Name: Kshitij Hundre  
Div:D15C

Roll No:18

**Exp 4 : Statistical Hypothesis Testing Using SciPy and Scikit-Learn**

**Aim:** Implementation of Statistical Hypothesis Test using Scipy and Sci-kit learn.

**Problem Statement:** Perform the following Tests:Correlation Tests:

a) Pearson’s Correlation Coefficient

b) Spearman’s Rank Correlation

c) Kendall’s Rank Correlation

d) Chi-Squared Test

### **Introduction to Hypothesis Testing**

Hypothesis testing is a statistical method used to make inferences about a population based on sample data. It helps in determining whether the observed results are due to chance or if there is a statistically significant relationship between variables.

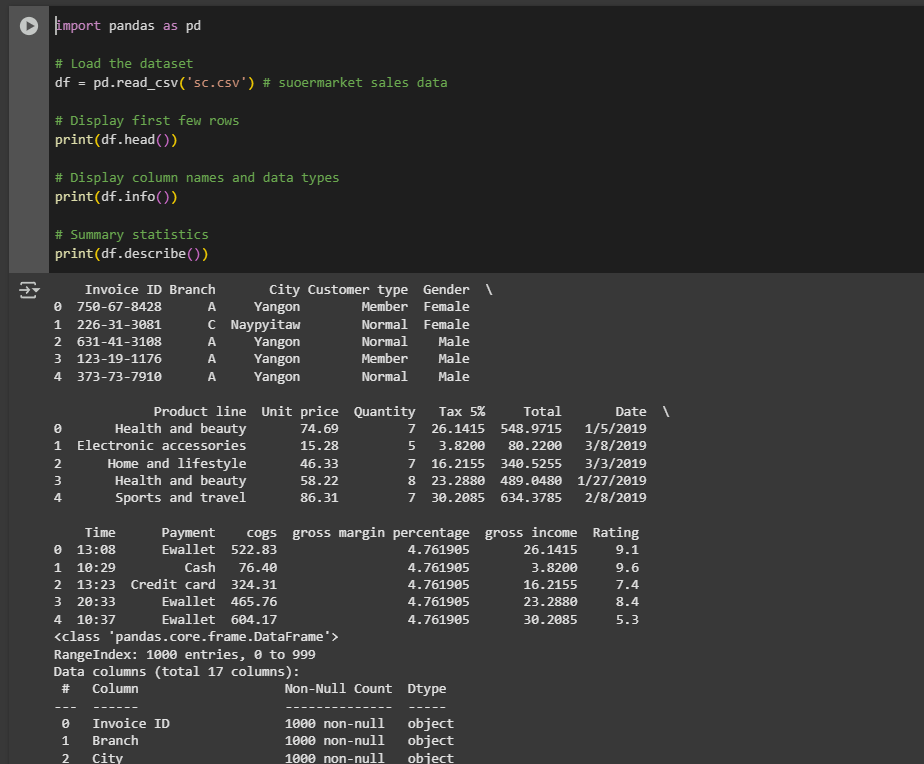
In this experiment, we will conduct **correlation tests and a chi-squared test** using Python's scipy.stats library.

**Theory and Output:**

**1.Loading dataset:**

Data loading is the first step in data analysis. The dataset is stored in a CSV file and read using pandas.read\_csv().

The first few rows are displayed to understand the dataset structure



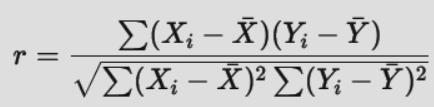
**2.Pearson’s Correlation Coefficient:**

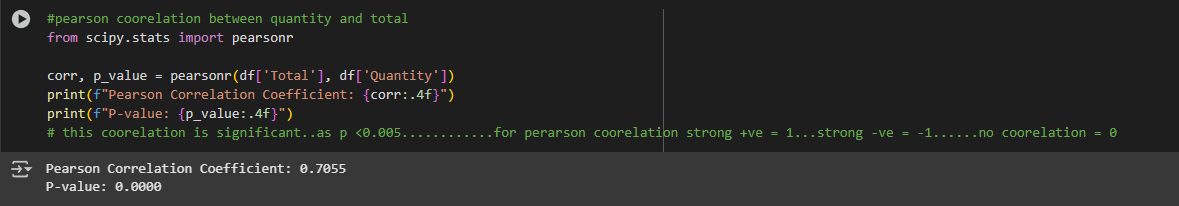
Pearson’s Correlation Coefficient (denoted as **r**) measures the **linear** relationship between two continuous variables.

Values range from **-1 to +1**:

* **+1**: Perfect positive correlation
* **0**: No correlation
* **-1**: Perfect negative correlation

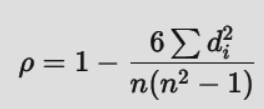
The formula for Pearson’s Correlation Coefficient is:

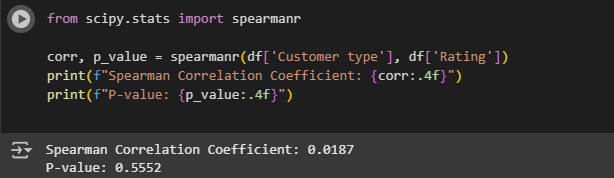




**3.Spearman’s Rank Correlation**

* Spearman’s Rank Correlation (denoted as ρ, rho) measures the monotonic relationship between two variables.
* It does not require normally distributed data.
* If ranks of two variables are related, it indicates correlation.
* The formula is:

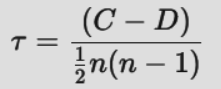


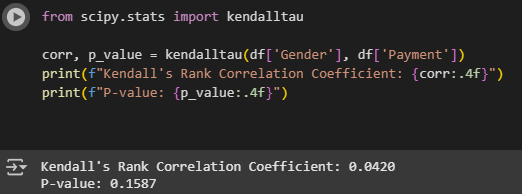


## **4.Kendall’s Rank Correlation**

### **Theory:**

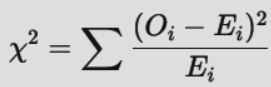
* Kendall’s Tau (τ) measures the **ordinal association** between two variables.
* It counts **concordant** and **discordant** pairs:
  + **Concordant pairs**: If one variable increases, the other also increases.
  + **Discordant pairs**: One increases while the other decreases.
* The formula is:

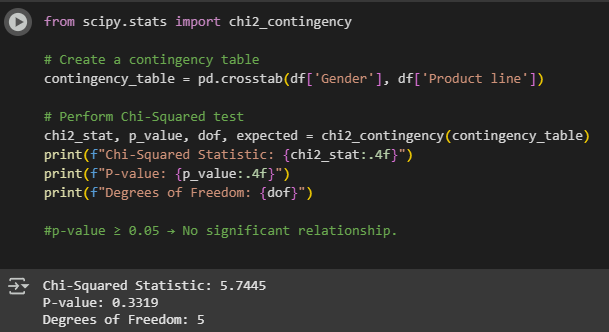




**5. Chi-Squared Test**

* The **Chi-Squared Test** is used for **categorical data** to check if two variables are independent.
* It compares **observed** and **expected** frequencies.
* The formula is:





## **Conclusion**

1. **Pearson’s Correlation**: Measures **linear relationship** between numerical variables. If **p < 0.05**, the correlation is significant.
2. **Spearman’s Correlation**: Checks for **monotonic relationship**. If **p < 0.05**, variables move together in a ranked order.
3. **Kendall’s Correlation**: Identifies **ordinal association**. A small **p-value** means a strong relationship.
4. **Chi-Square Test**: Determines **independence of categorical variables**. If **p < 0.05**, variables are dependent; otherwise, they are independent.

**Final Summary:**

* If **p < 0.05**, the test indicates a significant relationship.
* If **p > 0.05**, no strong relationship exists.

These tests help understand **associations** in the dataset for data-driven decisions.