BY :
ARWA

2022

TUWAIQ
DATA-
SCIENCE
BOOTCAMP

POINT BISERIAL

UNDERSTANDING :

→ POINT BISERIAL

* Concept & Definition :

is used to understand the strength of the relationship between two variables. Your variables of interest should include one continuous and one binary variable. See more below.

RESULT OF

POINT-BISERIAL CORRELATION?

$$r_{pb} = \frac{(\bar{y}_1 - \bar{y}_2) \cdot \sqrt{pq}}{s_y}$$

1) IF it's closer to 1 positive correlation

2) IF it's closer to −1 negative correlation

3) IF it's closer to 0 no correlation

EXAMPLE

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
a = np.array([0, 0, 0, 1, 1, 1, 1])
b = np.arange(7)
print(stats.pointbiserialr(a,b))
print(stats.pearsonr(a, b))
np.corrcoef(a, b)
```

```
PointbiserialrResult(correlation=0.866025403784386, pvalue=0.011724811003954649)
PearsonRResult(statistic=0.866025403784386, pvalue=0.011724811003954649)

array([[1.       , 0.8660254],
       [0.8660254, 1.       ]])
```

LET'S DISCUSS AND ANALYZE

ASSUMPTIONS FOR POINT-BISERIAL CORRELATION

Normally Distributed 01

No Outliers 02

Equal Variances 03