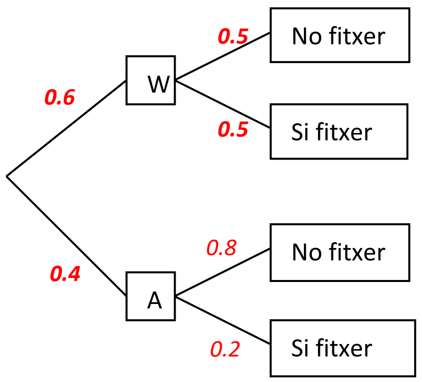
NOM:__________________________COGNOM:____

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

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



Ω	Probabilitats	
(W,No)	0.6*0.5 = <u>0.30</u>	
(W,Si)	0.6*0.5 = <u>0.30</u>	
(A,No)	**<u>0.32</u>** = 0.4 * P(No|A) -> P(No|A)=0.8	
(A,Si)	0.4*(1-0.8) = <u>0.08</u>	P(Si|A)=0.2

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

P(A,Si) = P(A)*P(Si|A) = 0.4 * 0.2 = **0.08** (un 8 %)

P(A|Si) = P((A,Si)) / P(Si) = 0.08 / (0.30+0.08) = 0.08 / 0.38 = **0.21** (aprox 21 %)

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

No hi ha independència entre fer la consulta via Web o via App i el fet de portar un fitxer adjuntat o no.
Per exemple P(Si|W) = 0.50 no és igual a P(Si) = 0.30+0.08=0.38, i no és igual a P(Si|A) = 0.2
La probabilitat que una consulta porti fitxer adjuntat és més gran si arriba via Web que si ho fa per App

El temps per resoldre les 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

T amb funció de densitat $f_T(x) = \ln(2)\ 2^{-x}$ i funció de distribució $F_T(x) = \int_0^x \ln(2) 2^{-t}\, dt = -2^{-t}|_0^x = 1 - 2^{-x}$

P(T>2)=1-P(T<=2) = $1-(1-2^{-2})$ = 2^{-2}= **0.25**

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

P(T<x) = 0.50 = $1-2^{-x}$ 0.50 = 2^{-x} log2(0.5)=-x log2(0.5)=-1 ---→ **x=1**

o bé ln(0.5)=ln(2^{-x}) -0.679 =-ln(2)*x

-0.679 = -0.679*x ---→ **x=1**

La mediana indica que al cap d'un dia es resolen el 50 % de les consultes

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	0,25	0,25	0	0.50
200	0	0,25	0,25	0.50
	0.25	0.50	0.25	

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

E(T) = 1*0.25+2*0.50+3*0.25 = **2**

V(T) = $(1-2)^2$*0.25 + $(2-2)^2$*0.50 + $(3-2)^2$*0.25 = **0.5**

E(C) = 100*0.5 + 200*0.5 = **150**

V(C) = $(100-150)^2$*0.5 + $(200-150)^2$*0.50 = **2500**

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

Cov(T,C) = (100-150)*(1-2)*0.25 + (100-150)*(2-2)*0.25 + (100-150)*(3-2)*0 +
(200-150)*(1-2)*0 + (200-150)*(2-2)*0.25 + (200-150)*(3-2)*0.25 =
= 12.5 + 0 + 0 + 0 + 0 + 12.5 = **25**

Cor(T,C) = Cov(T,C) / (sqrt(V(T))*sqrt(V(C))) = 25 / (sqrt(0.5)*sqrt(2500) = 25/(0.71*50) = 25/35.5 = **0.7**

La correlació és positiva i força alta. Indica una relació positiva entre el nombre de temes i el nombre de caràcters de les consultes: més temes implica més caràcters

NOM:________________________COGNOM:____

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

Problema 2 (B)

La biblioteca del barri està fent un estudi dels préstec 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)

Es considera la v.a. P:"nombre de pel·lícules prestades" P~Bin(n=80, p=0.06)

S'ha de calcular P(R=10) = $\binom{80}{10} \cdot 0.06^{10} \cdot 0.94^{70}$ = 0.0131

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

Es considera la v.a. R:"nombre de revistes prestades" R~Bin(n=80, p=0.15)

E(R) = 80· 0.15 = 12 revistes

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)

Sigui X la variable aleatòria, X:"nombre de documents prestat fins a prestar la primera revista" ~Geom(p=0.15)

S'ha de calcular P(X=10) = $0.15 \cdot 0.85^9$ = 0.0347

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)

Sigui la v.a. I: "Nombre de llibre infantils prestats en una hora".

P ~ P(5), P(P=3) = $\frac{5^3 \cdot e^{-5}}{3!}$ = 0.1404

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)

I4 ~ P(4×5), P(I4=25) = $\frac{20^{25} \cdot e^{-20}}{25!}$ = 0.0446

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)

T~exp(5/60)=exp(1/12), temps esperat: 12 min

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)

P(T > $t_{0.9}$)=0.9 . Calculem $e^{-t/12}$ = 0.9 ; -t/12 = ln(0.9) ; t= -12· ln(0.9) = 1.2643. Per tant, ens cal esperar (almenys) 1.26 minuts entre dos préstecs de llibres infantils amb un 90% de probabilitat. O bé: "Amb una probabilitat de 0.9 el segon llibre arribarà després de 1.26 minuts".

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)

P(almenys dues boles B) es pot calcular amb una binomial B~(5,2/10)

P(B≥2) = 1- P(B < 2) = 1 – P(B=0) – P(B=1) = 1 – $(8/10)^5$ – $C_{5,1}$·(2/10)·$(8/10)^4$ = 0.2627

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)

Considerem la VA T: “temps de préstec d’un llibre de la biblioteca” amb $T \sim N(\mu,\sigma)$.
A partir de la informació de l’enunciat, es té que $P(T<12) = 0.03$ i que $P(T>40) = 0.05$

Com que $P(T<12) = 0.03$; $P(Z < (12-\mu)/\sigma) = 0.03$; I per tant, $(12-\mu)/\sigma = -1.881$ [Com que qnorm(0,97) = 1.881, per la simetria de la normal estandarditzada se sap que qnorm(0.03) = –1.881]

De la mateixa manera, com que $P(T>40) = 0.05$, per tant, $P(T<40) = 0.95$ i estandarditzant:
$P(Z < (40-\mu)/\sigma) = 0.95$, es té que $(40-\mu)/\sigma = 1.645$ (qnorm(0.95))

Amb aquesta informació s’obté que:

$$\begin{cases} 12 = \mu - 1.881\sigma \\ 40 = \mu + 1.645\sigma \end{cases}$$

Del que es dedueix que els valors demanats són **μ =26.94** i **σ = 7.94**

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)
Sigui L :”temps, en dies, de préstec d’un llibre a la biblioteca estudiada”. $L \sim N(20,5)$
$P(L<15) = P((L-20)/5 < (15-20)/5) = P(Z < -1) = 0.1587$

c) Quina és la probabilitat que la mitjana dels temps de préstec de 100 llibres sigui menys de 19 dies ? (1 punt)
$L \sim N(20, 5)$ i pel TCL es té que $M_{100} \sim N(20, 0.5)$
$P(M_{100} < 19) = P(Z < (19-20)/0.5) = P(Z < -2) = 0.0228$ amb $Z \sim N(0,1)$

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)
Pel TCL es té que $S_{400} \sim N(480000, 4000)$ (en grams) = N(480,4) (en kg)
$P(S_{400} > 475) = P((S_{400} - 480)/4 > (475-480)/4) = P(Z > -1.25) = 1 - P(Z < -1.25) = 1 - 0.1056498 = 0.8944$

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

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

NOM:__________________________COGNOM:____

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

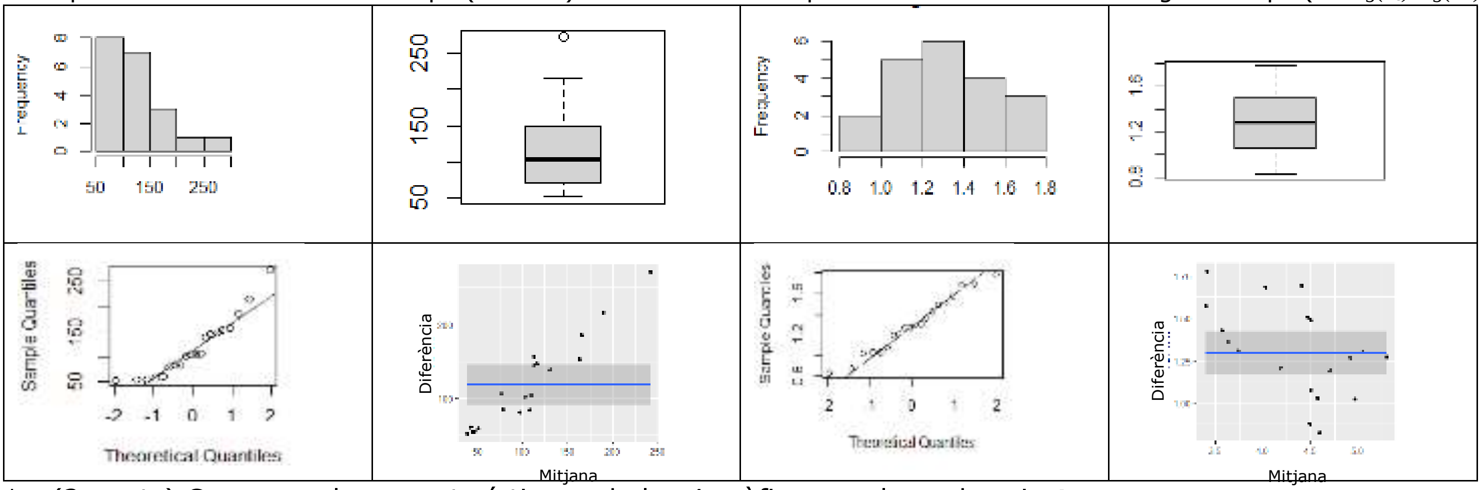
Problema 3 Bloc C

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 **V**elocitat **V** (en FPS o Fotogrames Per Segon, indicant fluïdesa de refresc, i per tant com més gran millor), i **T**emps **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):

$\sum(T_S - T_M) = 2361.07$ $\quad\sum(T_S - T_M)^2 = 344506.3$ $\quad\sum(\log(T_S) - \log(T_M)) = 25.89$ $\quad\sum(\log(T_S) - \log(T_M))^2 = 34.94$

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

Histograma: D no sembla Normal sinó exponencial decreixent. Dl sí sembla campana de Gauss

Boxplot: D no és molt simètric i té un outlier. Dl sí és simètric

Normal QQ Plot: D no s'acaba d'alinear. Dl sí queda més alineat seguint el model Normal

Blandt Altman: D mostra que la diferència entre S i M creix conforme la mitjana també creix, indicant un efecte multiplicatiu. En canvi a Dl no sembla seguir cap patró

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

Dl=$log(T_S)$-$log(T_M)$ diferència de logaritmes, μ_{Dl} és la diferència mitjana en log de temps en S menys M

$\overline{dl}$ = 25.89/20 = 1.295 (mitjana de la diferència de logaritmes)

s_{Dl} = sqrt((34.94-((25.89*25.89/20))) / (19)) = 0.27 (desviació de la diferència de logaritmes)

se = s_{Dl} / sqrt(20) = 0.06 (estàndard error)

IC(μ_{Dl},95%): $\overline{dl} \pm t_{19,0.975}\ se$=$1.295 \pm qt(0.975,19) * 0.06$ =$1.295 \pm 2.093 * 0.06$ =1.295 ± 0.126 =**[1.17, 1.42]**

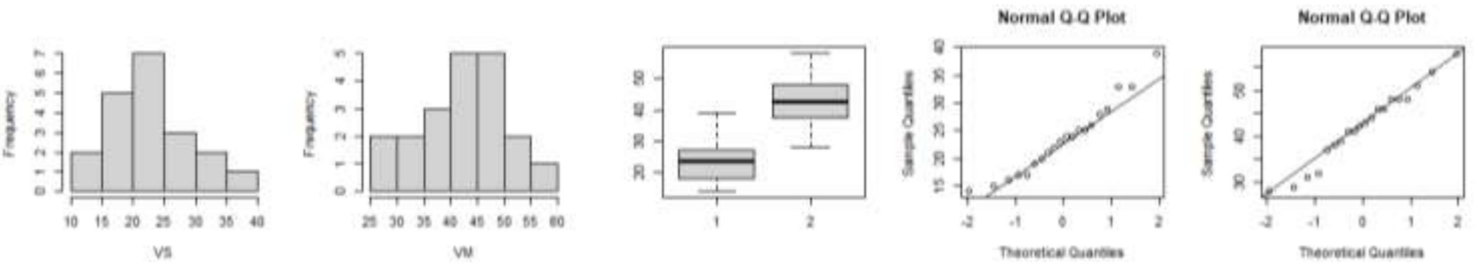
Com Dl=$log(T_S)$-$log(T_M)$=$log(T_S/T_M)$ i μ_{Dl} és $\mu_{log(T_S/T_M)}$ llavors **IC(μ_{T_S/T_M},95%)** és $[e^{1.17}, e^{1.42}]$ = **[3.22, 4.14]**

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

Amb confiança 95% la mitjana de temps en log és entre 1.17 i 1.42 superior en S que en M (S més lent)
Amb confiança 95% la mitjana del rati de temps entre S i M és entre 3.22 i 4.14 superior en S (S pot arribar a multiplicar per 4 la mitjana de temps en MSPT respecte M, S és 4 vegades més lent)

Fer la diferència de logaritmes pot arreglar problemes de normalitat (fent diferència de log i no log de la diferència ja que aquesta diferència podria tenir valors negatius). A més a més, la comparació es pot veure en termes de quocient (en mitjana quantes vegades és més gran o més petit) enlloc de diferència en mitjana que utilitza les unitats de les variables (MSPT)

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

Per comparar les mitjanes de velocitat, com són mostres independents, cal verificar gràficament la normalitat de les dues mostres i la homoscedasticitat entre elles.

Els dos histogrames i els dos Normal QQ Plot mostren que les dues mostres independents compleixen normalitat

Els box-plots mostren que les dues mostres independents estan centrades en valors de mediana diferents però tenen una variabilitat semblant

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

Els dos t.test difereixen en que en un s'assumeix igualtat de variàncies entre les mostres independents (en el de l'esquerra indicant var.equal=T) i l'altre no (en el de la dreta indicant var.equal=F)

Segons l'anàlisi gràfica dels boxplot podem assumir igualtat de variàncies i, per tant, el t.test de l'esquerra seria adequat

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

IC($\mu_{V_S} - \mu_{V_M}$,95%) és **[-23.45, -13.95]**

Amb confiança 95% la diferència de mitjanes de velocitat entre el cas S i M és entre -23.45 i -13.95 inferior en S que en M

La velocitat mitjana en S es pot esperar entre -23.45 i -13.95 FPS inferior a la velocitat mitjana en M

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

Com que l'IC és $(23.5 - 42.2) \pm t_{38,0.975} \cdot \mathbf{s}\sqrt{\frac{1}{20} + \frac{1}{20}}$ i ha donat [-23.44785, -13.95215]

Llavors (-23.45-(-13.95))/2 = qt(0.975,38) * s * sqrt(1/20+1/20)
4.75 = 2.024 * s * 0.316 → **s** = (4.75/(2.024*0.316)) = **7.4**

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

En quant a resultats, en els dos casos els IC quantifiquen la diferència de rendiment entre S i M. Es pot esperar temps en MSPT inferiors i velocitats en FPS superiors en el cas M (multi-core) respecte el S (single-core)

En quant a disseny i característiques, pel temps s'ha usat mostres aparellades i per velocitats mostres independents. El cas aparellat és més acurat ja que el control de les condicions augmenta l'eficiència per trobar diferències. Pel cas del temps ha calgut aplicar logaritmes per complir les premisses

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

Problema 4 Bloc D.

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

```
lm(formula = photo_sizes ~ iso_values)                             Model 1
             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
```

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))                   Model 2
                         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
```

1. Expliqueu breument quin és el primer model estadístic, quines les seves premisses, explicant cadascuna en aquesta situació (no analitzeu resultat).

   És un model lineal, on la resposta és la mida de la fotografia i la variable explicativa és el valor ISO. Les premisses són: homocedasticitat (variància de la mida no depèn de la ISO), normalitat (la mida condicionada per la ISO emprada es distribueix Normalment), linealitat (el valor esperat de la mida depèn linealment de la ISO emprada) i independència (les observacions recollides són independents unes d'altres).

2. Igualment, expliqueu breument quin és el segon model estadístic, i quines los seves premisses (no analitzeu resultat).

   Les variables són les mateixes, amb el mateix rol. En aquest cas no és un model lineal exactament perquè la variable explicativa no és quantitativa i no pot variar més que dins dels quatre valors indicats. Per tant tenim una categoria de referència (ISO 100) i s'estimaran els canvis en la mida per utilitzar una categoria distinta (200, 400 o 800). Les premisses són les mateixes excepte linealitat.

3. Compareu els dos valors d'"Estimate" que corresponen al paràmetre (Intercept): quines són les diferències entre ells?

   En el primer cas, Intercept significa el valor esperat per la mida de la foto amb ISO igual a 0, o un valor de referència que després s'afegirà a l'increment associat al ISO real utilitzat. En el nostre cas, la mida base és de 3.596MB amb error estàndard 0.256. En el segon cas, es tracta del valor esperat per la mida de la foto amb ISO igual a 100, igual a 3.467MB amb error estàndard 0.302.

4. Feu el mateix amb els valors corresponents a la variable explicativa de cada un dels models, i interpreteu-los. (1.5pts)

   Al primer model, l'estimació 0.0001316 ens diu que amb un increment d'1 en la sensibilitat ISO la mida augmentarà en $1.3\cdot10^{-4}$ MB però amb un error d'estimació de $5.6\cdot10^{-4}$, llavors no podem dir que un increment de mida estigui associat a canvis en la ISO.

   Al segon model tenim tres canvis:

   - passar de 100 a 200 incrementa la mida en 0.2867MB
   - passar de 100 a 400 incrementa la mida en 0.2333MB
   - passar de 100 a 800 incrementa la mida en 0.1933MB

   En tot cas, l'error d'estimació (0.4276) és gran en els tres casos, llavors tampoc podem afirmar res en concret: les variacions observades en la mida de la fotografia podrien ser resultat de l'atzar.

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)

   Utilitzar el logaritme de la ISO tindria més sentit, perquè és versemblant esperar que l'efecte en la mida sigui el mateix si canviem de 100 a 200, o de 800 a 1600 (dos increments iguals en termes logarítmics: log(100)=4.61, log(200)=5.30; log(800)=6.68, log(1600)=7.38). Però no podem esperar que l'efecte es mantingui quan es canvia de 100 a 200, o quan passa de 800 a 900.

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)              model A
             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,
```

```
lm(formula = y ~ x + z, data = D)                    model B
            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,
```

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 anàlisi aparellada seria possible? És a dir, prenent 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.
   No es pot analitzar unes dades que no han estat dissenyades per aquell propòsit. En aquest disseny, les dades són de dos grups independents (*Sí*, *No*) perquè la intervenció s’assigna a l’atzar. Cada observació no té una parella en correspondència en l’altre grup.

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)
   L’IC de l’efecte val $0.8333 \pm 2.074 \; 1.2920 = (-1.846, \; 3.513)$. El guany del mode d’estalvi és incert: no emprar-ho en lloc de sí emprar-ho representa un increment del consum entre -1.846% (menys consum que amb el mode d’estalvi), i 3.513% (consumeix més bateria). No tenim evidències per concloure.

8. Repetiu la resposta però amb els resultats del model B. (1.5pts)
   L’IC de l’efecte val $0.8333 \pm 2.080 \; 0.4395 = (-0.0806, \; 1.7472)$. Molta menys incertesa per un IC més estret, però no ens serveix per concloure amb determinació, ja que encara es podria admetre un efecte nul del mode d’estalvi. També l’efecte positiu del mode seria més modest, en cas de ser eficaç.

9. Interpreteu el valor 0.35000 de la variable “x” del model B. Quines unitats utilitza aquesta magnitud? (0.5pts)
   Les unitats són *%càrrega decrementat per minut* de treball. Cada minut treballant amb el portàtil representa un 0.35% menys de bateria, amb error estàndard 0.027%/min.

10. L’indicador R-squared pren valors molt diferents en els dos models avaluats. Raoneu pel que veieu als resultats sobre els motius d’aquesta situació. (0.5pts)
   La diferència entre els dos models és que el B inclou <u>una variable explicativa més</u>, el temps d’activitat. Aquesta variable té un coeficient estimat de 0.35, amb <u>un error estàndard molt petit</u> (0.027), llavors veiem que <u>influeix molt a les variacions de la resposta “y”</u>, al contrari que la variable “z”, que és escassament significativa. En realitat, quasi tota la <u>variabilitat</u> del descens de càrrega <u>es podria atribuir al temps d’activitat</u>, i molt poc al mode d’estalvi. Al primer model veiem que “z” explica menys d’un 2%, i en canvi “x” més “z” arriba al 89%.