1. **Estimació puntual de l'error tipus**

   Error tipus = $\frac{s}{\sqrt{n}} = \frac{0.4}{\sqrt{25}} = \frac{0.4}{5} = 0.08$

2. **IC al 95% pel valor esperat μ del temps de servei**

   a. **Estadístic:**

   $\frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{2.17-2}{0.4/5} = \frac{0.17}{0.08} = 2.125 \sim t_{24,0.95}$

   b. **Premisses:**

   Les dades son independents, segueixen una distribució normal amb σ desconeguda.

   c. **Distribució de probabilitat que ha de seguir l'estadístic:**

   t-student amb 24 graus de llibertat: $t_{24}$ per ser σ desconeguda.

   d. **Càlcul de l'IC al 95%**

   $\mu \in (\bar{x} \pm t_{24,0.975} * \frac{s}{\sqrt{n}})$= $(2.17 \pm 2.064 * \frac{0.4}{\sqrt{25}})$= $(2.17 \pm 2.064 * \frac{0.4}{\sqrt{25}})$= (2.17$\pm$0.165)=(2.005, 2.335)

   e. **Interpretació del resultats**

   Com que μ=2 no pertany a l'interval de confiança, encara que per molt poc, rebutgem la Hipòtesi nul·la que el temps d'espera és en mitjana 2 i per tant el temps d'espera és superior a 2, oscil·lant entre 2.005 i 2.335 minuts.

3. Quina distribució de referència podem usar per l'estadístic de l'apartat 2.a) si en comptes de tenir una mostra de 25 observacions tinguéssim una mostra de 100 observacions

   Quan n és de l'ordre de 100 observacions, la distribució de referència de la t_student s'aproxima a la distribució N(0,1)

4. Plantegeu i resoleu una prova d'hipòtesis bilateral (al 95%) que us permeti decidir si la mitjana del temps de servei és μ=2

   H0: μ=2

   H1: μ≠2

   Estadístic: $\frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{2.17-2}{0.4/\sqrt{25}} = \frac{0.17}{0.08} = 2.125$

   $t_{24,0.975} = 2.064$

   Com que el valor de l'estadístic (2.125) és superior al valor de referència de la distribució $t_{24,0.975}$ i per tant, rebutgem la Hipòtesi nul.la H0, és a dir, la mitjana del temps de servei és superior a 2.

5. **A) Doneu les estimacions per interval al 90, 95 i 99% de confiança amb σ=1 per estimar l'esperança del temps de servei SI (n=25)**

   En aquest cas, donat que la σ és coneguda, la distribució de referència és una Normal i per tant l'IC es calcula com $\bar{x} \pm z_{1-\alpha/2} * \sigma/\sqrt{n}$ on $\alpha/2$ serà 0.05, 0.025 i 0.005 respectivament.

IC al 90%: $\bar{x} \pm z_{0.95} * \sigma/\sqrt{n}$ =2.17 $\pm$ 1.65 $*$ 1/$\sqrt{25}$ = 2.17 $\pm$ 0.33=(1.84, 2.5)
IC al 95%: $\bar{x} \pm z_{0.975} * \sigma/\sqrt{n}$ =2.17 $\pm$ 1.96 $*$ 1/$\sqrt{25}$ = 2.17 $\pm$ 0.392=(1.78, 2.56)
IC al 99%: $\bar{x} \pm z_{0.995} * \sigma/\sqrt{n}$ =2.17 $\pm$ 2.575 $*$ 1/$\sqrt{25}$ = 2.17 $\pm$ 0.515=(1.66, 2.69)

**B) Quina distribució de referència usem en aquest estadístic? Per què?**
Una distribució normal per que es coneix la desviació tipus poblacional

**C) Comenteu quan i per què canvia l'IC quan augmentem la confiança del 90 a 95 i 99%**
A mesura que s'augmenta la confiança, l'interval s'eixampla, ja que a més confiança, més incertesa. Això ve donat per la distribució de referència, que en aquest cas és la N(0,1).

6. **A) Doneu les estimacions per interval al 90, 95 i 99% de confiança per estimar l'esperança del temps de servei SI, amb σ desconeguda (n=25)**

En aquest cas, donat que la σ és desconeguda, la distribució de referència és una t_student amb n-1 graus llibertat i per tant l'IC es calcula com $\bar{x} \pm t_{1-\alpha/2} * s/\sqrt{n}$ on $\alpha/2$ serà 0.05, 0.025 i 0.005 respectivament.
IC al 90%: $\bar{x} \pm t_{0.95} * s/\sqrt{n}$ =2.17 $\pm$ 1.711 $*$ 0.4/$\sqrt{25}$ = 2.17 $\pm$ 0.137=(2.033, 2.307)
IC al 95%: $\bar{x} \pm t_{0.975} * s/\sqrt{n}$ =2.17 $\pm$ 2.064 $*$ 0.4/$\sqrt{25}$ = 2.17 $\pm$ 0.165=(2.005, 2.335)
IC al 99%: $\bar{x} \pm t_{0.995} * s/\sqrt{n}$ =2.17 $\pm$ 2.797 $*$ 0.4/$\sqrt{25}$ = 2.17 $\pm$ 0.224=(1.946, 2.394)

**B) Quina distribució de referència usem en aquest estadístic? Per què?**
Una distribució t-student per que no es coneix la desviació tipus poblacional i per tant es treballa amb la mostral.

**C) Comenteu quan i per què canvia un IC si el calculem amb sigma coneguda o no coneguda**

Per poder realitzar aquestes comparacions ens caldrà que la desviació tipus poblacional sigui igual que la mostral. Assumint que **σ=s=1**, els intervals de confiança al 90, 95 i 99% amb sigma desconeguda seran
IC al 90%: $\bar{x} \pm t_{0.95} * s/\sqrt{n}$ =2.17 $\pm$ 1.711 $*$ 1/$\sqrt{25}$ = 2.17 $\pm$ 0.342=(1.828, 2.512)
IC al 95%: $\bar{x} \pm t_{0.975} * s/\sqrt{n}$ =2.17 $\pm$ 2.064 $*$ 1/$\sqrt{25}$ = 2.17 $\pm$ 0.413=(1.757, 2.583)
IC al 99%: $\bar{x} \pm t_{0.995} * s/\sqrt{n}$ =2.17 $\pm$ 2.797 $*$ 1/$\sqrt{25}$ = 2.17 $\pm$ 0.560=(1.610, 2.730)

Si comparem aquests resultats amb els obtinguts en l'apartat A d'aquest mateix problema (assumint que σ=s) s'observa que els IC d'aquest apartat C) son molt més amples que els de l'apartat A). Això és degut que aquests IC s'han calculat amb la distribució t_student que per n=25 observacions presenta més variabilitat que la distribució N(0,1).

NOM:________________________________________________________________

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

**Problema 2 (B5).** La màquina virtual de Java disposa d'una optimització anomenada JIT (*Just In Time*) que permet compilar les seccions de codi més utilitzades per reduir el temps d'execució. Estudis previs indiquen que els temps sense JIT són més de 7 vegades superiors als de amb JIT (anomenant els temps NO i JIT respectivament, indiquen que el rati NO/JIT ≈ 7). Per confirmar-ho, uns companys de PE han generat de forma aleatòria 100 vectors de mida 200000 i per a cada vector han obtingut els temps d'ordenació de Mergesort amb i sense JIT.

La taula mostra la seva descriptiva (JIT i NO) i la del seus logaritmes naturals (Log(JIT) i Log(NO)), així com de les diferències respectives (NO-JIT, i Log(NO)-Log(JIT)=Log(NO/JIT) que anomenarem D).

| N=100 | Mit-jana | Desv. est | Corr |
|---|---|---|---|
| *JIT* | 133.7 | 16.59 | 0.170 |
| *NO* | 962.8 | 96.56 | |
| *NO-JIT* | 829.1 | 95.15 | |
| *Log(JIT)* | 4.887 | 0.1228 | 0.154 |
| *Log(NO)* | 6.874 | 0.1009 | |
| *Log(NO)-Log(JIT)=D* | 1.976 | 0.1464 | |

**1)** [1 punt] D'acord amb la recollida i amb la descriptiva de les dades, es tracta de dues mostres independents o aparellades? Raoneu la resposta.

Aparellades, ja que el mateix vector és ordenat pels 2 sistemes en comparació.

Es pot fer la diferència ja que els tempsièsim tenen en comú el vector ordenat

Es pot esperar que els temps ièssims si semblin entre sí més que els temps obtinguts sobre vectors diferents: cap esperar una certa relació positiva entre les dues variables. Els coeficients de correlació indiquen una correlació positiva lleu en aquesta mostra.

**2)** [1 punt] Té avantatges haver recollit les dades de forma aparellada? Compareu els errors estàndard.

En principi, ha de tenir més precisió i el error estàndard d'estimació de l'efecte (diferències) hauria de ser millor. Però la petita correlació observada (0.170 i 0.154) podria portar un benefici modest. Per comprovar-ho el calculem sota els 2 supòsits (dades independents i dades aparellades):

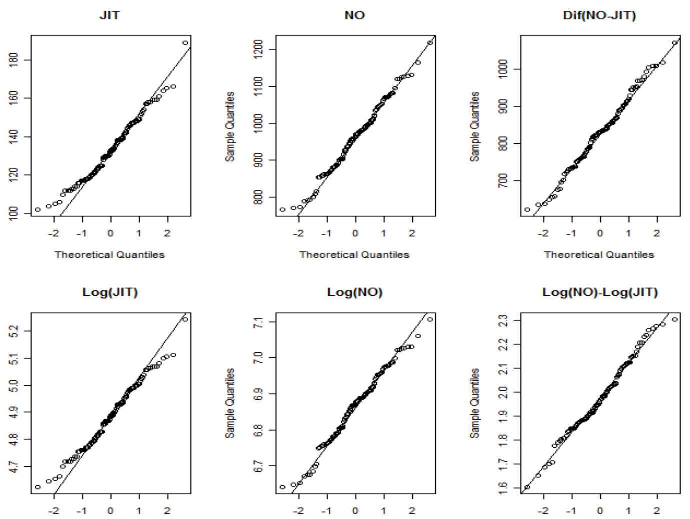
Independents: $SE(\tilde{Y}_{JIT} - \tilde{Y}_{NO}) = S_P\sqrt{1/n_{JIT} + 1/n_{NO}} \approx 62.28\sqrt{1/100 + 1/100} \approx 9{,}8$

$$S^2_{Pooled} \approx \frac{S^2_{JIT}(n_{JIT}-1) + S^2_{NO}(n_{NO}-1)}{(n_{JIT}-1)+(n_{NO}-1)} \approx \frac{16.59^2(99) + 96.56^2(99)}{99+99} \approx 62.28^2$$

Aparellades: $SE(\tilde{D}_{JIT-NO}) = S_D\sqrt{1/n_D} \approx 95.15\sqrt{1/100} \approx 9{,}5$

El disseny **aparellat és una mica més eficient, ja que el SE es menor (9,5<9.8). [I també pels logs (0'0146<0.0159)]**

Si compara ES de les mitjanes separades -> 0. Si compara les anterior amb l'aparellat -> 1/3.




**3)** [1 punt] Veient aquests Q-Q plots, digueu si les variables s'acosten al model Normal i quines s'acosten més.

Totes 6 s'ajusten prou bé.

Les dues variables de les diferències (columna 3ª) tenen un ajustament més bo al model Normal.

Aplicar logaritmes no sembla modificar la forma de la distribució.

**4)** [1 punt] Comenteu avantatges en interpretació de resultats i premisses per tractar les dades logo-transformades en aquest exemple.

**Interpretació**: Com el rendiment comparatiu ve expressat en forma de quocient o raó (JIT és més de 7 vegades més ràpid), treballar amb logaritmes permet estimar el valor d'aquest quocient o fer contrastos rellevants. OK: si 'eliminen efecte multiplicatiu"

**Normalitat**: *és indiferent*.

**Homoscedasticitat**: millor logo-transformades ja que les desviacions típiques sota els 2 sistemes en comparació són molt mes semblants i permeten interpretar millor la mitjana de les diferències.

**Independència** entre les observacions de les diferències: no es pot valorar.

**5)** [4 punts] Amb les dades de les diferències de logaritmes, contrasteu la $H_0$ que l'esperança poblacional ($\mu_D$) de D (on D=log(NO)-log(JIT)=log(NO/JIT)) sigui 1.946 (≈log(7)) enfront de l'alternativa que sense JIT trigui més (més de 7 vegades abans de fer logaritmes).

Indiqueu: hipòtesi, premisses, estadístic, distribució i punt crític, càlculs, resultats i interpretació.

- Prova d'hipòtesi unilateral

  Ho: $\mu_D = Log(7) = 1.946$ ó $\mu_D \le Log(7) = 1.946$

  $H_1$: $\mu_D > Log(7) = 1.946$

- Premisses: Les diferències D provenen d'una distribució normal i son independents entre si.

- Estadístic: $\hat{t} = \dfrac{(\tilde{D} - \mu_D)_0}{s_D\sqrt{1/100}} \sim t_{n-1}$ que en aquest cas serà $t_{99}$.

- Com els graus de llibertat son 99, amb unalfa=0.05=5% unilateral, punt crític $t_{0.95,99}$= 1.660391.

- El valor de l'estadístic és $SE(\tilde{D}_{Log(JIT)-Log(NO)}) = S_{DLogs}\sqrt{1/n_D} \approx 0.1465\sqrt{1/100} \approx 0{,}0146$

  $$\hat{t} = \frac{(1.976 - 1.946)}{0.0146} = \frac{0.030}{0.0146} \approx 2.055$$

- Com que 2.055 > 1.660, rebutgem $H_0$

- És a dir, hem demostrat que JIT és més de 7 vegades més ràpid.

Per cada error, -1p; Si no interpreta logs a rati, -1/2

**6)** [2 punts] Calculeu un IC per a $\mu_D$ amb una confiança del 90%, i interpreteu com ens informa del rati NO/JIT.
Indiqueu i calculeu com trobaríeu un valor *k* del qual podem afirmar, amb una confiança del 95%, que JIT és, com a mínim, *k* vegades més ràpid que sense JIT (NO/JIT ≈ k)

- $IC_{90\%}$ bilateral: Com $\tilde{D} \approx 1.976$, $t_{0.95,99}$= 1.660391i $SE(\tilde{D}_{Log(JIT)-Log(NO)}) \approx 0{,}0146$

  $IC_{unilateral,95\%}\ E(D_{Log(JIT)-Log(NO)}) = \tilde{D} + t_{0.05,99} \cdot SE(\tilde{D}_{Log(JIT)-Log(NO)}) \approx 1.976 \pm (-1.660391) \cdot 0.0146 \approx [1.95, 2.00]$

  $[e^{1.95}, e^{2.00}] \approx 7.041, 7.389$

  Amb una confiança del 90% l'autèntic rati és algun valor entre 7.04 i 7.38.

- Demana calcular un $IC_{95\%}$ unilateral entre X i infinit (és a dir, sense límit superior).
  Dada la simetria de l'IC90%, podem afirmar, amb una confiança del 95%, que
  JIT és 7.041 vegades més ràpid. O encara més.

Per cada error, -1/2p; Si no interpreta logs a rati, -1/2

| | | | |
|---|---|---|---|
| >qt(0.975,99) = 1.984217 | >qt(0.975,198) = 1.972017 | >pt(0.975,99) = 0.8340317 | >pt(0.975,198) = 0.834625 |
| >qt(0.95,99) = 1.660391 | >qt(0.95,198) = 1.652586 | >pt(0.95,99) = 0.8277872 | >pt(0.95,198) = 0.8283648 |
| >qt(0.90,99) =] 1.290161 | >qt(0.90,198) = 1.285842 | >pt(0.90,99) = 0.8148479 | >pt(0.90,198) = 0.8153932 |

FIB Q1 2013-2014 PARCIAL 2 de PE 18 de desembre de 2013

NOM:

## SOL.LUCIÓ PROBLEMA 3 BLOC B6

S'ha mesurat, en segons, el temps de processament de 7 programes en funció del nombre de milions d'accessos Input/Output al disc. Els resultats apareixen a la següent taula:

| Accessos I/O | Temps de CPU |
|---|---|
| 14 | 2 |
| 16 | 5 |
| 27 | 7 |
| 42 | 9 |
| 39 | 10 |
| 50 | 13 |
| 83 | 20 |

1) Especifiqueu quina és la variable explicativa o independent (X) i quina la variable resposta o depenent (Y). De quin tipus són aquestes variables? (**1 punt**).

La variable explicativa és el *Nombre d'accessos Input/Output* i la variable resposta és el *Temps de processament*. Ambdues variables són quantitatives. El *Nombre d'accessos Input/Output* és discreta ja que només pren valors naturals, i el *Temps de processament* és continua ja que pren valors en tots els reals positius.

2) Calculeu la recta de regressió que relaciona les dues variables anteriors. Per tal d'alleugerir els càlculs, sapigueu que: $\sum_i x_i = 271, \sum_i x_i^2 = 13855, \sum_i y_i = 66, \sum_i y_i^2 = 828$ i el coeficient de correlació és $r = 0.9867$ (**3 punts**).

De les sumes que ens donen tenim que: $\overline{x} = \frac{\sum_i x_i}{n} = \frac{271}{7} = 38,71$ i $\overline{y} = \frac{\sum_i y_i}{n} = \frac{66}{7} = 9.43$. De la fórmula $S_x^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$ i la seva equivalent en $y$, s'obté que: $S_x^2 = 560,26$ i $S_y^2 = 34,28$. Fent l'arrel quadrada obtenim que: $S_x = 23,67$ i $S_y = 5,85$.

Tenint en comte que la fórmula del pendent de la recta de regressió en funció del coeficient de correlació s'obté que: $b_1 = r \cdot \frac{S_y}{S_x} = 0.9867 \cdot 5,85/23,67 = 0.2438$. I de la fórmula del terme independent s'obté que $b_0 = \overline{y} - b_1 \cdot \overline{x} = 9,43 - 0.244 \cdot 38,71 = -0.015$.

3) Què passarà amb la variable resposta si incrementem en dos milions d'unitats el nombre

d'accessos I/O? I si l'increment és de 10 milions? (**2 punts**).

De la interpretació del pendent de la recta de regressió tenim que, si incrementem en un milió d'unitats els accessos Input/Output el temps incremetarà en 0.2438 segons. Per tant, si els accessos incrementen en dos milions, el temps s'incrementarà en 0.4876 segons, i si l'increment és de 10 milions d'accessos, el temps s'incrementarà en $2,43$ segons.

4) Calculeu l'estimació puntual de la variància residual (**1 punt**).

$$S^2 = \frac{(n-1)S_y^2(1-r^2)}{n-2} = \frac{6 \cdot 34,28 \cdot (1-0.986^2)}{5} = 1.08.$$

7) Calculeu la predicció puntual corresponent a 43 milions d'accessos I/O (**1 punt**).

L'estimació puntual serà igual a: $\hat{y} = b_1 \cdot 43 + b_0 = 0.2438 \cdot 43 - 0.015 = 10.4756.$

6) Calculeu un interval de confiança al 95% per el pendent de la recta de regressió. (**2 punts**).

La fórmula de la desviació estàndar estimada de $b_1$ és igual a: $S_{b_1} = \frac{S}{S_x}\sqrt{1/(n-1)}$ i, per tant $S_{b_1} = \frac{\sqrt{1.1437}}{23{,}67} \cdot \frac{1}{\sqrt{6}} = 0.0183$. Per tant, l'intèrval de confiança per $b_1$ demanat és igual a:

$$b_1 + -t_{5,0.975} \cdot S_{b_1} = 0.2438 + -2.571 \cdot 0.0183 = 0.2438 + -0.047.$$

La qual cosa és equivalent a: $(0.1988, 0.29)$.