PE Q1 2019-20. Examen parcial 2, 7 de gener de 2020

Exercici 1

Una empresa que fabrica portàtils presenta el seu nou portàtil LapTOP afirmant que les seves bateries són més potents que mai, que ara és possible treballar més de 10 hores i mig sense connectar-se a la xarxa elèctrica. Per demostrar-ho presenten les dades de funcionament (en hores) de 21 portàtils del tipus LapTOP que es poden resumir amb els següents estadístics:

$$\sum_{i=1}^{21} t_i = 225.05 \quad \text{i} \quad \sum_{i=1}^{21} t_i^2 = 2415.088.$$

Definim com a variable T el temps de funcionament (en hores) dels portàtils LapTOP sense connectar-se a la xarxa elèctrica i denotem per μ_T i σ_T el seu valor esperat i la seva desviació estàndard.

(a) (1 punt)

Doneu una estimació de μ_T i σ_T.

Solució:

$$\hat{\mu}_T = 225.05/21 = 10.72,$$
$$\hat{\sigma}_T = \sqrt{\frac{1}{20}\left(2415.088 - 21 \cdot 10.72^2\right)} = 0.41.$$

(b) (3 punts)

Calculeu els intervals de confiança d'un nivell de confiança igual al 95% tant de μ_T com de σ_T. Quines són les premisses en ambdós casos? Com s'interpreta l'interval de confiança de μ_T?

Solució:

Premisses: T ha de seguir una distribució normal i les dades han de ser d'una mostra aleatòria simple (m.a.s.).

- Interval de confiança de a la mitjana:

$$\mathrm{IC}(\mu_T; 0.95) = \bar{x} \mp t_{20;\,0.975} \cdot \frac{s_x}{\sqrt{n}} = 10.72 \mp 2.086 \cdot \frac{0.41}{\sqrt{21}} = [10.53,\ 10.9].$$

- Interval de confiança de la desviació estàndard:

$$\mathrm{IC}(\sigma_T^2; 0.95) = \left[\frac{s_T^2 \cdot (n-1)}{\chi^2_{20;\,0.975}},\ \frac{s_T^2 \cdot (n-1)}{\chi^2_{20;\,0.025}}\right] = \left[\frac{0.41^2 \cdot 20}{34.17},\ \frac{0.41^2 \cdot 20}{9.591}\right] = [0.1,\ 0.35],$$
$$\Longrightarrow \mathrm{IC}(\sigma_T; 0.95) = [0.32,\ 0.59].$$

Els dos intervals contenen el valor esperat i la desviació estàndard d'T, respectivament, amb una probabilitat de 0.95.

(c) (1 punt)

Basant-vos en els resultats de l'apartat anterior, es pot afirmar que el funcionament mitjà dels portàtils LapTOP és superior a 10.5 hores? Raoneu la resposta.

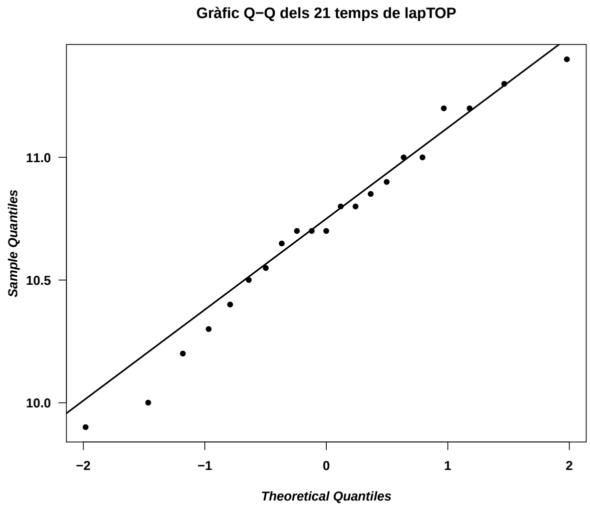
Solució:

El fet que el límit inferior d'IC(μ_T; 0.95) és superior a 10.5 implica que es pot rebutjar l'hipòtesi

$$\text{H: } \mu_T \leq 10.5$$

amb un valor p inferior a 0.025. Podem, per tant, suposar que el temps de funcionament mitjà és superior a 10 hores.

(d) (**0.5 punts**)
Veient el següent gràfic Q-Q dels temps del portàtil LapTOP, sembla raonable suposar que la variable T segueix una distribució normal? Raoneu la resposta.



Solució:
Els punts no formen una recta perfecta, però tampoc s'allunyen massa de la recta. Per tant, sembla raonable suposar que les dades provenen d'una distribució normal.

Després de portar sis mesos en el mercat, l'empresa fa una enquesta entre els seus clients per estudiar el grau de satisfacció amb els seus productes i en cas del portàtil LapTOP, 108 de 144 clients es mostren satisfets.

(e) (**1.5 punts**)
Doneu un estimació puntual de la proporció de clients satisfets amb LapTOP (π) i calculeu l'interval de confiança de π amb un nivell de confiança igual al 95%. En quines premisses es basa aquest càlcul.

Solució:

- **Estimació puntual**: $p_1 = 108/144 = 0.75$.
- **Interval de confiança**:

$$\mathrm{IC}(\pi; 0.95) = p_1 \mp z_{0.975} \cdot \sqrt{\frac{p_1 \cdot (1 - p_1)}{n}} = [0.68,\ 0.82].$$

  Alternativa:

$$\mathrm{IC}(\pi; 0.95) = p_1 \mp z_{0.975} \cdot \sqrt{\frac{0.5 \cdot (1 - 0.5)}{n}} = [0.67,\ 0.83].$$

- **Premisses**: Dades provenen d'una m.a.s. i mostra suficientment gran.

(f) ($\mathbf{0.3 + 0.5 + 0.6 + 0.5 + 0.6 + 0.5 = 3}$ **punts**)
En una nova enquesta tres mesos més tard, només 98 de 144 clients mostren la seva satisfacció amb els portàtils LapTOP. Basada en aquesta mostra, es pot suposar que el grau de satisfacció amb LapTOP entre tots els clients encara supera el 65%? Per contestar aquesta pregunta,

- doneu un estimació puntual de la proporció de clients satisfets amb LapTOP,

  Solució:
  $p_2 = 98/144 = 0.68$.

- indiqueu quines són la hipòtesi nul·la i la seva alternativa,

  Solució:

$$\mathrm{H}_0\colon \pi_2 \le 0.65 \quad \text{vs.} \quad \mathrm{H}_1\colon \pi_2 > 0.65.$$

- calculeu el valor de l'estadístic de contrast,

  Solució:

$$z = \frac{p_2 - \pi_2}{\sqrt{\pi_2 \cdot (1 - \pi_2)/n}} = 0.769.$$

- indiqueu la distribució que té l'estadístic de contrast sota la hipòtesi nul·la,

  Solució:
  Distribució sota H_0: $Z \overset{H_0}{\sim} \mathcal{N}(0, 1)$.

- resoleu el contrast d'hipòtesi amb $\alpha = 0.05$,

  Solució:
  Valor crític (unilateral): $z_{0.95} = 1.645$.
  Com $z = 0.769 < 1.645$, no podem rebutjar la hipòtesi nul·la utilitzant un nivell de significació igual a 0.05.

- i interpreteu el resultat.

  Solució:
  No obstant el grau de satisfacció a la mostra sigui superior al 65%, no podem afirmar que així sigui entre tots els clients de LapTOP (amb un nivell de confiança del 95%).

NOM:__________________________COGNOM:____

Problema 2 (B5)

A tots els apartats, justifiqueu les respostes, explicant el passos

Quan anem de vacances a un hotel utilitzem diferents eines per buscar el que més ens convenç. Volem comparar els cercadors Kay i Tri en quant al nombre d'hotels que ens ofereixen al cercar en ciutats d'Europa. Per fer l'estudi agafem en cadascun dels cercadors les 10 ciutats europees per les que ofereixen un nombre més gran d'hotels d'una certa categoria. Els resultats pel nombre d'hotels oferts a Kay (K) en les 10 primeres ciutats del seu rànquing, i el nombre d'hotels oferts a Tri (T) en les 10 primeres ciutats del seu rànquing (no tenen perquè ser les mateixes ciutats en cadascun dels cercadors) sabem que són "normals" i amb els següents valors numèrics:

K: 140 130 123 120 114 112 102 100 90 88 mean(K) = 111.9 sd(K) = 17

T: 199 170 165 160 150 142 138 128 120 110 mean(T) = 148.2 sd(T) = 26.43

1.- Justifiqueu si es tracta de dades aparellades o independents. Compareu com haurien de ser les dades en aquest cas per ser aparellades o independents (1 punt)

Són independents ja que són 10 ciutats de cada rànquing i poden ser diferents

Per ser dades aparellades haurien de ser les mateixes ciutats

2.- Primer compararem les variàncies en els dos cercadors:

a) plantegeu les hipòtesis de si es poden considerar les variàncies iguals o una és superior a l'altra (0.5 punts)

$H_0 : \sigma^2_T = \sigma^2_K$

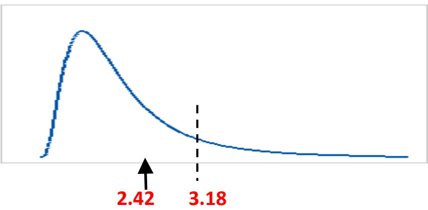
$H_1 : \sigma^2_T > \sigma^2_K$ **(prova unilateral)**

b) indiqueu i calculeu l'estadístic (0.5 punts)

$var(T)/var(K) = sd(T)^2/sd(K)^2 = 26.43^2/17^2 = 698.54/289 = \mathbf{2.42}$

c) representeu gràficament la distribució de l'estadístic i el/s punt/s crític/s suposant un α = 0.05 (1 punt)

Punt crític $F_{9,9,0.95} = \mathbf{3.18}$



d) analitzeu els resultats interpretant la conclusió a que arribeu (1 punt)

Com que el valor de l'estadístic (2.42) està dins la zona d'acceptació,

res s'oposa a acceptar igualtat de variàncies

Per tant en els dos cercadors podem esperar la mateixa variabilitat

3.- Comenteu què implica el resultat de la prova anterior a l'hora de fer la prova de comparar mitjanes que es demanarà a continuació (1 punt)

Implica que podrem assumir la premissa d'igualtat de variàncies tot i que desconegudes, ja que és una premissa necessària per la prova de comparar dues μ

4.- Ara compararem les mitjanes del nombre d'hotels oferts:
a) plantegeu les hipòtesis de si es poden considerar les mitjanes iguals o no (0.5 punts)

H_0: $\mu_T = \mu_K$

H_1 : $\mu_T \neq \mu_K$ (prova bilateral)

b) indiqueu i comenteu les premisses (0.5 punts)

Normalitat de K i T (ho podem assumir segons ho afirma l'enunciat)

Variàncies desconegudes però es poden assumir iguals per la prova anterior

c) indiqueu i calculeu la desviació "pooled" i l'error estàndard (0.5 punts)

s_pooled = 22.22   **s_pooled = sqrt(493.73) = sqrt((9*26.43² + 9*17²)/18) = 22.22**

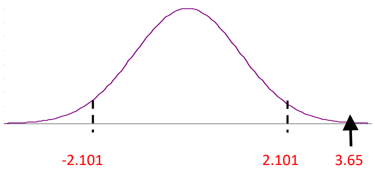
se = 9.94   **se = s_pooled*sqrt(1/10+1/10) = 22.22 * sqrt(2/10) = 9.94**

d) indiqueu i calculeu l'estadístic (0.5 punts)

estadístic = 3.65   **(mean(T)-mean(K))/se = (148.2-111.9)/9.94 = 36.3/9.94 = 3.65**

e) representeu gràficament la distribució de l'estadístic i el/s punt/s crític/s suposant un α = 0.05 (0.5 punt)

punts crítics $t_{18,0.975}$=2.101 i $t_{18,0.025}$=-2.101 i valor de l'estadístic 3.65



f) calculeu un interval de confiança al 95% de la diferència de mitjanes (1 punt) i interpreteu-lo

[15.42 , 57.18]
(mean(T)-mean(K))- $t_{18,0.975}$ * se = (148.2-111.9)-2.101*9.94 = 36.3 – 20.88 = 15.42
(mean(T)-mean(K))+ $t_{18,0.975}$ * se = (148.2-111.9)+2.101*9.94 = 36.3 + 20.88 = 57.18

Amb 95% de confiança la diferència de mitjanes del nombre d'hotels trobats està entre uns 15 i 57 per sobre en el cas de Tri

g) analitzeu els resultats interpretant la conclusió a que arribeu (1 punt)

Com que l'estadístic (3.65) està fora de la zona d'acceptació dels punts crítics (+-2.101), O bé 0 no pertany al IC, vol dir que les dades mostren que no és raonable creure la hipòtesis nul·la

Per tant no és raonable creure que la mitjana d'hotels que ofereixen els dos cercadors sigui la mateixa

5.- Amb quina de les dues plataformes decidiríeu treballar? Justifiqueu la vostra elecció tenint en compte les proves anteriors (0.5 punts)

Si ens interessa el màxim nombre d'ofertes triaríem Tri,

perquè, en mitjana la diferència és d'unes 36 ofertes més a Tri (mean(T)-mean(K)=36.3)
o bé en mitjana més nombre d'ofertes, amb 95% de confiança entre unes 15 i 57 més a Tri

NOM:____________________________COGNOM:____

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

Problema 3 (B6)

En un empresa el responsable de compres ha encarregat al departament d'informàtica un estudi sobre la relació entre el preu (P), en euros, i la capacitat (C) del disc dur, en GB, dels 28 ordinadors comprats en el darrer trimestre de l'any.
Les dades recollides es troben resumides en la següent taula:

	Mitjana	Desviació tipus	Covariància
Capacitat del disc dur (C)	162.86	34.73	3226.98
Preu (P)	788.39	105.24	

1. Calculeu el coeficient lineal de la recta de regressió del preu respecte la capacitat del disc dur (0'5 punts)

$b_1 = \frac{s_{CP}}{s_C^2} = 2.675$

2. Calculeu el coeficient del terme independent de la recta de regressió anterior (0'5 punts)

$b_0 = \bar{P} - b_1 \cdot \bar{C} = 352.74$

3. Escriviu la recta de regressió del preu respecte la capacitat i interpreteu-la en el context de l'estudi realitzat. (1 punt)

La recta de regressió és P=2.675C+352.74

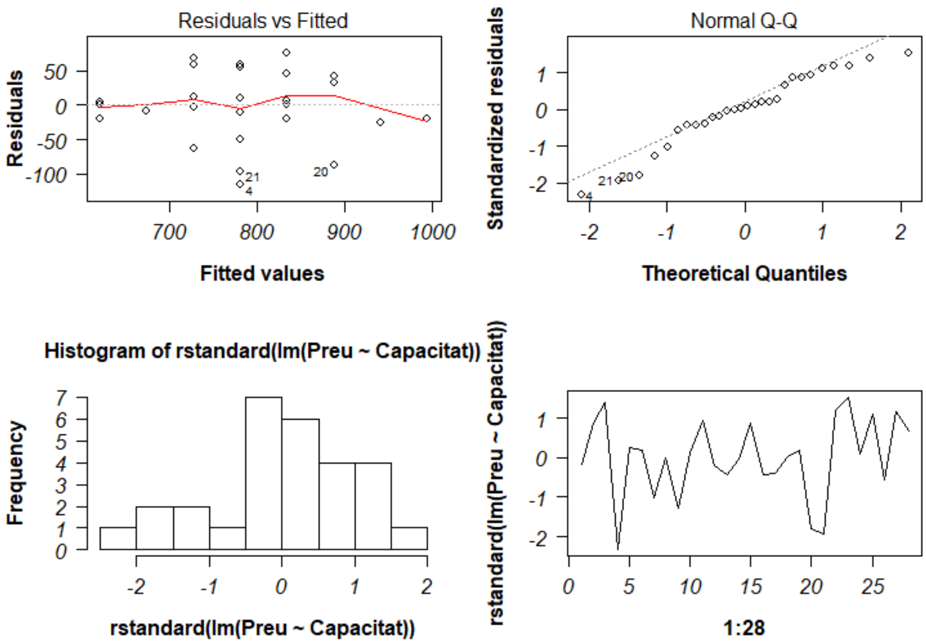
Per cada GB que s'augmenta en la capacitat del disc dur, el preu augmenta en 2.675€
En el cas de la constant seria el cas extrem en què consideréssim que un ordinador no té disc dur i aleshores tindria un preu de 352.74€. En aquest context, seria un ordinador que es ven sense disc dur.

4. Calculeu l'estimació puntual pel paràmetre de la desviació residual σ amb aquestes dades, i interpreteu-ho (1 punt)

$S^2 = \frac{(n-1)\cdot(S_P^2 - b_1 \cdot S_{CP})}{n-2}$ = 2537.26, per tant s=50.37.

Podem esperar, típicament, fluctuacions de 50.37€ del preu respecte als valors que ens doni el model, per a una capacitat del disc dur concreta.

5. A partir dels següents gràfics, feu l'anàlisi de les premisses del model lineal. Indiqueu si es compleixen o no cadascuna d'elles i argumenteu-ho en cada cas. (2 punts)



Linealitat: Es compleix. Els residus segueixen una tendència constant al voltant del 0.
Homocedasticitat: Es compleix. Els residus tenen la mateixa variabilitat al llarg dels valors predits.
Normalitat: Es compleix. En el *qqnorm*, els punts estan prou propers a la recta identitat.
Independència: Es compleix. No s'observa cap patró en el gràfic dels residus versus l'ordre.

6. Volem estudiar si per cada GB que té de més el disc dur, el preu augmenta en 4€ o no.

a) Indiqueu les hipòtesis, la fórmula de l'estadístic i quina és la distribució d'aquest sota la hipòtesi nul·la (1 punt)

$\begin{cases} H_0: \beta_1 = 4 \\ H_1: \beta_1 \neq 4 \end{cases}$, test bilateral

L'estadístic és $\hat{t} = \dfrac{b_1 - \beta_1}{S_{b1}}$ i $\hat{t} \sim t_{26}$

b) Calculeu el valor de l'estadístic (1 punt)

$$S_{b1} = \sqrt{\frac{s^2}{(n-1)\cdot s_C^2}} = 0.2791$$

$$\hat{t} = \frac{2.675 - 4}{0.2791} = -4.7474$$

c) Si considerem un risc del 5%, indiqueu els punts crítics (0'5 punts)

Els punts crítics són $t_{0.025,26}$ =-2.056 i $t_{0.975,26}$ =2.056

d) A partir de l'estudi i dels càlculs realitzats, interpreteu els resultats de la prova d'hipòtesi aplicada sobre el cas concret. (0'5 punts)

El valor de l'estadístic (-4'74) és més petit que el valor crític (-2'056); per tant, podem concloure que tenim evidències que en augmentar el disc dur en un GB el preu no augmenta en 4€.

7. Feu una predicció puntual i per interval de confiança del 95% pel preu d'un disc dur de capacitat 160GB (2 punts)

$$\hat{y}_{C=160} = 2.675 \cdot 160 + 352.74 = 780.74\ €$$

$$\text{IC}(y_{C=160}, 95\%) = \hat{y}_{C=160} \mp t_{0.975,26} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{(n-1)\cdot S_x^2}} = 780.74 \mp 2.056 \cdot 50.37 \cdot \sqrt{1 + \frac{1}{28} + \frac{(160-162.86)^2}{27 \cdot 34.73^2}} =$$
$$[675.33\ , 886.15]$$