2. (2 puntos) La siguiente tabla muestra las calificaciones obtenidas por cinco personas al realizar dos pruebas: X (inteligencia oral) e Y (expresión oral).

X	92.5	77.5	100	107.5	122.5
Y	0.5	3.5	5	6.5	9.5

Sabiendo que $\sum XY = 2680$, $\sum X^2 = 51125$ y $\sum Y^2 = 170$. Resuelva razonando y explicando su respuesta:

a) ¿Cuál de estas variables, X o Y, tiene mayor variabilidad relativa? **(0.5 puntos)**

b) ¿Qué puede decir sobre una relación lineal entre X e Y? Si cree que existe una relación lineal, ¿qué tan fuerte es? **(0.5 puntos)**

c) ¿Cuál es la calificación de la prueba de expresión oral que podemos predecir para una persona que ha obtenido 120 puntos en la prueba de inteligencia oral? **(0.5 puntos)**

d) ¿Cuál es la proporción de variabilidad de Y que el modelo de regresión lineal entre X e Y no explica? **(0.5 puntos)**

NOTA: *Puede usar cualquier tipo de calculadora. De ser así, además de los resultados de la calculadora, debe indicar las fórmulas que está aplicando.*

a) coeficientes de variación

$CV_x = \frac{\sigma_x}{\bar{x}}$; $CV_y = \frac{\sigma_y}{\bar{y}}$

$\bar{x} = \frac{\sum x_i}{N} = \frac{500}{5} = 100$

$\bar{y} = \frac{\sum y_i}{N} = \frac{25}{5} = 5$

$\sigma_x = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2} =$

$\sigma_x = \sqrt{\frac{51125}{5} - 100^2} = 15$

$\sigma_y = \sqrt{\frac{\sum y_i^2}{N} - \bar{y}^2} = \sqrt{\frac{170}{5} - 5^2} = 3$

$CV_x = \frac{15}{100} = 0'15$; $CV_y = \frac{3}{5} = 0'6$

a mayor CV, mayor dispersión relativa a la media y tiene mayor dispersión relativa la que significa que su media es menos representativa.

b) $r_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$; $\sigma_{xy} = cov(x,y) = \frac{\sum x_i y_i}{N} - \bar{x}\,\bar{y} = \frac{2680}{5} - 100 \cdot 5 = 36$

$r_{xy} = \frac{36}{15 \cdot 3} = 0'8$ → correlación fuerte directa entre X e Y.

Se espera que aumentos en X se relacionen con aumentos de Y y viceversa.

c) $\boxed{\hat{y} = a + bx}$; $b = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{36}{15^2} = 0'16$

$a = \bar{y} - b\bar{x} = 5 - 0'16 \cdot 100 = -11$

$\hat{y} = -11 + 0'16x$

$\boxed{\hat{y}(x=120) = -11 + 0'16 \cdot 120 = 8'2}$

d) ¿Cuál es la proporción de variabilidad de Y que el modelo de regresión lineal entre X e Y no explica? **(0.5 puntos)**

R^2 = coeficiente de determinación es lo que el modelo sí explica

$1 - R^2$ = proporción que el modelo NO explica.

$1 - 0'8^2 = 1 - 0'64 = 0'36$ → un 36% no es explicado por el modelo

3. (2 puntos) Hay 60 novelas y 20 libros de poesía en una estantería. Una persona (A) elige un libro al azar y se lo lleva. Después de eso, otra persona (B) hace lo mismo.

a) ¿Cuál es la probabilidad de que el libro seleccionado por la persona B sea una novela? **(0.75 puntos)**

b) Sabiendo que la persona B ha elegido una novela, ¿cuál es la probabilidad de que el seleccionado por la persona A sea un libro de poesía? **(0.75 puntos)**

c) Suponiendo que hay 10 lectores en lugar de dos, y cada uno de ellos se lleva un libro, al azar. ¿Podríamos aplicar un modelo binomial B(n, p) para calcular la probabilidad del suceso "6 de estos 10 libros que se han llevado son libros de poesía"? Justifique su respuesta. De ser así, indique los valores de los parámetros n y p. **(0.5 puntos)**

a)

N_A: $P(N) = \frac{60}{80} = \frac{3}{4} = 0'75$

P_A: $P(P) = \frac{20}{80} = \frac{1}{4} = 0'25$

N_B: $P(N|N) = \frac{59}{79}$; P_B: $P(P|N) = \frac{2}{79}$

N_B: $P(N|P) = \frac{60}{79}$; P_B: $P(P|P) = \frac{19}{79}$

$\boxed{P(N_B) = 0'75 \cdot \frac{59}{79} + 0'25 \cdot \frac{60}{79} = 0'75}$ → P. Total

b) $\boxed{P(P_A | N_B)} = \frac{P(P_A)\,P(N_B|P_A)}{P(N_B)} = \frac{1/4 \cdot 60/79}{0'75} = \boxed{\frac{20}{79}}$

c) ¿10 lectores seguirían Binomial (n,p)?

Al no tener una probabilidad de éxito estable no puede ser binomial

1.- (3 puntos) Rellene 8 respuestas de las siguientes 10 preguntas en la tabla de la parte inferior. Solo hay una respuesta correcta para cada pregunta **(0.375 puntos x8, sin penalización por respuestas incorrectas)**

i. La siguiente tabla recoge el número de respuestas correctas. ¿Cuál es el percentil correspondiente a una persona con 65 respuestas correctas? **a) 82.5**; b) 0; c) 80; d) 65.

(*En la tabla: Class = Clase, x_i = marca de clase, f_i = frecuencia absoluta, F_i = frecuencia acumulada*)

Class	x_i	f_i	F_i
[0,10)	5	40	40
[10,20)	15	60	100
[20,30)	25	75	175
[30,40)	35	90	265
[40,50)	45	105	370
[50,60)	55	85	455
[60,70)	65	80	535
[70,80)	75	65	600
		600	

$P_n = 65 = Li + \frac{\frac{N}{100} n - F_{i-1}}{f_i} \cdot a_i$

$65 = 60 + \frac{\frac{600}{100} n - 455}{80} \cdot 10$ ¿n?

$5 = \frac{6n - 455}{8}$; $40 + 455 = 6n$; $n = \frac{495}{6} = 82'5$

entre el P82 y el P83

ii. En un modelo de regresión lineal...

- a) Si la media de los residuos es cero, entonces las predicciones son fiables;
- **b)** El hecho de que la media de los residuos sea cero no implica que las predicciones sean fiables;
- c) La media de los residuos no se puede calcular;
- d) Lo importante es que el término medio de los residuos tome el valor cero.

iii. Al sacar una carta de una baraja de 40 cartas, determine la probabilidad de que sea un caballo o de espadas.
a) 0.325; b) 0.35; c) 0.5; d) 0.4.

$P(C \cup E) = P(C) + P(E) - P(C \cap E) = \frac{4}{40} + \frac{10}{40} - \frac{1}{40} = \frac{13}{40} = 0'325$

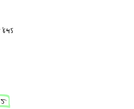
iv. Los resultados de varias pruebas de inteligencia siguen una distribución normal con media 100 y desviación típica 15. ¿Cuál es el intervalo, cuyo valor central es 100, que contiene al 50% de la población? a) $[80, 120]$; **b)** Un intervalo como ese no existe; c) $[100, +\infty)$; **d) $[90, 110]$**.

$P(\mu - \sigma < x < \mu + \sigma) = 68\%$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 95\%$

$P(\mu - 3\sigma < x < \mu + 3\sigma) = 99'7\%$

$[100 - 15, 100 + 15] = [85, 115]$



v. En una muestra de 200 individuos hay 60 graduados universitarios y 120 mujeres. Considerando que el muestreo aleatorio está bien realizado, ¿cuál es la afirmación más rigurosa sobre la muestra representativa obtenida? a) Incluye a más mujeres que hombres; **b)** Un 40% de la población son hombres y un 70% no es graduado universitario; c) La mitad de las mujeres es graduada universitaria; d) Es una población rural, debido al bajo porcentaje de graduados universitarios.

M: $\frac{120}{200} = 60\%$; $\bar{M}$: $\frac{80}{200} = 40\%$

G: $\frac{60}{200} = 30\%$; $\bar{G}$: $\frac{140}{200} = 70\%$

vi. ¿Cuál de las siguientes afirmaciones sobre el coeficiente de correlación lineal de Pearson (r) es verdadera?:
a) $r \approx \pm 1$ implica que existe una correlación lineal fuerte; b) $r = 0$ implica que existe una correlación lineal débil; **c)** Si existe una correlación lineal fuerte, entonces $r \approx \pm 1$; d) Tanto la a) como la c) son verdaderas.

vii. Un productor quiere anunciar que la efectividad de su producto es superior al 90%, en promedio. Antes, intenta contrastarlo. ¿Cuál es la prueba (contraste) adecuada? a) $H_0: \mu \leq 0.9$ frente a $H_1: \mu \leq 0.9$; **b)** $H_0: \mu \leq 0.9$ frente a $H_1: \mu > 0.9$; c) $H_0: \mu \geq 0.9$ frente a $H_1: \mu < 0.9$; d) $H_0: \pi \geq 0.9$ frente a $H_1: \pi \leq 0.9$.

Después de realizar una prueba de hipótesis (contraste) sobre diferencias de medias, la hipótesis nula no es rechazada. Entonces, el intervalo de confianza de la diferencia de medias es:

- a) (0.25, 1.56)
- b) (1.1, 1.2);
- c) (0.46, 0.92)
- **d)** (−0.2, 0.21)

No rechaza $H_0 \Rightarrow \mu_1 - \mu_2 = 0$

el 0 está en el dentro del intervalo d)

viii. La función de densidad de una variable aleatoria es $f(x) = cx$ si $0 < x < 4$, y $f(x) = 0$ en caso contrario. Entonces:
a) $c = 1/4$; **b)** $P(X \geq 1) = 15/16$; c) $P(X < 1) = 1/8$; d) $c = 1/16$.

Definición de función de densidad $\int_{-\infty}^{+\infty} f(x)dx = 1$

$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} \cancel{0}\,dx + \int_{0}^{4} cx\,dx + \int_{4}^{+\infty} \cancel{0}\,dx = c\frac{x^2}{2}\Big|_0^4 =$

$= c\left(\frac{4^2}{2} - \frac{0^2}{2}\right) = 8c = 1$; $c = \frac{1}{8}$

$P(x < 1) = \int_0^1 \frac{1}{8} x\,dx = \frac{1}{8}\frac{x^2}{2}\Big|_0^1 = \frac{1}{16}$; $P(x \geq 1) = 1 - \frac{1}{16} = \boxed{\frac{15}{16}}$

3.- (2 puntos) La vida útil de un líder semiconductor a potencia constante sigue una distribución normal con una media de 7000 horas y una desviación típica de 600 horas.

a) ¿Cuál es la probabilidad de que el líder falle antes de 5000 horas? **(0.67 puntos)**

b) ¿Cuál es la vida útil que supera el 95% de los líderes? **(0.67 puntos)**

c) Si se utilizan tres líderes en un producto y se asume que fallan de manera independiente, ¿cuál es la probabilidad de que los tres sigan funcionando después de 7000 horas? **(0.67 puntos)**

$X \to N(7000, 600)$

a) $P(x < 5000) = P(z < \frac{5000 - 7000}{600}) = P(z < -3.33) =$

$= 1 - P(z < 3.33) = 1 - 0'9996 = \boxed{0'0004}$

b) $P(x < k) = P(z < \frac{k - 7000}{600}) = 0'05$; $\frac{k - 7000}{600} = -1'645$

$\boxed{k = -1'645 \cdot 600 + 7000 = 6013}$

c) $P(x > 7000) = 0'5$

Si 3 líderes independientes

$\boxed{P(x > 7000 \cap x > 7000 \cap x > 7000) = 0'5 \cdot 0'5 \cdot 0'5 = 0'125}$