

NOM: _____ COGNOM: _____
(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs)

Problema 1 (B4)

(1) [2pts] Un experiment empíric recull dades (x_i) de la quantitat d'energia que capta un panell solar. Concretament, quanta energia produeix un panell en el temps de 60 minuts. Digueu quins dels següents elements corresponen a estadístics i quins no, justificant breument la resposta:

- $E(X)$, el valor esperat de la energia que el panell produeix en una hora [*wh: watts-hora, unitat d'energia*]
- n , el nombre de mesures que prenem per a l'experiment
- x_{\min} , la quantitat més petita d'energia que s'ha observat durant la recollida de dades
- $\sum x_i$, l'energia total produïda
- P , la potència nominal declarada pel fabricant [*w: watts, unitat de potència*]

(2) [2pts] Expliqueu el concepte d'error tipus (*standard error*) basant-se en dos casos concrets: l'error tipus d'una mitjana i l'error tipus d'una proporció. Utilitzeu un exemple per a cada cas amb nombres concrets inventats per vosaltres.

(3) [2pts] El resultat d'una enquesta pre-electoral ens diu que la intenció de vot per a determinat candidat està entre el 9.3% i el 12.9%. L'enquesta es basa en 1099 entrevistes telefòniques, però no es diu enlloc quin és el grau de confiança exacte d'aquest interval. Suposant que s'ha utilitzat una aproximació a la Normal per l'elevat nombre d'observacions, trobeu justificadament el grau de confiança emprat.

- (4) [1.5pts] A partir dels resultats de l'enquesta s'escolten les següents afirmacions a un programa de TV. Comenteu la validesa d'aquestes afirmacions:

TERTULIÀ A: "aquesta mostra no és fiable, hi ha milions d'electors, i només s'ha preguntat a poc més de mil"

TERTULIÀ B: "de cap manera, aquesta mostra és molt fiable, la forquilla té només 3.6 punts d'imprecisió"

TERTULIÀ C: "no sabeu què dieu, la mostra no és vàlida perquè (posa aquí que) *les entrevistes s'han fet només a telèfons fixes, de 9 del matí a 5 de la tarda*"

- (5) [2.5pts] Després de la realització de l'experiment del panell solar obtenim les següents dades: 40 mesures, amb una mitjana de 200 wh i una desviació tipus igual a 65 wh. Si la potència nominal del panell declara 180 w,

- tenim alguna evidència per a dir que la potència declarada no és la correcta?
- tenim alguna evidència per a afirmar que el panell produeix més energia que l'establerta?

Justifiqueu les dues qüestions amb les proves d'hipòtesis corresponents, i calculeu/aproximeu els p-valors de cada prova, junt amb un diagrama en el que representareu les solucions trobades.

a) proves d'hipòtesi corresponents als dos casos:

b) estadístics i punts crítics per a cada cas plantejat

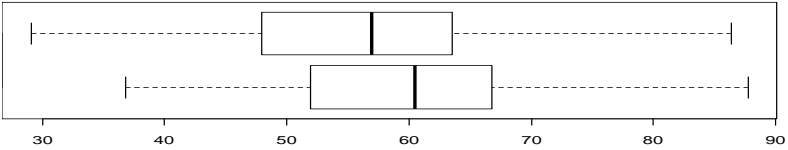
c) p-valors de cada prova i conclusions

d) Representeu les solucions trobades i els punts crítics en un diagrama

Problema 2 (B5)

Volem estudiar si la mida dels videojocs de pagament és més gran que la dels gratuïts i no tenim cap raó teòrica per explicar que els gratuïts puguin tenir mides més grans. Hem agafat 60 jocs de cada tipus, amb el següents resultats:

	mitjana	Desviació tipus
Gratuït	56.218	11.178
Pagament	59.993	12.1



..... **p-value = 0.04163**
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
0.189 Inf

1.- Si volem comparar les seves mitjanes, quin és el disseny (independent/aparellat) emprat? (0.5 punts)

- Expressau i justifiqueu les dues hipòtesis de la prova (0.5 punts)

- Valoreu com podríem argumentar amb els resultats de l’enunciat que es compleixen les premisses de normalitat i igualtat de variàncies (1 punt)

- Sota la hipòtesi d’igualtat, quin seria l’error tipus estimat per a la diferència de mitjanes mostrals? (1 punt)

- Indiqueu quin és l’estadístic de la prova i calculeu-lo (1 punt)

- Si no hi hagués cap diferència en la mida mitjana dels dos tipus de videojocs, com es distribuiria l’estadístic de la prova? Amb un risc $\alpha=5\%$, feu un gràfic per il·lustrar el o els punts crítics i situar les àrees d’acceptació i de rebuig de la hipòtesi nul·la (1 punt)

- A quina conclusió arribeu? Incorporeu l'interval de confiança del 95% a la discussió (1 punt)

2.- També volen estudiar la proporció de mesures per sota del llindar de 50. A la mostra n'hi ha 11 per A i 23 per B. Per comparar si els dos fabricants tenen igual proporcions (o no), indiqueu:

- Quines són les proporcions estimades de videojocs per sota de 50? (0.5 punts)

- Expresseu les hipòtesis de la nova prova (0.5 punts)

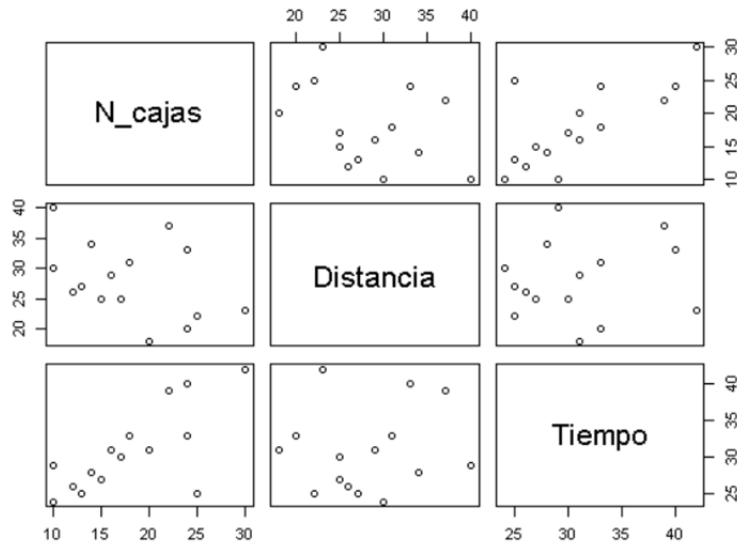
- Indiqueu quin és l'estadístic de la prova i calculeu-lo (1 punt)

- Com es distribueix l'estadístic sota la hipòtesi nul·la? Feu un gràfic indicant el o els punts crítics i la zona d'acceptació i de rebuig (1 punt)

- Interpreteu el resultat, i feu una conclusió global (1 punt)

Problema 3 (B6)

Disposem de les següents dades de nombre de caixes repartides, el temps invertit i la distància recorreguda i el model predictiu que vol estudiar quins són els factors importants per predir el temps d'enviament: l'exercici es base en els resultats de: <http://www.diegocalvo.es/analisis-de-regresion-lineal-multiple-en-r/>



a) Amb les dades que disposeu, comenta de forma descriptiva les variables fent èmfasi amb la relació que hi ha amb la variable resposta(1p):

S'ha realitzat un model lineal i s'han obtingut els resultats que veus a continuació:

```
call:
lm(formula = datos$Tiempo ~ datos$N_cajas, data = datos)

Residuals:
    Min       1Q   Median       3Q      Max
-10.6583  -1.6018  -0.1821   2.5262   5.3952

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.5452     3.4142   5.432 0.000115 ***
datos$N_cajas  0.6845     0.1805   3.791 0.002244 **
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.053 on 13 degrees of freedom
Multiple R-squared:  0.5251,    Adjusted R-squared:  0.4886 
F-statistic: 14.37 on 1 and 13 DF,  p-value: 0.002244
```

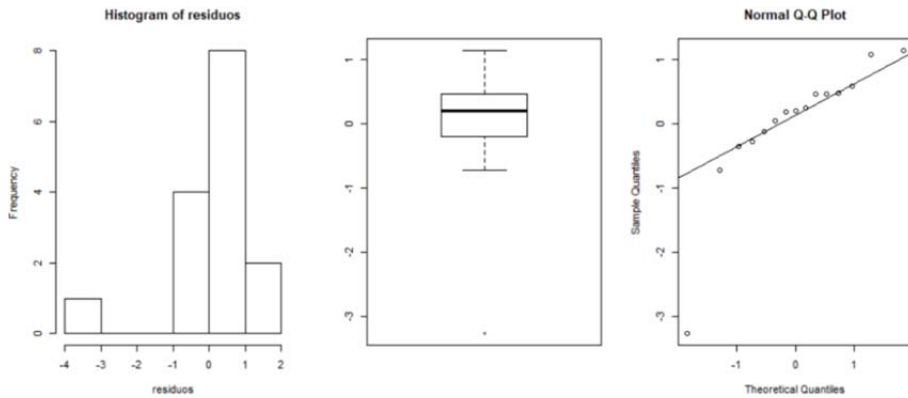
b) ¿És adient el model predictiu en termes dels coeficients estimats? Defineix quins són els contrastos de significació per als paràmetres del model (2p)

c) Calcula un IC per al terme β_1 (coeficient associat al nombre de caixes) del model amb un 95% de confiança (1p):

d) El temps està més relacionat amb el nombre de caixes o amb la distancia? (1p)

e) Comenta i valora la capacitat predictiva del model (1p):

Un cop analitzem els residus del model, veiem els següents gràfics amb els principals resultats:



f) Són els residus adients? Disposem de tota la informació? (1p)

g) Com es pot millorar la consistència dels resultats? Penses que seria interessant tenir més variables a l'estudi? Quines? (1p)

h) Quines consideracions hem fet a les variables del model? En cas que les hipòtesis no es compleixin, qué podem fer? (1p)

i) Quin seria el temps estimat per a 10 caixes? I per interval amb un 95% de confiança per al temps mitjà? (1p)