**Fundamentals of Data Science**

**SHARON ZACHARIA | 22074407**

**Data Description**

The analysis aims to understand the distribution of salaries, calculate the mean value (W̃), and determine a specific value X. The give data consists of 4000 observations containing a statistically representative sample of salaries of a European country.

**Distribution Description**

Descriptive Statistics of the given dataset involves the following.

**Mean**: The mean salary is calculated to be 44065.53 it represents the average of the salaries across the dataset.

**Median**: The median salary is found to be 40793.0 representing the middle value dividing the dataset into two equal halves.

**Mode**: The mode was found to be 27225 representing the most frequently occurring value in dataset.

**Range**: The Range of the distribution is 169039, it demonstrates the spread between the minimum and maximum salaries around the mean.

**Standard Deviation**: The standard deviation is 24017.82 indicating the extent of deviation or dispersion values from the mean.

**Skewness**: The skewness of the distribution was found to be 0.743 it measures the symmetry of the distribution around the mean.

**Kurtosis:** The Kurtosis of the distribution was found to be 0.542.

**Mean Value**

Mean is a measure of central tendency that represents a typical value in a dataset.

Mean can be calculated in the following way.

***A black and white math symbol

Description automatically generated***

- *n* is the total number of data points

- xi represents each individual salary

The Mean value of the given dataset is found to be 44065.53.

**Value of X**

The value X, which is the fraction of population with salaries between 0.75 ̃W and ̃W (where W is the mean) and is calculated using: -

Number of salary data between Upper and lower bounds divided by the count of total salary observations. (upper bound = mean salary and lower bound = .75 \* mean salary). Value of X = 0.19

A graph of a function

Description automatically generated with medium confidence

**Conclusion**

The PDF of the function is shown on the histogram and the specific value of X was found out to be 0.19. which means 0.19 of the total population’s salary falls between 0.75 of the mean and mean value of the data.