

**NOM:** \_\_\_\_\_

(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes)

### Problema 1.

Baixar una cançó del nostre grup de rock preferit d'un determinat servidor en Internet al disc dur del vostre ordinador tarda cert temps que podem considerar com una variable aleatòria  $X$  amb distribució normal amb una esperança de 10 segons i una desviació típica de 2 segons.

#### A. (2 punts per cadascun dels 5 apartats)

##### 1.-

1.1 Quina és la probabilitat de que baixar una cançó trigui més de 12 segons?

$$\begin{aligned} P(X > 12) &= P(Z > (12-10)/2) \\ &= P(Z > 1) \\ &= 1 - P(Z < 1) \\ &= 1 - 0.84513 \text{ (Valor taules Normal (0,1))} \\ &= 0,1587 \quad \leftarrow \mathbf{1 \text{ punt}} \end{aligned}$$

[R: 1-pnorm(12,10,2) ]

1.2 I la d'obtenir la cançó en un temps inferior a 7 segons?

$$\begin{aligned} P(X < 7) &= P(Z < (7-10)/2) = \\ &= P(Z < -1.5) \\ &= 1 - P(Z < 1.5) \\ &= 1 - 0.9332 = 0.0668 \quad \leftarrow \mathbf{1 \text{ punt}} \end{aligned}$$

[R: pnorm(7,10,2)]

**2.-** Si volem calcular les probabilitats de temps de baixada d'un CD recopilatori, amb 35 cançons (suposant en la baixada de les 35 cançons és independent) hem de considerar la variable:  $X = X_1 + X_2 + \dots + X_{35}$ .

2.1 ¿Quina és la llei de distribució d' $X$ ? Com es justifica? Indica els paràmetres de la llei de distribució.

Pel teorema del límit central,  $X \sim \text{Normal}$   $\leftarrow \mathbf{0,5 \text{ punt}}$

$$\Sigma X_i \approx N(35\mu, \sqrt{35}\sigma) \text{ O bé } N(35 \cdot 10 = 350; 5,9 \cdot 2 = 11,8) \quad \leftarrow \mathbf{0,5 \text{ punt}}$$

2.2 Quina és la probabilitat de que el CD complet trigui entre 320 segons i 360 segons?

$$\begin{aligned} \text{Càlcul } P(320 < X < 360) &= P\left(\frac{320-350}{11.8} < Z < \frac{360-350}{11.8}\right) \\ &= P(-2.54 < Z < 0.85) \\ &= P(Z < 0.85) - [1 - P(Z < 2.54)] \\ &= 0.8023 - (1 - 0.9945) \text{ (Valors taules Normal (0,1))} \\ &= 0.7968 \quad \leftarrow \mathbf{1 \text{ punt}} \end{aligned}$$

[R > pnorm(0.85,0,1)-pnorm(-2.54,0,1) [1] 0.7967948  
> pnorm(360,350,11.8)-pnorm(320,350,11.8) [1] 0.7961248 ]

**3.-** Suposant que el temps de baixada segueix una llei uniforme a l'interval de 8 a 16 segons, considerant 35 suficient per aplicar el TCL:

3.1 Quina és la distribució de la variable suma  $X = (X_1 + X_2 + \dots + X_{35})$ ?

Per a fer-ho segons la diapositiva del TCL al B3 i fer servir la fórmula:

$$\begin{aligned} \Sigma_{i=1}^n X_i &\approx N(n\mu, \sigma\sqrt{n}) \quad \leftarrow \mathbf{0,25 \text{ punts}} \\ a &= 8; b = 16; n = 35 \\ E(\mathbf{X}_i) &= (a+b)/2 = (8+16)/2 = 12 \\ V(\mathbf{X}_i) &= [(b-a)^2]/12 = ((16-8)^2)/12 = 5.333 \\ X &\sim N(35 \cdot 12; \text{arrel}(5.333 \cdot 35)) \quad \leftarrow \mathbf{0,75 \text{ punts}} \\ X &\sim N(420; 241.4) \end{aligned}$$

3.2 Quina és la distribució de la variable promig  $Y = (Y_1 + Y_2 + \dots + Y_{35})/35$ ?

Idem per al calcul de la distribució del promig.

$$\frac{\sum_{i=1}^n Y_i}{n} = \bar{Y}_n \approx N(\mu, \sigma / \sqrt{n}) \quad \leftarrow 0,25 \text{ punts}$$

$$X \sim N(12; \text{arrel}(5.333/35)) \quad \leftarrow 0,75 \text{ punts}$$

$$X \sim N(12; 0.39)$$

4.- D'altra banda, s'ha observat que 4 de cada 10 CDs surten defectuosos i no podem fer l'enregistrament de la nostra música d'una manera correcta.

4.1 Quina és la llei de distribució de la variable  $R \sim$  "nombre de CDs defectuosos per paquet de 10 CDs"?

Indica també el/s paràmetre/s d'aquesta llei.

$$R \sim \text{Binomial} \quad \leftarrow 0,5 \text{ punts}$$

$$\text{Paràmetres: } n=10; p=4/10 \quad \leftarrow 0,5 \text{ punts}$$

4.2 Quina és la probabilitat de trobar exactament 5 CDs defectuosos en 1 paquet de 10?

$$P(R=5) = \text{binomi}(10, 4) * 0,4^5 * 0,6^{10-5}$$

$$\text{o be } = F_R(5) - F_R(4) \quad \leftarrow 0,5 \text{ punts}$$

$$= 0.8338 - 0.6331 \text{ (Valors Taules Binomial, } n=10, p=0,4; k=5 \text{ i } k=4) \quad \leftarrow 0,5 \text{ punts}$$

$$= 0.2007$$

$$[ R \text{ pbinom}(5,10,0.4) - \text{pbinom}(4,10,0.4) [1] 0.2006581 ]$$

5.- Donada l'alt nombre de CDs defectuosos, els usuaris fan arribar les seves queixes al distribuïdor. El nombre de reclamacions és en promig de 4,5 reclamacions/dia.

5.1 Quina llei de distribució segueix la variable  $Y \sim$  "Nombre de reclamacions al dia"? Indica també el/s paràmetre/s d'aquesta llei. Quina és la probabilitat de recollir menys de 7 reclamacions en 2 dies?

$$Y \sim \text{Poisson}(\lambda=4,5) \quad \leftarrow 0,5 \text{ punts}$$

$$Y_2 \text{ (Nombre de reclamacions en 2 dies)} \sim \text{Poisson}(\lambda=4,5*2)$$

$$P(X < 7) = P(X \leq 6) = 0.1649 \text{ (Valor taula Poisson)} \quad \leftarrow 0,5 \text{ punts}$$

$$[ R \text{ ppois}(6,9.5) [1] 0.1649492 ]$$

5.2 Quina llei de distribució segueix la variable  $W \sim$  "Dies entre reclamacions"? Indica també el/s paràmetre/s d'aquesta llei. Quina és la probabilitat d'estar 2 dies o més dies sense rebre reclamacions?

$$W \sim \text{Exp}(\lambda=4,5) \quad \leftarrow 0,5 \text{ punts}$$

$$P(W >= 2) = 1 - P(W < 2) = 1 - F_W(2) = 1 - (1 - \exp(-4.5*2)) \quad \leftarrow 0,5 \text{ punts}$$

$$= 0.00012$$

**NOM:** \_\_\_\_\_

(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes).

Seguint amb l'exemple anterior d'una variable aleatòria  $X$  amb distribució normal que representa el temps de baixar una cançó d'un determinat servidor en Internet, ara considerem que conèixer l'esperança del temps de baixada i la desviació poblacionals és poc creïble.

**B.**

Durant 7 dies consecutius, i a la mateixa hora, s'ha baixat una cançó del servidor al disc dur, i s'ha observat el temps en segons que triga. Els resultats es mostren a continuació:

9,05 9,77 8,58 11,58 7,47 9,32 10,80

(amb alguns càlculs intermedis:  $\sum x = 66,57$   $\sum x^2 = 644,3715$ )

1) (0,5 punts) Doneu una estimació puntual de la mitjana i de la desviació del temps que triga en baixar una cançó.

$$\hat{\mu} = \bar{x} = 9,51; \hat{\sigma} = s = 1,371787$$

2) (0,5 punts) Amb aquesta mostra, ¿en quant estimeu l'error de la mitjana o error típic?

$$\text{error tipo} = 0,5184868$$

3) (2 punts) Estimeu per IC al 95% de confiança la mitjana poblacional.

$$IC(\mu, 1-\alpha) = (8.241308, 10.778692)$$

$$9.51 \pm t_{6,0.975} \frac{s}{\sqrt{7}} = 9.51 \pm 2.447 \frac{1.371787}{\sqrt{7}} = [8.24, 10.78]$$

4) (2 punts) Quin valor de  $n$  hauríem de recollir per obtenir un IC de la mitjana poblacional al 95% de confiança amb una amplada de 0,5 seg (en aquest cas sí que assumim una desviació poblacional de 2 sg)

$$n = 62$$

5) (2 punts) Estimeu per IC al 95% de confiança la desviació poblacional

$$IC(\sigma, 1-\alpha) = (0.884, 3.021)$$

6) (1,5 punts) Experiments realitzats amb anterioritat informen que el temps de baixada, utilitzant aquest servidor i aquest tipus d'ordinador, és en mitjana de 12 segons. Poseu a prova si aquesta afirmació es pot considerar creïble o no amb una PS. Considereu un risc del 5%.

- (0,25) Hipòtesis (indicant si la prova és bilateral o unilateral):

H:  $\mu = 12$  (bilateral)

- (0,25) Càlcul del valor de l'estadístic:

$t = -4.8024$

- (0,25) ¿Quin és el punt crític que utilitzareu per decidir si l'estadístic anterior permet creure'ns la hipòtesis?

$t_{crític} = 2,447$

- (0,25) Representeu gràficament en la distribució de l'estadístic el valor obtingut i el punt crític:

- (0,25) Decisió de PS:

Hay suficiente evidencia para rechazar la hipótesis de que el tiempo de bajada es 12 seg.

- (0,25) Relacionar la decisió anterior amb el resultat del IC de la pregunta 3

El valor 12 seg está fuera del IC

7) (1,5 punts) Se sap que la proporció de persones que no baixen cançons d'internet és d'un 20%. Es pren una mostra de 100 persones i s'obté que la proporció és del 22%. Hi ha evidència de que la proporció hagi augmentat? Per responder plantejeu un CH i considereu un risc del 5%.

- (0,25) Hipòtesis (indicant si la prova és bilateral o unilateral):

$H_0: \pi = 0,20$

$H_1: \pi > 0,20$

- (0,25) Càlcul del valor de l'estadístic:

$z = 0,50$

- (0,25) Càlcul del P-valor:

$p_{valor} = P(z > 0,50) = 0,3085$

- (0,25) Representeu gràficament en la distribució de l'estadístic el valor obtingut i el P-valor:

- (0,50) Decisió de CH i interpretació:

No hay evidencia para rechazar la hipótesis de que la proporción de personas que bajan música es del 20%.

**NOM:** \_\_\_\_\_

(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes).

**Problema 2. (tots els apartats valen igual)**

La diferència entre la velocitat de baixada contractada (VC) i la velocitat real (VR) en les connexions a Internet és una de les queixes més freqüents entre els usuaris de les línies ADSL. Els objectius d'aquest estudi són: (A) estudiar si són iguals (en mitjana) i estimar la magnitud del possible desfasament; i (B) establir una relació entre les dues. Coneixements previs aconsellen treballar amb el logaritme natural dels temps, anotats per C i R. Sigui  $D=R-C$ . Es disposa de les següents dades:

$$n=30 \quad \Sigma C=253 \quad \Sigma C^2=2150 \quad \Sigma R=245 \quad \Sigma R^2=2021 \quad \Sigma D=-7.6 \\ \Sigma D^2=5.7 \quad \Sigma CR=2080 \quad \Sigma CD=-63.5 \quad \Sigma RD=-65.3$$

**A.** Es pot considerar que els temps C i R són iguals? Per respondre-ho, fem inferència sobre la seva diferència:

1. D'acord amb l'enunciat, es tracta d'un disseny 'aparellat' o independent? Raoneu la resposta.

*Cada usuari té un 'parell' de dades, que venen 'juntos'.*

*Cal esperar que Ri s'assembli més a la seva parella Ci que a cap altre.*

*Cal esperar 'no independència' entre C i R*

2. Càlculu les mitjanes i desviacions típiques de C, R i D

$$\bar{R} = 245/30 = 8.167 \quad S_R^2 = 0.834^2 \\ \bar{C} = 253/30 = 8.433 \quad S_C^2 = 0.751^2 \\ \bar{D} = -7.6/30 = -0.253 \quad S_D^2 = (5.7 - ((-7.6)^2/30))/(30-1) = 0.1302 \rightarrow S_D = 0.3608$$

3. Calculeu la covariància i la correlació entre R i C

$$S_{RC} = (2080 - 253 \cdot 245/30)/29 = 0.477$$

$$r_{RC} = 0.477/(0.834 \cdot 0.751) = 0.7614$$

4. Sota la PS de si són iguals els temps C i R, escriviu la hipòtesi (indicant si és bilateral o unilateral), l'estadístic i la seva distribució

*Sota PS la hipòtesi és:  $H: \mu_D = 0$  (bilateral)*

$$\text{Estadístic: } \hat{t} = \frac{(\bar{D} - 0)}{S_D / \sqrt{n}} \rightarrow t_{n-1}$$

*Distribució estadístic:  $t_{29}$*

5. Sota PS, calculeu l'estadístic i (aproximat) el p-valor

$$\text{Estadístic} = (-0.253 - 0) / \sqrt{0.1302/30} = -3.846$$

$$P\text{-valor} = P(|t_{29}| > |3.846|) < P(|t_{29}| > |3.660|) = 2 \cdot (1 - P(t_{29} < 3.660)) = 2 \cdot (1 - 0.9995) = 0.001 \rightarrow P < 0.001$$

*Més exactament,  $P = 0.0006$*

6. Sota PS, indiqueu la conclusió i interpretació pràctica:

PS: donat  $P\text{-valor}=0.0006$ , NO és versemblant la igualtat

Conclusió pràctica: és pot dubtar que la velocitat de baixada real sigui igual a la velocitat de baixada contractada.

7. Calculeu un IC al 95% per a la mitjana de la diferència

$$IC(\mu_D, 0.95) = \bar{D} \pm t_{n-1, 1-\alpha/2} \sqrt{\frac{S_D^2}{n}} = -0.253 \pm t_{29, 0.975} \sqrt{0.1302/30} =$$

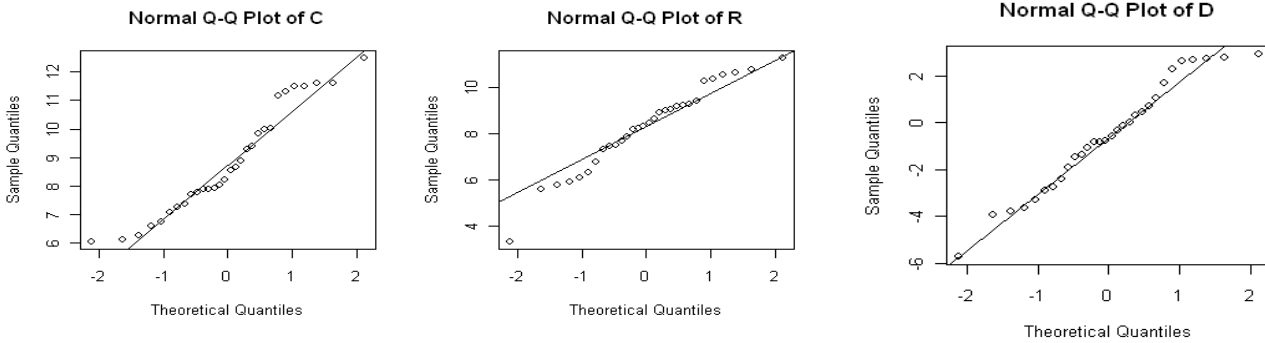
$$= -0.253 \pm 2.045 \cdot 0.066 = [-0.388 ; -0.119]$$

8. Interpreteu l'anterior IC

La diferència real entre les mitjanes dels logaritmes de la velocitat de baixada contractada i la velocitat de baixada real es troba dins del interval  $[-0.388 ; -0.119]$ .

No és versemblant la igualtat de les velocitats perquè el 0 no està inclòs dins del interval.

9. Indiqueu i valoreu la premissa de Normalitat. A quina variable s'aplica?



La premissa de normalitat s'aplica a la variable D.

Mirant el gràfic de la dreta, sembla raonable emprar el model Normal per estudiar aquesta variable (qualsevol de les tres s'ajusta bé a una Normal)

10. Interpreteu i comenteu globalment l'estudi d'aquest apartat A

En conjunt, les empreses donen una velocitat real que és més baixa que la contractada.

Com:

$$1 - e^{-0.253} \approx 1 - 0.776 = 0.224$$

$$1 - e^{-0.388} \approx 1 - 0.678 = 0.322$$

$$1 - e^{-0.119} \approx 1 - 0.888 = 0.112$$

Estimen que és en promig un 22.4% més baix, amb un interval d'incertesa entre un 32.2% i un 11.2%. Hem de tenir en compte, però, que els voluntaris no són una m.a.s. de tots els usuaris. El IC té en compte les oscil·lacions degudes a un atzar aleatori de mostreig, però no quantifica altres fluctuacions (com les degudes a un tipus especial d'usuari, hores, etc...).

**NOM:** \_\_\_\_\_  
 (Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes).

**B.** Si contractem C, què es pot esperar sobre R? Per respondre-ho, aplicarem el model de regressió lineal

1. Estimeu puntualment els coeficients de la recta de regressió:

$$b_1 = 0.7614 \cdot (0.834/0.751) = 0.8452 \quad b_0 = 8.167 - 0.8452 \cdot 8.433 = 1.0387$$

2. Ompleneu la taula de descomposició de variabilitat (indicant al costat el càlcul dels valors de la columna SQ):

	SQ	Graus lliberat GdL	QM = SQ/GdL	Rati	P-valor
<i>Explicada pel model</i>	$SQ_E = 0.8452^2 \cdot 29 \cdot 0.751^2 = 11.692$	1	$QM_E = 11.692$	$\hat{F} = 11.692/0.3027 = 38.63$	<0'001
<i>Residual</i>	$SQ_R = 20'167 - 11.692 = 8.475$	28	$QM_R = 8.475/28 = 0.3027$		
<i>Total</i>	$SQ_T = (30-1) \cdot 0.834^2 = 20'167$	29			

3. Estimeu i interpreteu el coeficient de determinació R<sup>2</sup>

$$R^2 = 11.692/20.167 = 0.5798$$

*La velocitat de baixada contractada explica en un 58% la variabilitat de la velocitat de baixada real.*

4. Poseu a prova H: V(R|C) = V(R): estadístic, distribució, càlcul, conclusió i interpretació:

*Estadístic:*  $\hat{F} = QM_E / QM_R = 38.63$  *Distribució sota H:*  $\hat{F} \rightarrow F_{1,28}$

*Conclusió:* Com  $P < 0.001$ , no és versemblant la hipòtesis que la velocitat de baixada contractada no aporti informació sobre la velocitat de baixada real.

*Interpretació:* La velocitat de baixada contractada explica en un 58% la variabilitat de la velocitat de baixada real.

5. Poseu a prova H:  $\beta_1 = 0$ : Estadístic, distribució, càlcul, conclusió i interpretació:

*Estadístic:*  $\hat{t} = \frac{(b_1 - \beta_1)}{\sqrt{\frac{s^2}{(n-1)S_c^2}}}$  *Distribució sota H<sub>0</sub>:*  $\hat{t} \rightarrow t_{28}$  Sent  $t_{28,0.975} = 2.048$

*Càlculs:*  $\hat{t} = (0.8452 - 0) / \sqrt{[0.3027 / ((30-1) \cdot 0.751^2)]} = 6.215$  *Rebutgem H amb P* ( $t_{28} > 6.215$ )  $< 0'0001$

*Conclusió pràctica:* El coeficient de la pendent ( $\beta_1$ ) és diferent de zero.

*Interpretació:* Una variació d'una unitat en C s'associa amb una variació de 0.845 (IC 0.567; 1.124) unitats en R.

6. Estimeu puntualment i per interval el valor de R predit per a un contracte de C=8.01 (ln(VC)=3010.9 KB/s)

$$\widehat{\ln(VR)} = b_0 + b_1 \cdot \ln(VC_0) = 1.039 + 0.845 \cdot 8.01 = 7.809$$

$$IC_{\ln(VC) = 8.01} (\ln(VR), 0.95) = \widehat{\ln(VR)} \pm t_{n-2,0.975} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(\ln(VC_0) - \bar{\ln(VC)})^2}{\sum_{i=1}^n (\ln(VC_i) - \bar{\ln(VC)})^2}} =$$

$$7.809 \pm 2.048 \cdot 0.550 \cdot \sqrt{1 + \frac{1}{30} + \frac{0.179}{16.367}} = [6.657 ; 8.960 ]$$

$$IC_{VC=3010,9}(VR, 0.95) = [e^{6.657}; e^{8.960}] = [778; 7789] \text{ KB/s}$$

7. De fet si el contracte es complís la constant seria 0 i la pendent 1. Creieu que es compleix  $H: \beta_1=1$  ?

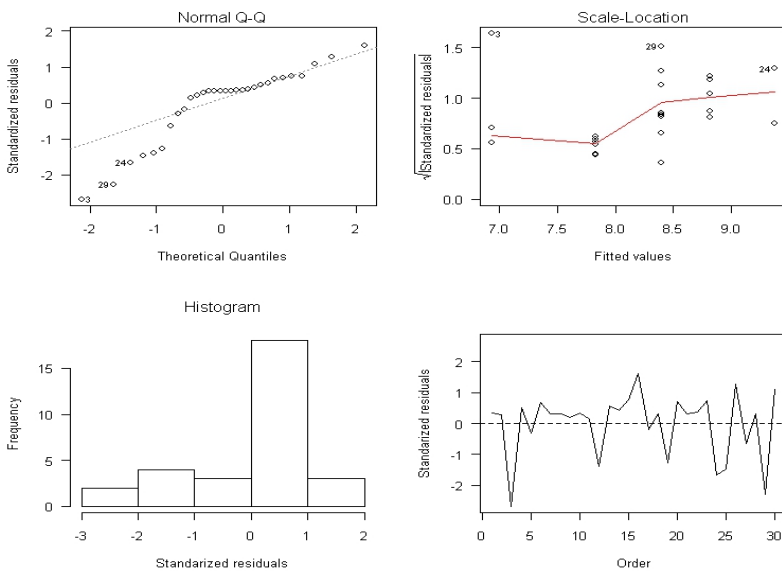
*Estadístics:* 
$$\hat{t}_{b_1} = \frac{(b_1 - \beta_1)}{\sqrt{\frac{S^2}{(n-1)S_c^2}}} \quad \text{Distribució sota } H_0: \hat{t} \rightarrow t_{28}$$

*Càlculs:* 
$$\hat{t}_{b_1} = (0'8452-1)/\sqrt{[0.3027/((30-1) \cdot 0'751^2)]} = -1.14$$

*Decisió:* No es pot rebutjar  $H_0$  amb  $P(t_{28} > |-1.14|) = 0.264$  (1.14 no és un punt extrem a una t Student)

*Conclusió pràctica:* El coeficient de la pendent ( $\beta_1$ ) podria no ser diferent de 1: no hi ha evidència per afirmar que el contracte no es compleix.

8. Valoreu les premisses.



*El gràfic QQ de Normalitat no mostra un bon ajustament. Més aviat mostra que la premissa de Normalitat pugui no aplicar a aquestes dades. El Histograma aniria en la mateixa línia.*

*De la mateixa manera, sembla que els residus tinguin valors absoluts més petits i més homogenis per valors predits al voltant de 7.8. (¿Es compleixen més els contractes en aquesta franja?)*

*Sí que sembla, en canvi, que les dades son independents entre elles.*

9. Interpreteu i comenteu globalment l'estudi d'aquest apartat B

*La velocitat de baixada contractada explica en un 58% la variabilitat de la velocitat de baixada real.*

*Una variació d'una unitat en C s'associa amb una variació de 0'845 unitats en la R (amb un IC que conté l'1).*

*Aquest resultats son encara temptatius perquè no es compleix la premissa de m.a.s. (són voluntaris) i la resta de premisses son molt dubtoses.*

10. Torneu a contestar la pregunta 1 de l'apartat A emprant ara la nova informació de que disposeu

*A la regressió hem rebutjat que C i R siguin independents, el que recolza l'aparellament de les dades: la velocitat real observada en un voluntari s'assembla més a la seva pròpia velocitat contractada que a la de un altre voluntari (afortunadament!!).*