FIB 2014-15 Q2. FINAL PE 17 de juny de 2015

Cognoms, nom: ....

Problema_1 (B1-B2). A primer curs d'uns estudis de Grau es matriculen 400 estudiants, 180 de les quals són dones i 220 homes. Curiosament, hi ha exactament 100 dones i 100 homes que disposen de MAC OS, mentre que la resta disposa d'ordinador amb Windows. Si s'escull un estudiant (home o dona) a l'atzar:

(a) Calculeu la probabilitat de que tingui MAC OS

$$P(MAC) = \frac{\#\ estudiants\ MAC}{\#\ estudiants\ totals} = \frac{100+100}{400} = 0.5$$

(b) Si resulta que té MAC OS, calculeu la probabilitat de que s'hagi escollit una dona

Dels que tenen MAC, 100 són dones i 100 són homes. Per tant,

$$P(DONA \mid MAC) = 0.5$$

Si anomenem X al sexe (0 = dona, 1 = home) i Y al SO (0 = Mac OS, 1 = Windows)

(c) Calculeu la funció de densitat conjunta $f_{X,Y}$

$$f_{X,Y}(0,0) = P(X=0 \wedge Y=0) = P(DONA \wedge MAC) = \frac{100}{400} = 0.25$$

$$f_{X,Y}(0,1) = P(X=0 \wedge Y=1) = P(DONA \wedge WIN) = \frac{80}{400} = 0.20$$

$$f_{X,Y}(1,0) = P(X=1 \wedge Y=0) = P(HOME \wedge MAC) = \frac{100}{400} = 0.25$$

$$f_{X,Y}(1,1) = P(X=1 \wedge Y=1) = P(HOME \wedge WIN) = \frac{120}{400} = 0.30$$

(d) Calculeu la funció de densitat condicionada $f_{X|Y=1}$

$$f_{X|Y=1}(0) = P(X=0 \mid Y=1) = P(DONA \mid Win) = \frac{80}{200} = 0.4$$

$$f_{X|Y=1}(1) = P(X=1 \mid Y=1) = P(HOME \mid Win) = \frac{120}{200} = 0.6$$

(e) Són independents els esdeveniments "ser home" i "disposar d'ordinador amb Windows"? Raoneu la resposta

$$\left.\begin{aligned} P(HOME \mid Win) &= 0.6 \\ P(HOME) &= \frac{220}{400} = 0.55 \neq 0.6 \end{aligned}\right\} \Longrightarrow \ \textit{no són independents}$$

(f) Si escollim dos estudiants diferents a l'atzar, calculeu la probabilitat de que els dos tinguin el mateix sistema operatiu (**3 decimals correctes**)

$$P(mateix\ SO) = P(els\ dos\ MAC) + P(els\ dos\ Win)$$

$$P(els\ dos\ MAC) = P(el\ primer\ MAC\ \wedge\ el\ segon\ MAC) =$$

$$= P(el\ primer\ MAC) \cdot P(el\ segon\ MAC \mid el\ primer\ MAC) = \frac{200}{400} \cdot \frac{199}{399} = 0.2494$$

$$P(els\ dos\ Win) = P(el\ primer\ Win\ \wedge\ el\ segon\ Win) =$$

$$= P(el\ primer\ Win) \cdot P(el\ segon\ Win \mid el\ primer\ Win) = \frac{200}{400} \cdot \frac{199}{399} = 0.2494$$

$$P(mateix\ SO) = P(els\ dos\ MAC) + P(els\ dos\ Win) = 2 \cdot 0.2494 = 0.499$$

Si, per terme mig, un MAC s'avaria 0.1 vegades/any i si el nombre d'avaries segueix una distribució de Poisson,

(g) Calculeu la probabilitat de que en dos any un MAC no s'avariï

Sigui U el nombre d'avaries/any d'un MAC. Sabem que $U \sim P(0.1)$. Per tant, si V denota el nombre d'avaries d'un MAC en un període de dos anys, llavors $V \sim P(0.2)$ i

$$P(V = 0) = e^{-0.2} \cdot \frac{0.2^0}{0!} = 0.81873$$

(h) Calculeu la probabilitat de que, en dos anys, 160 o més MAC's no s'hagin avariat

Hi ha 200 MAC. La probabilitat de que en 2 anys no s'avariï un MAC concret és $p = 0.81873$. Si W denota el nombre de MAC's, de entre els 200, que no s'avarien en un període de dos anys, llavors

$$W \sim Bin(200{,}0.81873)$$

Usant l'aproximació normal,

$$W \approx N\left(200 \cdot 0.81873; \sqrt{200 \cdot 0.81873 \cdot (1 - 0.81873)}\right)$$

$$W \approx N(163.746; 5.448)$$

$$P(W \geq 160) \approx P\left(Z \geq \frac{159.5 - 163.746}{5.448}\right) = 0.78$$

(Z designa una normal típica)

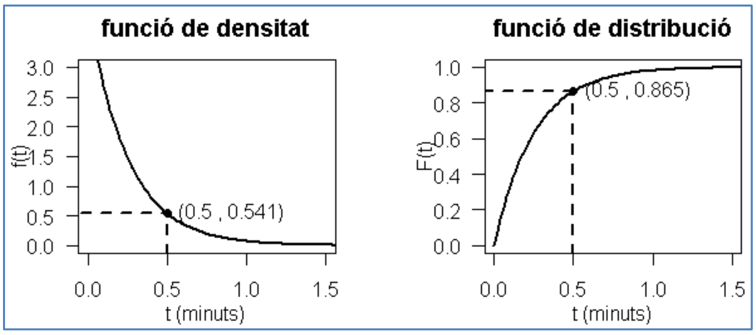
Puntuació: preguntes (c),(f),(g),(h): 1.5 punts (cada una); altres preguntes: 1 punt (c.u.)

NOM:____

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

Problema 2 (B3-B4). L'empresa de missatgeria YALLEGO,S.A. té una base de dades de clients que es consultada recurrentment pel personal de la companyia.

1. La durada de les consultes es distribueix segons una exponencial amb les següents funcions de densitat i de distribució. Troba el paràmetre λ de la distribució. *(0.5 punts)*



Usant la funció de distribució:

$F(t) = 1 - \exp(-\lambda \cdot t) \rightarrow 0.865 = 1 - \exp(-\lambda \cdot 0.5) \rightarrow$

$\rightarrow \exp(-\lambda/2) = 0.135 \rightarrow -\lambda/2 = \ln(0.135) \rightarrow$

$\rightarrow \lambda = -2 \cdot \ln(0.135) \approx 4$

2. Troba la probabilitat que una consulta trigui més de mig minut. Si no has trobat la λ a l'apartat anterior, fes servir λ = 3. *(0.5 punts)*

T: "Durada de les consultes fetes pels empleats de l'empresa"~Exp(λ=4 $minuts^{-1}$)

$P(T > 0.5) = 1 - F_T(0.5) = 1 - (1 - \exp(-4*0.5)) = \exp(-2) = 0.135$

3. Una consulta porta un minut executant-se i encara no ha acabat. Quina és la probabilitat que acabi en el següent mig minut? (és a dir, que en total duri menys de 1 minut i mig) *(1 punt)*

$P(T < 3/2 \mid T > 1)$ = (propietat de no memòria) = $P(T < 3/2 - 1) = P(T < 0.5) = 1 - P(T > 0.5) = 1 - 0.135 = 0.865$

4. El nombre de consultats diàries fetes pel Toni, un empleat de l'empresa es distribueix com una Poisson de mitjana 10. Quina és la probabilitat que un dia concret faci exactament 10 consultes? *(1 punt)*

X: "Nombre de consultes fetes pel Toni en un dia" ~ Poisson(λ=10)

$P(X = 10) = P(X \leq 10) - P(X \leq 9)$ = (Taules) = $0.583 - 0.458 = 0.125$

5. Quin nombre de consultes setmanals (en 5 dies laborables) no superarà el Toni amb una probabilitat de 0.95? *(1 punt)*

X_5: "Nombre de consultes fetes pel Toni setmanalment"~ Poisson(λ=50) ~ N(μ = 50, σ = √50)

$P(X_5 < x_M) = 0.95 \rightarrow P(Z < (x_M - 50)/\sqrt{50}) = 0.95 \rightarrow$ (Taules) $\rightarrow (x_M - 50)/\sqrt{50} = 1.645 \rightarrow x_M = 61.63 \approx 62$ consultes

6. En un dia en que el Toni fa 5 consultes, quina és la probabilitat de que almenys una d'elles trigui més de 30 segons? *(1 punt)*

N: "Nombre de consultes que excedeixen els 30 segons en 5 intents"~ Binomial (n=5, p = 0.135)
$P(N \geq 1) = 1 - P(N \leq 0) = 1 - P(N = 0) = 1 - (0.865)^5 = 1 - 0.484 = 0.516$

L'informàtic de la companyia decideix crear índexs sobre les taules de la base de dades per tal de millorar-ne el rendiment. Per saber si ho ha aconseguit recull una mostra de 100 temps de consultes fetes pels treballadors després d'haver creat els índexs. La suma dels temps (en segons) i dels temps al quadrat són:

$$\sum(x) = 1299 \qquad \sum(x^2) = 32689$$

7. Troba la mitjana, la desviació tipus i l'error estàndard de la mitjana mostral dels temps *(1 punt)*

$$\bar{x} = \frac{1299}{100} = 12.99$$
$$s^2 = \frac{32689 - 100 \cdot 1299^2}{100 - 1} = 159.74 \rightarrow s = 12.64$$
$$se = \frac{s}{\sqrt{n}} = \frac{12.64}{10} = 1.264$$

8. Posa a prova la hipòtesi que el rendiment de les consultes és inferior a 15 segons.

   a. Escriu formalment la prova d'hipòtesi i digues si és bilateral o unilateral *(0.5 punts)*

$$\begin{cases} H_0: \mu = 15 \\ H_1: \mu < 15 \end{cases} \rightarrow \text{Unilateral}$$

   b. Escriu l'expressió de l'estadístic, la seva distribució sota la hipòtesi nul·la i les premisses adients *(0.5 punts)*

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{99}$$ Premisses: m.a.s, n≥100

   c. Calcula el valor de l'estadístic i digues quin és el punt crític per a un nivell de significació α = 0.05 *(0.5 punts)*

$$Estadístic: t = \frac{12.99 - 15}{1.264} = -1.59$$
$$Punt\ crític: t_{99,0.05} \approx -1.66$$

   d. Conclou sobre la prova d'hipòtesi *(0.5 punts)*

   No tenim evidència per rebutjar la hipòtesi nul·la de que la mitjana dels temps sigui de 15 segons

9. Construeix un interval de confiança del 95% (IC95%) bilateral per la mitjana poblacional dels temps i interpreta'l *(1 punt)*

$$IC(\mu, 0.95) = \bar{x} \mp t_{99,0.975} \cdot se = 12.99 \mp 1.99 \cdot 1.264 = [10.47\,, 15.51]$$

La mitjana poblacional dels temps d'execució de les consultes es troba entre 10.5 i 15.5 segons amb una confiança del 95%

10. Calculeu la grandària mostral necessària per tenir un IC95% amb una amplada de 2.5 segons *(1 punt)*

$$\Delta: Amplitud\ de\ l'interval$$
$$\Delta = 2 \cdot t_{99,0.975} \cdot \frac{s}{\sqrt{n}} \rightarrow 2.5 = 2 \cdot 1.99 \cdot \frac{12.64}{\sqrt{n}} \rightarrow 2.5 = \frac{50.30}{\sqrt{n}} \rightarrow n = \left(\frac{50.30}{2.5}\right)^2 \approx 404.8$$

Es necessita una grandària mostral de 405

NOM:___________________________COGNOM:____

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

Problema 3 (B5,B6)

Un professor de Ciències de la Computació vol analitzar la relació que hi ha entre el temps de CPU de l'algorisme A i el de l'algorisme B. Per dur a terme aquesta investigació, recull una mostra, obtenint la següent informació:

A = [94.07, 96.79, 92.15, 92.30, 96.50, 83.11, 91.16, 90.81, 81.37, 89.81, 84.92, 84.43, 86.33, 87.60, 81.08]

B = [86.59, 93.08, 87.85, 86.83, 92.70, 76.80, 83.40, 86.74, 77.67, 85.70, 79.96, 79.80, 81.15, 81.92, 76.32]

El nombre d'observacions és 15 per cadascun i els estadístics a utilitzar son:

Mitjana A = 88.83;	Mitjana B = 83.77	Covariancia = 26.63
Variancia A = 26.70;	Variancia B = 28.21	Variancia diferencia =1.66

1.- **(2 punts)** Assumim que són dades aparellades i volem contrastar si l'esperança del temps dels dos algorismes és la mateixa o no:

a. Plantegeu quina és la hipòtesi nul·la i l'alternativa
b. Digueu quin és l'estadístic i les premisses
c. Es pot rebutjar la hipòtesi nul·la? Raoneu la vostra resposta

Error_tipus = sqrt(1.66) / sqrt(15) = 0.33
Mitjana_diferència = 88.83-83.77 = 5.06

t = (5.06 - 0) / 0.33 = **5.06 / 0.33 = 15.33**

$\mathbf{t_{14,\,0.975} = 2.145}$

2. (**2 punts**) I si les dades no son aparellades, per contrastar si l'esperança del temps dels dos algorismes és la mateixa o no:

a. Plantegeu quina és la hipòtesi nul.la i l'alternativa
b. Digueu quin és l'estadístic i les premisses en aquest cas
c. Es pot rebutjar la hipòtesi nul.la? Raoneu la vostra resposta. I raoneu si obtindríeu el mateix resultat que a l'apartat 1?

s_pooled = sqrt((14 *26.7 + 14*28.21) / (15+15-2)) = 5.24
Error_tipus = 5.24 sqrt(1/15+1/15) = 1.91

t = (88.83-83.77) / 1.91 = **5.06 / 1.91 = 2.65**

$\mathbf{t_{28,\,0.975} = 2.048}$

3. (1 **punt**) Calculeu, interpreteu i compareu els intervals de confiança al 95% per a la diferència de mitjanes pels 2 casos anteriors

IC apartat 1: $5.06 \pm t_{14,0.975}\, 0.33$ = **[4.35 , 5.77]**

IC apartat 2: $(88.83-83.77) \pm t_{28,0.975}\, 1.91$ = **[1.15 , 8.97]**

4. (**2 punts**) Per veure si hi ha relació entre els dos temps de CPU, volem calcular la recta de regressió de B (resposta) en funció de A (predictor). Estimeu puntualment el terme independent (β_0) i el pendent (β_1). Doneu un interval de confiança al 95% per a β_1

b_1 = 26.63 / 26.7 **= 0.997**

b_0 = 83.77 – 0.997*88.83 **= -4.79**

$b_1 \pm t_{13,0.975}\, Sb_1$ = 0.997 ± 2.160 * 0.07 = **[0.85 , 1.15]**

(s^2 = 14(28.21-0.997*26.63)/13 = 1.79 i Sb_1 = sqrt (1.79/(14*26.7)) = 0.07)

5. (**2 punts**) Sabent que el valor actual de A és 100, calculeu la predicció puntual del Temps de CPU de B i l'interval de confiança al 95% per al corresponent valor esperat.

B_{100} = -4.79 + 0.997*100 = 94.91

94.91 ± $t_{13,0.975}$ * sqrt(1.79) * sqrt(1/15 + $(100-88.83)^2$ / (26.7*14)) = 94.91± 2.160*1.34*0.63 = **[93.09 , 96.73]**

6. (**1 punt**) Interpreteu els resultats obtinguts als dos apartats anteriors