Annual Salary Distribution Analysis

Name: Sam Jacob ID:22073281

Data Overview:

The dataset contains individual annual salaries in euros, aimed at analyzing the distribution and identifying the mean salary (W~) and the specific value 'X' that separates the top 33% of earners. A histogram with thirty bins displays the salary data, alongside a fitted normal distribution for comparison. The distribution exhibits positive skewness, indicating a tendency towards higher salaries among a subset of individuals.

Distribution Characteristics:

The salary distribution, as visualized in the histogram, displays a positive skewness with a value of 0.6015, indicating a longer tail towards higher salaries. The kurtosis of the distribution is 0.1908, which is less than that of a normal distribution, suggesting a flatter peak and fewer extreme outliers. These characteristics imply that while most individuals earn around the lower salary range, there is a significant number of people with incomes above the average, leading to an asymmetric distribution.

Mean Calculation:

The mean annual salary is calculated as the average of all salary values. The mean annual salary for the dataset is calculated as €50,320.26.

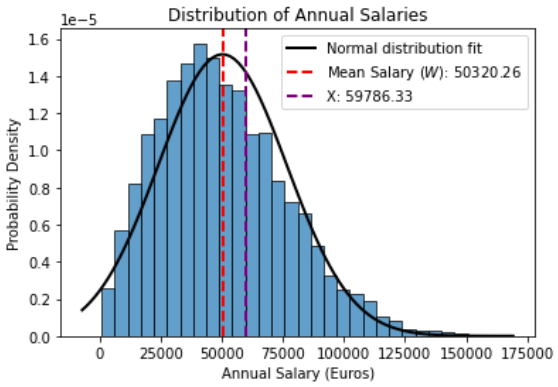
The mean (μ) of the salary data is calculated using the NumPy library's np. mean function.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Where x_i represents each individual's salary, and N is the total number of data points.

Value X Calculation:

To find the required value X, representing a specific percentile (67th percentile in this case), we use the formula for the percentile: X=Percentile(X) X is set to the 67th percentile, ensuring that 33% of people have a salary above X (since 100−67=33). Here the observed x value is 59786.33



Conclusion:

The dataset reveals an average annual salary of €50,320.26 and a 67th percentile value of €59,786.33, indicating that 33% of individuals earn above this threshold, reflecting a positively skewed salary distribution with a majority earning less and a smaller proportion significantly more. This analysis provides valuable insights into the distribution's central tendency and income disparities within the dataset.

.