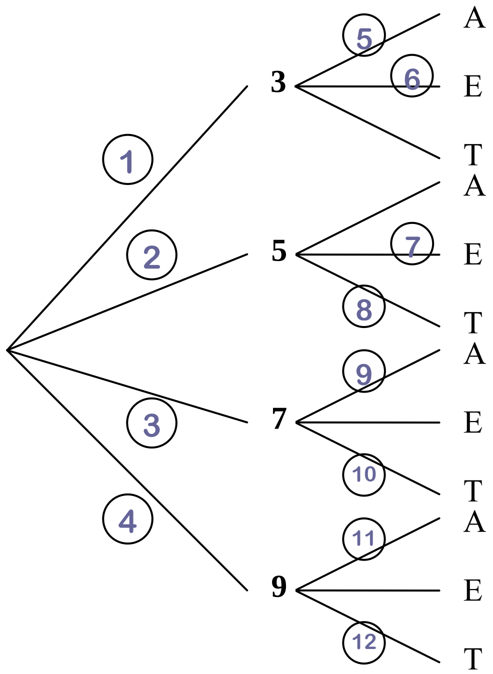
El Treball Final d'Estudis a Hogwarts (el col·legi on va estudiar Harry Potter) té tres possibles especialitats que l'alumne pot escollir: A: Arts Oscures; E: Encanteris; T: Transformacions, i la qualificació final pot ser 3 (suspés), 5 (aprovat), 7 (notable) o 9 (excel·lent). Com que el col·legi mai ha canviat el seu pla d'estudis disposa de molta informació dels resultats obtinguts en el passat, resumits a aquesta taula (valors relatius a 1000 treballs):

	3	5	7	9
A	36	184	45	25
E	10	60	94	136
T	38	80	238	54

Preg. 1) Ompliu la taula següent, d'acord amb les probabilitats indicades a la figura de la dreta (amb nombres dins d'un cercle). Poseu les probabilitats en forma de fracció. (3 pts)

	Forma algebraica	Solució
1	P(3)	84/1000
2	P(5)	324/1000
3	P(7)	377/1000
4	P(9)	215/1000
5	P(A | 3)	36/84
6	P(E | 3)	10/84
7	P(E | 5)	60/324
8	P(T | 5)	80/324
9	P(A| 7)	45/377
10	P(T | 7)	238/377
11	P(A | 9)	25/215
12	P(T | 9)	54/215



Preg. 2) Sabem d'un alumne que ha tret en el seu Treball Final d'Estudis un 7 almenys. Quina és la probabilitat que no hagi escollit Encanteris? (1 pt)

P("E" | nota=>7) = P("E" i nota =>7) / P(nota =>7) = (94+136)/(377+215) = 0.3885
P(¬"E" | nota=>7) = 1–P("E" | nota=>7) = **0.6115**

Preg. 3) Quina és la nota mitjana dels alumnes en el Treball Final d'Estudis? I si només prenem els que aproven? (1 pt)

Nota mitjana = (3*84+5*324+7*377+9*215)/1000 = 6.45
Si només prenem els que aproven serà
Nota mitjana aprov =(5*324 + 7*377+9*215)/916 = 6.76

Preg. 4) Suposem que un alumne no ha aprovat. Quina és l'especialitat que més probablement va escollir, i perquè? (1 pt)
S'han de calcular les 3 probabilitats i triar aquellla que tingui el valor més alt
P(A | "3") = P(A i "3") / P("3") = 36/84 = 0.43
P(E | "3") = P(E i "3") / P("3") = 10/84 = 0.12
P(T | "3") = P(T i "3") / P("3") = 38/84 = 0.45
Per tant, sembla ser que él més probable és que hagi escollit T: Transformacions
Preg. 5) Calculeu el valor esperat, la variància i la desviació estàndar de la qualificació a alguna de les especialitats disponibles. (2 pts)

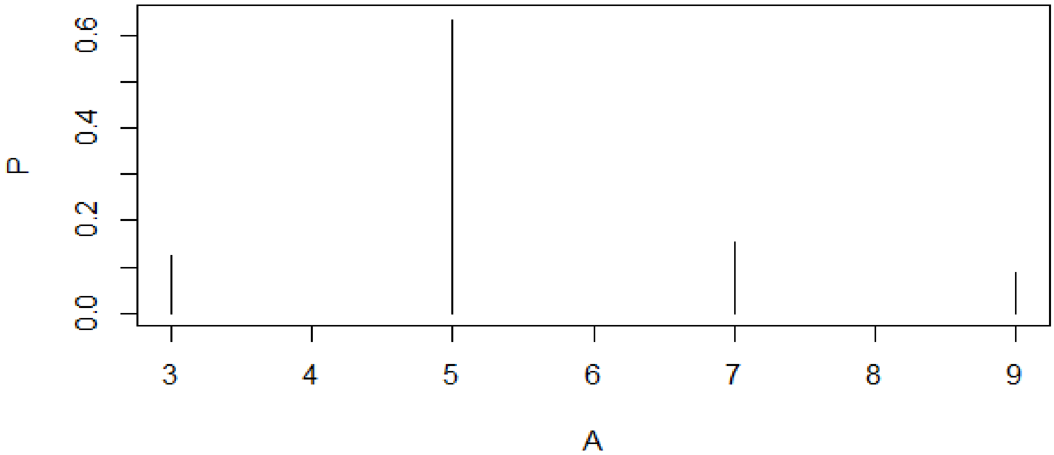
Prenem com especialitat A: Arts Oscures
El valor esperat serà E(A) = 3*36/290 + 5*184/290 + 7*45/290 +9*25/290 = 5.41

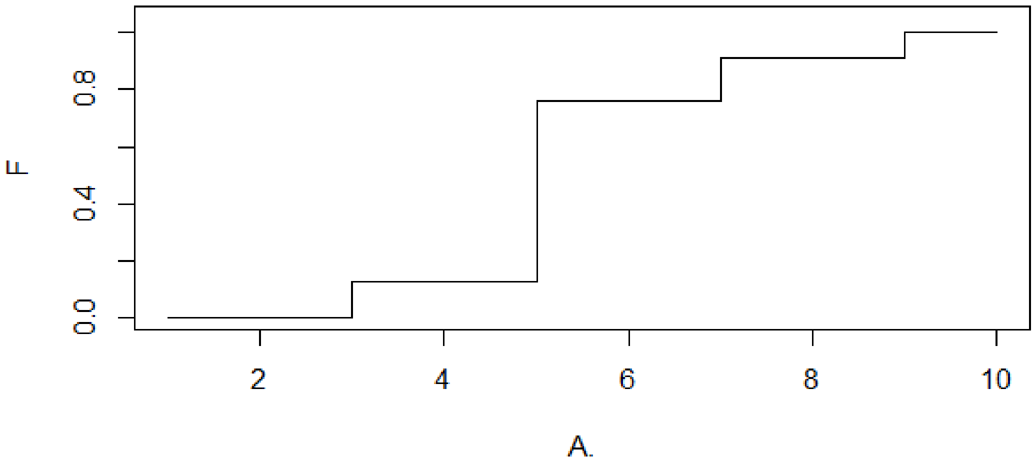
$E(A^2) = 3*3*36/290 + 5*5*184/290 + 7*7*45/290 + 9*9*25/290 = 31.57$
$V(A) = 31.57 – 5.41^2 = 2.33$
$\sigma = 1.527$

Preg. 6) Per a aquesta mateixa especialitat, representeu tabular i gràficament la funció de probabilitat i la funció de distribució. (1 pt)

Funció de probabilitat

Valors	Prob	Distr.
<3	0	0
3	0,12	0.12
5	0,63	0.76
7	0,16	0.91
9	0,09	1.00





Preg. 7) El col·legi utilitza un índex I = 100×(Q–3)/6, on Q és la qualificació descrita anteriorment i que pot prendre els valors 3, 5, 7, i 9. Què valen E(I) i V(I)? Per respondre no cal construir la funció de probabilitat de l'índex I. (2 pts)

$I = 100/6\ Q – 50$
$E(I) = 100/6\ E(Q) – 50 = 100/6\ 6.45 – 50 = 57.43$
$V(I) = (100/6)^2\ V(Q) = 886.97$

Una altra forma (no recomanada):
Q={ 3, 5, 7, 9}
I={ 0, 33.3, 66.7, 100}
Prob= {0.084, 0.324, 0.377, 0.215}

Q	I	prob	E(I)	V(I)
3	0	0.084	0	277.08
5	33,3	0.324	10.8	188.18
7	66,7	0.377	25.13	32.14
9	100	0.215	21.5	389.56
			57.43	886.97

Preg. 8) Justifiqueu que la qualificació al Treball Final <u>no</u> és independent de l'especialitat que segueix l'alumne. Feu una interpretació que es derivi d'aquest fet. (2 pts)

Per exemple, la distribució de la nota Q (global) no és la mateixa que per a l'especialitat A:
P(Q=3,5,7,9) = {0.084, 0.324, 0.377, 0.215}
P(Q=3,5,7,9 | A) = {0.12, 0.63, 0.16, 0.09}
Així, si un alumne fa l'especialitat A té moltes probabilitats que obtingui un 5, però en general la nota més probable és un 7.

NOM:____

Poseu el nom. Contesteu en el seu lloc reservat. Expliciteu i justifiqueu passos i càlculs.

Problema 2 (B3 i 4) Totes valen igual.

Els estudiants amb notable a la nota final tarden en resoldre els problemes d'estatus un temps variable amb una esperança de 20'. Contesteu les següents preguntes assumint que la seva variabilitat la podem modelar mitjançant:

X: una distribució Normal amb una desviació típica de 5';
Y: una exponencial; i
Z: una uniforme amb un límit inferior de 5'.

Dona els valors de les desviacions típiques de la exponencial Y i la uniforme Z:

1. Y

$20 = E(Y) = 1/\lambda \quad \rightarrow \quad \lambda = 0.05$

$V(Y) = 1/\lambda^2 \quad \rightarrow \quad \sigma_Y = 1/0.05 = 20$

2. Z

Per simetria, límit superior b = 35

$V(Z) = (b\text{-}a)^2/12 = 75 \rightarrow \quad \sigma_Y \approx 8.66$

Per cada distribució de probabilitat, amb una garantia del 95%, podem assegurar que el temps serà inferior a...?

3. X

$0.95 = P(X \leq x) = 1\text{-}P(Z \leq (x\text{-}20)/5) = 1\text{-}P(Z \leq 1645) \rightarrow$

$x = 20+1.645*5 \approx 28.225$

4. Y

$0.95 = 1 - e^{-1/20 * x} \quad \rightarrow \quad 0.05 = e^{-1/20 * y} \quad \rightarrow$

$\ln(0.05) = y/20 \quad \rightarrow \quad y = - 20 * \ln(0.5) \approx 59.9$

5. Z

$0{,}95 = (z\text{-}5)/(35\text{-}5) \rightarrow \quad 0.05 = e^{-1/20 * y} \quad \rightarrow$

$Z = 5 + 30*0.95 \approx 33.5$

Aquests alumnes fan uns 50 problemes per curs i es vol estudiar el temps total empleat. Doneu les distribucions (amb esperances i variàncies) resultants en els 2 primers casos (X, Y)

6. $SUM_{50}(X)$

$E(SUM_{50}(X)) = 50*E(X) = 1000$
$V(SUM_{50}(X)) = 50*V(X) = 50*25 = 1250$
$SUM_{50}(X) \rightarrow N\ (1000, 1250)$

7. $SUM_{50}(Y)$

$E(SUM_{50}(Y)) = 50*E(Y) = 1000$
$V(SUM_{50}(Y)) = 50*V(Y) = 50*400 = 20000$
$SUM_{50}(Y) \rightarrow N\ (1000, 20000)$

Pels 2 primers, entre quins 2 valors simètrics oscil·la el temps total que necessita el 95% central dels estudiants

8. X

$0.95 \text{ central} \in \mu \pm z_{1-\alpha/2}\, \sigma \approx 1000 \pm 1.96 * \sqrt{1250} = 1000 \pm 69.3 \approx$ **(930.7, 1069.3)**

9. Y

$0.95 \text{ central} \in \mu \pm z_{1-\alpha/2}\, \sigma \approx 1000 \pm 1.96 * \sqrt{20000} = 1000 \pm 277.2 \approx$ **(732.8, 1277.2)**

D'altra banda, hem recollit el temps de 6 problemes i aquest ha estat 3, 5, 7, 21, 38 i 58.

10. Calculi la mitjana

$\Sigma x = 132 \rightarrow$ $\boxed{\overline{X}} = 132/6 = 22$

11. Calculi la desviació tipus.

$\Sigma x^2 = 5332 \rightarrow$ $S^2 = (5332 - 22^2/6)/5$ **≈ 485.6**; → **S ≈ 22.04**

12. Mirant les dades, la mitjana i la desviació tipus, quina distribució (Normal, Expon. o Unif.) creu més raonable?

La exponencial, ja que (1) les dades es van dispersant progressivament; i (2) desviacions típiques semblants a les mitjanes així ho indiquen.

Faci l'IC80% de la mitjana poblacional

13. Estadístic, distribució i premisses necessàries (1p)

$$\hat{t} = \frac{(\bar{d} - \mu)}{s/\sqrt{n}} \rightarrow t_{n-1}$$

σ desconeguda i X ~ N

14. Càlcul i interpretació

IC (μ, 0.80) $= \boxed{\overline{X}} \pm t_{n-1,1-\alpha/2}$ $s/\sqrt{n} = 22 \pm 1.476 * 22.04/\sqrt{6} \approx 22 \pm 13.28 \approx$ **8.72, 35.28)**

Amb una confiança del 80%, E(D) és algun valor entre 10.5 i 33.5

Faci l'IC90% de la variància poblacional.

15. Estadístic, distribució i premisses necessàries

$(n-1)*S^2/\sigma^2 \sim \chi^2_{n-1}$ σ desconeguda i D ~ N

16. Càlcul

$$IC(\sigma^2, 0.90) = \left(\frac{s^2(n-1)}{\chi^2_{5,0.95}}, \frac{s^2(n-1)}{\chi^2_{5,0.05}} \right) \approx \left(\frac{485.6*5}{11.070}, \frac{485.6*5}{1.145} \right) \approx (219.33\ ,\ 2120.524)$$

17. Si volem obtenir un IC95% de la mitjana del temps que tingui una amplitud total (de límit superior, LS, a inferior, LI) que sigui la meitat de la desviació tipus, quants casos haurem de recollir? Assumiu σ coneguda

$\sigma/2 = LS - LI = 2\ Z_{1-\alpha/2}\ \ \sigma/\sqrt{n} \rightarrow n = [(2\ Z_{1-\alpha/2}\ \ \sigma)/\ (\sigma/2)]^2 \approx (4*1.96)^2 \approx 61.4656 \rightarrow n=62$

D'altra banda, s'ha observat que només 1% dels alumnes no tenen cap errada en e-status. Si tenim 171 alumnes

18. Quin és el model exacte de distribució i els paràmetres de la variable R~ "nombre d'alumnes amb cap fallo?

$R \sim B(n=171, \pi=0.01)$

19. Quina és la probabilitat exacte de no trobar cap alumne amb cap fallo?

$$P(R=0) = \binom{n}{k}\pi^k(1-\pi)^{n-k} = 1 \bullet 1 \bullet 0.99^{171} \approx 0.179316 \approx 18\%$$

20. Aproximi la solució del punt 19 amb la distribució de Poisson.

$\lambda = n \cdot \pi = 171 \cdot 0.01 = 1.71$

$P(R=0) = \lambda^k \cdot e^{-\lambda}/k! = 1.71^0 \cdot e^{-1.71}/0! = 1 \cdot e^{-1.71} \cdot 1 \approx 0.180866 \approx 18\%$

FIB Q1 2012-13 FINAL PE 21 de gener 2013

Poseu el nom. Contesteu en el seu lloc reservat. Expliciteu i justifiqueu passos i càlculs.

B5-B6

Totes les preguntes valen 1 punt excepte la pregunta 8 que val 2 punts

Una enquesta passada al mateix col·lectiu de professionals en dues universitats diferents (A i B) ha obtingut que aquests professionals han invertit en la seva formació (Grau + Master + Doctorat) en promig 13 anys a la Universitat A amb una desviació tipus de 1.2 anys mentre que a la Universitat B han estat 14 anys amb una desviació tipus de 1.5 anys. Es vol verificar si els temps en formació en ambdues universitat és igual o diferent, tenint en compte que el nombre d'enquestats ha estat 61 a cada universitat.

Pregunta 1: Plantegeu la prova d'hipòtesi adient per verificar si la variabilitat en la seva formació és similar en ambdues universitats, especificant:

Hipòtesis (indicar si la prova és unilateral o bilateral):

H_0: $\sigma_1^2 = \sigma_2^2$

H_1: $\sigma_1^2 \neq \sigma_2^2$

És una prova d'hipòtesi bilateral.

Estadístic:

F=$\frac{S_1^2}{S_2^2}$

Distribució de l'estadístic:

F~ $F_{n1-1,n2-1}$, és a dir F segueix una distribució F de Fisher-Snedecor amb n_1-1=61-1=60 graus en el numerador i n_2-1=61-1=60 graus de llibertat en el denominador

Càlcul del valor de l'estadístic i del p_valor (i la seva representació gràfica):

F=$\frac{1.5^2}{1.2^2} = \frac{2.25}{1.44} = 1.56$

Treballant amb un risc del 5%, el punt crític de la $F_{60,60}(0.975)$=1.67. Com que 1.56<1.67, no es pot rebutjar la H_0 i per tant es pot assumir que les dues variances poblacionals son similars

Pregunta 2: Plantegeu i resoleu la prova d'hipòtesi mitjançant l'interval de confiança per verificar si en promig la seva formació és similar en ambdues universitats, especificant:

Hipòtesis (indicar si la prova és unilateral o bilateral):

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

És una prova d'hipòtesi bilateral.

En primer lloc, cal calcular la variancia promitjada o $s_{pooled}^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$, que en aquest cas és

$$s_{pooled}^2 = \frac{(61-1)1.5^2+(61-1)1.2^2}{61+61-2} = \frac{60*(2.25+1.44)}{120} = 1.85$$

A continuació, el valor de l'estadístic *t*:

$$\hat{t} = \frac{\overline{y_1}-\bar{y}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{14-13}{\sqrt{1.85}\sqrt{\frac{1}{61}+\frac{1}{61}}} = \frac{1}{1.36*0.18} = 4.08$$

Distribució de l'estadístic:

$$\hat{t} \sim distribució\ t\ student\ amb\ n1+n2-2\ graus\ de\ llibertat, és\ a\ dir\ 120\ graus\ de\ llibertat$$

Càlcul de l' interval de confiança al 95%:

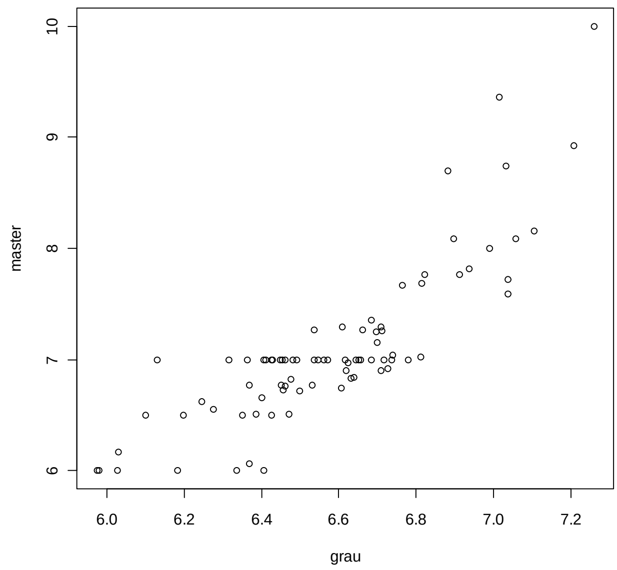
1±1.98*$\sqrt{1.85}\sqrt{\frac{1}{61}+\frac{1}{61}}$=1±1.98*1.36*0.18=1±0.48.

És a dir, l'interval de confiança al 95% és (0.52,1.48)

El col·lectiu de la Universitat A vol verificar si les notes que vam treure els estudiants al Grau eren equivalents o no a les notes obtingudes al Màster. Tot aquest col·lectiu ha realitzat el Grau i Màster a la mateixa universitat i s'ha observat que:

N (nombre d'estudiants) = 81
Mitjanes obtingudes en el Grau i Màster: 6.6 i 7.1 respectivament
Variancies obtingudes en el Grau i Màster: 0.08 i 0.53 respectivament
Correlació entre les notes de Grau i Màster: 0.85

Ajusteu una recta de regressió que permeti predir les notes obtingudes al Màster en funció de los notes obtingudes al Grau, calculant els següents valors i dibuixant la recta damunt de la figura 1:



Començarem amb l'estimació de **b1:**

b1=$\frac{s_{XY}}{s_X^2} = r\frac{s_Y}{s_X} = 0.85\frac{\sqrt{0.53}}{\sqrt{0.08}} = 2.19$

L'estimació de **b0 és**

b0: $= \overline{Y} - b_1\overline{X} = 7.1 - 2.19*6.6 = -7.354$

Plantegeu la prova d'hipòtesis adient per detectar si el pendent és significatiu o no amb un risc del 5% (la variancia residual s^2=0.3914^2=0.15)
Informació addicional:
qt(0.975,80)=1.99
qt(0.95,80)=1.66
qt(0.025,80)=-1.99
qt(0.05,80)=-1.66

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

Estadístic:
$\frac{(b_1-\beta_{1/0})}{s_{b1}} \sim t_{n-2}$, on $s_{b1} = \sqrt{s^2/(n-1)s_x^2}$ i s^2 és la variancia residual

La variancia residual"= 0.3914^2=0.15 i per tant

$$s_{b1} = \sqrt{s^2/(n-1)s_x^2} = \sqrt{\frac{0.15}{(81-1)*0.08}} = 0.15$$
$$\hat{t} = \frac{b_1}{s_{b1}} = \frac{2.19}{0.15} = 14.6$$

Com que $\hat{t}$ >qt(0.975,80)=1.99, rebutgem la Ho i per tant podem afirmar que el paràmetre β_1 és significativament diferent de zero.

Doneu el valor del coeficient de determinació R^2 i interpretar-lo

El coeficient de determinació R^2 és equivalent en aquest cas al coeficient de correlació al quadrat: r^2, que és 0.85^2 = 0.72. El que vol dir que el 72% de la variabilitat de les notes del Master ve explicada per les notes de Grau.

Doneu una predicció de quina serà, en promig, la nota de Màster pels estudiants que al Grau varen obtenir un 6.5 així com el seu interval de confiança al 95%

Predicció: $\widehat{Master} = b0 + b1 * Grau = -7.354 + 2.19 * 6.5 = 6.881$

Interval de confiança (es pot utilitzar la informació addicional anterior):
Informació addicional: la variancia residual s^2=0.3914^2=0.15

$$\hat{Y} \pm t_{n-2,0.975} s \sqrt{\frac{1}{n} + \frac{(X_h - \overline{x})^2}{\sum(x_i - \bar{x}^2)}} = 6.881 \pm 1.99 * 0.3914 * \sqrt{\frac{1}{81} + \frac{(6.5-6.6)^2}{(n-1)s_x^2}}$$
$$= 6.881 \pm 1.99 * 0.3914 * \sqrt{0.0123 + \frac{0.01}{80 * 0.08}} = 6.881 \pm 1.99 * 0.3914 * 0.12 = 6.881 \pm 0.125$$

L'interval de confiança al 95% és (6.756, 7.006). És a dir, la previsió de les notes de Màster obtingudes pels estudiant que van obtenir en promig al Grau 6.5 oscil·len des de 6.756 a 7.006.

Indiqueu quines premisses s'haurien de verificar per validar el model, quins del següents gràfics les validen i quins no. Justifiqueu la vostra resposta

Les premisses a verificar son Independència, homoscedasticitat i normalitat dels residus, més la linealitat entre les variables.

- La independència es pot afirmar a partir del gràfic "residuals vs Time", no sembla haver-hi cap tipus d'estructura
- La normalitat es pot deduir a partir dels gràfics QQ-plot i histograma, tot i que hi ha uns quants valors atípics, de 0.5 cap endavant que fa que no es pugui afirmar amb rotunditat que es compleixen aquestes premisses
- Observant el gràfic "residuals vs fitted values" es veu ben clar que la distribució d'aquests residus te una forma de V i no un núvol de punts sense estructura. Aquest comportament ve induït per la relació entre Notes de Grau i Màster (figura 1). Sembla que la millora en la nota de Màster és més ràpida quan l'alumne té bona nota de Grau (apreciable a partir de la nota de Grau = 6.6), com si la relació no fos lineal. Igualment, és bastant probable que la variància dels residus sigui major per la part dreta del núvol, quan les notes de Màster són altes.