Salary Distribution and Statistical Analysis



**Introduction:**

The dataset was subjected to a comprehensive analysis, which involved fitting it to a normal distribution model and evaluating the proportion of the population within specific limits

**Describe the Data:**

The given data is a set of salary values, which are loaded from a CSV file (data9-1.csv) into a Pandas DataFrame. The dataset is stored in the 'Data' column.

**Describe the Distribution:**

The distribution of the data is visualized using a histogram and a fitted normal distribution (Gaussian distribution) curve. The mean (μ) and standard deviation (σ) of the data are calculated using the norm.fit method from the SciPy library. The histogram is plotted with 25 bins, and the fitted normal distribution curve is overlaid on the histogram.

**Calculating the Mean (mu) and Standard Deviation (std):**

The mean (mu) and standard deviation (std) of the data are calculated using the norm.fit function from the SciPy library. The mean is the average value of the dataset, and the standard deviation measures the spread or dispersion of the values around the mean.

Mathematical Formula:

* Mean (mu): *µ* = *∑ Xi / n* , *i = 1,2,3…n*
* Standard Deviation (std): 𝜎 = √∑ ( X – μ )2

n

**Calculation of Lower and Upper Bounds (0.8W̄ and 1.2W̄ ):**

The required value X is calculated by defining lower and upper bounds. In this case, the lower bound is set to 0.8 times the mean (0.8 \* mu) and the upper bound is set to 1.2 times the mean (1.2 \* mu).

Mathematical Formula:

* Lower Bound=0.8×*μ*
* Upper Bound=1.2×*μ*

**Fraction of Population within Bounds (X):**

The fraction of the population within the specified bounds (between 0.8 and 1.2) is calculated by filtering the data. The fraction (X) is then obtained by dividing the length of the filtered data by the total length of the original data.

The mathematical formula for the fraction (X) is:

*X*= Number of Data Points within Bounds / Total Number of Data Points.

**Result Interpretation:**

The mean salary(μ) is 34,159.03, the standard deviation(σ) is 23,432.85 and the fraction of the population with salaries between 0.8W̄ and 1.2W̄ is 0.26575. This indicates that approximately 26.575% of the population falls within the specified salary range.