

login:

Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs.

Problema 1. (Blocs 1 i 2)

Un sistema d'emmagatzematge de fitxers de dades consta de dos servidors anomenats SMALL i LARGE. La mida (en Gb) dels fitxers que arriben al sistema és una variable aleatòria contínua X amb funció de distribució

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 2x - x^2 & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

1. (1 punt) Calculeu la funció de densitat d' X .

2. (1 punt) Calculeu el valor esperat d' X .

El mòdul de control del sistema fa que tots els fitxers de mida menor o igual que 0.25 Gb es desin a SMALL i tots els que siguin més grans de 0.5 Gb es desin a LARGE. Quan un fitxer té mida entre 0.25 i 0.5 Gb es sorteja, amb probabilitats $1/2$ i $1/2$, si es desa a un servidor o a l'altre.

3. (2 punts) Calculeu la probabilitat que un fitxer que arriba al sistema acabi desat a LARGE.

Indicació: Expliqueu primer perquè $\Pr(X \leq 0.25) = 7/16$ i $\Pr(X > 0.5) = 1/4$.

4. **(2 punts)** Si sabem que un fitxer està a **SMALL**, quina probabilitat hi ha que la seva mida sigui més petita o igual que 0.25 Gb?

Ens diuen que el temps d'accés a un fitxer és 1 milisegon si la mida del fitxer és més petita o igual que 0.25 Gb, és 2 milisegons si la seva mida està entre 0.25 Gb i 0.5 Gb i està desat a **SMALL**, és 3 milisegons si la seva mida està entre 0.25 Gb i 0.5 Gb i està desat a **LARGE**, i és 4 milisegons si la mida és més gran que 0.5 Gb.

5. **(1 punts)** Doneu la funció de probabilitat de la variable aleatòria T , que mesura el temps d'accés (en milisegons) a un fitxer del sistema.

6. **(2 punts)** Calculeu l'esperança i la desviació estàndard de T .

Definim la variable aleatòria B que val 0 si $X \leq 0.5$ Gb i val 1 si $X > 0.5$ Gb.

7. **(1 punts)** Doneu la funció de probabilitat conjunta de B i T , digueu si són independents i expliqueu per què.

NOM: _____

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

Problema 2 (B3-B4).

El temps en minuts que els estudiants empen per solucionar un problema d'e-status a PE en aquest quadrimestre ha estat:

```
> summary(minu)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     Sd.
0.2833  5.6830 10.5200 13.7600 19.1500 85.6800 10.9429
```

Admetem que la variable Y ="temps per resoldre un problema" es distribueix com una exponencial.

1. Trobeu les probabilitats que Y sigui inferior a 5 minuts,

- suposant que el paràmetre λ el preneu de la mitjana mostral. (1pt)
- suposant que el paràmetre λ el preneu de la desviació tipus. (1pt)
- suposant que el paràmetre λ el preneu de la mediana. (1pt)

Recordeu quines són les propietats del model exponencial per derivar el valor de λ a cada cas.

2. La Neus, en canvi, presenta una distribució molt diferent, ja que ella es pren amb calma cada exercici i això suposa que els seus temps segueixin un model Normal. El 35% dels casos els resol en menys de 22 minuts, i per 1 de cada 4 triga més de 33 minuts.

a. Quin és el temps esperat per la Neus per resoldre un exercici? Quina és la variància del seu temps? (2.5pt)

b. Si una tarda de diumenge es posa a fer e-status, resol 5 exercicis, amb una pausa intermèdia després de cada exercici, que es distribueix com $N(10\text{min}, 2.5\text{min})$: calcula la probabilitat que el temps total hagi estat de més de tres hores (suposeu que els temps són independents uns d'altres). (2.5pt)

3. 8 de cada 10 exercicis s'han fet entre les 20h i les 00:00. Si seleccionés de la base de dades (que és molt gran) una mostra a l'atzar de N exercicis, trobeu la probabilitat que almenys un 70% s'hagin resolt després de les 20h:

a. si N és 20. (1pt)

b. si N és 200 (trobeu una aproximació al resultat exacte). (1pt)

```
> round(qchisq(c(0.005,0.01,0.025,0.05,0.10,0.25,0.75,0.9,0.95,0.975,0.99,0.995), 109), 3)
[1] 74.724 77.616 82.000 85.903 90.558 98.712 118.563 128.298 134.369 139.784 146.257 150.774
```

Ara ja no suposarem que coneixem un model o un paràmetre concret per a cap variable: només disposem de dades.

4. Volem recollir una mostra per estimar la proporció dels exercicis que es fan després de les 20h, i es desitja obtenir un interval de confiança al 95% amb amplada inferior a 6% (és a dir, la precisió serà de $\pm 3\%$). Quin és el nombre de dades que necessitem recollir? Assumeix màxima incertesa sobre la proporció poblacional (2pt)

5. Com podríem aconseguir una mostra més petita (sense canviar l'objectiu, ni modificar la confiança o l'amplada)? (1pt)

6. Com és molt raonable que els temps de resolució d'exercicis no siguin una distribució Normal, es vol analitzar la variable $X = \log(Y)$, que seria molt menys problemàtica. Una mostra de 110 observacions independents de X es descriu a continuació:

```
> summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.    Sd.
-0.08701  1.70600  2.32100  2.19500  2.75700  4.09200  0.88212
```

a. Creieu que hem fet bé escollint X en lloc de Y ? Justifiqueu la resposta. (1.5pt)

b. En qualsevol cas, estimeu $E(X)$ per interval de confiança 90%. Feu també l'estimació de σ^2 per interval de confiança 90%. (1.5pt)

c. Volem posar a prova si la desviació de X és 1: desenvolueu la prova formal per contestar la qüestió, enfront de l'opció que diu que la desviació seria menor que 1. Amb la taula de valors del peu de pàgina procureu afitar el p-valor resultant. (2.5pt)

d. Hauria estat correcte realitzar aquesta prova amb les dades en forma de temps (és a dir, amb una mostra de Y)? Justifiqueu la resposta. (1.5pt)

```
> round(qchisq(c(0.005,0.01,0.025,0.05,0.10,0.25,0.75,0.9,0.95,0.975,0.99,0.995), 109), 3)
[1] 74.724 77.616 82.000 85.903 90.558 98.712 118.563 128.298 134.369 139.784 146.257 150.774
```

NOM: _____

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

Problema 3 (B5-B6).

Un FIBER molt avantatjat ha dissenyat un nou tipus de CPU i vol verificar si el seu prototipus superarà o no al model de CPU actual amb el que està treballant. Per verificar-ho pren una mostra en 32 CPUs, 16 de les antigues que anomenarem A, i 16 de les noves anomenades N, amb els següents resultats de mitjanes i desviacions d'una variable de temps:

$$\bar{t}_A = 3.10 \quad s_{T_A} = 1.66 \quad \bar{t}_N = 3.07 \quad s_{T_N} = 1.19$$

1.- (3 punts) Poseu a prova si es pot considerar les variàncies poblacionals iguals o una significativament superior a l'altra:

- hipòtesis (indiqueu si és prova unilateral o bilateral), estadístic, distribució i premisses adients

- càlcul de l'estadístic, punt crític amb confiança 95%, i representació gràfica (indicant el p_valor i com es calcula per exemple en R)

- es pot rebutjar la hipòtesis nul·la? Justifiqueu-ho i interpreteu-ne les conclusions

2.- (3 punts) Poseu a prova si es pot considerar que les dues mitjanes poblacionals són iguals o no amb confiança del 95%

- hipòtesis (indiqueu si és prova unilateral o bilateral), estadístic, distribució i premisses adients

- càlcul de l'estadístic, punts crítics, i representació gràfica (indicant el p_valor i el seu càlcul per exemple en R)

- es pot rebutjar la hipòtesis nul·la? Justifiqueu-ho i interpreteu-ne les conclusions

3.- (4 punts) Un professor li comenta a aquest estudiant que pot plantejar el problema com si fossin dades aparellades fent exactament les mateixes proves en els dos casos. Suposem que es fa així i que s'obtenen els resultats indicats a l'inici i, a més a més, es calcula la correlació entre les dades antigues i les noves obtenint $r_{T_A T_N} = 0.77$ amb una variància residual igual a 1.15 en el model de regressió lineal del temps del model nou en funció del temps del model antic.

- calculeu la recta de regressió

- poseu a prova si la recta es pot considerar que no és plana

- doneu un interval de confiança al 95% de la pendent de la recta

- doneu una estimació puntual de T_N quan T_A val 3 amb un interval de confiança al 95% d'aquesta predicció