

UOC - Primavera 2017 – Exàmens

Examen 1

Problema 1. En un centre de documentació es vol veure si les consultes d'un determinat diari es fan per Internet (des del mateix centre) o consultant-ho en paper. Es pren nota del nombre de consultes durant 8 dies i s'obtenen les següents dades:

12, 6, 7, 3, 15, 10, 18, 5 (consultes en paper)

9, 3, 8, 8, 9, 8, 9, 18 (consultes electròniques)

Es demana:

a) Trobeu el rang (o recorregut) en ambdós casos i comenteu la seva validesa com a mesura de dispersió (en aquest cas).

b) Trobeu la mitjana i els tres quartils de les dues mostres.

c) Feu un diagrama de caixa per a cada variable.

Criteris de puntuació i valoració (sobre 10): a) 3 punts; b) 4 punts i c) 3 punts.

SOLUCIÓ:

a) En ambdós casos, rang = nombre més gran - nombre més petit = 18 - 3 = 15

En aquest cas, si mirem les distribucions de les dades tenim:

Paper: 3, 5, 6, 7, 10, 12, 15, 18 i Electrònica: 3, 8, 8, 8, 9, 9, 9, 18

i veiem que la dispersió en el primer cas és més gran que en el segon. I com que el rang no posa de manifest cap diferència entre les dues distribucions, hauríem de dir que, en aquest cas, el rang no és una bona mesura de dispersió.

b) Mitjana (paper)

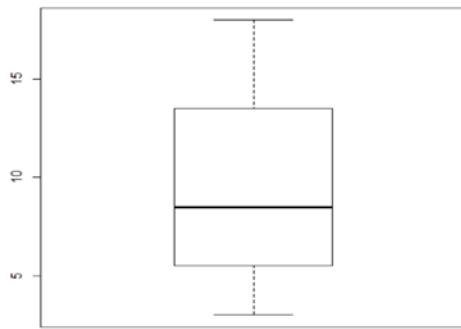
$$\bar{x} = \frac{\sum x_i}{n} = \frac{3+5+6+7+10+12+15+18}{8} = 9.5.$$

Mediana = (7+10)/2 = 8.5, Q1=5.5 Q3=13.5

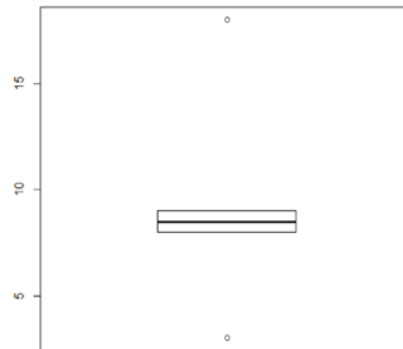
Mitjana (electrònic) = 9

Mediana = (8+9)/2 = 8.5 Q1=8 Q3=9

c) Boxplot de "Paper"



Boxplot d'"Electrònica"



Problema 2.

La taula següent presenta un resum de les característiques sol·licitades en 940 ordres de compra de computadores.

		Memòria Addicional	
		No	Sí
Processament opcional d'alta velocitat	No	514	68
	Sí	112	246

- Completeu la taula de contingència amb els totals.
- Calculeu la probabilitat que una ordre de compra sol·liciti un processador d'alta velocitat i no demani memòria addicional.
- Quina és la probabilitat que en una ordre de compra es demani un processador amb alta velocitat sabent que té memòria addicional?
- Els successos "demander processador d'alta velocitat" i "demander memòria addicional", són independents? Per què?

Criteris de puntuació i valoració (sobre 10): a) 1 punt; b) 3 punts, c) 3 punts i d) 3 punts.

SOLUCIÓ:

		Memòria Addicional		
		No	Sí	
Processament opcional d'alta velocitat	No	514	68	582
	Sí	112	246	358
		626	314	940

- $P(\text{SI Alta velocitat} \cap \text{NO memòria addicional}) = 112/940 = 0,1191 = 11,91\%$
- $P(\text{SI Alta velocitat} / \text{memòria addicional}) = 246/314 = 0,7834 = 78,34\%$

d) $P(\text{SI Alta velocitat})=358/940=0,38$; $P(\text{SI memòria addicional})=314/940=0,33$.
Com que $P(\text{SI Alta velocitat i SI memòria addicional})=246/940=0,26$ és diferent de $P(\text{SI Alta velocitat}) \cdot P(\text{SI memòria addicional})$ diem que els successos no són independents.

Problema 3. Suposem que l'examen d'una certa assignatura l'han suspès el 15% dels 200 estudiants matriculats. Anem escollint exàmens a l'atzar i ens preguntem:

- a) Quina és la probabilitat que el primer examen suspès que trobem sigui el setè?
- b) Quina és la probabilitat que trobem un examen **aprobat** estrictament abans del tercer examen?
- c) Quan s'espera que surti el primer examen suspès?

En cada cas indiqueu quina és la variable aleatòria, el seu tipus i els corresponents paràmetres.

Criteris de puntuació i valoració (sobre 10): S'han d'indicar les fórmules i els càlculs realitzats, així com els raonaments. a) 4 punts (1 punt distribució i paràmetres, 3 punts càlculs); b) 4 punts (1 punt distribució i paràmetres, 3 punts càlculs); c) 2 punts.

SOLUCIÓ:

- a) $X = \text{"Posició del primer examen suspès"}$. És geomètrica de paràmetre $p=0.15$.
 $P(X=7)=0.85^6 \cdot 0.15=0.0566$
- b) $Y = \text{"Posició del primer examen aprovat"}$. És geomètrica de paràmetre $p=0.85$.
 $P(Y \leq 2)=1-0.15^2=0.9775$
- c) $X = \text{"Posició del primer examen suspès"}$. És geomètrica de paràmetre $p=0.15$.
 $E(X)=1/0.15= 6.6667$

Problema 4. Disposem d'una mostra de 300 exàmens de la UOC, amb una nota mitjana de 78 (sobre 100) i una desviació estàndard de 20. Suposant que les notes segueixen una distribució normal,

- a) Calculeu un interval de confiança per a la mitjana de les notes, amb un nivell de confiança del 90%.
- b) Quina hauria de ser la mida de la mostra perquè la longitud de l'interval de confiança fos inferior a 2?

Valors de probabilitats que us poden ser útils; si no trobeu exactament el que necessiteu, useu el més proper

$p(X \geq x)$	$X \sim N(0,1)$	$X \sim t$ de Student amb 299 graus de llibertat	$X \sim t$ de Student amb 300 graus de llibertat
0.01	2.326	2.339	2.339
0.025	1.96	1.967	1.968
0.05	1.645	1.65	1.65

0.005	2.575	2.592	2.592
-------	-------	-------	-------

Criteris de puntuació i valoració (sobre 10): S'han d'indicar les fórmules i els càlculs realitzats, així com els raonaments. a) 6 punts: plantejament 2 punts, càlculs 4 punts; b) 4 punts (plantejament 2 punts, càlculs 2 punts).

SOLUCIÓ

a) És un interval de confiança de la mitjana d'una distribució normal amb la desviació estàndard desconeguda: $\left(\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}\right) = \left(78 \pm t_{0.05, 299} \frac{20}{\sqrt{300}}\right) = \left(78 \pm 1.65 \frac{20}{\sqrt{300}}\right) = (76.0947, 79.9053)$

b) Si la longitud és 2 aleshores el marge d'error és 1 i tenim que $n \geq \left(t_{\frac{\alpha}{2}, n-1}\right)^2 s^2 = 1.65^2 * 20^2 = 1089.12$

Problema 5. Volem contrastar si la mitjana de missatges rebuts per hora en un dia en un servidor de correu és superior a 1500 missatges. Agafem una mostra aleatòria de 100 dies, calculem les mitjanes del nombre de missatges rebuts per hora obtenint una mitjana de 1529.971 missatges amb una desviació típica de 155.791.

Suposant que la variable que ens dóna la mitjana del nombre de missatges rebuts per hora és normal,

a) Establiu el contrast a realitzar. Heu d'indicar: si es tracta d'un contrast d'una mostra o de dues mostres, de quin paràmetre hem de fer el contrast (mitjana, proporció, etc.) i la hipòtesi nul·la i la hipòtesi alternativa.

b) Quin és el valor de l'estadístic de contrast?

c) A un nivell de significació de $\alpha = 0.05$, calculeu el(s) valor(s) crític(s).

d) A quina conclusió arribaríeu? No és suficient afirmar si acceptem o rebutgem la hipòtesi nul·la, s'ha de raonar la conclusió en el context del problema.

Criteris de puntuació i valoració (sobre 10): a) 3 punts. Si indica si es tracta d'una mostra o dues, 1 punt. Si indica el paràmetre, 1 punt. Si indica el contrast, 1 punt. b) 2 punts. Indica l'expressió de l'estadístic de contrast, 1 punt. Valor del mateix, 1 punt. c) 1 punt. d) 4 punts. Indicar si es rebutja o s'accepta, 1 punt. Raonament, 3 punts.

Valors de les taules de la $N(0,1)$ i la t de Student (utilitzeu els que considereu necessaris, si necessiteu un valor que no és exacte, utilitzeu el valor més proper)

$p(X \geq x)$	$x(N(0,1))$	$x(t(99))$	$x(t(100))$
0.01	2.326	2.365	2.364
0.025	1.96	1.984	1.984
0.05	1.645	1.66	1.66

SOLUCIÓ

a) És tracta d'un contrast d'una mostra i el paràmetre és la mitjana. El contrast a fer és el següent:

$$\begin{aligned} H_0: & \mu = 1500, \\ H_1: & \mu > 1500. \end{aligned}$$

on μ representa la mitjana de missatges rebuts per hora en un dia pel servidor.

b) El valor de l'estadístic de contrast és

$$t = \frac{\bar{x} - 1500}{s} = \frac{1529.971 - 1500}{155.791} = 1.924.$$

c) El valor crític serà:

$$t_{\alpha,99} = t_{0.05,99} = 1.66.$$

d) Com que el valor de l'estadístic de contrast és major que el valor crític, rebutjaríem la hipòtesi nul·la i conclouríem que efectivament tenim prou evidències per afirmar que la mitjana del nombre de missatges per hora en un dia és superior a 1500.

Problema 6

Volem fer un estudi per veure si el temps d'execució d'un determinat algorisme d'anàlisi de grafs depèn o no linealment del nombre de nodes del graf. S'ha determinat el temps d'execució en un determinat conjunt de grafs que tenen de 10 a 100 nodes. Hem calculat la corresponent recta de regressió i els resultats es mostren a la sortida següent que ens ha donat R:

```
##
## Call:
## lm(formula = temps.execucio ~ nodes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -49.462 -19.211   0.748  23.930  55.709
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  33.3274     6.9128   4.821 0.00000522 ***
## nodes        5.2560     0.1099  47.825 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.32 on 98 degrees of freedom
## Multiple R-squared:  0.9589, Adjusted R-squared:  0.9585
## F-statistic: 2287 on 1 and 98 DF, p-value: < 0.00000000000000022
```

a) Trobeu la recta de regressió del temps d'execució (en mil·lisegons) en funció del nombre de nodes del graf i interpreteu els resultats.

b) Feu el contrast d'hipòtesi sobre el pendent de la recta, amb un nivell de significació de $\alpha = 0.01$ i interpreteu els resultats. Heu de donar el contrast sobre el pendent indicant la hipòtesi nul·la i alternativa i l'estadístic de contrast.

c) Calculeu el coeficient de determinació. Discutiu si és un bon model.

Criteris de puntuació i valoració (sobre 10): a) (3 punts). Donar la recta de regressió, 1.5 punts. Interpretació dels paràmetres de la recta, 1.5 punts. b) (5 punts). Donar el contrast d'hipòtesi indicant la hipòtesi nul·la i

alternativa, 2 punt. Donar el valor de l'estadístic, 3 punts. c) (2 punts). Donar el coeficient de determinació, 1 punt. Interpretació del mateix, 1 punt.

SOLUCIÓ

a) La recta de regressió és la següent:

$$\text{temps.execucio} = 33.32736 + 5.25602 \cdot \text{nodes}.$$

El coeficient 5.25602 és el pendent de la recta de regressió i representa l'augment del temps d'execució de l'algorisme quan el graf augmenta en un node.

El coeficient 33.32736 és l'ordenada a l'origen i no tendria interpretació real ja que vendria a ésser el temps d'execució corresponent a un graf buit.

b) El contrast sobre el pendent de la recta és el següent:

$$\begin{aligned} H_0: & \beta_1 = 0, \\ H_1: & \beta_1 \neq 0, \end{aligned}$$

on β_1 és el pendent de la recta de regressió. L'estadístic de contrast val: 4.8211.

c) El valor del coeficient de determinació val 0.95891. Es un valor molt proper a 1. Per tant, podem dir que la regressió és molt bona.

Examen 2

Problema 1. Les següents dades corresponen a la quantitat d'assignatures superades per 12 estudiants de la UOC:

x_i	n_i
3	2
10	4
15	6

- a) Calculeu la mitjana i la mediana.
- b) Feu el diagrama de barres.
- c) Calculeu la desviació típica mostral.

Criteris de puntuació i valoració (sobre 10): a) 4 punts; b) 3 punts; c) 3 punts.

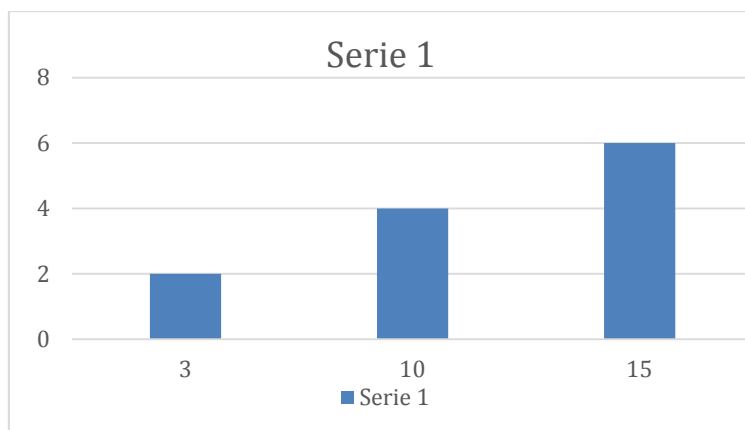
SOLUCIÓ:

$$\bar{x} = \frac{\sum n_i x_i}{n} = \frac{3 \cdot 2 + 10 \cdot 4 + 15 \cdot 6}{12} = 11.33.$$

a)

Mediana = 12.5 (nombre parell de dades, 12), ocupa el lloc 6 i 7 $(10+15)/2 = 12.5$

b)



c) La desviació típica mostral valdrà:

$$s = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum 2(3-11.33)^2 + 4(10-11.33)^2 + 6(15-11.33)^2}{11}} = 4.529.$$

Problema 2. Quatre indústries (A, B, C, i D) produeixen microprocessadors per a la fabricació d'un determinat tipus d'ordinador per a cotxe. Produeixen respectivament: A el 40%, B el 30%, C el 20% i D el 10% de la producció total. Dels que produeix A el 5% són defectuosos, així com el 4% dels de B, el 2% dels de C i el 1% dels de D.

a) Representeu l'arbre de probabilitats.

b) Si seleccionem a l'atzar un vehicle dotat d'ordinador a bord, quina és la probabilitat que el seu ordinador porti el microprocessador defectuós?

c) Els successos “microprocessador fabricat per A” i “microprocessador defectuós” són independents? Per què?

Criteris de puntuació i valoració (sobre 10): a) 4 punts; b) 3 punts i c) 3 punts.

SOLUCIÓ:

a) Considerarem els següents successos i les probabilitats que ens proporciona l'enunciat:

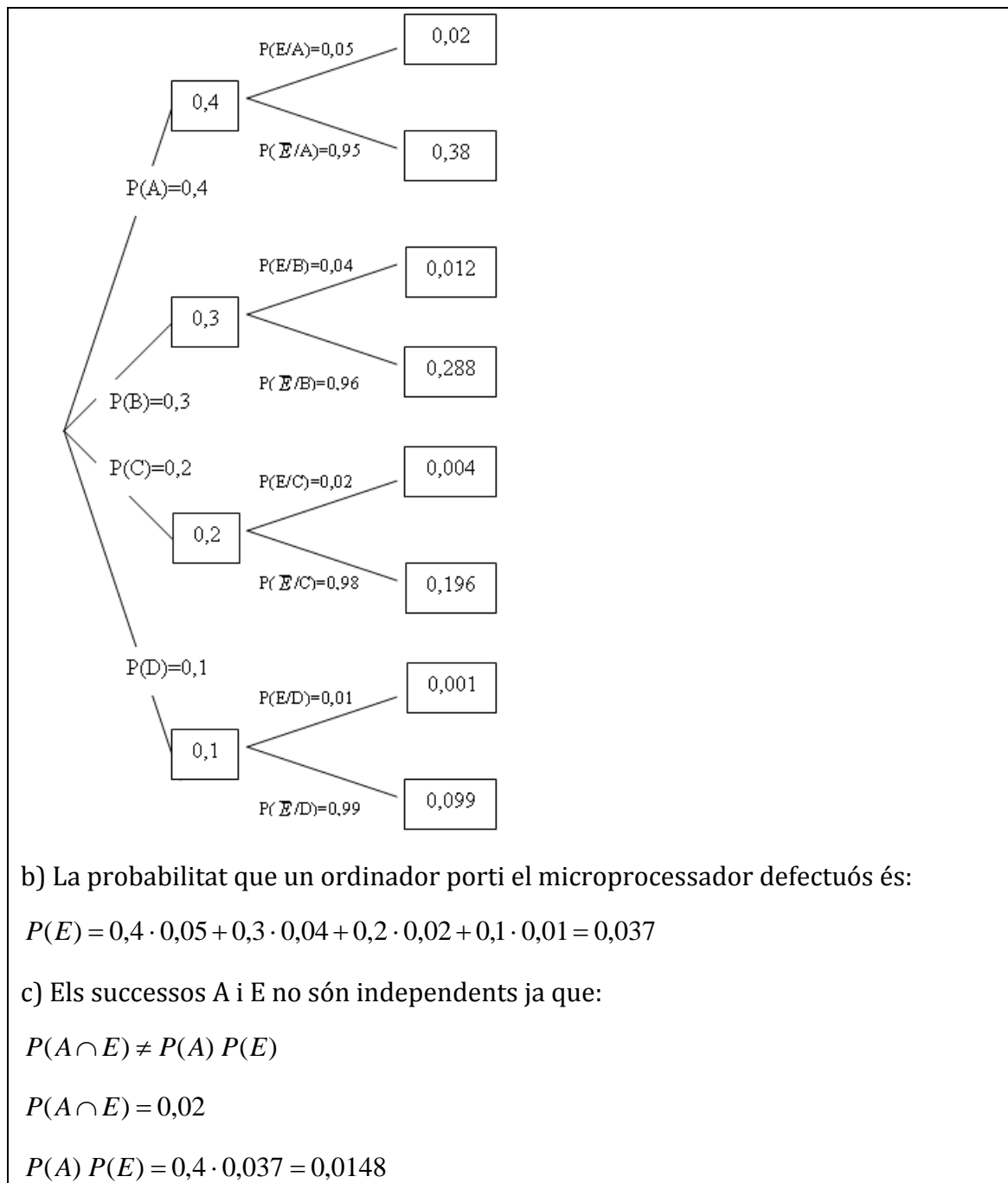
A = “Microprocessador fabricat per A” $P(A) = 0,4$ $P(E / A) = 0,05$

B = “Microprocessador fabricat per B” $P(B) = 0,3$ $P(E / B) = 0,04$

C = “Microprocessador fabricat per C” $P(C) = 0,2$ $P(E / C) = 0,02$

D = “Microprocessador fabricat per D” $P(D) = 0,1$ $P(E / D) = 0,01$

E = “Microprocessador defectuós”



Problema 3. Suposem que l'examen d'una certa assignatura l'han suspès el 15% dels 200 estudiants matriculats. Escollim una mostra aleatòria de 5 exàmens i ens preguntem

a) Quina és la probabilitat que entre ells hi hagi exactament 4 aprovats? Indiqueu quina és la variable aleatòria, el seu tipus i els corresponents paràmetres.

b) Quina és la probabilitat que entre ells hi hagi més d'un aprovat?

c) Quina és la variància del nombre d'exàmens aprovats quan escollim 5 exàmens a l'atzar?

Criteris de puntuació i valoració (sobre 10): S'han d'indicar les fórmules i els càlculs realitzats, així com els raonaments. a) 5 punts (2 la distribució de la variable aleatòria i 3 la probabilitat); b) 3 punts; c) 2 punts.

SOLUCIÓ

a) $X =$ "nombre d'exàmens aprovats si escollim aleatòriament 5 exàmens". Binomial de paràmetres $n=5$, $p=0.85$. $P(X=4)=5 \cdot 0.85^4 \cdot 0.15=0.3915$.

b) $P(X>1)=1-P(X=0)-P(X=1)=1-0.15^5-5 \cdot 0.15^4 \cdot 0.85=0.9978$

c) $\text{Var}(X)=np(1-p)=0.6375$

Problema 4. Un assistent de veu per mòbils ha processat 200 comandes i n'ha processat correctament 180 d'aquestes.

a) Calculeu un interval de confiança per a la proporció de vegades que l'assistent processa correctament una comanda, amb un nivell de confiança del 99%.

b) Quina hauria de ser la mida de la mostra perquè el marge d'error fos inferior a 0.01?

Valors de probabilitats que us poden ser útils; si no trobeu exactament el que necessiteu, useu el més proper

$p(X \geq x)$	$X \sim N(0,1)$	$X \sim t$ de Student amb 199 graus de llibertat	$X \sim t$ de Student amb 200 graus de llibertat
0.01	2.326	2.345	2.345
0.025	1.96	1.971	1.971
0.05	1.645	1.652	1.652
0.005	2.575	2.6	2.6

Criteris de puntuació i valoració (sobre 10): S'han d'indicar les fórmules i els càlculs realitzats, així com els raonaments. a) 6 punts: plantejament 2 punts, càlculs 4 punts; b) 4 punts: plantejament 2 punts, càlculs 2 punts.

SOLUCIÓ

a) És un interval de la proporció amb $\hat{p} = \frac{180}{200} = 0.9$ per tant tenim $\left(\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})} \right) = 0.9 \pm 2.576 \sqrt{0.9 \cdot 0.1} = (0.845, 0.955)$.

b) Tenim que $N \geq \left(z_{\alpha/2} \right)^2 \frac{\hat{p}(1-\hat{p})}{ME^2} = 2.576^2 \frac{0.9 \cdot 0.1}{0.01^2} = 5972.1984$, és a dir, 5973

Problema 5. Volem contrastar si el nombre de peticions a un servidor és el mateix a la tarda i al matí en un dia. Per fer l'estudi, varem escollir aleatòriament 90 dies i varem fer el contrast corresponent en R. Els resultats varen ésser els següents:

```
##
## Paired t-test
##
## data: servidor.mati and servidor.tarda
```

```
## t = -5.048, df = 89, p-value = 0.000002349
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -246.6335 -107.3136
## sample estimates:
## mean of the differences
## -176.9736
```

a) Establiu el contrast a realitzar. Heu d'indicar si es tracta d'un contrast d'una mostra o de dues mostres, de quin paràmetre hem de fer el contrast (mitjana, proporció, etc.), i la hipòtesi nul·la i la hipòtesi alternativa.

b) Quin és el valor de l'estadístic de contrast?

c) Quin és el p-valor del contrast?

d) A un nivell de significació de $\alpha = 0.05$, a quina conclusió arribaríeu? No és suficient afirmar si acceptem o rebutgem la hipòtesi nul·la, s'ha de raonar la conclusió en el context del problema.

Criteris de puntuació i valoració (sobre 10): a) 4 punts (una mostra o dues, 2 punts; paràmetre, 1 punt; contrast, 1 punt). b) 1 punt. c) 1 punt. d) 4 punts (acceptar o no, 1 punt; conclusió en el context, 3 punts).

SOLUCIÓ

a) Es tracta d'un contrast de mitjanes de dues mostres aparellades. El contrast a realitzar és el següent:

$$\left. \begin{array}{l} H_0: \mu_{\text{mati}} = \mu_{\text{tarda}} \\ H_1: \mu_{\text{mati}} \neq \mu_{\text{tarda}} \end{array} \right\}$$

on μ_{mati} i μ_{tarda} representen les mitjanes de les variables aleatòries que ens donen el nombre de peticions al servidor al matí i a la tarda.

b) El valor de l'estadístic de contrast és -5.0479892.

c) El p-valor del contrast és 0.0000023.

d) Com que el p-valor és menor que α , rebutjaríem la hipòtesi nul·la i conclouríem que efectivament tenim prou evidències per afirmar el nombre de peticions al matí i a la tarda és diferent.

Problema 6. El nombre de processadors i la memòria RAM de 5 ordinadors d'una aula d'informàtica és el següent:

RAM (Gb)	1	2	2	3	2
Nombre processadors	2	4	4	8	3

a) Calculeu la recta de regressió del nombre de processadors en funció de la RAM de cada ordinador.

b) Estimeu el nombre de processadors per un ordinador amb 2 Gb i mig de memòria RAM.

Criteris de puntuació i valoració (sobre 10): a) (8 punts). Càlcul de les mitjanes, 2 punts. Càlcul de les variàncies i covariància, 2 punts. Càlcul del pendent, 2 punts. Càlcul de l'ordenada a l'origen, 2 punts. b) 2 punts.

SOLUCIÓ

a) Primer calculem les mitjanes de les dues variables:

$$\overline{RAM} = 2, \overline{Nombre.proc} = 4.2.$$

Les variàncies i la covariància seran:

$$\begin{aligned} var(RAM) &= 0.5, var(Nombre.proc) = 5.2, \\ cov(RAM, Nombre.proc) &= 1.5. \end{aligned}$$

La recta de regressió serà: $Nombre.proc = \beta_0 + \beta_1 \cdot RAM$, on

$$\beta_1 = \frac{cov(RAM, Nombre.proc)}{var(RAM)} = \frac{1.5}{0.5} = 3.$$

$$\beta_0 = \overline{Nombre.proc} - \beta_1 \cdot \overline{RAM} = 4.2 - 3 \cdot 2 = -1.8.$$

La recta de regressió serà, doncs:

$$Nombre.proc = -1.8 + 3 \cdot RAM$$

b) El nombre de processadors estimat per un ordinador amb 2.5 Gb de RAM serà:

$$-1.8 + 3 \cdot 2.5 = 5.7.$$

Examen 3

Problema 1. Les següents dades representen el temps en segons que va tardar un programa en executar-se en WINDOWS durant 12 vegades:

2.5	7.1	5	8.5	7	8.1	2.3	7.1	4	8	6.6	5
-----	-----	---	-----	---	-----	-----	-----	---	---	-----	---

- Trobeu la mitjana i els tres quartils.
- Feu un diagrama de caixa amb aquestes observacions.
- Calculeu la desviació típica poblacional.

Criteris de puntuació i valoració (sobre 10): a) 4 punts; b) 3 punts i c) 3 punts.

SOLUCIÓ:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2.5 + 7.1 + 5 + 8.5 + 7 + 8.1 + 2.3 + 7.1 + 4 + 8 + 6.6 + 5}{12} = 5.93$$

a) La mitjana val:

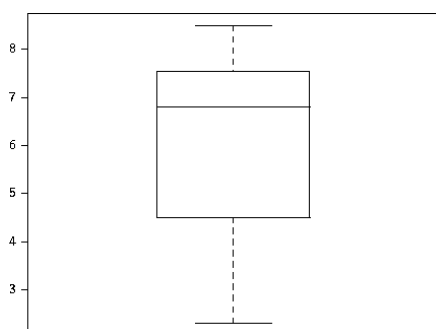
Per trobar els quartils, primer ordenem les dades: 2.3; 2.5; 4; 5; 5; 6.6; 7; 7.1; 7.1; 8; 8.1; 8.5.

El primer quartil serà la mediana de les 6 primeres dades: serà la semisuma dels valors que ocupen els llocs 3 i 4: $Q_1 = (4+5)/2 = 4.5$.

La mediana serà la semisuma dels valors que ocupen els llocs 6 i 7: $Q_2 = (6.6+7)/2 = 6.8$.

El tercer quartil serà la mediana de les 6 darreres dades: serà la semisuma dels valors que ocupen els llocs 9 i 10: $Q_3 = (7.1+8)/2 = 7.55$.

b)



c) La desviació típica poblacional valdrà:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum (2.5 - 5.93)^2 + L + (5 - 5.93)^2}{12}} = 2,0479.$$

Problema 2. Hem mesurat la velocitat de baixada (en KB/s) de 92 arxius allotjats en un servidor en un determinat moment del dia. Els resultats els hem tabulat segons la franja horària (matí, nit, tarda) i segons la velocitat (alta, baixa, mitjana, molt alta):

	V_alta	V_baixa	V_mitjana	V_molt alta
Matí	4	3	9	6
Nit	5	8	3	4
Tarda	13	13	11	13

- Calculeu la probabilitat que la velocitat de baixada sigui molt alta.
- Calculeu la probabilitat que la velocitat de baixada d'un arxiu sigui molt alta i es realitzi a la nit.
- Si sabem que l'arxiu s'ha baixat a la tarda, calculeu la probabilitat que s'hagi baixat a velocitat alta.
- Els successos "franja horària Nit" i "velocitat alta" són independents? Raoneu la resposta.

Criteris de puntuació i valoració (sobre 10): 2.5 punts cada apartat.

SOLUCIÓ:

- Calculeu la probabilitat que la velocitat de baixada sigui molt alta.

$$P(V_molt\ alta) = \frac{23}{92} = 0,25.$$

- Calculeu la probabilitat que la velocitat de baixada d'un arxiu sigui molt alta i es realitzi a la nit.

$$P(V_molt\ alta \cap Nit) = \frac{4}{92} = 0,04.$$

- Si sabem que l'arxiu s'ha baixat a la tarda, calculeu la probabilitat que s'hagi baixat a velocitat alta.

$$P(V_alta / Tarda) = \frac{P(V_alta \cap Tarda)}{P(Tarda)} = \frac{13/92}{50/92} = \frac{13}{50} = 0,26$$

d) Els successos “franja horària Nit” i “velocitat alta” són independents? Raonar la resposta.

Calculem les probabilitats següents:

$$P(\text{Nit}), P(V_alta), P(\text{Nit} \cap V_alta)$$

$$P(\text{Nit}) = \frac{20}{92} = 0,217, \quad P(V_alta) = \frac{22}{92} = 0,24, \quad P(\text{Nit} \cap V_alta) = \frac{5}{92} = 0,054$$

Com que es verifica que

$$P(\text{Nit} \cap V_alta) \neq P(\text{Nit}) \cdot P(V_alta), \text{ aleshores no són independents.}$$

Problema 3. Una variable aleatòria té la següent funció de massa de probabilitat:

Valors possibles $X=x$	$P(X=x)$
3	0.4
4	A
8	B

Sabent que la seva esperança és 4,

a) Calculeu A i B.

b) Calculeu la funció de distribució de X i la probabilitat que X sigui inferior o igual a 6.

c) Calculeu la variància de X.

Criteris de puntuació i valoració (sobre 10): S'han d'indicar les fórmules i els càlculs realitzats, així com els raonaments. a) 4 punts; b) 3 punts (2 punts la funció de distribució, 1 punt la probabilitat); c) 3 punts.

SOLUCIÓ:

a) $0.4 + A + B = 1$ i a més $3 \cdot 0.4 + 4 \cdot A + 8 \cdot B = 4$, per tant $A = 0.5$, $B = 0.1$

b) $F(3) = 0.4$, $F(4) = 0.9$, $F(8) = 1$. $P(X \leq 6) = 0.4 + A = 0.4 + 0.5 = 0.9$

c) $\text{Var} = (3-4)^2 \cdot 0.4 + (4-4)^2 \cdot 0.5 + (8-4)^2 \cdot 0.1 = 2$

Problema 4. Els ingressos mensuals d'una família segueixen una distribució normal de mitjana 1000 euros i desviació típica 80 euros. Suposem a més que l'ingrés de cada mes és independent dels altres mesos.

a) Quina distribució segueixen els ingressos anuals?

b) Calculeu la probabilitat que aquesta família ingressi més de 12300 euros en un any.

c) Quant haurien de gastar en un any per tenir una probabilitat del 95% de no entrar en nombres vermells (és a dir de no gastar més del que s'ingressa)?

Valors de probabilitats que us poden ser útils; si no trobeu exactament el que necessiteu, useu el més proper

$p(X \geq x)$	$X \sim N(0,1)$
0.0228	2
0.0359	1.8
0.05	1.645
0.0548	1.6
0.0876	1.4
0.1	1.2816
0.1151	1.2
0.1357	1.1
0.1587	1

Criteris de puntuació i valoració (sobre 10): S'han d'indicar les fórmules i els càlculs realitzats, així com els raonaments. a) 4 punts: plantejament 2 punts, càlculs 2 punts; b) 3 punts (plantejament 1 punt, càlculs 2 punts); c) 3 punts (plantejament 1 punt, càlculs 2 punts).

SOLUCIÓ:

a) Segueix una distribució normal de mitjana $12 \cdot 1000 = 12000$ i desviació típica $\sigma\sqrt{n} = 80 \cdot \sqrt{12} = 277.128129211$.

b) Demanen $p(X > 12300) = p(Z > (12300 - 12000) / 277.128129211) = p(Z > 1.082531755) = 0.1395$. Amb les taules de l'enunciat 0.1357.

c) Ara demanen $p(X > g) = 0.95$, o el que és el mateix, $p(Z > (g - 12000) / 277.128129211) = 0.95$ amb el que tenim que $(g - 12000) / 277.128129211 = -1.645$ i per tant $g = 11544.1$.

Problema 5. Volem comprovar si el percentatge de missatges SPAM rebuts per dia en dos servidors de correu és el mateix o no. Agafem una mostra aleatòria de 100 dies, calculem el percentatge de missatges SPAM rebuts en els dos servidors. Ens diuen que al primer servidor han arribat 50 missatges SPAM i al segon, 60 missatges SPAM.

a) Establiu el contrast a realitzar. Heu d'indicar si es tracta d'un contrast d'una mostra o de dues mostres, de quin paràmetre(s) hem de fer el contrast (mitjana(es), proporció(ns), etc.), i la hipòtesi nul·la i la hipòtesi alternativa.

b) Quin és el valor de l'estadístic de contrast?

c) A un nivell de significació del 5%, quin(s) és(són) el(s) valor(s) crític(s)?

d) A un nivell de significació de $\alpha = 0.05$, a quina conclusió arribaríeu? No és suficient afirmar si acceptem o rebutgem la hipòtesi nul·la, s'ha de raonar la conclusió en el context del problema.

Valors de les taules de la $N(0,1)$ (utilitzeu els que considereu necessaris, si necessiteu un valor que no és exacte, utilitzar el valor més proper)

x	p(X >= x)
2.06	0.0197
1.96	0.025
1.86	0.0314
1.76	0.0392
1.66	0.0485
1.56	0.0594
1.46	0.0721
1.36	0.0869
1.26	0.1038
1.16	0.123
1.06	0.1446
0.06	0.4761

Criteris de puntuació i valoració (sobre 10): a) 3 punts, si indica si es tracta d'una mostra o dues, 1 punt, si indica el paràmetre, 1 punt, si indica el contrast, 1 punt, b) 2 punts. Si indica la fórmula de l'estadístic, 1 punt. Valor del mateix, 1 punt. c) 1 punt. d) 4 punts (acceptar o no, 1 punt; conclusió en el context, 3 punts).

SOLUCIÓ:

a) Es tracta d'un contrast de proporcions de dues mostres independents. El contrast a realitzar és el següent:

$$\begin{aligned} H_0: & p_1 = p_2, \\ H_1: & p_1 \neq p_2, \end{aligned}$$

on p_1 i p_2 representen les proporcions de missatges SPAM rebuts en un dia pel servidor 1 i pel servidor 2 respectivament.

b) El valor de l'estadístic de contrast és

$$Z = \frac{p_1 - p_2}{\sqrt{p \cdot (1 - p) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.6}{\sqrt{0.55 \cdot (1 - 0.55) \cdot \left(\frac{1}{100} + \frac{1}{100}\right)}} = -1.421.$$

c) Els valors crítics del contrast són:

$$z_{0.025} = -1.96, z_{0.975} = 1.96.$$

d) Com que el valor de l'estadístic de contrast es troba entre els valors crítics, acceptaríem la hipòtesi nul·la i conclouríem que NO tenim prou evidències per afirmar que la proporció de missatges SPAM en els dos servidors NO sigui el mateix.

Problema 6. Volem estudiar si el percentatge de missatges SPAM rebuts en un dia en un determinat servidor de correu té relació amb el dia del mes en què els missatges són enviats. Per això hem estudiat durant dos mesos el percentatge de missatges SPAM rebuts durant els 60 dies i els resultats han estat els següents:

```
##
## Call:
## lm(formula = percentatge.spam ~ dies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.134565 -0.076360 -0.006697  0.068212  0.159768
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  0.084031   0.023317   3.604    0.000652 ***
## dies         0.020422   0.001313  15.548 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08806 on 58 degrees of freedom
## Multiple R-squared:  0.8065, Adjusted R-squared:  0.8032
## F-statistic: 241.7 on 1 and 58 DF,  p-value: < 0.0000000000000022
```

a) Trobeu la recta de regressió del percentatge de missatges SPAM per dia en funció del dia del mes i interpreteu els resultats.

b) Feu el contrast d'hipòtesi sobre el pendent de la recta, amb un nivell de significació de $\alpha = 0.01$ i interpreteu els resultats. Heu de donar el contrast sobre el pendent indicant la hipòtesi nul·la i alternativa, l'estadístic de contrast, el p-valor i la conclusió a què s'arriba.

c) Calculeu el coeficient de determinació. Discutiu si és un bon model.

Criteris de puntuació i valoració (sobre 10): a) (3 punts). Donar la recta de regressió, 1.5 punts. Interpretació dels paràmetres de la recta, 1.5 punts. b) (5 punts). Donar el contrast d'hipòtesi indicant la hipòtesi nul·la i alternativa, 1 punt. Donar el valor de l'estadístic, 1 punt. Donar el p-valor, 1 punt. Conclusió, 2 punts. c) (2 punts). Donar l'expressió del coeficient de determinació, 1 punt. Interpretació del mateix, 1 punt.

SOLUCIÓ:

a) La recta de regressió és la següent:

$$\text{percentatge.SPAM} = 0.08403 + 0.02042 \cdot \text{dies}.$$

El coeficient 0.02042 és el pendent de la recta de regressió i representa l'augment del percentatge de missatges SPAM en un dia quan el dia del mes augmenta en un dia.

El coeficient 0.08403 és l'ordenada a l'origen i no tendria interpretació real ja que vendria a ésser el percentatge de missatges SPAM el dia 0.

b) El contrast sobre el pendent de la recta és el següent:

$$\begin{aligned} H_0: & \beta_1 = 0, \\ H_1: & \beta_1 \neq 0, \end{aligned}$$

on β_1 és el pendent de la recta de regressió. L'estadístic de contrast val: 3.60378 amb un p-valor de 0.00065. Com que el p-valor és més petit que el nivell de significació, podem rebutjar la hipòtesi nul·la i afirmar que tenim prou indicis per acceptar que el pendent de la recta de regressió no és nul.

c) El valor del coeficient de determinació val 0.8065. Es un valor proper a 1. Per tant, podem dir que la regressió és moderadament bona.