NOM:____________________________COGNOM:____

(Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs)

Problema 1 (B1-B2)

Tenim un programa format per dos mòduls. El primer mòdul conté algun error amb probabilitat 0.3. El segon mòdul és més complex i té una probabilitat de 0.45 de contenir algun error, independentment del primer mòdul. Considerem l'experiència aleatòria d'observar si hi ha error en els dos mòduls.

1. Indiqueu el conjunt de resultats possibles i calculeu les seves probabilitats (1 punt)

Considerem E1: "El primer mòdul conté algun error" i E2: "El segon mòdul conté algun error".
Aleshores $\Omega = \{(E1,E2), (E1, \neg E2), (\neg E1,E2), (\neg E1, \neg E2)\}$
$P((E1,E2)) = P((E1 \cap E2)) = P(E1) \cdot P(E2) = 0.3 \cdot 0.45 = 0.135$, ja que E1 i E2 són esdeveniments independents.
$P((E1,\neg E2)) = P((E1 \cap \neg E2)) = P(E1) \cdot P(E2) = 0.3 \cdot 0.55 = 0.165$
$P((\neg E1,E2)) = P((\neg E1 \cap E2)) = P(E1) \cdot P(E2) = 0.7 \cdot 0.45 = 0.315$
$P((\neg E1,\neg E2)) = P((\neg E1 \cap \neg E2)) = P(E1) \cdot P(E2) = 0.7 \cdot 0.55 = 0.385$

Que hi hagi algun error només en el primer mòdul fa que el programa no funcioni amb una probabilitat de 0.6. Si hi ha algun error només en el segon mòdul aquesta probabilitat és 0.8. Si hi ha errors en els dos mòduls aleshores el programa falla amb una probabilitat de 0.9. *El programa no falla si no hi ha cap error.*

2. Calculeu la probabilitat que el programa falli. (2 punts)

Sabem que $P(Falli \mid (E1 \cap \neg E2)) = 0.6$, $P(Falli \mid (\neg E1 \cap E2)) = 0.8$, $P(Falli \mid (E1 \cap E2)) = 0.9$ i $P(Falli \mid (\neg E1 \cap \neg E2)) = 0$
$P(Falli) = P(Falli \cap E1 \cap E2) + P(Falli \cap E1 \cap \neg E2) + P(Falli \cap \neg E1 \cap E2) + P(Falli \cap \neg E1 \cap \neg E2) =$
$= P(Falli \mid (E1 \cap E2)) \cdot P((E1 \cap E2)) + P(Falli \mid (E1 \cap \neg E2)) \cdot P((E1 \cap \neg E2)) + P(Falli \mid (\neg E1 \cap E2)) \cdot P((\neg E1 \cap E2)) + 0 =$
$= 0.9 \cdot 0.135 + 0.6 \cdot 0.165 + 0.8 \cdot 0.315 = 0.4725$

3. Suposem que el programa falla. Quina és la probabilitat que hi hagi algun error en tots dos mòduls? (1 punt)

$$P((E1 \cap E2) \mid Falli) = \frac{P(E1 \cap E2 \cap Falli)}{P(Falli)} = \frac{P(Falli \mid (E1 \cap E2)) \cdot P((E1 \cap E2))}{P(Falli)} = \frac{0.9 \cdot 0.135}{0.4725} = 0.2571$$

4. Considera l'esdeveniment E: "hi ha algun error en algun dels dos mòduls". Són independents els esdeveniments E i que el programa falli ? Justifica-ho. (1 punt)

$P(E) = 1 - P(\neg E) = 1 - P((\neg E1,\neg E2)) = 1 - 0.385 = 0.615$
$P(Falli) = 0.4725$

Per calcular $P(Falli \cap E)$ notem que $P(Falli) = P(Falli \cap E) + P(Falli \cap \neg E)$. Tenim que $P(Falli \cap \neg E) = P(Falli \mid \neg E) \cdot P(\neg E)$ i com que $P(Falli \mid \neg E) = 0$, aleshores $P(Falli \cap \neg E) = 0$ i per tant, $P(Falli) = P(Falli \cap E) = 0.4725$

$0.4725 = P(Falli \cap E) \neq P(Falli) \cdot P(E) = 0.4725 \cdot 0.615 = 0.2906$
No són esdeveniments independents.

Considerem les variables aleatòries X1="nombre d'errors en el primer mòdul" i X2="nombre d'errors en el segon mòdul" que són independents entre elles i que tenen les funcions de probabilitat que es mostren a continuació:

k	$P_{X1}(k)$	$P_{X2}(k)$
0	0.7	0.55
1	0.2	0.25
2	0.1	0.15
3	0	0.05

5. Calculeu l'esperança i la variància d'X1 (1 punt)

$E(X1) = 0 \cdot 0.7 + 1 \cdot 0.2 + 2 \cdot 0.1 + 3 \cdot 0 = 0.4$
$E(X1^2) = 0^2 \cdot 0.7 + 1^2 \cdot 0.2 + 2^2 \cdot 0.1 + 3^2 \cdot 0 = 0.6$
$Var(X1) = E(X1^2) - E(X1)^2 = 0.6 - 0.4^2 = 0.44$

6. Anomenem Y="nombre total d'errors" obtinguda com Y=X1+X2. Calcula la funció de distribució d'Y i representa-la gràficament (2 punts)

$P_Y(0)$=P(Y=0) = P(X1=0, X2=0) = P(X1=0)·P(X2=0) = 0.7 · 0.55 = 0.385
$P_Y(1)$=P(Y=1) = P(X1=1, X2=0) + P(X1=0, X2=1) = P(X1=1)·P(X2=0) + P(X1=0)·P(X2=1) = 0.2 · 0.55 + 0.7· 0.25 = 0.285
$P_Y(2)$=P(Y=2) = P(X1=2, X2=0) + P(X1=1, X2=1) + P(X1=0, X2=2) = P(X1=2)·P(X2=0) + P(X1=1)·P(X2=1)· P(X1=0)·P(X2=2) = 0.1 · 0.55 + 0.2· 0.25 + 0.7·0.15 = 0.21
$P_Y(3)$=P(Y=3) = P(X1=2, X2=1) + P(X1=1, X2=2) + P(X1=0, X2=3) = P(X1=2)·P(X2=1) + P(X1=1)·P(X2=2)· P(X1=0)·P(X2=3) = 0.1 · 0.25 + 0.2· 0.15 + 0.7·0.05 = 0.09
$P_Y(4)$=P(Y=4) = P(X1=2, X2=2) + P(X1=1, X2=3) = P(X1=2)·P(X2=2) + P(X1=1)·P(X2=3) = 0.1 · 0.15 + 0.2· 0.05 = 0.025
$P_Y(5)$=P(Y=5) = P(X1=2, X2=3) = P(X1=2)·P(X2=3) = 0.1 · 0.05 = 0.005

La funció de distribució:

k	$F_Y(k)$
(-∞, 0)	0
[0,1)	0.385
[1,2)	0.67
[2,3)	0.88
[3,4)	0.97
[4,5)	0.995
[5,+∞)	1

7. Calculeu l'esperança i la variància d'Y (1 punt)

Solució 1:
E(Y)=1· 0.285 + 2· 0.21 + 3· 0.09 + 4·0.025 + 5·0.005 = 1.1
$E(Y^2)= 1^2· 0.285 + 2^2· 0.21 + 3^2· 0.09 + 4^2·0.025 + 5^2·0.005 = 2.46$
$Var(Y)= E(Y^2) - E(Y)^2 = 2.46-1.1^2 = 1.25$

Solució 2:
E(X2)=0·0.55 + 1· 0.25 + 2· 0.15 + 3· 0.05 = 0.7
$E(X2^2)=0^2·0.55 + 1^2· 0.25 + 2^2· 0.15 + 3^2· 0.05 = 1.3$
$Var(X2)= E(X2^2) - E(X2)^2 = 1.3 – 0.7^2 = 0.81$
Per les propietats de l'esperança i la variància tenim que:
E(Y)= E(X1+X2)= E(X1)+E(X2)= 0.4 + 0.7 = 1.1
Var(Y)= Var(X1+X2)= Var(X1)+ Var(X2) = 0.44 + 0.81 = 1.25 perquè X1 i X2 són independents.

8. Calcula la correlació d'X1 i Y (1 punt)

La taula amb la funció de probabilitats conjunta és

X1 \ Y	0	1	2	3	4	5
0	0.385	0.175	0.105	0.035	0	0
1	0	0.11	0.05	0.03	0.01	0
2	0	0	0.055	0.025	0.015	0.005

E(X1,Y)= 1·1·0.11 + 1·2·0.05 + 1·3·0.03 + 1·4·0.01 + 2·2· 0.055+ 2·3·0.025 + 2·4· 0.015 + 2·5·0.005 = 0.88
Cov(X1,Y)= E(X1,Y) – E(X1)E(Y)= 0.88 – 0.4·1.1 = 0.44

$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{0.44}{\sqrt{0.44}\cdot\sqrt{1.25}} = 0.5933$

Una altra possibilitat per calcular cov(X1, Y) seria
cov(x1, y) = cov(x1, x1+x2) = cov(x1, x1) + cov(x1, x2) = v(x1) = 0.44

NOMBRE: ____

(Conteste cada pregunta en su sitio. Explique y justifique los cálculos).

Problema 2 (B3-B4). (SOLUCIÓN)

2.1 El tiempo necesario para presentar un examen sigue una distribución normal con media de 70 minutos y desviación estándar de 12 minutos.

a) (1.0 pts.) Calcule la probabilidad de que un alumno finalice el examen antes de una hora.

$X \sim N(\mu = 70, \sigma = 12)$

$P(X \leq 60) = P\left(Z \leq \dfrac{60-70}{12}\right) = P(Z \leq -0.83)$

$P(Z \leq -0.83) = P(Z \geq 0.83) = 1 - 0.7967 = 0.2033$

b) (1.0 pts.) ¿Cuánto debe durar el examen si se desea que el 90 % de los estudiantes tenga tiempo suficiente para terminarlo?

$X \sim N(\mu = 70, \sigma = 12)$

$P(X \leq x) = 0.9 \Rightarrow P\left(Z \leq \dfrac{x-\mu}{\sigma}\right) = 0.9$

$z = F^{-1}(0.9) = 1.28$

$x = \mu + \sigma \cdot z = 70 + (12)(1.28) = 85.36 \approx 86 \;\; minutos$

2.2 El examen consta de 16 preguntas de opción múltiple, cada una con cinco respuestas posibles. Solamente existe una respuesta correcta para cada pregunta. Suponga que uno de los alumnos contesta cada una de las preguntas de forma aleatoria e independiente.

a) (1.0 pts.) ¿Cuál es la probabilidad de que el alumno apruebe el examen?

X = número de éxitos en n=16 pruebas

$X \sim B(n = 16, p = 1/5)$

$P(X \geq 8) = 1 - P(X \leq 7) = 1 - \sum_{k=0}^{7} \binom{16}{k} (0.2)^k (0.8)^{16-k} = 1 - 0.993 = 0.007$

b) (0.5 pts.) Calcule el valor esperado del número de respuestas correctas.

$E(X) = np = (16)(1/5) = 3.2 \approx 3$

2.3 (1.0 pts.) Suponga que sólo el 30 % de los estudiantes lograrán contestar correctamente a la pregunta No. 1 (ya sea por suerte o por haber estudiado). En el momento de corregir el examen, el profesor elige los exámenes de manera aleatoria e independiente. Calcule la probabilidad de que el 5to examen corregido sea el primero en tener correctamente la pregunta No.1.

X = número de pruebas hasta observar el primer éxito

$X \sim Geom(p = 0.3)$

$P(X = 5) = (0.3)(0.7)^{5-1} = 0.0720$

2.4 (1.0 pts.) El tiempo que tarda un profesor en revisar cada una de las preguntas del examen, constituye una variable aleatoria independiente con media de 1.5 minutos y varianza de 1.0. Aproxime la probabilidad de que el profesor pueda revisar 100 preguntas en menos de 2 horas.

X_i variable aleatoria independiente con $\mu = 1.5$ y $\sigma = 1$

$n = 100$

$$P\left(\sum_{i=1}^{100} x_i \leq 120\right) = P\left(\overline{X} \leq \frac{120}{100}\right) = P(\overline{X} \leq 1.2) \implies \text{(n grande)} \implies \overline{X} \sim N(\mu, \sigma/\sqrt{n})$$

$$P\left(Z \leq \frac{1.2 - 1.5}{0.1}\right) = P(Z \leq -3) = 0.0013$$

2.5 Suponga que hasta ahora se han revisado 250 exámenes. El profesor encuentra que la pregunta No. 2 ha sido respondida correctamente por 134 alumnos.

a) (0.5 pts.) Haga una estimación puntual para la probabilidad de que un alumno conteste correctamente dicha pregunta.

$\hat{\pi} = p = 134/250 = 0.536$

b) (1.0 pts.) Encuentre un intervalo de confianza del 98 % para dicha probabilidad. ¿Se puede decir que la mayoría de los alumnos saben la respuesta correcta a la pregunta No. 2?

$$\begin{aligned} IC(\pi, 0.98) &= p \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.536 \pm z_{0.99}\sqrt{\frac{(0.536)(1-0.563)}{250}} \\ &= 0.536 \pm 2.33(0.03154) \\ &= [0.4625, 0.6095] \end{aligned}$$

El intervalo de confianza contiene el 0.5, por lo tanto, al 98 % de confianza, no se puede asegurar que la mayoría de los alumnos saben la respuesta correcta a la pregunta No. 2.

c) (1.0 pts.) ¿Cuántos exámenes se tendrían que corregir para afirmar que la mayoría de los alumnos saben la respuesta correcta a la pregunta No. 2? Utilice un riesgo del 5 %. (Asuma que la probabilidad estimada anteriormente se mantiene).

$[v, w] = [0.4625, 0.6095] \implies$ queremos que $v = 0.5$

$$v = 0.5 = p - z_{0.975}\sqrt{\frac{0.5(1-0.5)}{n}}$$

$$0.563 - 0.5 = (1.96)\sqrt{\frac{0.5^2}{n}} \implies n = \left(\frac{(1.96)(0.5)}{0.036}\right)^2 = 741.05 \approx 742 \text{ exámenes}$$

2.6 Las edades de 5 profesores de la asignatura son: 39, 54, 61, 72, 59.

$$\sum x_i = 285 \qquad \sum (x_i)^2 = 16823$$

a) (1.0 pts.) Encuentre un intervalo de confianza del 99 % para la desviación estándar de la edad de los profesores de la asignatura. Asuma que las edades siguen una distribución Normal.

$X \sim N(\mu, \sigma)$ $\qquad n = 5$ $\qquad \alpha = 0.01$ $\qquad s^2 = \dfrac{16823 - \frac{285^2}{5}}{5-1} = \dfrac{578}{4} = 144.5$

$$IC(\sigma^2, 0.99) = \left(\frac{s^2(n-1)}{\chi^2_{4,0.995}}, \frac{s^2(n-1)}{\chi^2_{4,0.005}}\right) = \left(\frac{578}{14.86}, \frac{578}{0.207}\right) = (38.8964, 2792.27)$$

$$IC(\sigma, 0.99) = (6.2367, 52.84)$$

b) (1.0 pts.) Queremos poner a prueba que la desviación estándar de las edades es de 7.2 años. Desarrolle formalmente la prueba de hipótesis para contrastar esta cuestión con un riesgo del 5 %. Haced el contraste de que la desviación estándar es menor a 7.2. Presente gráficamente la distribución de probabilidad del estadístico y marque el área de rechazo. ¿Cuál es la conclusión de esta prueba? Dé una aproximación para el p-valor.

$$H_0: \quad \sigma = 7.2$$
$$H_1: \quad \sigma < 7.2$$

$$\hat{\chi^2} = \frac{s^2(n-1)}{\sigma_0^2} = \frac{(144.5)(4)}{7.2^2} = 11.1497 \sim \chi_4^2 \qquad > \quad 0.711 = \chi^2_{0.05,4}$$

$$\text{p-valor} = P\left(\hat{\chi^2} < 11.14\right) \approx 0.975$$

$\therefore$ No existe suficiente evidencia para rechazar H_0, i.e., no hay nada absolutamente a favor de la H_1.

NOM:_______________________________COGNOMS:____

(*Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs*)

Problema 3 (B5-B6)

Volem analitzar la relació que hi ha entre el temps de pujada d'una imatge a la web en el format TIFF (dades A) i en el format JPEG (dades B). Per dur a terme aquesta investigació es recull una mostra amb 32 observacions, 16 amb TIFF i 16 JPEG, i s'obtenen els següents temps [segons]:

A: 9.407 9.215 9.650 8.798 8.311 9.081 8.137 8.981 8.492 8.443 8.633 8.760 8.108 9.066 9.133 8.257
B: 7.657 9.178 8.659 9.308 8.683 9.270 7.680 8.340 8.674 8.462 7.767 8.570 7.996 8.115 7.632 8.175

Els estadístics a utilitzar són:
Mitjana A = 8.78; Mitjana B = 8.39; Covariància = 0.05
Variància A = 0.22; Variància B = 0.32; Variància diferència = 0.44

1. Assumim que són dades aparellades (mateixa imatge en dos formats); volem contrastar si l'esperança del temps de pujada per als dos formats és la mateixa o no:
   a. Plantegeu quina és la hipòtesi nul·la i l'alternativa
   b. Digueu quin és l'estadístic, la seva distribució sota H_0 i les premisses
   c. Es pot rebutjar la hipòtesi nul·la? Raoneu la vostra resposta i doneu la conclusió. Nota: No calculeu aquí l'interval de confiança (2 punts)

a.
H_0: $\mu_A = \mu_B$
H_1: $\mu_A \neq \mu_B$

b.
Estadístic t
Premisses m. a. aparellades, A-B segueix una normal

c.
Error_tipus = sqrt(0.44) / sqrt(16) = 0.166
Mitjana_diferència = 8.78 -8.39 = 0.39
t = (0.39 - 0) / 0.166 = 2.35

$t_{15, 0.975}$= 2.13145

2. Assumim ara que les dades no són aparellades, i resolem el mateix contrast:
   a. Plantegeu quina és la hipòtesi nul·la i l'alternativa
   b. Digueu quin és l'estadístic, la seva distribució sota H_0 i les premisses en aquest cas
   c. Es pot rebutjar la hipòtesi nul·la? Raoneu la vostra resposta i doneu la conclusió.
   Nota: No calculeu aquí l'interval de confiança. (2 punts)

a.
H_0: $\mu_A = \mu_B$
H_1: $\mu_A \neq \mu_B$

b.
Estadístic t
Premisses m.a.s. A i B normals, $\sigma_A = \sigma_B = \sigma$

c.
s_pooled = sqrt((15*0.22 + 15*0.32) / (16+16-2)) =0.52
Error_tipus= 0.520* sqrt(1/16+1/16) = 0.184
t = (8.78 -8.39) /0.184=2.12
$t_{30, 0.975}$= 2.042272

3. Calculeu, interpreteu i compareu els intervals de confiança al 95% per a la diferència de mitjanes pels 2 casos anteriors. Tant si els intervals són similars com si són diferents, digueu quina creieu que és la raó per la seva semblança o disparitat. (1 punt)

IC apartat 1:
0.39± t15,0.9750.166 = [0.036; 0.744]

IC apartat 2:
0.39±t30,0.9750.184 = [0.0142; 0.766]

Són semblants perquè les dades estan poc correlacionades entre sí (rho=cov(A,B)/(sd(A)*sd(B)) = 0.05/(0.47*0.57) = 0.19) i, per tant, el guany en eficiència del disseny aparellat és menor.

4. Per veure si hi ha relació lineal entre els dos temps de pujada, volem calcular la recta de regressió de B (resposta) en funció de A (predictor). Estimeu puntualment el terme independent (Beta 0) i el pendent (Beta 1). Doneu un interval de confiança al 95% per a Beta 1 i interpreteu-ho. (2 punts)

b_1= cov(A,B) / var A= 0.05/0.22=0.227
b_0= mean B- b_1*mean A=8.39-0.227*8.78=6.397
b_1±t14,0.975*S_{b1}= 0.227 ±2.145 * 0.317=[-0.453, 0.907]

(s^2=15*(0.32-0.227*0.05)/14 =0.331
S_{b1}= sqrt (0.331/(15*0.22)) = 0.317)
L'increment del temps de pujada en B per cada increment unitari d'A està entre -0.45 i 0.91. Com que el 0 està inclòs dins de l'interval no hi ha evidència per dir que hi ha relació lineal entre els 2 temps de pujada.

5. Sabent que pujar una imatge TIFF ha trigat 10 segons, calculeu la predicció puntual del temps de pujada pel format JPEG i l'interval de confiança al 95% per al corresponent valor esperat (2 punts)

B_{10}= 6.397+ 0.227*10 = 8.667
IC95% = 8.667±$t_{14,0.975}$* sqrt(0.331) * sqrt(1/16 + $(10\text{-}8.78)^2$ / (0.22*15))
= 8.667±2.145* 0.575* 0.717=[7.783, 9.551]

6. La validació del model lineal es realitza mitjançant gràfics dels residus. Quines són les premisses que s'avaluen? (Només heu de dir el nom de les premisses, no cal avaluar-les) ¿Per a quina premissa concreta es fa servir el gràfic dels residus en funció de l'ordre de recollida? Com s'interpreta? (1 punt)

Linealitat, homoscedasticitat, Independència i Normalitat. El gràfic dels residus en funció de l'ordre de recollida serveix per avaluar la premissa d'independència. Hi ha independència si no es veu cap patró en el gràfic