Solució del Problema 1

Durant alguns anys un professor d'estadística a la FIB ha estudiat el nombre de consultes per part dels estudiants al llarg dels quadrimestres. La informació recollida la resumeix en forma de la taula següent on X i Y son les variables:

- X: Setmana abans d'una prova d'estadística important ($X = 0$, No; $X = 1$, Sí).
- Y: Nombre d'estudiants que li fan setmanalment almenys una consulta al professor.

La taula presenta les probabilitats conjuntes de les dues variables:

$\mathrm{P}(\boldsymbol{X} \cap \boldsymbol{Y})$	$\boldsymbol{X}=\mathbf{0}$	$\boldsymbol{X}=\mathbf{1}$
$\boldsymbol{Y}=\mathbf{0}$	0,36	0
$\boldsymbol{Y}=\mathbf{1}$	0,24	0,02
$\boldsymbol{Y}=\mathbf{2}$	0,12	0,08
$\boldsymbol{Y}=\mathbf{3}$	0,03	0,1
$\boldsymbol{Y}=\mathbf{4}$	0	0,05

a) (1 punt)

Doneu les distribucions marginals d'X i d'Y.

$\mathrm{P}(\boldsymbol{X} \cap \boldsymbol{Y})$	$\boldsymbol{X}=\mathbf{0}$	$\boldsymbol{X}=\mathbf{1}$	$\mathrm{P}_{\boldsymbol{Y}}$
$\boldsymbol{Y}=\mathbf{0}$	0,36	0	0,36
$\boldsymbol{Y}=\mathbf{1}$	0,24	0,02	0,26
$\boldsymbol{Y}=\mathbf{2}$	0,12	0,08	0,2
$\boldsymbol{Y}=\mathbf{3}$	0,03	0,1	0,13
$\boldsymbol{Y}=\mathbf{4}$	0	0,05	0,05
$\mathrm{P}_{\boldsymbol{X}}$	0,75	0,25	

b) (1 punt)

Doneu les distribucions condicionades d'Y donat $X = 0$ i $X = 1$, respectivament

$\mathrm{P}(\boldsymbol{Y}\vert\boldsymbol{X})$	$\boldsymbol{X}=\mathbf{0}$	$\boldsymbol{X}=\mathbf{1}$
$\boldsymbol{Y}=\mathbf{0}$	0,48	0
$\boldsymbol{Y}=\mathbf{1}$	0,32	0,08
$\boldsymbol{Y}=\mathbf{2}$	0,16	0,32
$\boldsymbol{Y}=\mathbf{3}$	0,04	0,4
$\boldsymbol{Y}=\mathbf{4}$	0	0,2

c) (1 punt)

Podem suposar independència entre ambdues variables? Raoneu la resposta.

X i Y **no** són independents ja que $\mathrm{P}_{Y|X=0} \neq \mathrm{P}_{Y|X=1}$.

d) (0.5 punts)

Si $X = 0$, quina és la probabilitat que almenys dos alumnes li facin una consulta en una setmana?

$$\mathrm{P}(Y \geq 2|X = 0) = 0.16 + 0.04 = 0.2.$$

e) (1.5 punts)
Si el professor ha rebut consultes d'almenys dos alumnes, quina és la probabilitat que estem en una setmana sense proves d'estadística?

$$\mathrm{P}(X=0|Y\geq 2)=\frac{\mathrm{P}(X=0\cap Y\geq 2)}{\mathrm{P}(Y\geq 2)}=\frac{0.12+0.03}{0.2+0.13+0.05}=0.39.$$

f) (1.5 punts)
Suposant que el valor d'Y d'una setmana sigui independent del valor de la setmana anterior, llavors, si $X=0$, quina és la probabilitat que en dues setmanes rebi com a màxim una sola consulta?

$$\begin{aligned}\mathrm{P}(Y_1+Y_2\leq 1|X=0)&=\mathrm{P}(Y_1=0\cap Y_2=0|X=0)+\mathrm{P}(Y_1=0\cap Y_2=1|X=0)+\mathrm{P}(Y_1=1\cap Y_2=0|X=0)\\&=0.48^2+2\cdot 0.48\cdot 0.32=0.5376.\end{aligned}$$

g) (1.5 punts)
Si estem en èpoques d'exàmens, quin és el valor esperat i quina és la desviació estàndard d'Y? Com s'interpreten aquests valors?

$$\mu_{Y|X=1}=\sum_{y=0}^{4}y\cdot\mathrm{P}(Y=y|X=1)=2.72$$

$$\sigma_{Y|X=1}=\sqrt{\sum_{y=0}^{4}(y-2.72)^2\cdot\mathrm{P}(Y=y|X=1)}=0.87.$$

La mitjana i la desviació estàndard d'Y empíriques al llarg de moltes setmanes (en èpoques d'exàmens) estaran molt a prop de 2.72 i 0.87, respectivament.

h) (2 punts)
Suposem que en un quadrimestre hi ha 12 setmanes sense prova d'estadística important. Quina és la distribució de la variable Z: 'Nombre de setmanes en un quadrimestre amb almenys una consulta dels estudiants, sabent que no hi hagut cap prova d'estadística important'? Quant valen $\mathrm{E}(Z)$ i la desviació estàndard de Z?

$$Z\sim B(12,1-0.48=0.52)\Longrightarrow \mu_Z=12\cdot 0.52=6.24\quad\text{i}\quad \sigma_Z=\sqrt{12\cdot 0.52\cdot 0.48}=1.73.$$

NOM:____

(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els càlculs en les respostes)

PROBLEMA 2

A. Un examen d'una assignatura en un centre educatiu te una taxa mitjana d'aprovats per convocatòria de 0.06

A1. ***(0.5 punts)***. Quina distribució segueix el nombre d'aprovats, si només disposeu de la taxa mitjana d'aprovats? Justifiqueu la vostra resposta

X= variable nombre d'aprovats

X~Poisson, λ=0.06

A2. ***(1.5 punts)***. Quina és la probabilitat que aprovin 1 o mes estudiants?

$P(X\geq1) = 1-P(X<1) = 1-P(X=0) = 1-0.942 = 0.058$

B. Un examen d'un altra assignatura del mateix centre educatiu, a la que es presenten tots els estudiants (n=100) te una proporció històrica d'aprovats del 30%

B1. ***(1 punt).*** Doneu l'esperança i variància de la variable nombre d'aprovats. Quina distribució de probabilitat segueix aquesta distribució? Doneu la distribució exacta així com la seva aproximació per valors de n grans. Justifiqueu la vostra resposta

X=variable nombre d'aprovats i segueix una distribució Binomial amb n=100, p=0.3

$E(X)=n*p = 100*0.3 = 30$; $V(X)=n*p*(1-p)=100*0.3*0.7=21$

Distribució exacta: B(n=100, p=0.3)

Aproximació d'aquesta distribució: $N(\mu=30, \sigma^2=21)$

B2. ***(1.5 punts).*** Calculeu quina és la probabilitat de que si es presenten 20 estudiants, aprovin entre 4 i 6, tenint en compte que és manté la proporció d'aprovats (Utilitzeu la distribució Binomial)

Y=variable nombre d'aprovats i segueix una distribució Binomial amb n=20, p=0.3

$P(4\leq Y\leq 6) = 0.6080 - 0.1071 = 0.5009$

Aquesta resultat s'ha obtingut utilitzant les taules de la distribució Binomial

C. Per tenir una primera estimació de la nota mitjana de l'examen del punt B, es va prendre una mostra aleatòria de 20 exàmens i es van obtenir els següents resultats:

$$\sum_{i=1}^{20} x_i = 150 \qquad \sum_{i=1}^{20} x_i^2 = 1163$$

C1. ***(2 punts).*** Calculeu un IC del 95% per a la mitjana assumint que les dades segueixen una distribució Normal

=150/20=7.5; V(x)=

IC(μ, 95%)=

C2. Estudis realitzats en assignatures similars a aquestes d'altres anys indiquen que la nota mitjana és 6.5. Utilitzeu la mostra disponible per posar a prova si els estudiants d'aquesta assignatura formen part d'aquesta població. Seguiu els següents passos:

a) ***(0.5 punts).*** Plantegeu la prova d'hipòtesi. Digueu si és unilateral o bilateral i el per què

H_0: μ=6.5

H_1: $\mu \neq$6.5

Prova d'hipòtesi bilateral per què l'alternativa és diferent de 6.5

b) ***(1.5 punts).*** Calculeu l'estadístic i digueu la seva distribució sota H0

Estadístic:

L'estadístic segueix una distribució t_student amb ν=19 graus de llibertat

c) ***(1.5 punts).*** Quin és el valor que utilitzareu per decidir si l'estadístic anterior permet rebutjar la hipòtesi nul·la amb un risc alfa del 5%? Hi ha evidència per rebutjar-la?

Per ser una prova d'hipotesi bilateral i l'estadístic seguir una distribució t_student amb ν=19 graus de llibertat, els límits amb un risc alfa del 5% son (-2.093, 2.093).

Com que 3.212 > 2.093, rebutgem H0, per tant, els estudiants d'aquest assignatura pertanyen a una població diferent a la població dels anys anteriors.

NOM:____

(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els càlculs en les respostes)

Problema 3.

En una empresa els ordinadors personals tenen instal.lades dues marques d'antivirus. Les variables que compten el nombre de fitxers examinats per minut les designarem per X i Y respectivament per cadascuna de les dues marques. Admetrem que aquestes variables segueixen una distribució normal amb la mateixa variància.

A.- Per una part, s'escullen a l'atzar 10 ordinadors amb cadascun dels antivirus (i amb igualtat d'altres condicions) i es mesuren X i Y per tal de posar a prova si el seu comportament mitjà és igual o no:

X:	221	234	267	216	482	308	395	264	386	314
Y:	335	353	431	413	262	303	306	402	338	295

$\sum x_i = 3087 \quad \sum x_i^2 = 1022043$

$\sum y_i = 3438 \quad \sum y_i^2 = 1210126$

1.- (0.5 punts) Calculeu les mitjanes i desviacions de les variables X i Y corresponents a les dues marques

mean(X) 308.7 sd(X) 87.61412
mean(Y) 343.8 sd(Y) 55.91819

2.- (0.5 punts) Es tracta d'un disseny 'aparellat' o 'independent'? Raoneu la resposta

Independent perquè cadascun dels ordinadors té un dels dos antivirus

3.- (2 punts) Plantegeu la prova per concloure si les mitjanes de X i Y són iguals o no indicant:

- Hipòtesis, estadístic i premisses

H_0: $\mu_X = \mu_Y$
H_1: $\mu_X <> \mu_Y$

$$\hat{t} = \frac{(\bar{x} - \bar{y})}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

X , Y $\rightarrow$ N $\sigma_1 = \sigma_2$ m.a.s indep.

- Càlcul de l'estadístic i el punt crític

(308.7-343.8) / (73.5 sqrt(1/10+1/10)) = -1.07
Punts crítics t_{18} -2.101 i 2.101

- IC per la diferència de mitjanes

(308.7-343.8) +- 2.101 sqrt(73.5^2/10 + 73.5^2/10) = -35.1 +- 69.06 = [-104.2 , 34.0]

- Conclusió i interpretació global de la prova anterior

Acceptem H_0 No hi ha prou evidència per no acceptar que el comportament mitjà dels dos antivirus és igual

B.- Per altra part, es decideix fer un estudi a 10 ordinadors escollits a l'atzar, on a cadascun es mesura X i Y de les dues marques assegurant que es passen els dos antivirus a cada ordinador en les mateixes condicions:

X:	230	341	261	309	275	367	317	272	298	317
Y:	252	347	303	321	339	398	325	267	360	412
D:	-22	-6	-42	-12	-64	-31	-8	5	-62	-95

$\sum x_i = 2987 \quad \sum x_i^2 = 906863$

$\sum y_i = 3324 \quad \sum y_i^2 = 1128346 \quad \sum x_i y_i = 1007383$

$\sum d_i = -337 \quad \sum d_i^2 = 20443$

1.- (2 punts) Plantegeu la prova per concloure si les mitjanes de X i Y són iguals o no indicant:

- Hipòtesis, estadístic i premisses

H_0: μ_D=0
H_1: μ_D<>0

$(\bar{D}-\mu_0) / (s_D/\sqrt{n})$ D $\rightarrow$ N m.a.s aparellada

- Càlcul de l'estadístic i el punt crític

mean(D) -33.7 sd(D) 31.8
(-33.7 - 0) / (31.8/sqrt(10)) = -3.35
Punts crítics t_9 -2.262 i 2.262

- IC per la diferència de mitjanes de velocitat

(-33.7) +- 2.262 * 31.8 / sqrt(10) = -33.7 +- 22.75 = [-56.45 , -10.95]

- Conclusió i interpretació global de la prova anterior

No podem acceptar H_0

Hi ha prou evidència per no acceptar que el comportament mitjà dels dos antivirus sigui igual

2.- (5 punts) A partir de les dades d'aquesta prova B es decideix estudiar la relació lineal entre X (predictora) i Y (resposta).

- Covariància i correlació entre X i Y

mean(X) = 298.7 sd(X) = 40.34 mean(Y) = 332.4 sd(Y) = 51.04

Cov = (1007383 – (2987*3324/10)) / 9 = 1611.56

Corr = 1611.56 / (40.34*51.04) = 0.78

- Estimeu puntualment els coeficients de la recta de regressió i expliqueu què signifiquen

b_1 = 1611.56 / (40.34^2) = 0.99

b_0 = 332.4 – 0.99*298.7 = 36.69

- Ompliu la taula de descomposició de la variabilitat i calculeu i interpreteu el coeficient de determinació R^2

	SQ	Graus de llibertat (Gdl)	QM = SQ/Gdl	Rati
Explicada pel model	0.99*9*1611.56=14358.99	1	14358.99	12.64
Residual	9086.74	8	1135.84	
Total	9*51.04^2 = 23445.73	9		

R^2 = 61.2 %

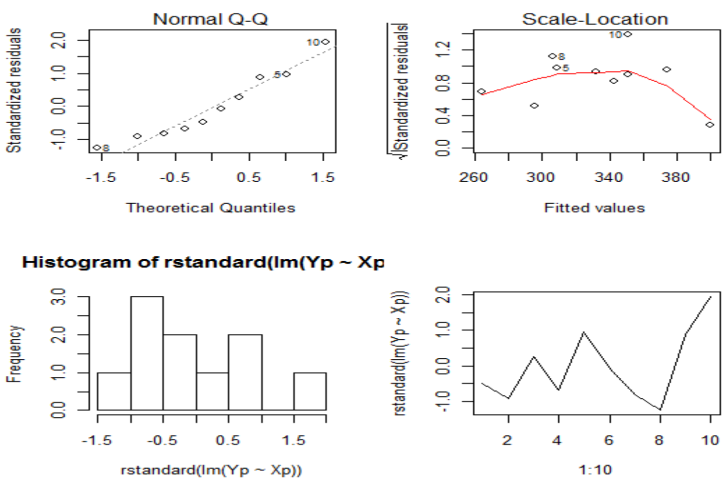
- Poseu a prova si el pendent val 1 o no, indicant les hipòtesis, l'estadístic i el seu càlcul, i la conclusió i interpretació

(0.99 - 1) / (sqrt(1135.84 /9*40.34^2)) = 0.036

Punts crítics t_8 -2.306 i 2.306

Acceptem H_0 No hi ha prou evidència per no acceptar que el pendent és igual a 1

- Valoreu les premisses amb els següents gràfics d'anàlisis de residus



El normal QQ plot i l'histograma mostre que hi ha força normalitat, no sembla haver-hi problemes de no independencia, i hi ha força homocedasticitat tot i que els residus semblen lleument majors en les últimes observacions.