

Lab 07 - ANÁLISIS DE REGRESIÓN MÚLTIPLE. ECUACIÓN DE REGRESIÓN MÚLTIPLE

Daniel Escriba Flores

Contents

Contexto:	1
Pregunta A: Datos, D-Dispersion e interpretacion	1
Lectura de datos	2
Medidas descriptivas	2
Diagrama de dispersion	3
Interpretacion de Correlaciones con la variable dependiente	3
Pregunta B: Regresion multiple y Ecuacion	4
Regresion multiple	4
Interpretación del Modelo	4
Ecuacion de regresion multiple	4
Pregunta C: Pronostico	5
Conclusion	5

Contexto:

El Times-Observer es un periódico en la ciudad Metro. Al igual que muchos periódicos en la ciudad, el Times-Observer pasa por dificultades financieras. La gerente de circulación estudia otros periódicos en ciudades similares en Estados Unidos y Canadá, con interés particular en las variables que se relacionan con el número de suscriptores. Esta reúne la informacion muestral de 25 periódicos en ciudades similares empleando la siguiente notación:

- Suscriptores : Número de suscriptores (en miles)
- Población : Población metropolitana (en miles)
- Presupuesto : Presupuesto en publicidad del periódico (miles de dólares)
- Ingreso : Ingreso familiar medio en el área metropolitana (miles de dólares)

Deseando predecir el número de suscriptores con base a las otras variables, haga lo siguiente:

Pregunta A: Datos, D-Dispersion e interpretacion

- Lea la base de datos, elabore un diagrama de dispersión para cada variable independiente con la dependiente indicando e interpretando sus respectivos valores de coeficiente de correlación.

Lectura de datos

```
library(readxl)
library(pander)
library(equatiomatic)

data = read_excel("periódicos.xlsx", sheet="Hoja1")
data = data[-1]
data = as.data.frame(data)

# Usar pander para formatear la tabla
pander(data, caption = "Tabla de datos de suscripciones a periodicos de USA y Canada")
```

Table 1: Tabla de datos de suscripciones a periodicos de USA y Canada

Suscriptores	Población	Presupuesto	Ingreso
37.95	588.9	13.2	35.1
37.66	585.3	13.2	34.7
37.55	566.3	19.8	34.8
38.78	642.9	17.6	35.1
37.67	624.2	17.6	34.6
38.23	603.9	15.4	34.8
36.9	571.9	11	34.7
38.28	584.3	28.6	35.3
38.95	605	28.6	35.1
39.27	676.3	17.6	35.6
38.3	587.4	17.6	34.9
38.84	576.4	22	35.4
38.14	570.8	17.6	35
38.39	586.5	15.4	35.5
37.29	544	11	34.9
39.15	611.1	24.2	35
38.29	643.3	17.6	35.3
38.09	635.6	19.8	34.8
37.83	598.9	15.4	35.1
39.37	657	22	35.3
37.81	595.2	15.4	35.1
37.42	520	19.8	35.1
38.83	629.6	22	35.3
38.33	680	24.2	34.7
40.24	651.2	33	35.8

Medidas descriptivas

```
summary(data)
```

##	Suscriptores	Población	Presupuesto	Ingreso
##	Min. :36.90	Min. :520.0	Min. :11.00	Min. :34.60
##	1st Qu.:37.81	1st Qu.:584.3	1st Qu.:15.40	1st Qu.:34.80
##	Median :38.28	Median :598.9	Median :17.60	Median :35.10
##	Mean :38.30	Mean :605.4	Mean :19.18	Mean :35.08
##	3rd Qu.:38.83	3rd Qu.:635.6	3rd Qu.:22.00	3rd Qu.:35.30

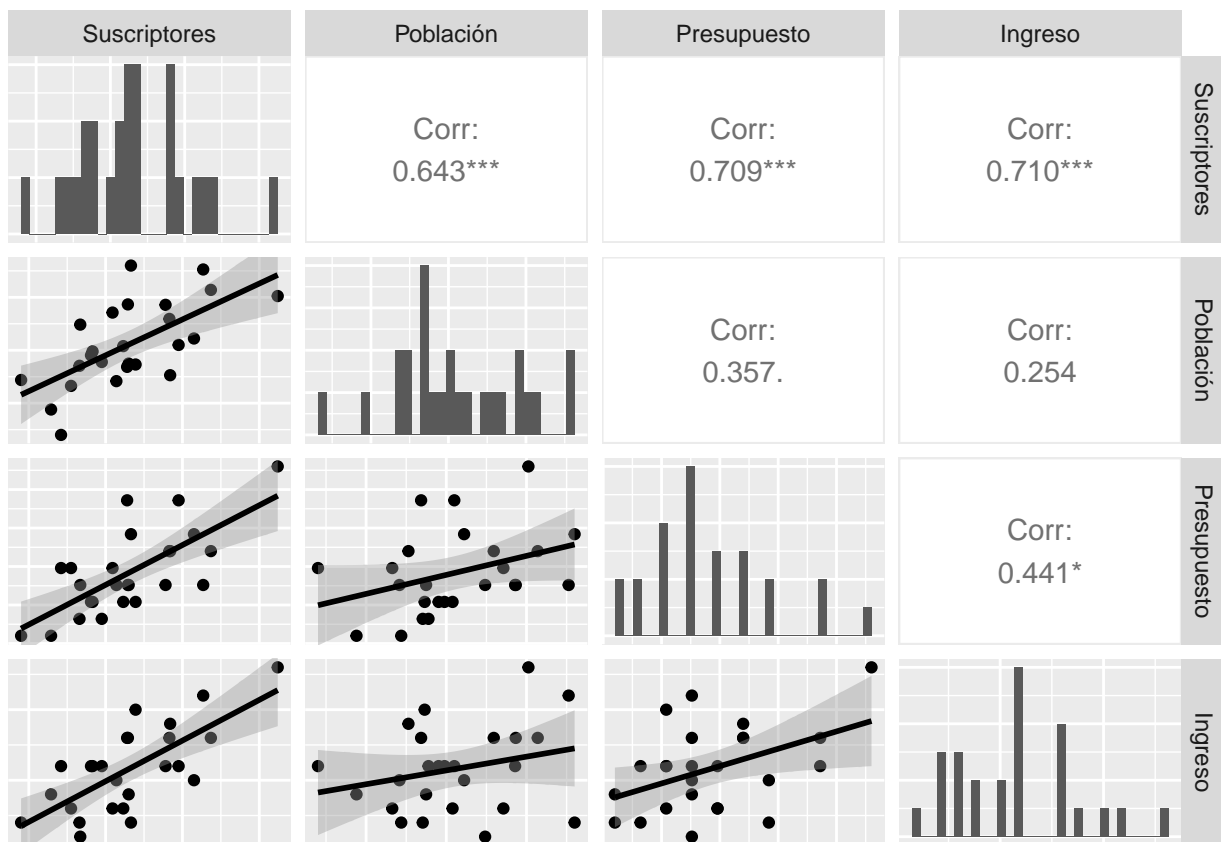
```
## Max. :40.24 Max. :680.0 Max. :33.00 Max. :35.80
```

Diagrama de dispersion

```
#install.packages ("GGally")

library(ggplot2)
library(GGally)

ggpairs(data,
  upper = list(continuous = "cor", na = "na"),
  columns=1:4,
  lower=list(continuous="smooth"),
  diag=list(continuous="barDiag"),
  axisLabels="none")
```



Interpretacion de Correlaciones con la variable dependiente

- Población (0.643): Correlación positiva fuerte y significativa. A mayor población, mayor número de suscriptores.
- Presupuesto (0.709): Correlación positiva muy fuerte y significativa. Los presupuestos más altos están asociados con más suscriptores.
- Ingreso (0.710): La correlación más fuerte, positiva y significativa. Los mayores ingresos predicen claramente más suscriptores.

Pregunta B: Regresion multiple y Ecuacion

b. Realice un análisis de regresión múltiple y escriba la ecuación de regresión múltiple.

Regresion multiple

```
modelo = lm(formula = Suscriptores ~ Población +
            Presupuesto +
            Ingreso,
            data= data)

summary(modelo)

##
## Call:
## lm(formula = Suscriptores ~ Población + Presupuesto + Ingreso,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58362 -0.24463 -0.03801  0.25393  0.63750
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.732813   8.427349  -0.680  0.503769
## Población    0.007537   0.001813   4.157  0.000446 ***
## Presupuesto  0.050883   0.014139   3.599  0.001689 **
## Ingreso      1.097381   0.245029   4.479  0.000207 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3269 on 21 degrees of freedom
## Multiple R-squared:  0.8348, Adjusted R-squared:  0.8112
## F-statistic: 35.38 on 3 and 21 DF,  p-value: 2.144e-08

modelo$coefficients

## (Intercept)    Población  Presupuesto      Ingreso
## -5.732813186  0.007536582  0.050883394  1.097380777
```

Interpretación del Modelo

Bondad de ajuste: $R^2 = 0.8348$ (83.48% de variabilidad explicada - excelente ajuste)

Variables significativas (todas con $p < 0.01$):

Población: cada aumento de 1 unidad aumenta 0.0075 suscriptores

Presupuesto: cada unidad adicional aporta 0.051 suscriptores

Ingreso: el predictor más fuerte (1.097 suscriptores por unidad)

Ecuacion de regresion multiple

Ecuación de regresión:

$\text{Suscriptores} = -5.732813 + 0.007536582 * \text{Población} + 0.05088339 * \text{Presupuesto} + 1.097381 * \text{Ingreso}$

```
# Versión básica
extract_eq(modelo)
```

$$\text{Suscriptores} = \alpha + \beta_1(\text{Población}) + \beta_2(\text{Presupuesto}) + \beta_3(\text{Ingreso}) + \epsilon \quad (1)$$

```
# Versión con valores de coeficientes
extract_eq(modelo, use_coefs = TRUE)
```

$$\hat{\text{Suscriptores}} = -5.73 + 0.01(\text{Población}) + 0.05(\text{Presupuesto}) + 1.1(\text{Ingreso}) \quad (2)$$

Pregunta C: Pronostico

- c. Realice un pronóstico cuando la población es de 600 000 habitantes, el presupuesto en publicidad del periódico es de 18 000 dólares y el ingreso familiar medio en el área metropolitana es de 36 000 dólares.

```
predictores = data.frame(Población=600, Presupuesto=18, Ingreso=36)
pander(predictores, caption = "Tabla de datos predictores")
```

Table 2: Tabla de datos predictores

Población	Presupuesto	Ingreso
600	18	36

```
costo_est = predict(object = modelo, newdata = predictores)
```

```
## Suscriptores Estimados con Población=600 000, Presupuesto=18 000, Ingreso=10 000 :
## 39.21075
```

Conclusion

El modelo muestra que el ingreso y el presupuesto son los factores más importantes para predecir suscriptores, con un 83.5% de explicación (R^2).

Para una ciudad con:

- 600,000 habitantes
- \$18,000 de presupuesto publicitario
- \$36,000 de ingreso familiar promedio

Se pronostican aproximadamente 39,210 suscriptores.

Estos resultados demuestran que las condiciones económicas y las inversiones en marketing son determinantes clave para el crecimiento de la base de suscriptores del periódico.