

# 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`