

NOM: \_\_\_\_\_  
 (Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliqueu i justifiqueu els càlculs en les respostes)

### Problema 1 (B4)

Se desea estudiar el tiempo que funciona cierta componente de un sistema hasta que comienza a presentar fallos sistemáticos. A fin de estimar el tiempo medio de funcionamiento a pleno rendimiento, se observaron 10 de estas componentes obteniéndose los siguientes tiempos –en miles de horas– de funcionamiento sin fallos:

1, 1.5, 0.8, 1.2, 0.9, 1.1, 1.4, 1.3, 0.7, 0.1

NOTA: al final del problema encontraréis los percentiles que necesitáis.

a) (2 puntos) Dar una estimación por intervalo para el tiempo medio de funcionamiento con una confianza del 90%.

- Estadístico, premisas y distribución

- Cálculos, resultado e interpretación.

b) (2 puntos) Dar una estimación por intervalo para la desviación estándar poblacional con una confianza del 90%.

- Estadístico, premisas y distribución

- Cálculos, resultado e interpretación.

c) (2 puntos) ¿Podemos decir que el tiempo de funcionamiento medio está por debajo de 1200 horas? Asumid varianza conocida igual a 0.12 y  $\alpha=0.05$ .

- Hipótesis, estadístico, premisas y distribución

- Cálculos, resultado e interpretación.

Una vez las componentes observadas fallaron, se realizó una intervención para arreglarlas y volver a hacerlas funcionar. El mecánico nos ha garantizado que obtendremos un mayor rendimiento del sistema. Así, se volvieron a poner en funcionamiento estas mismas componentes, obteniéndose los siguientes nuevos tiempos –en miles de horas– de funcionamiento sin fallos:

0.1, 0.3, 0.4, 1.3, 1.4, 1.2, 0.4, 1.4, 0.7, 1.3

d) (2 puntos) ¿Se ha alargado efectivamente el tiempo medio de funcionamiento al arreglar la componente?

- Hipótesis, estadístico, premisas y distribución (tomar  $\alpha=0.05$ )

- Cálculos, resultado e interpretación.

Datos posteriores recogidos por la empresa, muestran que de las 1000 componentes fabricadas en el primer mes del año, 216 duraron menos de 900 horas.

e) (2 puntos) Estimar por intervalo la probabilidad de que una componente dure menos de 900 horas, con una confianza del 95%.

- Estadístico, premisas y distribución.

- Cálculos, resultado e interpretación.

#### Datos auxiliares:

> qt(0.95,9)	> qchisq(0.05,10)
[1] 1.833113	[1] 3.940299
> qt(0.95,10)	> qchisq(0.95,9)
[1] 1.812461	[1] 16.91898
> qt(0.975,9)	> qchisq(0.95,10)
[1] 2.262157	[1] 18.30704
> qt(0.975,10)	> qchisq(0.025,9)
[1] 2.228139	[1] 2.700389
> qnorm(0.95)	> qchisq(0.025,10)
[1] 1.644854	[1] 3.246973
> qnorm(0.975)	> qchisq(0.975,9)
[1] 1.959964	[1] 19.02277
> qchisq(0.05,9)	> qchisq(0.975,10)
[1] 3.325113	[1] 20.48318

NOM:

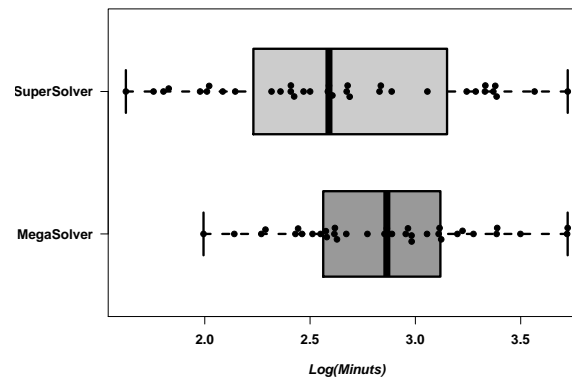
(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliqueu i justifiqueu els càlculs en les respostes)

## Problema 2 (B5)

Per comparar la velocitat amb la qual resolen dos servidors diferents, *SuperSolver* i *MegaSolver*, problemes d'optimització s'envia un total de 70 problemes de maximització diferents als dos servidors, 35 a cadascun. Pel fet que el temps que triguen els servidors per resoldre els problemes, és asimètrica cap a la dreta, treballem a continuació amb els logaritmes dels temps. Sigui  $X$  el logaritme del temps que triga el *SuperSolver* i  $Y$  el del *MegaSolver*.

Els valors descriptius a cada mostra són els següents i a més a més es mostra una representació gràfica:

	Mitjana	Mediana	Desv. est.	Mínim	Màxim
<i>SuperSolver</i>	2,63	2,59	0,57	1,63	3,72
<i>MegaSolver</i>	2,85	2,86	0,44	1,99	3,72



a) (0.5 punts)

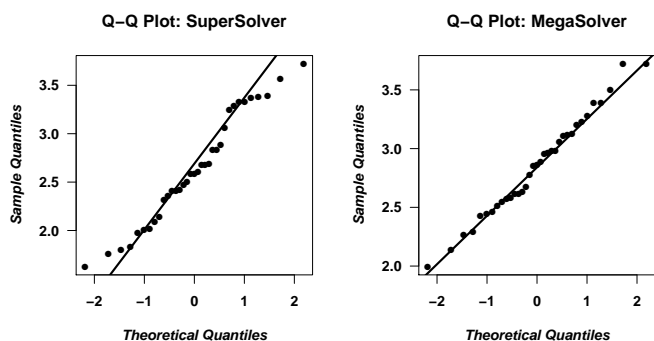
Es tracta de dues mostres independents o aparellades? Raoneu la resposta.

b) (0.75 punts)

Només basant-vos en els valors descriptius i en el gràfic, creieu que es pot parlar de superioritat d'un dels dos servidors?

c) (0.75 punt)

Veient els següents  $Q-Q$  plots, sembla raonable suposar que les variables  $X$  i  $Y$  segueixen una distribució normal? Raoneu la resposta.



NOM: \_\_\_\_\_

**d) (2 punts)**

Estudiem primer la variabilitat dels (logaritmes de) temps d'ambdós servidors. Es pot suposar que són iguals? Responen aquesta pregunta plantejant i resolent la hipòtesi adient (amb  $\alpha = 0.1$ ) i explicant quines són les premisses que cal fer per realitzar aquesta prova.

**Nota:** Quantils de la distribució F:  $qf(.9, 34, 34) = 1.56$ ;  $qf(.95, 34, 34) = 1.77$ ;  $qf(.975, 34, 34) = 1.98$ .

**e) (3 punts)**

Per saber si es pot suposar que hi ha diferències entre ambdós servidors, es vol calcular l'interval de confiança (al 95%) per a la diferència de les mitjanes dels logaritmes del temps ( $IC(\mu_X - \mu_Y; 0.95)$ ).

- Quines són les premisses?
  
- Quina és la distribució de  $\bar{X} - \bar{Y}$  (suposant igualtat de variàncies)?
  
- Calculeu la variància conjunta (*pooled variance*).
  
- Calculeu  $IC(\mu_X - \mu_Y; 0.95)$ .
  
- Quina és la interpretació d'aquest interval?

**f) (2 punts)**

En una altre experiment s'han enviat els mateixos 31 problemes d'optimització a ambdós servidors i s'han recollit les diferències dels logaritmes dels temps. La mitjana d'aquestes diferències ha estat igual a 0.22 (a favor del *SuperSolver*), la desviació estàndard igual a 0.37.

Calculeu l'interval de confiança (al 99%) per a la mitjana de les diferències. Es pot suposar la superioritat d'un dels dos servidors? Per què?

**g) (1 punt)**

Mantenint la mitjana igual, a partir de quina desviació estàndard ja no es podria parlar de superioritat (amb un nivell de confiança igual a 99%)?

NOM: \_\_\_\_\_  
 (Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els càlculs en les respostes)

### Problema 3 (B6)

Les variables Y i X mesuren una puntuació de 0 a 10 en un esforç físic i hores connectat a Facebook, respectivament. Hem observat els següents 9 valors: Y= (8,4,5,5,2,3,3,2,0) i X= (8,9,11,14,15,16,19,20,21) I alguns resultats intermedis són :

$$\sum_{i=1}^9 X_i = 133 \quad \sum_{i=1}^9 X_i^2 = 2145 \quad \sum_{i=1}^9 Y_i = 32 \quad \sum_{i=1}^9 Y_i^2 = 156 \quad \sum_{i=1}^9 X_i Y_i = 400$$

$$\bar{X} = 14.778 \quad \bar{Y} = 3.556 \quad s_X = 4.738 \quad s_Y = 2.297$$

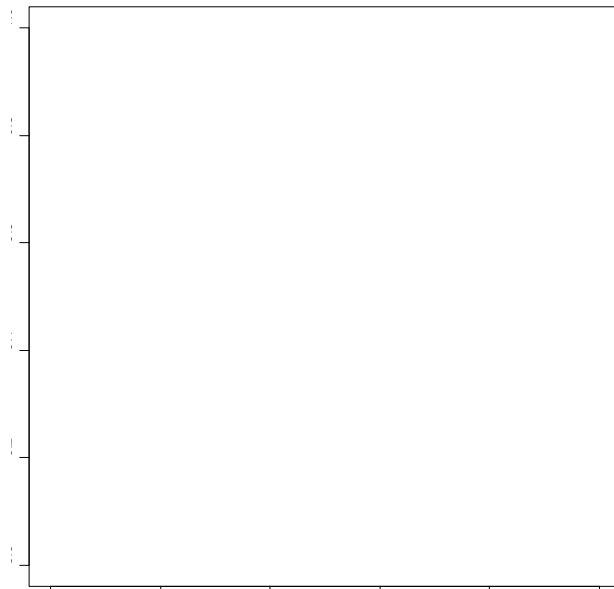
Volem explicar la variabilitat de Y en funció dels valors de X

Aplicant regressió lineal, en R alguns dels resultats que obtenim són:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	.....	1.5478	6.173	0.000457
X	.....	0.1003	-4.049	0.004877

(1 punt) Quina és la recta de regressió? Indiqueu el càlcul dels coeficients i dibuixeu la recta



(1 punt) Comenteu el significat dels coeficients de la recta:

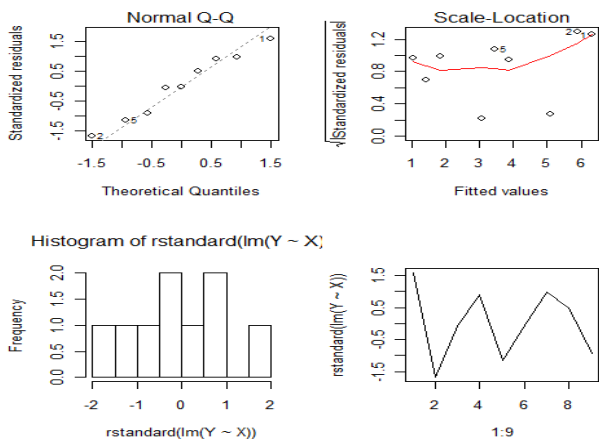
(1 punt) Completeu la taula de descomposició de la variabilitat. Comenteu aquesta descomposició i què indica?

Font Variabilitat	SQ	GdL	DF	QM	Rati
Explicada					
Residual					
Total					

(1 punt) Resoleu la prova d'hipòtesi per acceptar o no una recta com a model:

(1 punt) Calculeu el coeficient de determinació i feu una interpretació global del model

(1 punt) Feu l'anàlisi de les premisses.



(1 punt) Feu la prova d'hipòtesi adient per si es pot acceptar o no que el pendent de la recta sigui -0.5

**Ara considerem la variable hores de connexió només observada en 3 valors o categories: 10, 15 o 20 hores. Volem relacionar la mateixa variable Y anterior amb la nova variable X\_categories=(10,10,10,15,15,15,20,20,20).**

En R alguns dels resultats que obtenim són:

```

Response: Y
          Df Sum_Sq  Mean_Sq  F_value  Pr(>F)
X_categories ..... 24.222  .....  .....  0.07748
Residuals     .....  .....  .....
  
```

(1 punt) Completeu la taula de descomposició de la variabilitat. Comenteu aquesta descomposició i què indica?

Font Variabilitat	SQ	GdL DF	QM	Rati
Explicada				
Residual				
Total				

(1 punt) Resoleu la prova d'hipòtesi de si hi ha relació o no entre la puntuació Y i X\_categories

(1 punt) Calculeu el coeficient de determinació i feu una interpretació global del model