

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

Problema 1 (Bloc C)

Totes valen 1 pt menys 1^a (0.5 pts) i 7^a (1.5 pts)

Un grup de recerca en imatges digitals està considerant un possible índex, una escala de 0 a 1000, per mesurar la nitidesa (S , de "Sharpness") d'una imatge, un tema important per diferenciar imatges reals d'altres generades per IA. Un dels passos inicials en la validació de l'índex consisteix en aplicar-ho a imatges reals per veure si es distribueix de forma coherent. Aquests són els índexs en una primera mostra a l'atzar de 5 elements: 75, 200, 250, 400, 600.

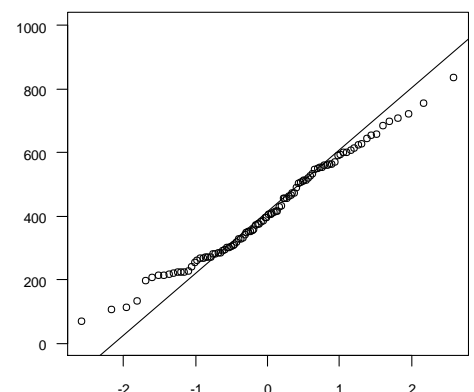
1) Heu sentit que les 5 imatges s'han obtingut agafant les 5 primeres imatges que han sortit posant a Google "random images". Critiqueu i/o milloreu aquest procediment d'obtenció de la mostra.

2) Amb la mostra disponible, calculeu una estimació per interval de confiança al 95% del valor mig de S .

3) Amb la mateixa mostra, calculeu una estimació per interval de confiança al 95% de la desviació tipus de S .

Una nova mostra més gran ($n=100$) ha resultat en un IC al 90% de confiança (365, 415) per a la esperança de S . 4) Obtingueu d'aquest resultat els valors de la mitjana i la desviació mostrals, i interpreteu l'interval de confiança.

5) Què és la figura de la dreta (obtinguda amb la segona mostra de S), i què ens diu?



Els investigadors d'aquest grup volen mostrar que les imatges generades per una IA tenen propietats diferents però que no són fàcils de percebre a simple vista, com la nitidesa S . Creuen que les imatges artificials són més nítides i molt més homogènies (menor dispersió). Per provar-ho, seleccionen aleatòriament 100 imatges creades amb una IA i apliquen el mateix índex que a les imatges reals, per analitzar els resultats anteriors amb els de les imatges IA.

6) Es tracta d'un estudi amb dades aparellades o amb dades independents? Raoneu la resposta.

En qualsevol cas, ara suposeu que es tracta de dues mostres independents. La segona mostra presenta un índex mitjà igual a 489 punts i una desviació tipus de 111 punts (si no heu resolt l'apartat 4, utilitzeu a partir d'ara els valors 400 i 125 com a mitjana i desviació tipus de la primera mostra).

7) Es demana un IC al 95% de confiança per a la diferència de mitjanes de S entre imatges IA i imatges reals. Comenteu sobre les premisses necessàries, i si creieu que són assumibles.

8) Es pot trobar l'error tipus de $\mu_{IA} - \mu_{real}$? I el de $\hat{\mu}_{IA} - \hat{\mu}_{real}$? Justifiqueu les respostes i expliqueu de què ens informa.

9) Avalueu la suposada major homogeneïtat de les imatges artificials respecte les reals i interpreteu.

10) L'índex pot ser emprat per classificar una imatge com a nítida o no nítida, segons algun llindar crític. Es vol determinar quantes imatges reals i quantes artificials s'han d'utilitzar a un estudi per trobar un IC 95% d'amplada 15% per la diferència de proporcions (o diferència de probabilitats) d'imatges nítides entre les dues tipologies. Tingueu en compte que, per determinades raons, en aquest estudi el nombre d'imatges generades per IA no pot ser més de la meitat que el nombre d'imatges reals.

qnorm(0.900) = 1.282	qnorm(0.975) = 1.960	qt(0.95,3)=2.353	qt(0.975,3)=3.182	qchisq(0.025,3)=0.216	qchisq(0.05,3)=0.352
qnorm(0.925) = 1.440	qnorm(0.990) = 2.326	qt(0.95,4)=2.132	qt(0.975,4)=2.776	qchisq(0.025,4)=0.484	qchisq(0.05,4)=0.711
qnorm(0.950) = 1.645	qnorm(0.995) = 2.576	qt(0.95,5)=2.015	qt(0.975,5)=2.571	qchisq(0.025,5)=0.831	qchisq(0.05,5)=1.145
qchisq(0.95,3)=7.815	qchisq(0.975,3)=9.348	qf(0.95,99,99)=1.3941	qf(0.975,99,99)=1.4862	qf(0.95,100,100)=1.3917	
qchisq(0.95,4)=9.488	qchisq(0.975,4)=11.14	qf(0.95,99,100)=1.3927	qf(0.975,99,100)=1.4844	qf(0.975,100,100)=1.4833	
qchisq(0.95,5)=11.07	qchisq(0.975,5)=12.83	qf(0.95,100,99)=1.3931	qf(0.975,100,99)=1.4850		

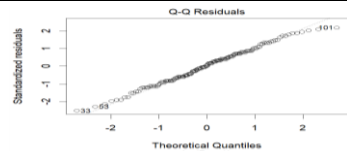
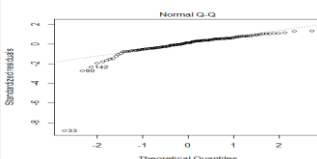
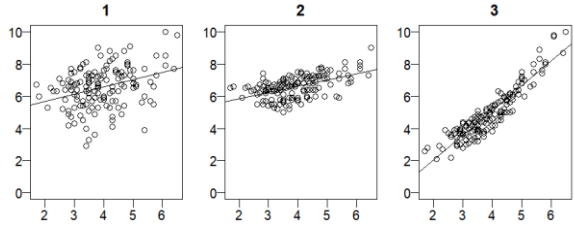
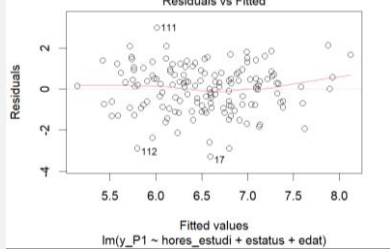
NOM: _____ COGNOM: _____

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

Problema 2 (Bloc D)

Nota: el separador decimal en tot l'exercici és el punt (".")

Per fer un estudi sobre els factors que poden predir el rendiment dels alumnes de PE, s'han recollit les següents variables corresponents a 150 alumnes: hores d'estudi setmanals, si va emprar e-status durant l'assignatura (sí/no) i l'edat a l'inici del curs. A més, s'han recollit les notes dels 2 exàmens parcials (y_{P1} i y_{P2}) que es consideren les variables que mesuren el rendiment. Finalment, es té una darrera variable calculada que és la diferència ($D = y_{P2} - y_{P1}$) de notes entre els exàmens parcials i la diferència de logaritmes ($LD = \log(y_{P2}) - \log(y_{P1})$). S'han ajustat amb R els següents 4 models:

Model 1	<p>Call: <code>lm(formula = D ~ 1)</code></p> <p>Coefficients:</p> <table><thead><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(> t)</th></tr></thead><tbody><tr><td>(Intercept)</td><td>-1.2093</td><td>0.1067</td><td>-11.34</td><td><2e-16 ***</td></tr></tbody></table> <p>---</p> <p>Residual standard error: [REDACTED] on 149 degrees of freedom</p>		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	-1.2093	0.1067	-11.34	<2e-16 ***																
	Estimate	Std. Error	t value	Pr(> t)																							
(Intercept)	-1.2093	0.1067	-11.34	<2e-16 ***																							
Model 2	<p>Call: <code>lm(formula = LD ~ 1)</code></p> <p>Coefficients:</p> <table><thead><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(> t)</th></tr></thead><tbody><tr><td>(Intercept)</td><td>-0.24500</td><td>0.03302</td><td>-7.419</td><td>8.3e-12 ***</td></tr></tbody></table> <p>---</p> <p>Residual standard error: 0.4044 on 149 degrees of freedom</p>		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	-0.24500	0.03302	-7.419	8.3e-12 ***																
	Estimate	Std. Error	t value	Pr(> t)																							
(Intercept)	-0.24500	0.03302	-7.419	8.3e-12 ***																							
Model 3	<p>Call: <code>lm(formula = y_P1 ~ hores_estudi)</code></p> <p>Coefficients:</p> <table><thead><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(> t)</th></tr></thead><tbody><tr><td>(Intercept)</td><td>4.7539</td><td>0.4066</td><td>11.69</td><td>< 2e-16 ***</td></tr><tr><td>hores_estudi</td><td>0.4554</td><td>0.1012</td><td>4.50</td><td>1.37e-05 ***</td></tr></tbody></table> <p>---</p> <p>Residual standard error: 1.185 on 148 degrees of freedom</p> <p>Multiple R-squared: 0.1204, Adjusted R-squared: 0.1144</p> <p>F-statistic: 20.25 on 1 and 148 DF, p-value: 1.365e-05</p>		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	4.7539	0.4066	11.69	< 2e-16 ***	hores_estudi	0.4554	0.1012	4.50	1.37e-05 ***											
	Estimate	Std. Error	t value	Pr(> t)																							
(Intercept)	4.7539	0.4066	11.69	< 2e-16 ***																							
hores_estudi	0.4554	0.1012	4.50	1.37e-05 ***																							
Model 4	<p>Call: <code>lm(formula = y_P1 ~ hores_estudi + estatus + edat)</code></p> <p>Coefficients:</p> <table><thead><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(> t)</th></tr></thead><tbody><tr><td>(Intercept)</td><td>4.70745</td><td>1.27938</td><td>3.679</td><td>0.000328 ***</td></tr><tr><td>hores_estudi</td><td>0.49170</td><td>0.09688</td><td>5.075</td><td>1.16e-06 ***</td></tr><tr><td>estatusSi</td><td>0.78500</td><td>0.19526</td><td>4.020</td><td>9.28e-05 ***</td></tr><tr><td>edat</td><td>-0.03132</td><td>0.06158</td><td>-0.509</td><td>0.611848</td></tr></tbody></table> <p>Residual standard error: 1.129 on 146 degrees of freedom</p> <p>Multiple R-squared: 0.2114, Adjusted R-squared: 0.1952</p> <p>F-statistic: 13.05 on 3 and 146 DF, p-value: 1.344e-07</p>		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	4.70745	1.27938	3.679	0.000328 ***	hores_estudi	0.49170	0.09688	5.075	1.16e-06 ***	estatusSi	0.78500	0.19526	4.020	9.28e-05 ***	edat	-0.03132	0.06158	-0.509	0.611848	
	Estimate	Std. Error	t value	Pr(> t)																							
(Intercept)	4.70745	1.27938	3.679	0.000328 ***																							
hores_estudi	0.49170	0.09688	5.075	1.16e-06 ***																							
estatusSi	0.78500	0.19526	4.020	9.28e-05 ***																							
edat	-0.03132	0.06158	-0.509	0.611848																							

- Emprant els models disponibles, quines són les estimacions puntual per a la diferència ($\mu_{y_{P2}} - \mu_{y_{P1}}$) i pel raci ($\mu_{y_{P2}} / \mu_{y_{P1}}$) de mitjanes del segon parcial respecte el primer? Interpreta-les.
- Emprant els models disponibles, calcula un interval de confiança (IC) del 95% per a la diferència de mitjanes de notes del segon parcial respecte el primer.
- Calcula la desviació estàndard residual del model 1.

> qt(0.90,148) = 1.287298	> qt(0.95,148) = 1.655215	> qt(0.975,148) = 1.976122
> qt(0.90,149) = 1.287259	> qt(0.95,149) = 1.655145	> qt(0.975,149) = 1.976013
> qt(0.90,150) = 1.287221	> qt(0.95,150) = 1.655076	> qt(0.975,150) = 1.975905
> qnorm(0.90) = 1.281552	> qnorm(0.95) = 1.644854	> qnorm(0.975) = 1.959964

4. Digues quina o quines premisses s'avaluen en els gràfics associats als models 1 i 2 i, basant-te en aquests gràfics, quin model sembla més idoni.
5. Segons el model 3, indica quina és la recta estimada pel model i la seva interpretació.
6. Segons el model 3, calcula un IC90% pel canvi esperat en la nota del primer parcial si incrementem en 2 hores el temps d'estudi setmanal, i interpreta'l.
7. Els 3 gràfics associats al model 3 representen les notes del parcial 1 (eix vertical) en funció de les hores d'estudi (eix horitzontal). Basant-te en la informació del model 3, digues quin del gràfics (1, 2 o 3) representa les dades reals. Argumenta la resposta.
8. Emprant el model 4, fes una predicció puntual de la nota del primer parcial per un estudiant de 20 anys que fa servir e-status i que estudia 5 hores a la setmana.
9. Interpreta el R^2 del model 4.
10. Digues quina o quines premisses s'avaluen al gràfic associat al model 4 i argumenta segons el que es veu si aquesta o aquestes premisses es compleixen i el perquè.