

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

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

## Problema 1 (B1-B2)

1. En una empresa determinada, el 50% dels documents estan escrits en anglès, 30% en català i 20% en castellà. A partir de les dades recollides sabem que el 40% del documents escrits en anglès tenen més de 15 pàgines; el 20% del documents escrits en català tenen més de 15 pàgines; i el 20% dels documents escrits en castellà tenen més de 15 pàgines.

1. a) Quina és la probabilitat d'escollir a l'atzar un document de més de 15 pàgines? (1 p)

1. b) Hem escollit un document a l'atzar i observem que té més de 15 pàgines, quina és la probabilitat que hagi estat escrit en castellà? (0.5 p)

2. Aquesta mateixa empresa també ha estudiat el nombre de fallades de hardware en un sistema informàtic. Per fer-ho estudien el nombre de fallades que ocorren en una setmana. Fan l'estudi durant les 52 setmanes de l'any. A partir de l'estudi troben que no ha ocorregut cap fallada en 9 setmanes de l'any. Aquesta informació i la de la resta de l'estudi la podeu trobar a la següent taula:

Nombre de fallades	0	1	2	3	4	5
Nombre de setmanes	9	13	15	10	5	1

2a) Indiqueu la funció de probabilitat del nombre de fallades en una setmana: (1 p)

2b) Indiqueu la funció de distribució del nombre de fallades en una setmana: (0.5 p)

2c) Quina és l'esperança d' $X$ ="Nombre de fallades en una setmana" (1 p)

2d) Calcula la variància i la desviació típica d' $X$ . (1 p)

2e) En la mateixa empresa han estudiat els errors d'una aplicació que corre en el mateix sistema. Denotem per  $Y$  el nombre d'errors de l'aplicació. A partir d'un estudi s'ha trobat que el nombre d'errors de l'aplicació està relacionat amb els errors del sistema mitjançant la funció  $Y = 5 \cdot X + 3$ .

e1) Calcula l'Esperança d' $Y$  (0.5 p)

e2) Calcula la desviació típica d' $Y$  (0.5 p)

3. En un laboratori mesuren el corrent en un circuit en ampers. A causa de diferents factors aleatoris, la mesura de  $Z$  varia. Els estudis realitzats indiquen que el corrent varia seguint la següent funció:

$$f(z) = \begin{cases} 0.025z + b, & 1 < z < 5 \\ 0, & \text{altrament} \end{cases}$$

3a) Troba  $b$  de manera que  $f(z)$  sigui una funció densitat i representa -la (2 p)

3b) Calcula la probabilitat que la mesura registrada sigui menor que 3. (1 p)

3c) Calcula l'esperança de les mesures registrades. (1 p)

NOM: \_\_\_\_\_ COGNOMS: \_\_\_\_\_  
(Contesteu cada pregunta en el seu lloc. Explíciteu i justifiqueu els càlculs)

## Problema 2 (B3-B4)

En una granja ecològica asseguren que la distribució del pes en grams d'un ou és Normal amb esperança 60 grams i variància 16 grams<sup>2</sup>.

Quina és la probabilitat que un ou pesi entre 55 i 65 grams?

Quina és la probabilitat que un ou pesi menys de 55.8 grams?

Si considerem una dotzena d'aquests ous, quina és la probabilitat de trobar-ne 6 que pesin menys de 55.8 grams? (definiu prèviament la variable nombre d'ous en una dotzena que pesin menys de 55.8 grams)

Quina variable representaria el nombre d'ous fins trobar-ne un de menys de 55.8 grams i quina és la probabilitat que sigui el primer

Quina variable representaria el pes total d'una dotzena d'ous i quina és la probabilitat que aquest pes total sigui inferior a 720 grams?

A la granja han decidit recollir dades del pes dels ous per tenir evidències quantitatives dels valors que assegurin que compleixen. Per això anoten el pes d'una dotzena d'ous i obtenen els següents valors:

$pes \leftarrow c(64, 58, 54, 53, 62, 55, 56, 60, 59, 54, 63, 63)$

$$\sum_{i=1}^{12} pes_i = 701 \qquad \sum_{i=1}^{12} pes_i^2 = 41125$$

Calculeu una estimació puntual de la mitjana i de la desviació poblacionals del pes dels ous

Calculeu i interpreteu un interval amb confiança del 95% de la mitjana poblacional

Calculeu i interpreteu un interval amb confiança del 95% de la desviació poblacional

Contrasteu si la mitjana poblacional del pes dels ous és 60 o no amb un risc del 5%. Indiqueu les hipòtesis, el càlcul de l'estadístic i justifiqueu la conclusió respecte les hipòtesis

Relacioneu els resultats de l'interval de confiança i de la prova d'hipòtesis anteriors

### Problema 3 (B5-B6)

Un enginyer informàtic es qüestiona quin algorisme pot emprar per trobar un cert node “marcat” en un arbre binari. Disposa de dues opcions, el BFS (Breadth-first search) o el DFS (Depth-first search), i voldria escollir el que en terme mitjà és més ràpid. Per prendre una decisió, utilitza un generador d'arbres binaris i marca un node aleatòriament. Amb aquests arbres, compta el nombre de nodes explorats per l'algorisme fins trobar el node marcat.

1. Si es sospita que, per a arbres d'una alçada determinada, quan el BFS triga molt a trobar el node, el DFS normalment és més ràpid, i viceversa: quin disseny us sembla més adient, el de dues mostres independents o el de mostres aparellades? Raoneu la resposta. (1 pt)
2. L'enginyer ha fet una prova amb dues mostres d'alçada 15, cadascuna amb els seus arbres (14 arbres en cada grup). Les mitjanes corresponents han estat 23920 i 31091 nodes, DFS i BFS respectivament. Les desviacions tipus són 17895.36 i 21749.15. Amb aquesta prova es pot confirmar que hi ha diferències entre els dos mètodes? Feu una prova d'hipòtesis formal per justificar la resposta. (*aquesta pregunta no té cap relació amb la resposta de la pregunta anterior*) (1 pt)
3. És correcte dissenyar la prova com a unilateral perquè hem vist que el BFS ha trigat més en mitjana? Doneu arguments amb la resposta. (1/2 pt)
4. Posteriorment, l'enginyer amplia el seu experiment amb arbres d'alçada variable: decideix generar alçades d'entre 5 i 24 de forma uniforme, i utilitzar un arbre amb aquesta alçada per ambdós algorismes. En aquest disseny, com serà la correlació entre les respostes BFS i DFS? Per què? (1/2 pt)
5. Observeu aquestes sortides de R: (2 pts)

Paired t-test	Paired t-test
<pre>data: Z\$n_DFS and Z\$n_BFS t = -1.365, df = 199, p-value = 0.1738 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -1153782.2 209852.8 sample estimates: mean of the differences -471964.7</pre>	<pre>data: log(Z\$n_DFS) and log(Z\$n_BFS) t = -0.54263, df = 199, p-value = 0.588 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.245322 0.139445 sample estimates: mean of the differences -0.05293849</pre>

- Comenteu breument quina prova s'està fent en cada cas.
- Calculeu la desviació tipus de la diferència que s'està emprant en cada cas.
- Com es pot interpretar l'interval (-0.2453, 0.1394)?
- A quin resultat s'arriba en cadascú dels anàlisis? Què es pot dir sobre la comparació dels dos algorismes?
- Dels dos anàlisis de dalt, hi ha alguna opció millor que l'altra? Per què?

(Contesteu cada pregunta en fulls apart. Poseu ben clar el número de l'apartat. Explíciteu i justifiqueu els càlculs)

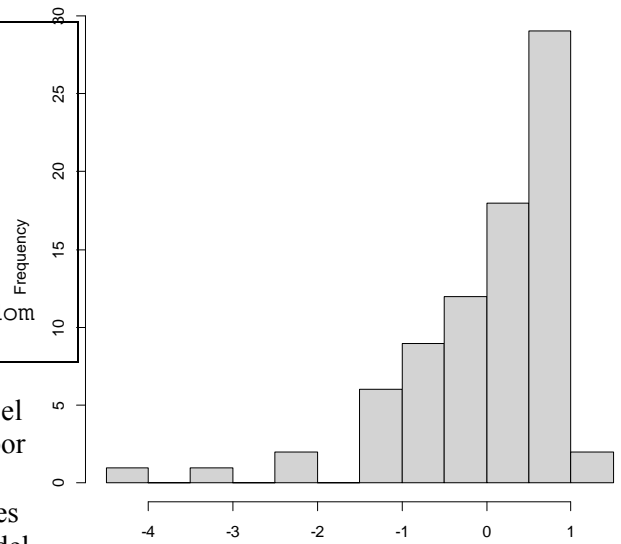
```

Residuals:
  Min       1Q   Median       3Q      Max
-4.4026 -0.4402  0.2295  0.7192  1.0144

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.11551    0.28713   -0.402   0.689
Altura       0.68359    0.01802  37.934 <2e-16 ***

Residual standard error: 0.9713 on 78 degrees of freedom
Multiple R-squared:  0.9486

```



Continuant amb el cas dels algorismes de cerca en arbres, hem investigat el DFS per modelar el logaritme del nombre de nodes recorreguts explicat per l'alçada del arbre (*Altura*) en un subconjunt de 80 casos. Amb aquesta transformació, la relació segueix una tendència lineal indiscutible. De totes formes, encara es plantegen dubtes respecte a les altres premisses del model.

6. Amb les sortides annexes, discutiu sobre el grau d'acompliment de les premisses amb aquestes dades (les vostres conclusions s'han de justificar amb elements concrets de les sortides). (1 pt)

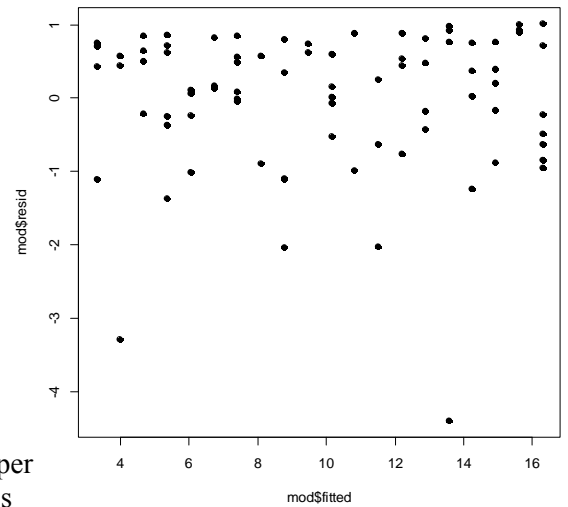
7. Expliqueu què representen els resultats següents: (1 pt)

- -0.11551
- 0.28713
- 0.68359
- 0.01802
- 0.9713
- 0.9486

*No es vàlid simplement prendre els mots que apareixen a la sortida.*

8. Teòricament, si un arbre té un nivell més, el nombre de nodes es duplica i per tant el cost de trobar un determinat node marcat també. El logaritme de 2 és 0.6931472; comprova fent una prova d'hipòtesis si amb les dades d'aquest experiment la teoria és coherent o no. Interpreta el resultat trobat. (2 pts)

9. El nostre algorisme DFS ha d'explorar un determinat arbre d'alçada igual a 20. Amb el model anterior, feu una estimació puntual i una per interval de confiança 90% del nombre de nodes que pot recórrer fins a trobar el node marcat (*Nota: l'alçada dels arbres té una mitjana de 14.75 i una desviació tipus de 6.064*). (1 pt)



*(Contesteu cada pregunta en fulls apart. Poseu ben clar el número de l'apartat. Expliqueu i justifiqueu els càlculs)*