NOM:_____COGNOM:_____

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

Problema 1 (B1-B2)

1. a) Un determinat sistema pot ser infectat pel Virus A. Aquest virus pot arribar a través del correu (C) o a través de la xarxa (X). El Virus A pot arribar a través del correu amb una probabilitat de 0'3 i amb una probabilitat de 0'4 de fer-ho través de la xarxa. A més a més, hi ha una probabilitat de 0'15 que arribi de manera simultània a través del correu i de la xarxa. Quina és la probabilitat que el virus A no entri en el sistema? (1 p)

   Si considerem C='El virus A arriba a través del correu' i X='El virus A arriba a través de la xarxa'.
   Tenim que $P(C)= 0'3$; $P(X) =0'4$ i $P(C \cap X)=0'15$

   Volem calcular $P(\overline{C} \cap \overline{X}) = P(\overline{C \cup X}) = 1 - P(C \cup X) = 1- [P(C) + P(X) – P(C \cap X)] = 1- (0'3 + 0'4 –0'15) = 0'45$

   Hi ha una probabilitat de 0'45 que el virus A no entri en el sistema.

   1b) Per procurar evitar la infecció pel virus B es contracta un sistema per detectar la seva entrada. En aquest cas el virus pot arribar només a través del correu o a través de la xarxa (no pot fer-ho de manera simultània). Pot fer-ho a través del correu amb una probabilitat de 0'6 i per la xarxa amb una probabilitat de 0'4. Si el virus B entra pel correu el sistema el detecta amb una probabilitat de 0'7. I si entra per la xarxa, aleshores el sistema detecta el virus amb una probabilitat del 0'8. Quina és la probabilitat que el virus B sigui detectat? (1 p)

   Si considerem C_B ='El virus B arriba a través del correu' i X_B ='El virus B arriba a través de la xarxa'.

   $P(\text{Detectat}) = P(\text{Detectat}| C_B) \cdot P(C_B) + P(\text{Detectat}| X_B) \cdot P(X_B) = 0'7 \cdot 0'6 + 0'8 \cdot 0'4 = 0.74$

   Per tant, hi ha una probabilitat de 0'74 que el virus B sigui detectat pel sistema contractat.

2. a) Tenim una carpeta amb dos fitxers. En aquest cas un virus C pot corrompre el primer fitxer amb una probabilitat de 0'5. De manera independent, el segon fitxer pot ser corromput pel virus C amb una probabilitat de 0'4. Considerem la variable aleatòria N:"Nombre de fitxers corromputs pel virus C". Calculeu la funció de probabilitat de la variable N. (1 punt)

   $P(N=0) = 0'5 \cdot 0'6 = 0'3$; $P(N=1) = 0'5 \cdot 0'6 + 0'5 \cdot 0'4 = 0'5$; $P(N=2) = 0'5 \cdot 0'4 = 0'2$

n	P_N (n)
0	0'3
1	0'5
2	0'2

   2b) Tenim una segona carpeta amb tres fitxers. En aquest cas el virus D pot corrompre el primer fitxer amb una probabilitat de 0'2, el segon (de manera independent) pot ser corromput amb una probabilitat de 0'3 i finalment el tercer fitxer (també de manera independent) té una probabilitat de 0'4 de ser infectat. Considerant ara la variable aleatòria M:"Nombre de fitxers corromputs pel virus D", calculeu la funció de probabilitat de M. (1 punt)

   $P(M=0) = 0'8 \cdot 0'7 \cdot 0'6 = 0'336$; $P(M=1) = 0'2 \cdot 0'7 \cdot 0'6 + 0'8 \cdot 0'3 \cdot 0'6 + 0'8 \cdot 0'7 \cdot 0'4 = 0'452$; $P(M=2) = 0'2 \cdot 0'3 \cdot 0'6 + 0'2 \cdot 0'7 \cdot 0'4 + 0'8 \cdot 0'3 \cdot 0'4 = 0'188$; $P(M=3)= 0'2 \cdot 0'3 \cdot 0'4 = 0'024$

m	P_M (m)
0	0'336
1	0'452
2	0'188
3	0'024

2c) Calculeu l'esperança de M (0'5 punts)

E(M)= 0 ·0'336 + 1·0'452 + 2·0'188 + 3·0'024 = 0'9

2d) Calculeu la variància i la desviació típica de N. (1 p)

E(N)= 0 ·0'3 + 1·0'5 + 2·0'2 = 0'9

$E(N^2)= 0^2 \cdot 0'3 + 1^2 \cdot 0'5 + 2^2 \cdot 0'2 = 1'3$

$Var(N) = E(N^2) - E(N)^2 = 1'3 - 0.9^2 = 0'49$

$\sigma_N = 0'7$

3. El temps, en minuts, que triga un sistema en reiniciar-se es pot modelar amb una variable aleatòria contínua, T, amb la següent funció de densitat:

$$f(t) = \begin{cases} a(10-t)^2, & 0 < t < 10 \\ 0 & , \ altrament \end{cases}$$

3a) Calculeu *a* perquè f(t) sigui una funció densitat (2 p)

S'ha de complir que $\int_{-\infty}^{+\infty} f(t)dt = 1$.

En aquest cas, tenim que $\int_0^{10} a(10-t)^2 dt = [-a\frac{(10-t)^3}{3}]_0^{10} = \frac{1000a}{3} = 1$

Per tant, a = 0'003

Per aquest valor d'a, la funció $f(t) = 0'003 \cdot (10-t)^2$ assoleix sempre valors positius

3b) Calculeu la funció de distribució de T (1'5 p)

Calculem $\int_0^t 0'003(10-z)^2 dz = [-0'003\frac{(10-z)^3}{3}]_0^t = -\frac{(10-t)^3}{1000} + 1$

$$F_T(t) = \begin{cases} 0, & t < 0 \\ 1 - \frac{(10-t)^3}{1000}, & 0 < t < 10 \\ 1 & , \ t > 10 \end{cases}$$

3c) Calculeu el temps esperat d'aquest sistema per reiniciar-se (1 p)

$\int_0^{10} 0'003(10-t)^2 \cdot t \, dt = 2'5$

NOM:_____COGNOMS:_____

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

Problema 2 (B3-B4)

A les *smart cities* hi ha milers de sensors per mesurar el trànsit del carrer, i abundància de dades per a analitzar. Per exemple, sobre l'entrada de vehicles per una gran avinguda de Barcelona entre 8 i 9 del matí se sap que:

- supera els 6.58 milers de vehicles el 14% dels dies,
- no arriba a 3.65 milers de vehicles el 12% dels dies.

1. Assumint que aquesta entrada segueix un model Normal, determineu el nombre esperat μ i la desviació tipus σ (en milers de vehicles/hora). *Pista: transforma les dues condicions anteriors en dues equacions lineals i resol el sistema.*

   P(X < 3.65) = 0.12, i P(X > 6.58) = 0.14, o P(X < 6.58) = 0.86. *Estandaritzant*:
   P(Z < (3.65 – μ)/σ) = 0.12, i P(Z < (6.58 – μ)/σ) = 0.86. *I prenent els percentils de la N(0,1):*
   (3.65 – μ)/σ = $Z_{0.12}$ = -1.175 *(complementari de 0.12, 0.88, està entre 1.17 i 1.18)*
   (6.58 – μ)/σ = $Z_{0.86}$ = 1.08
   Resposta: μ =5.176, σ =1.299

2. Posteriorment, la distribució del nombre de vehicles segueix sent Normal, però amb diferent valor dels paràmetres:

Hora	**9-10**	**10-11**	**11-12**	**12-13**	**Suma**	**Suma de quadrats (variància)**
Valor esperat	4.60	3.95	3.14	2.56	14.25	
Desviació tipus	1.15	0.94	0.83	0.92		3.7415

   Suposant que les diferents trams horaris són independents, deduïu la distribució del nombre de vehicles que entren en el període 9am-1pm, i calculeu la probabilitat que un matí qualsevol es detectin més de 16 mil vehicles en aquestes 4 hores.

   Resp: X_4 ~ N(14.25, 1.9343). Prob = P(X_4 > 16) = 0.1828

3. Entre 2 i 3 de la tarda, el temps que passa entre dos vehicles que entren consecutivament per aquesta avinguda segueix un model Exponencial amb mitjana 2 segons. Amb aquesta informació, trobeu la probabilitat que en un període de 10 segons d'aquesta franja comptem menys de 4 vehicles entrant-hi. Digueu també quina serà la variància del nombre de vehicles que hi entren entre 2 i 3 de la tarda. Raoneu les respostes.

   Resp: taxa per segon: λ= ½. Per a 10 segons: N_{10} ~ Poisson(5), P(N_{10} < 4) = F(3) = 0.265. Taxa per 3600 segons: 1800. Com per a variables Poisson la taxa també és la variància, resposta: 1.8 $(x1000)^2$

4. Una càmera especial pot comptar el nombre d'ocupants als vehicles. Es creu que un 10% dels turismes porta 3 o més ocupants a dins. Si es pren una mostra de 40 turismes a l'atzar, quina és la probabilitat que 8 turismes portin almenys 3 ocupants? Justifiqueu la resposta.

   Resp: Y ~ B(40, 0.1), P(Y=8) = 76904685 0.1^8 0.9^32 = 0.0264
   Per Poisson, amb λ=4, també surt semblant: 0.02977

5. Per cada 1000 turismes, trobeu una fita superior (amb risc d'error del 5%) per al nombre de turismes portant 3 o més ocupants. Si fossin 5000 turismes els que s'exploren, la fita amb el mateix risc també seria 5 vegades més gran? Contesteu sense fer càlculs.

   Resp: N ~ B(1000, 0.1), aproximem per TCL a N(100, $\sqrt{1000\ 0.1\ 0.9}$) = N(100, 9.49). Percentil 0.95 = 100 + $z_{0.95}$*9.49 = 115.6

   Amb 5000 turismes la fita no seria 5 vegades més gran perquè la desviació no és 5 vegades major, la desviació és menor i la fita quedaria per sota de 5 x 115.6; concretament, 535.

6. La limitació de velocitat a l'entrada de Barcelona és de 50 km/h, però a certes hores quan el trànsit és fluid sembla que els vehicles corren més del permès. Volem estudiar si es pot afirmar que la velocitat mitjana entre 2 i 3 de la tarda està per sota dels límits de velocitat. 287 vehicles escollits a l'atzar un determinat dia a aquesta franja han donat una velocitat mitjana de 48.9 km/h, i una desviació tipus mostral de 8.4 km/h. Amb un risc del 1% digueu quina conclusió es pot extreure de l'estudi, justificant breument la resposta.

Es unilateral, per l'esquerra ("... per sota dels límits..."). Estadístic t = (48.9 – 50)/8.4/sqrt(287) = -2.2184, però el punt crític al 1% és -2.33. No es pot rebutjar la hipòtesi nul·la (la velocitat mitjana és 50 o més).

7. Els turismes amb 3 o més ocupants tenen dret a circular per un carril VAO. El fet és que també es colen alguns vehicles que no arriben a aquesta ocupació. Es vol determinar la proporció de vehicles que cometen la infracció al carril VAO. Amb aquest objectiu, s'observen 400 vehicles a l'atzar dels que hi circulen i s'observa que 320 tenen almenys 3 ocupants. Calculeu un interval de confiança al 99% per la proporció d'infractors, i interpreteu.

P = 80/400 = 0.2, IC = $0.2 \pm z_{0.995}\sqrt{0.2(1-0.2)/400}$ = (0.1485 0.2515)

A partir de les dades, creiem que entre un 15% i un 25% dels conductors del carril VAO són infractors perquè no arriben a 3 ocupants, amb 99% de confiança.

Volem estudiar si una campanya de promoció del vehicle compartit ha tingut èxit incrementant la proporció de vehicles amb alta ocupació (en endavant, "la proporció"). S'ha observat a un tram normal d'autovia un total de 800 vehicles a l'atzar, dels quals 95 tenien 3 o més ocupants. Hem analitzat els resultats amb R, obtenint:

```
data:  95 out of 800, null probability 0.1
X-squared = 3.125, df = 1, p-value = 0.03855
alternative hypothesis: true p is greater than 0.1
95 percent confidence interval:
 0.1012102 1.0000000
```

(nota: X-squared és el estadístic X^2 de Pearson aplicat a una proporció, relacionat amb l'estadístic z ~ N(0,1) que utilitzem habitualment a aquesta prova: $\sqrt{X^2} \approx z$)

8a. Digueu com són les dues hipòtesis plantejades a aquesta prova:

H_0: π = 0.1
H_1: π > 0.1

8b. Justifiqueu la hipòtesi alternativa

Volem veure si s'ha incrementat la proporció de vehicles amb alta ocupació

8c. Digueu si les següents afirmacions són certes o no, i expliqueu per què:

- La probabilitat que la proporció sigui el 10% és 0.03855

És falsa perquè és incorrecte parlar de probabilitats per a dir si un paràmetre, com és π, pren un determinat valor. Un paràmetre es suposa constant, per tant no admet aleatorietat ni en conseqüència probabilitats.

- La probabilitat que la proporció observada ("mostral") valgui 0.1 és 0.03855

És falsa, la distribució de la proporció mostral és desconeguda perquè depèn d'un paràmetre desconegut. Inclús si s'assumeix H_0, es pot veure que la probabilitat d'obtenir exactament 80 vehicles de 800 tampoc és 0.03855.

8d. Finalment, expliqueu detalladament com es calcula el p-value (necessiteu les taules de la Normal).

L'estadístic és l'arrel de X^2, que assumirem que es distribueix com N(0,1) si la hipòtesi nul·la fos certa. És una prova unilateral cap a la dreta, per tant hem de trobar $P(Z > \sqrt{X^2}) = P(Z > 1.768) = 1 - 0.9616 = 0.0384$

8e. Doneu una conclusió global.

S'han trobat evidències febles de que la promoció pot haver incrementat la presència de vehicles amb 3 o més ocupants, però en tot cas l'extrem inferior de l'estimació (al 95% de confiança) continua estant molt a prop del 10%.

NOM:_____

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

Problema 3 (B5-B6).

Es pren una mostra de 15 estudiants de l'assignatura de PE a la FIB i s'analitza si el nombre d'hores (**H**) que han estudiat per a un examen afecta a les puntuacions que han obtingut (**P1**). A més, tenim les puntuacions d'un altre grup (**P2**).

Les dades són: **Hores (H)**: 2, 3, 3, 4, 4, 5, 5, 6, 6, 6, 7, 7, 7, 8, 8

Puntuació (P1): 5, 5, 6, 5, 7, 7, 8, 6, 9, 8, 7, 9, 10, 8, 9

Puntuació (P2): 4.4, 4.5, 5.3, 3.7, 7.3, 6.7, 6.8, 5.8, 8.2, 8, 7, 7.8, 8, 8.7, 8

i els estadístics obtinguts a partir d'aquestes tres variables són:

Mitjana de H: mean(H)=5.4 **Desviació tipus de H**: s_H=1.88

Mitjana de P1: mean(P1)=7. 27 **Desviació tipus de P1**: s_{P1}=1.62

Mitjana de P2: mean(P2)=6.68 **Desviació tipus de P2**: s_{P2}=1.58

Correlació entre H i P1: cor(H,P1)=0.78

A partir de les puntuacions P1 i P2 d'aquestes dades mostrals es creu que les mitjanes poblacionals podrien ser iguals. Plantegeu i resoleu la prova d'hipòtesi que permeti rebutjar o no aquesta hipòtesi**.** Les passes a seguir són:

1) (0.5 punts) Definiu la hipòtesi nul.la i l'alternativa per al test relacionat amb les mitjanes de les puntuacions P_1 i P_2

$H_0: \mu_{P1} = \mu_{P2}$

$H_1: \mu_{P1} \neq \mu_{P2}$

2) (1.5 punts) Calculeu el valor de l'estadístic d'aquesta prova d'hipòtesi explicitant les premisses que assumiu

Assumint distribució normal de les puntuacions i que les variàncies poblacionals desconegudes les podem considerar iguals

$s^2 = \frac{(15-1)S_1^2+(15-1)S_2^2}{15+15-2} = \frac{36.74+34.95}{28} = \frac{71.69}{28} = 2.56$

$\hat{t} = \frac{\overline{y_{p1}}-\overline{y_{p2}}}{S\sqrt{1/n1+1/n2}} = \frac{7.27-6.68}{1.6\sqrt{\frac{1}{15}+\frac{1}{15}}} = \frac{0.59}{1.6*0.37} = \frac{0.59}{0.59} = 1$

3) (1 punt) Indiqueu la conclusió i interpretació de la prova d'hipòtesi al nivell de significació α=0.05

Valor de la t (a taules): $t_{\upsilon=15+15-2,\alpha=0.975}$= 2.048 Com que $\hat{t}$ =1 < 2.048, no es pot rebutjar la H_0,

per tant no es pot rebutjar que les mitjanes poblacionals siguin iguals

4) (0.5 punts) Quin és el valor, en termes de valors absoluts, més petit de l'estadístic t pel qual la hipòtesi nul·la pot ser rebutjada?

$t_{\upsilon=28,\alpha=0.975}$= 2.048.

5) (0.5 punts) Indiqueu si les següents afirmacions són verdaderes o falses i justifiqueu la vostra resposta:

- Si es rebutja la hipòtesi nul.la H_0al nivell de 0.05, també podem rebutjar al nivell 0.1

Verdadera, perquè si hem pogut rebutjar al 5% és perquè el p-valor era inferior a 5% (i també serà inferior a 10%)

- Si el p-valor és igual a 0.15, podem rebutjar la hipòtesi nul·la al nivell del 10%

Falsa, ja que el 10% és menor que 0.15 i per tant no seriem a la zona de rebuig de la H_0

6.- (1 punt) Expliqueu com s'haurien de recollir les dades per a que fossin un conjunt de dades aparellades i com s'haurien de recollir si es vol treballar amb dos conjunts de dades independents

Quan tenim dues mostres de forma que els valors de cada mostra pertanyen al mateix individu, parlem de dades aparellades, és a dir, mesurem, per exemple, el que passa abans i després d'una experiència, com podria ser les notes obtingudes per un conjunt d'estudiants abans i després d'estudiar 8 hores diàries. Com exemple de dades independents en aquest cas seria comparar les notes dels estudiants d'un grup A amb les notes dels estudiants d'un grup B.

7) (1.5 punts) Plantegeu un model lineal i estimeu la recta de regressió que permet estimar la puntuació de **P1** a partir de les hores d'estudi **H**

$P1=\beta_0 + \beta_1 H+\varepsilon$

on β_0 i $\beta_1$son els paràmetres desconeguts i ε és el soroll, desconegut

La recta de regressió estimada és

$\widehat{P1}=b_0 + b_1 H$ i els paràmetres estimats són

$b_1=r\frac{s_{P1}}{s_H} = 0.78\frac{1.62}{1.88} = 0.67$

$b_0 = \overline{P1} - b_1\overline{H} = 7.27 -0.67*5.4=7.27-3.62 = 3.65$

Per tant, la recta de regressió estimada és:

$\widehat{P1}=3.65 + 0.67H$

8) (1 punt) Un alumne determinat ha estudiat 6.5 hores i ha tret un 7.5: calculeu el seu valor residual

$\widehat{P1}=3.65 + 0.67*6.5 = 8.007$

Valor residual= $7.5-8.007 = -0.507$

9) (1 punt) Calculeu i interpreteu el coeficient de determinació (R^2)

En el cas d'una regressió lineal simple, el coeficient de determinació $\boldsymbol{R^2}$ coincideix amb el coeficient de correlació al quadrat. En aquest cas serà $0.78^2 = 0.61$.

És una mesura de la qualitat de l'ajust en el model lineal. En aquest sentit, quan més a prop d'1 sigui el R^2, millor qualitat de l'ajust, en el sentit que els errors (diferència entre valor observat i el valor obtingut a la recta de regressió) siguin mínims.

10) (1.5 punts) Calculeu un interval de confiança, a un nivell de confiança del 95%, pel valor mitjà de la puntuació, sabent que les hores d'estudi han estat 5.5

$\widehat{P1}_h=3.65 + 0.67*5.5 = 7.34$

Interval de confiança al 95%

$\widehat{P1}_h \pm 2.16*1.05\sqrt{\frac{1}{15}+\frac{0.01}{49.48}} = 7.34 \pm 0.59 = [6.75, 7.93]$

1.05 prové de l'estimador no esbiaixat per a la variància dels errors d'aquest model

$\boldsymbol{s^2=\frac{\sum e_i^2}{n-2}=\frac{(n-1)s_{P1}^2(1-r^2)}{n-2}}=\frac{14*1.62^2*(1-0.78^2)}{13}=\frac{14*2.62*0.39}{13}=\frac{14.31}{13}=1.1$ (s=1.05)

0.01 prové de $(5.5 - 5.4)^2$

49.48 = S_H^2 (n-1) = 1.88^2 * 14