

Correlacion_PuntoBiserial.R

Usuario

2025-11-13

```
# Ejercicio correlacion de spearman
# JEGR
# 25/09/2025

set.seed(123) # Para fijar el Logaritmo y siempre que corra el codigo
sean los
# mismos datos

# Numero de observaciones
n <- 20

# Generar horas de estudio (entre 1 y 10)
Horas_estudio <- sample(1:10, n, replace = T)

# Asignar probabilidad de aprobar en funcion de horas de estudio

# A mas horas, mas alta probabilidad

Resultado <- sapply (Horas_estudio, function(horas) {
  ifelse(runif(1) < (horas / 10), "Aprobado", "Reprobado")
})

# Crear data frame

estudio <- data.frame(
  Estudiante = 1:n,
  Horas_estudio,
  Resultado
)
estudio

##   Estudiante Horas_estudio Resultado
## 1             1            3  Reprobado
## 2             2            3  Reprobado
## 3             3           10  Aprobado
## 4             4            2  Reprobado
## 5             5            6  Aprobado
## 6             6            5  Aprobado
## 7             7            4  Reprobado
## 8             8            6  Aprobado
## 9             9            9  Aprobado
## 10            10           10 Aprobado
```

```

## 11      11      5  Aprobado
## 12      12      3  Reprobado
## 13      13      9  Aprobado
## 14      14      9  Aprobado
## 15      15      9  Aprobado
## 16      16      3  Aprobado
## 17      17      8  Aprobado
## 18      18     10  Aprobado
## 19      19      7  Aprobado
## 20      20     10  Aprobado

# Convertir puntuacion de examen en numeros asignando valores de 1 (apr)
y 0(susp)
estudio$Resultado_bin <- ifelse(estudio$Resultado == "Aprobado", 1, 0)
head(estudio)

##   Estudiante Horas_estudio Resultado Rsultado_bin
## 1           1       3 Reprobado      0
## 2           2       3 Reprobado      0
## 3           3      10  Aprobado      1
## 4           4       2 Reprobado      0
## 5           5       6  Aprobado      1
## 6           6       5  Aprobado      1

cor.test(estudio$Horas_estudio, estudio$Rsultado_bin,
         method = "pearson")

##
## Pearson's product-moment correlation
##
## data: estudio$Horas_estudio and estudio$Rsultado_bin
## t = 4.5536, df = 18, p-value = 0.0002461
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4275568 0.8869888
## sample estimates:
##        cor
## 0.7316478

# Media de horas de estudio de Los aprobados
mean_aprobados <- mean(estudio$Horas_estudio[estudio$Resultado ==
"Aprobado"])
mean_aprobados

## [1] 7.733333

# Media de horas de estudio de Los reprobados
mean_reprobados <- mean(estudio$Horas_estudio[estudio$Resultado ==
"Reprobado"])
mean_reprobados

## [1] 3

```

```
sd(estudio$Horas_estudio)
## [1] 2.874113
N <- length(estudio$Estudiante)
n
## [1] 20
```