ID: 1805088

Ans. to the Q. No. - (a)

After running "task_a.py", we get the histogram diagram, mean, median, standard deviation & skewness values.
The output is as follows:

Mean: 14.54
Median: 10.32
Standard Deviation: 14.09
Skewness: 1.78

As the data shows a positive skewness (greater than 0), it indicates that the distribution of given data is "Right-Skewed". It means there are more values at the lower end of the scale & the tail is longer on the right.

Geiven these characteristics, we can hypothesize that the distribution can be a

gamma or Weibull distribution, as they are used for modeling skewed data with a longer tail on one side.

Ans. to the Q.No. -(b)

From task-a, we have chosen either gamma distribution or Weibull distribution.

Gamma Distribution:

It is characterized by 2 parameters.

(i) α : shape parameter

(ii) β : scale "

Maximum Likelihood Estimation (MLE) seeks to find the parameter (α, β) that maximize the likelihood function given the data. The optimization process iteratively adjusts the parameters to find the values that result in the highest likelihood.

Result:

Aftere reunning "task-b.py", we get,

$\alpha = 1.004578222$

$\beta = 11.473229687$

Weibull Distribution:

It is also characterized by 3 parameters:

(i) c : shape parameter

(ii) λ : Scale "

(iii) γ : Location "

However, the location parameter is usually fixed at zero for non-negative data. So, it becomes a two parameter Weibull distribution.

MLE for Weibull is also about maximizing the likelihood function. Optimization like Newton-Raphson or other gradient

ascent methods are used to converge to the parameter estimates.

Result:

$$c = 0.011025422$$

$$\lambda = 14.604990711$$

Ans. to the Q.No -(c)

To compare the observed data with the expect frequencies from the gamma distribution, we can use Chi-square statistic for the goodness-of-fit test.

$$\chi_c^2 = \frac{\sum (O_i - E_i)^2}{E_i}$$

Here,

c = degree of freedom

O = Observed value

E = Expected value.

Result:

Running "task-c.py", we get,

Chi-squared statistic : 36.1111

p-value : 0.867957

As the p-value is greater than the significant level of $\alpha = 0.05$, we "fail to reject the null hypothesis". This indicates that, there is not enough statistical-evidence to say that the gamma distribution doesn't fit the data well for the given bins & sample size.