NOM:__________________________COGNOM:____

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

Problema de B1

Suposem que en una ciutat concreta, l'aeroport A controla el 50% del trànsit aeri de passatgers, i que els aeroports B i C controlen el 30% i 20% restant respectivament. Les taxes de detecció d'armes (probabilitat que un passatger porti una arma) en aquests aeroports són de 0.10, 0.20 i 0.20 respectivament.

a) Dibuixa l'arbre de probabilitats que es dedueix de l'enunciat i el conjunt de resultats (1 punt)

{(A,SIa), (A,NOa), (B,SIa), (B,NOa), (C,SIa), (C,NOa)}

b) Quina és la probabilitat de que un passatger, en un dels aeroports qualsevol, no porti una arma? (1 punt)

P(NOa) = P((A,NOa))+ P((B,NOa))+ P((C,NOa)) = 0.5*0.90 + 0.3*0.80 + 0.2*0.80 = 0.45 + 0.24 + 0.16 = 0.85

c) S'informa que s'ha detectat un passatger que porta una arma. De quin aeroport és més probable que sigui ? (2 punts)

Max (P(A|SIa), P(A|SIa), P(A|SIa))

Max ((0.5*0.10)/0.15 , (0.3*0.20)/0.15 , (0.2*0.20)/0.15) =

Max (0.05/0.15, 0.06/0.15, 0.04/0.15) =

Max (0.334 , 0.40 , 0.267)

De l'aeroport B

d) Si un passatger ve de l'aeroport A o B, quina és la probabilitat de que no porti cap arma? (1 punt)

P(NOa | (A o B)) = ((0.5*0.90)+(0.3*0.80)) / (0.5+0.3) = 0.69 / 0.8 = 0.8625

e) En aquesta ciutat es proposa que independentment de l'aeroport i de si s'ha detectat arma o no, la meitat dels passatgers s'han de treure les sabates. Quina és la probabilitat que un passatger que no porta cap arma i no s'ha tret les sabates estigui a l'aeroport A? I que estigui a l'aeroport A si sabem que no porta cap arma però no sabem res de les sabates? (2 punts)

P(A | (NOa,NSab)) = (0.5*0.9*0.5) / (0.5*0.9*0.5+0.3*0.8*0.5+0.2*0.8*0.5) = 0.225 / 0.425 = 0.529

P(A | (NOa)) = (0.5*0.9) / (0.5*0.9+0.3*0.8+0.2*0.8) = 0.45 / 0.85 = 0.529

f) Suposem una altra ciutat que també té tres aeroports amb el mateix repartiment del trànsit aeri i amb les mateixes taxes de detecció d'armes en els dos primers aeroports. Quines són les taxes en el tercer aeroport si sabem que en aquesta ciutat la probabilitat de que un passatger no porti arma en un qualsevol dels aeroports és del 84% ? (2 punts)

P(NOa) = P((A,NOa))+ P((B,NOa))+ P((C,NOa)) = 0.5*0.90 + 0.3*0.8 + 0.2* x = 0.84

0.45 + 0.24 + 0.2x = 0.84   x = 0.15 / 0.2   x=0.75

A C prob NOa és 0.75, i la prob de SIa és 0.25

e) En aquesta segona ciutat, quina és la probabilitat d'estar a l'aeroport B o C i portar una arma? (1 punt)

P(B,SIa) + P(C,SIa) = P((B,SIa))+ P((C,SIa)) = 0.3*0.20 + 0.2*0.25 = 0.06 + 0.05 = 0.11

NOM:____________________________COGNOM:____

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

Problema de B2

Algunes assignatures de la FIB, donada la situació del COVID 19, han volgut quantificar la probabilitat de fer exercicis setmanalment en les plataformes online que posen a disposició dels estudiants. En una assignatura X la distribució de probabilitats de fer entre 0 i 4 exercicis és la següent:

X = x	0	1	2	3	4
P (X = x)	0.1	0.2	0.25	0.25	0.2

1) Quina és la probabilitat de fer algun exercici? Argumenta si es compleix la condició que espera el professorat de que quatre de cada cinc estudiants facin algun exercici (1 punt)

P(X > 0) = P(X=1)+P(X=2)+P(X=3)+P(X=4) = 0.9

Per les condicions de l'enunciat -> P(X = 0) = 0.1 i, per tant, el 90% des estudiants han fet servir la plataforma i es pot dir que sí es compleix pq 0.9 > 4/5 = 0.8

2) Suposem que sabem que un alumne ha fet algun exercici. Quina és la probabilitat que n'hagi fet més de dos? (1 punt)

P(X>2/X>0)=P(X>=3/X>=1) = P(X>=3 ∩ X>=1) / P(X>=1) = 0.45/ 0.9 = 0.5

3) Quin és el valor esperat i la desviació típica del nombre d'exercicis realitzats setmanalment? Comenteu i interpreteu els resultats (1.5 punts)

E(X) = 0*0.1+... = 2,25

V(X) = $(0-2.25)^2$*0.1+... = 1,5875 exercicis^2 i la Desviació típica és de 1,26 exercicis (es descompta 0.25 si només es dona la variància)

S'espera que un alumne faci entre 2 o 3 exercicis setmanalment amb una variabilitat d'entre 1 i 2 (No es compta els 0,5 punts si no es comenta)

4) Per a una assignatura amb 100 matriculats i amb l'anterior distribució de probabilitats per a cada estudiant (i suposant que el nombre d'exercicis que fa un estudiant no afecta als demés), calculeu, i justifiqueu formalment, quin és el valor esperat i la desviació del nombre d'exercicis que recollirà l'assignatura setmanalment (1.5 punts)

La idea és que són independents:

$S_{100} = X_1 + X_2 + ... + X_{100}$

$E(S100) = \sum_{i=1}^{100} E(Xi) = 100 * 2,25 = 225$ exercicis

$\sigma(S100) = \sqrt{\sum_{i=1}^{100} Var(Xi)} = \sqrt{100 * 1,5875} = 12.6$ exercicis

5) Dues assignatures, X i Y, han estudiat la distribució de probabilitat conjunta dels seus estudiants comuns per fer exercicis d'una i altra setmanalment, obtenint:

	X=0	X=1	X>1
Y=0	0.05	0	0.05
Y=1	0	0.1	0.1
Y>1	0.05	0.1	0.55

Quina és la probabilitat de fer 2 o més exercicis entre les dues assignatures? (1 punt)

1 – (P(X=0,Y=0)+ P(X=1,Y=0)+ P(X=0,Y=1)) = 1 – (0.05+0+0) = 0.95

Indiqueu i justifiqueu com és la relació entre fer exercicis d'una assignatura i l'altra (1 punt)

No hi ha relació lineal ni directa ni inversa pq les probabilitats es reparteixen per igual

6) Les dues assignatures anteriors també tenen estudiada la distribució del temps en minuts que dediquen els seus estudiants en resoldre els exercicis.Per a l'assignatura X els minuts dedicats segueixen la distribució exp(-k), i per a l'assignatura Y la distribució 1/9 k^2 (0<=k<=3)

Indiqueu les funcions de probabilitat i de distribució de probabilitat en cada cas i calculeu quina és la probabilitat d'estar menys de 2 minuts per fer un exercici en cada cas? (1.5 punts)

$f_{TX}(k)$ = exp(-k) $F_{TX}(k)$ = 1-exp(-k)
P(TX<2) = $F_{TX}(2)$ = 1- exp(-2) = 0.865

$f_{TY}(k)$ = 1/9 k^2 (0<=k<=3) $F_{TY}(k)$ = 1/27 k^3 (0<=k<=3)
P(TY<2) = $F_{TY}(2)$ = 1/27 2^3 = 8/27 = 0.296

En quin dels dos casos el valor esperat de minuts per fer un exercici és més gran? (1.5 punts)

E(TX) = {segons model exp, o resoldre integral per parts} = 1/1 = 1 minut

E(TY) = {integral de 1/9 k^3 entre 0 i 3} = 1/36 k^4 entre 0 i 3 = 81/36 = 2.25 minuts

NOM:__________________________COGNOM:____

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

Problema 3 (B3)

En una determinada marca de roba estan estudiant la quantitat de peces de roba de les comandes que reben online a través del nou web. En concret estan estudiant la quantitat de comandes que només compren una peça de roba i han trobat que, de mitjana, aquestes representen un 25% de les comandes.

Es defineix la variable aleatòria X: "nombre de comandes amb una sola peça de roba comprada". Si en una hora concreta han rebut 20 comandes,

1) Quina és la probabilitat que cap de les 20 comandes hagi estat d'una sola peça de roba? (explicita la distribució que segueix la variable X) (1 punt)

X~Bin(20, 0.25)
P(X=0) = $\binom{20}{0}\cdot 0.25^0\cdot(1-0.25)^{20}$=0.00317

2) Quina és la probabilitat que s'hagin rebut més de quatre comandes amb una sola peça de roba? (1 punt)

P(X>4) = 1- P(X≤3) =1 – 0,2252 = 0,7748 (TAULES)

3) Quin és el nombre esperat de comandes fins que arribi una comanda amb una sola peça de roba? (explicita la distribució que segueix la variable que definiu) (1 punt)

X1: "Nombre de comandes rebudes fins a rebre una comanda amb una sola peça de roba"
X1~Geom(0.25)
E(X1) = 1/0.25 = 4
El valor esperat per rebre una comanda amb una sola peça de roba és de 4 comandes.

A partir de l'estudi realitzat han recollit que la despesa de les comandes que reben segueixen una distribució normal amb μ=50€ i σ=15 €. Per incentivar les vendes online volen procurar no cobrar despeses d'enviament, però només els hi surt a compte per les comandes de més de 20€.

4) Quina és la probabilitat que una comanda sigui per un import menor de 20€? (1 punt)

Es defineix V: "despesa de les comandes online", V~N(50, 15)
P(V<20) = P(Z < $\frac{20-50}{15}$) = P(Z < −2) = 1- P(Z < 2) = 1 – 0.97725 = 0.2275

De 8:00 del matí a 8:00 de la nit dels dies laborables, l'empresa compta amb un servei personalitzat d'atenció a les comandes. El nombre de trucades que reben al servei de comandes segueix una llei de Poisson amb una mitjana 3 trucades per hora.

5) Quina és la probabilitat que en una hora es rebin dues trucades? (1 punt)

N: "nombre de trucades al servei personalitzat", N~Poiss(3)
P(N=2) = $\frac{3^2\cdot e^{-3}}{2!}$ = 0.2240

6) Quina és la desviació poblacional del nombre de trucades que es reben al servei? (0,5 punts)

Var(N) = 3 i per tant la desviació és de 1,73 trucades

Considerant que la quantitat de trucades rebudes al Servei personalitzat d'atenció a les comandes segueixen una distribució de Poisson amb una mitjana de 3 trucades per hora, es defineix la variable aleatòria T: "temps transcorregut entre dues trucades consecutives al servei".

6) Quina és la probabilitat que el temps transcorregut entre dues trucades sigui més petit que 30 minuts? (Explicita la distribució de la variable aleatòria T) (1 punts)

T~exp(3)
P(T<0.5) = 1 – e $^{-3\cdot 0.5}$ = 0.7769

7) Quin és el temps esperat entre dues trucades consecutives al servei? (0’5 punts)

E(T)= 1/3 hores, per tant, uns 33 minuts.

Es vol estudiar el conjunt de les trucades que es reben al Servei d’atenció personalitzat durant les 12 hores en què està operatiu.

8) Justifica quines condicions ha de complir la situació plantejada per tal que es pugui utilitzar el Teorema Central del Límit per a realitzar l’estudi desitjat. (1 punt)

Es consideren 12 variables aleatòries M_i = “nombre de trucades rebudes al servei d’atenció personalitzat” on totes segueixen la mateixa distribució, M_i ~ Poiss (3), per tant són idènticament distribuïdes. També cal considerar que són independents.

9) Determina el nombre màxim de queixes que es reben en un dia laborable de 8 del matí a 8 del vespre al Servei d’atenció personalitzat amb un error del 5%. (2 punts)

Es considera $S_{12}= M_1+...+M_{12}$ amb M_i independents i idènticament distribuïdes tal i com s’ha justificat a l’apartat anterior. Pel Teorema Central del Límit es té que S_{12} ~ N(12· 3 , $\sqrt{12 \cdot 3}$) = N(36,6)

$P(S_{12} \leq k) = 0.95$
P($Z \leq \frac{k-36}{6}$) = 0.95, amb Z ~N(0,1).
A partir de les dades de la taula., s’obté que: $\frac{k-36}{6}$ = 1.645, per tant k= 45.87

Amb un error del 5% es rebran durant les 12 hores d’atenció un màxim de 46 rucades al Servei d’atenció personalitzat

Problema de B4

1. Es vol dissenyar una nova app de pagament per mòbil que sigui un comparador de preus de productes informàtics a diferents portals (Amazon, Ali Express, eBay). Per saber quina acceptació tindria l'app al mercat s'ha fet una enquesta a 200 usuaris potencials. En aquesta enquesta es pregunta als clients quants diners estarien disposats a pagar per un producte informàtic concret (mateix per tots els enquestats). Es crea una variable de rendiment de l'app (R) que es la diferència entre el preu que pagaria el client (p_C) i el preu que oferiria l'app ($p_A = 500€$). Els creadors volen contrastar la hipòtesi de si el preu que pagarien els clients és igual al preu que oferiria l'app.

`mean(R,na.rm=TRUE)`	`sd(R,na.rm=TRUE)`
5.71	43.57

a) Planteja el contrast d'hipòtesis i digues explícitament si el contrast és unilateral o bilateral. (1 punt)

$$\begin{cases} H_0: \mu_R = 0 \\ H_1: \mu_R \neq 0 \end{cases}$$

El contrast seria **bilateral** ja que l'enunciat demana si són iguals o no.

b) Digues i) quin estadístic empraries; ii) fes el seu càlcul, iii) la distribució que segueix sota la hipòtesi nul·la i les premisses necessàries; i iv) digues quin o quins són el punt/s crític/s amb un nivell de significació α=5%. (1 punt)

Opció 1 correcta (premissa R~N)

$t = \frac{\bar{R}}{s_R/\sqrt{n}} = \frac{5.71}{43.57/\sqrt{200}} = \mathbf{1.85}$ Punts crítics = $[t_{0.025,199}, t_{0.975,199}] \approx \pm\mathbf{1.97}$

Opció 2 correcta (premissa n>100)

$Z = \frac{\bar{R}}{\frac{s_R}{\sqrt{n}}} = \frac{5.71}{\frac{43.57}{\sqrt{200}}} = \mathbf{1.85}$ Punts crítics =$[Z_{0.025}, Z_{0.975}] \approx \pm\mathbf{1.96}$

c) Conclou sobre la prova d'hipòtesi. (1 punt)

Com que l'estadístic és menor al punt crític, **no hi ha evidència per rebutjar la hipòtesi nul·la** i, per tant, no hi ha evidència per rebutjar que el preu mitjà que pagarien els usuaris serà diferent del que oferirà l'app.

d) Els creadors de l'aplicació, no saben massa d'estadística i no tenen clar, en el cas que la variable R no complís les premisses amb quina de les següents variables podrien dur a terme el contrast d'hipòtesi:

$$R' = |p_C - p_A| \qquad R'' = \log(p_C) - \log(p_A) \qquad R''' = \log(p_C - p_A)$$

Quina/es de les 3 variables et semblarien adients per dur a terme el contrast d'hipòtesis? Raona la resposta (1 punt)

Les variables R' i R''' es descartarien. La variable R' perquè no té sentit agafar el valor absolut si volem testar que la p_C és diferent a p_A en mitjana ja que tots els valors es decantarien cap al costat positiu. La variable R''' no es pot agafar ja que la diferència $p_C - p_A$ pot prendre valors negatius i, per tant, no es podria fer el logaritme. **La única opció raonable seria R''**.

2. Un altre pregunta important de l'enquesta és si estarien interessats en adquirir aquesta app en el futur. L'enquesta s'ha realitzat emprant quotes per gènere de tal manera que s'han obtingut els resultats per 100 dones i 100 homes. 32 dones i 18 homes han manifestat el seu interès en l'aplicació. Es decidirà treure l'app al mercat si el percentatge de gent interessada (dones o homes) és superior al 20% amb un nivell de significació α=5%.

a) Planteja el contrast de hipòtesis i digues explícitament si és unilateral o bilateral (1 punt)

$$\begin{cases} H_0: \pi = 0.20 \\ H_1: \pi > 0.20 \end{cases}$$

És **unilateral** perquè volem veure si la proporció de gent interessada és superior a 0.20.

b) Digues quin estadístic emprar, les premisses i la distribució que seguirà sota la hipòtesi nul·la (1 punt)

Estadístic: $\hat{\boldsymbol{Z}} = \frac{\boldsymbol{p-\pi_0}}{\sqrt{\frac{\boldsymbol{\pi_0 \cdot (1-\pi_0)}}{\boldsymbol{n}}}}$

Premisses: $\boldsymbol{n \cdot \pi_0 \geq 5, n \cdot (1-\pi_0) \geq 5}$**, m.a.s., independents**

Distribució: **N(0,1)**

c) Calcula l'estadístic i el punt crític (1 punt)

$$\hat{Z} = \frac{0.25 - 0.20}{\sqrt{\frac{0.20 \cdot (1-0.20)}{200}}} = \mathbf{1.77}$$

$$Punt\ crític = \mathbf{1.64}$$

d) Conclou sobre la prova d'hipòtesi. Trauran l'app al mercat? (1 punt)

Es **pot rebutjar la hipòtesi nul·la** i, per tant, acceptar, que la hipòtesi alternativa és certa. Podran treure l'app al mercat.

3. La pregunta de l'apartat 2 sobre si els usuaris estarien interessats en comprar l'app és clau i els creadors voldrien estimar la proporció poblacional de gent interessada en l'app amb un error de ±0.05 i una confiança del 90%. Digues quina hauria de ser la grandària mostral de l'enquesta per assolir aquest objectiu (2 punts)

El IC90% d'una proporció poblacional és:

$$IC(\pi, 1-\alpha) = p \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Donat que volem un 90% de confiança l'error (E) serà:

$$Opció\ 1 \rightarrow E = Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 1.645 \cdot \sqrt{\frac{0.5 \cdot 0.5}{n}} = 0.05 \rightarrow n = \frac{1.645^2 \cdot 0.5 \cdot 0.5}{0.05^2} \approx 270.6 \rightarrow n = \mathbf{271}$$

$$Opció\ 2 \rightarrow E = Z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 1.645 \cdot \sqrt{\frac{0.25 \cdot 0.75}{n}} = 0.05 \rightarrow n = \frac{1.645^2 \cdot 0.25 \cdot 0.75}{0.05^2} \approx 202.9 \rightarrow n = \mathbf{203}$$

A tots els exercicis: justifiqueu formalment les passes de les respostes

Problema B5

Un grup d'estudiants de PE es planteja estudiar les diferències de valoracions de jocs de mòbil gratuïts i de pagament. Amb aquest objectiu recullen les valoracions (sobre una escala de 0 a 100) d'una pàgina especialitzada en Internet de 31 jocs gratuïts i de pagament, respectivament. La taula a continuació mostra els indicadors numèrics més rellevants dels valors en ambdues mostres.

	n	Mitjana	Mediana	Desv. est.	Mínim	Màxim
Jocs de pagament	31	78,8	79,5	7,1	66,6	95,5
Jocs gratuïts	31	73,7	72,6	8,7	55,8	93,2

Definim com a variable V la valoració dels jocs i utilitzem com a subíndexs p i g per jocs de pagament i gratuïts, respectivament.

(a) (0.5 punts)
Es tracta de dues mostres independents o aparellades? Raoneu la resposta.

Solució:
Són mostres **independents** perquè es tracta de 62 jocs diferents, 31 gratuïts i 31 de pagament.

(b) (0.5 punts)
Només en base als valors descriptius, creieu que la variable V segueix una distribució normal en les dues mostres?

Solució:
El fet que les medianes i les mitjanes són gairebé iguals en les dues mostre fa pensar que la distribució de V sigui simètrica i per tant podria ser d'una distribució normal. Per altra banda, una distribució normal no té cap màxim, mentre les valoracions sí en tenen (100).

(c) (2.5 punts)
Abans d'estudiar si les valoracions mitjanes són diferents d'un tipus a un altre dels jocs de mòbil, comprovem si es compleix la igualtat de variàncies o no.

- Plantegeu la hipòtesi nul·la i l'alternativa **(0.5)**:

  Solució:

$$\mathrm{H}_0\colon \sigma_p^2 = \sigma_g^2 \quad \text{vs.} \quad \mathrm{H}_1\colon \sigma_p^2 \neq \sigma_g^2$$

- Calculeu el valor de l'estadístic de contrast **(0.6)**:

  Solució:

$$f = \frac{8.7^2}{7.1^2} = 1.5, \text{ on } F = \frac{s_g^2}{s_p^2} \overset{\mathrm{H}_0}{\sim} F_{30;30}.$$

- Resolució del contrast utilitzant un nivell de significació $\alpha = 0.1$ **(1)**:

  Solució:
  El quantil 0.95 de la distribució $F_{30;30}$ és 1.84, que és més gran que 1.5. Per tant, no rebutgem la hipòtesi nul·la.

- Interpretació pràctica **(0.4)**:

  Solució:
  Podem suposar que es compleix la homocedasticitat, és a dir la igualtat de variàncies.

(d) **(3 punts)**
Contrasteu ara la hipòtesi que les valoracions mitjanes de jocs de mòbil de pagament i gratuïts són iguals o no.

- Plategeu la hipòtesi nul·la i l'alternativa **(0.5)**:

  Solució:

  $$\mathrm{H}_0\colon \mu_p = \mu_g^2 \quad \text{vs.} \quad \mathrm{H}_1\colon \mu_p^2 \neq \mu_g^2$$

- Calculeu el valor de l'estadístic de contrast **(1)**:

  Solució:

  $$t = \frac{78.8 - 73.7}{\sqrt{0.5 \cdot (7.1^2 + 8.7^2) \cdot 2/31}} = 2.53, \text{ on } T \text{ segueix una distribució } t \text{ amb 60 graus de llibertat.}$$

- Resolució del contrast utilitzant un nivell de significació $\alpha = 0.05$ **(1)**:

  Solució:
  El quantil 0.975 de la distribució t_{60} és 2, que és més petit que 2.53. Per tant, sí podem rebutjar la hipòtesi nul·la.

- Interpretació pràctica **(0.5)**:

  Solució:
  Com rebutgem la hipòtesi i com $\bar{x}_p > \bar{x}_g$, podem afirmar que, **en mitjana**, jocs de mòbil de pagament són millor valorats que els jocs gratuïts.

(e) **(2.5 punts)**
Hem fet el càlcul del interval de confiança del 95% de la diferència de les valoracions mitjanes amb R i hem obtingut el següent resultat:

$$\mathrm{CI}(\mu_p - \mu_g; 0.95) = [1.12,\ 9.17].$$

Utilitzeu aquest resultat per calcular l'interval de confiança del 90% de la diferència de les valoracions mitjanes.

Solució:
El marge d'error del $\mathrm{CI}(\mu_p - \mu_g; 0.95)$ és $(9.17 - 1.12)/2 = 4.025$ i es calcula com $t_{60;0.975} \cdot s \cdot \sqrt{2/31}$, on s és la desviació estàndard combinada. Llavors, el marge d'error del $\mathrm{CI}(\mu_p - \mu_g; 0.9)$ és

$$4.025 \cdot \frac{t_{60;0.95}}{t_{60;0.975}} = 3.362$$

i, per tant

$$\mathrm{CI}(\mu_p - \mu_g; 0.9) = 78.8 - 73.7 \mp 3.362 = [1.74,\ 8.46].$$

(f) **(1 punt)**
Expliqueu cóm es pot utilitzar $\mathrm{CI}(\mu_p - \mu_g; 0.95)$ per contrastar la hipòtesi

$$\mathrm{H}_0\colon \mu_p = \mu_g \quad \text{vs.} \quad \mathrm{H}_1\colon \mu_p \neq \mu_g.$$

Solució:
Si $\mathrm{CI}(\mu_p - \mu_g; 0.95)$ no conté 0, es pot rebutjar $\mathrm{H}_0\colon \mu_p = \mu_g$ utilitzant un nivell de significació $\alpha = 0.05$.

NOM:____________________________COGNOM:____

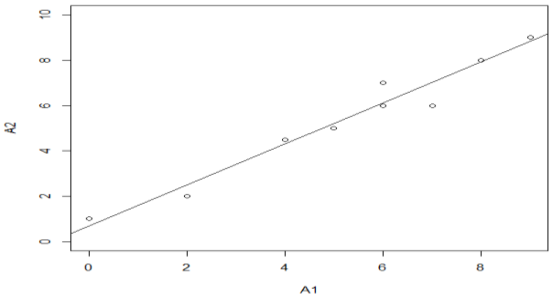
(*Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs*)

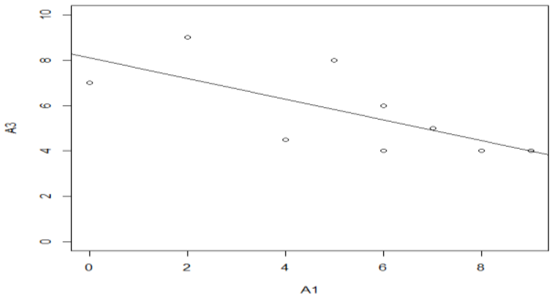
Problema de B6

Volem comparar una possible relació lineal entre les notes d'una assignatura A1 amb unes altres dues A2 i A3. Disposem de les següents dades i resultats en 10 dels seus estudiants:

Notes A1:	0	2	4	5	6	6	7	8	9	9	$\sum A1_i = 56$	$\sum A1_i^2 = 392$	
Notes A2:	1	2	4.5	5	6	7	6	8	9	9	$\sum A2_i = 57.5$	$\sum A2_i^2 = 397.25$	$\sum A1_i A2_i = 393$
Notes A3:	7	9	4.5	8	4	6	5	4	4	4	$\sum A3_i = 55.5$	$\sum A3_i^2 = 339.25$	$\sum A1_i A3_i = 275$

mean(A1)=5.6 sd(A1)=2.95 mean(A2)=5.75 sd(A2)=2.72 mean(A3)=5.55 sd(A3)=1.86 cov(A1,A2)=7.89 cov(A1,A3)= -3.98






```
lm(formula = A2 ~ A1)
Coefficients:
              Estimate    Std. Error    t value    Pr(>|t|)
(Intercept)   ______        0.3813        1.78     0.113
A1            ______        0.0609       14.87     0.0000004

Residual standard error: 0.5393 on 8 degrees of freedom
Multiple R-squared:  0.9651,    Adjusted R-squared:  0.9607
F-statistic: 221.1 on 1 and 8 DF,  p-value: 4.124e-07
```

1) Per comparar les rectes ajustades en els dos models indiqueu:

a) els valors dels coeficients ordenada a l'orígen i del pendent de la recta ajustada en cada model (2 punts)

Per A2=b0+b1*A1, dels resultats de lm: **b0=**0.3813*1.78 = **0.679** i **b1=**0.0609*14.87 = **0.906**

(o bé calucular r_{A1A2} = cov(A1,A2)/ $S_{A1}S_{A2}$ = 7.89/(2.95*2.72) = 0.98, i llavors
$b1=r_{A1A2}$ (S_{A2}/S_{A1})=0.98 (2.72/2.95)=0.90 b0=mean(A2) – b1*mean(A1)=5.75 – 0.90*5.6=0.71

Per A3=b0+b1*A1, **b1=** r_{A1A3} (S_{A3}/S_{A1}) =-0.72 (1.86/2.95)=**-0.45** i **b0 =** mean(A3) – b1*mean(A1) = 5.55 + 0.45*5.6 =**8.1**
havent calculat r_{A1A3} = cov(A1,A3)/ $S_{A1}S_{A3}$ = -3.98/(2.95*1.86) = -0.72

b) les equacions de les rectes estimades amb la seva interpretació (2 punts)

A2 = 0.7 + 0.9 * A1
b0 igual a 0.7 indica l'ordenada a l'orígen, és a dir quan a l'assignatura A1 la nota és 0 a l'assignatura A2 la nota és 0.7
b1 igual a 0.9 indica el pendent, és a dir per cada unitat que augmenta la nota de A1 la de A2 augmenta en 0.9

A3 = 8.1 – 0.45 * A1
b0 igual a 8.1 indica l'ordenada a l'orígen, és a dir quan a l'assignatura A1 la nota és 0 a l'assignatura A2 la nota és 8.1
b1 igual a -0.45 indica el pendent, és a dir per cada unitat que augmenta la nota de A1 la de A2 disminueix en 0.45

2) Pel cas del model entre A1 i A2, calculeu un interval de confiança amb risc 5% per al pendent de la recta, i interpreteu-lo (2 punts)

$IC(95\%,\beta_1) = b_1 \pm t_{n-2,0.975} \cdot S_{b1} = 0.9 \pm 2.31 * 0.0609 = 0.9 \pm 0.14 = [0.76, 1.04]$

Amb un 95% de confiança el valor del pendent està entre 0.76 i 1.04.
A més a més, 1 és un valor que pertany a l'interval, per tant amb un 95% de confiança el pendent de la recta és d'una unitat indicant que cada punt més en la nota de A1 correspon a un punt més de nota de A2

3) Indiqueu i compareu pels dos models els valors del coeficient de correlació i el de determinació (2 punts)

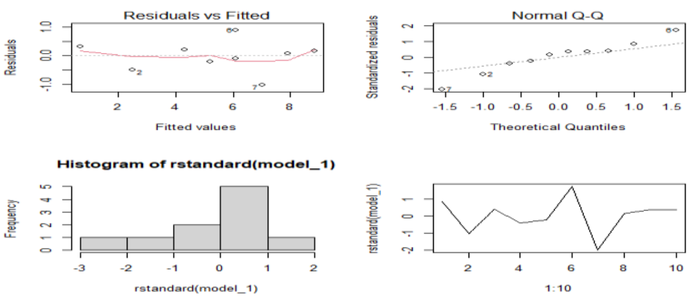
En la relació entre A1 i A2
el coeficient de correlació és 0.98 (calculat abans)
el coeficient de determinació $R2 = (0.98)^2 = 0.96$ (dels resultats de lm en R)

En la relació entre A1 i A3
el coeficient de correlació és -0.72 (calculat abans)
el coeficient de determinació $R^2 = (-0.72)^2 = 0.52$

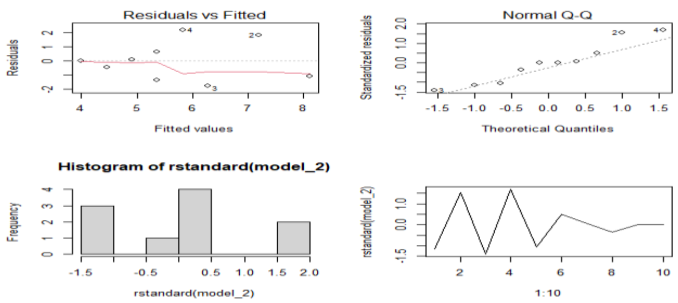
A1 i A2 tenen una relació positiva (correlació positiva) i intensa (coef det 96%)
A1 i A3 tenen una relació negativa (correlació negativa) i no tant intensa (coef det 52%)

4) Enuncieu les premisses de la regressió lineal i compareu-les en els dos models (indiqueu els gràfics d'on es dedueixen) (2 punts)

Model lineal entre A1 i A2



Model lineal entre A1 i A3



Les premisses són: linealitat (en el gràfic Y-X inicial si s'observa que les variables es disposen en forma de recta), normalitat (en l'histograma de residus i el Normal Q-Q s'observa si l'histograma té forma de campana i si en el Normal Q-Q els punts queden alineats), i independència i homocedasticitat (en els altres dos gràfics dels residus si s'observa que no hi ha cap patró que relacioni els punts i que no es disposen en "zones" amb variabilitat diversa)

En els dos casos s'observa força linealitat (una positiva i l'altra negativa)
En els dos casos s'observa força normalitat (en el model A1-A2 es veu millor en l'histograma i en el model A1-A3 millor en el Normal Q-Q)
En els dos casos hi ha independència doncs no s'observa cap patró de relació entre els punts
El model entre A1 i A2 és més homocedàstic, ja que en el model entre A1 i A3 hi ha primer una zona amb més variabilitat i després disminueix