**SMA 2453: SIMULATION AND MODELLING**

**GROUP 7 CAT 2**

**THE FREQUENCY TEST**

**GROUP MEMBERS**

1. SCM 211-0368/2020 MICHAEL OWISO
2. SCM 211-0632/2020 HAWKINS LUMASIA
3. SCM 211-0750/2020 CYNTHIA MOGERE
4. SCM 211-0639/2020 HELEN PATIENCE WANGARI
5. SCM 211-0334/2020 WANGAI GLORIOUS MELODY
6. SCM 211- 0282/2020 NGOBO ANNASTACIA NJERI

### Frequency Test in Testing for Random Numbers

**Repository:**

[GItHub](https://github.com/Okwizi/Frequency-Test-for-Random-Numbers)

**Definition:**

Frequency test is a statistical method used to assess the randomness of a sequence of numbers or data points. It examines the distribution of values within the sequence to determine if they follow an expected pattern or if there are any deviations that suggest non-randomness.

**History:**

The frequency test is one of the earliest methods developed for testing randomness. It dates back to the work of Karl Pearson and Charles Spearman in the late 19th and early 20th centuries. They laid the groundwork for statistical methods to analyze data patterns, including tests for randomness.

**Developers:**

While Pearson and Spearman contributed to the theoretical foundations of frequency testing, the specific methods and tests used today have been further refined and developed by various statisticians and researchers in the field of probability and statistics.

**Assumptions:**

- The sequence of numbers under test is assumed to be independent and identically distributed (i.i.d.).

- The expected distribution of values is known or can be determined.

- The sample size is sufficiently large to draw meaningful conclusions.

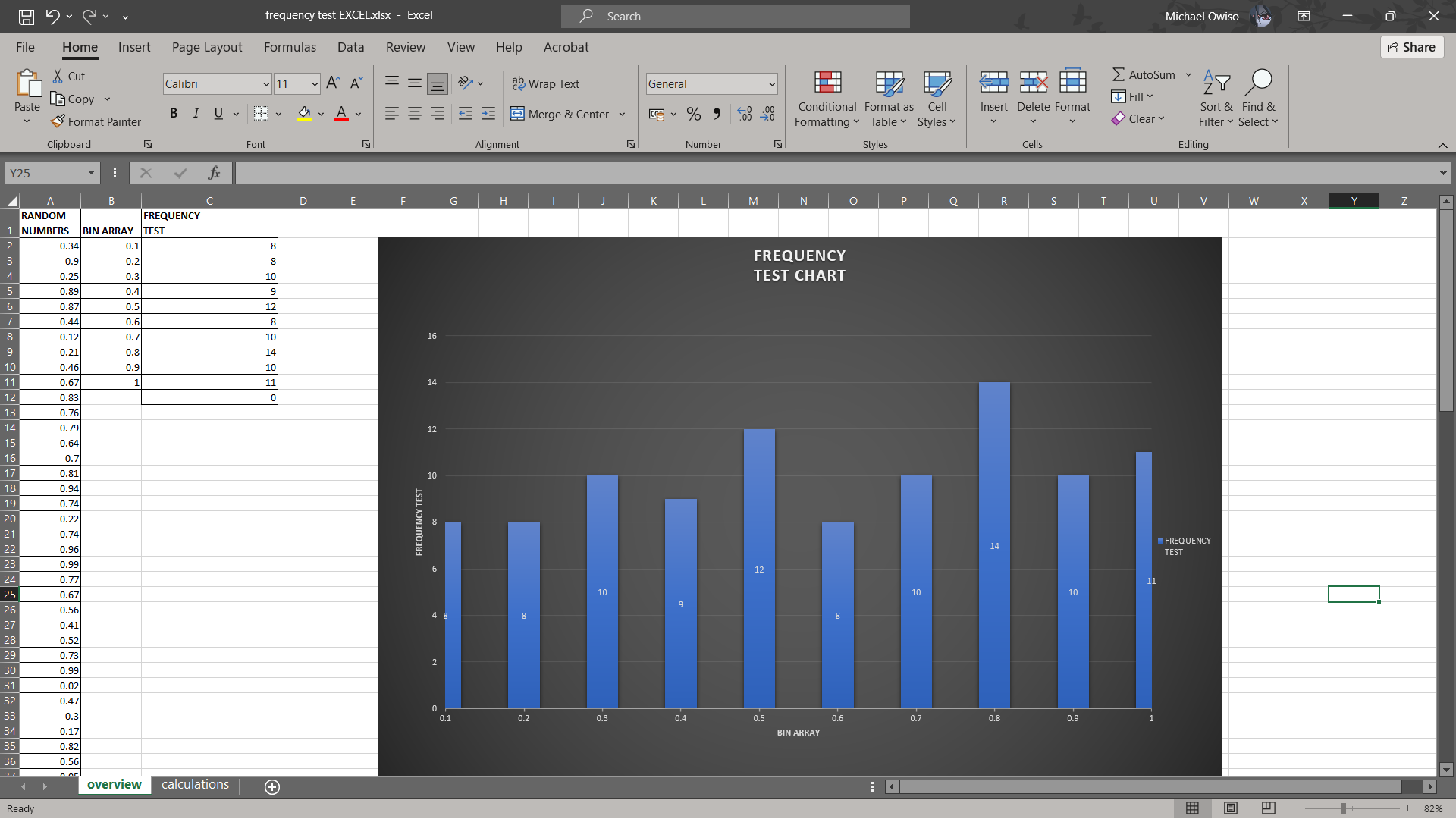
- The level of significance is 0.05.

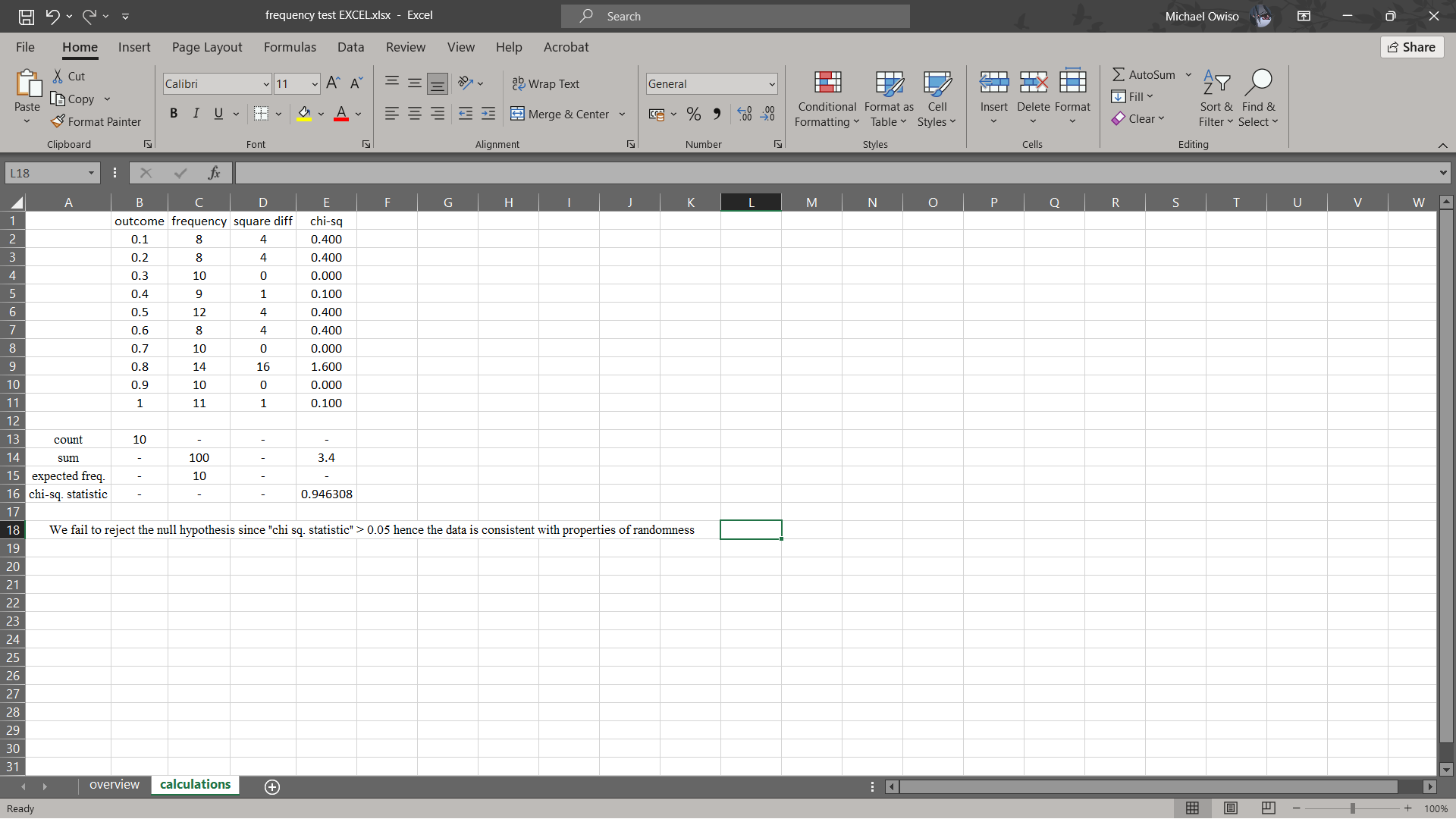
**Ways to Perform Frequency Test:**

**1.FREQUENCY TEST BY USE OF EXCEL**

STEPS TO FOLLOW

1. Generate the random numbers using RAND() or enter the data set of random numbers in one column
2. Enter an array of bins or intervals. This is the score ranges you wish to analyze
3. Create a column on ins adjacent to the random numbers
4. Calculate the frequency using the formula =FREQUENCY(data\_array, bin\_array).
5. Display the frequency in the adjacent cell.
6. You can generate a histogram to display a table of the generated information.
7. Now calculate the expected frequency for the data and degrees of freedom.
8. Calculate the sq. difference of the data.
9. Calculate the sq. square values for each data.
10. Add the chi-sq. values to get the chi-square statistic.
11. Use =CHISQ.DIST.RT(param 1, param 2) function to calculate the p-value for the data.
12. If the p-value is > than 0.05, we fail to reject the null hypothesis i.e. the data is consistent with the properties of randomness. Otherwise we reject the null hypothesis.





**2.FREQUENCY TEST USING PYTHON**

This is the frequency test equivalent in python code.

**Python Dependencies Setup Instructions**

**Prerequisites**

Ensure that Python is installed on your system.

**Installation Steps**

Follow these steps to install the necessary Python libraries:

1. Open a terminal or command prompt.
2. Execute the following commands to install the dependencies:

pip install --user numpy

pip install --user pandas

python -m pip install --user scipy

**Code Explanation**

The provided code snippet defines a function named frequency\_test that performs a frequency test on a dataset against a specified distribution. It calculates the expected frequencies based on the chosen distribution, normalizes these frequencies, and then conducts a chi-square test to compare the observed frequencies with the expected frequencies. The function returns the chi-square statistic and the p-value from this test.

**Inputs**

* data :A NumPy array containing the dataset to be analyzed.
* distribution : A string indicating the type of distribution (e.g., 'uniform', 'binomial', 'poisson', 'exponential', 'geometric') to use for calculating expected frequencies.
* \*\*dist\_params : Keyword arguments that provide additional parameters required by the specified distribution.

**Process Flow**

1. Data Preparation: The input data is flattened, and the unique values along with their counts in the dataset are identified.
2. Expected Frequency Calculation: Depending on the specified distribution, the function calculates the expected frequencies for each unique value in the dataset.
3. Normalization: The expected frequencies are normalized to ensure that their sum matches the sum of the observed frequencies.
4. Chi-Square Test: A chi-square test is performed using the observed frequencies and the normalized expected frequencies.
5. Results: The function returns the chi-square statistic and the p-value resulting from the chi-square test.

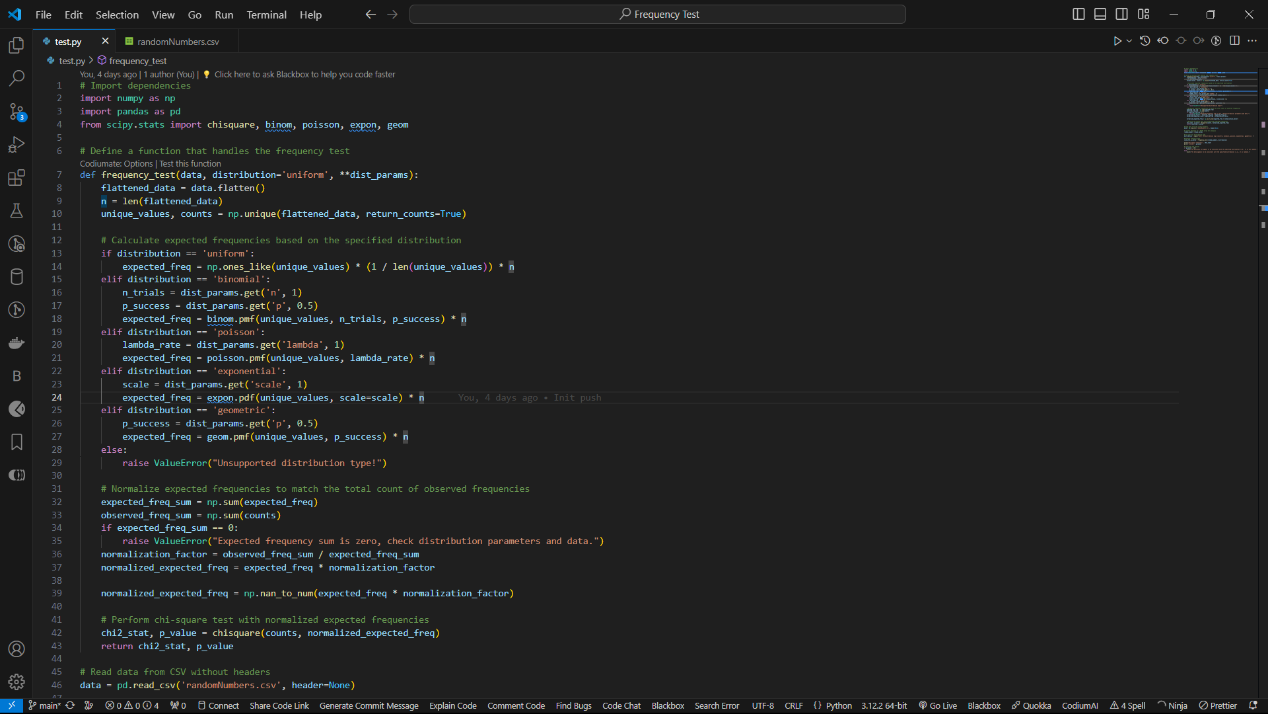
**Outputs:**

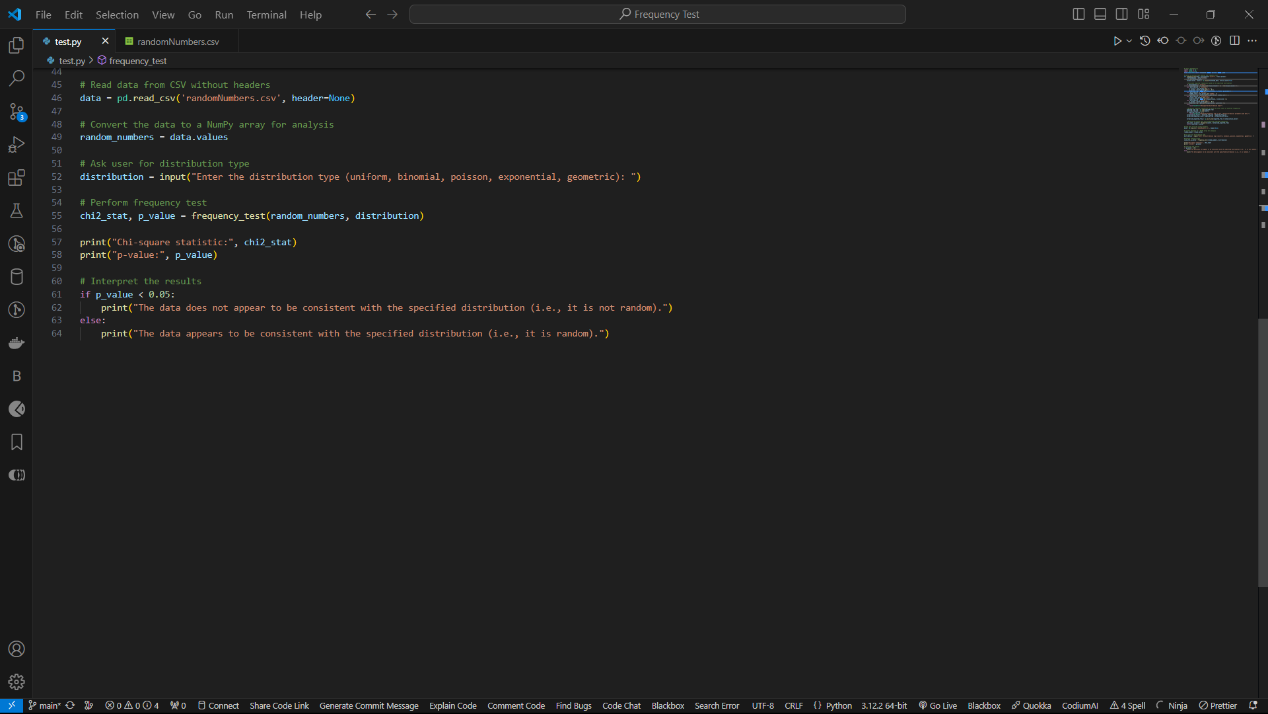
* chi2\_stat : The chi-square statistic, indicating the difference between observed and expected frequencies.
* p\_value : The p-value associated with the chi-square statistic, used to determine the statistical significance of the observed difference.

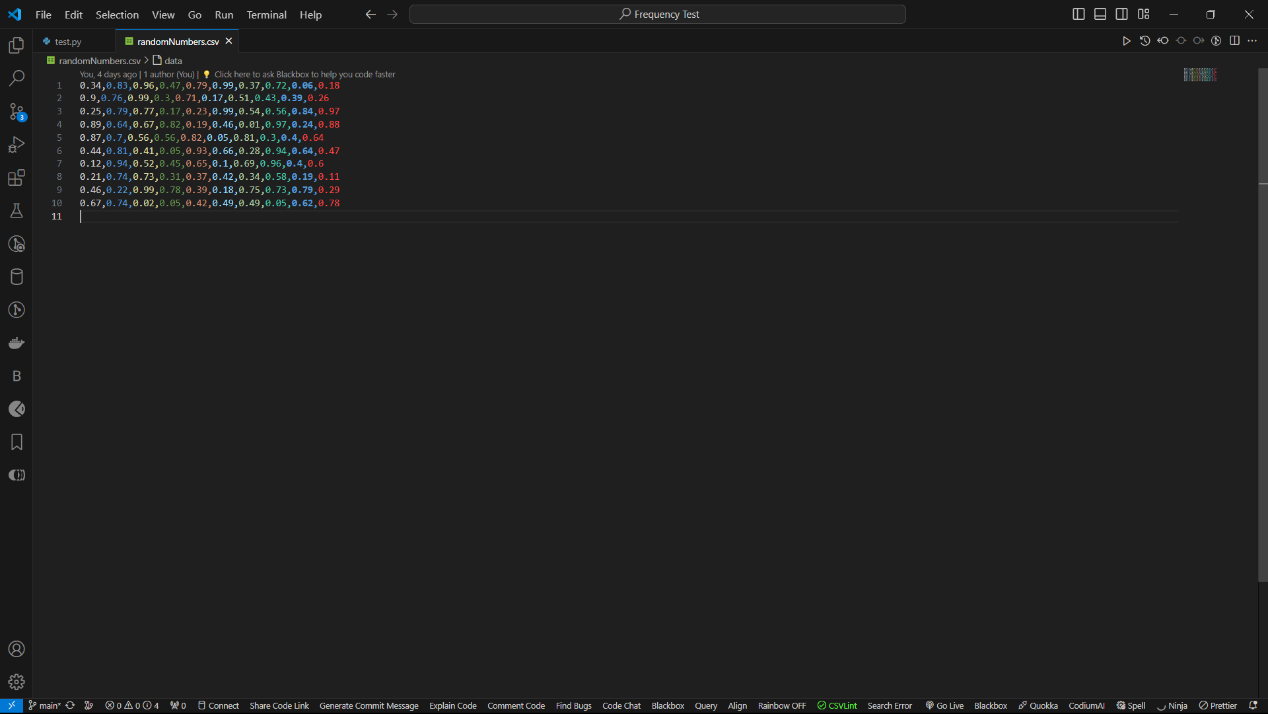
**NOTE:** The dataset provided will only work for a uniform distribution since the value for expected\_freq in frequency\_test gives a mismatch between the sum of observed frequencies and the sum of expected frequencies; causing a discrepancy beyond a certain tolerance level.

**Usage:**

In the project directory, open a terminal and run python test.py







**Advantages:**

- Simple and easy to understand.

- Can provide valuable insights into the randomness of a sequence.

- Widely applicable across various fields, including cryptography, finance, and quality control.

**Disadvantages:**

- Limited in detecting certain types of non-randomness, such as subtle patterns or dependencies.

- Results may be influenced by sample size and the chosen significance level.

- Assumes a known or easily determinable expected distribution, which may not always be the case.

**Suitability for Use:**

Frequency tests are suitable for assessing the randomness of large datasets with known or easily determinable expected distributions. They are commonly used in quality control processes, cryptographic applications, and scientific research where randomness is essential. However, they may not be sufficient on their own and are often complemented by other tests and methods for a more comprehensive analysis of randomness.

References:

1. Marsaglia, George. "Random number generators." Journal of Modern Applied Statistical Methods 2.1 (2003): 2.

2. NIST Special Publication 800-22, "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications," available at: https://csrc.nist.gov/publications/detail/sp/800-22/rev-1a/final.

3. Pearson, Karl. "Mathematical contributions to the theory of evolution: On the law of ancestral heredity." Proceedings of the Royal Society of London 59.350-352 (1896): 69-71.

4. Knuth, Donald E. "The Art of Computer Programming, Volume 2: Seminumerical Algorithms." Addison-Wesley Professional, 1997.

5. L’Ecuyer, Pierre. "Random number generation." Handbook of computational statistics: Concepts and methods. Springer, Berlin, Heidelberg, 2004. 35-71.

6. Ahrens, J., Dieter, U., & Stoll, D. "Software Development of a Linear Congruential Random Number Generator." Computing 3.2 (1987): 85-106.

These articles cover various aspects of random number generation, frequency tests, and their applications in different fields.