DA Assignment 3

K.Kalyan Reddy, 21361

1 Importing the dataset

- Reads data from a file specified by the `filepath` parameter.
- Processes the data by splitting it and converts the filtered data into a NumPy array and returns it.

2 Implementation Summary

1. The `get_N_and_D_matrices` function generates matrices N and D.These are generated in according to the order of m_s,m_ns,f_s,f_ns given in the dataset .txt file
2. These matrices are reshaped into 48x4 arrays for further calculations.
3. The `get_p_values` function computes p-values for a given each row in the data. We use F-statistic formula for this.

$$\frac{1/(\text{rank}(D) - \text{rank}(N))}{1/(n - \text{rank}(D))} \times \left(\frac{X^T \left(I - N \left(N^T N \right)^\dagger N^T \right) X}{X^T \left(I - D \left(D^T D \right)^\dagger D^T \right) X} - 1 \right)$$

4. We will calculate p-value by calculating the probability right to F-statistic
5. Finally, the function returns the list of computed p-values and plots a histogram for the p-values.

3 Results

3.1 Histogram Plot

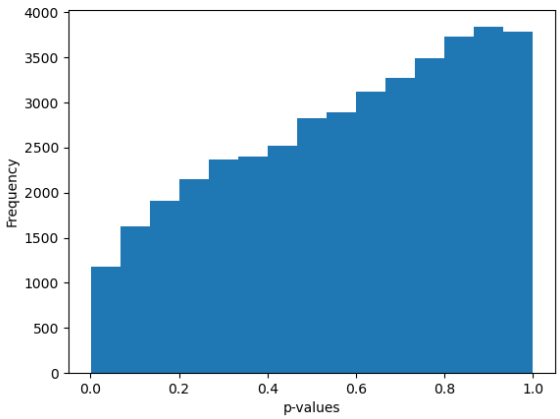

Figure 1: p-values Histogram

3.2 Inferences

A greater number of rows of particular interest implies a higher frequency of rejections of the null hypothesis, as seen by the significant concentration of rows with p-values approaching 1 in the histogram.