

NOM: _____ COGNOM: _____

Problema 1 (B4)

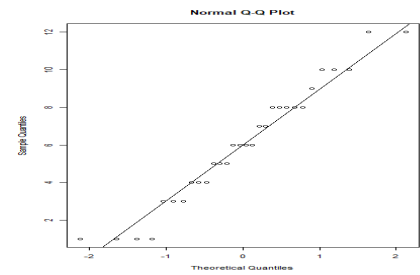
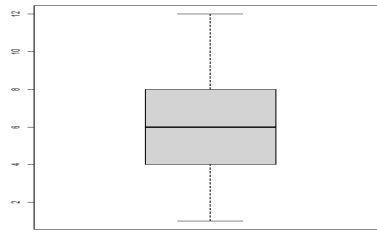
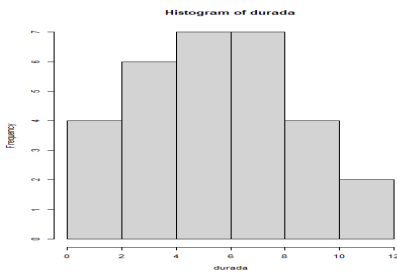
(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs)

Per avaluar el funcionament d'un aplicatiu de salut, una de les dades que es recullen és la durada en minuts en que els usuaris estan connectats. Just abans de la pandèmia els valors poblacionals dels paràmetres que s'assumien eren de 5 minuts de mitjana i 3 minuts de desviació

Per comprovar si l'ús de l'aplicatiu ha canviat molt, es recull una nova mostra de 30 durades:

$$\sum_{i=1}^{30} \text{durada}_i = 181$$

$$\sum_{i=1}^{30} \text{durada}_i^2 = 1385$$



1.- (0.5 punts) Amb els resultats de la mostra, justifiqueu si es pot considerar que segueixen el model normal.

2.- (0.5 punts) Calculeu una estimació puntual de l'esperança i de la desviació de la durada

3.- (1.5 punt) Calculeu un interval de confiança al 95% de l'esperança de la durada assumint la desviació poblacional de 3 minuts, i interpreteu-lo

4.- (1.5 punts) Calculeu un interval de confiança al 95% de l'esperança de la durada sense assumir que la desviació poblacional és coneguda, i compareu-lo amb el calculat a l'apartat 3

5.- (1.5 punt) Calculeu un interval de confiança al 99% de l'esperança de la durada sense assumir que la desviació poblacional és coneguda, i compareu-lo amb el calculat a l'apartat 4

6.- (2 punts) Uns dels responsables de l'anàlisi d'aquestes dades tenien la sospita que la pandèmia havia fet incrementar la mitjana de la durada de la connexió dels usuaris. Per això van calcular aquests resultats:

```
t.test(durada,mu=5,alternative="greater")
t = 1.7807, df = 29, p-value = 0.04272
alternative hypothesis: true mean is greater than 5
95 percent confidence interval:  5.047337      Inf
sample estimates:  mean of x      6.033333
```

Amb aquests resultats plantegeu la prova d'hipòtesis que representa (indiqueu hipòtesis, càlculs i conclusió), i interpreteu l'interval de confiança

7.- (1 punt) Amb aquesta mateixa mostra de 30 durades calculeu un interval de confiança al 95% per a la desviació i interpreteu-lo

8.- Per altra part, en un moment de molts intents d'accés, es vol quantificar el percentatge d'èxits accedint a l'aplicatiu. Per això es recull una nova mostra de 100 intents, obtenint que en 68 sí s'ha aconseguit l'accés.

(1.5 punts) Calculeu un interval de confiança per a la proporció d'intents que sí aconsegueixen accedir-hi i interpreteu-lo relacionant-lo amb el fet de que es voldria assegurar un 80% d'èxit d'accés

NOM: _____ COGNOM: _____

Problema 2 (B5)

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

Per comparar la velocitat de transferència de discos HDD (H) i SSD (S), s'han copiat fitxers, de 1 MB a 4 GB, amb un mateix ordinador. Es mesura la velocitat a MB/s de copiar uns mateixos 80 fitxers amb ambdós tipus de disc i decidint a l'atzar quin s'usarà primer. La taula següent proporciona el nombre de rèpliques, la mitjana i la desviació tipus o estàndard, per a cada tecnologia de disc i per a la seva diferència:

| | Nº observacions | Velocitat en MB/s | |
|------------|-----------------|-------------------|-----------|
| | | Mitjana | Desviació |
| H | 80 | 28 | 3 |
| S | 80 | 120 | 9 |
| S-H | 80 | 92 | 9 |

Indiqueu i justifiqueu si es tracta d'un disseny de dades aparellades o independents (0.5 punts).

Comenteu en cada cas (o disseny) què implica en quant a la variància de la diferència. (0.5 punts)

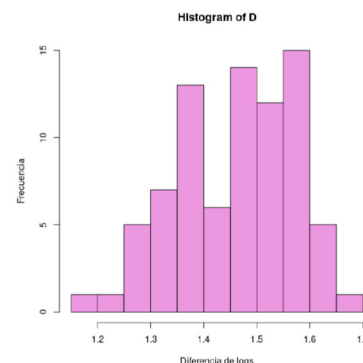
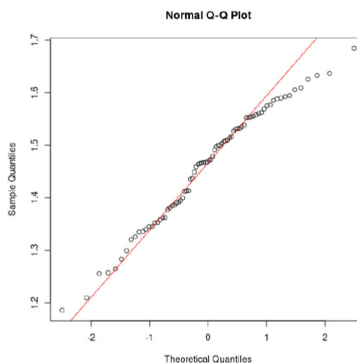
Si es tracten com a mostres independents (assumint normalitat i igualtat de variàncies poblacionals), calculeu:

- la desviació pooled i l'error estàndard de la diferència de mitjanes. (1 punt)

- un interval de confiança al 95% de la diferència de mitjanes (podeu utilitzar la convergència a la Normal per 'n' grans) (1 punt)

Es considera ara la diferència dels logaritmes $\ln(S)-\ln(H)$, obtenint aquests dos gràfics:

(1 punt) Interpreteu i indiqueu de què ens informen aquests dos gràfics.



Per la variable "ln(S)-ln(H)" els resultats han estat $\sum_{i=1}^{80} x_i = 116'88$ i $\sum_{i=1}^{80} x_i^2 = 171'72$.

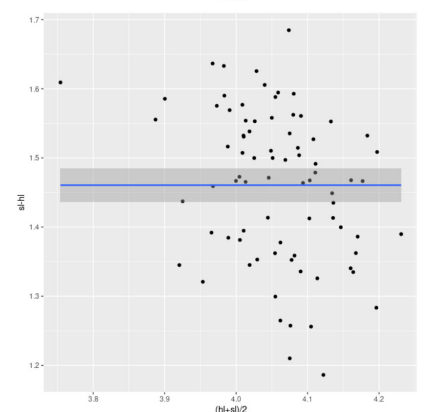
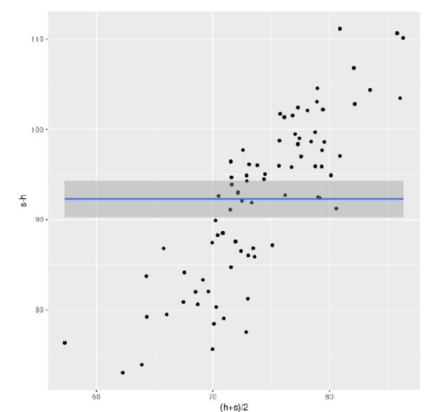
(1 punt) Feu una estimació puntual de l'esperança i de la desviació.

(1 punt) Interpreteu les estimacions anteriors.

(2 punts) Sabent que la funció de distribució corresponent a 1'664371 en una t de Student amb 79 graus de llibertat val 0.95, useu aquest valor per donar un interval simètric de confiança per a la diferència i indiqueu amb quina confiança s'haurà calculat.

(1 punt) Interpreteu l'interval anterior en relació a la comparació de les velocitats dels discs H i S.

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



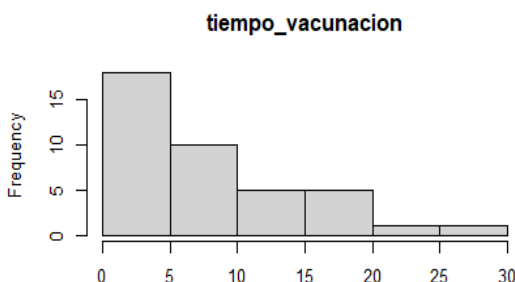
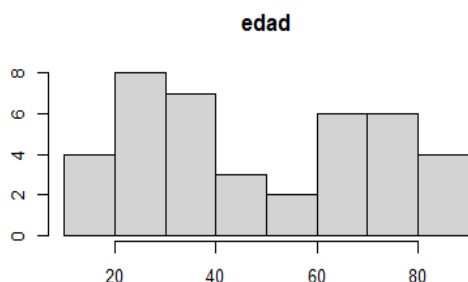
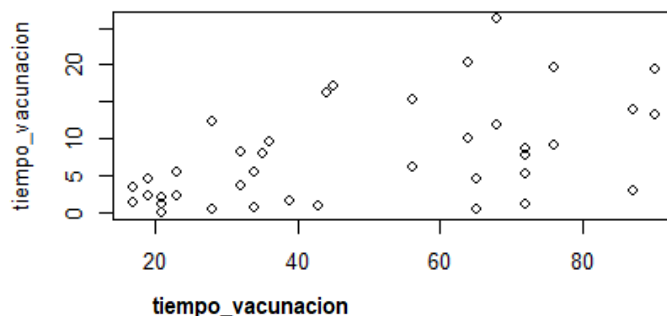
NOM: _____ COGNOM: _____

Problema 3 (B6)

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

El temps de vacunació s'entén com els minuts que passen entre la arribada d'una persona al centre sanitari, amb l'objectiu de vacunar-se, i la sortida amb la vacuna administrada. Donada l'actual situació de la COVID19 el personal sanitari es planteja si el temps de vacunació depèn de l'edat de les persones que s'han de vacunar, ja que s'han observat patrons que poden portar a pensar que sí que hi ha relació. A continuació, es proporciona alguna informació sobre les dades recollides en un centre sanitari que ha fet un estudi observacional:

tiempo de vacunació en funció de la edad



1 (1 punt) Amb la informació prèvia, creieu que seria raonable plantejar un model lineal per explicar el temps de vacunació en funció de l'edat?

2 (2 punts) Donat la següent sortida de R, comenteu les estimacions de tots els paràmetres del model lineal i què signifiquen:

```
Call:
lm(formula = tiempo_vacunacion ~ edad)

Residuals:
    Min       1Q   Median       3Q      Max
-10.265  -3.495  -1.191   2.466  15.661

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.55417    2.10631   0.263  0.793894
edad         0.14626    0.03917   3.734  0.000615 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

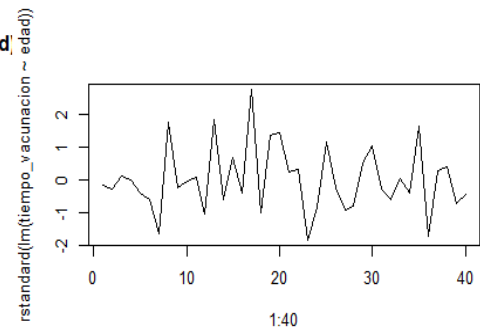
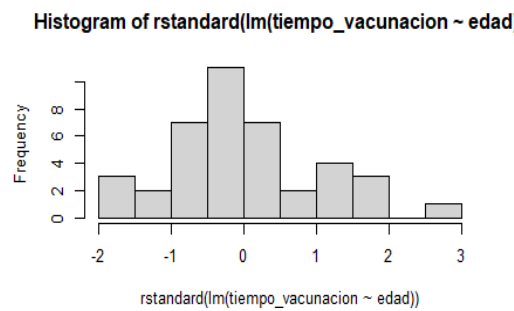
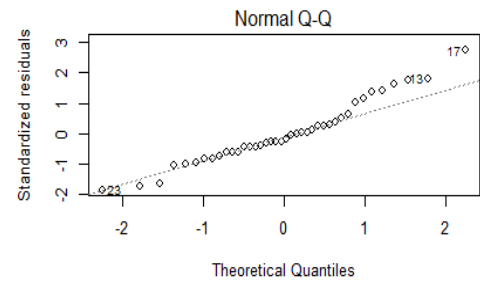
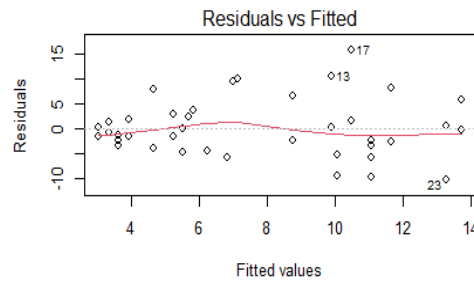
Residual standard error: 5.858 on 38 degrees of freedom
Multiple R-squared:  0.2685,    Adjusted R-squared:  0.2492
F-statistic: 13.95 on 1 and 38 DF,  p-value: 0.0006155
```

3 (1 punt) Valoreu la capacitat del model per explicar la variabilitat de la variable resposta.

4 (1 punt) Calculeu el IC del 90% de confiança per al pendent de la recta i interpreteu-lo.

5 (2 punts) Es vol analitzar si, donat el model anterior, podem afirmar que el temps de vacunació s'incrementa 1 minut per cada 3 anys del pacient (amb un risc del 5%):

6 (2 punts) Donats els següents gràfics de residus, valideu les premisses del model.



7 (1 punt) Si haguéssiu de fer alguna transformació sobre les dades, quina seria aparentment una transformació que podria millorar els resultats? Justifiqueu la proposta.