

Data Analytics Assignment 3

B Srinath Achary (SR No: 21441)

1 Implementation

The implementation of the analysis process involves several key steps:

1. **Data Extraction:** The first step involves the extraction of the required columns from the dataset. Specifically, columns 1 to 49 are extracted, where the first 12 columns represent data for male non-smokers, the next 12 for male smokers, the subsequent 12 for female non-smokers, and the remaining columns for female smokers.
2. **Matrix Computation:** Following data extraction, matrices N and D are computed as defined in class slides. These matrices are fundamental in calculating the F-statistic used in the Two-Way Analysis of Variance (ANOVA) analysis.
3. **F-Statistic Calculation:** In this step, the F-statistic is calculated for each row (probe) of the dataset. The F-statistic is computed using the following formula:

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

where X is 48×1 vector containing the intensities of a certain probe corresponding to individuals and $n = 48$ (for this dataset). Note that rows (probes) for which the denominator in the F-statistic is zero are ignored. In the analysis, there were seven such rows.

4. **P-Value Computation:** Subsequently, p-values for the F-distribution are computed. The parameters for the F-distribution are determined as follows: $a = \text{Rank}(D) - \text{Rank}(N)$ and $b = n - \text{Rank}(D)$. These p-values are computed based on the F-statistics obtained in the previous step.
5. **Histogram Generation:** Finally, a histogram of the computed p-values is generated. This histogram provides a visual representation of the distribution of p-values obtained from the Two-Way ANOVA analysis.

2 Plot

Figure 1 shows the histogram of p-values obtained from the Two-Way ANOVA analysis. The red dashed line indicates the significance threshold at $p = 0.05$.

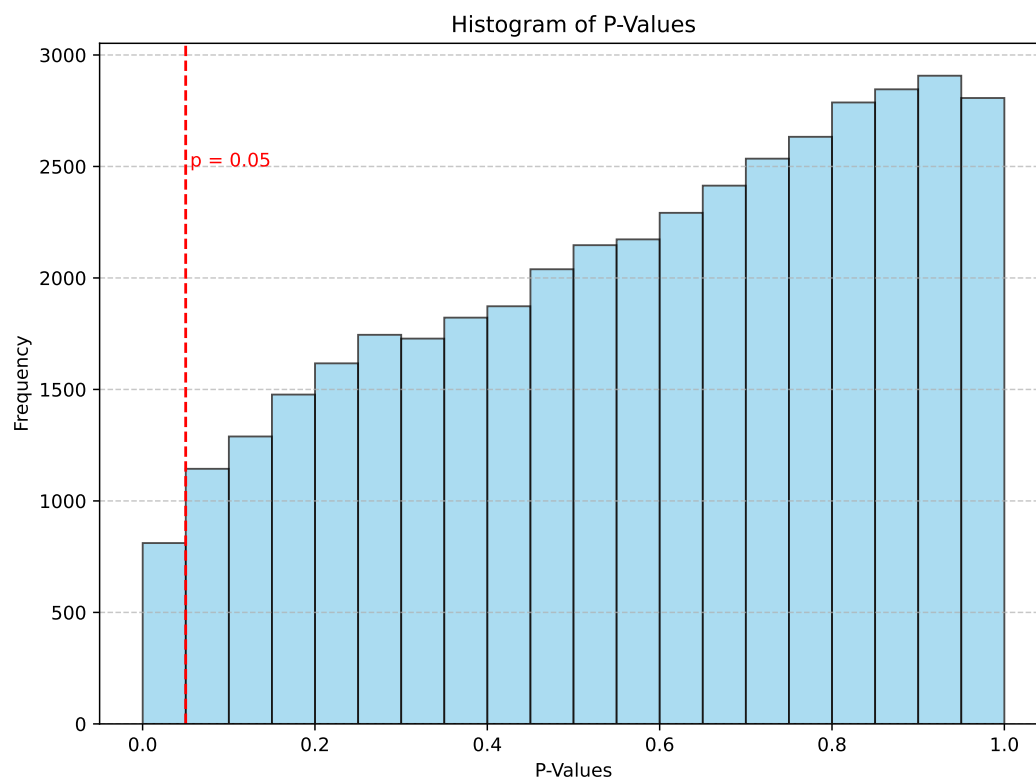


Figure 1: Histogram of P-Values