

## Problema 1 Bloc A

Les consultes *on-line* a un proveïdor d'un servei arriben per diverses línies de comunicació, principalment via Web i via App, repartides en 60% i 40% respectivament. Les consultes via Web es reparteixen igual entre les que porten un fitxer adjuntat i les que no; i un 32% són consultes via App i no tenen un fitxer adjuntat.

1.- (1 punt) Indiqueu l'arbre i el conjunt de resultats amb les seves probabilitats

2.- (1 punt) Calculeu (a) la probabilitat que una consulta es faci via App i tingui fitxer adjuntat; (b) la probabilitat que una consulta amb fitxer adjuntat hagi estat feta via App

3.- (1 punt) Podem assumir independència entre la via de consulta i el fet de portar fitxer adjuntat o no? Raoneu la resposta

El temps per resoldre consultes pot ser d'unes hores o allargar-se dies. Es coneixen les funcions de densitat i distribució del temps per resoldre consultes:  $\ln(2) 2^{-x}$  i  $1 - 2^{-x}$  per  $x > 0$  ( $x$  en dies)

4.- (1 punt) Calculeu la probabilitat d'estar més de 48 hores per resoldre una consulta

5.- (1 punt) Calculeu el valor de la mediana i interpreteu el valor

També es té quantificat els % de consultes segons el nombre de temes que impliquen (1, 2 o 3, tenint en compte que el màxim és 3) i el nombre de caràcters típic en que s'expressen (100 o 200 caràcters). Qualsevol combinació de nombre de temes i longitud és igual de probable, excepte dues que no es donen mai: consultes curtes amb 3 temes, o consultes llargues amb 1 tema.

6.- (1 punt) Indiqueu les probabilitats conjuntes i marginals entre nombre de temes i nombre de caràcters

	1	2	3
100			
200			

7.- (2 punts) Calculeu l'esperança i la variància del nombre de temes en les consultes i del nombre de caràcters

8.- (2 punts) Calculeu la correlació entre el nombre de temes i els caràcters, i interpreteu-la

NOM: \_\_\_\_\_ COGNOM: \_\_\_\_\_

Contesteu cada pregunta en el seu lloc. Explicieu justifiqueu els càlculs.

**Problema 2 (B)**

La biblioteca del barri està fent un estudi dels préstecs de documents que es realitzen. En una primera fase han estudiat els diferents documents que s'agafen en préstec a través de les màquines d'autogestió i han trobat que, de mitjana, en un 15% dels préstecs l'usuari demana una revista i en un 6% en demana una pel·lícula. Si en un matí en una de les màquines s'han efectuat 80 préstecs,

1.- Quina és la probabilitat que s'hagin prestat 10 pel·lícules? [Indica el model de probabilitat emprat] (1 punt)

2. Quin és el nombre esperat de revistes prestades durant el matí? (0.5 punts)

3. Quina és la probabilitat que el desè document en ser prestat sigui la primera revista prestada? [Indica el model de probabilitat emprat] (0.5 punts)

A més de les revistes, la biblioteca també recull els llibres infantils que es presten. En concret han trobat que el préstec dels llibres infantils es distribueixen d'una manera aleatòria seguint un model de Poisson amb una mitjana de 5 llibres l'hora.

4a) Quina és la probabilitat que en una hora s'hagin prestat 3 llibres infantils? (0.5 punts)

4b) Suposem que el nombre de llibres infantils prestats es comporta de manera independent segons l'hora del dia. En un matí, la biblioteca està oberta 4 hores. Quina és la probabilitat que s'hagin prestat 25 llibres infantils al llarg de tot el matí? (0.5 punts)

Considerem ara el temps, en minuts, entre préstecs de llibres infantils.

5a) Quin és el model de probabilitat que modelitza aquest fenomen i indica quin és el temps esperat (en minuts) entre dos préstecs de llibres infantils. (0.5 punts)

5b) Quin és el temps d'espera, en minuts, entre dos préstecs de llibres infantils que podem garantir que se superarà amb un 90% de probabilitat? (1 punt)

6.- Entre tots els nens que han participat en l'estudi se sorteja la participació en un taller especial. Per això hi ha una bossa amb 10 boles on a cada bola hi ha una lletra de la paraula BIBLIOTECA. Cada participant ha d'extreure 5 boles havent retornat a la bossa cada bola extreta abans d'agafar la següent. Podran participar en el taller especial aquells nens que treguin almenys dues boles amb lletra B. Quina és la probabilitat de participar en el taller? (1 punt)

7. Es vol estudiar el temps de préstec de les novel·les de la biblioteca. S'ha trobat que un 3% de les novel·les es tornen abans de 12 dies i que per l'altra banda, un 5% dels llibres es tornen més enllà dels 40 dies. Sabent que el temps de préstec segueix un model normal, determineu els valors de la mitjana i de la desviació tipus. (2 punts)

b) El temps de préstec d'un llibre infantil segueix un model normal amb mitjana 20 dies i desviació 5 dies. Quina és la probabilitat que un usuari tingui un llibre infantil menys de 15 dies en préstec? (0.5 punts)

c) Quina és la probabilitat que la mitjana dels temps de préstec de 100 llibres sigui menys de 19 dies ? (1 punt)

8. El préstec interbibliotecari té cada vegada més usuaris. L'empresa que el gestiona factura el servei en funció del pes dels llibres que transporta. S'ha estudiat que el pes dels llibres segueix una normal de mitjana 1200 g i una desviació de 200 g. Quina és la probabilitat que un total de 400 llibres pesin més de 475 kg? (1 punt)

Valors per la distribució normal estandaritzada  $Z(0,1)$

<code>pnorm(0,25)=</code>	0,5987	<code>pnorm(1,75)=</code>	0,9599	<code>qnorm(0,55)=</code>	0,126	<code>qnorm(0,85)=</code>	1,036
<code>pnorm(0,5)=</code>	0,6915	<code>pnorm(2)=</code>	0,9772	<code>qnorm(0,6)=</code>	0,253	<code>qnorm(0,9)=</code>	1,282
<code>pnorm(0,75)=</code>	0,7734	<code>pnorm(2,25)=</code>	0,9878	<code>qnorm(0,65)=</code>	0,385	<code>qnorm(0,95)=</code>	1,645
<code>pnorm(1)=</code>	0,8413	<code>pnorm(2,5)=</code>	0,9938	<code>qnorm(0,7)=</code>	0,524	<code>qnorm(0,97)=</code>	1,881
<code>pnorm(1,25)=</code>	0,8944	<code>pnorm(2,75)=</code>	0,9970	<code>qnorm(0,75)=</code>	0,674	<code>qnorm(0,99)=</code>	2,326
<code>pnorm(1,5)=</code>	0,9332	<code>pnorm(3)=</code>	0,9987	<code>qnorm(0,8)=</code>	0,842	<code>qnorm(0,995)=</code>	2,576

### Problema 3 Bloc C

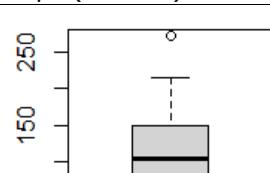
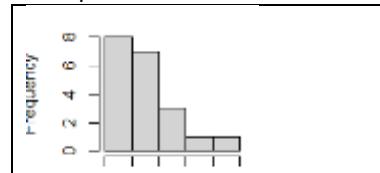
Un objectiu del Bloc transversal d'uns estudiants de PE ha estat "analitzar la diferència de rendiment del videojoc Minecraft executat en single-core (**S**) i multi-core (**M**) en diverses proves". Van recollir les variables **Velocitat V** (en FPS o Fotogrames Per Segon, indicant fluïdesa de refresh, i per tant com més gran millor), i **Temps T** (en MSPT o MiliSegons Per Tic, i per tant com més petit millor).

Per una part varen reproduir en S i M unes mateixes condicions: mateixes proves amb ordre a l'atzar amb codi reproduïble i correctament documentat. Varen recollir 20 temps en S ( $T_S$ ) i 20 en M ( $T_M$ ) amb els següents resultats de les diferències de temps i les diferències dels logaritmes dels temps (log és logaritme natural en R):

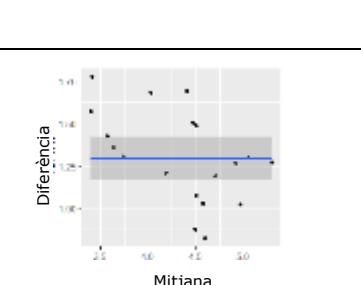
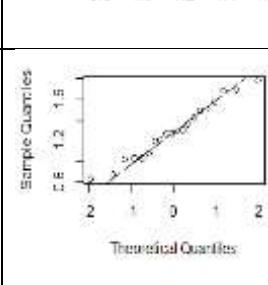
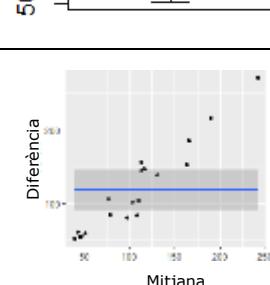
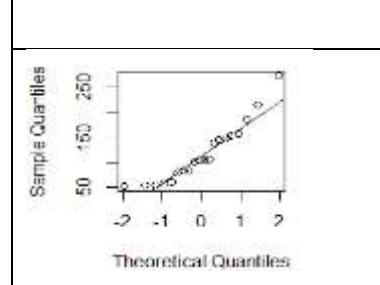
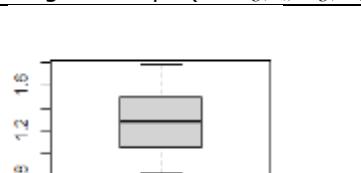
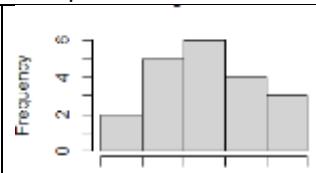
$$\sum(T_S - T_M) = 2361.07 \quad \sum(T_S - T_M)^2 = 344506.3$$

$$\sum(\log(T_S) - \log(T_M)) = 25.89 \quad \sum(\log(T_S) - \log(T_M))^2 = 34.94$$

Gràfiques amb la diferència de temps ( $D=T_S-T_M$ ):



Gràfiques amb la diferència dels log de temps ( $DL=\log(T_S)-\log(T_M)$ ):

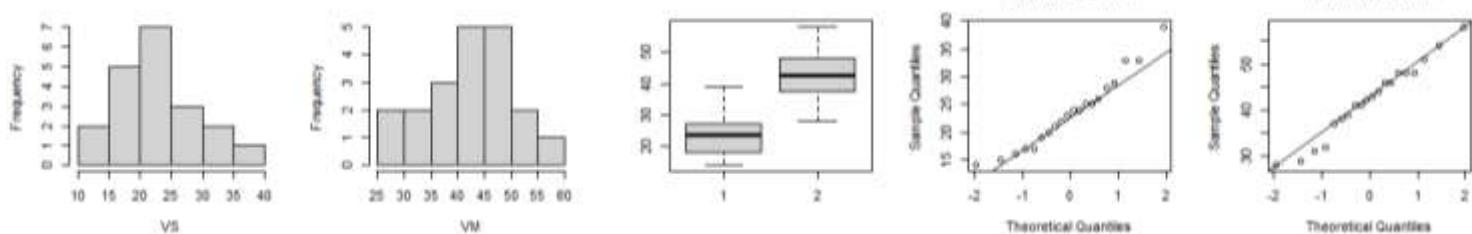


1.- (2 punts) Compareu les característiques de les 4 gràfiques pel cas de  $D$  i  $DL$

2.- (2 punts) Pel cas que compleix les premisses, calculeu un IC al 95% per a la diferència mitjana

3.- (1 punt) Interpreteu l'IC anterior. Expliqueu què implica i quins avantatges té fer la diferència de log

Prèviament als resultats anteriors havien recollit observacions en S en M en condicions semblants però sense un disseny assegurant unes mateixes proves amb ordre a l'atzar, i es van centrar en les velocitats. Varen recollir 20 velocitats en S (**VS**) i 20 en M (**VM**) amb els següents resultats per comparar mitjanes:



```
t.test(VS,VM,var.equal=T)
Two Sample t-test
data: VS and VM
t = -7.9733, df = 38, p-value = 1.235e-09
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
-23.44785 -13.95215
sample estimates: mean of x      mean of y
23.5          42.2
```

```
t.test(VS,VM,var.equal=F)
Welch Two Sample t-test
data: VS and VM
t = -7.9733, df = 36.414, p-value = 1.672e-09
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
-23.45465 -13.94535
sample estimates: mean of x      mean of y
23.5          42.2
```

4.- (1 punt) Comenteu els gràfics en relació a les premisses per comparar les mitjanes de velocitat

5.- (1 punt) Comenteu les diferències entre els dos *t.test* i justifiqueu quin dels dos és adequat

6.- (1 punt) Indiqueu i interpreteu un IC 95% per a la diferència de velocitats mitjanes

7.- (1 punt) Calculeu l'estimació de la desviació estàndard comú usada sota la premissa d'homoscedasticitat

8.- (1 punt) Compareu l'estudi dels temps i el de velocitats pels resultats i també pel disseny i característiques

Valors que poden ser útils pels blocs C i D:

qt(0.95,18)= 1.734	qt(0.95,19)= 1.729	qt(0.95,20)= 1.725	qt(0.95,21)= 1.721	qt(0.95,22)= 1.717	qt(0.95,38)= 1.686	qt(0.95,40)= 1.684
qt(0.975,18)= 2.101	qt(0.975,19)= 2.093	qt(0.975,20)= 2.086	qt(0.975,21)= 2.080	qt(0.975,22)= 2.074	qt(0.975,38)= 2.024	qt(0.975,40)= 2.021

NOM: \_\_\_\_\_

COGNOM: \_\_\_\_\_

**Problema 4 Bloc D.***Contesteu cada pregunta en el seu lloc. Explicieu i justifiqueu els càlculs. Si no s'indica, la pregunta 1 punt*

Repassem alguns treballs del bloc transversal. Aquesta sortida de R mostra un primer model:

```
lm(formula = photo_sizes ~ iso_values)
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.5956522  0.2564544 14.021   <2e-16
iso_values  0.0001316  0.0005563  0.237    0.814
Residual standard error: 1.155 on 58 degrees of freedom
```

**Model 1**

La variable `photo_sizes` és la mida en MB de fotos capturades amb telèfon mòbil, i `iso_values` el valor de la sensibilitat ISO seleccionada pel fotògraf (valors entre 100, 200, 400 i 800: quan es duplica la sensibilitat el sensor necessita la meitat de llum per capturar la mateixa imatge). La següent sortida utilitza les mateixes dades per un segon model:

```
lm(formula = photo_sizes ~ as.factor(iso_values))
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.4667    0.3023 11.466 2.55e-16
as.factor(iso_values)200 0.2867    0.4276  0.670   0.505
as.factor(iso_values)400 0.2333    0.4276  0.546   0.587
as.factor(iso_values)800 0.1933    0.4276  0.452   0.653
Residual standard error: 1.171 on 56 degrees of freedom
```

**Model 2**

1. Expliqueu breument quin és el primer model estadístic, quines les seves premisses, explicant cadascuna en aquesta situació (no analitzeu resultat).
2. Igualment, expliqueu breument quin és el segon model estadístic, i quines les seves premisses (no analitzeu resultat).
3. Compareu els dos valors d’”Estimate” que corresponen al paràmetre (Intercept): quines són les diferències entre ells?
4. Feu el mateix amb els valors corresponents a la variable explicativa de cada un dels models, i interpreteu-los. (1.5pts)
5. Un membre del equip està dubtós entre el primer model lineal, o un nou model lineal prenent com a variable explicativa el logaritme de la ISO. Doneu una recomanació justificada entre una opció o altra. (0.5pts)

Un altre treball investiga si el mode “estalvi de bateria” és eficaç per disminuir el consum del portàtil. Partint sempre del mateix nivell inicial de càrrega, s’assigna a l’atzar el mode (*Sí*, *No*) i el temps d’activitat amb el portàtil (15, 25, 35 minuts), quatre mesures per combinació. Es proven dos models:

```
lm(formula = y ~ z, data = D)
  Estimate Std. Error
(Intercept) 6.8333   0.9136
zn          0.8333   1.2920
Residual standard error: 3.165 on 22 d.f.
Multiple R-squared:  0.01856,
```

**model A**

```
lm(formula = y ~ x + z, data = D)
  Estimate Std. Error
(Intercept) -1.91667  0.74107
x            0.35000  0.02691
zn          0.83333  0.43946
Residual standard error: 1.076 on 21 d.f.
Multiple R-squared:  0.8916,
```

**model B**

La variable “y” registra el descens de càrrega de la bateria en el període d’activitat (en percentatge); “x” el temps d’activitat en minuts, i “z” el mode d’estalvi (“n” és *No*).

6. Una analisi aparellada seria possible? És a dir, prenen la diferència de descens de càrrega entre un valor en mode *Sí* i un altre valor en mode *No*, amb el mateix temps d’activitat. Raoneu breument la resposta.

7. Segons els resultats del model A, podem dir que val la pena emprar el mode d’estalvi de bateria? Doneu justificació de la resposta, incorporant-hi l’interval de confiança del 95% per l’efecte del mode. (1.5pts)

8. Repetiu la resposta però amb els resultats del model B. (1.5pts)

9. Interpreteu el valor 0.35000 de la variable “x” del model B. Quines unitats utilitza aquesta magnitud? (0.5pts)

10. L’indicador R-squared pren valors molt diferents en els dos models evaluats. Raoneu pel que veieu als resultats sobre els motius d’aquesta situació. (0.5pts)