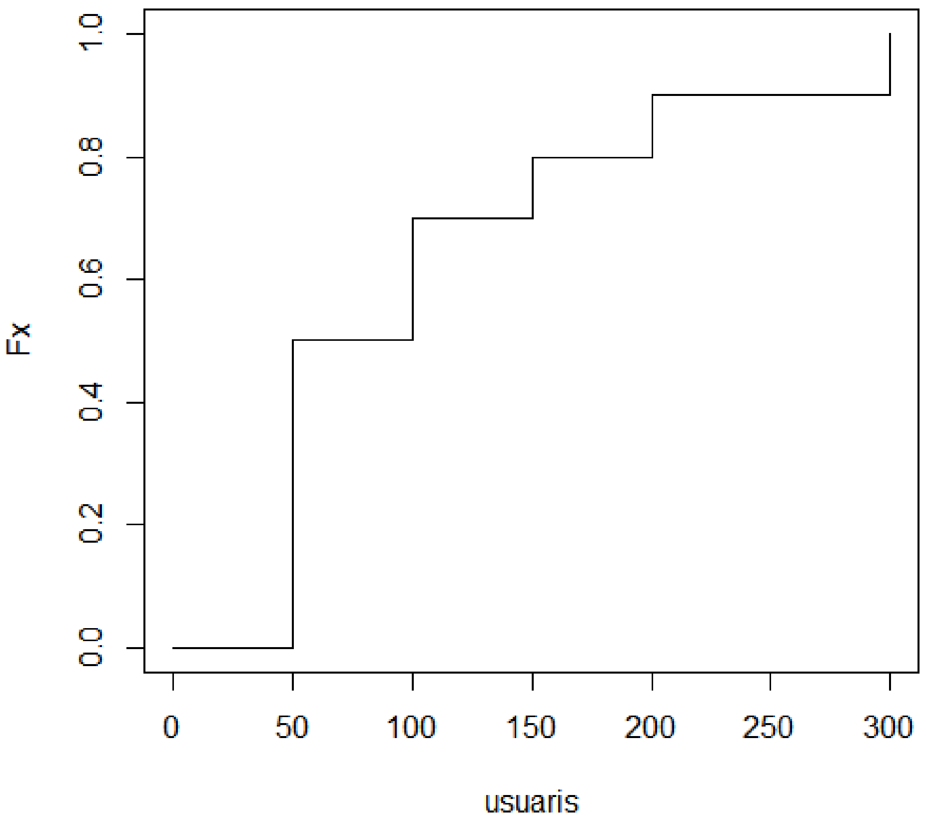
NOM: **COGNOM:**

Problema 1 (B1-B2)

S'està estudiant el canvi de màquina d'una web de serveis, i s'ha fet l'estudi del nombre d'usuaris connectats (U), amb els següents resultats: de cada 10 dies, durant 5 dies es dóna servei a 50 usuaris; durant dos dies hem tingut 100 usuaris; un altre dia a 150, un altre a 200, i el restant a 300 (per simplicitat, assumirem que aquests nombres són els únics nombres possibles per a U).

1. Considereu que s'agafa un dia a l'atzar i es comprova quants usuaris hi ha connectats aquest dia. Dibuixeu la gràfica de la funció de distribució del nombre d'usuaris. (1 punt)



2. Calcula i explica el significat de l'esperança i la variància de la variable aleatòria "nombre d'usuaris connectats". (2 punts)

Esperança = 110 El valor mig d'usuaris servits al llarg de 10 dies.

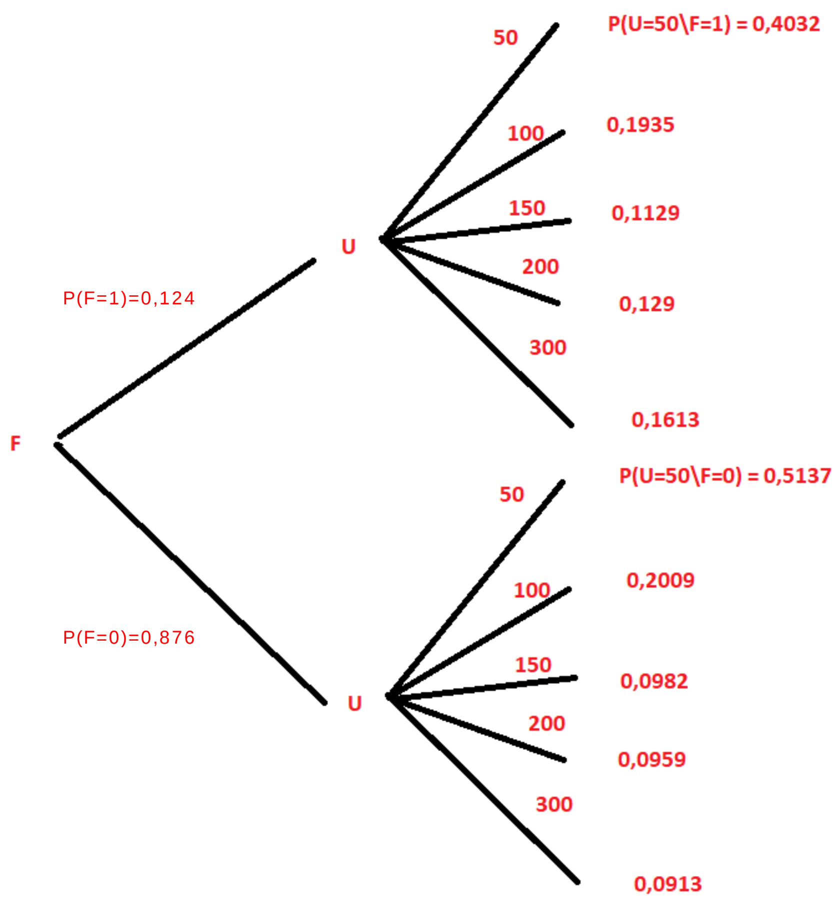
Variància = 6400 Mesura de dispersió que indica la mitjana de les desviacions entre observacions i valor esperat, al quadrat.

S'està considerant una màquina que té una probabilitat de fallida de un 10% si hi ha 50 usuaris i que s'incrementa en un 1% per cada 25 usuaris addicionals connectats –fins un màxim de 300. La variable nombre d'usuaris segueix el model de probabilitat descrit prèviament.

3. Escriviu la taula de probabilitats conjuntes de les variables aleatòries nombre d'usuaris (U) i fallida (F, 0 no falla; 1, sí falla), amb les seves marginals. Justifiqueu totes les passes. (1 punt)

	Fallida	F=1	F=0	
# Usuaris				
50		0,05	0,45	**0,5**
100		0,024	0,176	**0,2**
150		0,014	0,086	**0,1**
200		0,016	0,084	**0,1**
300		0,02	0,08	**0,1**
		0,124	**0,876**	1

4. Dibuixa l'arbre d'esdeveniments i probabilitats associat a l'experiència anterior, ficant al primer nivell la variable F. (1 punt)



5. Si es sap que el servidor ha donat error, quina és la probabilitat que hagi estat quan servia a 200 clients? (1 punt)

P (U=200\F=1) = P(U=200 i F=1) / P(F=1) = 0,129

6. Si sabem que no ha fallat, quina és la probabilitat que hi hagi com a màxim 150 usuaris connectats? (1 punt)

P(U<=150 i F=0) = P(U=50 i F=0) + P(U=100 i F=0) + P(U=150 i F=0)
P (U<=150\F=0) = P(U<=150 i F=0) / P(F=0) = 0,8128

7. Ara fem la següent consideració: les connexions es realitzen a l'inici del dia i si el servidor dóna algun error tots els usuaris connectats es queden sense servei. Defineix la variable aleatòria "nombre d'usuaris que es quedaran sense servei", i calcula el seu nombre esperat. (1.5 punts)

S = V.A. que indica el nombre d'usuaris que es queden sense servei.
P(S=0) = P(F=0) = 0,876
P(S=50) = 0,05
P(S=100) = 0,024, etc. (veure taula ex. 3)
Esperança = 0 × 0,876 + 50 × 0,05 + 100 × 0,024 + ... = 16.2

8. Un tècnic vol trobar un model molt més acurat sobre la distribució del nombre d'usuaris, utilitzant un model de variable contínua. Prenent els límits entre 0 i 300, el tècnic valora aquesta funció de densitat:

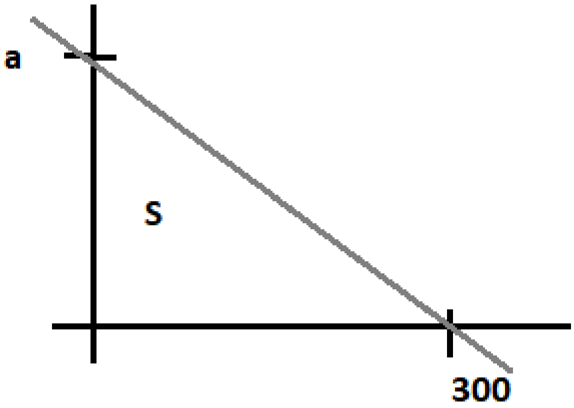
$$f(x) = a - bx; \qquad 0 < x < 300, \quad a > 0, \ b > 0$$

Sabent que $f()$ s'anul·la a l'extrem dret del domini, trobeu el valors de les constants a i b, i representeu gràficament la funció. Justifiqueu la resposta. (1.5 punts)

$f\,(300) = 0$

$f(0) = a - b*0 = a$

S = 1 = a*300/2
a = 0.0067
0.0067-b*300 = 0
b = 2.22*(10^-5)



També es pot resoldre igualant la integral entre 0 i 300 a 1.

Exercici 2

Un grup d'alumnes ha fet una enquesta entre els alumnes d'una assignatura del segon any del Grau d'Enginyeria Informàtica sobre l'ús que fan d'estatus. La pregunta vegada va ser quantes vegades van utilitzar l'estatus la setmana anterior i les respostes van ser les següents:

$$1, 2, 2, 4, 1, 4, 5, 3, 3, 0.$$

Definim, per tant, com a variable X el nombre d'exercicis d'estatus realitzats a la setmana.

(a) (0,5 punts)

Quines són les estimacions puntuals de la mitjana i de la desviació estàndard d'X?

Solució:

$$\hat{\mu}_X = \bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 2.5,$$
$$\hat{\sigma}_X = s_x = \frac{1}{9}\sum_{i=1}^{10}(x_i - 2.5)^2 = 1.58.$$

(b) (0,75 punts)

Per quines raons es podria suposar que X segueix una distribució de Poisson?

Solució:

Per una banda, X recull el nombre d'events durant un interval de temps i per l'altra banda tenim $\mu_X \approx \sigma_X^2$.

Suposem a continuació que X segueix una distribució de Poisson amb paràmetre $\lambda = 2.8$.

(c) (1 punt)

Si sabem que un alumne ha fet almenys 4 exercicis d'estatus a la setmana, quina és la probabilitat que finalment n'hagi fet més de 6?

Solució:

$$\mathrm{P}(X > 6 | X \geq 4) = \frac{\mathrm{P}(X > 6)}{\mathrm{P}(X \geq 4)} = \frac{1 - \mathrm{P}(X \leq 6)}{1 - \mathrm{P}(X \leq 3)} = \frac{1 - 0.976}{1 - 0.692} = 0.079.$$

En total hi ha 200 alumnes que fan l'assignatura. Si suposem que cadascun utilitza l'estatus independentment dels altres, llavors podem suposar que la variable X_{200} 'Nombre d'exercicis fets en una setmana pels 200 alumnes' segueix aproximadament una distribució normal.

(d) (0,5 punts)

Quins són els paràmetres de X_{200}?

Solució:

$$X_{200} \sim \mathrm{Po}(200 \cdot 2.8 = 560) \approx \mathcal{N}(\mu = 560, \sigma = \sqrt{200} \cdot \sqrt{2.8} = 23.7).$$

(e) (1,5 punts)

Quines són les probabilitats que entre tots els 200 alumnes es facin més de 560 i menys de 520 exercicis en una setmana, respectivament?

Solució:

$$\mathrm{P}(X_{200} > 560) = 0.5,$$
$$\mathrm{P}(X_{200} < 520) = \mathrm{P}\big(Z < (520 - 560)/23.7\big) = \mathrm{P}\big(Z < -1.69) = 0.05.$$

(f) (1,5 punts)
Calculeu l'interval centrat en 560 que conté el nombre d'exercicis setmanals fets pels 200 estudiants amb una probabilitat del 99%.

Solució:

$$560 \mp z_{0.995} \cdot \sigma = 560 \mp 2.58 \cdot 23.7 = 499 - 621.$$

Amb una probabilitat superior a 0.99, cada setmana es fan entre 499 i 621 exercicis.

Poques setmanes abans de l'examen, els alumnes comencen a fer-li moltes consultes per correu electrònic al seu professor. Es pot suposar que el nombre de consultes fetes pels estudiants entre les 8 del matí i les 8 del vespre segueix una distribució de Poisson amb mitjana igual a 24 consultes.

(g) (0,5 punts)
Quina és la distribució del temps entre les arribades de dues consultes al correu electrònic del professor?

Solució:

$$T \sim \mathcal{E}(\lambda = 24/12 = 2\,\text{hora}^{-1}).$$

(h) (1,5 punts)
Sabent que en mitja hora abans d'anar a dinar no li ha arribat cap consulta al professor, quina és la probabilitat que tingui una nova consulta a la seva bústia del correu electrònic quan torna de dinar exactament una hora després?

Solució:

$$\begin{aligned} \text{P}(T < 90\,\text{min} | T > 30\,\text{min}) &= \text{P}(T < 60\,\text{min}) = \text{P}(T < 1\,\text{hora}) \\ &= 1 - e^{-2 \cdot 1} = 1 - e^{-2} = 0.865. \end{aligned}$$

A la mateixa enquesta d'abans es va preguntar també pel temps dedicat setmanalment a l'estudi de l'assignatura i es van obtenir els següents estadístics: $\hat{\mu}_T = \bar{t} = 3.3$ hores i $\hat{\sigma}_T = s = 0.6$ hores.

(i) (1,5 punts)
Calculeu els intervals de confiança (amb un nivell de confiança igual al 95%) tant de la mitjana com de la desviació estàndard de la variable T: Hores dedicades setmanalment a l'estudi de l'assignatura. Suposem que T segueix una distribució normal i recordeu que $n = 10$.

Solució:

(j) Interval de confiança per a la mitjana:

$$\text{IC}(\mu_T; 0.95) = \bar{t} \mp t_{9;0.975} \cdot \frac{s}{\sqrt{n}} = [2.87,\ 3.73].$$

(k) Interval de confiança per a la desviació estàndard:

$$\text{IC}(\sigma_T^2; 0.95) = \Big[\frac{s^2 \cdot (n-1)}{\chi^2_{9;\,0.975}},\ \frac{s^2 \cdot (n-1)}{\chi^2_{9;\,0.025}}\Big] = [0.17,\ 1.2],$$

$$\Longrightarrow \text{IC}(\sigma_T; 0.95) = [0.41,\ 1.1].$$

(l) (0,75 punts)
Basant-vos en els resultats de l'apartat anterior, es pot afirmar que el temps de dedicació mitjà és superior a les 3 hores? Raoneu la resposta.

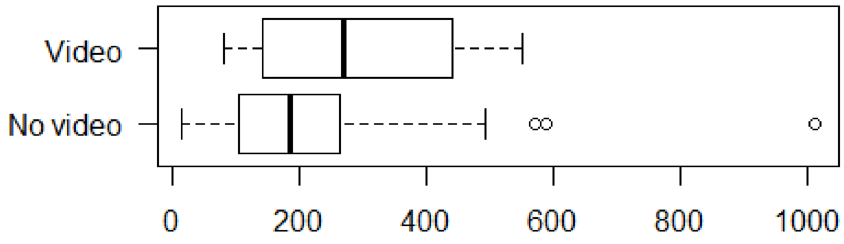
Solució:
Com el límit inferior d'IC$(\mu_T; 0.95)$ és inferior a 3 hores i utilitzant un nivell de significació igual a 0,05, no podem afirmar que la dedicació setmanal mitjà sigui superior a 3 hores.

FIB 2017-18 Q1. FINAL PE. 26 de gener de 2018
NOM:__________________COGNOM:_____________________________
(Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs)

Problema 3 (B5 – B6)

Una empresa de productes informàtics empra Facebook com una via de difusió de la seva marca. Actualment, volen analitzar com influeix el format dels continguts que posen al seu mur a traves de mesurar el nombre de interaccions (*likes*, *shares* i comentaris) en funció de si insereixen un vídeo a la publicació. La descriptiva del nombre d'interaccions està a continuació:

	n	Mitjana	Desviació tipus
Publicacions amb vídeo	7	295.9	184.0
Publicacions sense vídeo	45	217.0	178.5



1. Digues si creus que es tracten de mostres aparellades o independents (1 punt)

Les mostres són independents. No hi ha cap vincle entre les publicacions amb o sense vídeo, de fet, la grandària mostral en ambdues mostres és diferent el que impossibilita que siguin dades aparellades.

Planteja un contrast d'hipòtesi per decidir si la mitjana del nombre d'interaccions de les publicacions amb vídeo (V) és superior a la de les publicacions sense vídeo (NV)

2. Digues quina és la hipòtesi nul·la i alternativa i si el contrast és bilateral o unilateral. (1 punt)

El contrast és <u>unilateral</u> ja que volem veure si la mitjana és superior en les publicacions amb video.

$$\begin{cases} H_0: \mu_V - \mu_{NV} = 0 \\ H_1: \mu_V - \mu_{NV} > 0 \end{cases}$$

3. Digues quines són les premisses per realitzar aquest test i si les veus raonables veient la descriptiva de les dades. (1 punt)

Les premisses són tenir una m.a.s, normalitat i variàncies iguals a les dues mostres. El fet de ser una m.a.s no es pot verificar amb la descriptiva perquè depèn de com s'han recollit les dades. La normalitat la podem avaluar a través de la simetria dels boxplots: en el grup amb vídeo es veu força simetria, però en les publicacions sense vídeo es veu una lleugera cua cap a la dreta que posaria en dubte aquesta premissa. Respecte a les variàncies iguals, malgrat que els boxplots semblen tenir una amplada diferent, les desviacions típiques de la taula són força semblants.

4. Calcula l'estadístic per realitzar aquest test. Digues quina és la seva distribució sota la hipòtesi nul·la i el punt crític amb un α = 0.05. Finalment, conclou sobra la prova d'hipòtesi. (2 punts)

$$s_{pool}^2 = \frac{(n_V - 1)\cdot S_V^2 + (n_{NV} - 1)\cdot S_{NV}^2}{n_V + n_{NV} - 2} = \frac{6\cdot 184^2 + 44\cdot 178.5^2}{50} = 32101.5 \rightarrow s_{pool} = 179.2$$

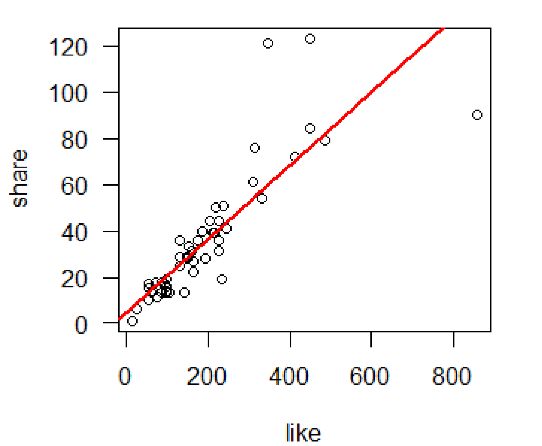
$$\hat{t} = \frac{\bar{y}_V - \bar{y}_{NV}}{s_{pool}\sqrt{\frac{1}{n_V} + \frac{1}{n_{NV}}}} = \frac{295.9 - 217}{179.2\cdot\sqrt{\frac{1}{7} + \frac{1}{45}}} = \frac{78.9}{72.8} = 1.08$$

Distribució: t_{50}

Punt crític: $t_{0.95,50} = 1.675$

Com que l'estadístic és positiu i inferior al punt crític, llavors no és pot rebutjar la hipòtesi nul·la i considerem que no hi ha evidència per tenir que les publicacions amb vídeo tenen un nombre superior d'interaccions en mitjana.

Entre les interaccions possibles es creu que hi ha una relació entre el nombre de “likes” (m’agrada) i el nombre de “shares” (compartir) d’una publicació. Per estudiar aquesta relació, s’ajusta un model lineal. A continuació, tens un gràfic descriptiu i la sortida del model amb R.




```
Call:
lm(formula = share ~ like, data = d)

Residuals:
    Min      1Q  Median      3Q     Max
-50.732  -5.759  -1.233   3.319  61.391

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.15939    3.29026   1.568    0.123
like         0.15783    0.01393  11.330 2.04e-15 ***
---
Residual standard error: 14.87 on 50 degrees of freedom
Multiple R-squared:  0.7197,Adjusted R-squared:  0.7141
F-statistic: 128.4 on 1 and 50 DF,  p-value: 2.038e-15
```

5. Segons el gràfic, hi ha alguna de les premisses del model lineal que sembli no complir-se? (1 punt)

   Sí. La premissa d’homoscedasticitat no es compleix per la forma d’embut que fa que els punts al final de la recta estiguin més allunyats.

6. Calcula un interval de confiança del 90% pel terme constant (*intercept*) (1 punt)

$$IC(90\%, \beta_0) = b_0 \mp t_{0.95,50} \cdot s_{b_0} = 5.16 \mp 1.675 \cdot 3.29 = [-0.35\,, 10.7]$$

7. Calcula l’estadístic per resoldre la prova d’hipòtesi sobre si el pendent poblacional de la recta és 0.15 o diferent. Dóna el punt crític per un α=0.05 i conclou sobre la prova. (1 punt)

$$\hat{t} = \frac{b_1 - 0.15}{s_{b_1}} = \frac{0.15783 - 0.15}{0.01393} = 0.56$$

   $t_{0.975,50} = 2.009$

   Com que l’estadístic és positiu i inferior al punt crític, llavors no és pot rebutjar la hipòtesi nul·la i no es pot rebutjar que el pendent poblacional sigui 0.15.

8. Fes una predicció puntual i per interval de confiança del 95% pel nombre de “shares” per una publicació amb 100 “likes”. [La mitjana del nombre de “shares” és 34.21] (2 punts)

   Puntual: $\hat{y}_h = b_0 + b_1 \cdot X_h = 5.15939 + 0.15783 \cdot 100 = 20.94$

   Mitjana del nombre de “*likes*”: $\bar{x} = \frac{\bar{y} - b_0}{b_1} = \frac{34.21 - 5.159}{0.15783} = 184.1$

   Variància de “*shares*”: $S^2 = \frac{(n-1) \cdot S_y^2 \cdot (1-r^2)}{n-2} \rightarrow S_y^2 = \frac{S^2 \cdot (n-2)}{(n-1) \cdot (1-r^2)} = \frac{14.87^2 \cdot 50}{51 \cdot (1-0.7197)} = 773.39$

   Variància de “*likes*”: $b_1 = r \cdot \frac{S_y}{S_x} \rightarrow S_x = \frac{r \cdot S_y}{b_1} = \frac{\sqrt{0.7197 \cdot 773.39}}{0.15783} = 149.48 \rightarrow S_x^2 = 22344.3$

$$IC(95\%, Y_h) = \hat{y}_h \mp t_{0.975,50} \cdot s\sqrt{1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{(n-1) \cdot S_x^2}} = 20.94 \mp 2.01 \cdot 14.87\sqrt{1 + \frac{1}{52} + \frac{(100 - 184.1)^2}{51 \cdot 22344.3}}$$
$$= [-9.3\,, 51.2]$$

   Com que no pot haver-hi un nombre negatiu de ”*shares*”, l’interval de confiança del 95% per a la predicció és [0, 51.2]