

NOM: _____

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

Problema 1 (B1-B2)

Es volen estudiar característiques d'una eina de resolució de problemes online que s'usa en una assignatura. Els problemes de la col·lecció estan etiquetats com a fàcils (F) mitjans (M) o difícils (D). Al llarg dels quadrimestres s'ha recollit informació del seu funcionament en quant a si les execucions dels problemes un cop corregides obtenen una nota d'aprovat (5 o superior) o no, i del nombre d'execucions dels problemes que es realitzen i del temps que es triga en cada execució.

Per una part anem a estudiar si obtenir un aprovat (A) o no (NA) en l'execució d'un problema té a veure o no amb l'etiqueta de dificultat (hi ha un 50% de problemes fàcils i un 20% de difícils). Sabem que de les execucions de problemes fàcils un 80% obtenen nota aprovada, mentre que dels mitjans només un 50%. I sabem que en global les execucions aprovades són un 60%.

Indiqueu i justifiqueu les probabilitats de F, M i D. I les probabilitats de A i NA (1pt)

Calculeu la probabilitat que el problema sigui difícil i aprovat (1pt)

Indiqueu l'arbre d'esdeveniments tenint en compte l'etiqueta i l'aprovat o no. Indiqueu el conjunt de resultats amb les seves probabilitats (1pt)

Calculeu la probabilitat de que una execució aprovada sigui de problema fàcil (1pt)

Un estudiant afirma que l'etiqueta de dificultat no té res a veure amb la probabilitat d'aprovar-ne o no una execució. Indiqueu i justifiqueu si creieu que té raó o no (1pt)

Per altra part anem a estudiar, només pels problemes fàcils i difícils, el nombre i el temps de les execucions.

En el cas del nombre d'execucions s'ha recollit (només pels problemes fàcils i pels difícils i pels casos de 1, 2 o 3 execucions) les següents proporcions

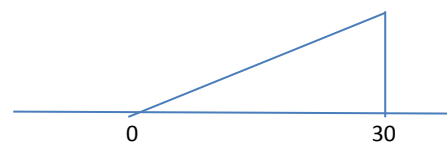
	Fàcils	Difícils
1 execució	0.30	0.20
2 execucions	0.10	0.07
3 execucions	0.20	0.13

Calculeu pels problemes fàcils la probabilitat de fer 1 execució. I la de fer-ne 2. I la de fer-ne 3 (1pt)

Calculeu l'esperança i la variància del nombre d'execucions pels problemes fàcils (1pt)

Indiqueu i justifiqueu formalment si és independent o no l'etiqueta fàcil/difícil del fet de fer 1,2 o 3 execucions (1pt)

I finalment pel temps de les execucions de problemes fàcils i el dels difícils s'ha recollit la distribució de les execucions que han trigat entre 0 i 30 minuts, obtenint respectivament:



Indiqueu la funció de distribució de probabilitat i calculeu la probabilitat de trigar menys de 15 minuts en els dos casos (1pt)

Interpreteu i comenteu les diferències dels dos casos tenint en compte la distribució i els càlculs anteriors (1pt)

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

Problema 2 (B3-B4)

La tecnologia d'un model concret de discs d'estat sòlid de 512 GB consisteix en 8 sectors de 64 GB. Per la seva part un sector es compon d'un gran nombre de "pistes". Anomenarem X al nombre de pistes que al cap d'un any d'ús s'inutilitzen a un sector: assumiu que els errors són independents uns dels altres i que el nombre mitjà és de 2.25/any/sector.

1. Doneu el model de probabilitat que segueix la variable aleatòria X , justificant la resposta. Quina és la probabilitat que en un any un sector donat no presenti cap pista inutilitzada? (1pt)
2. Calculeu la probabilitat que un disc d'un any d'ús tingui més d'un sector amb totes les pistes correctes. (1pt)
3. Amb un error del 5%, quin és el màxim nombre de pistes inutilitzades que pot haver a un disc d'un any d'ús? (1pt)
4. Si el disc ha funcionat 4 anys, digueu un model de probabilitat aproximat per el nombre de pistes inutilitzades, amb els paràmetres corresponents. Trobeu la probabilitat que un disc d'aquests presenti més de 100 pistes fora d'ús. (1pt)
5. Escolliu la resposta més encertada: "Un disc de cada ____ (a: 100; b: 2000; c: 20) presenta més de 50 pistes inútils al cap de 2 anys", i justifiqueu la resposta. (1pt)

6. Amb els discs d'un altre fabricant, s'assumeix en principi que la variància de la variable Y: "nombre de pistes inutilitzades després de sis mesos d'ús" val 9, i es pretén estimar el valor esperat de la variable Y. Calculeu el nombre de discs que caldria observar per obtenir un interval de confiança al 90% amb una amplitud no major de 1. (1pt)
7. Es decideix prendre una mostra de 15 observacions, mantenint per similitud amb el primer fabricant una suposada variància poblacional igual a 9. Entre els 15 discs, seguits durant mig any i completament independents uns dels altres, hi havia 160 pistes inutilitzades. Poseu a prova formalment si la mitjana de la variable Y pot ser igual a 9. Trobeu el p valor de la prova i expliqueu la conclusió. (1.5pt)
8. Repetiu la prova sense adoptar una variància per a la població. La variància mostral valia 9.53. No cal trobar el p valor, però representeu en la distribució de l'estadístic les zones de rebuig i acceptació. Quina és la conclusió? (1.5pt)
9. Per a l'anàlisi anterior, critiqueu les premisses següents: (1pt)
- Mostra aleatòria simple
 - Normalitat de la variable resposta
 - Mida de la mostra
 - Homocedasticitat

Problema 3 (B5-B6)

En un estudi per comparar les notes de batxillerat (B) i les notes a l'examen de selectivitat (S) es fa una enquesta preguntant a 16 alumnes què han tret en ambdues notes. *Per contestar els següents apartats considereu un risc $\alpha=0.05$ i una confiança del 95%.*

	Mean	Var
B	7.03	
S	6.19	
D = B-S	0.84	2.07

- 1) Quin tipus de mostres estaríem tractant, aparellades o independents? Justifiqueu la resposta. **[1p]**

- 2) Imagineu que tractem les dades aparellades.
 - a) Plantegeu les hipòtesis nul·la i alternativa per estudiar si la nota de Batxillerat és superior a la de Selectivitat. Indiqueu si la prova és bilateral o unilateral. **[0.5p]**

 - b) Indiqueu la distribució de l'estadístic corresponent sota la hipòtesi nul·la i les premisses. **[1p]**

 - c) Calculeu l'estadístic segons les dades i raoneu si podem rebutjar la hipòtesi nul·la. **[1p]**

- 3) Imagineu que considerem les dades com a mostres independents. Volem posar a prova si les variàncies de les notes de Batxillerat i de Selectivitat són iguals.
 - a) Plantegeu el contrast d'hipòtesi. **[0.5p]**

L'output de R que obtenim és el següent:

```
F test to compare two variances
```

```
data: data$S and data$B
F = --- , num df = 15, denom df = 15, p-value = 0.5
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.486693  3.986782
```

- b) A partir de l'output, calculeu l'estadístic F i interpreteu-ho. Suggeriment, dibuixeu el gràfic de la F **[1p]**

c) Observant només l'interval de confiança, a quina conclusió arribaríem? Justifiqueu-ho. **[0.5p]**

d) Quin seria el mínim rati de variàncies que donaria lloc a rebutjar la hipòtesi nul·la? **[0.5p]**

4) Volem estudiar si existeix relació entre les dues notes. Per fer-ho, hem estimat un model lineal amb R que ens ha donat el següent output:

```
Call:
lm(formula = S ~ B, data = data)
[...]
Coefficients:

                Estimate      Std. Error  t value    Pr(>|t|)
(Intercept)      ---            ---      1.413     0.179
B                 ---            ---      1.606     0.131

Residual standard error: 1.335 on 14 degrees of freedom
Multiple R-squared:  0.1555,    Adjusted R-squared:  0.09519
F-statistic: 2.578 on 1 and 14 DF,  p-value: 0.1307
```

a) Calculeu els estimadors de β_0 i β_1 , i escriviu l'equació de la recta sabent que $S_S / S_B = 1.2$. **[1p]**

b) Podem considerar que la nota de Batxillerat està associada amb la nota de selectivitat(Justifica la resposta amb arguments d'inferència estadística)? Calculeu també l'interval de confiança del pendent. **[1.5p]**

c) Podem dir que el model fa un bon ajustament de les dades? **[0.5p]**

d) Segons els plots següents, podem validar les premises? **[1p]**

