Coding project

1. Data description

The dataset comprises annual salary values in Euros. The data, loaded from the CSV file "data8-1.csv," consists of 4000 salary entries, ranging from 452 to 148526.

2. Description of Distribution:

The y-axis indicates the probability of each range, while the x-axis represents value ranges. This graph, which is a histogram, illustrates the distribution of values in a dataset. With more instances of lower values and fewer of higher values, the data is skewed to the right. Lower values are more prevalent in this dataset, as shown by the majority of values being grouped on the left side of the graph. Despite the term "Histogram with Mean" on the graph, there is no visible indication of the mean. There are no extra annotations or grid lines, and the bars are all the same shade of blue.

3. Calculation of Mean of salaries (˜W):

The mean salary (˜W) is calculated using the formula

$$\tilde{W} = \frac{\sum_{i-1}^{n}(x_i \cdot f_i)}{\sum_{i-1}^{n} f_i}$$

where
midpoint of each bin is X_i,
and f_i is the frequency
The mean salary for the given data 34125

4. Calculate another value, X

First variance is using mid point and mean and probaility of bins

$$\sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2 \cdot P(x_i)$$

Then standard deviation is figured out through variance.

STD = square root of variance