

# Regresión Lineal Múltiple (RLM)

2022-10-12

## Contents

<b>1. Operaciones con matrices</b>	<b>2</b>
1.1 Definición de matriz . . . . .	2
1.2 Transpuesta de una matriz . . . . .	2
1.3 Multiplicación de matrices . . . . .	3
1.4 Matriz identidad . . . . .	3
1.5 Inversa de matriz . . . . .	3
<b>2. Regresión Lineal Múltiple (RLM)</b>	<b>4</b>
2.1 Definición de RLM . . . . .	4
2.2 Forma matricial . . . . .	4
<b>3. Mínimos Cuadrados Ordinarios para RLM</b>	<b>5</b>
3.1 Estimador de varianza poblacional . . . . .	5
3.3 Intervalos de confianza para los coeficientes de regresión . . . . .	5
3.4 Anova para Regresión Lineal Múltiple . . . . .	5
3.5 Prueba de hipótesis sobre los coeficientes . . . . .	5
3.6 Prueba F de Fisher . . . . .	6
3.7 $R^2$ y $R^2$ ajustada . . . . .	6
3.8 Intervalo de confianza para valor de Y . . . . .	6
3.9 Intervalo de confianza para la predicción media . . . . .	6
<b>Ejercicio:</b>	<b>6</b>

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
library(tidyverse)
library(tidymodels)
write_matex2 <- function(x) {
  begin <- "\\begin{bmatrix}"
  end <- "\\end{bmatrix}"
  X <-
    apply(x, 1, function(x) {
      paste(
        paste(x, collapse = "&"),
        "\\\\"
      )
    })
  paste(c(begin, X, end), collapse = "")
}
```

## 1. Operaciones con matrices

### 1.1 Definición de matriz

Definamos una matriz como un arreglo bidimensional que contiene tanto renglones como columnas:

```
# Definamos en R la matriz:
A <- matrix(c(1,3,0,1,4,5,1,2,3),nrow = 3, ncol = 3)
```

Como se definen por renglones y columnas, decimos que es una  $Matrix_{n \times m}$  donde  $n$  es el número de renglones y  $m$  el número de columnas:

$$\begin{bmatrix} 1 & 1 & 1 \\ 3 & 4 & 2 \\ 0 & 5 & 3 \end{bmatrix}$$

### 1.2 Transpuesta de una matriz

Al transponer una matriz, lo que buscamos es que ahora los renglones se vuelvan columnas y las columnas renglones, por ejemplo:

```
A
```

```
##      [,1] [,2] [,3]
## [1,]    1    1    1
## [2,]    3    4    2
## [3,]    0    5    3
```

```
t(A)
```

```
##      [,1] [,2] [,3]
## [1,]    1    3    0
## [2,]    1    4    5
## [3,]    1    2    3
```

$$\begin{bmatrix} 1 & 3 & 0 \\ 1 & 4 & 5 \\ 1 & 2 & 3 \end{bmatrix}$$

### 1.3 Multiplicación de matrices

Así como con números reales conocemos las multiplicaciones, para matrices también se puede multiplicar (en R con el comando `%*%`):

```
B <- matrix(c(3,2,1,1,5,5,1,7,3),nrow = 3, ncol = 3)
C = A %*% B
C
```

```
##      [,1] [,2] [,3]
## [1,]    6   11   11
## [2,]   19   33   37
## [3,]   13   40   44
```

$$\begin{bmatrix} 1 & 1 & 1 \\ 3 & 4 & 2 \\ 0 & 5 & 3 \end{bmatrix} \times \begin{bmatrix} 3 & 1 & 1 \\ 2 & 5 & 7 \\ 1 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 11 & 11 \\ 19 & 33 & 37 \\ 13 & 40 & 44 \end{bmatrix}$$

### 1.4 Matriz identidad

La matriz identidad está dada por la matriz con 1 en la diagonal

```
D <- matrix(c(1,0,0,0,1,0,0,0,1),nrow = 3, ncol = 3)
D
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    0    1    0
## [3,]    0    0    1
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

### 1.5 Inversa de matriz

La inversa de una matriz es aquella matriz B' tal que multiplicada por B del lado derecho te devuelve la matriz identidad

```
invB <- solve(B)
round(B %*% invB, 1)
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    0    1    0
## [3,]    0    0    1
```

$$\begin{bmatrix} 3 & 1 & 1 \\ 2 & 5 & 7 \\ 1 & 5 & 3 \end{bmatrix} \times \begin{bmatrix} 0.4 & 0 & 0 \\ 0 & -0.1 & 0.4 \\ -0.1 & 0.3 & -0.2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

## 2. Regresión Lineal Múltiple (RLM)

### 2.1 Definición de RLM

El modelo de regresión múltiple es una generalización del modelo de regresión simple, ahora en lugar de tener solo a  $\beta_0$  y  $\beta_1$ , contamos con más variables independientes:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$$

Generalmente denotamos con  $n$  el número de observaciones que tenemos en los datos y con  $k$  el número de regresores que tenemos. De la misma forma que el modelo simple,  $\beta_0$  representa el intercepto y  $\beta_1, \beta_2, \dots, \beta_k$  representan los coeficientes de regresión asociados a las variables independientes  $X_i$

La interpretación de cada  $\beta_i$  ahora es como cambia en promedio  $Y$  por unidad de cambio en la variable independiente, manteniendo todas las demás constantes.

### 2.2 Forma matricial

Podemos expresar el modelo de regresión múltiple de forma matricial, por ejemplo, en el caso donde  $n = 5$  y  $k = 3$ , podemos traducirlo de la siguiente forma:

```
y <- matrix(c("Y1", "Y2", "Y3", "Y4", "Y5"), ncol = 1, nrow = 5)
X <- matrix(c(1,"X11","X12", "X13",
              1,"X21","X22", "X23",
              1,"X31","X32", "X33",
              1,"X41","X42", "X43",
              1,"X51","X52", "X53"), ncol = 4, nrow = 5, byrow = T)
betas <- matrix(c("Beta1", "Beta2", "Beta3", "Beta4", "Beta5"), ncol = 1, nrow = 5)
error <- matrix(c("error1", "error2", "error3", "error4", "error5"), ncol = 1, nrow = 5)
```

$$\begin{bmatrix} Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \end{bmatrix} = \begin{bmatrix} 1 & X11 & X12 & X13 \\ 1 & X21 & X22 & X23 \\ 1 & X31 & X32 & X33 \\ 1 & X41 & X42 & X43 \\ 1 & X51 & X52 & X53 \end{bmatrix} \times \begin{bmatrix} Beta1 \\ Beta2 \\ Beta3 \\ Beta4 \\ Beta5 \end{bmatrix} + \begin{bmatrix} error1 \\ error2 \\ error3 \\ error4 \\ error5 \end{bmatrix}$$

Cuando lo vemos expresado en forma matricial podemos verlo como  $Y = X\beta + \hat{e}$

Fuente de variación	Suma de Cuadrados	Grados de libertad	Suma de cuadrados medios
TSS	$\sum (y_i - \bar{y})^2$	$n - 1$	$\frac{\sum (y_i - \bar{y})^2}{n-1}$
RSS	$\sum (y_i - \hat{y})^2$	$n - k - 1$	$\frac{\sum (y_i - \hat{y})^2}{n-k-1}$
ESS	$\sum (\hat{y}_i - \bar{y})^2$	$k$	$\frac{\sum (\hat{y}_i - \bar{y})^2}{k}$

### 3. Mínimos Cuadrados Ordinarios para RLM

#### 3.1 Estimador de varianza poblacional

$$\hat{\sigma}^2 = \frac{\sum e^2}{n - k - 1} = \frac{e^T e}{n - k - 1}$$

## 3.2 Estimador de coeficientes de regresión

Estimadores puntuales:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

De esta manera, podemos decir que los estimadores siguen una distribución:

$$\hat{\beta} \sim \text{Normal}(\beta, \sigma^2 (X^T X)^{-1})$$

Se tiene como resultado que los estimadores también están distribuidos de manera normal y son un estimador insesgado. En particular a la matrix  $(X^T X)\sigma^2$  se le conoce como la matriz de varianzas-covarianzas de  $\hat{\beta}$

Al igual que en caso simple, estimamos la varianza de las  $\hat{\beta}$ :

$$\sigma_{\hat{\beta}_i}^2 = \hat{\sigma}^2 (X^T X)^{-1}$$

#### 3.3 Intervalos de confianza para los coeficientes de regresión

Los IC de los coeficientes de regresión siguen una fórmula muy similar al caso simple. Al estar estimando la varianza poblacional, entonces no podemos utilizar la distribución normal, por lo que empleamos la t-student:

$$\text{IC para } \beta_i : (\hat{\beta}_i \pm t_{\alpha/2, n-k-1} \sqrt{\hat{\sigma}_{\hat{\beta}_i}^2})$$

#### 3.4 Anova para Regresión Lineal Múltiple

#### 3.5 Prueba de hipótesis sobre los coeficientes

Sea el estadístico:

$$t_{\beta_i} = \frac{\hat{\beta}_i - \beta_0}{\sqrt{\sigma_{\hat{\beta}_i}^2}}$$

Tendremos una prueba de hipótesis para cada variable independiente que contenga nuestro modelo + una por el intercept. En este estadístico es común que el valor bajo  $H_0 : \beta_i = 0$ , por que quisieramos probar si podemos rechazar que  $\beta_i = 0$ , es decir afirmar que toma otro valor. Esta prueba la realizaremos con la t-student de dos colas al igual que en caso simple.

De esta manera, utilizando el valor-p alcanzado, se concluiría.

### 3.6 Prueba F de Fisher

Para la prueba de Fisher también se ajustará el estadístico:

$$F = \frac{\frac{ESS}{k}}{\frac{RSS}{n-k-1}}$$

El cual se distribuye  $F_{k,n-k-1}$

### 3.7 $R^2$ y $R^2$ ajustada

La  $R^2$  la podemos ver de la misma forma:

$$R^2 = \frac{ESS}{TSS}$$

y la  $R^2_{adj}$ :

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

$R^2_{adj}$  penaliza a  $R^2$  de acuerdo al número de regresores

### 3.8 Intervalo de confianza para valor de Y

### 3.9 Intervalo de confianza para la predicción media

**Ejercicio:**