Estadística / 10 de enero de 2026 / José Carlos López Henestrosa

1) a) Para calcular la varianza poblacional manualmente, aplicamos la fórmula de la varianza poblacional:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}$$

Sabemos que $N = 10$, x_i es cada valor individual y μ la media poblacional, la cual desconocemos. Por lo tanto, la calculamos con su fórmula:

$$\boxed{\mu = \frac{\sum x_i}{N} = \frac{43+45+50+42+43+57+39+53+60+41}{10} = \frac{473}{10} = 47,3}$$

Ahora, calculamos las desviaciones cuadráticas $(x_i - \mu)^2$:

x_i	$x_i - \mu$	$(x_i - \mu)^2$
43	$43 - 47,3 = -4,3$	18,49
45	$45 - 47,3 = -2,3$	5,29
50	$50 - 47,3 = 2,7$	7,29
42	$42 - 47,3 = -5,3$	28,09
43	$43 - 47,3 = -4,3$	18,49
57	$57 - 47,3 = 9,7$	94,09
39	$39 - 47,3 = -8,3$	68,89
53	$53 - 47,3 = 5,7$	32,49
60	$60 - 47,3 = 12,7$	161,29
41	$41 - 47,3 = -6,3$	39,69

Calculamos la suma de las desviaciones al cuadrado:

$$\sum_{i=1}^{10} (x_i - \mu)^2 = 18,49 + 5,29 + 7,29 + 28,09 + 18,49 + 94,09 + 68,89 + 32,49 + 161,29 + 39,69 = \boxed{474,1}$$

Dividimos el resultado por el número total de datos (N):

$$\boxed{\sigma^2 = \frac{474,1}{10} = 47,41}$$

V

① b) # Definimos el vector de datos

```
tiempos <- c(43,45,50,42,43,57,39,53,60,41)

# Calculamos la desviación típica muestral con sd()
sd_muestral <- sd(tiempos)

# Obtenemos el tamaño de la muestra
n <- length(tiempos)

# Ajustamos para obtener la desviación típica poblacional
sd_poblacional <- sd_muestral * sqrt((n-1)/n)

# Mostramos el resultado redondeando a 4 decimales
round(sd_poblacional, 4)

[1] 6.8855
```

V

① c) Debemos utilizar el teorema de Chebyshev:

$$1 - \frac{1}{k^2}$$

El intervalo dado nos indica que k es 10, por lo que sustituimos k en la fórmula:

$$1 - \frac{1}{10^2} \Rightarrow 1 - \frac{1}{100} \Rightarrow 1 - 0{,}01 \Rightarrow 0{,}99$$

Por lo tanto, la proporción mínima de observaciones que han de pertenecer al intervalo $(\bar{x} - 10 s_x, \bar{x} + 10 s_x)$ es $\boxed{0{,}99}$.

V

① d) Utilizamos la fórmula de la tipificación. Como en los apartados a y b se pedían la varianza y desviación típica poblacional, voy a utilizar esos parámetros para mantener la coherencia:

$$z = \frac{x - \mu}{\sigma}$$

Sabemos lo siguiente:

- $x = 30$ (valor del nuevo usuario)
- $\mu = 47,3$ (media del apartado a)
- $\sigma = 6,885492$ (desviación típica poblacional).

Procedemos a los cálculos:

$$x - \mu = 30 - 47,3 = -17,3$$

$$z = \frac{-17,3}{6,885492} = -2,512529$$

V

Redondeando a 4 decimales: $\boxed{z = -2,5125}$

② a) Utilizamos el Teorema de la probabilidad total. Definimos los sucesos:

- A: $P(A) = 0,39$
- B: $P(B) = 0,26$
- C: $P(C) = 0,35$
- +: El resultado es positivo.

Las probabilidades condicionales son:

- $P(+|A) = 0,18$
- $P(+|B) = 0,2$
- $P(+|C) = 0,07$

Por lo tanto, la fórmula de la probabilidad total es:

$$P(+) = P(A) \cdot P(+|A) + P(B) \cdot P(+|B) + P(C) \cdot P(+|C)$$

Desglose de operaciones

- Método A: $0,39 \cdot 0,18 = 0,0702$
- Método B: $0,26 \cdot 0,2 = 0,052$
- Método C: $0,35 \cdot 0,07 = 0,0245$

Suma total:

$$\boxed{P(+)} = 0,0702 + 0,052 + 0,0245 = \boxed{0,1467}$$ V

② b) Utilizamos el Teorema de Bayes a partir de que sabemos que el resultado ha sido positivo: $P(C|+)$.

$$P(C|+) = \frac{P(C) \cdot P(+|C)}{P(+)}$$

$P(C) \cdot P(+|C) = 0,0245$ (calculado en el apartado a)
$P(+) = 0,1467$ (calculado en el apartado a)

$$P(C|+) = \frac{0,0245}{0,1467} = 0,167007\ldots$$

V

Redondeado a 4 decimales: $\boxed{P(C|+) = 0,1670}$

② c) Utilizamos el teorema de Bayes nuevamente, para aplicarlo al suceso negativo (-). Podremos calcular $P(A|-)$:

$$P(A|-) = \frac{P(A) \cdot P(-|A)}{P(-)}$$

Determinamos la probabilidad de negativos para el método A y la probabilidad total de negativos:

- $P(-|A)$: Como $P(+|A) = 0{,}18$, entonces $\boxed{P(-|A)} = 1 - 0{,}18 = \boxed{0{,}82}$
- $\boxed{P(A) \cdot P(-|A)} = 0{,}39 \cdot 0{,}82 = \boxed{0{,}3198}$
- $\boxed{P(-)} = 1 - P(+) = 1 - 0{,}1467 = \boxed{0{,}8533}$

Sustituimos en la fórmula:

$$P(A|-) = \frac{0{,}3198}{0{,}8533} = 0{,}374780\ldots$$

Redondeado a 4 decimales: $\boxed{P(A|-) = 0{,}3748}$ V

③ a) La variable aleatoria es X, que representa el tiempo de ejecución del algoritmo en milisegundos. La distribución es normal $X \sim N(\mu, \sigma)$.

Sabemos que $\sigma = 25{,}5$, pero desconocemos μ.

Nos dicen que la probabilidad de que el tiempo sea menor o igual que 450,5 es 0,19: $P(X \leq 450{,}5) = 0{,}19$

Para encontrar la media, tenemos que estandarizar la variable usando la fórmula $z = \frac{X - \mu}{\sigma}$ y buscar el valor z que acumule una probabilidad de 0,19 en una distribución normal estándar.

$$\frac{450,5 - \mu}{25,5} = z_{0,19} \Rightarrow \mu = 450,5 - (z_{0,19} \cdot 25,5)$$

Realizamos el cálculo con R utilizando la función qnorm() para encontrar el valor de $z_{0,19}$:

V

```
#Encontramos el valor z
z_val <- qnorm(0,19)

#Despejamos la media
sigma <- 25,5
x <- 450,5
mu <- x-(z_val*sigma)

#Mostramos valores
z_val                # -0,877896
mu                   # 472,886348
```

Podemos seguir con el cálculo de μ planteado anteriormente:

$$\boxed{\mu = 450,5 - (-0,877896 \cdot 25,5) = 450,5 - (-22,386348) = 472,886348}$$

Redondeado a 4 decimales: $\boxed{\mu = 472,8863}$ V

3b) La variable aleatoria es X, que es el número de fallos en las ejecuciones del algoritmo. La distribución es binomial, ya que los ensayos son independientes, con una probabilidad constante de éxito (que ocurra un fallo).

$X \sim B(n,p)$. $n=19$ (total ejecuciones) y $p=0{,}37$ (probabilidad de fallo)

$$P(x=9) = \binom{19}{9} \cdot 0{,}37^9 \cdot (1-0{,}37)^{19-9}$$

V

Lo calculamos en R con dbinom():

V

```
probabilidad <- dbinom(x=9, size=19, prob=0,37)
```

Esto da un resultado de 0,1182468... redondeado a 4 => $\boxed{0{,}118247}$

Desglose de los cálculos intermedios:

- Coeficiente binomial: $\frac{19!}{9!\,(19-9)!} = 92378$
- Término $\boxed{p^k}$ $(0{,}37^9) = \boxed{0{,}00013}$
- Término $\boxed{(1-p)^{n-k}}$ $(0{,}63^{10}) = \boxed{0{,}009849}$

V

- $\boxed{\text{Multiplicación}}$: $92378 \cdot 0{,}00013 \cdot 0{,}009849 = \boxed{0{,}118247}$

③ c) La variable aleatoria es X, que es el número de la ejecución en la que ocurre el primer fallo. La distribución es geométrica, ya que estamos buscando el número de pruebas hasta conseguir el primer éxito (algoritmo fallido).

$X \sim Geom(p)$ $p = 0{,}37$ (probabilidad de fallo).

Se pide la probabilidad de que la primera ejecución fallida sea después de la número 8: $P(X > 8)$

Para una distribución geométrica, la probabilidad de que el suceso ocurra después del ensayo k es igual a la probabilidad de no tener el suceso en los primeros k intentos:

$P(x) > k = (1-p)^k \Rightarrow P(X > 8) = (1-0{,}37)^8$

Cálculo en R con pgeom():

```
# Probabilidad de fallo
p <- 0,37
# Usamos la funcion pgeom()
prob <- pgeom(7, prob = p, lower.tail = FALSE)
# Mostramos resultado
prob
```

Cálculos intermedios:

$-(1-p) = 1-0{,}37 = 0{,}63$

$-\ 0{,}63^8 = 0{,}0248\ldots16$

Redondeado a 4 decimales: $\boxed{P(X>8) = 0{,}0248}$ V

(4) a) - $n = 282$ juegos "Strategy"

- $\bar{x} = 0{,}41$ millones de euros
- $s = 0{,}05$ millones de euros
- $(1-\alpha): 80\% = 0{,}80$

Dado que el tamaño de la muestra es $n > 30$, usamos la distribución normal. La fórmula del intervalo de confianza es:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$ X

Como se pide el límite inferior, tomamos el signo negativo.

Desglose de cálculos:

Sigma es desconocida por lo que hay que trabajar con la t de Student

- $z_{0,90} \approx 1{,}281552$ X
- $\boxed{\sqrt{n}} = \sqrt{282} = \boxed{16{,}792856}$
- $\boxed{\text{Error estándar}}\ \left(SE = \frac{s}{\sqrt{n}}\right) = \frac{0{,}05}{16{,}792856} = \boxed{0{,}002977}$ X
- Margen de error $(E = z_{\alpha/2} \cdot SE) = 1{,}281552 \cdot 0{,}002977 = \boxed{0{,}003816}$

Restamos el margen de error a la media muestral para obtener el límite inferior:

$\boxed{\text{inferior} = 0{,}41 - 0{,}003816 = 0{,}406184}$

Redondeado a 4 decimales: $\boxed{0{,}4062}$

(3) b) Calculamos el intervalo de confianza al 80% hallando el límite superior:

- $\bar{x} = 0,41$

Valoro razonamiento

- $E \approx 0,0038$
- Intervalo de confianza (80%): $(0,41 - 0,0038 \,,\, 0,41 + 0,0038)$
$(0,4062 \,,\, 0,4138)$

Observamos que 0,4119124 está dentro del intervalo, pero la respuesta es "No", porque el intervalo de confianza nos brinda un rango de valores posibles, pero no nos permite afirmar que la media poblacional sea exactamente un valor concreto dentro de este rango.

(4) c) - $n = 1689$
- $x = 282$
- Nivel de confianza: 80%.

Utilizamos la aproximación normal, ya que n es grande.

$$\text{Límite superior} = \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Desglose de cálculos:

- $\hat{p} = \frac{282}{1689} = 0,166963$
- $z_{0,80} \approx 1,281552$
- SE: $\hat{p}(1-\hat{p}) = 0,166963(1 - 0,166963) = 0,139086$

Dividimos por n: $\frac{0,139086}{1689} = 0,000082$

Raíz cuadrada: $\sqrt{0,000082} = 0,009075$

- Margen de error: $E_n = 1{,}281552 \cdot 0{,}009078 = \boxed{0{,}011630}$

Sumamos el margen de error a la proporción muestral:

$\boxed{\text{Superior}} = 0{,}166963 + 0{,}011630 = \boxed{0{,}178593}$

Redondeado a 4 decimales: $\boxed{0{,}1786}$ V

9 d)
- $L = 0{,}04$
- $E = 0{,}02$
- Nivel de confianza: $80\% \Rightarrow z_{\alpha/2} \approx 1{,}281552$
- $\boxed{\hat{p}} = \frac{282}{1689} \approx \boxed{0{,}166963}$

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow n = \frac{z^2_{\alpha/2} \cdot \hat{p}(1-\hat{p})}{E^2}$$

Desglose cálculos:

- $\boxed{z^2_{\alpha/2}} = 1{,}281552^2 = \boxed{1{,}642374}$
- $\boxed{\hat{p}(1-\hat{p})} = 0{,}166963 \cdot (1 - 0{,}166963) = \boxed{0{,}139086}$
- $\boxed{E^2} = 0{,}02^2 = \boxed{0{,}0004}$
- División (n): $\boxed{n} = \frac{0{,}228432}{0{,}0004} = \boxed{571{,}078863}$ V

El ejercicio pide el valor entero más próximo, el cual es $\boxed{572}$

5 a) - Hipótesis nula (H_0): $\mu_A \leq \mu_B$ (el tiempo medio de A es menor o igual al de B).

- Hipótesis alternativa (H_1): $\mu_A > \mu_B$
- Estadístico de contraste: t de Student con $n_A + n_B - 2$ grados de libertad, dado que asumimos varianzas poblacionales iguales pero desconocidas,

Datos:

- Algoritmo A: $n_A = 25$, $s_A = 143$
- " B: $n_B = 35$, $s_B = 129{,}8$
- Suposición: Varianzas iguales.

Cálculos: $$s_p^2 = \frac{(n_A - 1)\, s_A^2 + (n_B - 1)\, s_B^2}{n_A + n_B - 2}$$

Desglose:

Numerador

- $(25-1)\,143^2 = 24 \cdot 20449 = 490776$
- $(35-1)\,129{,}8^2 = 34 \cdot 16848{,}04 = 572833{,}36$
- Suma $= 1063609{,}36$

Denominador

- Grados de libertad: $25+35-2 = 58$

- $$\boxed{s_p^2} = \frac{1063609{,}36}{58} = \boxed{18338{,}092414}$$

$$\boxed{SE} = \sqrt{s_p^2 \left(\frac{1}{n_A} + \frac{1}{n_B}\right)} = \sqrt{1257{,}469194} = \boxed{35{,}460812}$$

Redondeado a 4 decimales: $\boxed{35{,}4608}$

5 b) - Hipótesis nula (H_0) : $\mu_A \leq \mu_B$
- " alternativa (H_1): $\mu_A > \mu_B$

- Distribución del estadístico: t de Student con $n_A + n_B - 2$ grados de libertad, ya que las varianzas poblacionales son desconocidas pero se asumen iguales y las muestras son independientes.

Cálculo ~~de~~ de los grados de libertad:

$$gl = n_A + n_B - 2 = 25 + 35 - 2 = 58$$

Obtenemos el valor crítico:

- $\alpha = 0,05$
- $gl = 58$
- Tipo de test: Cola derecha

Usando ~~sol~~ una tabla de distribución t de Student, hallamos que el valor es $\simeq 1,67155$. Redondeado a 4 decimales: $\boxed{1,6716}$ V

5 c) $$t = \frac{\bar{x}_A - \bar{x}_B}{S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

Sustituimos valores (s_p fue calculado en el apartado a):

$$t = \frac{533 - 434,6}{135,4182 \sqrt{\frac{1}{25} + \frac{1}{35}}} = \frac{98,4}{135,4182 \sqrt{0,04 + 0,02857}} =$$

$$= \frac{98,4}{135,4182 \cdot 0,26186} = \frac{98,4}{35,4607} \simeq 2,774902\ldots$$

Redondeado a 4 decimales: 2,7749 V

5 d) P-valor = $P(T_{58} > 2{,}7744)$

Para obtener el valor, consultamos una tabla estándar de la distribución t de Student con gl = 58:

- Para t = 2,663, el área (p) es 0,005.
- Para t = 2,920, el área (p) es 0,0025.

Como t = 2,7744 está dentro de estos dos puntos, sabemos que p-valor está entre 0,0025 y 0,005. Dado que 2,77 está aproximadamente en la mitad del intervalo, el valor será cercano al promedio geométrico de los extremos. Se calcula con F

X

Calculando el valor, sale que p-valor es ≈ 0,00366...

Redondeado a 4 decimales: $\boxed{0{,}0037}$

V

5 e) Como el p-valor es menor que el nivel de significación, rechazamos H_0.

6 a) Utilizamos la covarianza dividida por la varianza de la variable independiente (X). La pendiente b se calcula como $b = \frac{S_{xy}}{S_{xx}}$. Donde:

- $S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$
- $S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

$\boxed{S_{xy} = 373{,}5 - 307{,}730769}$
$= 65{,}769231$

X

Cálculo S_{xy}:

$\frac{(\sum x_i)(\sum y_i)}{n} = \frac{63 \cdot 127}{26} = \frac{8001}{26} \approx 307{,}730769$

Cálculo S_{xx}:

$\frac{(\sum x_i)^2}{n} = \frac{63^2}{26} = \frac{3969}{26} \approx 152{,}653846$

$\boxed{S_{xx} = 203 - 152{,}653846}$
$= 50{,}346154$ X

Cálculo de la pendiente: $\boxed{b = \frac{65,769231}{50,346154}} \approx \boxed{1,306341}$

Estas cantidades no son correctas

X

Redondeando a 4 decimales: $\boxed{1,3063}$

15 b) Utilizamos las medias muestrales de ambas variables y la pendiente obtenida en el apartado b.

La ecuación de la recta de regresión es $y = a + bx$. Despejando a: $a = \bar{y} - b\bar{x}$

Donde:

- $\bar{y}$ es la media de la variable dependiente (pacientes)
- $\bar{x}$ es la media de la variable independiente (horas de sueño)
- b es la pendiente.

$$\boxed{\bar{x} = \frac{\sum x_i}{n} = \frac{63}{26}} \approx \boxed{2,423077}$$

$$\boxed{\bar{y} = \frac{\sum y_i}{n} = \frac{127}{26}} \approx \boxed{4,884615}$$

Sustituimos en la fórmula: $(b \approx 1,306341)$:

$$\boxed{a = 4,884615 - (1,306341 \cdot 2,423077) = 1,719250}$$

Redondeamos a 4 decimales: $\boxed{1,7193}$

V

c) Se calcula elevando al cuadrado el coeficiente de correlación de Pearson:

$$R^2 = \frac{(S_{xy})^2}{S_{xx} \cdot S_{yy}}$$

Nos falta por calcular $S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$

$$\frac{(\sum y_i)^2}{n} = \frac{127^2}{26} = \frac{16129}{26} \approx 620,346154$$

$$S_{yy} = 772,5 - 620,346154 = 152,153846$$

Sustituimos valores en la fórmula:

$$\boxed{R^2} = \frac{(65,764231)^2}{50,346154 \cdot 152,153846} = \frac{4325,591744}{7660,361095} \approx 0,564671$$

Redondeado a 4 decimales: $\boxed{0,5647}$

Las cantidades no son corre

X