2 prefiere máquina M_2 porque porque hace un 76.9860 % de piezas buenas para él (la otra máquina hace un 70.5882 % de piezas buenas para él, que es inferior).
10. **Resultado:** $n \geq 38416$
Razonamiento de por qué no hace falta conocer \hat{p} :

$$\text{Error} = 2z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.01. \text{ Despejamos y sale } n \geq \left(\frac{2z_{1-\alpha/2}}{0.01} \right)^2 \hat{p}(1-\hat{p}).$$

Lo cierto es que, sea cual sea el valor de $\hat{p} \in [0, 1]$, podemos deducir que $\hat{p}(1-\hat{p}) \leq 0.25$.

Por tanto si $n \geq \left(\frac{2 \cdot 1.96}{0.01} \right)^2 \cdot 0.25 = 38416$ queda garantizado que el error máximo es menor que 0.01.