

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

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

**Problema 1 (B4).** Per comparar M i N, els dos proveïdors de pel·lícules més populars, hem calculat la diferència D entre els seus temps de baixada de 120 films escollits a l'atzar de la llista de la CSMS. Encara que estava planificat inicialment comparar els temps en tots els casos conjuntament, els resultats mostren 2 grups clarament diferenciats: 100 casos amb valors al voltant de la igualtat (Empat E:  $n_E=100$ , mitjana $_E=0$ ,  $sd_E=4$ ); i 20 casos amb clara avantatge per N (G=Guanya N,  $n_G=20$ , mitjana $_G=40$ ,  $sd_G=12$ ). Dins de cada grup,  $D \sim N$ . (Totes valen igual)

<b>Dades auxiliars:</b>	<pre>&gt; qt(0.975,99) [1] 1.984217 &gt; qnorm(0.995) [1] 2.575829 &gt; qnorm (0.975) [1] 1.959964</pre>	<pre>&gt; qt(0.95,19) [1] 1.729133 &gt; pt(8,19) [1] 0.9999999</pre>	<pre>&gt; qchisq(0.05,20) [1] 10.85081 &gt; qchisq(0.95,20) [1] 31.41043 &gt; qchisq(0.025,20) [1] 9.590777</pre>	<pre>&gt; qchisq(0.975,20) [1] 34.16961 &gt; qchisq(0.05,19) [1] 10.11701 &gt; qchisq(0.95,19) [1] 30.14353</pre>	<pre>&gt; qchisq(0.025,19) [1] 8.906516 &gt; qchisq(0.975,19) [1] 32.85233</pre>
<b>Usage</b>	<pre>pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE) qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)</pre>				

1) Estimeu per interval (99%) i interpreteu la proporció de casos en que guanya N.

Estimeu per interval (95%) el valor de la esperança de D en els casos 'E' d'empat.

2) Estadístic, distribució, premisses.

3) Càlcul, interpretació.

4) Estimeu per interval (90%) el valor poblacional de la sd en els casos 'G' on guanya N. Estadístic, distribució, premisses.

5) Podeu demostrar (amb  $\alpha=0.05$  unilateral) que és cert que N guanya en els casos G?

6) Creieu que el model Normal podria ser una bona aproximació per la distribució conjunta dels 120 casos? (Per ajudar-te, dibuixa el gràfic amb la funció de densitat en l'eix d'ordenades i el valor de D en el d'abscisses):

7) Quant valen els estadístics mitjana i variància del 120 casos junts?

8) Posa a prova la hipòtesi d'igualtat (amb  $\alpha=0.05$  bilateral) en els 120 casos conjuntament. [Si no has contestat la pregunta anterior, utilitza mitjana=5 i  $sd=15$ ]. Estadístic, premisses, càlcul i interpretació.

9) Faci una interpretació global

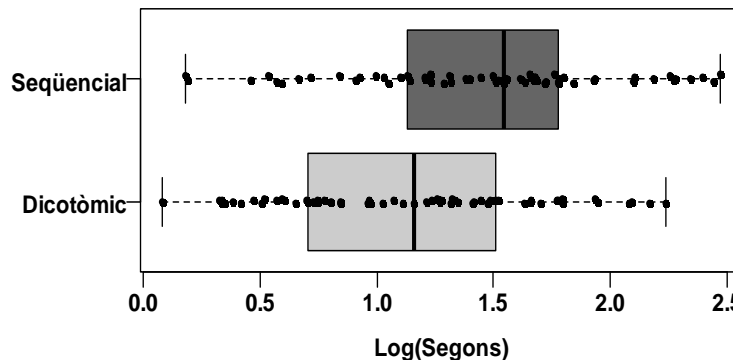
10) Vostè treballa per N i vol fer públics els resultats. Però el revisor de una revista informàtica ha dit que els subgrups E i G han estat suggerits pels resultats i per tant algunes hipòtesis no eren prèvies a les dades. A més a més, la seva filiació deixa clar cert conflicte de interessos. Segons el revisor aquests arguments fan que les troballes no tinguin cap valor. Redacti una resposta al revisor.

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

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

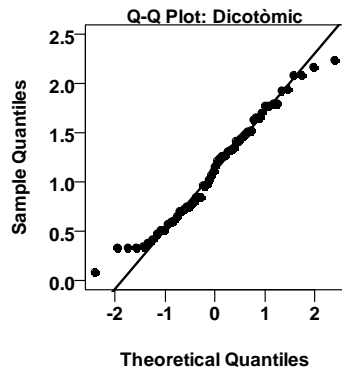
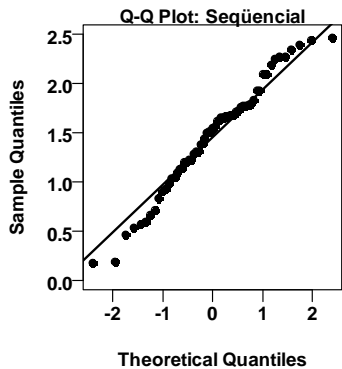
**Problema 2 (B5).** Es vol comparar la velocitat de dos algoritmes de cerca, *Seqüencial* i *Dicotòmic*, en vectors prèviament endreçats. Es generen 122 vectors endreçats i s'envien per fer una cerca a un dels dos algoritmes de forma aleatòria. Com que el temps de la cerca té una distribució asimètrica amb cua cap a la dreta, treballem amb els logaritmes dels temps. Sigui  $X$  el logaritme del temps que triga el algorisme *Seqüencial* i  $Y$  el del *Dicotòmic*. Els estadístics descriptius per a cada mostra estan a la següent taula. A més a més, es mostra el boxplot dels logaritmes dels temps:

	N	Mitjana	Desv. est.	Mediana	Mínim	Màxim
<i>Seqüencial</i>	61	1.47	0.56	1.55	0.18	2.47
<i>Dicotòmic</i>	61	1.14	0.54	1.16	0.08	2.24



**1) (1 punt).** Es tracta de dues mostres independents o aparellades? Raoneu la resposta.

**2) (1 punt).** Veient els següents Q-Q plots, sembla raonable suposar que les variables  $X$  i  $Y$  segueixen una distribució normal? Raoneu la resposta.



**3) (2 punts).** Estudiem primer la variabilitat dels (logaritmes dels) temps d'ambdós algoritmes. Es pot suposar que són iguals? Responen aquesta pregunta plantejant i resolent la hipòtesi adient (amb  $\alpha = 0.1$ ) i explicant quines són les premisses que cal fer per realitzar aquesta prova.

4) (3 punts). Per saber si es pot suposar que hi ha diferències entre ambdós algoritmes, es vol plantejar una prova d'hipòtesi d'igualtat de mitjanes:  $H_0: \mu_1 = \mu_2$  vs.  $H_1: \mu_1 \neq \mu_2$  (nivell de significació del 5%)

a) Calcula la variància conjunta (pooled variance) i el valor de l'estadístic (Desenvolupa els càlculs).

b) Digues quin és el punt crític amb un 5% de significació i treu conclusions sobre la prova d'hipòtesi.

c) Calcula l'interval de confiança del 90% per a la diferència de mitjanes dels logaritmes i interpreta'l.

5) (3 punts). A l'hora d'implementar els algoritmes s'han programat de manera que si l'execució triga més de 5 segons s'avorta l'execució i dona un missatge d'error. Volem comparar si la probabilitat d'error en ambdós algoritmes és pot considerar la mateixa. La taula de contingència de la dreta resumeix els resultats obtinguts:

	Temps < 5 s.	Temps > 5 s.
Seqüencial	32	29
Dicotòmic	47	14

a) Planteja la prova d'hipòtesi i especifica si és unilateral o bilateral. Quina és l'expressió de l'estadístic? Quina és la distribució de l'estadístic i sota quines premisses?

b) Quant val l'estadístic? (Desenvolupa els càlculs)

c) Podem rebutjar la hipòtesi nul·la amb un grau de significació del 5%?

La copisteria CUT&PASTE amb la intenció de verificar la eficiència de les seves fotocopiadores, analitza els temps d'impressió (en segons) en funció de la grandària del fitxer (en kb) i això ho fa per 100 fitxers diferents. La representació gràfica d'aquestes dades és a la figura 1

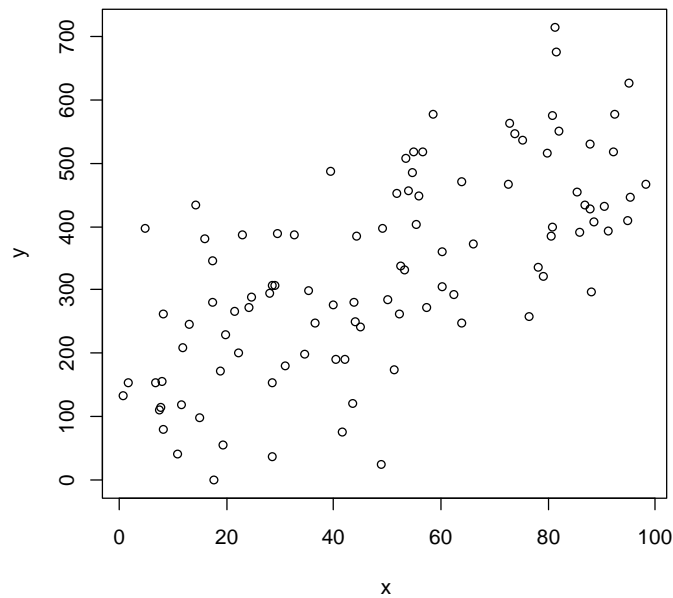


Figura 1

**a) (3 punts)**

Sabent que  $X$ =grandària del fitxer (en kb) i  $Y$ =temps d'impressió (en segons), en R s'han calculat els següents estadístics:

$$\text{Mean}(X) = 48.76 \text{ kb} \quad \text{Var}(X) = 796.76 \text{ kb}^2$$

$$\text{Mean}(Y) = 330.44 \text{ seg} \quad \text{Var}(Y) = 24769.21 \text{ seg}^2$$

i la correlació entre  $(X,Y)$ ,  $r=0.68$

Estimeu els coeficients de la recta de regressió  $b_0$  i  $b_1$  i dibuixeu la recta de regressió damunt de la figura 1

**b) (2 punts)**

Quant val la desviació tipus residual? Quina interpretació te en el model de regressió? Raoneu la vostra resposta.

**c) (2 punts)**

Poseu a prova mitjançant la prova d'hipòtesi adient si es pot admetre que aquesta recta passa pel punt  $(0,0)$ , és a dir, si el paràmetre  $\beta_0$  és significatiu. (Assumiu un risc  $\alpha=0.05$ )

**d) (1 punt)**

A la figura 2, teniu els gràfics necessaris per a poder validar la qualitat d'aquest model. Interpreteu quina premissa tracta de verificar cadascun d'aquests gràfics i si es verifica o no.

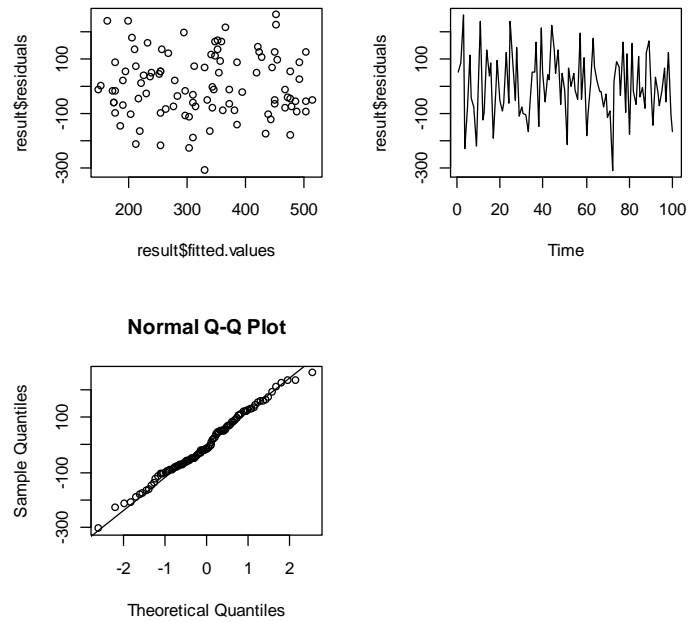


Figura 2. Anàlisi de les premisses

**e) (2 punts)**

Doneu una previsió puntual amb el seu interval de confiança al 95% de quants segons trigarà en imprimir-se un fitxer de 60 kb.