

NOM: _____ COGNOM: _____
(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs. Totes les preguntes valen igual)

Primer Problema (B4)

Primera part

Suposem que volem portar a terme un test d'hipòtesis bilateral per a comparar una esperança amb un valor concret. Explica els conceptes següents i **il.lustra** la teva explicació amb un gràfic.

1) Definiu *punt-crític*, i expliqueu per a què serveix. Calculeu el punt crític associat a una distribució t-Student amb 10 graus de llibertat, amb un $\alpha=0.01$.

2) Definiu *p-valor* i expliqueu per a què serveix. Calculeu el p-valor que correspon a un valor de 3.169 en una distribució t-Student amb 10 graus de llibertat.

Segona part

La diferència entre el temps de processament de dues implementacions diferents d'un algorisme es va mesurar en 7 workloads diferents. Les diferències de temps obtingudes varen ser de: 1.5, 2.6, -1.8, 1.3, -0.5, 1.7 i 2.4.

1) Explica com creus que s'han obtingut aquestes dades. Creus que provenen de dues mostres independents o aparellades? Perquè?

2) Quina és la distribució de probabilitat de la v.a. Y corresponent a la diferència de temps de processament de les dues implementacions? Com comprovaries que és l'acertada?

- 3) El nostre objectiu és portar a terme un test d'hipòtesis per a testar si les diferències de processament de les dues implementacions són o no estadísticament significatives.
Quin és el test d'hipòtesis escaient en aquest cas? Especifiqueu-ne les hipòtesis.
- 4) Quin és el valor de l'estadístic pel nostre conjunt de dades?
- 5) Conclueu si la hipòtesi nul.la pot o no ser rebutjada. Primer feu-ho per mitjà del punt crític, i després per mitjà del p-valor (calculeu un p-valor aproximat amb les taules).
- 6) Calculeu un interval de confiança al 95% per l'esperança de la diferència dels temps. Us porta a la mateixa conclusió que heu tret abans respecte al test d'hipòtesis?
- 7) Calculeu un interval de confiança per al valor esperat de Y menys 1 ($\mu - 1$).
- 8) A partir de l'interval de confiança del punt anterior, compareu si el valor esperat de Y és o no diferent de la unitat.

Nom:

(B5) Efecte d'una actualització del hardware a l'empresa informàtica PCaccount.com

El director de l'empresa PCaccount.com està interessat en demostrar l'efectivitat d'una important actualització del hardware existent en aquesta companyia. Per fer-ho es pren una mostra de 61 observacions abans de l'actualització i 61 més després de l'actualització. Basat en aquestes dades, el temps mitjà de funcionament és de $\bar{A} = 8.5$ minuts abans de l'actualització i $\bar{B} = 7.2$ minuts després de l'actualització. A més, es sap que $s_A^2 = 3.24$ i $s_B^2 = 2.25$.

- 1) Quina distribució de probabilitat segueix el quocient de variàncies. Raoneu la vostra resposta (1.5 punts)

- 2) Plantegeu i resoleu la prova d'hipòtesi d'igualtat de variàncies amb una confiança del 95%, sent l'alternativa $\sigma_A^2 > \sigma_B^2$ (1.5 punts)

- 3) Obtindríeu el mateix resultat de la pregunta 2 si l'alternativa a la prova d'hipòtesi d'igualtat de variàncies fos $\sigma_A^2 \neq \sigma_B^2$? Trebal·leu amb una confiança del 95%, (1.5 punts)

- 4) Si assumim que les variàncies poblacionals son iguals ($\sigma_A^2 = \sigma_B^2$), calculeu la variància "pooled" (comuna), així como la desviació estàndard "pooled" (1.5 punts)

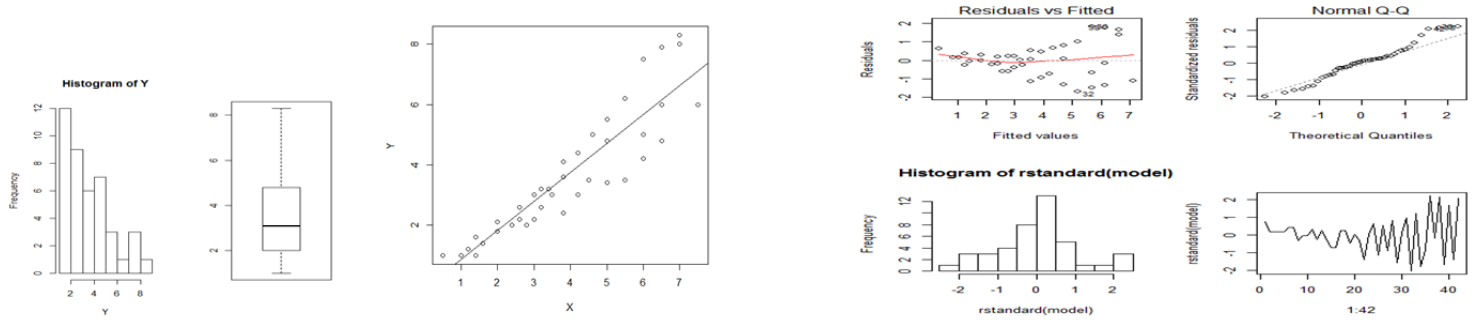
5) Calculeu un interval de confiança al 95% per a la diferència de mitjanes. (1.5 punts)

6) Ara, el director de l'empresa PCaccount.com se n'ha donat que potser treballar amb dues mostres independents no és la millor opció i pensa que és millor recollir un únic conjunt de dades i aplicar dos algorismes diferents per veure si realment un és més efectiu que l'altre. Quines avantatges/inconvenients representa treballar amb dades aparellades? Com es calcula, en aquest cas, la variància de la mostra? Raoneu la vostra resposta (1 punt)

7) Finalment, aquesta companyia decideix treballar amb dos conjunts de dades i està interessada en poder estimar la proporció d'èxits (un cert indicador particular) de cadascun d'aquests algorismes. Pel que fa al primer algorisme la proporció és 0.8 i pel segon algorisme és 0.5. El nombre d'observacions utilitzat pel primer algorisme és 50 i pel segon és 30. Es pot assumir que aquests dos algorismes tenen un rendiment similar? Raoneu la vostra resposta (1.5 punts)

Problema 3 (B6)

Volem estudiar la relació lineal entre les notes de dos parcials d'una assignatura. Es recullen les notes del primer parcial (X) i del segon parcial (Y) de 42 estudiants, i obtenim els següents resultats:



Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.10195	0.28491	-0.358	0.722
X	0.96198	0.06718	14.320	<2e-16

Residual standard error: 0.8467 on 40 degrees of freedom
 Multiple R-squared: 0.8368, Adjusted R-squared: 0.8327

1.- Indiqueu l'equació de la recta de regressió estimada i interpreteu-ne els coeficients

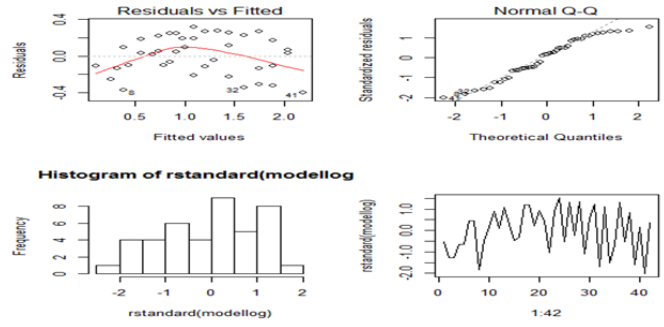
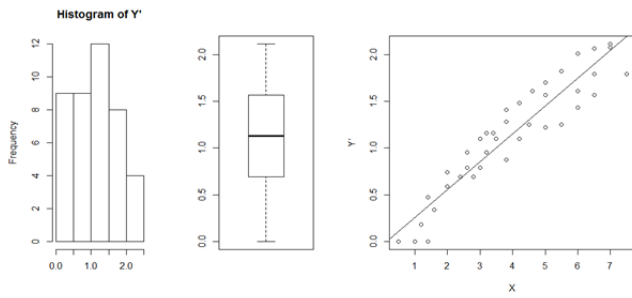
2.- Calculeu un interval de confiança al 95% pels paràmetres β_0 i β_1

3.- Comenteu què diuen els intervals anteriors respecte a la interpretació dels coeficients

4.- Calculeu el coeficient de determinació i el coeficient de correlació i interpretant-los

5.- 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.

Apliquem una transformació logarítmica a les notes del segon parcial (Y) obtenint unes noves dades (Y') pels 42 estudiants i n'estudiem la relació lineal amb X:



$$\sum x_i = 158.3 \quad \sum y'_i = 45.427 \quad \sum x_i^2 = 755.49 \quad \sum (y'_i)^2 = 65.056 \quad \sum x_i y'_i = 218.617$$

6.- Calculeu la covariància i la correlació entre X i Y'

7.- Calculeu la recta de regressió i interpreteu-ne els coeficients

8.- Poseu a prova si la recta de regressió es pot considerar plana amb un risc del 5%

9.- Calculeu una predicció puntual pel cas de X=5 i calculeu un interval de confiança d'aquesta predicció

10.- Compareu els dos models en quant al coeficient de determinació, la variabilitat residual i les premisses. Quin model creieu que és més encertat?