Learning Objectives

Learning Objectives:

1. Perform a correlation(Pearson test)
2. Learn about parametric and non-parametric test
3. Perform a chi-square test
4. Coding Exercises

Correlation test

Correlation is a statistical method to measure and describe the relationship between two variables.

Note: unlike previous tests like (t, tests or anova). it does not compare group differences but rather test relationships between two variables

Pearson

Pearson Correlation: one of the most commonly used correlation tests. It measures the direction and the strength (= degree) of the linear relationship between two variables. The result always has a value between -1 and 1.

- Form of the relationship: linear
- Direction of the relationship: positive or negative
- Strength of the relationship: little, weak, moderate, strong, perfect

| (a) Perfect | (b) No or little | (c) Moderate to strong | (d) Weak |
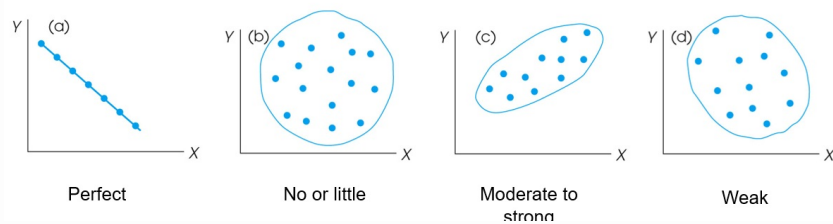
```python
# Import  libraries
import pandas as pd
from scipy.stats import pearsonr

# given data
x=[27, 32, 38, 94, 70, 29, 17, 8, 48, 82, 52, 14, 91, 22, 58,
96, 73, 99, 75, 76]
y=[78, 35, 28, 75, 2, 22, 25, 17, 72, 45, 86, 96, 24, 41, 73,
51, 58, 76, 90, 77]

# Convert data to dataframe
data = pd.DataFrame (x,y)

# Apply the pearsonr()
corr, _ = pearsonr(x, y)
print('Pearsons correlation: %.3f' % corr)
```

This is an image of the data above
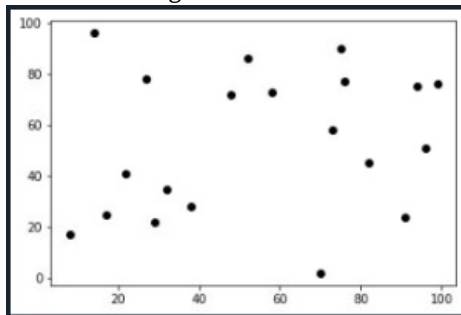
Chi-square

Parametric tests: Test about population parameters
* T-test
* ANOVA
* Correlation
* Regression

Non-Parametric tests: Test about population proportions / frequencies

- Chi-square tests

Chi-Square

A Pearson's Chi-Square test is used to determine if two independent sets of data are **associated** with each other. The Chi-Square test is best used for categorical data such as gender/sex, day of the week, etc. This is determined using the **p-value** that is calculated from the test. A p-value of less than or equal to 0.05 means the data sets are **associated** with each other while a higher p-value means the data sets are **not associated** with each other. Chi-square does not require satisfying specific assumptions about the population parameters.

example of chi-square test

```
# Defining the table
data = [[27, 32, 38, 94, 70, 29, 17, 8], [78, 35, 28, 75, 2, 22,
25, 17]]
stat, p, dof, expected = chi2_contingency(data)
```