

Problemes-e-status-Models-estadi...



fibernauta



Probabilidad y Estadística



2º Grado en Ingeniería Informática



**Facultad de Informática de Barcelona (Fib)
Universidad Politécnica de Catalunya**

antes

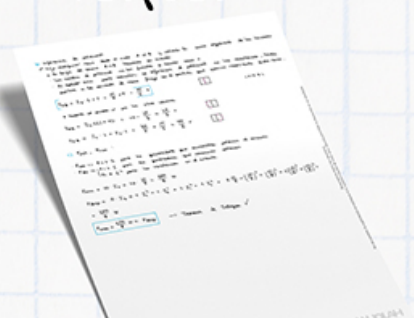


**Descarga sin publi
con 1 coin**



Después

WUOLAH



Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

pierdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH



Probabilitat i Estadística FIB-UPC

Problemes d'e-status:
B6 – Models estadístics i previsió



Modelo lineal

Corrección: Modelo lineal

Se ha estudiado la relación existente entre un factor X y una variable Y, y con los datos de una muestra obtenida con este propósito se ha hallado la siguiente modelización:

Call:
lm(formula = y ~ x)

Residuals:

Min	1Q	Median	3Q	Max
-0.86108	-0.30631	0.05851	0.31077	0.87816

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	87.808666	0.000000	197.035	<2e-16 ***
x	0.015952	0.007923	2.013	0.0497 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4084 on 48 degrees of freedom
Multiple R-squared: 0.07787, Adjusted R-squared: 0.05866
F-statistic: 4.053 on 1 and 48 DF, p-value: 0.04971


✓	1. En la salida hay un valor corrompido. ¿Puede recuperar el valor original? El valor desaparecido era 0.4456501.	0.4456501
✓	2. Responda con el estadístico de la prueba $\beta_0 = 0$. La solución es 197.034728546884	197.035
✓	3. ¿Qué vale el error de estimación del término lineal? La solución es 0.00792318073106572	0.007923
✓	4. Responda con el estadístico de la prueba $\beta_1 = 0$. La solución es 2.01329927612341	2.013
⚠	5. Extremo inferior de un intervalo de confianza 95% para la pendiente. La solución es 2.11114616302313e-05	0.00002174
✓	6. P-valor de la prueba $\beta_1 = 0$. La solución es 0.0497090581084531	0.0497
✓	7. Halle el coeficiente de determinación asociado al modelo. La solución es 0.0778695724347688	0.07787

Resultat

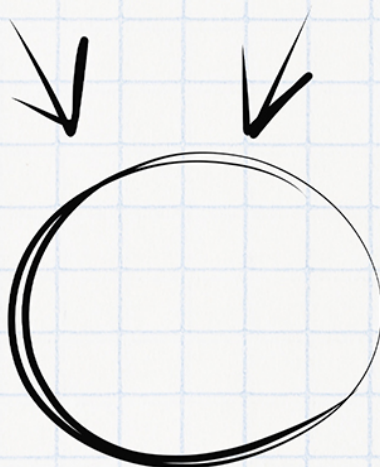
Nota 9.429

Imagínate aprobando el examen

Necesitas tiempo y concentración

Planes	 PLAN TURBO	 PLAN PRO	 PLAN PRO+
 Descargas sin publi al mes	10 	40 	80 
 Elimina el video entre descargas			
 Descarga carpetas			
 Descarga archivos grandes			
 Visualiza apuntes online sin publi			
 Elimina toda la publi web			
 Precios Anual <input type="checkbox"/>	0,99 € / mes	3,99 € / mes	7,99 € / mes

Ahora que puedes conseguirlo,
¿Qué nota vas a sacar?



WUOLAH

Script en R

```
b0 = 87.808666; tb0 = 197.035; pb0 = 0
b1 = 0.015952; sb1 = 0.007923; tb1 = 2.013; pb1 = 0.0497
desv = 0.4084; coef = 0.07787; df = 48

# Valor corrompido
sb0 = b0/tb0; sb0
# Estadístico de la prueba  $\beta_0=0$ 
tb0
# Error de estimación del término lineal
sb1
# Estadístico de la prueba  $\beta_1=0$ 
tb1
# Extremo inferior de un intervalo de confianza 95% para la
pendiente ( $\beta_1$ )
b1 - qt(0.975,df) * sb1
# P-valor de la prueba  $\beta_1=0$ 
#  $P(|t| > tb1) = 2*(1-P(t < tb1))$  [R: 2*(1-pt(tb1,df))]
pb1
# Coeficiente de determinación
coef

# Preguntas de otras ejecuciones
#           Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.76750  0.20624    66.755 <2e-16 ***
x           -0.01143  0.02865     ?@?*#@  0.691
# Valor corrompido -> tb1 = b1/sb1 = -0.01143/0.02865
#           Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.083e+02  4.765e-02 -2273.754 <2e-16 ***
x           @|?|@&!|?  2.230e-03  1.559      0.134
# Valor corrompido -> b1 = tb1 * sb1 = 1.559 * 2.230e-03
# Extremo superior de un intervalo de confianza 95% para la
pendiente ( $\beta_1$ ) -> b1 + qt(0.975,df) * sb1
# ¿Cuántos grados de libertad se asocian a la t-Student
presente en este modelo? -> df
```

```

# ¿Rechazaría la hipótesis de que la recta realmente es
horizontal? (0: no, 1: sí) -> ifelse(tbl > qt(0.975,df), 1, 0)

# ¿Rechazaría la hipótesis de que la recta pasa por el
origen? (0: no, 1: sí) -> ifelse(tb0 > qt(0.975,df), 1, 0)

# Diga el valor que se ha estimado para la desviación típica
del término aleatorio del modelo -> desv

# P-valor de la prueba  $\beta_0=0$  -> pb0

# Error de estimación del término independiente -> sb0

# Estimación del término independiente -> b0

# Pendiente estimada -> b1

# ¿Podríamos rechazar con estos datos  $\beta_0=-108.39$ ? (use riesgo
 $\alpha=2\%$ ; 0: no, 1: sí) -> t = (b0-(-108.39))/sb0
                        ifelse(t > qt(0.99,df), 1, 0)

# ¿Podríamos rechazar con estos datos  $\beta_1=0.001$ ? (use riesgo
 $\alpha=2\%$ ; 0: no, 1: sí) -> t = (b1-0.001)/sb1
                        ifelse(t > qt(0.99,df), 1, 0)

```

Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

pierdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH

Consola de R

```
> b0 = 87.808666; tb0 = 197.035; pb0 = 0
> b1 = 0.015952; sb1 = 0.007923; tb1 = 2.013; pb1 = 0.0497
> desv = 0.4084; coef = 0.07787; df = 48
> sb0 = b0/tb0; sb0
[1] 0.4456501
> tb0
[1] 197.035
> sb1
[1] 0.007923
> tb1
[1] 2.013
> b1 - qt(0.975,df) * sb1
[1] 2.174082e-05
> pb1
[1] 0.0497
> coef
[1] 0.07787
```

Rectas para todo

Corrección: Rectas para todo

En cierto estudio se pretende observar la asociación entre una variable X y una respuesta Y. Se dispone de 53 observaciones de la respuesta, con sus correspondientes valores de X, que han proporcionado los siguientes indicadores:

$\bar{X} = 71.68168$
 $\bar{Y} = -8.30196$
 $s_X = 25.90671$
 $s_Y = 0.71573$
 $s_{X,Y} = 0.46941$

Se le pide que conteste las preguntas que vienen a continuación.

✓	1. ¿Cuánto vale la pendiente estimada? Nota: 1.25	0.006994035
✓	2. De acuerdo con la hipótesis: $\beta_1=0.005$, diga qué valor toma el estadístico de la prueba. Nota: 1.25	-1.112
✓	3. Obtenga una estimación de la variancia residual. Nota: 1.25	0.5219792
✓	4. ¿Se podría rechazar la hipótesis: $\beta_0=-8.1$? (use riesgo $\alpha=10\%$; 0: no, 1: sí). Nota: 1.25	0
✓	5. Diga el valor que se ha estimado para la desviación típica del término aleatorio del modelo. Nota: 1.25	0.7224813
✓	6. De acuerdo con la hipótesis: $\beta_0=-8.1$, diga qué valor toma el estadístico de la prueba. Nota: 1.25	-0.8561661
✓	7. ¿De cuántos grados de libertad dispone este modelo para estimar los parámetros? Nota: 1.25	51
✓	8. Teniendo en cuenta las características del modelo que has estimado, diga cuál de las siguientes gráficas es la que corresponde a las variables analizadas.	2

Nota: 1.25

Resultat
Nota: 10

Script en R

```
x = 71.68168
y = -8.30196
sx = 25.90671
sy = 0.71573
sxy = 0.46941
n = 53

# ¿Cuánto vale la pendiente estimada?
b1 = sxy/sx^2; b1 # rxy = sxy/(sx*sy)
      # b1 = rxy*(sy/sx)

# De acuerdo con la hipótesis:  $\beta_1=0.005$ , diga qué valor toma
el estadístico de la prueba
s2 = ((n-1)*(sy^2-b1*sxy))/(n-2)
sb1 = sqrt(s2/((n-1)*sx^2))
t = (b1-0.005)/sb1; t

# Obtenga una estimación de la variancia residual
s2

# ¿Se podría rechazar la hipótesis:  $\beta_0=-8.1$ ? (use riesgo
 $\alpha=10\%$ ; 0: no, 1: sí)
b0 = y - b1*x
sb0 = sqrt(s2*(1/n + x^2/((n-1)*sx^2)))
t = (b0-(-8.1))/sb0
alfa = 10/100
ifelse(t > qt(1-alfa/2,n-2), 1, 0)

# Diga el valor que se ha estimado para la desviación típica
del término aleatorio del modelo
s = sqrt(s2); s

# De acuerdo con la hipótesis:  $\beta_0=-8.1$ , diga qué valor toma
el estadístico de la prueba
t

# ¿De cuántos grados de libertad dispone este modelo para
estimar los parámetros?
n-2
```

Consola de R

```
> x = 71.68168
> y = -8.30196
> sx = 25.90671
> sy = 0.71573
> sxy = 0.46941
> n = 53
> b1 = sxy/sx^2; b1
[1] 0.0006994035
> s2 = ((n-1)*(sy^2-b1*sxy))/(n-2)
> sb1 = sqrt(s2/((n-1)*sx^2))
> t = (b1-0.005)/sb1; t
[1] -1.112029
> s2
[1] 0.5219792
> b0 = y - b1*x
> sb0 = sqrt(s2*(1/n + x^2/((n-1)*sx^2)))
> t = (b0-(-8.1))/sb0
> alfa = 10/100
> ifelse(t > qt(1-alfa/2,n-2), 1, 0)
[1] 0
> s = sqrt(s2); s
[1] 0.7224813
> t
[1] -0.8561661
> n-2
[1] 51
```

Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

perdo
espacio

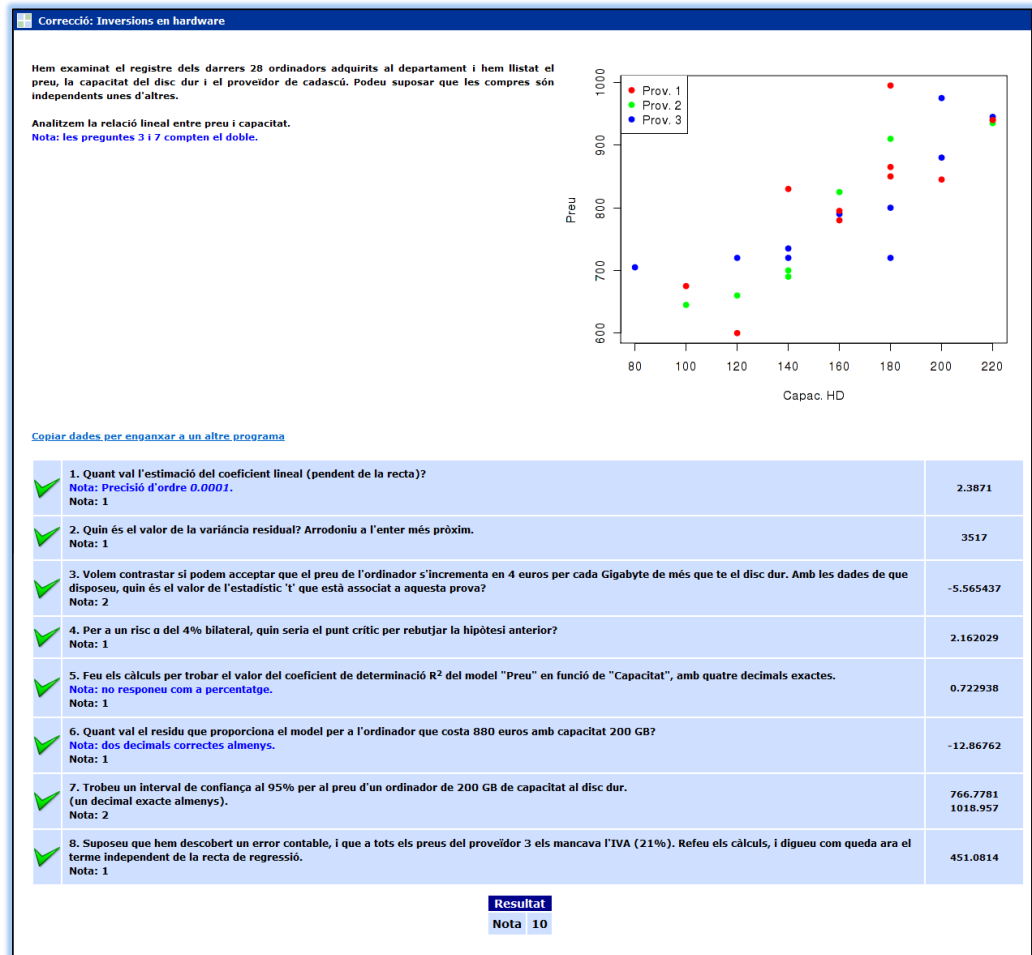


Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH

Inversions en hardware



Script en R

Antes de ejecutar el script:

- Clicar en "Copiar dades per enganxar a un altre programa"
- Seleccionar los datos
- Pulsar Ctrl+C

Preu Capacitat Proveïdor

```
1 940 220 1
2 720 120 3
3 690 140 2
4 660 120 2
5 705 80 3
6 845 200 1
7 910 180 2
8 830 140 1
9 700 140 2
10 850 180 1
11 720 180 3
12 825 160 2
13 975 200 3
14 720 140 3
15 600 120 1
16 780 160 1
17 645 100 2
18 945 220 3
19 735 140 3
20 935 220 2
21 790 160 3
22 800 180 3
23 865 180 1
```

© <http://fibernalia.blogspot.com>

WUOLAH

```

24 940 220 1
25 880 200 3
26 675 100 1
27 795 160 1
28 995 180 1

dades = read.table("clipboard", header=TRUE)
model = summary(lm(dades$Preu ~ dades$Capacitat))

# Quant val l'estimació del coeficient lineal (pendent de la
recta)?

b1 = model$coefficients[2]; b1

# Quin és el valor de la variància residual? Arrodoniu a
l'enter més pròxim.

varRes = round((model$sigma)^2); varRes

# Quin és el valor de l'estadístic 't' que està associat a
aquesta prova (el preu de l'ordinador s'incrementa en 4 euros
per cada Gigabyte de més que te el disc dur)?

sb1 = model$coefficients[4];
t = (b1-4)/sb1; t

# Per a un risc  $\alpha$  del 4% bilateral, quin seria el punt crític
per rebutjar la hipòtesi anterior?

alfa = 4/100

df = 28-2 # df = length(dades$Preu)-2 = model$df[2]

puntCritic = qt(1-alfa/2,df); puntCritic

# Feu els càlculs per trobar el valor del coeficient de
determinació  $R^2$  del model "Preu" en funció de "Capacitat",
amb quatre decimals exactes.

r2 = model$r.squared; r2

# Quant val el residu que proporciona el model per a
l'ordinador que costa 880 euros amb capacitat 200 GB?

b0 = model$coefficients[1];
residu = 880 - b0 - b1*200; residu

# Trobeu un interval de confiança al 95% per al preu d'un
ordinador de 200 GB de capacitat al disc dur.

preu = dades$Preu

```



```

capacitat = dades$Capacitat
interval.lm = lm(preu~capacitat)
dataframe = data.frame(capacitat = 200)
p7 = predict(interval.lm, dataframe, interval="predict")
p7[2]; p7[3]

# Supposeu que hem descobert un error contable, i que a tots
els preus del proveïdor 3 els mancava l'IVA (21%). Refeu els
càlculs, i digueu com queda ara el terme independent de la
recta de regressió.

for (i in 1:length(dades$Preu)) { if (dades$Proveïdor[i] ==
3) dades$Preu[i] = dades$Preu[i] + 0.21*dades$Preu[i] }

model = summary(lm(dades$Preu ~ dades$Capacitat))
b0 = model$coefficients[1]; b0

# Preguntas de otras ejecuciones

# Trobeu un interval de confiança al 95% per al preu mitjà
d'un ordinador de 120 GB de capacitat al disc dur.

# predict(interval.lm, dataframe, interval="confidence")

```

Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

pierdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH

Consola de R

```
> dades = read.table("clipboard", header=TRUE)
> model = summary(lm(dades$Preu ~ dades$Capacitat))
> b1 = model$coefficients[2]; b1
[1] 2.387069
> varRes = round((model$sigma)^2); varRes
[1] 3517
> sb1 = model$coefficients[4];
> t = (b1-4)/sb1; t
[1] -5.565437
> alfa = 4/100
> df = 28-2
> puntCritic = qt(1-alfa/2,df); puntCritic
[1] 2.162029
> r2 = model$r.squared; r2
[1] 0.7229379
> b0 = model$coefficients[1];
> residu = 880 - b0 - b1*200; residu
[1] -12.86762
> preu = dades$Preu
> capacitat = dades$Capacitat
> interval.lm = lm(preu~capacitat)
> dataframe = data.frame(capacitat = 200)
> p7 = predict(interval.lm, dataframe, interval="predict")
> p7[2]; p7[3]
[1] 766.7781
[1] 1018.957
> for (i in 1:length(dades$Preu)) { if (dades$Proveedor[i] ==
3) dades$Preu[i] = dades$Preu[i] + 0.21*dades$Preu[i] }
> model = summary(lm(dades$Preu ~ dades$Capacitat))
> b0 = model$coefficients[1]; b0
[1] 451.0814
```

Reservados todos los derechos. No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

Te has descargado este apunte gracias a la publicidad. También puedes eliminarla con 1 coin.

Script en R

Antes de ejecutar el script:

- Clicar en "Copiar dades per enganxar a un altre programa"
- Seleccionar los datos
- Pulsar Ctrl+C

seg	score
40	0
50	6
30	3
25	20
40	9
60	16
35	8
35	13
20	7
25	6
45	10
45	4
55	4
25	0
35	0
20	4
35	14
60	9
40	21
45	5
15	0
45	3
40	18
15	15
60	9
40	14
55	14
55	14
55	17
20	10
20	12
15	8
30	0
30	1
20	0
15	11
25	12
25	14
15	11
30	15
35	9
50	6
60	15
55	14
50	7
50	8
45	12
50	0
30	5
60	9

```

dades = read.table("clipboard", header=TRUE)

model = summary(lm(dades$score ~ dades$seg))

# Dar la estimación del parámetro lineal (pendiente de la
recta) del modelo de regresión lineal para la relación entre
la longitud del anuncio y el resultado de la encuesta

b1 = model$coefficients[2]; b1

# Dar el error estándar del estimador anterior

sb1 = model$coefficients[4]; sb1

# Con un riesgo alfa del 5%, ¿podemos concluir que la
longitud del anuncio y la puntuación del test de memoria
están linealmente relacionados? (1: si, 0: no)

alfa = 5/100

n = length(dades$seg)

tb1 = model$coefficients[6] # tb1 = b1/sb1

ifelse(tb1 > qt(1-alfa/2,n-2), 1, 0)

# ¿Qué longitud tendría un anuncio visto por cierto
individuo, el cual le hubiera asignado una puntuación de 0?

b0 = model$coefficients[1]

(0-b0)/b1

# ¿Qué puntuación se obtendría de un individuo que ve un
anuncio de 35 segundos?

b0 + b1*35

# Dar un intervalo de confianza del 95% de la predicción
anterior

seg = dades$seg

score = dades$score

interval.lm = lm(score~seg)

dataframe = data.frame(seg = 35)

p6 = predict(interval.lm, dataframe, interval="predict")

```


Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato
→ Planes pro: más coins

pierdo
espacio



Necesito
concentración

ali ali ooh
esto con 1 coin me
lo quito yo...

WUOLAH



```
p6[2]; p6[3]
```

```
# Estimar por IC del 95% la puntuación esperada de todos los  
anuncios que duran 35 segundos
```

```
p7 = predict(interval.lm, dataframe, interval="confidence")
```

```
p7[2]; p7[3]
```

```
# Obtenga una estimación del coeficiente de determinación  $R^2$ 
```

```
r2 = model$r.squared; r2
```

```
# Obtenga el diagrama de residuos estandarizados vs la  
duración del anuncio, y diga cuál de las siguientes es la  
salida correcta
```

```
plot(lm(dades$score ~ dades$seg))
```

```
# Preguntas de otras ejecuciones
```

```
# Dar la estimación del parámetro independiente (término  
constante) del modelo de regresión lineal para la relación  
entre la longitud del anuncio y el resultado de la encuesta
```

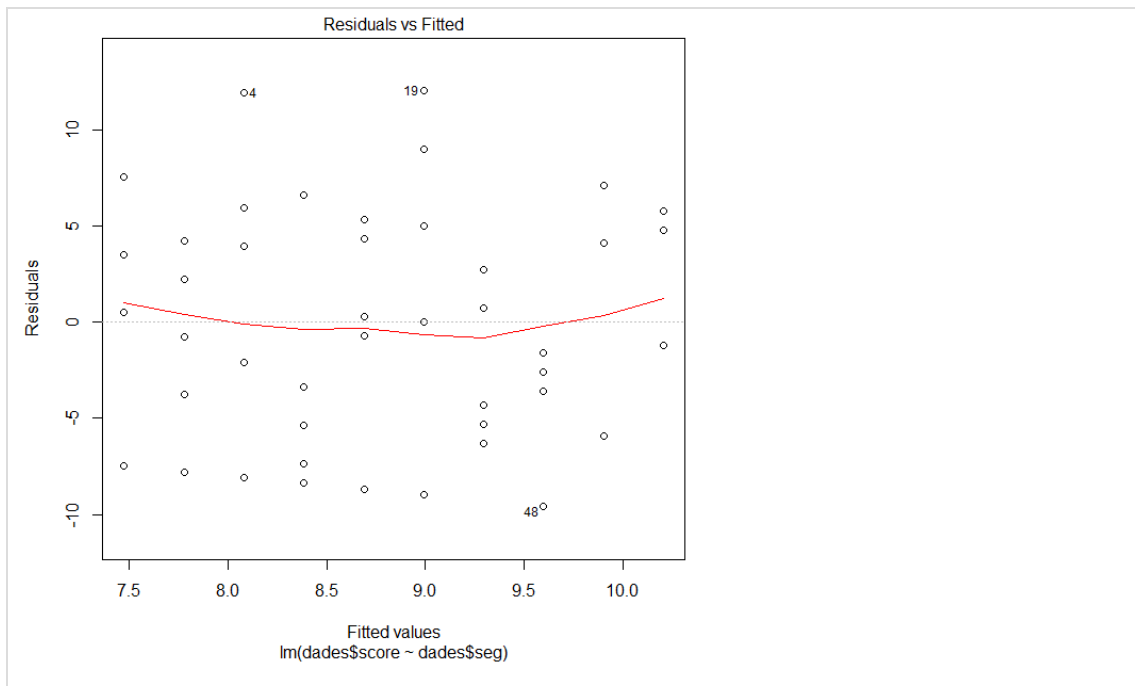
```
# b0 = model$coefficients[1]; b0
```

```
# Dar el error estándar del estimador anterior
```

```
# sb0 = model$coefficients[3]; sb0
```

Consola de R

```
> dades = read.table("clipboard", header=TRUE)
> model = summary(lm(dades$Preu ~ dades$Capacitat))
> b1 = model$coefficients[2]; b1
[1] 0.06060606
> sb1 = model$coefficients[4]; sb1
[1] 0.05647635
> alfa = 5/100
> n = length(dades$seg)
> tb1 = model$coefficients[6]
> ifelse(tb1 > qt(1-alfa/2,n-2), 1, 0)
[1] 0
> b0 = model$coefficients[1]
> (0-b0)/b1
[1] -108.36
> b0 + b1*35
[1] 8.688485
> seg = dades$seg
> score = dades$score
> interval.lm = lm(score~seg)
> dataframe = data.frame(seg = 35)
> p6 = predict(interval.lm, dataframe, interval="predict")
> p6[2]; p6[3]
[1] -2.961111
[1] 20.33808
> p7 = predict(interval.lm, dataframe, interval="confidence")
> p7[2]; p7[3]
[1] 7.033175
[1] 10.34379
> r2 = model$r.squared; r2
[1] 0.02342941
> plot(lm(dades$score ~ dades$seg))
```



Análisis de residuos

Corrección: Análisis de residuos		
<p>El crecimiento de un bebé en los primeros meses puede estar condicionado por varios factores. Uno de ellos puede ser el peso al nacer (X), pero no necesariamente la relación entre esta variable y, por ejemplo, el incremento de peso entre los días 70 y 100 de vida (Y) puede ajustarse a través de un modelo de regresión lineal. Para comprobar que se cumplen las premisas se debe hacer un análisis exhaustivo de los residuos. En cada pregunta, observe los siguientes gráficos:</p>		
1. Y vs. X (Nótese que la pendiente negativa denota que los recién nacidos más pesados experimentan un menor crecimiento posterior)	2. Residuos vs. valores predichos	3. QQ-plot de Normalidad
<p>1. ¿Cuál es la premisa que no se cumple?</p> <ol style="list-style-type: none"> 1. Todas se cumplen 2. Linealidad 3. Homocedasticidad 4. Normalidad de los residuos 		
Nota: 2.5		
<p>2. ¿Cuál es la premisa que no se cumple?</p> <ol style="list-style-type: none"> 1. Todas se cumplen 2. Linealidad 3. Homocedasticidad 4. Normalidad de los residuos 		
Nota: 2.5		
<p>3. ¿Cuál es la premisa que no se cumple?</p> <ol style="list-style-type: none"> 1. Todas se cumplen 2. Linealidad 3. Homocedasticidad 4. Normalidad de los residuos 		
Nota: 2.5		
<p>4. ¿Cuál es la premisa que no se cumple?</p> <ol style="list-style-type: none"> 1. Todas se cumplen 2. Linealidad 3. Homocedasticidad 4. Normalidad de los residuos 		
Nota: 2.5		

Resultat
Nota 10