

# Lab04 - COEFICIENTE DE CORRELACIÓN

Escriba Flores, Daniel Agustin

2025-04-13

## Contents

<b>Contexto:</b>	<b>1</b>
<b>Pregunta A: Diagrama de dispersion</b>	<b>1</b>
Lectura de datos . . . . .	2
Diagrama . . . . .	2
<b>Pregunta B: Correlacion y determinacion</b>	<b>3</b>
Interpretacion: . . . . .	4
<b>Pregunta C: Prueba de normalidad</b>	<b>4</b>
Interpretación: . . . . .	4
Conclusión: . . . . .	4

## Contexto:

El director del Zoológico de Tampa estudia la relación entre el número de visitantes, en miles, y la temperatura alta, en grados Fahrenheit. Selecciona una muestra de 15 días y la información muestral recopilada se tabula a continuación:

Visitantes (miles)	Temperatura (°F)	Visitantes (miles)	Temperatura (°F)
2.0	86	2.2	84
0.6	71	2.5	66
2.0	89	1.3	76
2.1	73	3.6	84
2.2	76	1.0	75
2.1	75	1.8	72
0.5	68	2.1	76
0.3	72		

De acuerdo a lo anterior, haga lo siguiente:

## Pregunta A: Diagrama de dispersion

- Traslade la información de la tabla mostrada a un documento de Excel, asígnele el nombre 'zoológico.xlsx', lea la base de datos y elabore un diagrama de dispersión para las variables numéricas.

## Lectura de datos

```
library(readxl) #Llamamos a la libreria

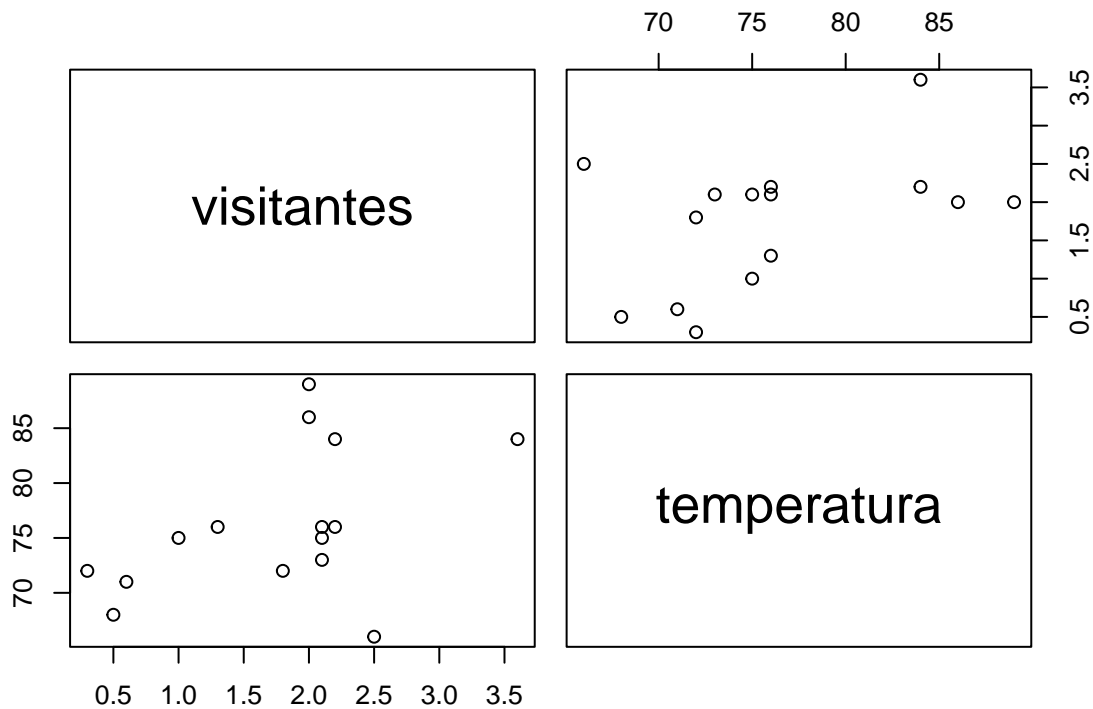
data = read_excel("zoologico.xlsx", sheet="Hoja1")
data = data[-1]
data = as.data.frame(data)
data
```

```
##      visitantes temperatura
## 1          2.0          86
## 2          0.6          71
## 3          2.0          89
## 4          2.1          73
## 5          2.2          76
## 6          2.1          75
## 7          0.5          68
## 8          0.3          72
## 9          2.2          84
## 10         2.5          66
## 11         1.3          76
## 12         3.6          84
## 13         1.0          75
## 14         1.8          72
## 15         2.1          76
```

## Diagrama

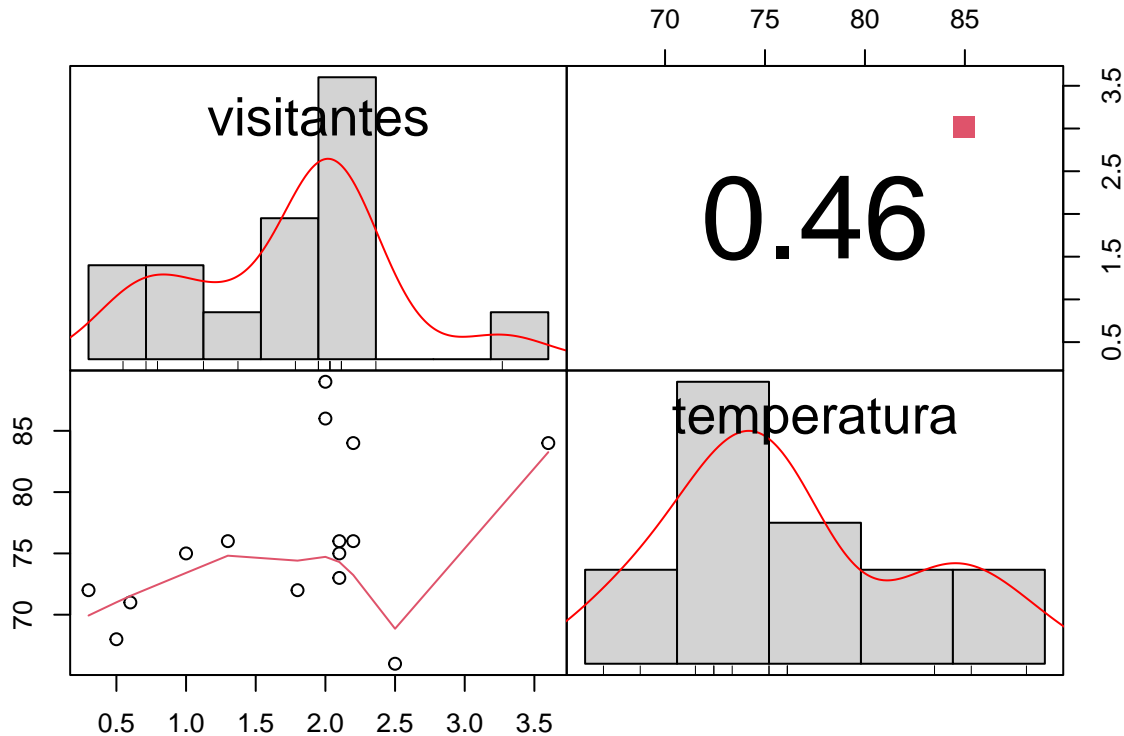
elaboramos

```
pairs(data)
```



```
# llamamos a la libreria para utilizar PerformanceAnalytics
library(PerformanceAnalytics)
```

```
chart.Correlation(data)
```



## Pregunta B: Correlacion y determinacion

b. Calcule e interprete el coeficiente de correlación y el de determinación.

```
cor(data) # Matriz de correlación
```

```
##           visitantes temperatura
## visitantes  1.0000000  0.4567093
## temperatura 0.4567093  1.0000000
```

Para calcular el coeficiente de determinación solo debemos elevar al cuadrado el valor del coeficiente de correlación.

```
x = data$temperatura
y = data$visitantes
```

```
coef_corr = cor(x, y, method = "pearson")
coef_corr # Coeficiente de correlacion
```

```
## [1] 0.4567093
```

```
coef_deter = coef_corr^2
coef_deter # Coeficiente de determinacion
```

```
## [1] 0.2085834
```

### Interpretacion:

- Coeficiente de correlación ( $r = 0.457$ ): Indica una relación positiva, pero moderadamente débil, entre la temperatura y el número de visitantes. Esto sugiere que, aunque días más cálidos tienden a atraer más público al zoológico, el efecto no es particularmente fuerte ni consistente.
- Coeficiente de determinación ( $R^2 = 0.209$ ): Revela que solo el 20.9% de las variaciones en la asistencia pueden atribuirse a cambios en la temperatura. La mayor parte (79.1%) depende de otros factores no considerados en este análisis, como días festivos, promociones, condiciones climáticas adicionales (lluvia, humedad) o temporadas escolares.

## Pregunta C: Prueba de normalidad

c. Realice la prueba de normalidad Shapiro-Wilk y diga si estas siguen o no una distribución normal.

```
shapiro.test(x)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  x  
## W = 0.92746, p-value = 0.25
```

### Interpretación:

El valor p (0.25) es mayor que el nivel de significancia típico ( $\alpha = 0.05$ ). Esto indica que no hay evidencia suficiente para rechazar la hipótesis nula de normalidad.

### Conclusión:

Los datos analizados siguen una distribución normal según la prueba de Shapiro-Wilk.