

Introduction

The dataset under consideration pertains to annual salaries, containing an array of salary values. The objective is to understand the distribution of salaries and calculate specific statistical measures.

Data Description:

The dataset comprises a sample of annual salaries, assumed to be representative of a larger population. Each data point represents an individual's annual income.

Distribution Analysis:

To characterize the distribution of salaries, a probability density function (PDF) was employed. A Gaussian (normal) distribution was fitted to the data using the mean (μ) and standard deviation (σ) obtained through the `norm.fit` function.

Distribution Parameters:

The Gaussian distribution is defined as:

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x - \mu)^2}$$

Mean Salary Calculation:

The mean annual salary (μ) was calculated as the estimated mean (μ) obtained from the fitted distribution. The mean is given by:

$$\mu \sim \mu$$

Fraction Calculation (X):

To find the fraction (X) of the population with salaries between $0.75\mu \sim 1.25\mu$, the cumulative distribution function (CDF) of the Gaussian distribution was utilized. The fraction is calculated as the difference between the CDF values at the upper and lower bounds:

$$X = \text{CDF}(1.25\mu) - \text{CDF}(0.75\mu)$$

Results:

- The mean annual salary (μ) is calculated as (μ)
- The fraction (X) of the population with salaries between $(0.75\mu \sim 1.25\mu)$ is calculated using the Gaussian distribution CDF.

Conclusion:

The analysis provides insights into the distribution of salaries and allows for the calculation of meaningful statistical measures. The mean salary represents the central tendency of the dataset, while the fraction (X) characterizes the proportion of the population within a specified salary range. These findings contribute to a comprehensive understanding of salary distribution, facilitating informed decision-making and further analysis.

