

La Compañía Toluca produce equipos de refrigeración así como diversos repuestos para estos. En el pasado, una de estas piezas de repuesto se ha producido periódicamente en lotes de diversos tamaños. Se desea determinar el tamaño óptimo del lote para producir este repuesto. La producción de este repuesto involucra modificar el proceso de producción (lo cual debe hacerse no importa el tamaño de lote que se desee producir) y también involucra operaciones de ensamblaje y movimiento de maquinaria. 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 trabajadas requeridas para producir dicho lote. Para determinar tal relación se utilizaron los datos de 25 corridas recientes de producción. Sean:

X_i : tamaño del lote de la corrida i ésima

Y_i : cantidad de horas trabajadas en la producción de la corrida i ésima

Las condiciones de producción fueron estables durante el periodo de seis meses en el que las 25 corridas se realizaron, y se esperaba que continuaran así durante los siguientes 3 años. A continuación los datos:

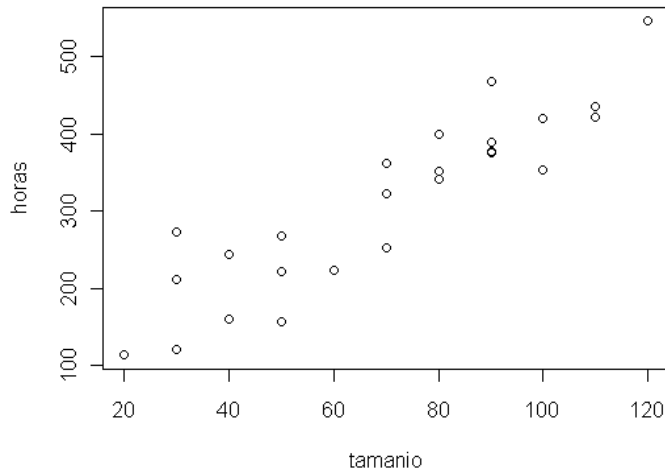
Tamaño	Horas				
		50	268	90	468
20	113	60	224	90	377
30	121	70	252	90	376
30	212	70	361	100	353
30	273	70	323	100	420
40	160	80	399	110	435
40	244	80	342	110	421
50	221	80	352	120	546
50	157	90	389		

Hacemos un scatter plot (gráfico de Y versus X) para ver si tiene sentido ajustar una regresión lineal a los datos:

Para ello, en R: cargamos los datos

```
> toluca<-read.table("toluca.txt", header=T)
> attach(toluca)
> plot(tamano, horas, main="Scatter plot de horas versus tamano")
```

Scatter plot de horas versus tamaño



Ajuste de mínimos cuadrados a los datos.

Ponemos “tamaño” como variable independiente, “horas” como variable dependiente y obtenemos:

```
> salida<-lm(horas ~ tamaño)
> summary(salida)
```

```
Call:
lm(formula = horas ~ tamaño)
```

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 *
tamaño	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

```
> anova(salida)
Analysis of Variance Table
```

Response: horas

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tamaño	1	252378	252378	105.88	4.449e-10 ***
Residuals	23	54825	2384		

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

Para calcular los valores ajustados por el modelo, y los residuos correspondientes a los valores medidos, usamos la salida del lm.

Datos con los valores ajustados y los residuos calculados:

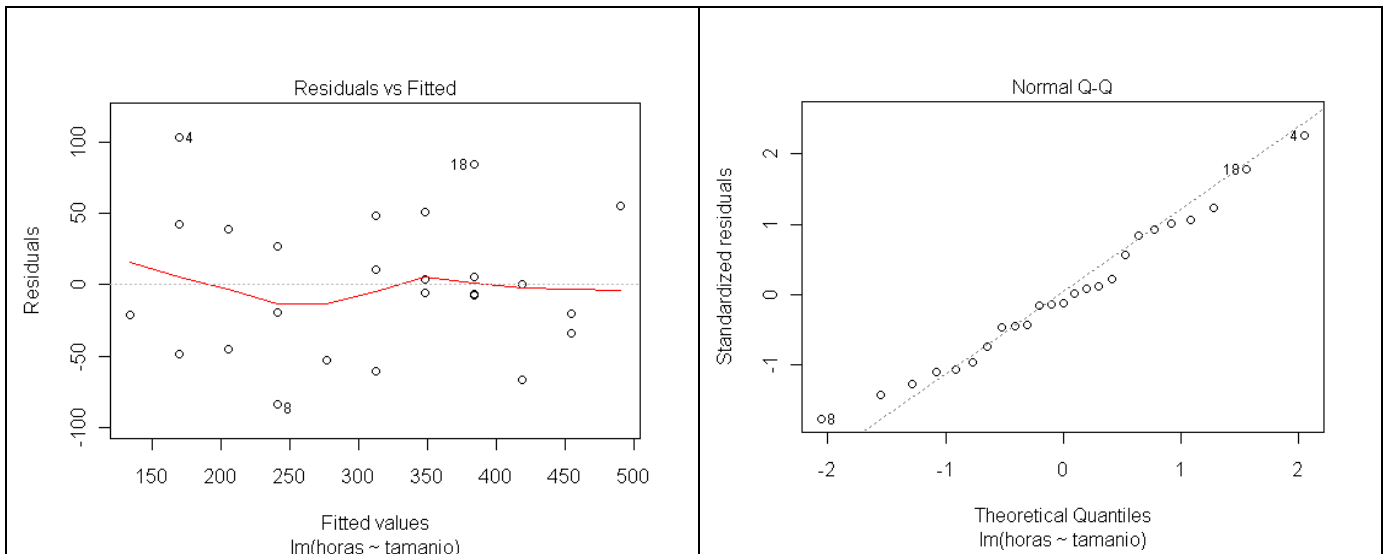
```

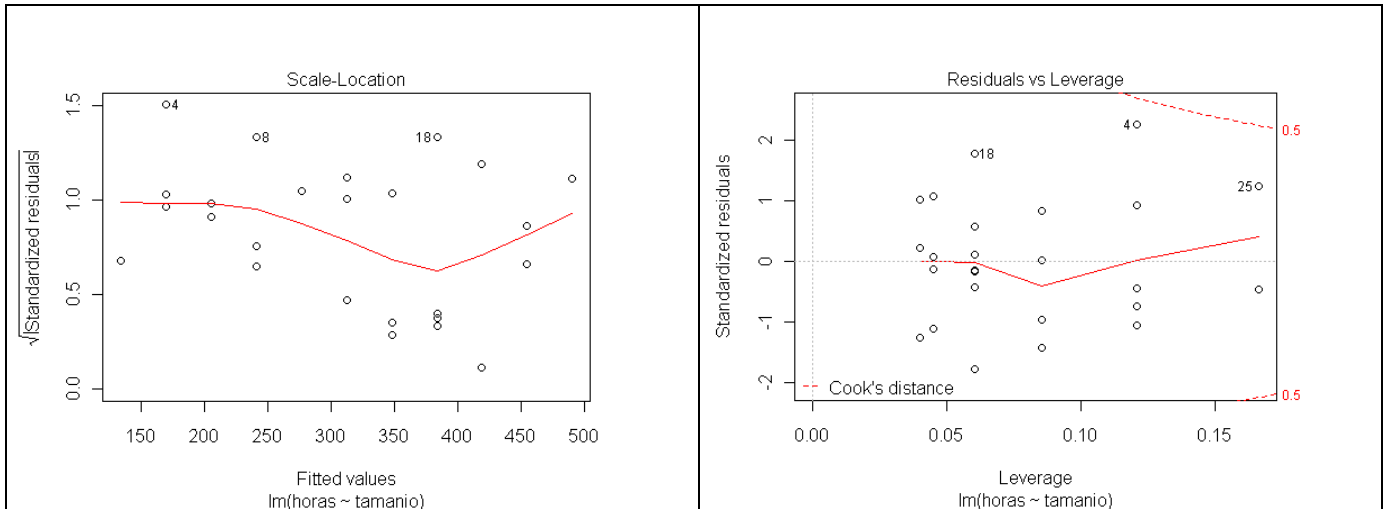
> tabla<-
cbind(tamano, horas, salida$fi tted. val ues, salida$resi dual s, rstandard(sal ida))
> tabla
  tamano horas
1      20   113 133.7699 -20.7698990 -0.46589980
2      30   121 169.4719 -48.4719192 -1.05881760
3      30   212 169.4719  42.5280808  0.92898076
4      30   273 169.4719 103.5280808  2.26146098
5      40   160 205.1739 -45.1739394 -0.96751587
6      40   244 205.1739  38.8260606  0.83155975
7      50   221 240.8760 -19.8759596 -0.41993651
8      50   157 240.8760 -83.8759596 -1.77211959
9      50   268 240.8760  27.1240404  0.57307295
10     60   224 276.5780 -52.5779798 -1.10201238
11     70   252 312.2800 -60.2800000 -1.26011565
12     70   361 312.2800  48.7200000  1.01846109
13     70   323 312.2800  10.7200000  0.22409489
14     80   399 347.9820  51.0179798  1.06931543
15     80   342 347.9820  -5.9820202 -0.12538063
16     80   352 347.9820   4.0179798  0.08421517
17     90   389 383.6840   5.3159596  0.11231485
18     90   468 383.6840  84.3159596  1.78141585
19     90   377 383.6840  -6.6840404 -0.14121947
20     90   376 383.6840  -7.6840404 -0.16234734
21    100   353 419.3861 -66.3860606 -1.42182789
22    100   420 419.3861   0.6139394  0.01314909
23    110   435 455.0881 -20.0880808 -0.43880279
24    110   421 455.0881 -34.0880808 -0.74461792
25    120   546 490.7901  55.2098990  1.23844036

```

Gráficos de los residuos

Para hacerlos con el R: `> plot(sal ida)`



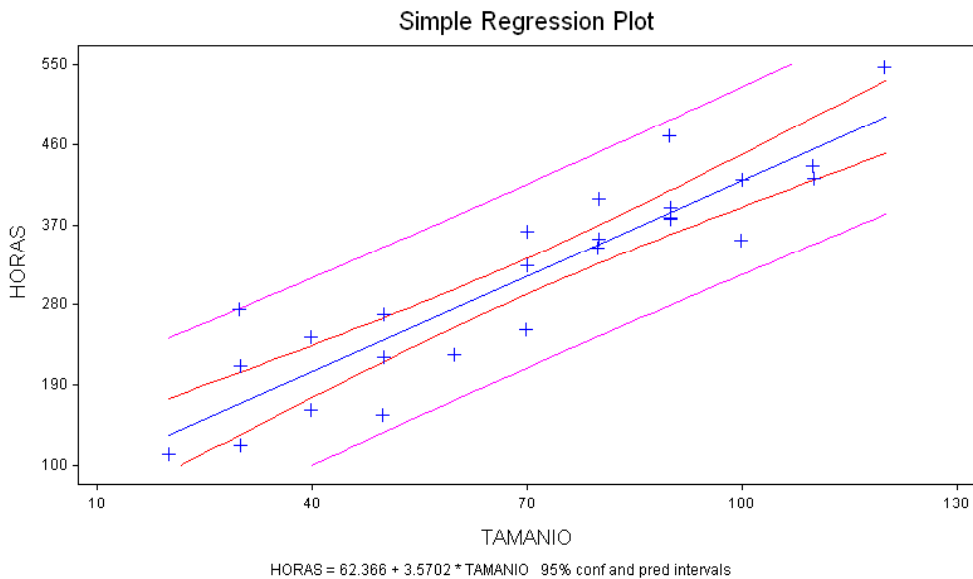


1. Residuos versus valores ajustados
2. QQ plot de los residuos estandarizados

Test de Shapiro-Wilk (de los residuos) $W = 0.9786$; p valor 0.8563

3. Gráfico de la raíz cuadrada del modulo de los residuos estándar versus los valores ajustados
4. Leverage.

Gráfico de “horas” (Y) versus “tamaño” (X), con la recta ajustada e intervalos de confianza para el valor esperado de Y conociendo X, y de predicción para una nueva observación de Y con nivel prefijado de X (ambos del 95%).



DESCRIPTIVE STATISTICS

	TAMANIO	HORAS	RESIDUOS	STD_RES	AJUSTADOS
N	25	25	25	25	25
MEAN	70.000	312.28	-2.956E-14	2.235E-03	312.28
SD	28.723	113.14	47.795	1.0243	102.55
VARIANCE	825.00	12800	2284.4	1.0491	10516
MINIMUM	20.000	113.00	-83.876	-1.7721	133.77
MAXIMUM	120.00	546.00	103.53	2.2615	490.79

Percentil de la t:

$$p = 0.975, df = n - 2 = 23$$

obtenemos

```
> qt(0.975, df = 23 )
[1] 2.068658
```

Intervalos de confianza para el valor esperado de Y dado X = x*, e Intervalos de predicción para el valor Y de una nueva observación realizada cuando X = x*:

```
> predict(salida, interval = "confidence")
> predict(salida, interval = "prediction", level = 0.95)
Juntamos ambas salidas en una para ahorrar espacio:
>
cbind(predict(salida, interval = "confidence", level = 0.95), predict(salida, interval = "prediction", level = 0.95))
  fit      lwr      upr      fit      lwr      upr
1 133.7699  92.58736 174.9524 133.7699  24.69771 242.8421
2 169.4719 134.36734 204.5765 169.4719  62.54638 276.3975
3 169.4719 134.36734 204.5765 169.4719  62.54638 276.3975
4 169.4719 134.36734 204.5765 169.4719  62.54638 276.3975
5 205.1739 175.64938 234.6985 205.1739  99.94828 310.3996
6 205.1739 175.64938 234.6985 205.1739  99.94828 310.3996
7 240.8760 216.09481 265.6571 240.8760 136.88151 344.8704
8 240.8760 216.09481 265.6571 240.8760 136.88151 344.8704
9 240.8760 216.09481 265.6571 240.8760 136.88151 344.8704
10 276.5780 255.14090 298.0151 276.5780 173.32931 379.8267
11 312.2800 292.08026 332.4797 312.2800 209.28112 415.2789
12 312.2800 292.08026 332.4797 312.2800 209.28112 415.2789
13 312.2800 292.08026 332.4797 312.2800 209.28112 415.2789
14 347.9820 326.54494 369.4191 347.9820 244.73335 451.2307
15 347.9820 326.54494 369.4191 347.9820 244.73335 451.2307
16 347.9820 326.54494 369.4191 347.9820 244.73335 451.2307
17 383.6840 358.90290 408.4652 383.6840 279.68959 487.6785
18 383.6840 358.90290 408.4652 383.6840 279.68959 487.6785
19 383.6840 358.90290 408.4652 383.6840 279.68959 487.6785
20 383.6840 358.90290 408.4652 383.6840 279.68959 487.6785
21 419.3861 389.86150 448.9106 419.3861 314.16040 524.6117
22 419.3861 389.86150 448.9106 419.3861 314.16040 524.6117
23 455.0881 419.98350 490.1927 455.0881 348.16254 562.0136
24 455.0881 419.98350 490.1927 455.0881 348.16254 562.0136
25 490.7901 449.60756 531.9726 490.7901 381.71792 599.8623
Mensajes de aviso perdidos
In predict.lm(salida, interval = "prediction", level = 0.95) :
  Predictions on current data refer to _future_ responses
```