

## **Report: Analysis of Annual Salary Data**

### **1. Description of the Data**

The dataset under consideration comprises annual salary values extracted from the 'data5-1.csv' file. This dataset, with 4000 salary entries, lacks a header, implying that it is a single-column dataset representing annual salaries in an unspecified currency.

### **2. Distribution Analysis**

The distribution of annual salaries was explored through the creation of a histogram. This graphical representation illustrates the frequency distribution of salaries across predefined bins. From the histogram, we observe a relatively diverse distribution of salaries, indicating variations in income levels within the dataset.

### **3. Mean Value Calculation**

The mean annual salary ( $\bar{w}$ ) was calculated as a measure of central tendency for the dataset. The mean is computed using the formula:

$$\bar{w} = \frac{\sum_{i=1}^n x_i}{n}$$

Where  $x_i$  represents individual salary values and  $n$  is the total number of values. For this dataset, the calculated mean salary is  $\bar{w}=27884.54$ .

### **4. X Value Calculation**

The calculation of the required value  $X$  was based on the 10th percentile. This implies that 10% of the population has a salary above this value. The formula to determine  $X$  at a certain percentile is:

$$X = \text{percentile}(X_{\text{data}}, p)$$

Where  $X_{\text{data}}$  is the dataset, and  $p$  is the desired percentile. In our analysis, the calculated  $X$  value at the 10th percentile is  $X=8760.30$ .

### **Comments**

The diversity observed in the salary distribution highlights the heterogeneity of income levels within the dataset. The mean salary provides a central reference point, while the calculated  $X$  value offers insights into income levels below which a specific percentage of the population falls.