

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

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

## Problema 1 (Bloc A)

1. En un taller de màgia matemàtica, es dipositen tres monedes no trucades en una caixa i se n'extreu una a l'atzar. Després es llença dues vegades, quina és la probabilitat que surtin dues cares? [0,5 punts]

Tot seguit se substitueix una de les monedes no trucades, per una de trucada amb dues cares. Ara a la caixa hi ha dues monedes no trucades (amb cara i creu) i una moneda que té cara a ambdues bandes. Tornem a repetir l'experiment d'extreure una moneda a l'atzar.

2a. Si es llença la moneda dues vegades, quina és la probabilitat que surtin dues cares? [0.75 punts]

2b. Si es llença tres vegades, quina és la probabilitat que surtin almenys dues cares? [0,75 punts]

2c. Finament es llença la moneda tres vegades i s'obtenen tres cares, quina és la probabilitat que la moneda llençada hagi estat la trucada? [1 punt]

La següent activitat es realitza amb dues bosses. Una bossa A conté 2 boles negres i 3 boles blanques i una segona bossa B conté 4 boles negres i 2 boles blanques. Es tria una de les bosses a l'atzar i se n'extreuen tres boles de manera consecutiva (sense reposició). Volem observar el nombre de boles blanques extretes (N)

3a. Trobeu la funció de probabilitat de la variable aleatòria N: "Nombre de boles blanques extretes". [2 punts]

3b. Calculeu l'esperança i la desviació típica de la variable aleatòria N. [1 punt]

3c. Si el nombre de boles blanques extretes és 2, quina és la probabilitat que la bossa escollida hagi estat l'A? [0,5 punts]

4. Considereu la variable aleatòria X, amb la següent funció de densitat:

$$f(x) = \begin{cases} x^2 & 0 \leq x \leq 1 \\ ax + b & 1 < x \leq \frac{7}{3} \\ 0 & \text{altrament} \end{cases}$$

Trobeu els valors de a i b per a que f(x) sigui una funció de densitat contínua. [2 punts]

Considereu ara la variable T que estudia el temps (en segons) d'un determinat experiment aleatori i que segueix una distribució amb la següent funció de densitat:

$$f(t) = \begin{cases} 3t^2 & \text{si } 0 \leq t \leq 1 \\ 0 & \text{altrament} \end{cases}$$

5. a) Calculeu la funció de distribució de T (només cal calcular la funció de distribució sense haver de comprovar que f(t) és una funció de densitat) [0,5 punts]

b) Calculeu el valor esperat de T. [0,75 punts]

c) Calculeu la variància de T. [0,75 punts]

Problema bloc B

Una investigació sobre un cert tipus d’imatges (reals i generades per IA) estudia característiques com l’índex de nitidesa, la mida i el format dels fitxers corresponents. Per a la nitidesa s’ha establert un índex (un valor més alt indica més nitidesa) amb distribució Normal, que pel cas de la **nitidesa d’imatges reals (Nr)** té mitjana 450 i desviació estàndard 100, i pel cas de **nitidesa d’imatges generades (Ng)** té mitjana 500 i desviació estàndard 50. La **mida d’imatges (M)**, tant reals com generades, no té distribució Normal, i té mitjana 100 MB i variància 2500.

Definiu explícitament variables i models on calgui, i expliciteu formalment els passos en el càlcul de probabilitats

1.- (1 punt) Dibuixeu, pels dos índexs de nitidesa, les dues funcions de densitat en un gràfic, i les dues funcions de distribució en un altre gràfic. Indiqueu, en els dos gràfics, els valors de l’eix horitzontal i només en el cas de les funcions de distribució, els valors de l’eix vertical

2.- (1 punt) Calculeu la probabilitat que per a una **imatge real**, l’índex de **nitidesa** sigui superior a 400. Calculeu-la també per a una **imatge generada** i dibuixeu-les al gràfic anterior

3.- (1 punt) Calculeu la probabilitat que l’índex de **nitidesa mitjà** d’una col·lecció de 25 **imatges reals** sigui superior a 400.

4.- (1 punt) Calculeu la probabilitat que la **mida total** d’una col·lecció de 25 **imatges** sigui inferior als 2250 MB. Indiqueu també quin valor de MB ens assegura que el 50% de les mides totals no el superaran

pnorm(0.1) = 0.5398278	pnorm(0.5) = 0.6914625	pnorm(1) = 0.8413447	pnorm(1.5) = 0.9331928	pnorm(2) = 0.9772499	pnorm(2.5) = 0.9937903
qnorm(0.1) = -1.281552	qnorm(0.5) = 0	qnorm(0.01) = -2.32635	qnorm(0.15) = -1.03643	qnorm(0.2) = -0.841621	qnorm(0.25) = -0.67449

Ara estudiarem el format i el nombre d'imatges de fenòmens meteorològics rebudes a la pàgina web d'un canal televisiu. Definiu explícitament variables i models on calgui, i explicitau formalment els passos en el càlcul de probabilitats

5.- (3 punts) Si es reben una mitjana de 42 imatges cada setmana, calculeu la probabilitat que un dia es rebin 8 imatges

I quin és el temps esperat en hores entre arribades de dues imatges consecutives?

I calculeu el nombre d'hores amb que podem assegurar que no arribaran imatges amb una probabilitat de 0.6

6.-(2 punts) Si per cada 20 imatges que es reben, l'esperança del nombre d'imatges en **format jpeg** és 10, calculeu la probabilitat de trobar-ne 12 d'aquest format en 20 imatges

I la probabilitat d'haver de revisar 5 imatges fins observar la primera en format jpeg

7.- (1 punt) Dibuixeu de forma aproximada les 4 funcions de probabilitat o de densitat de cadascun dels models implicats en les preguntes 5 i 6 anteriors. Indiqueu el model i valor de paràmetre concret i senyaleu el valor de l'esperança en cada gràfic

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

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

### **Problema 3 Bloc C**

1. A partir d'una mostra de fitxers volem donar un interval de confiança per a l'esperança de la variable "temps de compressió", assumint distribució Normal. Si la mostra obtinguda és de mida 10 amb mitjana de 7.2, i considerem que la variància poblacional és 1.4, l'interval calculat al 95% és: (1 punt)

2. En dos programes de compressió diferents que tenen una durada amb distribució normal de  $\sigma_1 = 6$  segons i  $\sigma_2 = 4$  segons respectivament, realitzem una operació de mostreig amb 30 casos per al primer programa i  $n_2$  per al segon. Si volem estimar la diferència en la mitjana dels temps, quina ha de ser la mida de la mostra del 2n programa per tal que l'error tipus d'aquest estudi sigui, com a màxim, 1.25? Justifiqueu el procés de càlcul. (2 punts)

3. En 12 fitxers, la diferència  $D$  entre els temps d'execució de dos programes de compressió de fitxers ha estat de mitjana 3.1 segons i desviació 4. Acceptant que  $D \sim \text{Normal}$ , per saber com són de diferents, calculeu l'interval de confiança al 99%, i plantegeu si és raonable pensar que els dos compressors triguen el mateix en mitjana. (1 punt)

4. Escriu la definició, en paraules, de la variància de la v.a.  $X$  i la variància de la variable mitjana mostral. Eviteu posar només fórmules. Poseu un exemple de cadascun d'ells. (1 punt)

Estudiarem el temps de compressió mesurat en 30 documents i realitzem 3 càlculs amb els següents resultats:

A: t = -3.3853, df = 29, p-value = 0.002058 alternative hypothesis: true mean is not equal to 8 90 percent confidence interval: 6.758417 7.588250 sample estimates: mean of x 7.173333	B: t = -3.3853, df = 29, p-value = 0.002058 alternative hypothesis: true mean is not equal to 8 95 percent confidence interval: 6.673901 7.672766 sample estimates: mean of x 7.173333	C: t = -3.3853, df = 29, p-value = 0.002058 alternative hypothesis: true mean is not equal to 8 99 percent confidence interval: 6.500241 7.846426 sample estimates: mean of x 7.173333
--	--	--

5. Què són els valors t de l'output de R i perquè són iguals? (1 punt)

6. Com s'ha calculat el p-valor de la sortida A. Feu un gràfic i representeu -aproximadament- aquest p-valor. (1 punt)

7. S'obtenen 60 compressions realitzades, en 52 no es supera el llindar de 8 segons. Indiqueu l'estimació puntual i un interval de confiança al 90% pel percentatge de compressions que no superen el llindar. Justifiqueu com feu el càlcul i interpreteu el que heu trobat (2 punts)

8. En les mesures preses a les 60 compressions hem obtingut un valor  $s^2 = 2^2$ . Calculeu l'interval de confiança del 95% per a la desviació tipus del temps de compressió (0.5 punts), i l'interval de confiança del 90% per a la variància (0.5 punts).

#### VALORS Xi-quadrat

qchisq(0.025, 60)= 40.48175	qchisq(0.975, 60)= 83.29767	qchisq(0.95, 59)= 77.93052	qchisq(0.05, 59)= 42.33931
qchisq(0.05, 60)= 43.18796	qchisq(0.99, 59)= 87.16571	qchisq(0.95, 60)= 79.08194	qchisq(0.95, 59)= 77.93052
qchisq(0.1, 59)= 45.57695	qchisq(0.99, 60)= 88.37942	qchisq(0.95, 60)= 79.08194	qchisq(0.975, 59)=82.11741
qchisq(0.1, 60)= 46.45889	qchisq(0.025, 59)= 39.66186		

Problema bloc D

Ja hem treballat el tema de la qualitat dels vins, ara ens ocuparem d'un altre plaer gastronòmic: el pernil. Disposem d'una base de dades amb avaluacions de diferents experts de 250 marques de pernil: la qualitat Q és un número entre 0 i 10; el tipus T pot ser 1) bodega, 2) pinso (*recebo*), 3) gla (*bellota*); també es recull el pes P en kilograms de la peça, i finalment l'origen R (Ext: Extremadura, Alt: altre). Volem analitzar la relació d'aquests factors amb la qualitat del pernil.

Estimate Std. Error			A	Estimate Std. Error			B
(Intercept)	5.5192	0.1028		(Intercept)	4.9382	0.1326	
Residual standard error: 1.625 on 249 deg. of fr.				T2	0.6972	0.2121	
				T3	1.8785	0.2503	
				Residual standard error: 1.471 on 247 deg. of fr.			
				Multiple R-squared: 0.1876,			
				F-statistic: 28.52 on 2 and 247 DF, p-value: 7.168e-12			
Estimate Std. Error			C	Estimate Std. Error			D
(Intercept)	3.2959	0.9218		(Intercept)	3.29784	0.92351	
T2	0.6758	0.2114		T2	0.67926	0.21214	
T3	1.8989	0.2494		T3	1.90608	0.25105	
P	0.2342	0.1301		P	0.23744	0.13079	
Residual standard error: 1.464 on 246 deg. of fr.				RExt	-0.05607	0.18715	
Multiple R-squared: 0.1982,				Residual standard error: 1.467 on 245 deg. of fr.			
F-statistic: 20.27 on 3 and 246 DF, p-value:				Multiple R-squared: 0.1985,			
9.028e-12				F-statistic: 15.17 on 4 and 245 DF, p-value: 4.298e-11			

1. Expliqueu que hem obtingut en el output A: quin és el model, quines variables intervenen, què resultats hem de destacar (no cal que trobeu intervals de confiança si no es demanen, però utilitzeu alguna mesura de la incertesa de l'estimació quan sigui oportú). (1pt)
2. D'acord amb la sortida que veiem al output B, quins canvis s'han introduït al model? Com s'interpreten les noves estimacions? Són canvis rellevants, és a dir, poden explicar (part de) la valoració del pernil? Justifiqueu la resposta. (2pts)
3. Comenteu de manera semblant el que es veu al output C. Perquè ha canviat tant l'estimació del terme "Intercept"? (2pts)

4. Continueu amb l'analítica del output D. (1pt)

5. Si tenim dues peces del mateix pes i origen, però una és bodega i l'altra és pernil de gla, quina és la diferència de valoració mitjana entre ambdues peces? Responeu amb un (o varis) interval(s) de confiança al 95%. (1pt)

6. Podem dir que una peça més grossa té més possibilitats de rebre millors valoracions? I una peça d'Extremadura? Raoneu les respostes, i quantifiqueu els "efectes" de cadascú dels factors. (2pts)

7. Perquè al primer anàlisi no tenim un indicador "Multiple R-squared"? Què ens diu a les altres anàlisis? (1pt)