EXAMEN 1 (19/1/2019)

EXERCICI 1

La taula recull el nombre de vegades que ha sortit cadascuna de les puntuacions en 20 llançaments d'un dau.

PUNTUACIÓ	1	2	3	4	5	6
N.º DE VEGADES	5	3	2	6	2	2

a) Calculeu la mitjana.

b) Calculeu la seva desviació típica i la variància poblacional.

c) Calculeu la mediana i els quartils.

d) Feu el diagrama de caixes de la variable.

Observació: cal escriure les fórmules i els passos intermedis en cada apartat.

Criteris de puntuació i valoració (sobre 10): a) 2 punts; b) 3 punts; c) 3 punts; d) 2 punts.

SOLUCIÓ:

a)Mitjana

$$\acute{x}=\frac{\sum n_i x_i}{N}=\frac{63}{20}=3.15$$

b)La variància i la desviació típica poblacional són:

$$S_x^2=\frac{\sum n_i x_i^2}{N}-\acute{x}^2=\frac{253}{20}-3.15^2=12.65-9.9225=2.7275$$

$$S_x=\sqrt{2.7275}=1.651514$$

c)Escrivim les 20 dades ordenades:

1 1 1 1 1 ↑ 2 2 2 3 3 ↑ 4 4 4 4 4 ↑ 4 5 5 6 6

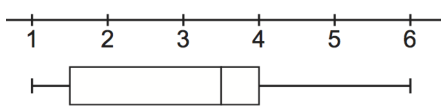
Q_1 Me Q_3

20 : 2 = 10. La mediana està entre les posicions 10.º i 11.º, Me = 3.5.

20 : 4 = 5. Q1 està entre els llocs 5.º i 6.º, Q1= 1,5.

5 · 3 = 13.5. Q3 està entre els llocs 15.º i 16.º, Q3= 4.

d)



EXERCICI 2

A un congrés de noves tecnologies hi assisteixen 1000 persones repartides tal i com es mostra a la taula següent:

	HOME	DONA
PARLEN ANGLÈS	515	310
NO PARLEN ANGLÉS	95	80

Anomenem H = home, D= dona, A = parla anglès, NO A = no parla anglès.

a) Calculeu les probabilitats següents:

P [H], P [D], P [A], P [NO A]

b) Descriviu els esdeveniments següents i calculeu les seves probabilitats:

D∩ A ; (NO A)|H ; H|(NO A)

c) Són independents els esdeveniments D i A? Raoneu la resposta.

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

SOLUCIÓ:

a)

$$P(H)=\frac{515+95}{1000}=\frac{610}{1000}=0.61$$

$$P(D)=\frac{310+80}{1000}=\frac{390}{1000}=0.39$$

$$P(A)=\frac{310+515}{1000}=\frac{825}{1000}=0.825$$

$$P(\text{NO}\quad A)=\frac{95+80}{1000}=\frac{175}{1000}=0.175.$$

b) D∩ A = Dona que parla anglès:

$$P(D\cap A)=\frac{310}{1000}=0.31.$$

(NO A)|H = No parla anglès sabent que és home:

$$P(NO\,A|H)=\frac{95}{610}=0.156.$$

H|(NO A) = Home sabent que no parla anglès

$$P(H|\text{NO}\quad A)=\frac{95}{175}=0.543.$$

c) Hem vist a l’apartat b) que P(D∩ A) = 0,31 i a l’apartat a) que P(D) = 0,39 i P(A) =

0,825. Per tant:

$P(dona \cap parla\,anglès) \neq P(dona) \cdot P(parla\,anglès)$

I per tant no són independents.

EXERCICI 3

Disposem d'un dau no perfecte on les probabilitats de cada cara venen donades per la taula següent:

Valor	1	2	3	4	5	6
Probabilitat	a	$2a$	$3a$	3/12	2/12	1/12

Anomenem X al resultat obtingut quan tirem aquest dau.

a) Determineu a per tal que sigui una distribució de probabilitat.
b) Calculeu la probabilitat d'obtenir un resultat parell quan tirem el dau
c) Si anomenem X a una variable discreta amb aquesta distribució, determineu el valor de x tal que $P(X \geq x) = 0.750$.
d) Calculeu l'esperança d'aquesta variable.

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

SOLUCIÓ:

a)Hem d'imposar que $\sum_{i=1}^{6} P(X = x_i) = 1$. Per tant

$$P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)+P(X=6)$$

$$a+2a+3a+\frac{3}{12}+\frac{2}{12}+\frac{1}{12}$$

$$6a+\frac{6}{12}=1$$

Per tant $a=1/12$.

b)$P(\text{parell}) = P(X=2)+P(X=4)+P(X=6)=\frac{2}{12}+\frac{3}{12}+\frac{1}{12}=0.5$

c) Com que

$$P(X \geq x) = 1 - P(X < x) \Rightarrow P(X < x) = 1 - 0.75 = 0.25 = \frac{3}{12}$$

Està clar que es compleix per a tot $x \in [2, 3)$.

d)

$$E[X]=\sum_{\forall x_i} x_i \cdot P(X=x_i)$$

$$1\cdot P(X=1)+2\cdot P(X=2)+3\cdot P(X=3)+4\cdot P(X=4)+5P(X=5)+6P(X=6)$$

$$1\frac{1}{12}+2\frac{2}{12}+3\frac{3}{12}+4\frac{3}{12}+5\frac{2}{12}+6\frac{1}{12}=\frac{1+4+9+12+10+6}{12}=\frac{42}{12}=\frac{7}{2}.$$

EXERCICI 4

Hem de fer un control de l'acceleració d'un nou model de cotxe. Mesurem 10 vegades el temps (en segons) que tarda a arribar als 100km/h. Obtenim les dades següents:
20, 19.2, 19.5, 16, 20.1, 20.2, 21, 19.8, 19.9, 18.
Suposem que la variable temps segueix una distribució normal. Aleshores:

a) Trobeu un interval de confiança al 98% per la mitjana del temps que tarda en arribar a la velocitat de 100 km/h.
b) Suposem ara que coneixem la desviació típica de la variable anterior temps i que val 1.4segons. Determineu un interval de confiança al 95% per la mitjana del temps que tarda en arribar a la velocitat de 100 km/h.

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

p(X>= x)	X~ N(0,1)	X~ t de Student amb 9 graus de llibertat	X~ t de Student amb 10 graus de llibertat
0.01	2.326	2.821	2.763
0.025	1.96	2.262	2.228
0.05	1.645	1.833	1.812
0.005	2.575	3.249	3.169

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

SOLUCIÓ:

a) Hem de trobar un interval de confiança per la mitjana quan la població és normal i desconeixem la desviació típica. Calculem la mitjana i la desviació típica mostra:

$$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n}=19.37$$

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}=1.41$$

L'interval de confiança és

$$\left(\bar{X}-t_{0.01,9}\frac{s}{\sqrt{n}},\bar{X}+t_{0.01,9}\frac{s}{\sqrt{n}}\right)=\left(19.37-2.821\frac{1.41}{\sqrt{10}};19.37+2.821\frac{1,41}{\sqrt{10}}\right)$$

$$(18.1108;20.6292).$$

b) Ara hem de trobar un interval de confiança per la mitjana quan la població és normal i coneixem la desviació típica. En aquest cas $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ segueix una distribució normal estàndard. La condició

$P\left(-z \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right)=0.95,$ ens dóna $z=1.96$.

L'interval de confiança és aleshores

$\left(\overline{X}-z\frac{\sigma}{\sqrt{n}}, \overline{X}+z\frac{\sigma}{\sqrt{n}}\right)=\left(19.37-1.96\frac{1.4}{\sqrt{10}}; 19.37+1.96\frac{1.4}{\sqrt{10}}\right)=(18.51; 20.23).$

EXERCICI 5

Una empresa sospita que la memòria RAM que té el servidor del Departament de ``Big Data'' es molt justa i s'ha d'ampliar. Per veure si efectivament la memòria és justa, s'han escollit a l'atzar 50 dies i s'ha vist que uns 38 dies el servidor ha sofert retards importants. Els directius de l'empresa prenen la decisió d'ampliar la memòria RAM si més del 70% dels dies el servidor sofreix retards importants. Tenim prou evidències, en vista de les dades obtingudes, per aconsellar a l'empresa que ha d'ampliar la memòria RAM del servidor?

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 d'enraonar la conclusió en el context del problema.

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

P(X>= x)	x~ N(0,1)
0.0197	2.06
0.025	1.96
0.0314	1.86
0.0392	1.76
0.0495	1.65
0.0594	1.56
0.0721	1.46
0.0869	1.36
0.1038	1.26
0.123	1.16
0.1446	1.06
0.4761	0.06

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. Si només diu que s'accepta o rebutja la hipòtesi nul·la, 1 punt.

SOLUCIÓ:

a) Es tracta d'un contrast d'una proporció. El contrast a realitzar és el següent:

$$\left.\begin{aligned} H_0 &: \quad p = 0.7, \\ H_1 &: \quad p > 0.7, \end{aligned}\right\}$$

on p representa el percentatge (proporció) de dies per mes en que el servidor està saturat.

b) El valor de l'estadístic de contrast és

$$Z = \frac{\hat{p} - 0.7}{\sqrt{\frac{0.7 \cdot (1 - 0.7)}{n}}} = \frac{0.76 - 0.7}{\sqrt{\frac{0.7 \cdot (1 - 0.7)}{50}}} = 0.926.$$

c) El valor crítics del contrast són:

$z_{0.05} = 1.645.$

d) Com que el valor de l'estadístic de contrast és menor que el valor crític, acceptaríem la hipòtesi nul·la i aconsellaríem a l'empresa no ampliar la memòria RAM del servidor.

EXERCICI 6

Se sospita que el temps en que un servidor web va saturat en un dia depèn del nombre de missatges SPAM que ha rebut aquell dia. Per això, s'agafa una mostra de 100 dies aleatòriament i es calcula la recta de regressió del nombre de missatges SPAMS rebuts cada dia en funció del temps de saturació (en segons) en el mateix dia. Els resultats obtinguts amb R són els següents:

```
##
## Call:
## lm(formula = missatges.spam ~ temps.saturació)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.889  -2.293  -0.522   2.195  16.942
##
## Coefficients:
##                  Estimate Std. Error t value            Pr(>|
t|)
## (Intercept)     0.4070307  0.9634604   0.422
0.674
## temps.saturació 0.0192476  0.0003218  59.808
<0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1
##
## Residual standard error: 4.762 on 98 degrees of freedom
## Multiple R-squared:  0.9733, Adjusted R-squared:  0.9731
## F-statistic:  3577 on 1 and 98 DF,  p-value: <
0.00000000000000022
```

a) Trobeu la recta de regressió del nombre de missatges spam en funció del temps de saturació i interpreteu els resultats.
b) Feu el contrast d'hipòtesi damunt el pendent de la recta, amb un nivell de significació de $\alpha = 0.05$ 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 la que 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:

$missatges.spam = 0.40703 + 0.01925 \cdot temps.saturació$.

El coeficient 0.01925 és el pendent de la recta de regressió i representa l'augment del nombre de missatges spam quan el temps de saturació augmenta en un segon un dia qualsevol.

El coeficient 0.40703 és l'ordenada a l'origen i vindria a ésser el nombre de missatges spams rebuts en un dia quan el servidor no ha estat saturat aquell dia.

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

$$\left.\begin{array}{ll} H_0: & \beta_1 = 0, \\ H_1: & \beta_1 \neq 0, \end{array}\right\}$$

on β_1 és el pendent de la recta de regressió. L'estadístic de contrast val: 59.8077 amb un p-valor de 0. 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 0.97333. És un valor proper a 1. Per tant, podem dir que la regressió és bastant bona.

EXAMEN 2 (12/1/2019)

EXERCICI 1

Les notes d'una classe en un examen d'estadística venen recollides a la taula següent:

Nota	1	2	3	4	5	6	7	8	9	10
N. d'alumnes	1	1	2	2	6	4	5	3	3	2

a) Calculeu la mitjana i la desviació típica poblacional ($\sigma = S_x$).

b) Quin percentatge d'alumnes hi ha entre $\acute{x}-\sigma$ i $\acute{x}+\sigma$.

c) Quin percentatge d'alumnes està per sobre de la mitjana?

d) Feu un histograma de la variable.

Observació: cal escriure les fórmules i els passos intermedis en cada apartat.

Criteris de puntuació i valoració (sobre 10): a) 4 punts; b) 2 punts; c) 2 punts, d) 2 punts.

SOLUCIÓ:

a)

x_i	n_i	$n_i x_i$	$n_i x_i^2$
1	1	1	1
2	1	2	4
3	2	6	18
4	2	8	32
5	6	30	150
6	4	24	144
7	5	35	245
8	3	24	192
9	3	27	243
10	2	20	200
	29	177	1 229

$$\acute{x}=\frac{\sum n_i x_i}{N}=\frac{177}{29}=6.1$$

$$S_x^2=\frac{\sum n_i x_i^2}{N}-\acute{x}^2=\frac{1229}{29}-6.1^2=5.17$$

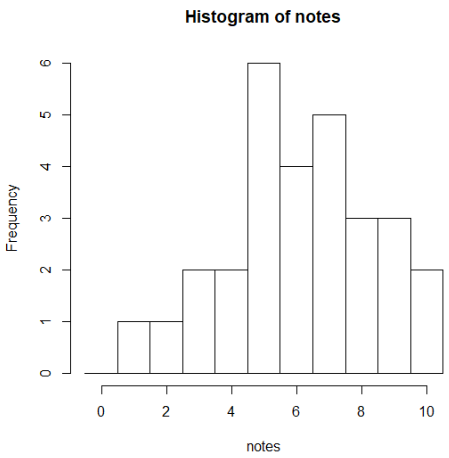
$$S_x=\sqrt{5.17}=2.27$$

La nota mitjana de la classe és de 6.1, amb una desviació típica de 2.27.

b) $\bar{x}-\sigma=3.83$ i $\bar{x}+\sigma=8.37$, entre 3.83 i 8.37 hi ha 20 alumnes que representen un 68.97% del total.

c) Per sobre de 6.1 hi ha 13 alumnes, que representen un 44,83 % del total.

d)



EXERCICI 2

En una cadena de televisió es va fer una enquesta a 2 500 persones per saber l'audiència d'un debat i d'una pel·lícula que es van emetre en hores diferents: 2 100 van veure la pel·lícula, 1 500 van veure el debat i 350 no van veure cap dels dos programes.

a) Organitzeu la informació en una taula de contingència, completant les dades que falten.

Si triem a l'atzar a un dels enquestats:

b) Quina és la probabilitat que veiés la pel·lícula i el debat?
c) Quina és la probabilitat que veiés la pel·lícula, sabent que va veure el debat?
d) Sabent que va veure la pel·lícula, quina és la probabilitat que veiés el debat?

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

SOLUCIÓ:

a)

	DEBAT	NO DEBAT	
PEL.LÍCULA	1 450	650	2 100
NO PEL.LÍCULA	50	350	400
	1 500	1 000	2 500

Diem D = "Va veure el debat" i P = "Va veure la pel.lícula".

b)

$$P[D \cap P] = \frac{1\,450}{2\,500} = \frac{29}{50} = 0{,}58$$

c)

$$P[P/D] = \frac{1\,450}{1\,500} = \frac{29}{30} = 0{,}97$$

d)

$$P[D/P]=\frac{1\,450}{2\,100}=\frac{29}{42}=0{,}69$$

EXERCICI 3

El temps en minuts que dura un determinat procés es pot modelitzar utilitzant una variable aleatòria X amb distribució uniforme U(1,3).

a) Determineu la funció de distribució d'aquesta variable.

b) Calculeu la probabilitat que duri menys de 2.5 minuts.

c) Determineu el valor de x tal que $P(X\geq x)=0.750$.

d) Calculeu l'esperança d'aquesta variable.

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

SOLUCIÓ:

a) Hem de calcular $F(x)=P(X\leq x)$. Per $x\in(1,3)$ obtenim

$$F(x)=P(X\leq x)=\int_1^x \frac{1}{2}\,\mathrm{dx}=\frac{1}{2}(x-1).$$

A més, per $x\leq 1$ tenim que $F(x)=0$ i per $x\geq 3$ tenim que $F(x)=1$.

b) $P(X\leq 2.5)=F(2.5)=\frac{1}{2}(2.5-1)=0.75$

c) Com que

$$P(X\geq x)=1-P(X<x)\Rightarrow P(X<x)=1-0.75=0.25$$

I per tant

$$F(x)=\frac{1}{2}(x-1)=0.25\Rightarrow$$
$$x-1=0.5\Rightarrow$$
$$x=1.5$$

d) $E[X]=\int_1^3 \frac{1}{2}\,\mathrm{xdx}=\left.\frac{x^2}{4}\right|_1^3=\frac{9}{4}-\frac{1}{4}=2$

EXERCICI 4

Durant un dia una botiga de mobles rep 25 clients. La despesa que fa cada client en euros, segueix una distribució normal de mitjana 800 euros i desviació típica 100 euros. Suposem, a més, que la despesa de cada client és independent de la resta de clients.

a) Quina distribució segueix l'ingrés total de la botiga en un dia.

b) Calculeu la probabilitat que l'ingrés total en un dia sigui inferior a 19000 euros.

c) Per mantenir la botiga oberta necessitem uns ingressos totals de y euros. Quin és el y màxim que podem suportar per tenir una probabilitat del 95% de mantenir la botiga oberta?

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

P(X>= x)	X~ 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; b) 3 punts (plantejament 1 punt, càlculs 2 punts); c) 3 punts (plantejament 1 punt, càlculs 2 punts).

SOLUCIÓ:

a) L'ingrés total en un dia segueix una distribució normal de mitjana 25*800=20000 i desviació típica$\sigma\sqrt{n}=100*\sqrt{25}=500$.

b) Demanen

$$P(X<19000)=P\left(Z<\frac{19000-20000}{500}\right)=P(Z<-2)=P(Z>2)=0.0228$$

on Z indica una normal estàndard.

c) Ara demanen P(X>y)=0.95, o el que és el mateix,

$$P(X>y)=P\left(Z>\frac{y-20000}{500}\right)=0.95.$$

És a dir,

$$P\left(Z>-\frac{y-20000}{500}\right)=0.05.$$

i $-\frac{y-20000}{500}=1.645$ és a dir que y=19177.5

EXERCICI 5

Volem contrastar si el percentatge màxim d'ús diari de la targeta de xarxa en un determinat servidor supera el 75%. Escollim 10 dies a l'atzar de forma independent i obtenim que la mitjana del percentatge màxim d'ús diari de la targeta de xarxa ha estat de 80% amb una desviació típica de $s=15\%$. Suposant que la variable que ens dóna el percentatge màxim d'ús diari de la targeta de xarxa del servidor 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 d'enraonar la conclusió en el context del problema.

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

p(X>= x)	X~ N(0,1)	X~ t de Student amb 9 graus de llibertat	X~ t de Student amb 62 graus de llibertat
0.01	2.326	2.821	2.764
0.025	1.96	2.262	2.228
0.05	1.645	1.833	1.812

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.

SOLUCIÓ:

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

$$\left.\begin{array}{l} H_0: \quad \mu = 75, \\ H_1: \quad \mu > 75. \end{array}\right\}$$

on μ representa el percentatge màxim d'ús diari de la tarjeta de xarxa del servidor.

b) El valor de l'estadístic de contrast és

$$t = \frac{\overline{x} - 75}{\frac{s}{\sqrt{n}}} = \frac{80 - 75}{\frac{15}{\sqrt{10}}} = 1.054.$$

c)El valor crític serà:

$$t_{\alpha,9} = t_{0.05,9} = 1.833.$$

d) Com que el valor de l'estadístic de contrast és menor que el valor crític, acceptaríem la hipòtesi nul·la i conclouríem que efectivament no tenim prou evidències per afirmar que la mitjana del percentatge màxim d'ús diari de la targeta de xarxa del servidor és major que el 75%.

EXERCICI 6

Volem estudiar si el nombre de "likes" rebuts per dia pels usuaris de Facebook està relacionat amb el nombre d'amics de cada usuari. Per això, hem escollit aleatòriament 50 usuaris de Facebook i hem calculat la mitjana per dia de "likes" i la mitjana per dia del nombre d'amics durant un mes de cada usuari. Hem calculat la recta de regressió amb R i els resultats obtinguts són els següents:

```
##
## Call:
## lm(formula = likes ~ amics)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.26062 -0.05293 -0.01281  0.06573  0.22618 
## 
## Coefficients:
##                Estimate Std. Error t value            Pr(>|t|)    
## (Intercept) -0.2127122  0.0555192  -3.831             0.00037 ***
## amics        0.0999490  0.0005345 186.997 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.1093 on 48 degrees of freedom
## Multiple R-squared:  0.9986, Adjusted R-squared:  0.9986 
## F-statistic: 3.497e+04 on 1 and 48 DF,  p-value: < 
0.00000000000000022
```

a) Trobeu la recta de regressió de la mitjana d'amics mensual en funció de la mitjana de "likes" mensual i interpreteu els resultats.

b) Feu el contrast d'hipòtesi damunt 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 la que 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:

$likes = -0.21271 + 0.09995 \cdot amics$

El coeficient 0.09995 és el pendent de la recta de regressió i representa l'augment del nombre mitjà mensual de "likes" quan el nombre mitjà mensual d'amics augmenta en un.

El coeficient -0.21271 és l'ordenada a l'origen i no tendria interpretació real ja que vendria a ésser el nombre de "likes" d'un usuari sense cap amic i això és impossible.

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

$$\left. \begin{array}{l} H_0: \quad \beta_1 = 0, \\ H_1: \quad \beta_1 \neq 0, \end{array} \right\}$$

on β_1 és el pendent de la recta de regressió. L'estadístic de contrast val: 186.99743 amb un p-valor de 0. 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 0.99863. Es un valor molt proper a 1. Per tant, podem dir que la regressió és molt bona.

EXAMEN 3 (19/1/2019)

EXERCICI 1

S'ha demanat a un grup de joves pel nombre d'hores que va dedicar diumenge passat a veure la televisió. Les respostes s'han presentat a la següent taula:

HORES	N. DE JOVES
0	1
1	2
2	8
3	10
4	7
5	2

a) Trobeu la mitjana i la variància mostral.
b) Construïu la taula de freqüències acumulades i, utilitzant-la, trobeu la mitjana i els quartils.
c) Feu el diagrama de caixes.

Observació: cal escriure les fórmules i els passos intermedis en cada apartat.

Criteris de puntuació i valoració (sobre 10): a) 4 punts; b) 4 punts; c) 2 punts.

SOLUCIÓ:

a) Mitjana

$$\acute{x}=\frac{\sum n_i x_i}{30}=\frac{86}{30}=2.867$$

La variància poblacional:

$$S_x^2=\frac{\sum n_i x_i^2}{N}-\acute{x}^2=\frac{286}{30}-2.8667^2=9.5333-2.8667^2=1.315556$$

Però ens demanen la mostral:

$$\frac{NS_x^2}{N-1}=\frac{30*1.315556}{29}=1.36092$$

b)

HORES	F. ABSOLUTA	F. ACUMULADA
0	1	1
1	2	3
2	8	11
3	10	21

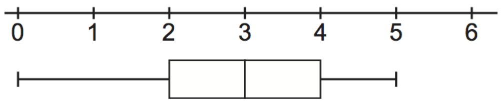
4	7	28
5	2	30

El total de dades s 30, i 30 : 2 = 15. La mediana està entre els llocs 15.º y 16.º. Mirant en la columna de las freqüències acumulades, des de el jove 12.º al 21.º van estar veient la televisió 3 hores. Por tant, $Me = 3$.

30 : 4 = 7.5. El primer quartil està en el 8.º lloc. Mirant a la columna de freqüències acumulades, des de el jove 4.º al 11.º van estar veient la televisió 2 hores, $Q_1 = 2$.

7,5 · 3 = 22.5. El tercer quartil està en el lloc 23.º. Mirant la columna de freqüències acumulades, des de el jove 12.º fins el 28.º van estar veient la televisió 4 hores, $Q_3 = 4$.

c)



EXERCICI 2

Es fa un estudi d'un nou tractament sobre 120 persones amb una certa malaltia. Se sap que 30 d'elles ja havien patit aquesta malaltia amb anterioritat. Entre les que l'havien patit, al 80% els funciona el nou tractament. Entre les que no l'havien patit, els funciona en el 90% dels casos.

a) Feu l'arbre de probabilitats.

b) Si triem dos pacients a l'atzar, quina és la probabilitat que els dos hagin patit la malaltia?

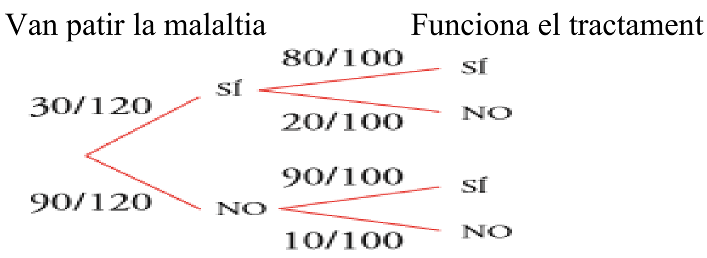
c) Determineu la probabilitat que en triar un pacient a l'atzar, no li funcioni el nou tractament.

d) Si a un pacient no li funciona el nou tractament, quina és la probabilitat que no hagi patit la malaltia amb anterioritat?

Criteris de puntuació i valoració (sobre 10): a) 4 punts; b), c) i d) 2 punts cada un.

SOLUCIÓ:

a) Ens basem en el següent diagrama en arbre:



b) $P(\text{Els dos van patir la malatia}) = \frac{30}{120} \cdot \frac{29}{119} = \frac{29}{476} = 0.061$

c) $P(\text{No li funciona}) = \frac{30}{120} \cdot \frac{20}{100} + \frac{90}{120} \cdot \frac{10}{100} = \frac{1}{8} = 0.125$

d) $P(\text{No va patir la malatia} / \text{no li funciona}) = \frac{\frac{90}{120} \cdot \frac{10}{100}}{\frac{1}{8}} = \frac{3}{5} = 0.6$

EXERCICI 3

El 20% dels treballadors d'una empresa va a la vaga. Se seleccionen 10 treballadors d'aquesta empresa. Anomenem X a la variable que dóna el número d'assistents a la vaga entre els 10 seleccionats.

a) Digueu quina llei segueix la variable X i doneu la seva esperança.

b) Calculeu la probabilitat que almenys dos vagin a la vaga.

c) Calculeu la probabilitat que la meitat vagin a la vaga.

d) Calculeu $E(3X+2)$ i $\mathrm{Var}(2X)$.

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

SOLUCIÓ:

a) X seguirà una distribució binomial de paràmetres 10 i 0.2. És a dir tenim $X \approx B(10, 0.2)$. Per tant, la seva esperança serà $E(X) = 10 * 0.2 = 2$.

b) $P(X \geq 2) = 1 - (P(X=0) + P(X=1)) = 1 - (0.8^{10} + 10 * 0.2 * 0.8^{9}) = 0.6241$.

c) $P(X=5) = \binom{10}{5} 0.8^5 0.2^5 = 0.02642$

d) Calculem
$$E(3X+2) = 3E(X) + 2 = 8,$$
$$\mathrm{Var}(2X) = 4\,\mathrm{Var}(X) = 4(10 * 0.2 * 0.8) = 4 * 1.6 = 6.4$$

EXERCICI 4

El pes d'una població juvenil s'ajusta a una llei normal. Agafem una mostra de 62 joves, i obtenim una mitjana mostral de $\bar{x} = 70$Kg., mentre que la desviació típica mostral és $s = 6$Kg.

a) Quin és l'interval de confiança per a la mitjana mostral 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>= x)	X~ N(0,1)	X~ t de Student amb 61 graus de llibertat	X~ t de Student amb 62 graus de llibertat
0.01	2.326	2.389	2.388
0.025	1.96	1.999	1.998
0.05	1.645	1.670	1.669
0.005	2.575	2.658	2.657

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 típica desconeguda: $\left(\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}\right) = \left(70 \pm t_{0.05,61} \frac{6}{\sqrt{62}}\right) = \left(70 \pm 1.67 \frac{6}{\sqrt{62}}\right) = (68{,}728, 71.272)$

b) Si la longitud és 2 aleshores el marge d'error és 1 i tenim que

$n \geq \left(t_{\frac{\alpha}{2}, n-1}\right)^2 \left(\frac{s}{1}\right)^2 = 1.67^2 \times 6^2 = 100.4$

La mida de la mostra haurà de ser de 101 individus.

EXERCICI 5

Tenim dos algorismes que calculen els principals estadístics de les profunditats de les fulles d'un arbre multiforcat (mitjana, moda, variància, etc.) Volem contrastar si el temps d'execució en microsegons dels dos algorismes és el mateix o no. Per fer el contrast agafem un arbre d'un nombre gran de fulles i executem els dos algorismes unes 100 vegades en R. Els resultats varen ésser els següents:

```
##
##  Welch Two Sample t-test
##
## data:  algorisme1 and algorisme2
## t = -44.95, df = 145.09, p-value < 0.0000000000000022
## alternative hypothesis: true difference in means is not equal
to 0
## 95 percent confidence interval:
##  -5.338969 -4.889231
## sample estimates:
## mean of x mean of y
##   15.0111   20.1252
```

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) Quin és l'interval de confiança associat al contrast anterior?

e) 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 d'enraonar 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) 1 punt. c) 1 punt. d) 1 punt. e) 4 punts. Si només diu que s'accepta o rebutja la hipòtesi nul·la, 1 punt.

SOLUCIÓ:

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

$$\left.\begin{array}{ll} H_0: & \mu_1 = \mu_2, \\ H_1: & \mu_1 \neq \mu_2, \end{array}\right\}$$

on μ_1 i μ_2 representen les mitjanes del temps d'execució dels algorismes 1 i 2, respectivament.

b) El valor de l'estadístic de contrast és -44.95.

c) El p-valor del contrast és 0.00000000000000022

d) L'interval de confiança associat al contrast anterior és:

```
## [1] -5.338969 -4.889231
## attr(,"conf.level")
## [1] 0.95
```

e) Com que el p-valor és menor que α, rebutjaríem la hipòtesi nul·la i conclouríem que efectivament tenim prou evidències per afirmar que el temps d'execució dels dos servidors és diferent.

EXERCICI 6

Volem saber si el temps d'execució en microsegons d'un programa que ens troba el mínim d'un determinat índex de balanç d'un arbre binari sobre el conjunt d'arbres binaris amb n fulles depèn o no linealment de n. Per això, es fa executar el programa 5 vegades amb els resultats següents:

Nombre de fulles	10	15	20	25	30
Temps d'execució	3.5	5	5.25	6.75	8.25

a) Calculeu la recta de regressió del temps d'execució en funció del nombre de fulles dels arbres.

b) Estimeu el temps d'execució pel conjunt d'arbres binaris de 22 fulles.

c) Es tracta d'un bon model? Raoneu la resposta.

Criteris de puntuació i valoració (sobre 10): a) (5 punts). Càlcul de les mitjanes, 1 punt. Càlcul de les variàncies i covariància, 2 punts. Càlcul del pendent, 1 punt. Càlcul de l'ordenada a l'origen, 1 punt. b) 2 punts. c) (3 punts) Càlcul del coeficient corresponent, 2 punts. Interpretació, 1 punt.

SOLUCIÓ:

a) Primer calculem les mitjanes de les dues variables:
$\overline{nombre.fulles}=20, \overline{temps.execucio}=5.75$.

Les variàncies i la covariància seran:
$var(nombre.fulles)=62.5, var(temps.execucio)=3.28125,$
$cov(nombre.fulles, temps.execucio)=14.0625.$

La recta de regressió serà: $temps.\hat{execucio}=\hat{\beta}_0+\hat{\beta}_1\cdot nombre.fulles$, on

$$\hat{\beta}_1=\frac{cov(nombre.fulles, temps)}{var(nombre.fulles)}=\frac{14.0625}{62.5}=0.225.$$

$$\hat{\beta}_0=\overline{temps}-\hat{\beta}_1\cdot\overline{nombre.fulles}=5.75-0.225\cdot 20=1.25.$$

La recta de regressió serà, doncs:
$temps.execucio=1.25+0.225\cdot nombre.fulles$

b) El temps d'execució pel conjunt d'arbres de 22 fulles val:
$1.25+0.225\cdot 22=6.2$.

c) El coeficient de determinació serà:

$$R^2=\frac{cov(nombre.fulles, temps)^2}{var(nombre.fulles)\cdot var(temps)}=\frac{14.0625^2}{62.5\cdot 3.28125}=0.9643.$$

És un coeficient molt bo. La regressió seria molt bona, és a dir, per un nombre concret de fulles podem obtenir una bona estimació del temps d'execució.