

## Estadística (Q) (7/11/2013)

### Un ejemplo de Regresión lineal

---

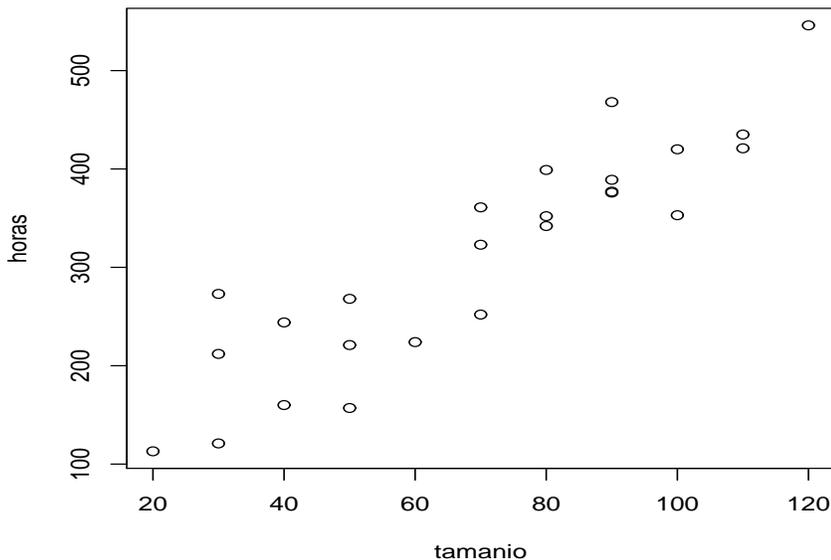
1. La Compañía Toluca produce repuestos para equipos de refrigeración. En el pasado, uno de estos repuestos se ha producido periódicamente en lotes de diversos tamaños. Se desea determinar el tamaño óptimo del lote para producir este repuesto. Una cuestión importante a la hora de determinar el tamaño de lote óptimo consiste en estudiar la relación entre el tamaño de lote y la cantidad de horas de trabajo requeridas para producir dicho lote. Para determinar tal relación se utilizaron los datos del tamaño de lote y cantidad de horas de trabajo correspondientes a 25 corridas de producción recientes. Las condiciones de producción fueron estables durante el período de seis meses en que fueron hechas las 25 corridas y se espera que continúen estables durante los próximos tres años. Las siguiente tabla contiene los datos de las 25 corridas:

Tamaño	Horas
20	113
30	121
30	212
30	273
40	160
40	244
50	221
50	157
50	268
60	224
70	252
70	361
70	323
80	399
80	342
80	352
90	389
90	468
90	377
90	376
100	353
100	420
110	435
110	421
120	546

- a) Representar los datos en un diagrama de dispersión y decidir si es razonable suponer que existe una relación lineal entre tamaño y horas.
- b) Proponer un modelo de regresión lineal para estos datos, definiendo claramente las variables y los parámetros involucrados junto con los supuestos del modelo.
- c) Calcular los coeficientes de la recta de cuadrados mínimos y graficarla sobre el diagrama de dispersión

- d) Calcular intervalos de confianza de nivel 0.95 para la pendiente y la ordenada al origen de la recta.
- e) Decidir si la pendiente es significativa al 5% (es decir, si es significativamente distinta de 0 al 5%). Escribir las hipótesis, el estadístico del test y su distribución bajo la hipótesis nula. Hallar el p-valor. Interpretar
- f) Calcular el coeficiente de determinación  $R^2$  e interpretar su significado en este contexto.
- g) Se desea producir un nuevo lote de 75 unidades. Predecir la cantidad de horas de trabajo que requerirá su producción y dar el intervalo de predicción correspondiente. ¿Cuál es la cantidad esperada de horas que requerirá la producción del lote? Dar un intervalo de confianza de nivel 0.95 para dicha cantidad esperada.

```
#para leer el archivo
# antes de empezar cambiar el directorio de trabajo al directorio donde
# está el archivo así:
setwd( "C:/Users/Dropbox/Estadística/segundo2013")
# o bien clickeando en Archivo y después en Cambiar dir y eligiendo el directorio
toluca=read.table("toluca.txt",header=TRUE)
# otra opción es poner directamente toluca=read.table(file.choose(),header=TRUE)
# y elegir el archivo a mano
attach(toluca)
#item a)
plot(tamano,horas)
```



```
#item c)
tolucalm=lm(horas~tamano)
summary(tolucalm)
```

Call:

```
lm(formula = horas ~ tamano)
```

Residuals:

Min	1Q	Median	3Q	Max
-83.876	-34.088	-5.982	38.826	103.528

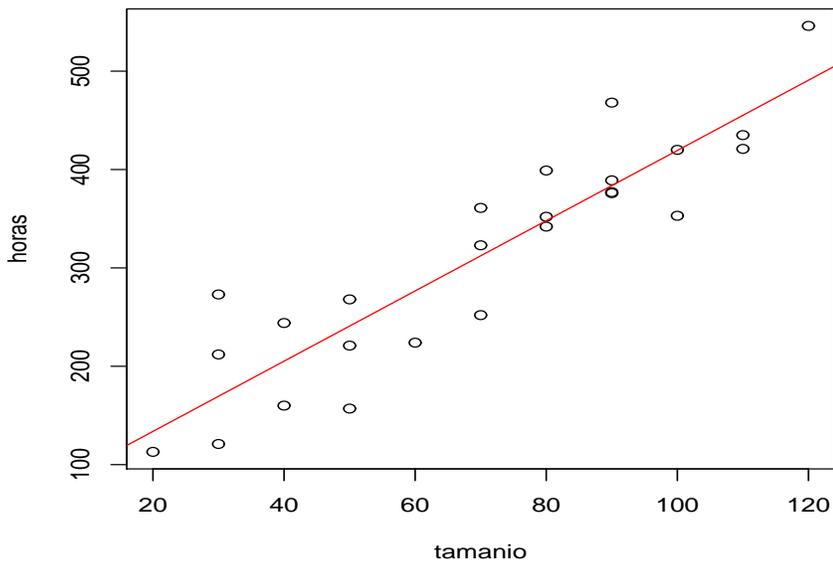
Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	62.366	26.177	2.382	0.0259 *
tamano	3.570	0.347	10.290	4.45e-10 ***

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

Residual standard error: 48.82 on 23 degrees of freedom  
Multiple R-squared: 0.8215, Adjusted R-squared: 0.8138  
F-statistic: 105.9 on 1 and 23 DF, p-value: 4.449e-10

```
abline(tolucalm,col=2)
```



```
#item g)
nuevo <- data.frame(tamano = 75)
predict(tolucalm, nuevo, interval="prediction")
  fit      lwr      upr
1 330.131 227.0696 433.1924
predict(tolucalm, nuevo, interval="confidence")
  fit      lwr      upr
1 330.131 309.6149 350.6471
```