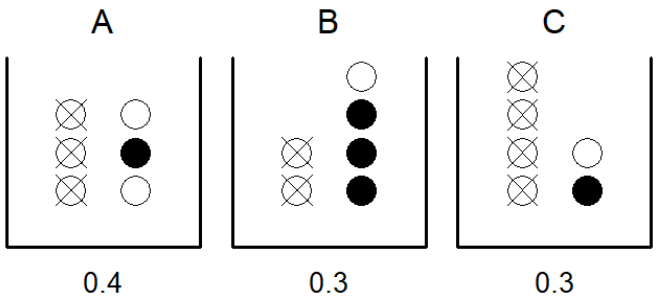
NOM:_____COGNOMS:_____

Problema 1 (Bloc A)

(*heu de justificar formalment el desenvolupament de les respostes*)

Tenim tres urnes A, B I C que contenen cadascuna sis monedes, com es veu a l'esquema de la dreta. Les monedes marcades amb ⊗ són monedes normals, amb la mateixa probabilitat per cara que per creu. El cercle blanc representa monedes trucades: la cara és 3 vegades *més* probable que la creu. El cercle negre també, però la cara és 3 vegades *menys* probable que la creu. Una urna es escollida a l'atzar, amb probabilitat 40% per a A, i les altres amb probabilitat 30% cadascuna. De l'urna escollida traiem una moneda (totes són indistingibles al tacte), que es llença a l'aire per observar un dels seus costats aleatòriament.



1. Representeu a la dreta l'arbre de esdeveniments i probabilitats de l'experiència descrita anteriorment. (1.5pt)

2. Poseu un exemple de *resultat* i un altre d'*esdeveniment*, clarament diferenciats, extrets de l'experiència anterior. (1pt)

3. Si l'urna escollida és la B, calculeu la probabilitat que de la moneda extreta observem cara. (1pt)

4. Quina és la probabilitat d'extreure una moneda trucada? (1pt)

5. El costat de la moneda que hem extret d'una urna mostra una *cara*. Trobeu la probabilitat que l'urna hagi estat la A. (1.5pt)

Suposeu que les monedes no trucades (⊗) són de 1€; les monedes (○) són de 50 cèntims, i les (●) són de 2€. Com abans, la mà que treu la moneda de l'urna seleccionada no és capaç de diferenciar unes monedes d'altres.

6. Especifiqueu la distribució de probabilitat de la variable aleatòria X "valor de la moneda"; representeu gràficament les funcions de probabilitat i de distribució. (1.5pt)

7. Trobeu el valor esperat de X i la seva desviació tipus. (1pt)

8. Demostreu utilitzant les probabilitats adients si la variable X és independent o no de si surt cara o creu amb la moneda extreta. (1.5pt)

NOM:_____COGNOMS:_____

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

Problema 2 (Bloc B)

A la imaginària ciutat de X, els delictes poden donar-se en qualsevol època de l'any, però són poc freqüents, i no relacionats entre ells. La policia estima que s'hi produeixen 9 delictes anualment en mitjana. L'alcalde vol conèixer com de tranquil es pot considerar a X un mes qualsevol (per simplificar, preneu que els 12 mesos d'un any tenen 30 dies).

1. Sigui D el nombre de delictes comesos a X en un any. Quin model de probabilitat és adequat per a D, i perquè? Especifiqueu tots els paràmetres que siguin necessaris. (1pt)

2. Sigui R el nombre de mesos d'un període d'un any en els que la policia de X ha registrat algun delicte. Determineu el model de probabilitat adient a R, amb tots els paràmetres necessaris. Calculeu el nombre esperat de mesos *conflictius* al cap d'un any. (1.5pts)

3. Calculeu la probabilitat que en un any tinguem un màxim de tres delictes a X. (1pt)

4. Calculeu la probabilitat que en un any tinguem un màxim de tres mesos amb algun delicte. (1pt)

El trajecte recomanat per una aplicació per anar de la ciutat de X a la ciutat de Y consisteix en: 1) F_1, agafar un tren, que suposa un temps amb distribució U[100, 200]; 2) F_2, caminar, que suposa un temps amb distribució N(100, 25); 3) F_3, agafar un autobús, que suposa un temps amb distribució Exp(1/60). Tots els temps són minuts. Per qüestions diverses, els temps del tren i el autobús estan correlacionats amb ρ=0.5; caminar és independent dels altres mitjans.

5. Es demana que trobeu el temps esperat del trajecte complet, així com la seva desviació tipus, detallant el desenvolupament necessari. (2pts)

6. La distribució de T és una Normal, o aproximadament Normal? Justificar la resposta. Tant si la resposta és sí o si és no, suposeu-ho cert per trobar la probabilitat que un viatge de X a Y es pugui fer en menys de quatre hores. (1.5pts)

7. Un viatger fa aquest trajecte 25 vegades en un any. Un dia, aquest individu vol calcular quant temps passa en total a l'any en el tren (comptant només els viatges d'aquests trajectes). Com el temps de cada viatge és aleatori, busca un interval centrat a la mitjana que contengui amb probabilitat 0.80 el temps *en hores* total anual dels seus viatges en tren. Calculeu-ho i expliqueu tot el procediment. (2pts)

```
> p = seq(0.60, 0.99, by=0.01)
> round(qnorm(p), 4)
 [1] 0.2533 0.2793 0.3055 0.3319 0.3585 0.3853 0.4125 0.4399 0.4677 0.4959
[11] 0.5244 0.5534 0.5828 0.6128 0.6433 0.6745 0.7063 0.7388 0.7722 0.8064
[21] 0.8416 0.8779 0.9154 0.9542 0.9945 1.0364 1.0803 1.1264 1.1750 1.2265
[31] 1.2816 1.3408 1.4051 1.4758 1.5548 1.6449 1.7507 1.8808 2.0537 2.3263
```

17-06-2026

NOM: _____ COGNOM: _____

Contesteu cada pregunta en els espais en blanc. Expliqueu i justifiqueu els càlculs.

BLOC C. Una empresa està desenvolupant un sistema basat en intel·ligència artificial per respondre preguntes sobre documentació jurídica. Abans de desplegar-lo, es duen a terme diversos estudis per avaluar-ne el rendiment.

Es realitza una primera prova pilot on es selecciona una mostra aleatòria de 10 documents jurídics i es calcula el percentatge de respostes correctes del sistema en cadascun d'ells.

Document	1	2	3	4	5	6	7	8	9	10
Precisió (%)	78	82	75	89	84	77	86	80	83	78

$\sum x_i = 812$

$\sum x_i^2 = 66108$

1. (1.5 pts.) Calcula els estimadors puntuals de la mitjana poblacional i de la desviació estàndard poblacional.

2. (1.5 pts.) Calcula un interval de confiança del 95% per a la precisió mitjana del sistema.

Després de l'estudi pilot, l'empresa decideix ampliar la mostra. El nou interval de confiança del 95% per a la precisió mitjana del sistema és:

$$IC_{95\%}(\mu) = [79.2,\ 83.2]$$

3. (1.5 pts.) Suposa que en el nou estudi s'assumeix coneguda la desviació estàndard poblacional i que aquesta coincideix amb la desviació estàndard estimada en l'estudi pilot. Determina la mida mostral utilitzada en aquest nou estudi.

Per analitzar si el comportament del sistema és igual d'estable en diferents entorns, es prova en dos departaments de l'empresa. Les puntuacions obtingudes són les següents:

	1	2	3	4	5	6	7	8	9	10
Dept. A	82	76	85	79	88	81	74	83	80	86
Dept. B	91	72	84	69	95	78	88	73	81	98

$\sum x_A = 814, \sum x_A^2 = 66332$

$\sum x_B = 829, \sum x_B^2 = 69629$

4. (2 pts.) Calcula un interval de confiança del 99% per al quocient de variàncies σ_A^2/σ_B^2.

5. (1 pt.) A partir de l'interval obtingut, és raonable assumir que ambdós departaments tenen la mateixa variabilitat? Justifica la resposta.

Finalment, l'empresa compara dues versions del sistema d'IA. Per a cadascuna d'elles es registra si la resposta a una pregunta és correcta o incorrecta. L'anàlisi realitzada amb R ha produït el següent resultat:

```
2-sample test for equality of proportions without continuity correction

data:  x out of n
X-squared = 2.5606, df = 1, p-value = 0.1096
alternative hypothesis: two.sided

95 percent confidence interval:
 -0.0779 0.0079

sample estimates:
   prop 1    prop 2
    0.840     0.875
```

6. (1 pt.) Quin paràmetre està estimant l'interval de confiança anterior?

7. (1.5 pts.) Interpreta l'interval de confiança en el context del problema.

$qnorm(0.80) = 0.8416$	$qnorm(0.90) = 1.2816$	$qt(0.90, 9) = 1.383$	$qt(0.95, 9) = 1.833$	$qf(0.95, 9, 9) = 3.18$
$qnorm(0.95) = 1.6449$	$qnorm(0.975) = 1.9600$	$qt(0.975, 9) = 2.262$	$qt(0.99, 9) = 2.821$	$qf(0.995, 9, 9) = 6.54$
$qnorm(0.99) = 2.3263$	$qnorm(0.995) = 2.5758$	$qt(0.995, 9) = 3.250$	$qt(0.999, 9) = 4.297$	$qf(0.005, 9, 9) = 0.153$

Nom i cognoms:
Contesteu cada pregunta en els espais en blanc. Expliqueu i justifiqueu els càlculs

BLOC D. Un grup d'estudiants de l'UPC estan estudiant diverses formes per comprimir eficientment fitxers GIS (dades geoespecials). En concret, han proposat 2 mètodes: A i B.
Per veure quin dels dos mètodes funciona millor, han decidit realitzar un estudi aparellat:

$$D = \%\text{ comprimit mètode A} - \%\text{ comprimit mètode B}$$

Utilitzant fitxers escollits de forma equiprobable i aleatòria. D'aquests fitxers sabem el tipus de terreny que descriuen: Ciutat, Camp o Muntanya, i la mida del fitxer (MB).
Inicialment, han proposat aquest model:

```
Call:
lm(formula = D ~ 1, data = dades)
Residuals:
     Min       1Q   Median       3Q      Max
-12.9333  -3.6933  -0.2383   4.0142  13.3367
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.5733     0.4460   1.285    0.201
Residual standard error: 5.463 on 149 degrees of freedom
```

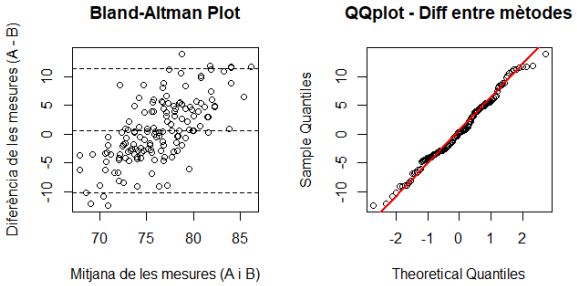


Figura 3

1 Podem afirmar estadísticament que existeix un mètode millor que l'altre? Indiqueu també l'interval del valor de la diferència del percentatge comprimit entre el mètode A i B amb un 95% de confiança. (1.5 punts)

2 Llisteu les premisses que assumeix el model. (1 punt)

3 Donats els gràfics anteriors, penseu que aquestes premisses es compleixen? Argumenteu la resposta. (1.5 punts)

Donat el següent fragment de codi: (El model 2 es descriu a la següent pàgina)

```
nova_obs <- data.frame( Entorn = "Muntanya", Mida_fitxer_MB = 667)
predict(model2, newdata = nova_obs, interval = "confidence")
predict(model2, newdata = nova_obs, interval = "prediction")
```

4 Quina és la diferència entre aquestes dues comandes? Què podem esperar dels resultats? (1.5 punts)

A continuació, els estudiants han decidit afegir més variables explicatives al model:

```
Call:
lm(formula = D ~ Entorn + Mida_fitxer_MB, data = dades)

Residuals:
   Min 1Q Median 3Q Max
-6.022 -1.761 -0.060 1.735 6.752

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.7608680 0.5384028 -16.272 <2e-16 ***
EntornCiutat -0.5208539 0.5368944 -0.970 0.334
EntornMuntanya 0.8015401 0.5295355 1.514 0.132
Mida_fitxer_MB 0.0168833 0.0007702 21.920 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
0.05 '.' 0.1 ' ' 1

Residual standard error: 2.647 on 146 degrees of freedom
Multiple R-squared: 0.77
Adjusted R-squared: 0.7652
F-statistic: 162.9 on 3 and 146 DF
p-value: < 2.2e-16
```

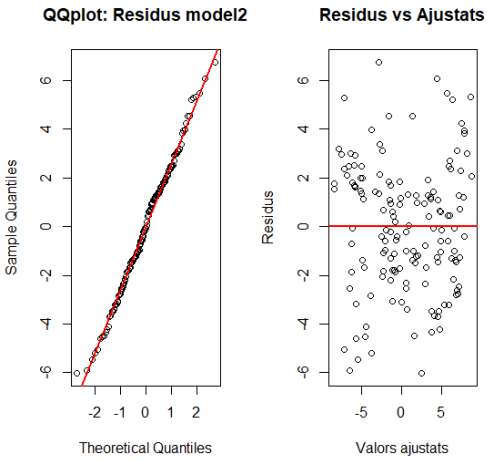


Figura 4

5 Es pot calcular el percentatge de compressió esperat del mètode A per un fitxer que descriu una Ciutat i té una mida de 426 MB únicament amb la informació proporcionada pel model? (1 punt)

6 Interpreteu el coeficient que fa referència a la mida del fitxer, com afecta en el valor final de D? Relacioneu-lo amb el gràfic Bland-Altam del primer model (1.5 punts)

7 Quines premisses està assumint el model? Donats els gràfics i la informació disponible, creieu que es compleixen? (2 punts)