

NOM: _____ COGNOM: _____

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

Problema 1 (B1 B2)

Volem estudiar les probabilitats i el que es pot guanyar en un joc que consisteix en llençar una moneda desequilibrada (creu surt el doble de cops de cara), un màxim de tres intents, seguint els següents passos: si surt cara en el primer intent es guanyen 100 euros i acaba el joc, i si surt creu hi ha un segon intent; si surt cara en el segon intent es guanya la meitat (i acaba el joc) i si surt creu hi ha un tercer intent; si surt cara en el tercer es guanya una tercera part i, per tant, si surten tres creus no es guanya res.

1.- (1 punt) Dibuixeu l'arbre de l'experiència aleatòria i expliqueu el conjunt de resultats i les seves probabilitats.

2.- (1 punt) Calculeu (expressant-ho formalment) la probabilitat de guanyar el màxim i la de no guanyar res.

3.- (1 punt) Calculeu (expressant-ho formalment) la probabilitat de guanyar alguna cosa.

4.- (1 punt) Quina és la probabilitat de guanyar alguna cosa si a la primera tirada ha sortit creu?

5.- (1 punt) Quines haurien de ser les probabilitats de cara i creu d'una moneda amb la qual la probabilitat de no guanyar res fos del 50%?

6.- (2 punts) Definiu la variable aleatòria que indica les probabilitats de les diferents quantitats a guanyar (incloent no guanyar res), i calculeu-ne l'esperança i la desviació.

7.- (3 punts) Considereu ara una altra variable aleatòria indicant el nombre de creus obtingudes en l'experiència aleatòria anterior. Indiqueu la taula de probabilitat d'aquesta nova variable, i la seva esperança i desviació. I també la taula de probabilitats conjuntes (d'aquesta variable nombre de creus i l'anterior de guanyats), i calculeu la covariància i la correlació, i comenteu la relació entre les dues variables relacionant-ho amb el valor de la correlació entre elles.

Un determinat algorisme consisteix en N etapes que cal superar satisfactòriament. Com es veu al pseudocodi, l'etapa i -èsima es computa a la funció `proces(i)`, que retorna un indicador d'estat Z , que es avalua a `condicio(Z)` per determinar si l'etapa s'ha superat. La funció `proces` fa uns càlculs molt complexos, i l'etapa no es supera el 40% de les vegades que s'executa, obligant a repetir-ho provant una altra estratègia.

```
for (i=0; i<N; i++) {  
    repeat {  
        Z=proces(i);  
    } until (condicio(Z));  
}
```

Quan la funció `proces` s'ha d'executar novament a la mateixa etapa es genera un indicador independent de l'anterior. Les etapes diferents també generen indicadors independents.

- [1pt] Fixem-nos en una etapa i fixa. Anomenem B_i a la variable aleatòria que compta el nombre de cicles del bucle `repeat`. Quin és el model probabilístic per a aquesta variable, i perquè? Expressen la funció de probabilitat de B_i i diguen quin és el nombre esperat de vegades que la funció `proces` serà emprada.
- [1pt] Calculeu la probabilitat que a l'etapa i la funció `condicio` sigui avaluada almenys 4 vegades.
- [1pt] Volem estudiar el comportament de l'algorisme, concretament a les etapes 5, 10, 15, 20, 25, 30, 35 i 40. Quin és el nombre esperat d'etapes on el nombre de cicles serà superior a 3? Trobeu la probabilitat que observem com a molt a una etapa aquest fet (més de 3 cicles). *Ajut: definiu una variable aleatòria adient, especificant el seu model.*
- [1pt] Suposem que l'algorisme precisa exactament $N=100$ etapes. A) Diguen quin és el model adequat per descriure el nombre total d'execucions de la funció `proces` (amb els seus paràmetres); B) Proposeu un model que pugui aproximar raonablement el model exacte, i expliqueu perquè es pot suposar raonable.

5. [1.5pt] Amb el model aproximat anterior, trobeu una fita superior pel nombre de vegades que s'executa `proces` i que sigui vàlida amb un error de l'1%. Per a $N=200$, la fita seria el doble d'aquesta? Justifiqueu-ho.
6. [2pt] Hem afegit una variable per comptar el nombre de vegades que s'executa `proces` i hem fet córrer l'algorisme, obtenint que han calgut 145 avaluacions, amb $N=100$.
- Amb aquestes dades, doneu una estimació puntual per a la probabilitat que la funció `proces` superi l'etapa corresponent.
 - Quin és l'error tipus de l'estimació anterior? Expliqueu el significat de l'error tipus trobat.
 - Estimeu amb un interval de confiança del 95% la probabilitat que una avaluació de la funció `condicio` retorni CERT.
 - Expliqueu com s'interpreta l'interval que heu trobat a l'apartat anterior, i si és compatible amb les assumpcions inicials del problema.

També hem incorporat al programa anterior codi que obté el temps de CPU emprat per l'ordinador a la funció `proces`:

```
for (i=j=s=sq=0; i<N; i++) {
  repeat {
    t = CPU.time( Z=proces(i) );
    s += t;
    sq += t*t;
    j++; // el comptador d'iteracions
  } until (condicio(Z));
}
```

A més del resultat $j=145$, també hem obtingut $s=1125$ (segons) i $sq=11600$ (s^2).

7. [1.5pt] Doneu una estimació puntual i per interval de confiança del 95% per al temps que en mitjana empra l'execució de la funció `proces`.
8. [1pt] Sabent que el temps de CPU no pot ser negatiu, raoneu si es pot admetre que la variable recollida (`CPU.time`) es distribueix com a Normal, i també perquè (o perquè no) es necessita Normalitat de la variable per a calcular intervals de confiança per al valor esperat.

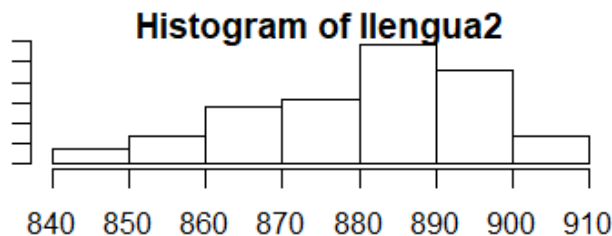
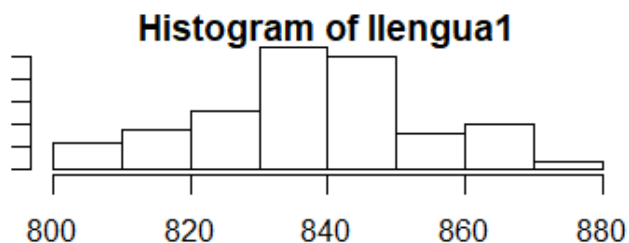
NOM: _____ COGNOM: _____
(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs)

Problema 3 (B5-B6)

Un grup d'estudiants de la UPC volen fer un estudi per comparar si hi ha diferència en l'ús de les vocals entre el català i el castellà. Per fer-ho cerquen de manera aleatòria en el repositori de la UPC 100 documents en cadascun dels dos idiomes. Per cada document, compten el nombre de vocals que hi ha en els primers 2000 caràcters sense tenir en compte els espais. Amb aquesta resposta es defineixen les corresponents variables aleatòries: T pels documents en català i S pels documents en castellà, obtenint-se les dades següents:

$$\bar{T} = 837.4 \text{ i } s_T = 16.5523$$

$$\bar{S} = 881.28 \text{ i } s_S = 15.3149$$



1. Indiqueu quin és l'histograma corresponent a T i a S. Raoneu la resposta. (1 punt)

2. Es tracta de dues mostres independents o aparellades? Raoneu la resposta (1 punt)

Es vol estudiar si les dues llengües usen el mateix nombre de vocals o si és el castellà el qual n'usa més. Per contestar els següents apartats considereu un risc $\alpha=0.05$ i una confiança del 95%.

3. a) Indiqueu les hipòtesis i les premisses de la prova a realitzar. (0'5 punts)

b) Indiqueu la fórmula de l'estadístic i quina és la distribució d'aquest quan se suposa que les dues respostes tenen la mateixa mitjana (0'5 punts)

c) Calculeu l'estadístic i raoneu si podem rebutjar la hipòtesi nul·la. (1 punt)

També es vol estudiar si hi ha relació entre el nombre de vocals emprades en ambdues llengües. Per això se seleccionen aleatòriament textos de 2000 caràcters sense comptar espais de 20 textos en català i es tradueixen al castellà mitjançant un traductor automàtic. Anomenant les dues variables aleatòries *Nombre de vocals en català* (T) i *Nombre de vocals en castellà* (S), es té que:

```
> mean(T)
[1] 837.55
> mean(S)
[1] 879.4
```

```
> sd(T)
[1] 16.58304
> sd(S)
[1] 19.47846
```

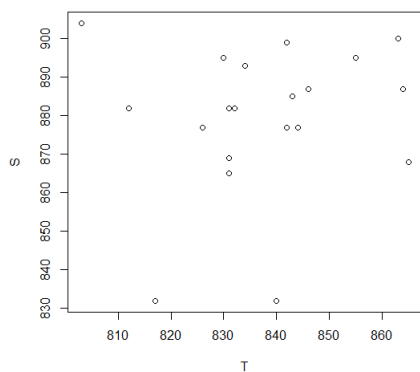
```
> sum(T)
[1] 16751
> sum(S)
[1] 17588
> sum(T*S)
[1] 14731604
```

4. a) Calculeu la covariància i el coeficient de correlació (1 pt)

b) Interpreteu el coeficient de correlació lineal (0'5 punts)

5.a) Calculeu la recta de regressió del Nombre de vocals en castellà (S) respecte el Nombre de vocals en català (T) (1 pt)

b) Representeu la recta de regressió en el gràfic següent. Expliciteu la representació de manera raonada. (1 pt)



c) És versemblant que el coeficient del pendent sigui zero? Realitzeu la prova d'inferència estadística apropiada per donar resposta a la pregunta. (1'5 punts)

d) Calculeu l'interval de confiança pel pendent i interpreta'l. (1 punt)