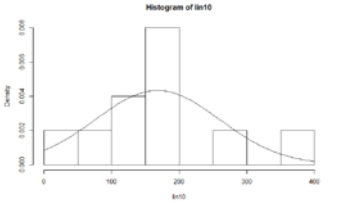
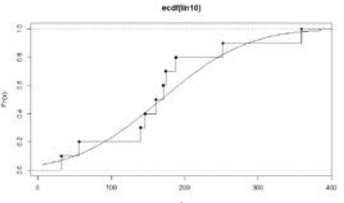
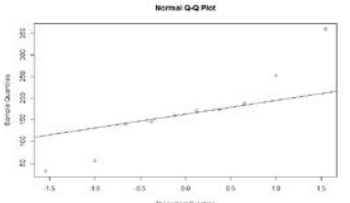


Problema 1 (B4)

Un grup d'estudiants de la FIB està estudiant els temps de cerca d'un nombre en un vector de 10^5 elements de l'algorisme de cerca lineal. Ha generat un únic vector amb 10^5 naturals de l'1 al 100000. Escullen deu vegades a l'atzar l'element a buscar entre 1 i 10^5 ; i obtenen els resultats següents expressats en ns.

$\sum_{i=1}^{10} x_i = 1679$ $\sum_{i=1}^{10} x_i^2 = 358243$			
Histograma (funció densitat empírica) vs funció densitat teòrica segons model Normal		Funció distribució empírica vs F. distr. teòrica segons model Normal	QQPlot (quantils empírics vs Qs. teòrics segons model Normal)

1. Calculeu les estimacions puntuals del temps mitjà i de la desviació tipus (1 punt)

2. Calculeu l'error tipus de la mitjana. Interpreteu la desviació i l'error tipus i digueu quina utilitat poden tenir en aquestes dades (p.e., per a què utilitzaríeu cadascú?) (1 punt)

3. A partir dels gràfics argumenteu si podem suposar que el temps segueix una distribució normal. (1 punt)

4. Assumint que les dades segueixen la distribució normal (independentment de la resposta a les preguntes anteriors), calculeu un interval bilateral amb 90% de confiança per la mitjana de temps. Interpreteu. (2 punts)
5. Se suposa que la mitjana poblacional del temps de cerca és 200 ns o superior, i els estudiants volen estudiar amb una prova d'hipòtesi al 5% si el temps de cerca ha millorat. Indiqueu:
- a) Les hipòtesis, les premisses, la fórmula de l'estadístic i quina és la distribució d'aquest sota la hipòtesi nul·la (1 punt)
- b) Calculeu el valor de l'estadístic (1 punt)
- c) Representeu gràficament el(s) punt(s) crític(s), les zones d'acceptació i de rebuig i el valor de l'estadístic (1 punt)
- d) A partir de l'estudi i del càlculs realitzats, interpreteu els resultats de la prova d'hipòtesi (1 punt)
6. A l'apartat 4 hem fet un IC bilateral al 10%; i al 5, una PH unilateral al 5%. Té sentit? Compareu els resultats (1 punt)

Problema 2 (B5). Es vol comparar els diners que es gasta el pare Noel i els reis mags d'orient per comprar regals. En una mostra aleatòria de 30 famílies, per a cadascuna d'elles, es recull el cost econòmic dels regals portats pel pare Noel i els regals portats pels reis. Siguin C1 i C2 els diners gastats pel primer i pels segons i D la diferència entre ells ($D=C2-C1$). Aquesta és la descriptiva d'aquestes variables:

Nota: les unitats són euros de despesa per persona

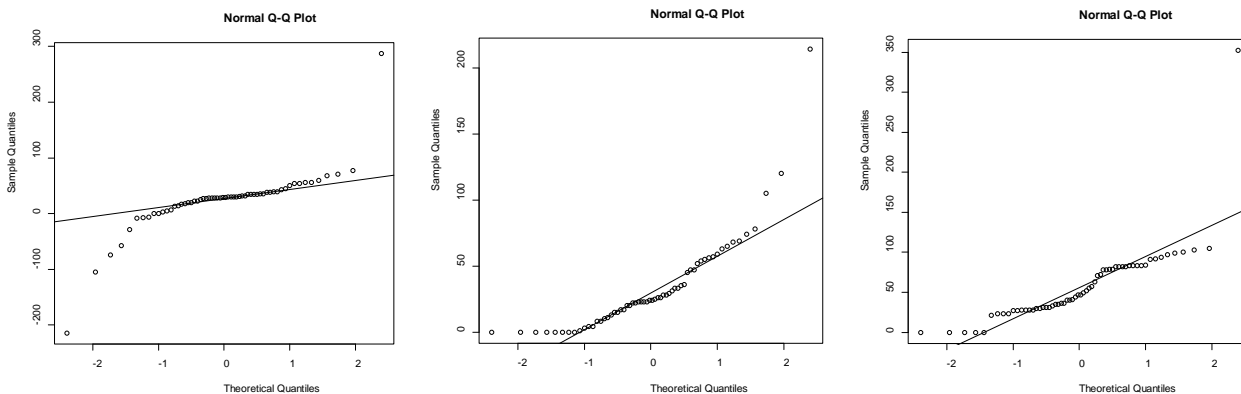
```
> summary(x)
  C1PareNoel      C2ReisMags      Diferencia
Min.   : 0.00    Min.   : 0.00    Min.   : -214.00
1st Qu.: 10.75   1st Qu.: 29.50   1st Qu.: 16.25
Median : 24.00   Median : 47.00   Median : 29.00
Mean   : 33.47   Mean   : 57.55   Mean   : 24.03
3rd Qu.: 48.25   3rd Qu.: 82.00   3rd Qu.: 38.00
Max.   :214.00   Max.   :352.00   Max.   : 287.00

> sd(x$C1PareNoel)
[1] 35.60207
> sd(x$C2ReisMags)
[1] 49.06081
> sd(x$Diferencia)
[1] 55.62525
```

1) (1 punt). Es tracta de dues mostres independents o aparellades? Raoneu la resposta.

2) (2 punts). Fixeu-vos en els següents *qqnorms*.

QQ norms per `x$Diferencia`, `x$C1PareNoel`, `x$C2ReisMags`



- a) La funció `qqnorm` o `qqplot` grafica las funcions quantil d'una mostra versus una distribució normal teòrica. Expliqueu el gràfic i què permet observar. (0.67 punts)
- b) Comenteu si es compleix la premissa de Normalitat en tots ells. N'hi ha algun d'ells en el que la normalitat sembli més creïble? (0.67 punts)
- c) En funció de la resposta de la pregunta 1), Quins són el gràfics rellevants per l'anàlisi que s'ha de realitzar?(0.67 punts) Raoneu les respostes.

3) (4 punts). Per saber si es pot suposar que hi ha diferències entre els diners que gasten el pare Noel i els reis, es vol plantejar una prova d'hipòtesi d'igualtat de mitjanes: $H_0: \mu_{C1} = \mu_{C2}$ vs. $H_1: \mu_{C1} \neq \mu_{C2}$

- a) Calcula l'estadístic per fer la comparació, digues quina distribució segueix sota la hipòtesi nul·la i amb quines premisses.

b) Digues quin és(són) el(s) punt(s) crític(s) amb un 5% de significació i treu conclusions sobre la prova d'hipòtesi.

c) Calcula l'interval de confiança del 99% per a la diferència de mitjanes i interpreta'l.

d) Interpreta que vol dir l'interval de confiança del 99% (si no has trobat els valors numèrics, fes servir x i y per a resoldre aquest apartat). Si l'interval de confiança es calcula amb un nivell del 90%. Com creieu que serà aquest: més ample o més estret que el calculat amb el 99% de confiança? Raoneu la resposta.

4) (3 punts). Ara interessa conèixer si la proporció de famílies que rep regals amb un cost superior a 20 euros per persona és el mateix quan els regals són del pare Noel o són dels reis mags.

a) Es decideix emprar dues mostres independents de grandària 100 (una en cada moment de temps). Les proporcions p_1 i p_2 de despesa superior als 20 euros per persona són 0.65 i 0.90 del pare Noel i dels reis respectivament.

Calcula l'estadístic per dur a terme la comparació i digues quina distribució segueix sota la hipòtesi nul·la i sota quines premisses.

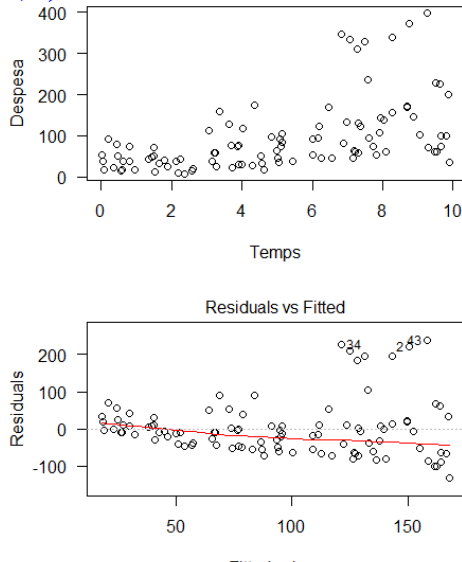
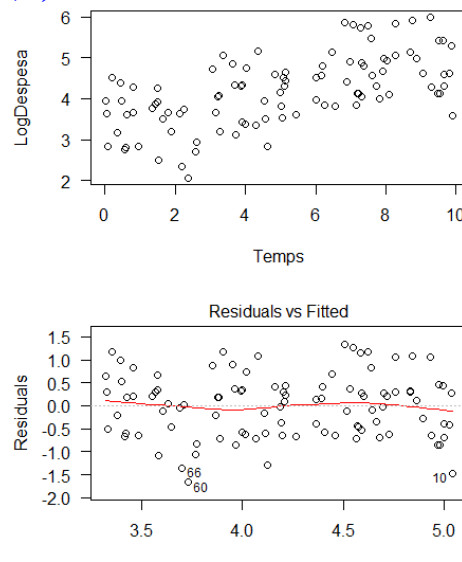
b) Digues quin és el punt crític amb un 5% de significació i treu conclusions sobre la prova d'hipòtesi.

c) Calcula l'interval de confiança del 95% per la diferència de proporcions π_1 i π_2

Problema 3 (B6)

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

El temps que navega un client dins d'una plana web de venda online de productes d'informàtica es creu que pot estar relacionat amb la seva despesa a l'hora d'adquirir productes. S'han recollit dades de clients que han realitzat alguna compra. A continuació, es pot trobar la descriptiva de les variables Temps (minuts), Despesa (euros) i LogDespesa (logaritme natural de la despesa). A sota, hi ha part de la sortida de R dels 2 models lineals que ajusten la despesa (transformada o no) segons el temps de navegació.

Descriptiva																															
<pre>> summary(dades)</pre> <table border="1"> <thead> <tr> <th>Temps</th> <th>Despesa</th> <th>LogDespesa</th> </tr> </thead> <tbody> <tr> <td>Min. : 0.010</td> <td>Min. : 7.89</td> <td>Min. : 2.066</td> </tr> <tr> <td>1st Qu.: 2.535</td> <td>1st Qu.: 38.48</td> <td>1st Qu.: 3.650</td> </tr> <tr> <td>Median : 5.075</td> <td>Median : 67.39</td> <td>Median : 4.209</td> </tr> <tr> <td>Mean : 5.094</td> <td>Mean : 95.12</td> <td>Mean : 4.205</td> </tr> <tr> <td>3rd Qu.: 7.580</td> <td>3rd Qu.: 118.06</td> <td>3rd Qu.: 4.771</td> </tr> <tr> <td>Max. : 9.900</td> <td>Max. : 397.89</td> <td>Max. : 5.986</td> </tr> </tbody> </table>	Temps	Despesa	LogDespesa	Min. : 0.010	Min. : 7.89	Min. : 2.066	1st Qu.: 2.535	1st Qu.: 38.48	1st Qu.: 3.650	Median : 5.075	Median : 67.39	Median : 4.209	Mean : 5.094	Mean : 95.12	Mean : 4.205	3rd Qu.: 7.580	3rd Qu.: 118.06	3rd Qu.: 4.771	Max. : 9.900	Max. : 397.89	Max. : 5.986	<pre>> sd(dades\$Temps); sd(dades\$Despesa); sd(dades\$LogDespesa)</pre> <pre>[1] 3.007195</pre> <pre>[1] 86.4399</pre> <pre>[1] 0.8550065</pre> <pre>> cov(dades\$Temps, dades\$Despesa); cov(dades\$Temps, dades\$LogDespesa)</pre> <pre>[1] 136.9266</pre> <pre>[1] 1.574313</pre>									
Temps	Despesa	LogDespesa																													
Min. : 0.010	Min. : 7.89	Min. : 2.066																													
1st Qu.: 2.535	1st Qu.: 38.48	1st Qu.: 3.650																													
Median : 5.075	Median : 67.39	Median : 4.209																													
Mean : 5.094	Mean : 95.12	Mean : 4.205																													
3rd Qu.: 7.580	3rd Qu.: 118.06	3rd Qu.: 4.771																													
Max. : 9.900	Max. : 397.89	Max. : 5.986																													
Model 1	Model 2																														
<pre>> mod1 <- lm(Despesa~Temps, dades); summary(mod1)</pre> <p>Coefficients:</p> <table border="1"> <thead> <tr> <th></th> <th>Estimate</th> <th>Std. Error</th> <th>t value</th> <th>Pr(> t)</th> </tr> </thead> <tbody> <tr> <td>(Intercept)</td> <td>(1)</td> <td>(2)</td> <td>1.234</td> <td>0.22</td> </tr> <tr> <td>Temps</td> <td>(3)</td> <td>(4)</td> <td>(5)</td> <td>1.81e-08 ***</td> </tr> </tbody> </table> <p>Residual standard error: (6) on 98 degrees of freedom Multiple R-squared: 0.2775, Adjusted R-squared: 0.2701 F-statistic: 37.64 on 1 and 98 DF, p-value: 1.806e-08</p> <pre>> par(mfrow=c(2, 1)); plot(Despesa~Temps, dades); > plot(mod1, 1)</pre> 		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	(1)	(2)	1.234	0.22	Temps	(3)	(4)	(5)	1.81e-08 ***	<pre>> mod2 <- lm(LogDespesa~Temps, dades); summary(mod2)</pre> <p>Coefficients:</p> <table border="1"> <thead> <tr> <th></th> <th>Estimate</th> <th>Std. Error</th> <th>t value</th> <th>Pr(> t)</th> </tr> </thead> <tbody> <tr> <td>(Intercept)</td> <td>(1)</td> <td>(2)</td> <td>24.737</td> <td>< 2e-16 ***</td> </tr> <tr> <td>Temps</td> <td>(3)</td> <td>(4)</td> <td>(5)</td> <td>1.3e-11 ***</td> </tr> </tbody> </table> <p>Residual standard error: (6) on 98 degrees of freedom Multiple R-squared: 0.3749, Adjusted R-squared: 0.3685 F-statistic: 58.78 on 1 and 98 DF, p-value: 1.296e-11</p> <pre>> par(mfrow=c(2, 1)); plot(LogDespesa~Temps, dades); > plot(mod2, 1)</pre> 		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	(1)	(2)	24.737	< 2e-16 ***	Temps	(3)	(4)	(5)	1.3e-11 ***
	Estimate	Std. Error	t value	Pr(> t)																											
(Intercept)	(1)	(2)	1.234	0.22																											
Temps	(3)	(4)	(5)	1.81e-08 ***																											
	Estimate	Std. Error	t value	Pr(> t)																											
(Intercept)	(1)	(2)	24.737	< 2e-16 ***																											
Temps	(3)	(4)	(5)	1.3e-11 ***																											

a) Amb els darrers gràfics que apareixen al final de cada model, quines premisses es poden avaluar? Quin dels dos models s'ajusta millor a les premisses del model lineal? (1 punt)

b) En el model escollit en l'apartat anterior i amb les dades proporcionades a l'enunciat, calcula els valors que falten a la sortida dins dels quadres 1 a 6 (3 punts)

En els següents apartats, si no has pogut resoldre l'apartat anterior empra els valors de la següent taula

(1)	(2)	(3)	(4)	(5)	(6)
3	0.2	0.2	0.02	8	0.7

c) Interpreta que suposa un increment d'un minut més navegant respecte a la despesa en compres (2 punts)

d) Calcula un interval de confiança del **90%** per la pendent del model escollit (2 punts)

e) Fes la predicció puntual i per interval de confiança del 95% per un usuari que navega 1 minut. (2 punts)