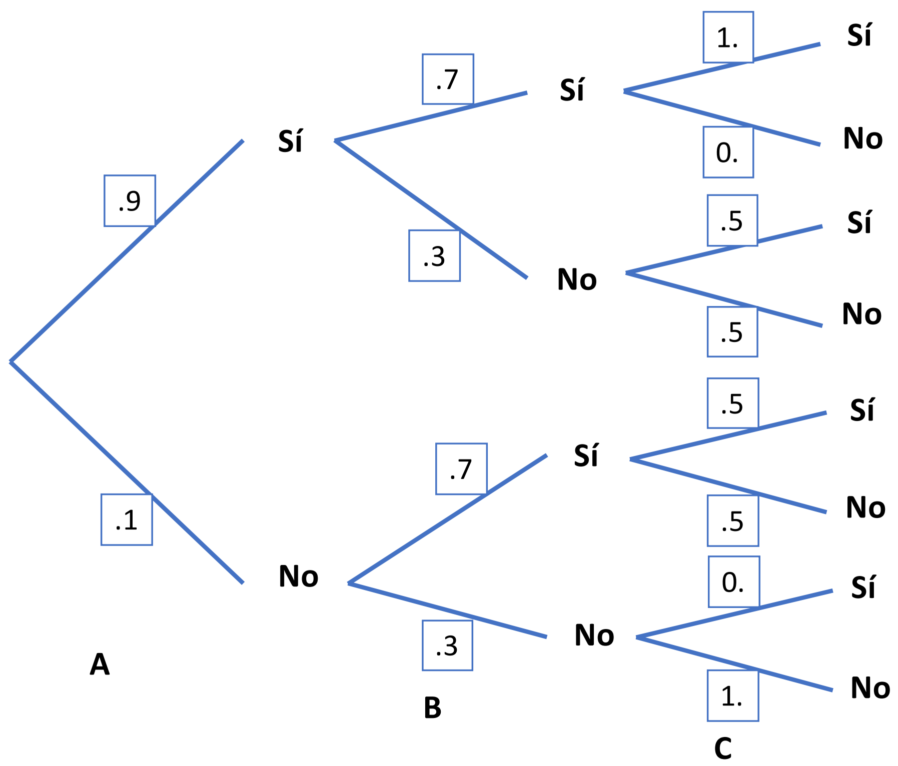
NOM:________________________COGNOM:____ 02/06/2025

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

Problema 1 (Bloc A)

Un jurat popular consta de tres membres, A, B i C. El membre A és una persona experta i, quan pren una decisió sobre el tema en el qual el jurat és competent, en el 90% de les vegades es tracta de la decisió correcta (“Sí”). El membre B és totalment independent del membre A, una mica menys expert, i els seus encerts es limiten a un 70%. Per últim, el membre C és més insegur: si veu que els seus companys hi coincideixen (tant per “Sí” com per “No”), s’hi afegeix a la mateixa decisió; i si veu que no hi ha acord entre A i B, pren una decisió amb una moneda equilibrada. La decisió final del jurat és per majoria: almenys dos vots en un sentit concret (que tant pot ser una decisió correcta com incorrecta).

1. Dibuixar l’arbre dels esdeveniments i probabilitats corresponent. Afegiu a cada branca terminal una indicació de quina seria la decisió del jurat en cada cas.



w1: 3 a favor (correcta)

w2: 2 a favor, 1 en contra (corr.)

w3: 2 a favor, 1 en contra (corr.)

w4: 1 a favor, 2 en contra (incorr.)

w5: 2 a favor, 1 en contra (corr.)

w6: 1 a favor, 2 en contra (incorr.)

w7: 1 a favor, 2 en contra (incorr.)

w8: 3 en contra (incorrecta)

2. Trobeu la probabilitat que la decisió final del jurat sigui la correcta. Utilitzeu un desenvolupament formal de les expressions de la probabilitat.

P(decisió correcta) = P({w1, w2, w3, w5}) = P(w1) + P(w2) + P(w3) + P(w5) =

= P(C=Sí | B=Sí ∩ A=Sí) P(B=Sí | A=Sí) P(A=Sí) + P(C=No | B=Sí ∩ A=Sí) P(B=Sí | A=Sí) P(A=Sí) +

+ P(C=Sí | B=No ∩ A=Sí) P(B=No | A=Sí) P(A=Sí) + P(C=Sí | B=Sí ∩ A=No) P(B=Sí | A=No) P(A=No) =

= 1 · 0.7 · 0.9 + 0 · 0.7 · 0.9 + 0.5 · 0.3 · 0.9 + 0.5 · 0.7 · 0.1 = 0.63 + 0.135 + 0.035 = **0.80**

Imagineu que a un altre jurat molt similar, amb les mateixes probabilitats anteriors, els membres A i B no se sap si són independents o no, però sí es coneix que un 8% de les vegades prenen simultàniament la decisió incorrecta (“No”).

3. Justifiqueu si es pot dir que A i B són independents o que no ho són.

Ens diuen que P(A=No ∩ B=No) = 0.08. Si fossin independents, P(A=No ∩ B=No) = P(A=No) P(B=No)

Com coneixem les probabilitats marginals del jutge A i del B perquè són les mateixes, P(A=No) = 0.1 i P(B=No)=0.3. Però aquest producte és igual a 0.03. diferent de 0.08 i, per tant, els dos jutges del segon jurat no són independents.

4. Ha canviat en aquest jurat la probabilitat de una decisió final correcta? Raoneu la resposta amb els càlculs adients desenvolupats formalment.

Les probabilitats que són diferents són les del jutge B; les del jutge A no han variat i les del jutge C tampoc perquè els seus criteris continuen sent els mateixos. Hem d’obtenir P(B=Sí | A=Sí), P(B=No | A=Sí), P(B=Sí | A=No) i P(B=No |

A=No). Sabem que P(A=No ∩ B=No) = 0.08 = P(B=No | A=No) P(A=No). Per tant, P(B=No | A=No) = 0.08/0.1 = 0.8, i P(B=Si | A=No) = 1–0.8 = 0.2. D'altra banda, P(B=No) = 0.3 = P(A=No ∩ B=No) + P(A=Sí ∩ B=No), d'on traiem que P(A=Sí ∩ B=No) = 0.3 – 0.08 = 0.22 = P(B=No | A=Sí) P(A=Sí); finalment, les altres probabilitats condicionades de B són: P(B=No | A=Sí) = 0.22 / 0.9 = 0.2444... i P(B=Sí | A=Sí) = 1–0.2444... = 0.7555...

Amb aquests nous resultats podem refer el càlcul d'una decisió correcta:

P(decisió correcta) = P(w1) + P(w2) + P(w3) + P(w5) =

1 · 0.7555... · 0.9 + 0 · 0.2444... · 0.9 + 0.5 · 0.2444... · 0.9 + 0.5 · 0.2 · 0.1 = **0.8**

La probabilitat no ha canviat.

Imaginem una aplicació que utilitza un byte (8 bits), però només un dels bits està a 1, i la resta valen 0. El bit activat (que val 1) és aleatori, i el valor del byte és una variable aleatòria X d'acord amb la següent funció de probabilitat:

$$P(X = 2^i) = (8 - i)\,k, \qquad i = 0, \ldots, 7$$

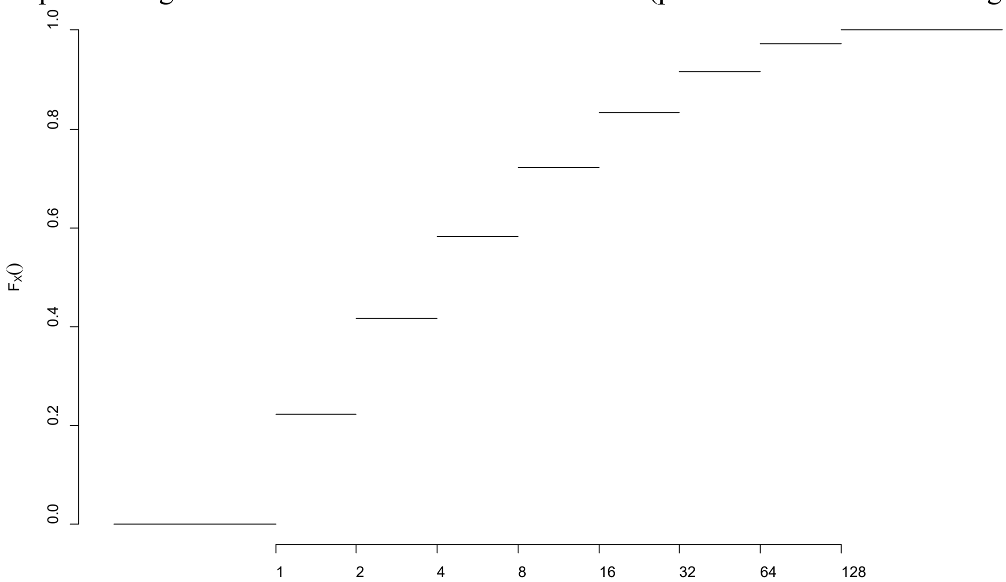
5. Deduïu el valor de la constant k, i trobeu el valor esperat de la variable X.

La variable pren vuit valors: 1, 2, 4, 8, 16, 32, 64 i 128, amb probabilitats: $8k$, $7k$, ..., k. Com la suma de totes les probabilitats ha de donar 1, tenim que:

$(8 + 7 + ... + 1)k = 36k = 1$, és a dir, $k=1/36$

$E(X) = \sum_{i=0}^{7} 2^i\,(8-i)/36$, que es pot fer amb la calculadora simplement, i resulta en: E(X) = **13.9444...**

6. Representeu gràficament la funció de distribució de X (podeu utilitzar una escala logarítmica pel eix d'abscisses).



I	F2
-2	0.0000000
0	0.2222222
1	0.4166667
2	0.5833333
3	0.7222222
4	0.8333333
5	0.9166667
6	0.9722222
7	1.0000000
9	1.0000000

7. Quina és la probabilitat que la variable X prengui un valor entre 20 i 50?

$P(20 \le X \le 50) = P(X = 32) = 3/36 = 1/12$

8. X' és una variable aleatòria amb la mateixa distribució que X, i ambdues són independents. Calculeu la probabilitat que X + X' sigui igual a 33. Expliqueu totes les passes del càlcul.

P(X + X' = 33) = P({X = 32 ∩ X' = 1} ∪ {X = 1 ∩ X' = 32}) = 2 · 3/36 · 8/36, ja que X i X' són independents i les dos parts separades per la ∪ tenen la mateixa probabilitat.

Resultat: P(X + X' = 33) = **1/27** = 0.0370...

NOM:____________________________COGNOM:____

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

Problema B

Els bots i els comptes falsos són eines clau en la difusió de *fake news* per grups organitzats. Aquests comptes automatitzats amplifiquen missatges falsos mitjançant interaccions massives, com ara likes, comparticions i comentaris. Un equip dissenya una campanya de fake news enviant de mitjana 1 missatge per bot cada quinze minuts segons una distribució de Poisson, utilitzant 5 bots independents per accelerar el procés.

1. Quina és la probabilitat que en una hora concreta no s'hagin superat els 10 missatges? (1 punt)

   X: num missatges per hora per bot ~ Poiss(4)

   X_5 : num missatges per hora 5 bots ~ Poiss(20)

   $P(X_5 < 10) = 0.0108$

2. Quina és la probabilitat que passin més de 5 minuts entre missatges enviats per tots 5 bots (per qualsevol d'ells)? (1 punt)

   T_5 : temps entre missatges per hora 5 bots ~ Exp(20)

   $P(T_5 > 5/60) = 0.189$

3. Després d'una hora de campanya s'han superat els 20 missatges. Quina és la probabilitat d'arribar als 40 amb els missatges enviats en la següent hora? (1 punt)

   $P(X_5 > 40\text{-}20) = 1 – F(20) = 1 – 0.559 = 0.441$

Per facilitar que els *bulos* es viralitzin, l'equip programa els bots per també amplificar fake news generant interaccions com likes o comentaris. Durant una hora el bot revisa 1000 publicacions i, cada vegada que es troba amb un contingut objectiu, fa un like a la publicació amb probabilitat 0.05.

4. Quina és la mitjana de publicacions revisades per un bot fins fer el primer like? Relaciona la resposta amb un model probabilístic apropiat. (1 punt)

   R = nombre de publicacions revisades per un bot ~ Geom(p=0.05)

   E(R) = 1/p = 20

5. Quina és la variància del nombre de likes que farien els 5 bots en una hora de funcionament? Quina és la probabilitat que entre els 5 facin entre 250 i 300 likes durant una hora? Expliqueu el procediment aplicat per arribar a la solució. (2 punts)

   L = nombre de likes durant una hora de funcionament ~ Binomial(5000, 0.05)

   Var(L) = npq = 5000 * 0.05 * 0.95 = 237.5

   E(L) =250

   $L \approx N(250, \sqrt{(5000\ p\ (1\text{-}p))}) = N(250, 15.411)$ ... $P(250 < L < 300) = F_L(300) – 0.5 = F_Z(3.24) – 0.5 = 0.4994$

En una xarxa social podem distingir entre dos tipus d'usuari. Un usuari estàndard interactua amb contingut divers i variat, fa de mitjana 15 likes per hora amb desviació tipus 10. En canvi, un usuari *echo chamber*, centrat principalment en continguts que reforcen les seves idees prèvies, en fa de mitjana 20 per hora amb desviació 5.

6. A una població amb 70 usuaris estàndard i 30 usuaris *echo chamber*, quina és la probabilitat de superar conjuntament els 1500 likes en una hora? (2 punts)

   TCL Normal

   Z: nombre de likes usuari estandard on E(Z) = 15, sd(Z) = 10

   Z1+...+Z70: nombre likes usuaris estandard ~ N(15*70, 10*sqrt(70))

   W: nombre de likes usuari echo chamber on E(W) = 20, sd(Z) = 5

   W1+...+W30: nombre likes usuaris echo chamber ~ N(20*30, 5*sqrt(30))

U: Nombre likes tots usuaris ~ N(15*70 + 20*30, sqrt(10^2*70 + 5^2*30)) = N(1650, 88.03)

P(U> 1500) = 1 – F_Z(-1.70) = F_Z(1.70) = 0.9554

7. Suposant una població amb 100 usuaris i suposant desviació 10 a la probabilitat de fer like per qualsevol usuari (estàndard o *echo chamber*), quina és la proporció d'usuaris estàndard i usuaris *echo chamber* si hem determinat (d'alguna manera) que amb probabilitat 95% el nombre total de likes per hora a la población és inferior a 1930? (2 punts)

   Denotem per p la proporció d'usuaris estàndard, ara U ~N(m,s) amb

   m=15*p*100 + 20*(1-p)*100 = 100(20-5p)

   s=10*sqrt(100)

   Volem P(Z<zalpha)=0.95 on Z= (1930-m)/s i zalpha=1.645, aïllem i obtenim m=1765.5 i p=0.47, per tant a la població hi ha 47 usuaris estàndard i 53 usuaris *echo chamber*.

ppois(20,20)= 0.55909	ppois(10,20)= 0.01081	dpois(20,20)= 0.08884	dpois(10,20)= 0.0058163
ppois(40,20)= 0.99997	ppois(10,4)= 0.99716	dpois(40,20)= 2.78e-05	dpois(10,4)= 0.0052925
ppois(30,20)= 0.98653	ppois(5,4)= 0.78513	dpois(30,20)= 0.0083435	dpois(5,4)= 0.15629
pnorm(c(1.2,1.3,…,2.1))= 0.8849 0.9032 0.9192 0.9332 0.9452 0.9554 0.9641 0.9713 0.9772 0.9821 0.9861 0.9893			
pnorm(c(2.8,2.85,…,3.3))= 0.9974 0.9978 0.9981 0.9984 0.9987 0.9989 0.9990 0.9992 0.9993 0.9994 0.9995			

NOM:___________________________COGNOM:____

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

Problema C

Es vols estudiar el temps de càrrega de pàgines web usant o no un bloquejador d'anuncis. Per una part, usant un primer navegador, es tenen els temps en segons d'una mostra aleatòria de 30 pàgines seleccionades d'acord a certs criteris d'inclusió. TN són els temps *no* usant el bloquejador, i TS *sí* usant-lo, i $D1=TN\text{-}TS$

$\sum_{i=1}^{30} TN_i = 73.17$ $\quad\sum_{i=1}^{30} TN_i^2 = 188.57$ $\quad\sum_{i=1}^{30} TS_i = 72.33$ $\quad\sum_{i=1}^{30} TS_i^2 = 184.35$ $\quad\sum_{i=1}^{30} D1_i = 0.85$ $\quad\sum_{i=1}^{30} D1_i^2 = 0.04$

Per a un segon navegador i per a una altra mostra de pàgines diferents però de característiques equivalents a les anteriors, es tenen també 30 diferències de temps (D2) entre no usar el bloquejador i usant-lo en aquest segon navegador

$$\sum_{i=1}^{30} D2_i = 1.115 \qquad \sum_{i=1}^{30} D2_i^2 = 0.05$$

1.- (2 punts) Per una part volem, només pel primer navegador, comparar les mitjanes de temps usant o no bloquejador. Per altra part volem comparar les mitjanes de la diferència (d'usar o no bloquejador) entre els dos navegadors. Indiqueu i justifiqueu si cadascuna d'aquestes dues proves de comparació de mitjanes son de mostres aparellades o independents. Indiqueu en cada cas qui seria la variable de resposta que caldria usar

Per comparar les mitjanes de temps usant o no bloquejador usarem mostres aparellades ja que tenim els 30 temps per a unes mateixes pàgines. Caldria usar els valors de D1 que són els 30 valors de les diferències de temps

Per comparar les mitjanes de la diferència (usant o no bloquejador) entre navegadors usarem mostres independents ja que tenim 30 diferències en cada navegador per unes pàgines diferents però de característiques equivalents. Caldria usar els valors de D1 i de D2 que són els 30 valors de les diferències de temps per a cada navegador

2.- (2 punts) Pel cas del primer navegador, calculeu un interval de confiança del 95% per quantificar la diferència mitjana usant o no el bloquejador. Indiqueu l'estimació de la diferència, de la seva desviació, de l'error tipus i l'IC. Interpreteu l'IC i la seva conclusió

```
mean(D1) = 0.85/30 = 0.028        sd(D1) = sqrt((0.04-((0.85^2)/30))/29) = 0.024
                                  se_D1 <- sd(D1)/sqrt(30)    0.004

0.028 - qt(0.975,29) * 0.004 = 0.028-2.045*0.004 = 0.02
0.028 + qt(0.975,29)* 0.004 = 0.028-2.045*0.004 = 0.04
```

Amb 95% de confiança la **diferència mitjana** de temps usant o no el bloquejador està entre 0.02 i 0.04 segons
No usar bloquejador d'anuncis fa l'accés més lent entre 0.02 i 0.04 segons amb confiança 95%

3.- (2 punts) Per comparar si la diferència usant o no bloquejador entre els dos navegadors és equivalent o no, calculeu un interval de confiança al 95% i interpreteu-lo amb la seva conclusió

```
mean(D1) =  0.028                 sd(D1) = sqrt((0.04-((0.85^2)/30))/29) = 0.024
mean(D2) = 1.115/30 = 0.037  sd(D2) = sqrt((0.05-((1.115^2)/30))/29) = 0.017

s_pool <- sqrt((0.024*0.024*29+0.017*0.017*29)/60) = 0.02
se <- 0.02*sqrt((1/30)+(1/30)) = 0.005

(0.028-0.037)-qt(0.975,58)*0.005 = -0.009-2.002*0.005 = -0.02
(0.028-0.037)+qt(0.975,58)*0.005 = -0.009+2.002*0.005 = 0.001
```

Amb 95% de confiança la diferència mitjana de temps entre navegadors està entre -0.02 i 0.001
Amb 95% de confiança la diferència de temps usant o no bloquejadors va des de 0.02 a favor d'un navegador fins a 0.001 a favor de l'altre, per tant no hi ha evidència de diferència entre navegadors

4.- (2 punts) Per a cadascun dels dos navegadors hem anotat quantes de les 30 pàgines tenen *banners* estàtics: 12 pel primer navegador i 24 pel segon. Indiqueu l'estimació de la diferència de proporcions amb l'error tipus i interpreteu-los. Calculeu un interval de confiança al 95% de la diferència de proporcions entre els navegadors i interpreteu-lo

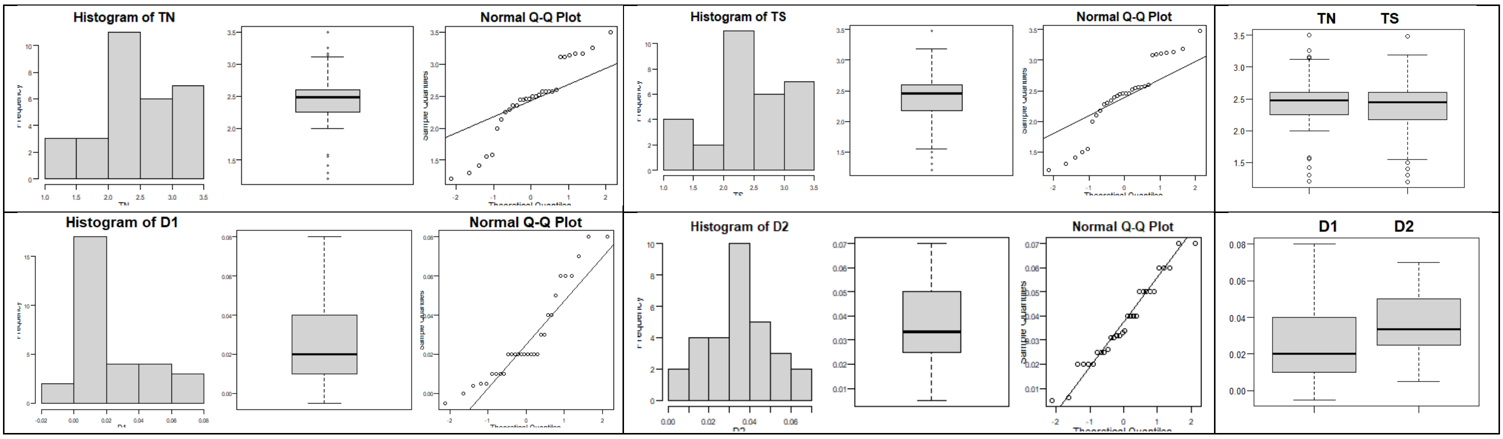
```
P1=12/30=0.4      P2=24/30=0.8
se = sqrt((0.4*0.6)/30+(0.8*0.2)/30) = 0.115
```

Hi ha un diferència de proporcions de 0,4 (40%) amb un error tipus de 0.115 (11.5%)

```
(0.4-0.8)-qnorm(0.975)*0.115 = -0.4-1.96*0.115 = -0.4-0.225 = -0.625
(0.4-0.8)+qnorm(0.975)*0.115 = -0.4+1.96*0.115 = -0.4+0.225 = -0.175
```

Amb 95% de confiança la diferència de proporcions està entre -0.625 i -0.175
Amb 95% de confiança el primer navegador presenta entre un 17.5% i un 62.5% menys de *banners* estàtics respecte el segon navegador

5.- (2 punts) Per a les anteriors preguntes 2 i 3 indiqueu les premisses que s'han de complir i indiqueu si es compleixen o no i, si cal, en quin dels següents gràfic es veu i com es veu



Per a la pregunta 2 cal m.a.s, i normalitat de la diferència D1 entre TN i TS. Als gràfics de D1 (histograma, boxplot i qqnorm veien que es compleix per assemblar-se a una normal, amb força simetria i ben alineada al qqnorm)

Per a la pregunta 3 cal m.a.s, i normalitat de D1 i D2 i homoscedasticitat entre elles. Per a D1 hem vist abans que podem acceptar normalitat, i per a D2 també en els seus gràfics. El boxplot entre D1 i D2 mostra que tenen un rang de valors i amplitud semblant que permet assumir homoscedasticitat entre elles

qt(0.95,30)=1.69726	qt(0.95,29)= 1.69913	qt(0.95,28)= 1.70113	qt(0.95,60)= 1.67065	qt(0.95,59)= 1.67109	qt(0.95,58)= 1.67155	qnorm(0.95)= 1.645
qt(0.975,30)= 2.0423	qt(0.975,29)= 2.0452	qt(0.975,28)= 2.0484	qt(0.975,60)= 2.0003	qt(0.975,59)= 2.001	qt(0.975,58)= 2.002	qnorm(0.975)= 1.96

DNI:_______________________________

*(Contesteu cada pregunta en el seu lloc, amb **lletra molt clara**. Expliciteu i justifiqueu els càlculs)*

Problema D

Hem aprofitat la idea d'un grup d'estudiants que comparen dues plataformes de música (Spotify i Youtube Music), prenent el nombre de (milions de) reproduccions a l'any 2024 en tres gèneres: anys 80, 90 i 2000. Les cançons es poden seleccionar aleatòriament per aquests gèneres amb una pàgina web (no sabem com ho fa, però ho acceptem), hem pres 50 de cada tipus i recollim la variable resposta a les dues plataformes. A la primera figura veiem el logaritme del nombre de reproduccions.

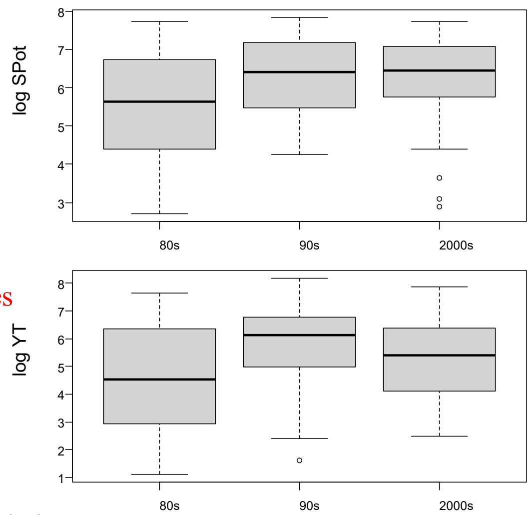
1. Aquest és un estudi experimental o observacional? Justifiqueu la resposta. (1punt)

   És observacional, perquè no hem decidit posar una cançó a una plataforma a l'atzar i mesurar quantes reproduccions es produeixen, sinó que ens hem limitat a observar el que ha passat durant l'any 2024, sense haver intervingut directament.

2. Quin disseny hem aplicat a l'estudi? Quines són les principals variables que hi intervenen, i quin paper hi juguen? Com reescriuríeu l'objectiu més precisament? (1.5 punts)

   Com que cada cançó es mesura a les dues plataformes, és un disseny aparellat. La resposta (outcome) és el nombre de reproduccions anual; la plataforma i el gènere són les variables explicatives (ja que no es poden considerar intervencions). Per el que es dedueix de l'enunciat, la plataforma seria la variable independent principal, així que l'objectiu es podria dir que volem estudiar com es comporta el nombre de reproduccions segons la plataforma (Spotify o Youtube), i secundàriament si aquesta relació es modifica segons el gènere musical.

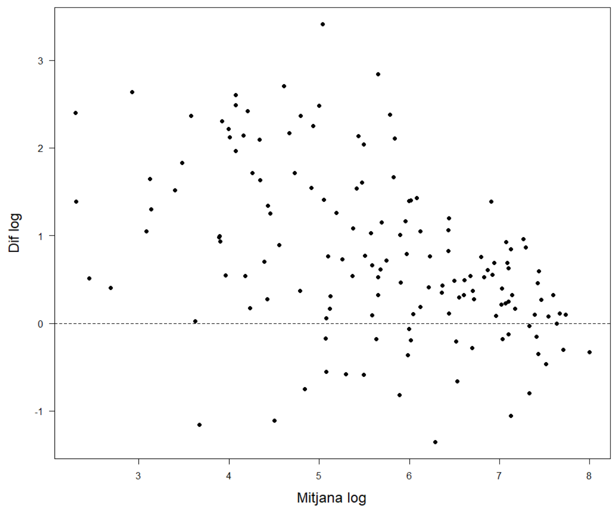
3. Descriviu breument la figura 1, i la informació que explica d'aquest estudi. Ens és útil per les premisses de l'anàlisi que hi aplicarem? (1 punt)

   Es representa el logaritme del nombre de reproduccions, però cada plataforma es veu a una escala distinta. Per aquesta raó no s'aprecien grans diferències entre Spotify i Youtube, però realment el logaritme és lleugerament més alt per Spotify en qualsevol gènere (Spotify més utilitzat).

   També veiem que en ambdues plataformes el nombre de reproduccions per cançons dels 80 és menor, i que 90s i 2000s són més semblants. En tot cas, com que hem d'utilitzar una mesura com el rati o la diferència de logaritmes (que no es mostra aquí), perquè el disseny és aparellat, no és una representació molt útil.

4. La figura 2 és un plot Bland-Altman de les variables log-transformades (el sentit de la diferència és *logSpot – logYT*). Què es pot concloure del diagrama, en aquest cas particular? (1.5 punts)

   A l'esquerra els punts solen estar per sobre de la línia 0, i molt dispersos. Quan ens movem cap a la dreta, hi apareixen més valors negatius i després es concentren més.

   Sembla que quan el nombre de reproduccions és més aviat baix (esquerra) Spotify guanya clarament, però amb cançons amb nombre de reproduccions alt (dreta) les plataformes semblen molt més igualades. Parlar d'un únic efecte constant potser no és adequat del tot.

5. A la dreta veiem el resultat de modelar el logaritme del nombre de reproduccions per la plataforma. Expliqueu el més important de la sortida, com s'interpreta i, si veieu algun punt de crítica, exposeu-ho raonadament. (1.5 punts)

```
> summary(lm(ln.nrep~plat))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.0087     0.1179  50.982  < 2e-16
platYtub     -0.7783     0.1667  -4.669 4.58e-06

Residual standard error: 1.443 on 298 degrees of freedom
Multiple R-squared:  0.06818,	Adj. R-squared:  0.06505
```

El primer "Estimate" (6.0087) correspon a la mitjana de la Y per a la categoria Spotify, amb error tipus 0.1179. El segon (-0.7783, e.t.=0.1667) seria el canvi en la mitjana per a la categoria Youtube. El rati senyal/soroll o valor t és -4.669, així que no sembla una diferència explicable només per l'atzar del mostreig: Youtube obtindria menys reproduccions que Spotify.

Desfent la transformació, en termes de (milions de) reproduccions, tindríem 407 per Spotify i 187 per Youtube (serien més aviat les mitjanes geomètriques). La variació residual, atribuïble a les cançons particulars, val 1.443. En escala logarítmica, aquest valor equival a multiplicar o dividir la mitjana geomètrica per 4.23 ($e^{1.443}$). El coeficient de determinació és només 6.8%: la plataforma explicaria molt poc de la variació en el nombre de reproduccions, en tot cas. Malgrat tot, l'anàlisi no és correcta, perquè s'assumeix que els dos grups (Spotify i Youtube) són independents, i realment no ho són, les observacions són aparellades, com hem dit a la pregunta 2.

6. A continuació, comenteu el resultat que es mostra en aquesta sortida de R. (2 punts)

```
> dat$dif = dat$lnS - dat$lnY   # ln* són logaritmes neperians
> mod = lm(dif ~ gen, dat)
> summary(mod)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.91237    0.13085   6.973 9.79e-11
gen2000s     0.02586    0.18505   0.140   0.8890
gen90s      -0.42817    0.18505  -2.314   0.0221

Residual standard error: 0.9252 on 147 degrees of freedom
Multiple R-squared:  0.04913,	Adj. R-squared:  0.03619
```

La variable dependent ara és el logaritme del rati del nombre de reproduccions, spotify/youtube, i utilitzem un model amb el gènere musical com a explicativa. La categoria de referència és els anys 80, amb mitjana estimada de 0.91237 (e.t.=0.13085): el rati equival a una mitjana de 2.49 vegades més reproduccions a Spotify que a Youtube.

El canvi en el logaritme del rati per als 2000 és petit (0.02585, amb e.t.=0.185), per tant un rati similar al dels 80, però el canvi per als 90 sembla més important (-0.428, e.t.=0.185), equivalent a un rati de només 1.62, igualment favorable a Spotify però més modest. En qualsevol cas, molta variabilitat residual (s=0.9252) i per tant poca capacitat predictora del gènere (R^2=4.9%).

7. Interpreteu aquesta sortida de R, en relació amb el model utilitzat: a) relacioneu la resposta amb els paràmetres del model estadístic formal, i b) expliqueu com obteniu els valors de la fila "Intercept". (1.5 punts)

```
> confint(mod)
                 2.5 %      97.5 %
(Intercept)  0.6537847  1.17096465
gen2000s    -0.3398374  0.39156549
gen90s      -0.7938707 -0.06246786
```

El model es pot escriure como $Y_j = \mu + \theta_j + \varepsilon$; a R hi ha una categoria (en aquest cas, el 80s) de referència per la que $\theta = 0$, i la seva mitjana es correspon amb la mitjana general μ.

La sortida ens diu que les respectives estimacions de μ, $\theta_{2000} + \theta_{90}$, amb IC al 95%, són (0.65, 1.17), (-0.34, 0.39) i (-0.79, -0.06). El paràmetre μ és la mitjana del logaritme del rati de reproduccions S/Y a la categoria 80s; el paràmetre θ_{2000} és el canvi en la mitjana de la categoria 2000s respecte la referència; i similar per al paràmetre θ_{90}.

L'IC de la fila "Intercept" es pot calcular del *summary* del model a partir de: $0.91237 \pm t_{0.975,\,147} \cdot 0.13085$. Traient els logaritmes, la mitjana del rati S/Y estaria (conf. 95%) entre 1.92 i 3.23 per als 80.