

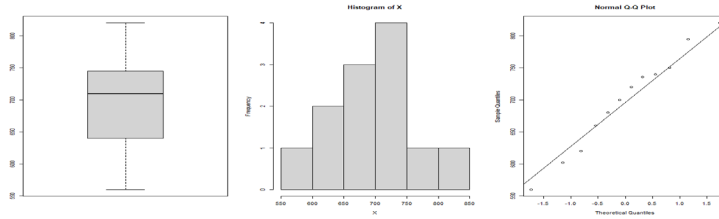
NOM: \_\_\_\_\_ COGNOM: \_\_\_\_\_

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

## Problema 1 (B4)

Una Universitat ha iniciat un pla de recollida de dades monitoritzant la concentració de CO<sub>2</sub> a les aules, en ppm (parts per milió). Obtenim, a l'atzar, 12 mesures de CO<sub>2</sub> en aules amb característiques, horari i ocupacions equivalents:

$X \leftarrow c(560,750,660,740,620,720,680,700,820,736,602,795)$       $\sum X_i = 8383$       $\sum X_i^2 = 5923025$



Les mesures de CO<sub>2</sub> presenten una fluctuació natural, amb desviació tipus de 70 ppm- Ens diuen també que 750 ppm és un valor de referència com a llindar superior per considerar òptima la qualitat de l'aire.

1.- Calculeu l'estimació puntual de la mitjana i la desviació de la concentració de CO<sub>2</sub> (1 punt)

2.- Amb els valors anteriors calculats, i els anteriors gràfics descriptius, comenteu la informació que donen sobre la qualitat de l'aire i sobre si es compleixen les premisses per calcular intervals de confiança (1 punt)

3.- Assumint el valor de la fluctuació natural de les mesures de CO<sub>2</sub> com a desviació poblacional, calculeu un interval de confiança al 95% per a la concentració mitjana de CO<sub>2</sub> (1 punt)

4.- I calculeu l'interval anterior si no assumim el valor anterior com a poblacional (1 punt)

5.- Interpreteu i compareu els dos intervals anteriors (2 punt)

Ara ens centrarem en unes dades d'una inspecció un dia i hora concrets en la que es prenen les mesures a 30 aules, i es tenen els dos resultats (A i B) següents:

A: t = -2.04, df = 29 alternative hypothesis: true mean is not equal to 750 95 percent confidence interval: 677.1729 750.0937 sample estimates: mean of x 713.6333	B: t = -2.04, df = 29 alternative hypothesis: true mean is less than 750 95 percent confidence interval: -Inf 743.9237 sample estimates: mean of x 713.6333
---	--

Una de les dues proves aporta evidència que la mitjana de ppm de les aules és inferior al llindar de 750 amb una confiança del 95%. Indiqueu quina és la prova i indiqueu hipòtesis, conclusió de la prova i interpretació de l'interval de confiança (2 punts)

Seguint amb aquestes dades de la inspecció, s'obté que de les 30 aules en 23 no es supera el llindar de 750 ppm. Indiqueu un interval de confiança al 95% pel percentatge d'aules que no superen el llindar (2 punts)

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

## Problema 2 (B5)

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

Volem comparar el temps d'execució de dos algorismes per ordenar vectors amb la mateixa complexitat de temps  $O(n^2)$ : Bubble sort, i Insertion sort. Hem generat a l'atzar 50 vectors amb una mida entre  $10^1$  i  $10^5$  i hem calculat el temps que triga cada algoritme. La taula següent proporciona la mitjana i la dispersió (desviació típica o estàndard) per a cada algoritme i per a la seva diferència. Per la resposta 'temps', esquerra; i pel seu logaritme natural, dreta. **Cada pregunta 1 punt.**

Temps en segons			Log(temps)		
Var	Mitjana	Dispersió	Var	Mitjana	Dispersió
<b>B</b>	193'4	175'7	<b>ln(B)</b>	4'2	2'4
<b>I</b>	91'6	83'1	<b>ln(I)</b>	3'4	2'4
<b>B-I</b>	101'8	98'4	<b>ln(B)-ln(I)</b>	0'74	0'03

1.- Indiqueu i justifiqueu si es tracta d'un disseny de dades aparellades o independents

2.- Comenteu què implica cada disseny (independent o aparellat) en quant a la variància de la diferència. Dona això alguna pista sobre el grau d'aparellament (dependència) de les dades?

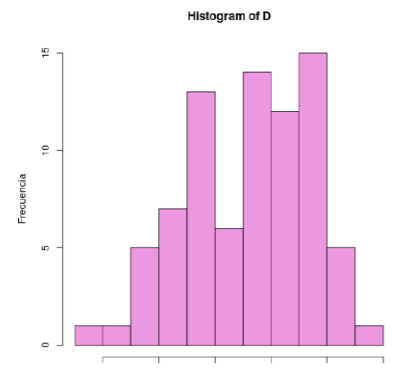
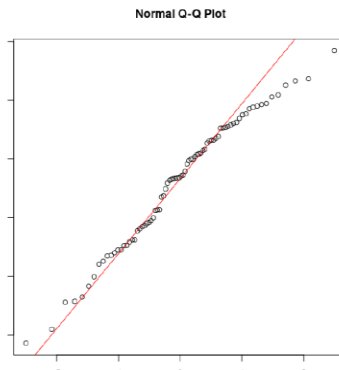
Si es tracten com a mostres independents (assumint normalitat i igualtat de variàncies poblacionals), calculeu:

3.- la desviació pooled i l'error estàndard de la diferència de mitjanes

4.- un interval de confiança al 95% de la diferència de mitjanes (podeu utilitzar la convergència a la Normal per 'n' grans)

5.- Opineu sobre les premisses

6.- Es considera ara la diferència dels logaritmes  $D = \ln(B) - \ln(I)$ , obtenint aquests dos gràfics. Interpreteu i indiqueu de què ens informen aquests dos gràfics

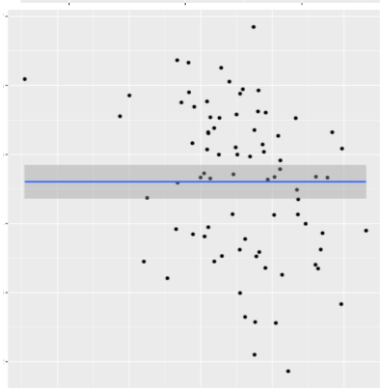
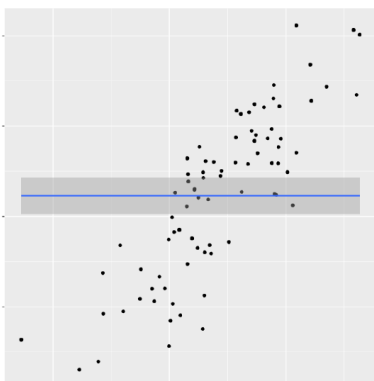


7.- Interpreteu els resultats numèrics descriptius (mitjana i desviació) de la diferència dels logaritmes. Quin triga menys? Quin és més ràpid? Quant més ràpid?

8.- Sigui T una v.a. amb distribució t de Student amb 49 graus de llibertat i  $P(T < 0.68) = 0.75$ ; doneu un interval simètric de confiança per a la diferència (0.5punts) i indiqueu amb quina confiança s'haurà calculat (0.5punts).

9.- Interpreteu el interval de confiança en la escala del temps (desfeu els logs)

10.- Els següents dos gràfics mostren les diferències B-I per cada fitxer en ordenades en funció de les mitjanes  $[(B+I)/2]$  en abscisses. Primer, el gràfic inicial, sense transformar; i després, el gràfic amb la transformació logarítmica. Sabent que ordenar fitxers grans pot resultar en diferències mes grans, interpreteu aquests gràfics. Té sentit estimar una diferència única per aplicar a tots els casos amb les dades sense transformar (primer gràfic); i amb les dades transformades (segon)?



**B6** NOM: \_\_\_\_\_ COGNOMS: \_\_\_\_\_

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

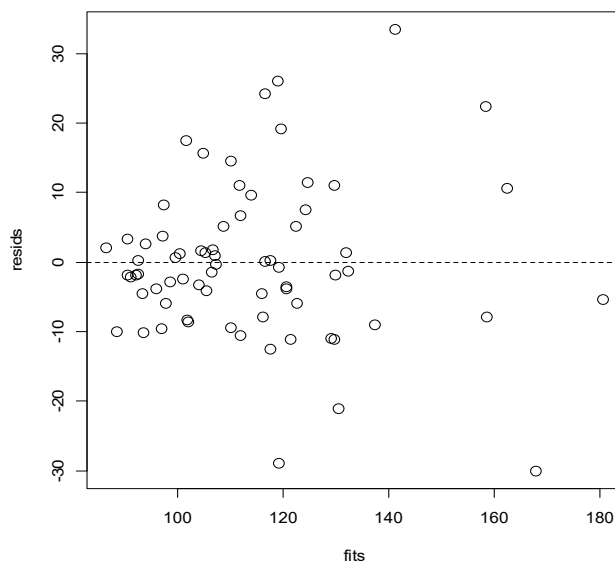
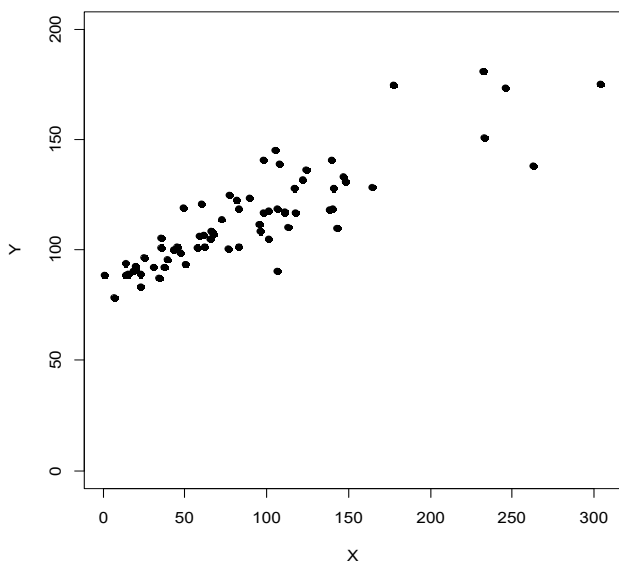
El temps emprat per a visualitzar una pàgina web a un navegador es pot descompondre en el temps destinat a la connexió amb el servidor remot, i el temps de processament del codi de la pàgina (descàrrega i mostrar a pantalla). Hem dissenyat un petit estudi, les dades del qual es troben a la dreta.

	1	2	3	4	5	6	$\sum v^2$	$\sum v$	Covar
Temps (cs)	84	96	135	112	98	136	75141	661	1365,6
Mida dades (KB)	16	45	125	76	22	190	60266	474	

La *mida de les dades* és l'espai que ocupa el fitxer HTML en kilobytes, i el *temps* és el temps mesurat en centèsimes de segon des de que es llença la petició fins a que la pàgina apareix completa al navegador (per tant, el temps total).

- [2pts] El primer objectiu de la recerca és el temps de la primera part, el destinat a la connexió, i que no depèn de la mida de les dades. L'anàlisi estadístic definit serà un model lineal amb les variables de la taula. Heu de trobar el valor de les estimacions pels paràmetres del model: 1) terme independent, 2) terme lineal, 3) desviació residual.
- [1.5pts] Expliqueu el significat de les estimacions anteriors (sigueu curosos amb les unitats corresponents a cada cas).
- [1.5pts] A partir de les estimacions resultants del model anterior, calculeu un interval de confiança al 95% per al temps esperat que es precisa per a connectar amb el servidor remot, i doneu una interpretació per a complementar el resultat.
- [1.5pts] Si es vol trobar un interval més estret per al paràmetre anterior, comenteu sobre l'eficàcia de les següents estratègies, justificant les respostes (preferiblement de manera formal):
  - Empraria una mostra més gran (per exemple, 12 observacions)
  - No augmentaria la mida de mostra, però augmentaria la mida de les pàgines
  - No augmentaria la mida de mostra, però disminuiria la mida de les pàgines

5. [1.5 pts] En base a la mida de la pàgina, quina és la capacitat de predicció del temps total del model anterior? Expliqueu com l'heu deduït. Com es coneix a l'indicador que heu utilitzat i què mesura?



Hem replicat l'estudi amb moltes més dades, els resultats del qual es mostren a les figures superiors.

6. [1 pt] Expliqueu breument cada un dels dos gràfics.

7. [1 pt] Amb l'ajut dels mateixos, valideu el model lineal aplicat a aquest cas: quines premisses es podrien valorar? Quines semblen admissibles, i perquè (o perquè no)? Si veieu alguna que no admetríeu, comenteu les possibles causes i solucions.