



UNIVERSIDAD SIMÓN BOLÍVAR
DECANATO DE ESTUDIOS PROFESIONALES
DEPARTAMENTO DE CÓMPUTO CIENTÍFICO y ESTADÍSTICA

Laboratorio 6: Regresión Lineal.

Por:

Andrés Navarro #11-10688

Nabil Marquez #11-10683

REPORTE

Presentado ante la Ilustre Universidad Simón Bolívar

Sartenejas, Agosto del 2016.

Caso de estudio (Ventas)

Se estudia el potencial de los agentes de ventas luego de un entrenamiento en el que se les califica en varias áreas.

Solución:

Primero, se realizarán los modelos para las dos variables respuesta, donde el resto de las variables tendrán $|p| > 0.5$. Para esto se calculó la matriz de correlación para ambos modelos, para identificar cuáles variables son realmente significativas para el modelo a construir. Finalmente, se realizarán modelos para ambas variables con regresión paso a paso, donde en cada paso se eliminará la variable menos significativa. En los modelos finales todas las variables tendrán un nivel de significancia menor a 0.05.

Modelos para el potencial de los agentes de ventas

En la Tabla 1 se encuentra la matriz de correlación obtenida para el modelo del potencial de los agentes de ventas, representado por la variable JOBPER. Como queremos ver cuál es la correlación de las variables del modelo con la variable respuesta, tomaremos en cuenta sólo los valores ubicados en la última fila de la tabla. Se puede observar que las variables que no cumple con la condición de $|p| > 0.5$ son EN y SI, por lo tanto, el resto de las variables estarán incluidas en el modelo a generar.

	JOBPER	AG	EN	AM	HC	SI	IN
JOBPER	10.000.000	0.5425388	0.4244794	0.8786946	0.5874404	0.15554909	-0.57770506
AG	0.5425388	10.000.000	0.3785963	0.4691164	0.5743583	0.29890710	-0.29792886
EN	0.4244794	0.3785963	10.000.000	0.5962339	0.4454779	0.14723314	-0.31522535
AM	0.8786946	0.4691164	0.5962339	10.000.000	0.6662114	0.14033089	-0.47050883
HC	0.5874404	0.5743583	0.4454779	0.6662114	10.000.000	0.37688297	-0.26877889
SI	0.1555491	0.2989071	0.1472331	0.1403309	0.3768830	100.000.000	-0.06199245
IN	-0.5777051	-0.2979289	-0.3152254	-0.4705088	-0.2687789	-0.06199245	100.000.000

Tabla 1

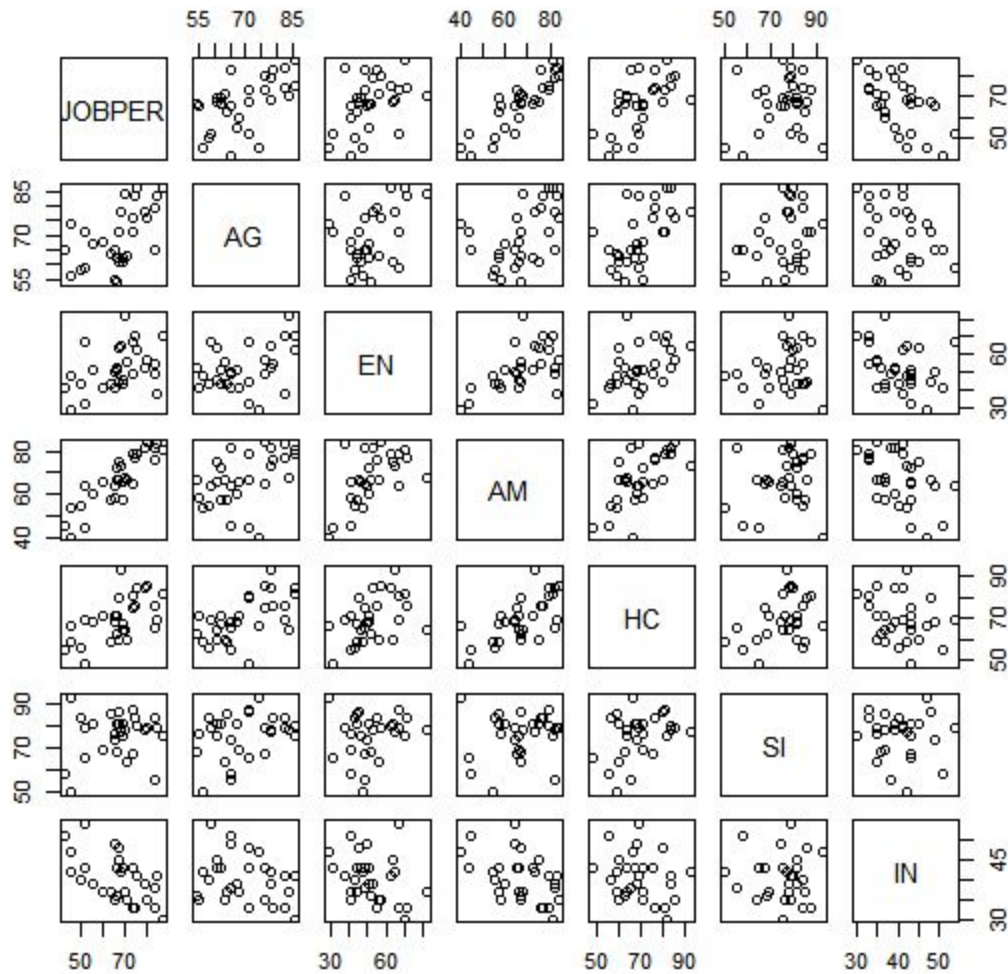


Figura 0

En la Figura 0 podemos observar que los datos están muy desordenados, pocos siguen una distribución normal.

Modelo inicial

Luego, el primer modelo para del potencial de los agentes de ventas: $m1 = \text{lm}(\text{JOBPER} \sim \text{AG} + \text{AM} + \text{HC} + \text{IN})$ con coeficientes 22.51710 0.21056 0.75758 -0.07045 -0.40345

El R cuadrado ajustado para este modelo es 0.8253, que es un valor bueno debido a que está cercano a 1, aunque puede que aún existan modelos mejores a este. Los p-valores obtenidos se encuentran en la Tabla 2. De aquí podemos ver que HC posee el mayor valor, por lo que esta variable parece no ser significativa para nuestro modelo.

	(Intercept)	AG	AM	HC	IN
P-valor	0.1213	0.1183	2.25e-06	0.6283	0.0576

Tabla 2

En la Figura 1 se pueden observar cómo se ajusta el modelo a cada una de las variables. A su vez, podemos observar normalidad de los residuos. Podemos observar que el modelo intenta llegar a cada uno de los datos, sin embargo hay ciertas excepciones, por lo que el modelo puede mejorar. En Q-Q Plot podemos ver que muchos puntos parecen estar cercanos a la recta, pero estos pueden ser más. El modelo puede mejorar.

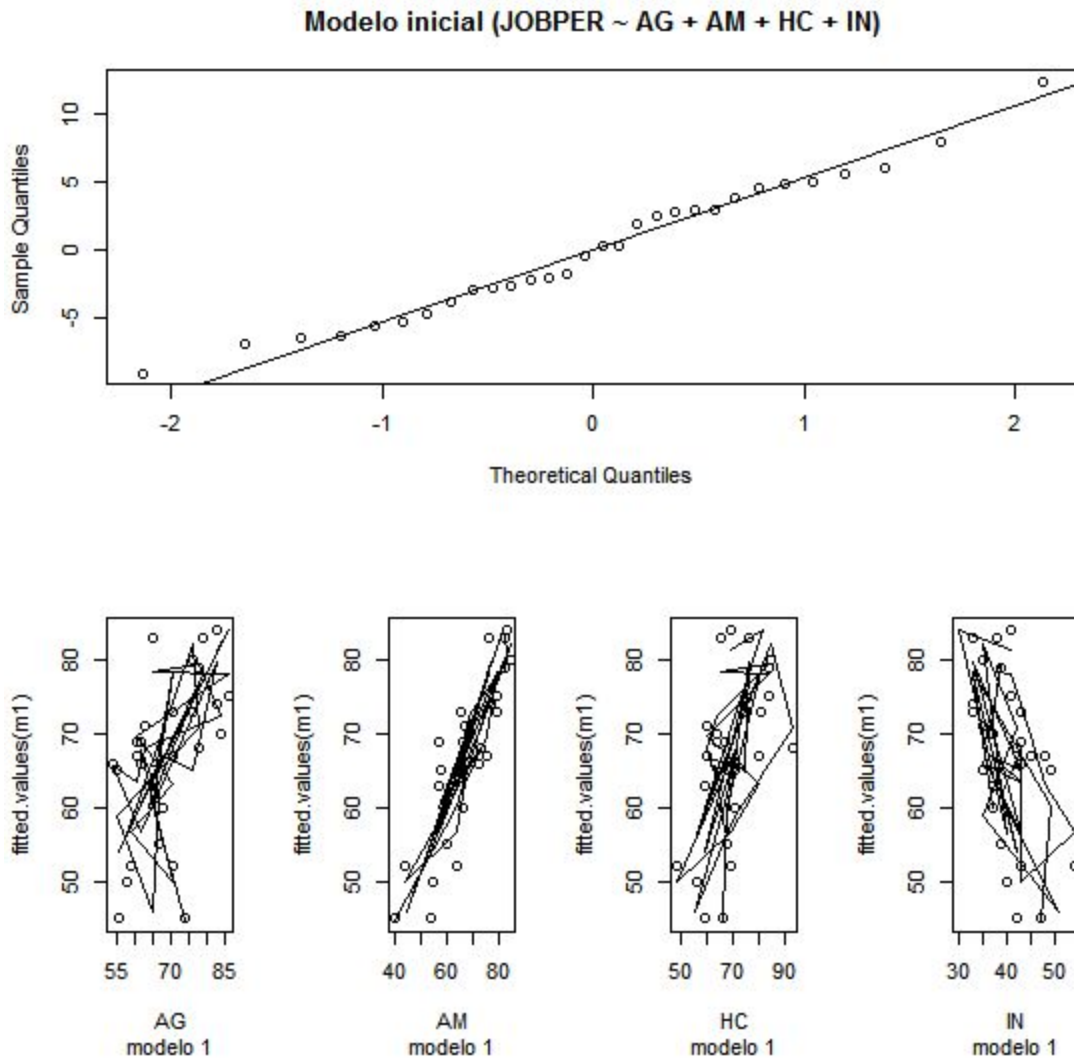


Figura 1

Modelo 2

Luego, el segundo modelo para el potencial de los agentes de ventas será : $m2 = \text{lm}(\text{JOBPER} \sim \text{AG} + \text{AM} + \text{IN})$ con coeficientes 22.1166 0.1846 0.7242 -0.4151

El R cuadrado ajustado para este modelo es 0.8236, que es un valor bueno debido a que está cercano a 1, aunque puede mejorar. Presenta una desmejora con respecto al R cuadrado del modelo anterior. Los p-valores obtenidos se encuentran en la Tabla 3. De aquí podemos ver que AG posee el mayor valor, por lo que esta variable parece no ser significativa para nuestro modelo.

	(Intercept)	AG	AM	IN
P-valor	0.1213	0.1272	1.65e-07	0.0463

Tabla 3

En la Figura 2 se pueden observar, nuevamente, cómo se ajusta el modelo a cada una de las variables. A su vez, podemos observar normalidad de los residuos. Podemos observar que el modelo presenta una mejora en el gráfico Q-Q y en el alcance a los datos para cada variable.

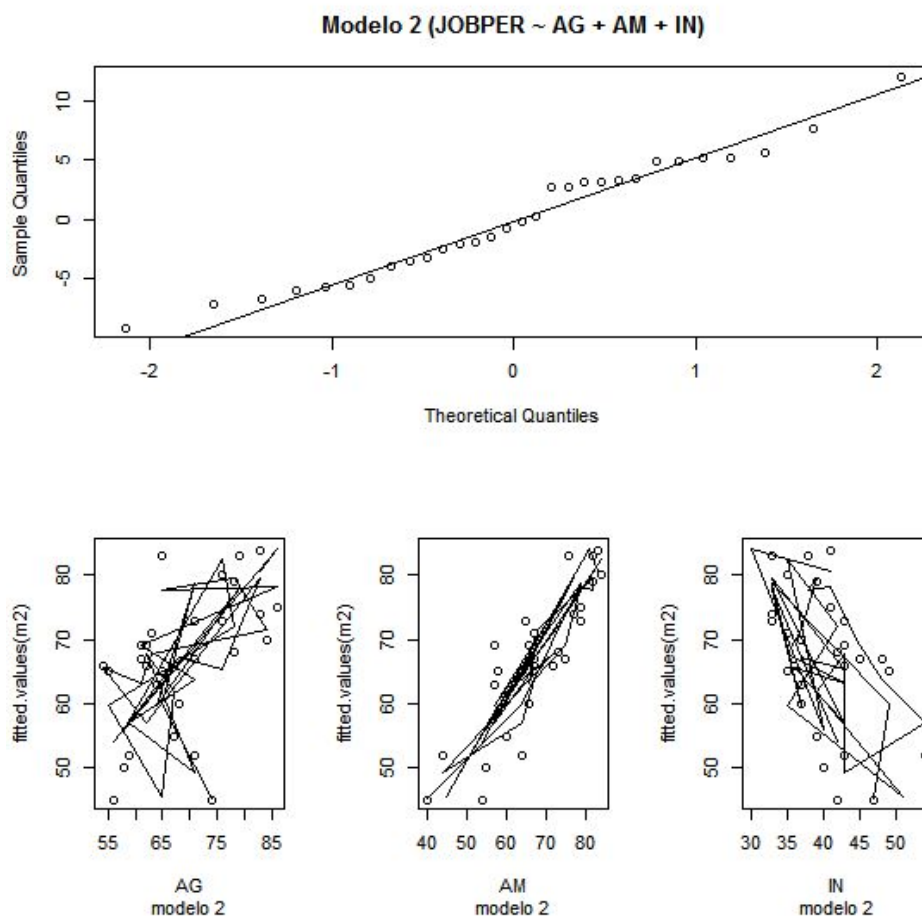


Figura 2

Modelo 3

Luego, el segundo modelo para el potencial de los agentes de ventas será : $m2 = \text{lm}(\text{JOBPER} \sim \text{AG} + \text{AM} + \text{IN})$ con coeficientes 22.1166 0.1846 0.7242 -0.4151

El R cuadrado ajustado para este modelo es 0.8068, que es un valor bueno debido a que está cercano a 1, aunque puede mejorar. Presenta una desmejora con respecto al R cuadrado del modelo anterior. Los p-valores obtenidos se encuentran en la Tabla 4. De aquí podemos ver todas poseen un nivel de significancia menor a 0.05.

	(Intercept)	AM	IN
P-valor	0.0177	9.86e-09	0.0365

Tabla 4

En la Figura 3 se pueden observar, finalmente, cómo se ajusta el modelo a cada una de las variables. A su vez, podemos observar normalidad de los residuos. Podemos observar que el modelo presenta una mejora en el gráfico Q-Q y en el alcance a los datos para cada variable (pequeña, pero existente).

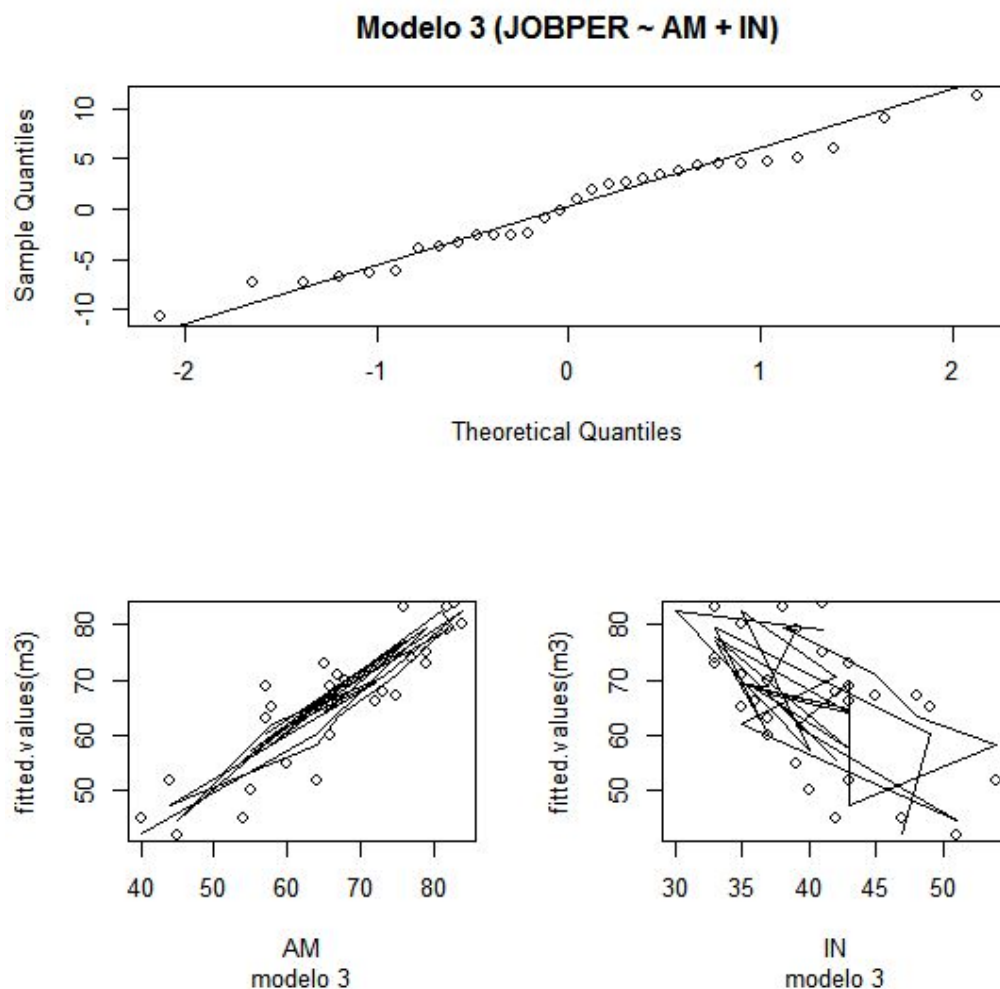


Figura 3

Predicción sobre el modelo original y final

En la Tabla 5 y podremos observar las predicciones para cada ambos modelos. Podemos observar que los intervalos de predicción con 95% de confianza, se encuentran aproximadamente en el mismo rango, por lo que el modelo final parece mantener la calidad.

Modelo	fit	lwr	upr
Original	47.3175	34.29835	60.33665
Final	46.78284	34.04716	59.51853

Tabla 5

Código utilizado:

Código disponible en el archivo “Lab6-Marquez-Navarro.R”

```
1 ##Andres Navarro 11-10688
2 ##Nabil J. Marquez 11-10683
3 ##Estadística Intensivo 2016 - Laboratorio 6
4 datos = read.csv(file="ventas.csv", header=T, sep=";")
5 variables = datos[31,2:7]
6 datos = datos[1:30,1:7]
7 attach(datos)
8 names(datos)
9 plot(datos)
10 # Matriz de Correlación
11 mcor= cor(datos)
12 print(mcor)
13
14 ##### Modelo 1
15 # Modelo donde |ro| > 0.5
16 m1 = lm(JOBPER ~ AG + AM + HC + IN)
17 summary(m1)
18 # Gráficos para ver cómo se ajustó el modelo
19 layout(matrix(c(1,1,1,1,2,3,4,5), 2, 4, byrow = TRUE))
20 qqnorm(resid(m1), main = "Modelo inicial (JOBPER ~ AG + AM + HC + IN)")
21 qqline(resid(m1))
22 plot(AG, fitted.values(m1), sub="modelo 1", type="l")
23 points(AG, JOBPER)
24 plot(AM, fitted.values(m1), sub="modelo 1", type="l")
25 points(AM, JOBPER)
26 plot(HC, fitted.values(m1), sub="modelo 1", type="l")
27 points(HC, JOBPER)
28 plot(IN, fitted.values(m1), sub="modelo 1", type="l")
29 points(IN, JOBPER)
30
31 ##### Modelo 2
32 # Modelo 2 eliminando HC
33 m2 = lm(JOBPER ~ AG + AM + IN)
34 summary(m2)
35 # Gráficos para ver cómo se ajustó el modelo
36 layout(matrix(c(1,1,1,2,3,4), 2, 3, byrow = TRUE))
37 qqnorm(resid(m2), main = "Modelo 2 (JOBPER ~ AG + AM + IN)")
38 qqline(resid(m2))
39 plot(AG, fitted.values(m2), sub="modelo 2", type="l")
40 points(AG, JOBPER)
41 plot(AM, fitted.values(m2), sub="modelo 2", type="l")
42 points(AM, JOBPER)
43 plot(IN, fitted.values(m2), sub="modelo 2", type="l")
44 points(IN, JOBPER)
45
46 ##### Modelo 3
47 # Modelo 3 eliminando AG
48 m3 = lm(JOBPER ~ AM + IN)
49 summary(m3)
50 # Gráficos para ver cómo se ajustó el modelo
51 layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
52 qqnorm(resid(m3), main = "Modelo 3 (JOBPER ~ AM + IN)")
53 qqline(resid(m3))
54 plot(AM, fitted.values(m3), sub="modelo 3", type="l")
55 points(AM, JOBPER)
56 plot(IN, fitted.values(m3), sub="modelo 3", type="l")
57 points(IN, JOBPER)
58
59 ##### Predicción sobre los modelos
60 predi = data.frame(AG=variables["AG"], EN=variables["EN"], AM=variables["AM"], HC=variables["HC"], SI=variables["SI"], IN=variables["IN"])
61 p1 <- predict(m1, newdata = predi, interval = "prediction")
62 predf = data.frame(AM=variables["AM"], IN=variables["IN"])
63 p2 <- predict(m3, newdata = predf, interval = "prediction")
```