

From Univariate to Multivariate Analysis

01

Univariate Analysis

Examining variables individually, one at a time

Example: Looking at a class's average test scores alone.

02

Multivariate Analysis

Studying interactions between multiple variables simultaneously

Example: Analyzing how test scores change based on the number of study hours.

03

Association

Observing one variable reveals information about another

Example: Discovering that knowing a student's study hours helps predict their test performance.

Examples of Associated Variables

Height & Weight

Taller individuals tend to weigh more than shorter individuals

Education & Income

Higher education levels are often linked to higher earnings

Age & Blood Pressure

Older individuals tend to have elevated blood pressure

Study Time & Performance

More hours studied typically correlate with better exam scores

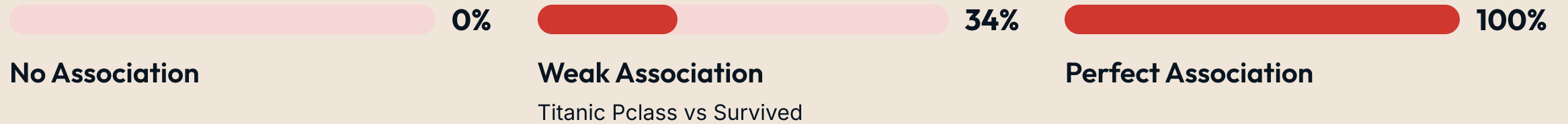
Cramér's V Statistic

Normalised version of Pearson's χ^2 that adjusts for sample size:

$$V = \sqrt{\frac{\chi^2}{n \cdot (\min(k, l) - 1)}}$$

```
from scipy.stats.contingency import association
print(f"{association(pd.crosstab(titanic['Pclass'], titanic['Survived'])):0.2f}")
```

0.34



Scatterplots & Scatter Matrix

Visualising bivariate relationships through scatterplots reveals patterns:

- Positive correlation (upward trend)
- Negative correlation (downward trend)
- Linear vs non-linear relationships
- No apparent correlation

