

DNI: _____

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

Problema 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.

2. Trobar la probabilitat que la decisió final del jurat sigui la correcta.

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.

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

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:

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

5. Dedueix el valor de la constant k , i trobeu el valor esperat de la variable X.
6. Representeu gràficament la funció de distribució de X (podeu utilitzar una escala logarítmica pel eix d'abscisses).
7. Quina és la probabilitat que la variable X prengui un valor entre 20 i 50?
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 càcul.

DNI:

(Contesteu cada pregunta en el seu lloc, amb lletra molt clara. Explicieu 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)
 2. Quina és la probabilitat que passin més de 5 minuts entre missatges enviats per tots 5 bots (per qualsevol d'ells)? (1 punt)
 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)

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)

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)

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)

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 població és inferior a 1930? (2 punts)

<code>ppois(20,20)= 0.55909</code>	<code>ppois(10,20)= 0.01081</code>	<code>dpois(20,20)= 0.08884</code>	<code>dpois(10,20)= 0.0058163</code>
<code>ppois(40,20)= 0.99997</code>	<code>ppois(10,4)= 0.99716</code>	<code>dpois(40,20)= 2.78e-05</code>	<code>dpois(10,4)= 0.0052925</code>
<code>ppois(30,20)= 0.98653</code>	<code>ppois(5,4)= 0.78513</code>	<code>dpois(30,20)= 0.0083435</code>	<code>dpois(5,4)= 0.15629</code>
<code>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</code>			
<code>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</code>			

DNI: _____

(Contesteu cada pregunta en el seu lloc, amb lletra molt clara. 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 $D1 = TN - TS$

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

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

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

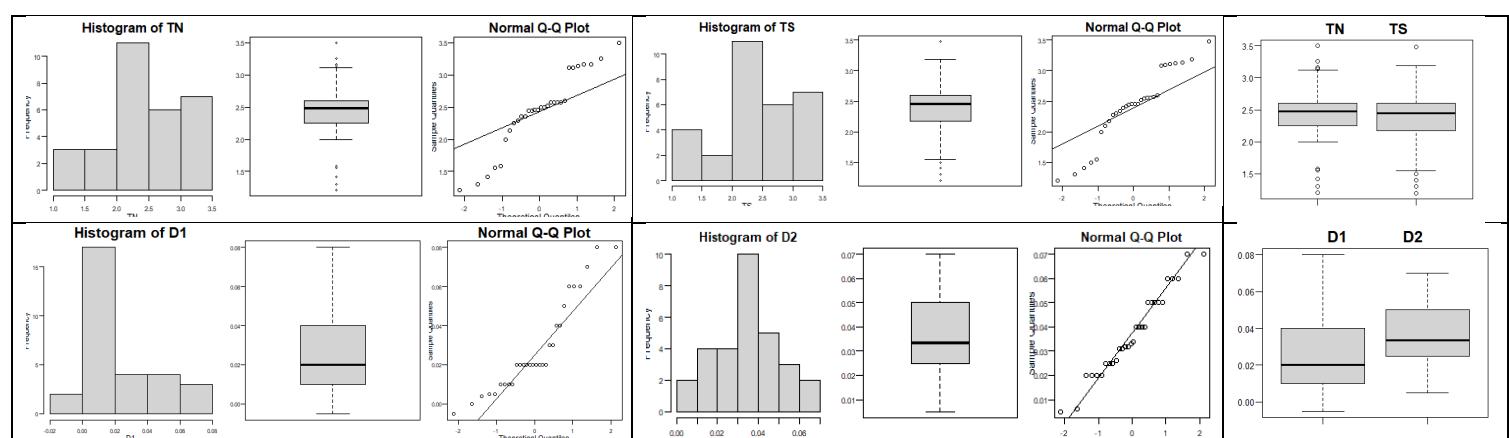
1.- (2 punts) Per una part volem, només pel primer navegador, comparar les mitjanes de temps usant o no bloquejador. Per altra part volem comparar les mitjanes de la diferència (d'usar o no bloquejador) entre els dos navegadors. Indiqueu i justifiqueu si cadascuna d'aquestes dues proves de comparació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

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

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

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

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

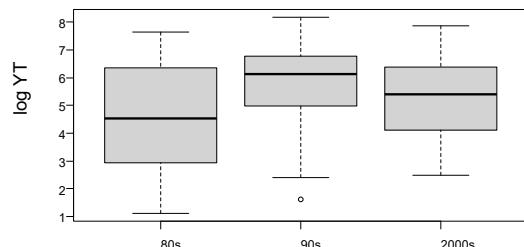
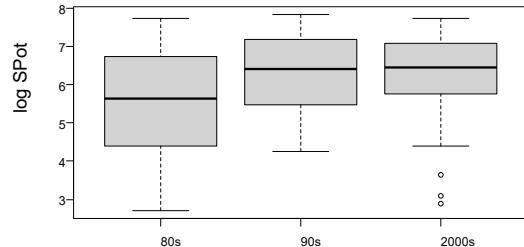


$qt(0.95, 30) = 1.69726$	$qt(0.95, 29) = 1.69913$	$qt(0.95, 28) = 1.70113$	$qt(0.95, 60) = 1.67065$	$qt(0.95, 59) = 1.67109$	$qt(0.95, 58) = 1.67155$	$qnorm(0.95) = 1.645$
$qt(0.975, 30) = 2.0423$	$qt(0.975, 29) = 2.0452$	$qt(0.975, 28) = 2.0484$	$qt(0.975, 60) = 2.0003$	$qt(0.975, 59) = 2.001$	$qt(0.975,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

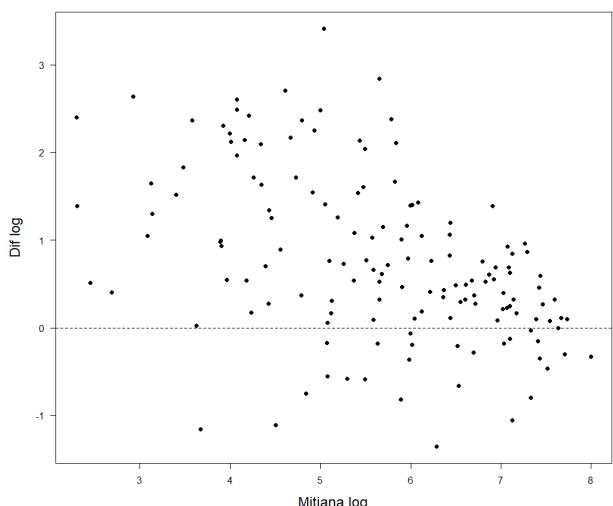
Problema D

Hem aprofitat la idea d'un grup d'estudiants que comparen dues plataformes de música (Spotify i Youtube Music), prenent el nombre de (millions de) reproduccions a l'any 2024 en tres gèneres: anys 80, 90 i 2000. Les cançons es poden seleccionar aleatoriament 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. (1 punt)
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)
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)



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



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
```

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
```

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 obtenuï 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
```