NOM:_____COGNOMS:_____

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

Problema 1 (B1-B2)

Totes les qüestions valen 1 punt

Malgrat els esforços de Twitter per controlar els missatges generats per robots (*bots*), cada cop més estan presents a la xarxa. Per intentar identificar de forma més precisa aquest missatges, Twitter ha contactat amb 2 empreses externes (Emp1 i Emp2) que es dediquen a detectar aquests missatges. La probabilitat de que un *tweet* (missatge) hagi estat generat per un Bot (B) és de **0.1**.

A més tenim les següents probabilitats:

- La probabilitat que l'empresa 1 classifiqui un missatge com a bot (B1) és de **0.17**
- La probabilitat que l'empresa 1 classifiqui un tweet generat per un bot (B) com a bot (B1) és de **0.8**
- La probabilitat que l'empresa 2 classifiqui un tweet generat per un bot (B) com a bot (B2) és de **0.78**
- La probabilitat que un tweet sigui un bot (B) i, a més, ambdues empreses el classifiquin com a bot és de **0.076**
- En un tweet que no és un bot ($\bar{B}$), la probabilitat que l'empresa 1 el classifiqui com a bot (B1) i l'empresa 2 (B2) el classifiqui com no bot ($\overline{B2}$) és **0.095**.
- En global, la probabilitat d'encert (correcta classificació) d'Emp2 és **0.0754** superior a la probabilitat d'encert d'Emp1.

Pista: Construeix l'arbre de probabilitat i afegeix valors a mesura que vas fent els següents apartats.

1. Quina és la probabilitat que l'empresa 1 classifiqui com a *bot* ($B1$) un tweet que no havia estat generat per un bot ($\bar{B}$)?

$$P(B1) = P(B1|B) \cdot P(B) + P(B1|\bar{B}) \cdot P(\bar{B})$$
$$0.17 = 0.8 \cdot 0.1 + P(B1|\bar{B}) \cdot 0.9 \rightarrow P(B1|\bar{B}) = \frac{0.17 - 0.8 \cdot 0.1}{0.9} = \mathbf{0.1}$$

2. Quina és la probabilitat que un tweet generat per un bot (B) i que l'empresa 1 l'ha classificat com un bot (B1), l'empresa 2 també el classifiqui com un bot (B2)?

$$P(B2|B \cap B1) = \frac{P(B2 \cap B \cap B1)}{P(B \cap B1)} = \frac{P(B2 \cap B \cap B1)}{P(B1|B) \cdot P(B)} = \frac{0.076}{0.8 \cdot 0.1} = \mathbf{0.95}$$

3. Quina és la probabilitat que un tweet sigui un bot i l'empresa 1 el classifiqui malament ($\overline{B1}$) però l'empresa 2 el classifiqui correctament (B2)?

$$P(B \cap \overline{B1} \cap B2) = P(B \cap B2) - P(B \cap B1 \cap B2) = P(B2|B) \cdot P(B) - P(B \cap B1 \cap B2) = 0.78 \cdot 0.1 - 0.076 = \mathbf{0.002}$$

4. Quina és la probabilitat que l'empresa 2 classifiqui correctament ($\overline{B2}$) un tweet que no és un bot ($\bar{B}$)?

$$P(\text{"Encert empresa 1"}) = P(B1|B) \cdot P(B) + P(\overline{B1}|\bar{B}) \cdot P(\bar{B}) = 0.8 \cdot 0.1 + 0.9 \cdot 0.9 = 0.89$$
$$P(\text{"Encert empresa 2"}) = P(B2|B) \cdot P(B) + P(\overline{B2}|\bar{B}) \cdot P(\bar{B}) \rightarrow (0.89 + 0.0754) = 0.78 \cdot 0.1 + P(\overline{B2}|\bar{B}) \cdot 0.9$$
$$\rightarrow P(\overline{B2}|\bar{B}) = \frac{(0.89 + 0.0754) - 0.78 \cdot 0.1}{0.9} = \mathbf{0.986}$$

5. Un tweet concret, l'empresa 1 el classifica com a bot (B1) i l'empresa 2 el classifica com a no bot ($\overline{B2}$). Quina és la probabilitat que realment no provingui d'un bot ($\bar{B}$)?

$$P(\bar{B}|B1 \cap \overline{B2}) = \frac{P(B1 \cap \overline{B2}|\bar{B}) \cdot P(\bar{B})}{P(B1 \cap \overline{B2}|B) \cdot P(B) + P(B1 \cap \overline{B2}|\bar{B}) \cdot P(\bar{B})} = \frac{0.095 \cdot 0.9}{0.04 \cdot 0.1 + 0.095 \cdot 0.9} = \mathbf{0.9553}$$

$$P(B1 \cap \overline{B2}|B) = \frac{P(\overline{B2} \cap B \cap B1)}{P(B)} = \frac{P(B \cap B1) - P(B2 \cap B \cap B1)}{P(B)} = \frac{P(B1|B) \cdot P(B) - P(B2 \cap B \cap B1)}{P(B)}$$
$$= \frac{0.8 \cdot 0.1 - 0.076}{0.1} = 0.04$$

Un aspecte important a tenir en compte a l'hora de valorar si un *tweet* ha estat generat per un *bot* és el temps que fa que l'usuari que va publicar el *tweet* va crear el seu compte. Aquest temps (en mesos) es distribueix segons la funció de densitat descrita a continuació:

$$f(t) = k \cdot e^{-\frac{t}{100}} \qquad per\ 0 < t < 150$$

6. Troba el valor de la constant *k* per a que *f(t)* sigui una funció de densitat.

$$\int_0^{150} f(t)dt = \int_0^{150} k \cdot e^{-\frac{t}{100}}dt = k\left[-100 \cdot e^{-\frac{t}{100}}\right]_0^{150} = k[-100 \cdot e^{-1.5} + 100] \rightarrow k = \frac{1}{100 \cdot (1 - e^{-1.5})} \approx \mathbf{0.0129}$$

7. Un algoritme classifica com a probable usuari "fals" aquell que fa menys de mig any que va crear el compte. Calcula la probabilitat de que un usuari estigui classificat com a tal.

$$\int_0^{6} 0.0129 \cdot e^{-\frac{t}{100}}dt = \left[-0.0129 \cdot 100 \cdot e^{-\frac{t}{100}}\right]_0^{6} = 1.29 \cdot \left[1 - e^{-\frac{6}{100}}\right] \approx \mathbf{0.075}$$

Les dues empreses de l'inici cobren per tweet analitzat. El preu per tweet analitzat depèn del nombre de tweets generats per bots reals detectats i va des de 0.01 € a 0.03 € i. La funció de probabilitat conjunta dels costos unitaris de cada empresa estan a la següent taula.

Costos Unitaris		**Empresa 2**		
		0.01 €	**0.02 €**	**0.03 €**
Empresa 1	**0.01 €**	0.1	0.05	0
	0.02 €	0.1	0.2	0.15
	0.03 €	0	0.1	0.3

8. Troba el cost unitari (per tweet) esperat per ambdues companyies. (4 decimals de precisió)
   CU_1="Cost Unitari per tweet per l'empresa 1"
   CU_2="Cost Unitari per tweet per l'empresa 2"

$$E(\mathrm{CU}_1) = 0.15 \cdot 0.01 + 0.45 \cdot 0.02 + 0.4 \cdot 0.03 = \mathbf{0.0225}$$
$$E(\mathrm{CU}_2) = 0.2 \cdot 0.01 + 0.35 \cdot 0.02 + 0.45 \cdot 0.03 = \mathbf{0.0225}$$

9. Twitter encarrega l'anàlisi d'1 milió de Tweets a l'empresa 1 i de 2 milions de tweets a l'empresa 2. Troba el cost unitari (per tweet) esperat dels 3 milions de tweets. (4 decimals de precisió)
   CU="Cost unitari per Tweet global"

$$E(CU) = E\left(\frac{1 \cdot 10^6 \cdot \mathrm{CU}_1 + 2 \cdot 10^6 \cdot \mathrm{CU}_2}{3 \cdot 10^6}\right) = \frac{1 \cdot 10^6 \cdot \mathrm{E}(\mathrm{CU}_1) + 2 \cdot 10^6 \cdot E(\mathrm{CU}_2)}{3 \cdot 10^6} = \frac{1 \cdot 0.0225 + 2 \cdot 0.0225}{3} = \mathbf{0.0225}$$

10. Sense fer cap càlcul digues si creus que la covariància entre aquestes dues variables és positiva, negativa o nul·la i argumenta-ho.

    Serà positiva ja que es veu a la taula de probabilitat conjunta que hi ha una tendència a tenir probabilitats més altes de cost elevat en una companyia quan a l'altre companyia el cost també és elevat.

NOM:_____COGNOMS:_____

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

Problema 2 (B3-B4)

Per decidir sobre normatives de control de contaminació un dels elements més controvertits són els diversos indicadors del nivell d'emissions de gasos contaminants dels vehicles. Una determinada marca indica per a un del seus models que la mesura d'opacitat (absorció en un filtre en acceleració forçada seguint una escala anomenada de Bacharach, en la qual quant més alt més contaminant) segueix una distribució Normal centrada a 1.5 i amb desviació 0.6

1.- (0.5 punts) Calculeu la probabilitat que un d'aquests vehicles doni una mesura d'opacitat per sobre de 2

O és N(mu=1.5, sigma=0.6)

P(O>2) = P(Z > (2-1.5)/0.6) = P(Z > 0.83) = 1 – P(Z<=0.83) = 1 – 0.7967 = 0.2033 **(20%)**

2.- (1 punt) Indiqueu el valor d'opacitat pel qual s'assegura que un 97.5% d'aquests vehicles no el superarà. I entre quins valors, centrats en la mitjana, es pot assegurar la opacitat amb una probabilitat del 95%

1.96 = (?-1.5)/0.6 -> ? és (1.96*0.6)+1.5 = 2.676 **Valor 97.5% és 2.676**

-1.96 = (??-1.5)/0.6 -> ?? és (-1.96*0.6)+1.5 = 0.324 **95 % entre Min de 0.324 i Max de 2.676**

3.- Habitualment, enlloc de fer una sola mesura, es realitzen repetides mesures independents. Si en realitzem 4, indiqueu:

a) (1 punt) la probabilitat que la mitjana aritmètica d'aquestes 4 repeticions doni un valor d'opacitat mitjana per sobre de 2 (expliciteu la variable, model i paràmetres que useu pel càlcul d'aquesta probabilitat i compareu-la amb la de l'apartat 1)

Om és N(mu=1.5, sigma=0.6/sqrt(4)) N(1.5,0.3)

P(Om >2) = P(Z > (2-1.5)/0.3) = P(Z > 2.5) = 1 – P(Z<=2.5) = 1 – 0.9525 = 0.0475 = 0.05

Baixa de ser un 20% a un 5% la probabilitat de valors d'opacitat per sobre de 2

b) (1 punt) la probabilitat que en la mitat d'aquestes repeticions la mesura estigui per sota de 2 (expliciteu la variable, model i paràmetres que useu pel càlcul d'aquesta probabilitat)

Nsota és Bin(n=4,p=0.8) Nsobre és Bin(n=4,p=0.2)

P(Nsota=2) = P(Nsobre=2) = P(Nsobre<=2) – P(Nsobre<=1) = 0.9728 – 0.8192 = 0.1536 **(prob 15%)**

4.- (0.5 punts) Ara ens interessa el nombre de repeticions que cal fer fins obtenir mesures d'opacitat superior a 2. Indiqueu la variable, model, paràmetres, esperança i variància de la variable nombre de repeticions fins a una primera mesura d'opacitat superior a 2. I el mateix per la variable nombre de repeticions fins a dues mesures d'opacitat superiors a 2

Nfins1 és Geom(p=0.2) E(Nfins1) = 1/p = 1/0.2 = 5 V(Nfins1) = 0.8/(0.2*0.2) = 20

Nfins2 és BinNeg(r=2,p=0.2) E(Nfins2) = r/p = 2/0.2 = 10 V(Nfins2) = (0.8*2)/(0.2*0.2) = 40

5.- (1 punt) En una determinada estació d'ITV tenen estudiat que vehicles amb mesures molt altes d'opacitat (superiors a 5) cada any en passen una mitjana de 8. Calculeu la probabilitat que un any en passin només 6, i l'esperança del nombre de mesos en que no hi passi cap d'aquests vehicles

NAny és Poisson(lambda=8) P(NAny=6) = P(NAny<=6)-P(NAny<=5) = 0.313 – 0.191 = 0.122 (12%)

NMes és Poisson(lambda=8/12=0.666) Mesos és Exp(lambda=0.666) E(Mesos) = 1 / lambda = 1/0.666 = 1.5
(un mes i mig és el temps esperable entre que arribin vehicles amb opacitat màxima)

6.- Es decideix prendre 17 mesures a un vehicle per fer inferència del seu nivell d'opacitat, i els resultats han estat:
Opa = c(0.2, 0.6, 0.8, 0.9, 1.2, 1.3, 1.5, 1.5, 1.5, 1.5, 1.6, 1.6, 2.0, 2.0, 2.2, 2.3, 2.8) sum(Opa) = 25.5 sum(Opa*Opa) = 45.07

a) (1 punt) Calculeu una estimació puntual de la mitjana, la desviació i l'error estàndard. Interpreteu-los

mean(Opa) sum(Opa)/17 = 25.5/17 = **1.5**
sd(Opa) sqrt((45.07-(25.5*25.5)/17)/16) = **0.653**
se = sd(Opa)/sqrt(17) = 0.653/4.12 = **0.158**

La mitjana o valor de tendència central del nivell d'opacitat és de 1.5
Com a variabilitat tenim una desviació de 0.65, i l'estàndard error o desviació de la mitjana baixa a menys d'una quarta part fins 0.16

b) (1 punt) Calculeu una estimació per interval amb una confiança del **95%** per a la mitjana de la opacitat. Interpreteu-lo

1.5 - (qt(0.975,16)* se) = 1.5 - 2.120*0.158 = **1.165**
1.5 + (qt(0.975,16)* se) = 1.5 + 2.120*0.158 = **1.835**

Amb un 95 % de confiança el valor de l'opacitat estarà entre 1.165 i 1.835

c) (1 punt) Quantes mesures hauríem de repetir si haguéssim volgut una amplada de 0.5 (0.25 cada costat de l'interval) per a l'interval de confiança de la mitjana d'opacitat al 95% i suposant una desviació poblacional coneguda i igual a 0.7

$n = (qnorm(0.975)*0.7/0.25)^2$ = **30.12**

n ha de ser més gran de 30 per aconseguir aquesta amplada de l'interval

d) (1 punt) A partir d'aquestes 17 dades i sabent que un valor inferior a 2.0 equival a una marca positiva, i sinó una de negativa. Calculeu un interval de confiança al 95% de la proporció de positius
(+, +, +, +, +, +, +, +, +, +, +, +, -, -, -, -, -)
p = 12/17 = 0.706
se = sqrt(0.5*0.5/17) = 0.12 (o se = sqrt(p*(1-p)/n) = sqrt(0.706*0.294/17) = 0.11)

Min: 0.706 – 1.96*0.12 = **0.47** (o 0.706 – 1.96*0.11 = 0.49)
Max: 0.706 + 1.96*0.12 = **0.94** (o 0.706 + 1.96*0.11 = 0.92)

e) (1 punt) Realitzeu la prova amb risc 5% de si la proporció de positius es d'un 50% o superior i indiqueu: hipòtesis, estadístic, punt/s crític/s, resolució i conclusió

H0: π=0.5 H1: π >0.5

Estadístic: (0.706 – 0.50)/se = 0.206/sqrt(0.5*0.5/17) = 0.206/0.12= **1.71** (o 0.206 / sqrt(0.706*0.294/17)=0.206/0.11=1.86)

Punt crític, qnorm(0.95), 1.645

Com que l'estadístic (1.71) > punt crític (1.645) vol dir que estem a la zona de rebuig de H0 i no a la d'acceptació

Evidència per rebutjar una proporció de 50% de positius, sinó que la proporció de positius és superior al 50%

NOM:_____COGNOMS:_____

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

Problema 3 (B5-B6)

Uns alumnes estudien les diferències de la valoració donada a (A) jocs de pagament i (B) jocs gratuïts, prenent una mostra (seleccionats de la recopilació feta per la revista "Supergamer") de 30 jocs de cada classe que tinguin almenys 50 valoracions dels usuaris. Els usuaris valoren els jocs en una escala entera de 1 a 5, i aquests alumnes recullen el nombre de valoracions i la valoració mitjana (Y) per cada joc.

$$\text{A: } \sum y_{A,i} = 136.0517 \quad \sum y_{A,i}^2 = 617.0739 \qquad \text{B: } \sum y_{B,i} = 135.2725 \quad \sum y_{B,i}^2 = 610.0058$$

L'objectiu és comprovar si els jocs de pagament o gratuïts tenen diferent valoracions. Expliqueu pas a pas i detalladament com arribeu a la resposta, que ha d'incloure a més a més un interval de confiança al 95% i la seva interpretació. (3pt)

Variable resposta: Y "valoració mitjana dels usuaris"

Disseny: dos grups (jocs de pagament / jocs gratuïts), amb dues mostres independents

Hipòtesis: H_0: $\mu_A = \mu_B$ contra H_1: $\mu_A \neq \mu_B$

Premisses: dues m.a.s. independents; Y és Normal; variàncies desconegudes i iguals.

Sota H_0 $t = \frac{\bar{y}_B - \bar{y}_A}{s\sqrt{\frac{1}{n_B}+\frac{1}{n_A}}} \sim t_{nA+nB-2}$; prenem risc α bilateral del 5%.

Càlculs. $\overline{y_A} = \frac{136.0517}{30} = 4.535057$ $\overline{y_B} = \frac{135.2725}{30} = 4.509083$ $s_A^2 = 0.00247369$ $s_B^2 = 0.001757671$ $s = 0.04599652$

$$t = \frac{4.509083 - 4.535057}{0.04599652\sqrt{\frac{1}{30}+\frac{1}{30}}} = -2.187$$

Rebutgem si $|t| > t_{58,\,0.975} = 2.002$ (es pot aproximar per $t_{60,\,0.975} = 2.000$) i, en aquest cas, es cert. Podem rebutjar que *els valors esperats de les valoracions mitjanes de jocs gratuïts i de pagament siguin els mateixos*. La diferència entre els valors esperats es situa entre:

$$4.509083 - 4.535057 \pm 2.002 \times 0.04599652\sqrt{\tfrac{1}{30}+\tfrac{1}{30}} = (-0.04975, -0.00220) \text{ (IC 95\% } \mu_B - \mu_A)$$

És a dir: creiem amb un 95% de confiança que el valor esperat de les valoracions mitjanes dels jocs gratuïts està entre 0.0022 i 0.0497 punts per sota que el dels jocs de pagament.

Calia posar un nombre mínim de valoracions als jocs que anaven a ser inclosos a l'estudi? Quina relació té aquest criteri amb les premisses de la prova que heu fet anteriorment? (1pt)

La resposta que s'observa és la mitjana dels usuaris que han donat una valoració (no cada valoració individual). Si cada dada fos la valoració individual, la resposta es distribuiria segons una variable discreta prenent valors d'1 a 5, i no seria Normal. Per a poder acceptar que la mitjana de diferents valoracions sigui aproximadament Normal hem de considerar la mitjana d'un nombre suficientment alt d'usuaris, i creiem que 50 és un nombre apropiat, perquè la valoració és discreta i possiblement bastant asimètrica cap a la dreta (els usuaris tendeixen a valorar alt), factors que aconsellen aplicar el TCL amb mides superiors a la pràctica habitual, per exemple, 30.

Un dels autors de l'estudi escriu al final: "El valor P obtingut, 4%[1], ens diu que la probabilitat d'equivocar-nos rebutjant la igualtat de mitjanes és només un 4%". Podeu comentar si aquesta afirmació és correcta o tractar de millorar-la? (1pt)

Aquesta frase no és correcta. El valor P és la probabilitat de trobar un resultat mes extrem que el que hem trobat (és a dir, un estadístic *t* més gran en valor absolut que 2.187), partint del fet que els dos tipus de jocs no tenen diferències en les valoracions. En aquest cas particular vindria a dir que només un 4% de les repeticions de l'estudi (suposant igualtat entre els jocs) per atzar obtindríem un resultat mostrant més discrepància que en aquesta realització. La frase de l'enunciat no diu enlloc que el 4% vingui d'una suposada igualtat entre jocs de pagament i gratuïts.

La segona part de l'estudi tracta de veure si un joc amb més valoracions té millor (o pitjor) valoració mitjana. S'agafa la mostra de jocs gratuïts, dels quals s'obtenen uns estadístics del logaritme natural del nombre de valoracions (X), ja que hi ha grans diferències entre uns jocs i altres:

$$\bar{x} = 6.489353 \quad \max = 8.451694 \quad s_X = 1.337217 \quad \text{cov} = 0.01695832$$

(1pt) Amb aquesta informació, justifiqueu que l'equació de la recta que relaciona X amb Y és: Y = 4.44754 + 0.009484 X

Comprovem: pendent, $b_1 = \frac{s_{X,Y}}{s_X^2} = \frac{0.016958}{1.3372^2} = 0.0094837$

$$b_0 = \bar{y} - b_1\bar{x} = 4.509083 - 0.0094837 \times 6.489353 = 4.44754$$

(0.5pt) Si es duplica el nombre de valoracions d'un joc, quin increment (puntual) podem esperar a la variable resposta?

Si V és el nombre de valoracions, diem què veurem en la valoració mitjana si passem de V=v a V=2v. En termes de logaritmes, de X=log(v) a X=log(2v) = log(v)+log(2). Per tant, l'increment esperat a Y serà de log(2)b_1 = 0.00657

(1pt) Trobeu l'error tipus associat al pendent, i resoleu la prova d'hipòtesis per determinar si amb més valoracions esperem una valoració del joc més alta. Estimeu per IC al 95% de confiança el pendent.

a) Trobar valor de la desviació residual s_R:

$$s_R^2 = \frac{(n-1)(S_Y^2 - b_1 s_{X,Y})}{n-2} = \frac{29\,(0.001757671 - 0.0094837 \times 0.016958)}{28} = 0.001653877 = 0.04067^2$$

b) Trobar error tipus del pendent estimat:

$$s_{b1} = \sqrt{\frac{s_R^2}{(n-1)\cdot s_X^2}} = \sqrt{\frac{0.001653877}{29 \cdot 1.78815}} = 0.005647$$

c) Obtenir l'estadístic per a la prova d'hipòtesi $\beta_1 = 0$:

$t = \frac{b_1}{s_{b1}} = \frac{0.0094837}{0.005647} = 1.679$, no podem rebutjar la hipòtesi, és versemblant que el nombre de valoracions no afecta al valor esperat de la valoració del joc (límit per no rebutjar: 2.048).

d) Interval de confiança per a β_1:

$$b_1 \pm t_{28,0.975}\, s_{b1} = 0.0094837 \pm 2.048 \times 0.005647 = (-0.00208, 0.02105)$$

(1pt) Quant val el coeficient de determinació d'aquest estudi, i com s'interpreta?

$$R^2 = r_{X,Y}^2 = \left(\frac{s_{X,Y}}{s_X\, s_Y}\right)^2 = \left(\frac{0.01695832}{1.337217 \times \sqrt{0.001757671}}\right)^2 = 0.0915$$

El 9.15% de la variabilitat en la valoració mitjana dels jocs es pot atribuir al nombre de valoracions (i més del 90% és d'origen desconegut). Ja hem vist abans que el nombre de valoracions (més exactament, el seu logaritme) no és estadísticament significatiu.

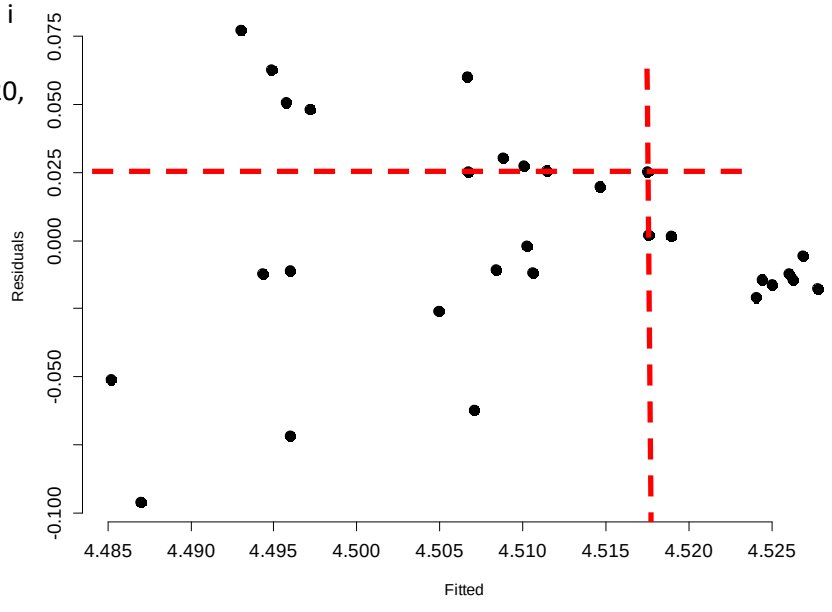
(1pt) El joc nº 20 de la mostra va tenir 1590 valoracions i una mitjana de 4.542767. Amb aquesta informació, localitzeu al gràfic del costat quin punt és el del joc nº 20, justificant la resposta.

Resposta estimada amb la recta (fitted):

4.44754 + 0.009484 * log(1590) = 4.517451

Residu = 4.542767 - 4.517451 = 0.025316

El punt és el que es troba a la intersecció de les rectes.



(0.5pt) A partir del gràfic, comenteu les premisses del model que veieu.

A partir d'aquest gràfic es pot comentar sobre linealitat i homoscedasticitat. Els residus no presenten cap estructura notable amb forma no lineal (còncava o convexa, o similar), llavors la premissa de linealitat seria acceptable, encara que sabem possiblement la resposta no depèn de X.

Per altra banda, s'observa clarament que a la dreta els residus estan agrupats. Donat que el pendent és positiu, el costat dret correspon a jocs amb moltes valoracions. Per tant, és normal que la resposta Y (que és la valoració mitjana) tingui menys dispersió. Per aquest motiu, tal com és manifest al gràfic, no es pot acceptar la premissa d' homoscedasticitat.