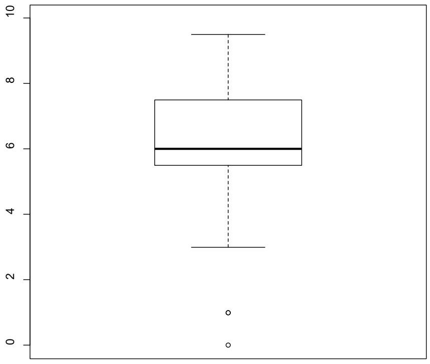
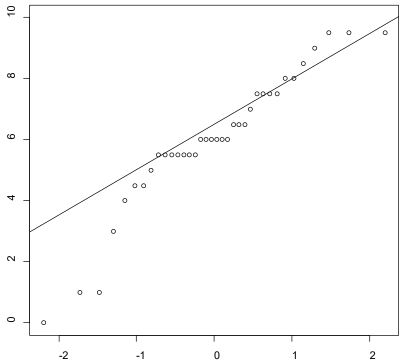
SOLUCIÓ

Problema 1 (B4)

Tenim les notes x_i de 36 estudiants d'una assignatura i es vol estudiar si el seu rendiment és comparable al que es venia obtenint en cursos anteriors.





$\sum x_i = 216$ $\sum x_i^2 = 1473$

1) (2.5 punts) Suposem que la desviació tipus de cursos anteriors (2.5 punts) equival a la desviació tipus poblacional:
a) Calculeu un interval de confiança pel valor esperat de la nota de la assignatura amb una confiança del 95%

σ = 2.5 punts

mean(x) ± $z_{0.975}$ σ / sqrt(36) = 216/36 ± 1.96 2.5 / sqrt(36) = 6 ± 0.8167 = **[5.18 , 6.82]**

(amplada IC és 1.633)

b) Trobeu el nombre d'observacions necessari per tenir un IC del valor esperat amb amplada igual a 1.2 i la mateixa confiança del 95%

amplada_total /2 = $z_{0.975}$ σ / sqrt(n) 0.6 = 1.96 2.5 /sqrt(n) sqrt(n) = 8.17 n = 66.69 **mínim 67 observacions**

2) (2.5 punts) L'equip docent no sembla convençut i opta per no assumir una desviació coneguda; aleshores, calculeu:
a) un IC de μ amb confiança del 95% i un amb confiança del 99%

Desviació mostral s = sqrt((1473 - 216^2/36) / 35) = **2.25** (sd(x))

mean(x) ± $t_{35,0.975}$ sd(x) / sqrt(36) = 6 ± 2.030 2.25 / sqrt(36) = 6 ± 0.76 = **[5.24 , 6.76]**

mean(x) ± $t_{35,0.995}$ sd(x) / sqrt(36) = 6 ± 2.724 2.25 / sqrt(36) = 6 ± 1.02 = **[4.98 , 7.02]**

b) comentar com i perquè han canviat els IC anteriors i també respecte el IC de l'apartat a) del punt 1)

L'IC amb confiança 99% és més ample que el de 95% perquè al voler assegurar més, s'eixampla el rang de valors i som menys precisos.

Aquests dos IC, comparats amb el del punt 1, presenten una situació una mica més complexa. Per una banda utilitzen un valor d'una distribució t-Student en lloc d'una normal, que sempre són més grans i per tant fan que l'interval sigui més ample. Però per l'altra banda casualment la desviació a aquesta mostra es més petita que la desviació poblacional assumida, pel que l'interval és més estret. Globalment, l'interval al 95% del punt 2 és més estret (més precís) que el del punt 1, i si comparéssim l'interval al 99% potser passaria el mateix.

4) (1 punts) Calculeu un IC al 95% per la desviació tipus. Trobeu discutible que s'assumeixi el valor de 2.5 punts que s'havia pres per la desviació tipus poblacional?

[$((sd(x))^2 (n-1) / \chi^2_{35,0.975}$, $((sd(x))^2 (n-1) / \chi^2_{35,0.025}$] = [2.25^2 35 / 53.2033* , 2.25^2 35 / 20.5694*] = [3.327 , 8.605]

La variància està entre 3.327 i 8.605 punts al quadrat, amb confiança del 95%
La desviació està entre **1.82 i 2.93** punts (arrels quadrades dels valors de la variància anteriors)

És raonable assumir un valor de 2.5 punts per a la desviació: 2,5 està en el IC [1.82,2.93]. Per tant, dubtar del valor 2,5 o discutir que pugui ser el valor de la desviació poblacional no sembla tenir molts arguments.

5) (4 punts) Finalment decidim fer una Prova d'Hipòtesis sobre si es pot acceptar que la nota mitjana està entorn el 5 (no s'assumeix cap valor per la variabilitat poblacional).
a) Realitzeu la PH adient:
Hipòtesis, estadístic i distribució:

H_0: μ **=** 5
H_1: μ **≠** 5

t = $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ amb una distribució t_{n-1}

Càlcul de l'estadístic i punt crítics:

t = (mean(x) – 5) / (sd(x) / sqrt(36)) = (6 - 5) / 0.3748 = **2.668**

$t_{35,0.975}$ = **2.0301** (entre **-2.0301** i **2.0301**** hi ha la zona d'acceptació, o no rebuig, de H_0)

Conclusió:

El valor de l'estadístic no està dins de la zona d'acceptació, per tant podem rebutjar que la nota mitjana es manté entorn el 5

b) Sospitant que el rendiment ha millorat, si ens plantegéssim la PH per decidir si la nota mitjana d'aquests estudiants és igual o és superior al valor 5, indica en la nova prova els apartats anteriors que canvien i comenta què implica fer una o altra prova.

Estadístic i distribució, i càlcul de l'estadístic, NO CANVIEN

Hipòtesis i punt crític i conclusió SÍ CANVIEN:
H_0: μ **=** 5
H_1: μ **>** 5

$t_{35,0.95}$ = **1.6896** (fins a 1.6896 hi ha la zona d'acceptació de H_0, valors més gran componen la zona de rebuig)

El valor 2.668 continua estant fora de la zona d'acceptació, per tant no acceptem que la nota mitjana es manté en el 5 sinó que seria superior.

La prova unilateral parteix de sospitar que el rendiment ha millorat i, per tant, planteja unes hipòtesis més concretes, el risc es reparteix diferent baixant el punt crític a partir del qual estem disposats a no acceptar la hipòtesi nul.la. Per tant, amb la prova unilateral es pot rebutjar la hipòtesi nul·la amb una diferència menor.

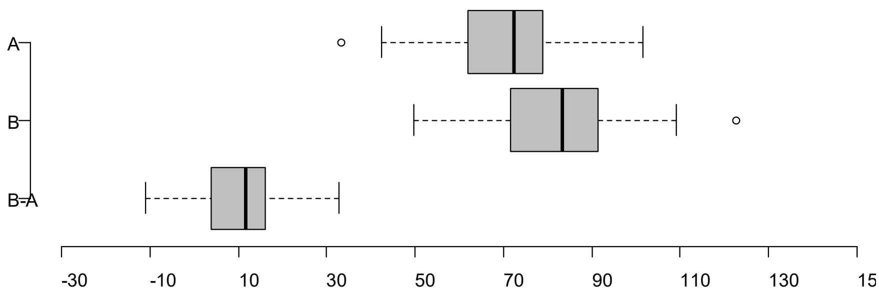
* Els valors presents de la khi-quadrat són exactes. Quan no es poden obtenir amb R ni emprar les taules es pot utilitzar l'aproximació a una N(df, 2df), on df són els graus de llibertat de la variable. En aquest cas els quantils que es prendrien són: 35 +/- z_0.975 sqrt(70) = [18.6, 51.4]. L'aproximació és millor quan més gran és df.
L'interval de confiança per la desviació seria (1.86, 3.09)

** Els valors presents de la t-Student són exactes. La distribució amb 35 graus de llibertat no hi és a les taules, però es podria fer una interpolació amb els valors de 30 i 40 que sí que hi són (2.042 i 2.021, respectivament)

Problema 2 (B5)

A Google Analytics (GA), hi ha una eina per assignar a certes pàgines un determinat import econòmic per visita (per exemple, a una pàgina de petició de servei de 20 euros que el contracten 1 de cada 20 persones, es pot assignar el valor de 1 euro). Els estudiants han dissenyat un experiment en que durant 2 mesos (61 dies) recullen els guanys estimats segons dos prototipus de pàgina (A i B). La descriptiva es mostra a continuació:

	A	B	B-A
Mitjana	70.82	81.66	10.83
Desv. Tipus	13.66	14.09	10.67
Mín	33.18	49.81	-11.01
Q1	61.85	71.57	3.93
Mediana	72.29	83.20	11.53
Q3	78.98	91.32	16.01
Màx	101.50	122.62	32.81



- Justifiqueu si es tracta de dades aparellades o independents (1 punt)

Aparellades pel dia (uns mateixos dies es recullen les dades del prototipus A i del B)

- Plantegeu una prova d'hipòtesis per decidir si la mitjana de guanys diaris és la mateixa en ambdós dissenys o si en el cas B són majors. Indiqueu les hipòtesis i si ha de ser unilateral o bilateral (1 punt)

$$\begin{cases} H_0: \mu_d = 0 \\ H_1: \mu_d > 0 \end{cases} \text{(unilateral)}$$

- Indiqueu l'expressió de l'estadístic, la seva distribució sota H_0 i les premisses assumides (1 punt)

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \sim t_{60}$$ Premisses: d = B-A ~ N i m.a.s

- Calculeu el valor de l'estadístic i el/s punt/s crític/s suposant un α = 0.05 (1 punt)

$$t = \frac{10.83}{10.67/\sqrt{61}} = 7.93$$
$$Punt\ crític \rightarrow t_{0.95,60} = 1.671$$

- Indiqueu, justifiqueu i interpreteu quina és la conclusió de la prova d'hipòtesis? (1 punt)

Hi ha evidència per rebutjar H_0,
ja que el valor de l'estadístic està més enllà del punt crític
Per tant, hi ha prou evidència per dir que la mitjana dels guanys són superiors en el cas B

Ara suposem que les dades anteriors no s'han recollit els mateixos dies pel cas A i B sinó en diferents mesos, i plantegem la prova d'hipòtesis per decidir si la mitjana de guanys diaris és la mateixa en ambdós prototipus o no.
- Indiqueu les hipòtesis i si ha de ser unilateral o bilateral (1 punt)

H_0: $\mu_B - \mu_A = 0$
H_1: $\mu_B - \mu_A \neq 0$
bilateral

- Indiqueu l'expressió de l'estadístic, la seva distribució sota H_0 i les premisses assumides. Indiqueu si la descriptiva de les dades recolza assumir les premisses (1 punt)

t = (mean(B) – mean(A)) / Error_tipus (és t_{120})
i Error_tipus és s_pooled sqrt(1/61+1/61) amb s_pooled = sqrt((60 s_B^2 + 60 s_A^2)/120)

Premisses:
A i B normals (boxplots força simètrics tot i que el de B menys)

Sigma_B=Sigma_A (s_B de magnitud similar a s_A i també boxplot de B no molt diferent que el de A. Per tant la premissa sembla acceptable)

- Calculeu l'error tipus estimat per a la diferència de mitjanes mostrals i el valor de l'estadístic (1 punt)

Error_tipus = s_pooled sqrt(1/61+1/61) = 13.877 sqrt(2/61) = 2.51

($s\_pooled^2 = (60 \; 14.09^2 + 60 \; 13.66^2) / (61+61-2) = (11917.67+11197.26)/120 = 192.56$

(s_pooled = sqrt(192.56) = 13.877)

t = (81.66 – 70.82) / Error_tipus = 10.84 / 2.51 = 4.32

- Calculeu i interpreteu l'interval de confiança del 95% per a la diferència d'esperances (1 punt)

(mean(B) – mean(A)) $\pm$ $t_{120,0.975}$ Error_tipus = 10.84 $\pm$ 1.98 2.51 = 10.84 $\pm$ 4.97 = [5.87 , 15.81]

Amb una confiança del 95% la diferència de mitjanes poblacionals (B-A) dels guanys en els prototipus A i B podria estar entre 5.87 i 15.81 [euros].

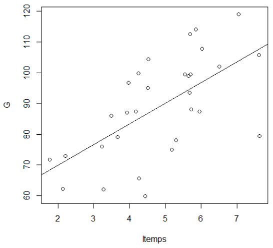
- Indiqueu, justifiqueu i interpreteu quina és la conclusió de la prova d'hipòtesis? (1 punt)

Hi ha evidència per rebutjar H_0 (l'estadístic estaria fora dels punts crítics en zona de rebuig)

Per tant, hi ha prou evidència per dir que les mitjanes dels guanys d'A i B no són iguals, i l'interval de confiança ens diu que la diferència és favorable al prototipus B.

Problema 3 (B6)

Els nous emprenedors volen estudiar si existeix alguna mena de relació lineal entre la durada de les visites i el guany monetari obtingut en visites en que es fa alguna compra. Ho estudien a partir del logaritme del temps (ltemps) i els guanys (G) de 30 visites. Els resultats de la descriptiva, de la covariància entre ltemps i G, i de la regressió lineal corresponents són:



	Mitjana	Desviació tipus	Min	Max
G	88.90	16.37	58.2	119.33
ltemps	4.83	1.51	1.5	7.8

$\text{cov(G,ltemps)} = S_{G,\text{ltemps}} = 15.34$

```
summary(lm(G ~ ltemps)):  Coefficients:      Estimate     Std.Error     t value      Pr(>|t|)
                            (Intercept)        ----         8.110        6.957       1.45e-07 ***
                                ltemps         ----         ----         4.191       0.000251 ***
                          Residual standard error: 13.06 on 28 degrees of freedom
```

- Calculeu els coeficients de la recta de regressió dels guanys G en funció del logaritme del temps. I trobeu també el coeficient de determinació, la desviació residual i el terme mancant de la columna "Std.Error" (2 punts)

$b_1 = 15.34 / 1.51^2 = 6.73$ (pendent)
$b_0 = 88.90 – 6.73\ \ 4.83 = 56.4$ (terme independent)
Corr(G, ltemps) = 15.34 / (16.37 1.51) = 0.62 R^2 = sqr(0.62) = 0.385 (coeficient de determinació)

La desviació residual és: arrel{ (30-1) 16.37^2 (1-0.385) / (30-2) } = 13.06

L'error tipus del terme lineal és: arrel{ 13.06^2 / (30-1) / 1.51^2 } = 1.6061

- Expliqueu què ens indiquen el pendent de la recta, la desviació residual i l'error tipus del pendent (2 punts)

Pendent: cada increment de 1 en el logaritme natural del temps implica augmentar 6.73 els guanys

Desviació residual: la desviació en les prediccions serà de l'ordre de 13 euros

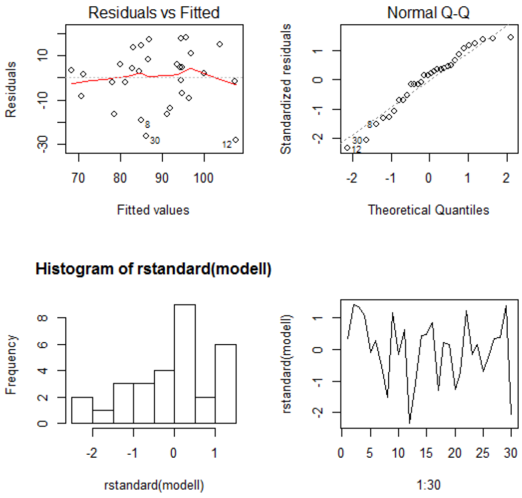
Error tipus del pendent: l'error típic d'estimació del pendent és aproximadament 1.6, d'acord a la dimensió de la mostra i com s'ha determinat la desviació residual.

- Calculeu un IC 95% del pendent de la recta i resoleu la prova d'hipòtesis de si la recta és horitzontal o no. (2 punts)

$IC_{95\%}(\beta_1) = b_1 \pm t_{28,0.975}\, Sb_1 = 6.73 \pm 2.048 \cdot 1.606 \approx 6.73 \pm 3.29 \approx [3.44; 10.02]$

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$
Hi ha evidència per rebutjar H_0,
ja que el valor posat a prova no està al IC (l'estadístic estaria fora dels punts crítics en zona de rebuig)
Per tant, hi ha prou evidència per dir que la recta no és horitzontal

- Enuncieu les premisses o hipòtesis de la regressió lineal i comenteu si es compleixen o no per aquest cas concret. Especifiqueu de quins resultats i/o gràfics es dedueixen els vostres comentaris. (2 punts)



Linealitat: prou clara al plot entre G i ltemps de l'enunciat, encara que amb força dispersió. Al gràfic dalt-esquerra la distribució dels punts és dispersa però entorn de la recta

Homoscedasticitat: raonable no hi ha patró, no hi ha grans zones amb més i menys variabilitat (gràfic dalt-esquerra i baix-dreta)

Normalitat: l'ajust a una normal és força correcte en el NormalQQ, tot i que l'histograma mostra certa asimetria amb cua a l'esquerra

Independència: molt raonable, ja que no hi ha patró que indiqui dependència (gràfic dalt-esquerra i baix-dreta)

- Es podría aceptar que si el logaritme del temps augmenta en una unitat els guanys s'incrementen en 10? Justifiqueu la resposta i valoreu-la (2 punt)

D'acord amb l'IC al 95%, una prova bilateral amb risc α=0.05 no es podria rebutjar, perquè el valor 10 és a dintre de l'IC.

Però si es vol verificar amb rigor, hauríem de plantejar les hipòtesis i comprovar què dóna l'estadístic en relació a les zones de rebuig.

H_0: $\beta_1 = 10$
H_1: $\beta_1 \neq 10$

L'estadístic de la prova és: (6.73 – 10) / 1.6061 = -2.036

Els punts crítics són: ±2.048. Llavors, l'estadístic cau a la zona d'acceptació (per poc), i no es pot rebutjar un increment de 10 euros por unitat del logaritme del temps.