NOM:_______________________________COGNOMS:____

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

Problema 1 (B1-B2)

S'ha estudiat els resultats de 250 alumnes de les assignatures de BD i PE. Per simplificar s'han arrodonit les notes per rangs superiors i finalment s'ha obtingut la següent taula:

BD \ PE	**0**	**4**	**6**	**9**	**10**	
0	10	15	5	3	2	35
4	7	20	10	7	0	44
6	3	18	20	10	12	63
9	5	10	15	30	15	75
10	0	5	10	7	11	33
	25	68	60	57	40	250

1. Considereu PE i BD com a variables aleatòries resultants de l'experiència "escollir a un alumne a l'atzar". Trobeu la funció de probabilitat de la variable PE, i calculeu-ne el valor esperat i la desviació tipus. Interpreteu el resultat. ? (1.5 punts)

PE				
0	**4**	**6**	**9**	**10**
0,10	0,272	0,240	0,228	0,16

E(PE) = 0*0.1+4*0.272+6*0.24+9*0.228+10*0.16 = 6.18

V(PE) = $E(X^2) - \mu_X^2$ = 47.46 - 38.1924 = 9.2676 --- σ = 3.044

2. Trobeu la distribució de la variable PE sabent que la qualificació és de més de 4 a l'assignatura BD). Quina és la qualificació esperada per a aquests alumnes? (1.5 punts)

PE/AprovatBD				
0	**4**	**6**	**9**	**10**
0,0467	0,193	0,2631	0,2748	0,2222

Numerador = sumatori de las proporciones o totals por nota de PE. Ex : Nombre d'alumnes que han tret un 0 a PE i han aprovat BD = 8

Denominador = proporció o totals de aprovats de BD : 171

E(PE | BD>4) = 0*0.0467+4*0.193+6*0.2631+9*0.2748+10*0.2222 = 7.0458

3. Si s'obté un 10 en alguna de les dues assignatures, el següent curs es guanya beca pels crèdits equivalents. A més, en el cas que es tregui un 9 en ambdues assignatures també s'obté una beca (però només una). Quina probabilitat té un alumne a pertànyer al segon cas (haver tret dos "9"), sabent que ha obtingut beca? (1 punt)

B = "alguna beca (1 o 2)"

P(B) = P(BD=10)+P(PE=10)-P(BD=10 i PE=10) + P(PE=9 i BD=9) = 0.132+0.16-0.044+0.12 = 0.368

P(BD=9 i PE=9)/P(B) = 0.12/0.368 = 0.326

4. Expliqueu la diferència existent entre trobar la probabilitat que un alumne amb una (i només una) beca hagi tret un 9 a PE o que un alumne amb un 9 a PE hagi obtingut una (i només una) beca. Calculeu-les. (1 punt)

La diferència consisteix en establir quin és el conjunt sobre el que ens fixem en els casos favorables (que anirà al denominador).

B = Beca=1 o Beca=2

P(Beca=1) = P(B)-P(Beca=2) = 0.368 - 0.044 = 0.324

P(PE=9 i (BD=9 o BD=10)) = 0.12 + 0.028 = 0.148

P(PE=9 | Beca= 1) = P(PE=9 i (BD=9 o BD=10)) / P(Beca=1) = 0.148/0.324 = 0.4568

P(Beca=1 | PE=9) = P(PE=9 i (BD=9 o BD=10)) / P(PE=9) = 0.148/0.228 = 0.6491

5. Trobeu i representeu gràficament les funcions de probabilitat de la nota de PE tal que han tret un 4 a BD i la funció de PE tal que han tret un 6 a BD. Són independents les qualificacions de PE i de BD? Justifica la resposta. (1.5 punts)

PE/BD=4				
0	4	6	9	10
0,16	0,45	0,23	0,23	0
PE/BD=6				
0	4	6	9	10
0,04	0,29	0,32	0,16	0,19

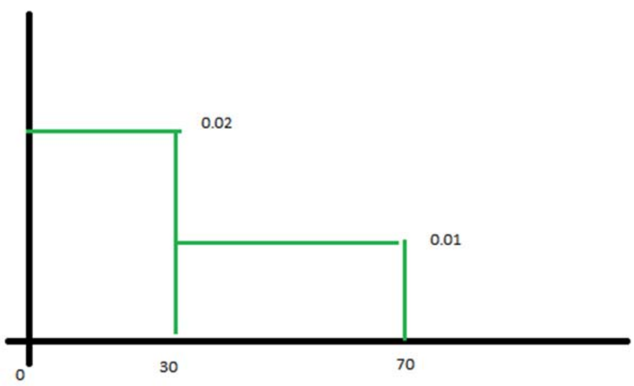
No són independents perquè la nota d'una assignatura fa variar la nota que es treu de l'altre, un exemple les dues funcions representades (si condicionem per BD=4 o per BD=6 la distribució de la nota de PE canvia).

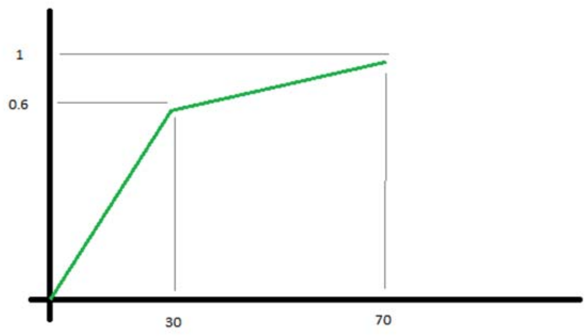
6. Si S és la suma de les qualificacions de PE i BD, calcula la correlació entre ambdues qualificacions, coneixent que $E(BD^2) = 49.388$ i $E[S^2] = 182.48$ (1.5 punt)

E(PE) = 6.18 sigma(PE) = 3.044273
E(BD) = 6.236 → E(S) = 12.416 sigma(BD) = sqrt(49.388 – 6.236^2) = 3.240417
V(S) = 182.48 – 12.416^2 = 28.32294
V(S) = V(PE) + V(BD) + 2 Cov(PE,BD) → Cov = 8.55504/2 = 4.27752
Corr = 4.27752/3.044273/3.240417 = 0.4337

Estem analitzant les hores dedicades per aprovar una assignatura, i es sap que en el cas d'una assignatura de 7 crèdits s'ha de dedicar un mínim de 60 hores. Es vol estudiar les hores addicionals que dedica cada estudiant.

7. Es sap que la distribució d'aquest temps de dedicació és així: un 60% dels alumnes dedica entre 0 i 30 hores addicionals, i la resta dedica entre 30 i 70, en ambdós casos amb perfil uniforme. Dibuixa les funcions de probabilitat i de distribució i descriu el significat de cadascuna d'elles. (1 punt)





La proporció d'hores dedicades és constant entre 0 i 30. I a partir del 30 fins al 70 també és constant, amb probabilitat menor ja que hi ha menys proporció d'alumnes i més diferència d'hores. La funció de distribució, com és l'acumulació, té una pendent constant per cada un dels dos intervals.

8. Calcula la probabilitat que un alumne dediqui en total un màxim de 100 hores a preparar l'assignatura. Troba quin és el temps de dedicació que correspon al percentil 95%. (1 Punt)

T: temps de dedicació; HA: hores addicionals (T = 60+HA)
T < 100 <-> HA < 100-60 = 40: P(HA<40) = 0.6 + 0.1 = 0.7
F(t) = 0.95. La recta de la segona part té com a equació Y = 0.01X + 0.3, resolent: X_{95} = 65h (en total 125h)

NOM:_______________________________COGNOMS:____

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

Problema 2 (B3-B4)

El temps de lliurament dels paquets d'una determinada empresa de missatgeria (24 hores/7 dies) es distribueix segons una exponencial. L'empresa considera que un temps raonable per completar el lliurament són 2 dies. Malgrat això, es coneix que només el 10% dels lliuraments d'aquesta empresa es fan dins d'aquest termini de 2 dies.

1. Quin és el temps esperat (en dies) de lliurament? (1 punt)

   T="Temps de lliurament" ~ Exp(λ)

$$F(t) = 1 - e^{-\lambda \cdot t}$$

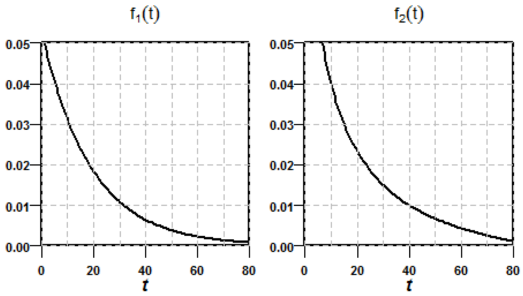
$$F(2) = 1 - e^{-\lambda \cdot 2} = 0.10 \rightarrow e^{-2\lambda} = 0.90 \rightarrow -2\lambda = \ln(0.90) \rightarrow \lambda = \frac{\ln(0.90)}{-2} = 0.053 \frac{lliurament}{dia}$$

$$E(T) = \frac{1}{\lambda} = \frac{1}{0.053} = 18.98\ dies$$

2. Quina és la probabilitat que un lliurament que no s'ha lliurat durant el primer dia, es lliuri en els dos dies següents, és a dir que es lliuri abans de 3 dies? (0.5 punts)

   Per la propietat sense memòria de l'exponencial, no influeix el que hagi passat en el passat. Per tant, la probabilitat és de 0.1

3. Digues quin dels 2 gràfic següents creus que representa la funció de densitat d'aquesta variable exponencial i argumenta-ho. *t*: "temps en dies" (0.5 punts)



La funció de densitat és la de l'esquerra. Per veure-ho es pot calcular un el valor en algun punt que estigui bastant allunyat en les 2 gràfiques, per exemple, per t=10, $f_1(10) \approx 0.03$ i $f_2(10) \approx 0.04$. Si fem els càlculs amb la lambda de l'apartat anterior, obtenim:

$$f(t = 10) = -\lambda \cdot \mathrm{e}^{-\lambda \cdot t} = -0.053 \cdot e^{-0.053 \cdot 10} \approx 0.03$$

Per tant, la funció de densitat correcte és la primera

4. En una determinada sucursal han rebut 20 ordres de lliurament en un dia concret. Quina és la probabilitat que almenys 2 d'aquests lliuraments arribin abans de 2 dies? (1 punt)

   X="Nombre de lliuraments lliurats abans de 2 dies" ~ B(n=20,p=0.1)

$$P(X \geq 2) = 1 - P(X \leq 1) = (Taules) = 1 - 0.3917 = 0.6083$$

5. En 2000 lliuraments fets per aquesta empresa, quina és la probabilitat que almenys 200 arribin en els 2 primers dies? (0.5 punts)

   Y="Nombre de lliuraments que arriben abans de 2 dies" ~ B(n=2000,p=0.1)~N(μ=200,σ=13.41)

   $P(Y \geq 200) \approx 0.5$ per simetria de la distribució Normal.

6. Digues quan val ***y*** en la següent frase i justifica els càlculs: "La companyia garanteix que almenys ***y*** paquets (dels 2000 anteriors) es lliuraran abans de dos dies amb una probabilitat de 0.95" (1 punt)

$$P(Y < y) = 0.05 \rightarrow P\left(\frac{Y-200}{13.41} < \frac{y-200}{13.41}\right) = 0.05 \rightarrow P\left(Z < \frac{y-200}{13.41}\right) = 0.05 \rightarrow \frac{y-200}{13.41} = -1.645 \rightarrow y = 177.9$$

7. Els paquets de pes entre 1 kg i 5 kg (suposem que el pes és uniforme) es posen a banda en un carretó per ser classificats. Si en un moment donat hi han 30 paquets al carretó, calcula la probabilitat que el pes de la càrrega del carretó superi els 100 kgs. (1 punt)

U = "Pes de 1 paquet" ~ U(1,5)

$$E(U) = \frac{b+a}{2} = \frac{5+1}{2} = 3 \;\; ; \;\; V(U) = \frac{(b-a)^2}{12} = \frac{(5-1)^2}{12} = 1.33$$

S = "Suma dels pesos de 30 paquets" ~ (TCL) ~ N(μ=30·3, σ=$\sqrt{30 \cdot 1.33}$)= N(μ=90, σ=6.32)

$$P(S \geq 100) = P(S \geq 100) = P\left(Z > \frac{100-90}{6.32}\right) = P(Z > 1.58) = 1 - P(Z < 1.58) = (Taules) = 1 - 0.943 = 0.057$$

8. En un estudi posterior amb 60 lliuraments, posant a prova si el temps mitjà és inferior a 15 dies s'ha trobat un p-valor igual a 0.1. Si la mitjana mostral era idèntica a la desviació mostral, pots trobar quin ha estat la mitjana del temps de lliurament a la mostra? (1 punt)

$$\frac{\bar{x}-\mu}{s\sqrt{\frac{1}{n}}} \sim t_{59} \rightarrow \frac{\bar{x}-15}{\bar{x}\sqrt{\frac{1}{60}}} = -1.296 \rightarrow \bar{x} - 15 = -0.167\bar{x} \rightarrow \bar{x} = 12.85$$

Per tal de millorar el rendiment del temps de lliurament, l'empresa ha introduït una millora en la logística. La suma i la suma al quadrat dels temps de lliurament en dies dels 120 primers enviaments han estat:

$$\sum_{i=1}^{120} y_i = 2082.2 \qquad \sum_{i=1}^{120} y_i^2 = 147992.5$$

9. Dóna una estimació puntual de la mitjana i de la desviació poblacionals dels nous temps de lliurament. (1 punt)

$$\hat{\mu} = \bar{x} = \frac{2082.2}{120} = 17.35$$

$$\hat{\sigma} = s = \sqrt{\frac{147992.5 - 120 \cdot 17.35^2}{119}} = 30.66$$

10. Calcula un interval de confiança del 95% del temps mitjà de lliuraments poblacional (1 punt)

$$IC(95\%, \mu) = \bar{y} \mp t_{119,0.975}\frac{s}{\sqrt{n}} = 17.35 \mp 1.98 \cdot \frac{30.66}{\sqrt{120}} = [11.8, 22.9]$$

11. Revisant les dades, es donen compte que entre aquests 120 lliuraments hi havia un d'un paquet que es va quedar oblidat dins d'un magatzem i que no es va entregar fins a 300 dies més tard des de que es va donar l'ordre. Recalcula l'anterior interval de confiança sense considerar aquest lliurament. (1 punt)

$$\sum_{i=1}^{119} y_i = 2082.2 - 300 = 1782.2 \qquad \sum_{i=1}^{119} y_i^2 = 147992.5 - 300^2 = 57992.5$$

$$\hat{\mu} = \bar{x} = \frac{1782.2}{119} = 14.98$$

$$\hat{\sigma} = s = \sqrt{\frac{147992.5 - 119 \cdot 17.35^2}{118}} = 16.29$$

$$IC(95\%, \mu) = \bar{y} \mp t_{118,0.975}\frac{s}{\sqrt{n}} = 14.98 \mp 1.98 \cdot \frac{16.29}{\sqrt{119}} = [12.02, 17.93]$$

12. Assenyala quin és el límit inferior de l'interval de confiança del 95% per σ després d'haver eliminat aquesta observació entre les següents opcions. *Pista: no facis càlculs però justifica la resposta breument* (0.5 punts)

a) 1234.8
b) 739.7
c) 348.5
d) 208.7
e) 35.1
f) 27.2
g) 18.7
h) 14.5 → Correcte

Es pot saber perquè és l'únic valor inferior a l'estimació puntual

NOM:____________________________COGNOM:____

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

Problema 3 (B5-B6)

Volem fer un estudi per comparar el temps d'encriptació (en µs) de dos algorismes AES i CAM. Per fer-ho es creen 30 fitxers de mida aleatòria i es mesura el temps d'encriptació per l'algorisme AES (A) i l'algorisme CAM (C). Les dades en µs es recullen en la següent taula:

	A (AES)	C (CAM)	D = C - A
Mitjana	5'1674	7'6076	2'4402
Desviació tipus	0'6363	2'1781	1'5558

Podem suposar que les variables A, C i D segueixen una distribució normal. Per contestar els següents apartats considereu un risc α=0.05 i una confiança del 95%.

1. Es tracta de dues mostres independents o aparellades? Raoneu la resposta. (1 punt)

   Es tracta de mostres aparellades, ja que per cada fitxer es pren tant la dada del temps d'encriptació per l'algorisme AES (A) com pel CAM (C)

   Suposeu que es tracta d'un disseny aparellat. Volem estudiar si els dos algorismes triguen o no el mateix temps en encriptar.

2. a) Indiqueu les hipòtesis i les premisses de la prova a realitzar. (0'5 punts)

   $\begin{cases} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{cases}$, test bilateral . D segueix una Normal. Mostra aleatòria simple.

   b) Indiqueu la fórmula de l'estadístic i quina és la distribució d'aquest quan se suposa que els dos algorismes tenen la mateixa mitjana. (0'5 punts)

   $$\hat{t} = \frac{\bar{D} - \mu_0}{\frac{s_D}{\sqrt{n}}}, \quad \hat{t} \sim t_{n-1} = t_{29}$$

   c) Calculeu l'estadístic i raoneu si podem rebutjar la hipòtesi nul·la. (1 punt)

   $$\hat{t} = \frac{\bar{D} - \mu_0}{\frac{s_D}{\sqrt{n}}} = \frac{2'4402}{\frac{1'5558}{\sqrt{30}}} = 8'5908$$

   $t_{29,\,0'975}$ = 2'045. Per tant, com que l'estadístic és més gran que el punt crític, tenim evidències per rebutjar que els dos algorismes triguin el mateix temps en encriptar els fitxers.

   Considereu que es tracta d'un disseny amb mostres independents.Ara volem estudiar si l'algorisme CAM té una dispersió més gran que l'algorisme AES posant a prova les variàncies del temps d'encriptació.

3. a) Indiqueu les hipòtesis i les premisses. (0'5 punts)

   $\begin{cases} H_0: \sigma_C^2 = \sigma_A^2 \\ H_1: \sigma_C^2 > \sigma_A^2 \end{cases}$, Mostra aleatòria simple. A i C segueixen una normal.

   b) Indiqueu la fórmula de l'estadístic i quina és la distribució d'aquest sota la hipòtesi nul·la. (0'5 punts)

   $$\hat{F} = \frac{s_C^2}{s_A^2}, \quad \hat{F} \sim F_{nC-1,nA-1} = F_{29,29}$$

   c) Calculeu l'estadístic i raoneu si podem si podem rebutjar la hipòtesi nul·la. (1 punt)

$\hat{F} = \frac{s_C^2}{s_A^2} = \frac{4'7441}{0'4049} = 11'7167$

$F_{29,29,\,0'95}$ = 1'86. Per tant, com que l'estadístic és més gran que el punt crític, la variància de C és més gran que la d'A

Volem estudiar si hi ha relació entre els temps d'encriptació dels dos algorismes.

4. a) A partir de les propietats de la variància, deduïu una expressió per calcular el coeficient de correlació lineal amb les dades donades a l'exercici, és a dir, les desviacions tipus de les variables A, C i D. Calculeu el coeficient de correlació lineal. (1'5 punts)

Com que sabem que Var(D) = Var(C-A)=Var(C) + Var(A) – 2·Cov(C,A), aleshores tenim que 2·Cov(C,A)= Var(C) + Var(A) – Var (D)

El coeficient de correlació lineal compleix que:

$$r_{AC} = \frac{Cov(A,C)}{s(A)\cdot s(C)} = \frac{2\cdot Cov(A,C)}{2\cdot s(A)\cdot s(C)} = \frac{Var(C)+Var(A)-Var\ (D)}{2\cdot s(A)\cdot s(C)} = \frac{s(A)^2+s(C)^2-s(D)^2}{2\cdot s(A)\cdot s(C)}$$

$$r_{AC} = \frac{s(A)^2+s(C)^2-s(D)^2}{2\cdot s(A)\cdot s(C)} = \frac{0'6363^2+2'1781^2-1'5558^2}{2\cdot 0'6363\cdot 1'5558} = 0'9843$$

Si no heu pogut calcular el coeficient de correlació, useu el valor r=0'9 pels següent apartats.

b) Interpreteu el coeficient de correlació lineal. (0'5 punts)

El valor del coeficient de correlació és molt proper a 1, per tant tenim un bon ajustament lineal directe.

5. a) Calculeu la recta de regressió de l'algorisme CAM respecte l'algorime AES. (1 punt)

b1= $r \cdot \frac{s(C)}{s\ (A)}$ = 3'3693 (3'08 amb r=0'9)

b0= $\bar{C} - b1 \cdot \bar{A}$ = -9'8029 (-8'31 amb r=0'9)

La recta de regressió és C=3'3693A – 9'8029

b) Podem considerar el temps d'encriptació de CAM està associat al d'AES? Realitza la prova d'inferència estadística apropiada per donar resposta a la pregunta. (1 punt)

Plantegem si el pendent de la recta de regressió és diferent de zero.

$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$, tenim que $\hat{t} = \frac{b1}{S_{b1}}$ = 29'5293, $\hat{t} \sim t_{28}$, per tant el punt crític és 2'048

On $S_{b1} = \sqrt{\frac{S^2}{(n-1)\cdot s(A)^2}}$ =0'1141 (S^2 = 0'1531)

A partir de la prova realitzada podem concloure que tenim evidències que el temps d'encriptació de CAM està associat al d'AES.
(Amb r=0'9, t=10'92 arribant a la mateixa conclusió)

6. Fes una predicció puntual i per interval de confiança pel temps d'encriptació de CAM per un temps d'encriptació d'AES de 5 µs. (1'5 punts)

Estimació puntual: $\widehat{C_h}$ = b0+b1·A_h = -9'8029 + 3'3693 · 5 = 7'0436 µs

Estimació per interval: $\widehat{C_h} \pm t_{n-2,0\prime975} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(A_h - \bar{A})^2}{(n-1)\cdot s(A)^2}} =$

$$7'0436 \pm 2'048 \cdot 0'3913 \cdot \sqrt{1 + \frac{1}{30} + \frac{(5-5\prime1674)^2}{29\cdot 0\prime6363^2}} = (6'2280, 7'8591)$$

Per a 5 µs de temps d'encriptació d'AES podem esperar un temps d'encriptació entre 6'22 i 7'85 µs amb una confiança del 95%