

Informe Final

Ana Julia Escobar

3 de agosto de 2015

Introducción

Análisis de nuestro modelo de regresión lineal múltiple.

Descripción información

Cargamos nuestras datas

```
options(warn=-1)
library(readxl)
pob1<-read_excel("poblacion1.xlsx", sheet=1, na=" ")
pob2<-read_excel("poblacion2.xlsx", sheet=1, na=" ")
View(pob1)
str(pob1)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   44 obs. of  4 variables:
## $ identificador : num  1001 1002 1003 1004 1005 ...
## $ poblacion      : num  18.7 13.8 13.4 11.4 10.5 10.3 10.3 10.3 10.3 9.4 ...
## $ var.pobl.mayor: num   6.2 7.2 7.1 2.6 2.8 1.4 -0.5 3 -1.9 -1.5 ...
## $ menores.18    : num  39.7 10.4 20.4 38.7 36.4 29.8 31.8 31.1 15.9 22.4 ...
```

```
View(pob2)
str(pob2)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   40 obs. of  7 variables:
## $ identificador : num  1044 1042 1010 1009 1008 ...
## $ part.almz.escl : num  73.8 34.4 22.2 39.9 51.7 57.2 43.7 72.3 78.2 13.9 ...
## $ var.ingresos   : num  50.5 24.2 33.5 38.5 26.2 27.2 29.4 26 18.4 22.5 ...
## $ tasa.crimen    : num  704.1 179.9 61.5 86.4 42.4 ...
## $ var.tasa.crimen: num  -40.9 12.3 -32.7 -13.5 -31.9 -17.6 -10 -21.6 -12.8 -33.6 ...
## $ region         : chr   "B" "B" "A" "A" ...
## $ serv.bas.compl : chr   "SI" "NO" "NO" "SI" ...
```

Disponemos de 44 observaciones de 4 variables. Disponemos de 40 observaciones de 7 variables.

Tendremos nuestras datas en una: población.

```
options(warn=-1)
poblacion <- merge(pob1,pob2,by="identificador",suffixes=c(", "))
str(poblacion)
```

```
## 'data.frame':   40 obs. of  10 variables:
## $ identificador : num  1001 1002 1003 1004 1005 ...
## $ poblacion      : num  18.7 13.8 13.4 11.4 10.5 10.3 10.3 10.3 10.3 9.4 ...
```

```
## $ var.pobl.mayor : num 6.2 7.2 7.1 2.6 2.8 1.4 -0.5 3 -1.9 -1.5 ...
## $ menores.18 : num 39.7 10.4 20.4 38.7 36.4 29.8 31.8 31.1 15.9 22.4 ...
## $ part.almz.escl : num 55.8 57.9 13.9 78.2 72.3 43.7 57.2 51.7 39.9 22.2 ...
## $ var.ingresos : num 28.7 26.2 22.5 18.4 26 29.4 27.2 26.2 38.5 33.5 ...
## $ tasa.crimen : num 52.6 111 38.3 86.6 77.5 54 61.2 42.4 86.4 61.5 ...
## $ var.tasa.crimen: num -2.9 -22.6 -33.6 -12.8 -21.6 -10 -17.6 -31.9 -13.5 -32.7 ...
## $ region : chr "A" "A" "A" "A" ...
## $ serv.bas.compl : chr "SI" "SI" "NO" "NO" ...
```

Modelo de regresión lineal simple

Explicaremos la variable poblacion en función de los regresores que tengan mayor correlación con nuestra variable dependiente.

```
options(warn=-1)
core <- 3:dim(poblacion)[2]
core[1] <- NA

for(i in 2:dim(poblacion)[2]){
  if(is.numeric(poblacion[,i])==TRUE){
    core[i] <- cor(poblacion[,2], poblacion[,i])
  } else {
    core[i] <- NA
  }
}
core
```

```
## [1] NA 1.00000000 -0.15774031 0.19709059 -0.04686172
## [6] -0.07677708 -0.37492254 -0.07645301 NA NA
```

Ajustamos un modelo de regresión lineal múltiple para las variables poblacion con var.pobl.mayor, menores.18, tasa.crimen

La regresion regresión.

```
reg <- lm(poblacion~var.pobl.mayor+menores.18+tasa.crimen, data=poblacion)
summary(reg)
```

```
##
## Call:
## lm(formula = poblacion ~ var.pobl.mayor + menores.18 + tasa.crimen,
##     data = poblacion)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.0239 -2.6986 -0.3656  2.0947 10.2333
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.388436   2.048573   3.607 0.000933 ***
## var.pobl.mayor  0.029694   0.082263   0.361 0.720238
## menores.18     0.039498   0.064026   0.617 0.541179
```

```
## tasa.crimen    -0.012813    0.006367   -2.012 0.051730 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.468 on 36 degrees of freedom
## Multiple R-squared:  0.1533, Adjusted R-squared:  0.08277
## F-statistic: 2.173 on 3 and 36 DF,  p-value: 0.1081
```

Es decir nuestro modelo resultante estaría dado por la expresión

$$\hat{p}o = 7.3884361 + 0.0296939\hat{v}a + 0.0394981\hat{m}e - 0.0128125\hat{t}a$$

Interpretación

Interpretación modelo: Si var.pobl.mayor se incrementa en una unidad, entonces poblacion aumenta en promedio 0.0296939 unidades.

Si menores.18 se incrementa en una unidad, entonces poblacion aumenta en promedio 0.0394981 unidades.

Si tasa.crimen se incrementa en una unidad, entonces poblacion disminuye en promedio 0.0128125 unidades.

Interpretación R^2 :

El modelo de regresión lineal obtenido explica el 15.3323897111369 % de la variabilidad total.

Pruebas de hipótesis

Para β_4 Como $t = -2.0121753$ es menor que $t_{38}(\frac{\alpha}{2}) = 2.0243942$ no rechazo $H_0 : \beta_4 = 0$.

Para β_3 Como $t = 0.6169066$ es menor que $t_{38}(\frac{\alpha}{2}) = 2.0243942$ no rechazo $H_0 : \beta_3 = 0$.

Para β_2 Como $t = 0.3609615$ es menor que $t_{38}(\frac{\alpha}{2}) = 2.0243942$ no rechazo $H_0 : \beta_2 = 0$.

Para β_1 Como $t = 3.6066254$ es mayor que $t_{38}(\frac{\alpha}{2}) = 2.0243942$ rechazo $H_0 : \beta_1 = 0$.

Análisis de varianza

Realizando la tabla **ANOVA** tenemos los siguientes resultados:

```
anova <- aov(reg)
summary(anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## var.pobl.mayor  1   12.7    12.73   1.058 0.3105
## menores.18     1   17.0    16.99   1.412 0.2424
## tasa.crimen    1   48.7    48.70   4.049 0.0517 .
## Residuals     36  433.0    12.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

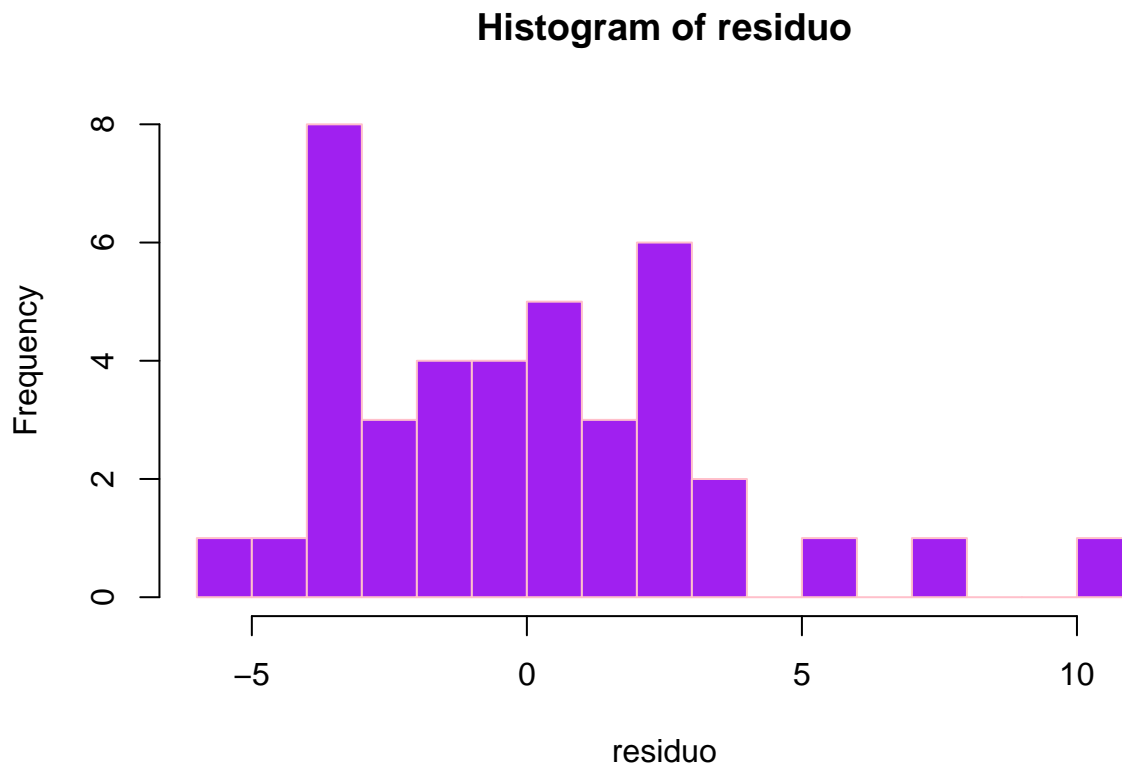
Intervalos de confianza

```
confint(reg,level = 0.95)
```

```
##              2.5 %      97.5 %  
## (Intercept)  3.23373707 1.154314e+01  
## var.pobl.mayor -0.13714393 1.965317e-01  
## menores.18   -0.09035265 1.693488e-01  
## tasa.crimen  -0.02572639 1.013625e-04
```

Gráficos residuales

```
library(DT)  
residuo <- reg[["residuals"]]  
prediccion <- reg[["fitted.values"]]  
hist(residuo,15,col = "purple",border = "pink")
```



```
mean(residuo)
```

```
## [1] -1.394284e-16
```

```
qqnorm(residuo)  
qqline(residuo,col="pink")
```

Normal Q-Q Plot

