

## Problema 1 (B4)

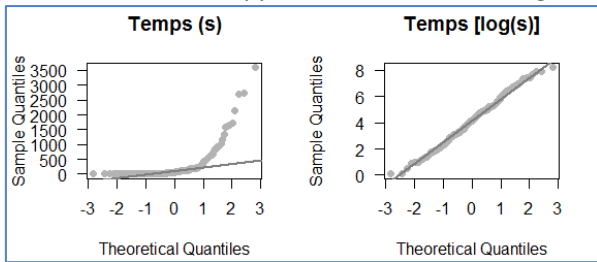
Un grup d'estudiants de la FIB ha creat una empresa de resolució de problemes informàtics on-line. Mitjançant un complex algoritme de resolució dels problemes més comuns, l'usuari pot arribar a l'arrel del seu problema sense la necessitat de que intervingui un tècnic. El primer pas ha estat crear una plana web que ja porta activa uns mesos. Amb l'ajuda de l'eina de Google Analytics (GA) obtenim estadístiques del rendiment de la pàgina.

La nostra primera preocupació és la durada de les visites. Excloent els usuaris que reboten (aquells que no interactuen amb la pàgina), el temps mitjà de les visites segons GA és de 120 segons. Com que no estan segurs de la fiabilitat d'aquesta dada, han creat un script propi que mesura precisament aquest indicador. Per les 30 primeres visites després d'iniciar el script, han obtingut els següents resultats pels temps i pel logaritme natural dels temps:

Temps						
<b>&gt; summary(temps)</b>						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
2.386	51.260	221.100	487.400	408.800	3775.000	
<b>&gt; sd(temps) [1] 793.1788</b>						
<b>&gt; summary(l(temps))</b>						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.8694	3.9370	5.3940	4.9890	6.0130	8.2360	
<b>&gt; sd(l(temps) [1] 1.764887</b>						

Volen testear a través d'una prova d'hipòtesis si el valor de 120 segons que dona GA pot ser la mitjana de temps poblacional o no emprant aquestes dades.

- Observant els dos *qqnorms* de les dades, argumenteu si s'han d'emprar les dades transformades o sense transformar (1 punt)



- Plantegeu les hipòtesis de la prova (a les dades transformades o no segons la resposta anterior) tot mencionant si ha de ser unilateral o bilateral (1 punt)

- Indiqueu l'expressió de l'estadístic per resoldre el test d'hipòtesis i la seva distribució sota la hipòtesi nul·la (1 punt)

- Calculeu el valor de l'estadístic i el/s punt/s crític/s suposant un  $\alpha = 0.05$  (1 punt)

- Representeu gràficament la distribució de l'estadístic, el seu valor, el/s punt/s crític/s, i les zones d'acceptació i rebuig (1 punt)

- Indiqueu, justifiqueu i interpreteu quina és la conclusió de la prova d'hipòtesis? (1 punt)

- Calculeu un Interval de Confiança del 95% (IC95%) per a la mitjana poblacional (1 punt)

- Interpreteu el IC anterior i relacioneu-lo amb la conclusió de la prova d'hipòtesis (1 punt)

Un altre preocupació que tenen són els visitants que romanen menys d'un *temps mínim* (que tenen estipulat en 15 segons) a la pàgina perquè consideren que són usuaris als que alguna cosa no els hi ha agradat. En les dades obtingudes hi ha un 20% de persones que romanen menys d'aquest *temps mínim*. Calculeu un IC 95% per a aquesta proporció de persones que romanen menys del *temps mínim* a la pàgina i interpreteu-lo. (1 punt)

I calculeu quan hauria de ser la  $n$  de la mostra per obtenir l'anterior IC amb la meitat d'amplada (1 punt)

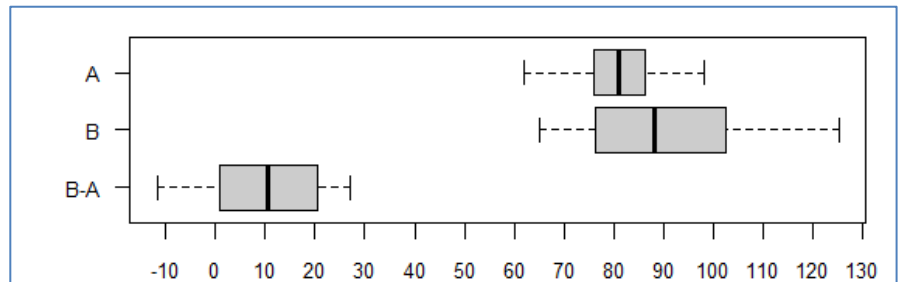
NOM: \_\_\_\_\_ COGNOM: \_\_\_\_\_

(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs)

## Problema 2 (B5)

A Google Analytics (GA), hi ha una eina per assignar a certes pàgines un determinat import econòmic per visita (per exemple, a una pàgina de petició de servei de 20 euros que el contracten 1 de cada 20 persones, es pot assignar el valor de 1 euro). Els estudiants han dissenyat un experiment en que durant 2 mesos (61 dies) recullen els guanys estimats segons dos prototipus de pàgina (A i B). La descriptiva es mostra a continuació:

	A	B	B-A
<b>Mitjana</b>	80.77	89.80	9.04
<b>Desv. Tipus</b>	8.34	16.52	12.09
<b>Mín</b>	61.82	64.93	-11.62
<b>Q1</b>	76.38	76.87	1.37
<b>Mediana</b>	80.95	88.26	10.52
<b>Q3</b>	86.12	100.60	20.00
<b>Màx</b>	98.17	125.10	26.97



- Justifiqueu si es tracta de dades aparellades o independents (1 punt)

- Plantegeu una prova d'hipòtesis per decidir si la mitjana de guanys diaris és la mateixa en ambdós dissenys o si en el cas B són majors. Indiqueu les hipòtesis i si ha de ser unilateral o bilateral (1 punt)

- Indiqueu l'expressió de l'estadístic, la seva distribució sota  $H_0$  i les premisses assumides (1 punt)

- Calculeu el valor de l'estadístic i el/s punt/s crític/s suposant un  $\alpha = 0.05$  (1 punt)

- Indiqueu, justifiqueu i interpreteu quina és la conclusió de la prova d'hipòtesis? (1 punt)

Ara suposem que les dades anteriors no s'han recollit els mateixos dies pel cas A i B sinó en diferents mesos, i plantegem la prova d'hipòtesis per decidir si la mitjana de guanys diaris és la mateixa en ambdós prototipus o no.

- Indiqueu les hipòtesis i si ha de ser unilateral o bilateral (1 punt)

- Indiqueu l'expressió de l'estadístic, la seva distribució sota  $H_0$  i les premisses assumides. Indiqueu si la descriptiva de les dades recolza assumir les premisses (1 punt)

- Calculeu l'error tipus estimat per a la diferència de mitjanes mostrals i el valor de l'estadístic (1 punt)

- Calculeu i interpreteu l'interval de confiança del 95% per a la diferència d'esperances (1 punt)

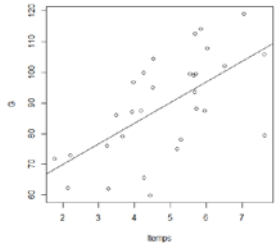
- Indiqueu, justifiqueu i interpreteu quina és la conclusió de la prova d'hipòtesis? (1 punt)

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

(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs)

### Problema 3 (B6)

Els nous emprenedors volen estudiar si existeix alguna mena de relació lineal entre la durada de les visites i el guany monetari obtingut en visites en que es fa alguna compra. Ho estudien a partir del logaritme del temps (ltemps) i els guanys (G) de 30 visites. Els resultats de la descriptiva, de la covariància entre ltemps i G, i de la regressió lineal corresponents són:



	Mitjana	Desviació tipus	Min	Max
G	88.90	16.37	58.2	119.33
ltemps	4.83	1.51	1.5	7.8

$$\text{cov}(G, \text{ltemps}) = S_{G, \text{ltemps}} = 15.34$$

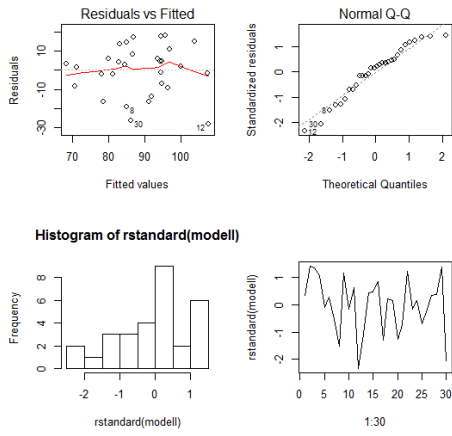
```
summary(lm(G ~ ltemps)):  Coefficients:      Estimate   Std. Error  t value    Pr(>|t|)
      (Intercept)      -----      8.110      6.957     1.45e-07 ***
           ltemps      -----      1.606      4.191     0.000251 ***
Residual standard error: 13.06 on 28 degrees of freedom
```

- Calculeu els coeficients de la recta de regressió dels guanys G en funció del logaritme del temps. I indiqueu el coeficient de determinació i la desviació residual (2 punts)

- Expliqueu què ens indiquen el pendent de la recta, el coeficient de determinació, la correlació i la desviació residual (2 punts)

- Calculeu un IC 95% del pendent de la recta i resolcu la prova d'hipòtesis de si la recta és plana o no. (2 punts)

- Enuncieu les premisses o hipòtesis de la regressió lineal i comenteu si es compleixen o no per aquest cas concret. Especifiqueu de quins resultats i/o gràfics es dedueixen els vostres comentaris. (2 punts)



- Calculeu una predicció del guany esperat per visites que romanen 120 segons amb un interval de confiança al 95% (1 punt)

- Feu una valoració global d'aquest model de regressió en quant a la validació, el coeficient de determinació i la desviació residual (1 punt)