Regresión: Tarea por Grupos

Cuervo Fernández, Esther García Prado, Sergio Martín Villares, Pablo

26 de noviembre de 2017

Parte I

Ejercicio Kutner: Mantenimiento de Copiadoras

1.20. Copier maintenance. The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data below have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and Y is the total number of minutes spent by the service person. Assume that first-order regression model (1) is appropriate.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1}$$

a. Obtain the estimated regression function.

[TODO]

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	46
Number of Observations Used	45
Number of Observations with Missing Values	1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	76960	76960	968.66	<.0001
Error	43	3416.37702	79.45063		
Corrected Total	44	80377			

Root MSE	8.91351	R-Square	0.9575
Dependent Mean	76.26667	Adj R-Sq	0.9565
Coeff Var	11.68729		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.58016	2.80394	-0.21	0.8371
X	1	15.03525	0.48309	31.12	<.0001

Figura 1: *Salida SAS*: Mantenimiento de Copiadoras - Resumen de Regresión Lineal Simple

$$\widehat{\beta}_0 = -0{,}58016 \tag{2}$$
$$\widehat{\beta}_1 = 15{,}03525 \tag{3}$$
$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i + \epsilon_i = -0{,}58016 + 15{,}03525 X_i + \epsilon_i \tag{4}$$

b. Plot the estimated regression function and the data. How well does the estimated regression function fit the data?

[TODO]

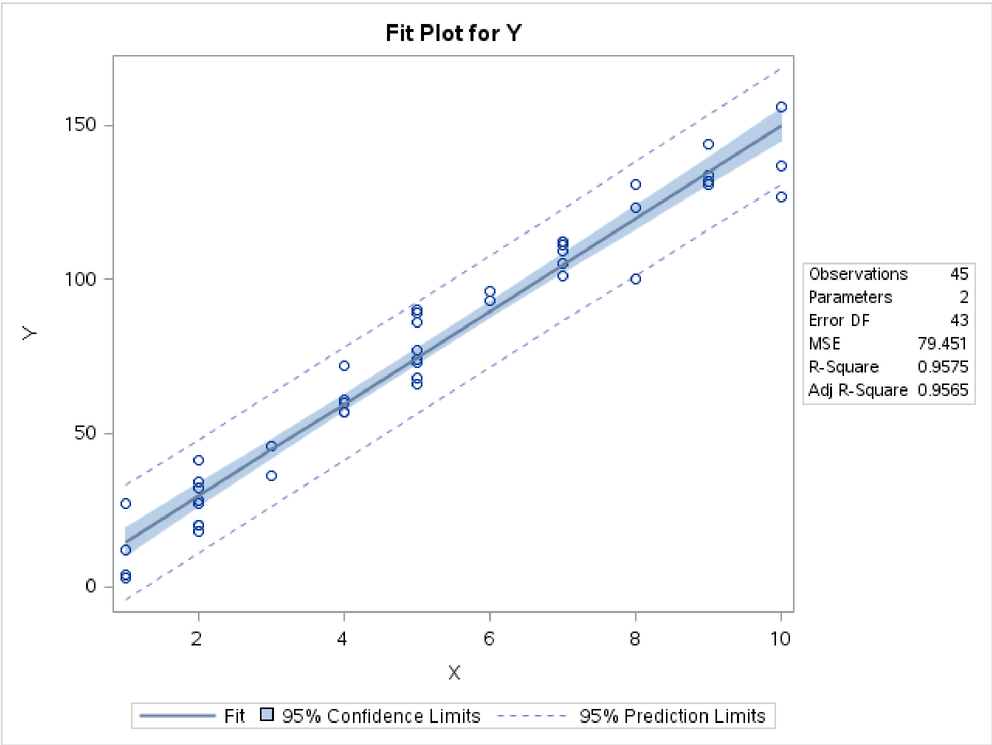

Figura 2: *Salida SAS*: Mantenimiento de Copiadoras - Gráfico de Regresión Lineal Simple

c. Interpret $\widehat{\beta}_0$ in your estimated regression function. Does $\widehat{\beta}_0$ provide any relevant information here? Explain.

[TODO]

d. Obtain a point estimate of the mean service time when $X = 5$ copiers are serviced.

[TODO]

$$E\left[\widehat{Y} \mid X = 5\right] = E\left[\widehat{\beta}_0 + \widehat{\beta}_1 X + \epsilon \mid X = 5\right] = -0{,}58016 + 15{,}03525 * 5 + 0 = 74{,}5961 \tag{5}$$
$$Var\left[\widehat{Y} \mid X = 5\right] = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right) = 1{,}1531^2 = 1{,}3298 \tag{6}$$

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Output Statistics								
Obs	X	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	Cook's D
1	2	20	29.4903	2.0061	-9.4903	8.685	-1.093	0.032
2	4	60	59.5608	1.4331	0.4392	8.798	0.050	0.000
45	5	77	74.5961	1.3298	2.4039	8.814	0.273	0.001
46	5	.	74.5961	1.3298	.	.	.	.

Figura 3: *Salida SAS*: Mantenimiento de Copiadoras - Prediccion de Regresión Lineal Simple para $X = 5$

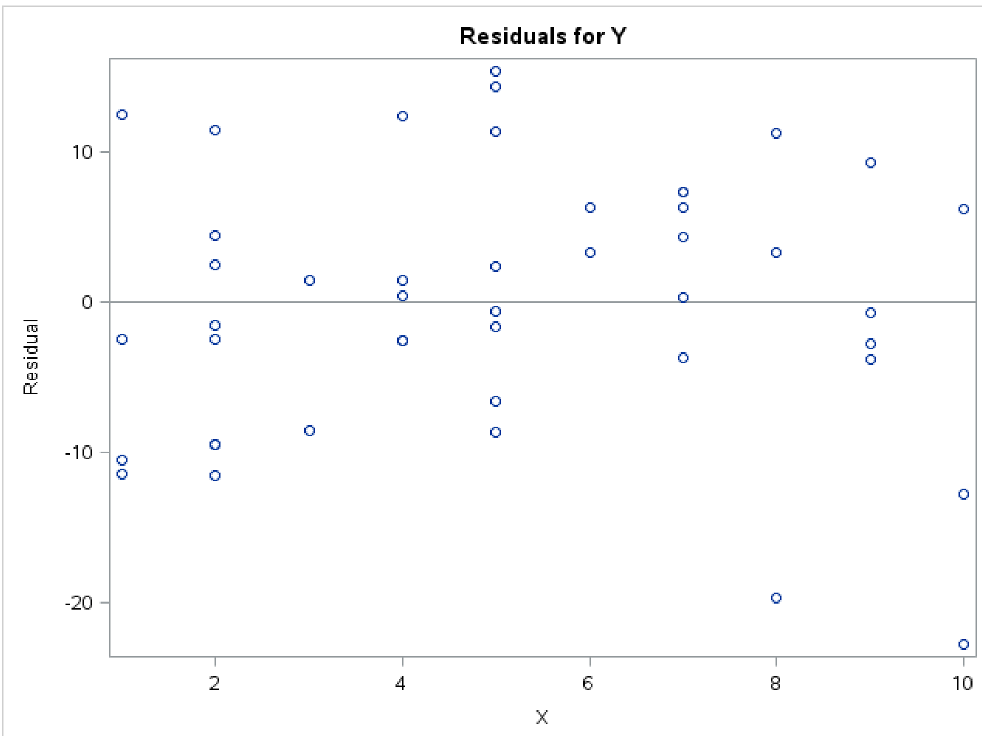

Figura 4: *Salida SAS*: Mantenimiento de Copiadoras - Gráfico de Residuos de Regresión Lineal Simple

1.24. Refer to Copier maintenance Problem 1.20.

a. Obtain the residuals e_i and the sum of the squared residuals $\sum_i e_i^2$. What is the relation between the sum of the squared residuals here and the quantity Q in (7)?

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 = 3416{,}37702 \qquad (7)$$

[TODO]

b. Obtain point estimates of σ^2 and σ. In what units is σ expressed?

[TODO σ en las unidades de Y]

$$\widehat{\sigma}^2 = MSE = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n-2} = \frac{3416{,}37702}{45-2} = 79{,}45063 \tag{8}$$

$$\widehat{\sigma} = \sqrt{\widehat{\sigma}^2} = \sqrt{79{,}45063} = 8{,}91351 \tag{9}$$

2.5. Refer to Copier maintenance Problem 1.20.

a. Estimate the change in the mean service time when the number of copiers serviced increases by one. Use a 90 % confidence interval. Interpret your confidence interval.

[TODO]

$$\widehat{\beta}_1 = 15{,}03525 \tag{10}$$

$$t_{n-2;1-\frac{\alpha}{2}} = \tag{11}$$

$$Var\left[\widehat{\beta}_1\right] = \tag{12}$$

$$\text{IC} = \left[\widehat{\beta}_1 \pm t_{n-2;1-\frac{\alpha}{2}}\sqrt{Var\left[\widehat{\beta}_1\right]}\right] \tag{13}$$

$$= \tag{14}$$

b. Conduct a *t-test* to determine whether or not there is a linear association between X and Y here; control the α risk at 0,10. State the alternatives, decision rule, and conclusion. What is the *P-value* of your test?

[TODO]

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned} \tag{15}$$

$$\text{P-value} = Pr\left[\frac{\widehat{\beta}_1 - 0}{\sqrt{Var\left[\widehat{\beta}_1\right]}} > t_{n-2;1-\frac{\alpha}{2}}\right] \tag{16}$$

c. Are your results in parts (a) and (b) consistent? Explain.

[TODO]

d. The manufacturer has suggested that the mean required time should not increase by more than 14 minutes for each additional copier that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Control the risk of a Type I error at 0,05. State the alternatives, decision rule, and conclusion. What is the *P-value* of the test?

[TODO]

$$\begin{aligned} H_0 &: \beta_1 \leq 14 \\ H_1 &: \beta_1 > 14 \end{aligned} \tag{17}$$

$$\text{P-value} = Pr\left[\frac{\widehat{\beta}_1 - 14}{\sqrt{Var\left[\widehat{\beta}_1\right]}} > t_{n-2;1-\alpha}\right] \tag{18}$$

e. **Does $\widehat{\beta}_0$ give any relevant information here about the start-up time on calls-i.e., about the time required before service work is begun on the copiers at a customer location?**

[TODO]

$$\begin{aligned} H_0 &: \beta_0 = 0 \\ H_1 &: \beta_0 \neq 0 \end{aligned} \tag{19}$$

$$\text{P-value} = Pr\left[\frac{\widehat{\beta}_0 - 0}{\sqrt{Var\left[\widehat{\beta}_0\right]}} > t_{n-2;1-\frac{\alpha}{2}}\right] \tag{20}$$

2.14. Refer to Copier maintenance Problem 1.20.

a. **Obtain a 90 % confidence interval for the mean service time on calls in which six copiers are serviced. Interpret your confidence interval.**

[TODO intervalo de confianza para $X = 6$]

$$\widehat{Y}_h = \tag{21}$$
$$t_{n-2;1-\frac{\alpha}{2}} = t_{43;0,95} = \tag{22}$$
$$Var\left[\widehat{Y}_h\right] = \sigma = \tag{23}$$
$$\tag{24}$$

$$\text{IC} = \left[\widehat{Y}_h \pm t_{n-2;1-\frac{\alpha}{2}} Var\left[\widehat{Y}_h\right]\right] \tag{25}$$

b. **Obtain a 90 % prediction interval for the service time on the next call in which six copiers are serviced. Is your prediction interval wider than the corresponding confidence interval in part (a)? Should it be?**

[TODO intervalo de predicción para $X = 6$]

$$E\left[\widehat{Y}_h\right] = \tag{26}$$
$$z_{1-\frac{\alpha}{2}} = z_{0,95} = \tag{27}$$
$$Var\left[\widehat{Y}_h\right] = \sigma = \tag{28}$$
$$\tag{29}$$

$$\text{IP} = \left[E\left[\widehat{Y}_h\right] \pm z_{1-\frac{\alpha}{2}} Var\left[\widehat{Y}_h\right]\right] \tag{30}$$

2.24. Refer to Copier maintenance Problem 1.20.

b. Conduct an *F-test* to determine whether or not there is a linear association between time spent and number of copiers serviced; use $\alpha = 0{,}1$. State the alternatives, decision rule, and conclusion.

[TODO]

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned} \tag{31}$$

$$\text{P-value} = Pr\left[\frac{MSM}{MSE} > F_{1;n-2;1-\frac{\alpha}{2}}\right] \tag{32}$$

$$= Pr\left[\frac{76960}{79{,}45063} > F_{1;43;0{,}95}\right] \tag{33}$$

$$< 0{,}0001 \tag{34}$$

c. By how much, relatively, is the total variation in number of minutes spent on a call-reduced when the number of copiers serviced is introduced into the analysis? Is this a relatively small or large reduction? What is the name of this measure?

[TODO]

d. Calculate r and attach the appropriate sign.

[TODO]

Parte II
Ejercicios Montgomery:

[TODO]

Parte III
Código Fuente

```
filename reffile '/folders/myshortcuts/sas/regression-group-task/data/CH01PR20.csv';

proc import datafile=reffile dbms=csv out=copiers;
  getnames=yes;
run;

proc print data=copiers;
run;
```

Figura 5: *Código SAS:* Mantenimiento de Copiadoras - Importación del conjunto de datos.

```
proc reg data=copiers;
  model y=x;
  id x;
run;
```

Figura 6: *Código SAS:* Mantenimiento de Copiadoras - Modelo de regresión simple.

```
data copiers_new_observation;
  x = 5;
run;

proc append base=copiers data=copiers_new_observation;
run;

proc reg data=copiers;
  model y = x /r clm;
  id x;
run;
```

Figura 7: *Código SAS:* Mantenimiento de Copiadoras - Predicción para $X = 5$.

Referencias

[1] Barba Escribá, L. Regresión y ANOVA, 2017/18. Facultad de Ciencias: Departamento de Estadística.

[2] Montgomery, D. C., Peck, E. A., and Vining, G. G. *Introduction to linear regression analysis*, vol. 821. John Wiley & Sons, 2012.

[3] Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. *Applied linear statistical models*, vol. 4. Irwin Chicago, 1996.

[4] SAS® Software Institute. Sas. https://www.sas.com/.