

Fórmulas de Covarianza y Correlaciones

Tu Nombre

2025-03-14

En este documento se presentan las expresiones matemáticas de:

- Covarianza
- Matriz de covarianzas
- Coeficiente de correlación de Pearson
- Coeficiente de correlación de Spearman

Covarianza

La **covarianza** de dos variables aleatorias X e Y se define como:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

También puede expresarse de forma equivalente como:

$$\text{Cov}(X, Y) = E[XY] - E[X] E[Y].$$

De manera intuitiva, la covarianza mide cuánto varían conjuntamente dos variables. Si $\text{Cov}(X, Y)$ es positiva, significa que, en promedio, cuando X aumenta, Y también tiende a aumentar, y viceversa.

Matriz de Covarianzas

Para un vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$, la **matriz de covarianzas** Σ se define como:

$$\Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n) \end{pmatrix},$$

donde $\Sigma_{ij} = \text{Cov}(X_i, X_j)$. Esta matriz describe las relaciones de variación conjunta entre todas las variables consideradas.

Coeficiente de Correlación de Pearson

El **coeficiente de correlación de Pearson** entre dos variables aleatorias X e Y se define como:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Aquí, $\rho_{X,Y}$ toma valores en el intervalo $[-1, 1]$.

- $\rho_{X,Y} = 1$: correlación lineal positiva perfecta.
- $\rho_{X,Y} = -1$: correlación lineal negativa perfecta.
- $\rho_{X,Y} = 0$: no existe correlación lineal.

Coeficiente de Correlación de Spearman

El **coeficiente de correlación de Spearman** se basa en los rangos de los datos en lugar de sus valores originales. Se calcula aplicando el coeficiente de Pearson a los rangos (ordenamientos) de los datos. Para dos variables X e Y con rangos $R(X)$ y $R(Y)$:

$$\rho_s = \frac{\text{Cov}(R(X), R(Y))}{\sqrt{\text{Var}(R(X)) \text{Var}(R(Y))}}.$$

Este método es menos sensible a la presencia de valores atípicos (*outliers*) y puede capturar relaciones monótonas (pero no necesariamente lineales).

Ejemplo de Cálculo en R

Para ilustrar estas medidas, considera el siguiente código en R:

```
# 1) Instalar (si es necesario) y cargar las librerías
# install.packages("ggplot2")
# install.packages("plotly")
library(ggplot2)
library(plotly)

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout

set.seed(123)
x <- rnorm(100)
y <- 2*x + rnorm(100, sd = 0.5) # Relación aproximadamente lineal con algo de ruido

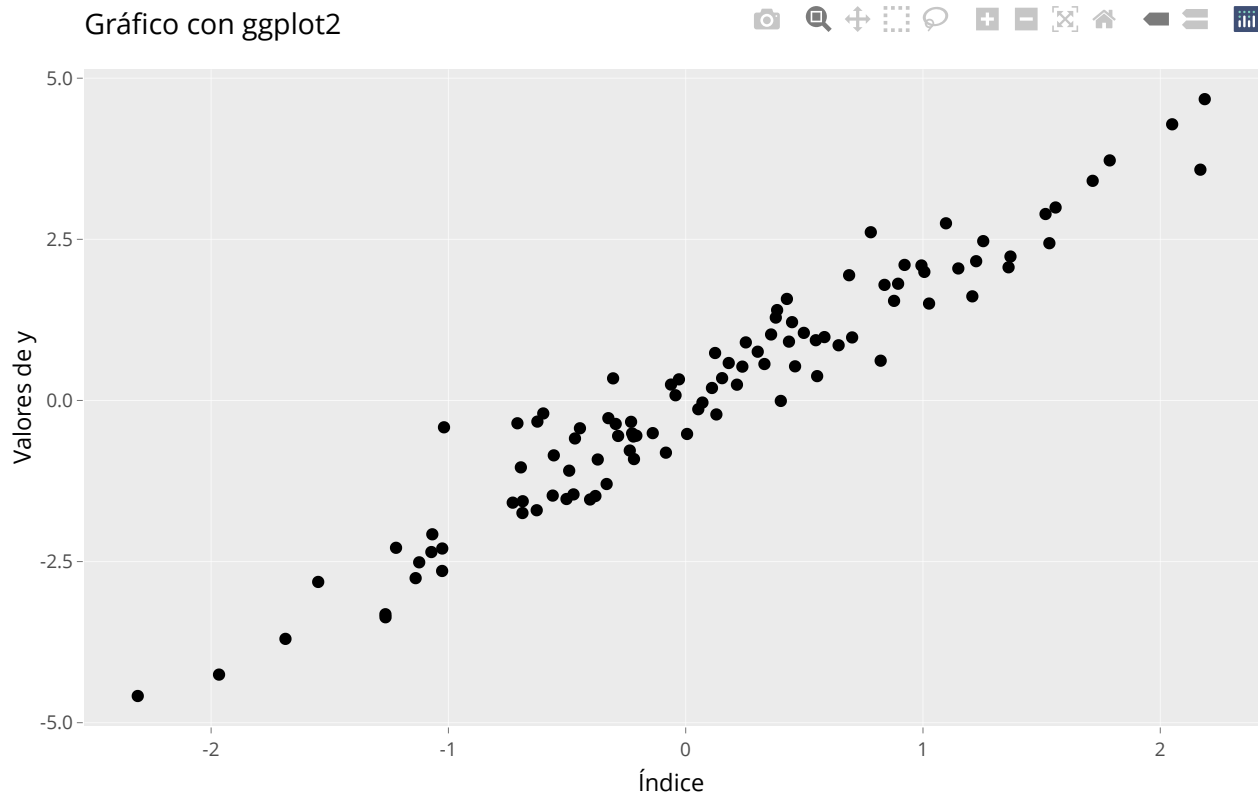
# Gráfico de y
# 4) Crear un data.frame para ggplot
df <- data.frame(Value_x=x, Value_y = y)

# 5) Gráfico con ggplot2
p <- ggplot(df, aes(x = Value_x, y = Value_y)) +
  geom_point() +
  ggtitle("Gráfico con ggplot2") +
  xlab("Índice") +
  ylab("Valores de y")

# 6) Hacer el gráfico interactivo con plotly
p_interactivo <- ggplotly(p)

# 7) Mostrar el gráfico interactivo en el visor de R
p_interactivo
```

Gráfico con ggplot2



```
# Covarianza
covarianza_xy <- cov(x, y)

# Coeficiente de correlación de Pearson
pearson_xy <- cor(x, y, method = "pearson")

# Coeficiente de correlación de Spearman
spearman_xy <- cor(x, y, method = "spearman")

list(
  Covarianza = covarianza_xy,
  Pearson = pearson_xy,
  Spearman = spearman_xy
)
```

```
## $Covarianza
## [1] 1.644605
##
## $Pearson
## [1] 0.965907
##
## $Spearman
## [1] 0.9576118
```

En el código anterior:

- `cov(x, y)` calcula la covarianza muestral entre X e Y .
- `cor(x, y, method = "pearson")` calcula el coeficiente de correlación de Pearson.
- `cor(x, y, method = "spearman")` calcula el coeficiente de correlación de Spearman.

Para calcular la matriz de covarianzas de dos o más variables (por ejemplo, x , y y otra variable z), se puede usar la instrucción:

```
# Generamos datos de ejemplo
x <- rnorm(100)
y <- 2 * x + rnorm(100, sd = 0.5)
z <- 3 * x + rnorm(100, sd = 0.3)

# Matriz de covarianzas de x, y, z
matriz_cov_xyz <- cov(cbind(x, y, z))
matriz_cov_xyz

##           x           y           z
## x 1.124521 2.295732 3.394896
## y 2.295732 4.955565 6.946020
## z 3.394896 6.946020 10.325457

# Matrices de correlación para x, y, z
# Matriz de correlaciones (Pearson)
matriz_cor_pearson_xyz <- cor(cbind(x, y, z), method = "pearson")
matriz_cor_pearson_xyz

##           x           y           z
## x 1.0000000 0.9725025 0.9962952
## y 0.9725025 1.0000000 0.9710349
## z 0.9962952 0.9710349 1.0000000

# Matriz de correlaciones (Spearman)
matriz_cor_spearman_xyz <- cor(cbind(x, y, z), method = "spearman")
matriz_cor_spearman_xyz

##           x           y           z
## x 1.0000000 0.9657006 0.9934473
## y 0.9657006 1.0000000 0.9603720
## z 0.9934473 0.9603720 1.0000000
```