

iNeuron Assessment (Statistics)

Qn.1) What is the meaning of six sigma in statistics? Give proper example

Ans) In statistics, Six Sigma is a quality management methodology that aims to achieve near-perfect quality by minimizing defects or variations in processes. The term "Six Sigma" refers to the statistical concept of standard deviations from the mean in a normal distribution. Specifically, Six Sigma strives for less than 3.4 defects per million opportunities, indicating a highly efficient and reliable process.

Example:

Definition: Six Sigma is about achieving a level of performance where only 3.4 defects per million opportunities occur, showcasing exceptional quality.

Application: Consider a surgeon with a 5.51 sigma ranking. This means that if they operate on 1,000,000 patients per year, only 30 patients are likely to experience issues

Goal: A process operating at Six Sigma has virtually zero defects, ensuring high-quality outcomes consistently.

By focusing on reducing defects and variations through statistical analysis and process optimization, Six Sigma helps organizations enhance customer satisfaction, reduce costs, and improve overall efficiency.

Qn.2) What type of data does not have a log-normal distribution or a Gaussian distribution? Give proper example

Ans)

Discrete distributions: Data that takes on discrete values and does not follow a continuous distribution. Examples include the Poisson distribution, which models the number of events occurring in a fixed interval of time or space, and the binomial distribution, which models the number of successes in a fixed number of independent Bernoulli trials.

Skewed distributions: Data that is skewed, meaning it is asymmetric and has a longer tail on one side than the other. Examples include the gamma distribution, which is often used to model waiting times, and the exponential distribution, which is often used to model the time between events in a Poisson process.

Mixed distributions: Data that is a mixture of different distributions. For example, a dataset that combines data from two different populations with different underlying distributions.

Some more example:

1. Bacteria Growth: Bacteria growth naturally follows an exponential distribution, which is a type of non-normal distribution
2. Income Distribution: In economics, the income of 97%–99% of the population is distributed log-normally, with higher-income individuals following a Pareto distribution.

Qn.3) What is the meaning of the five-number summary in Statistics? Give proper example

Ans) The five-number summary in statistics is a set of descriptive statistics that provides key information about a dataset. It consists of five values that are presented in the following order:

1. Minimum: The smallest observation in the dataset.
2. First Quartile (Q1): The value below which 25% of the data falls.
3. Median: The middle value that splits the dataset into two halves.
4. Third Quartile (Q3): The value below which 75% of the data falls.
5. Maximum: The largest observation in the dataset.

Example:

Consider a set of observations representing the number of moons for each planet in the Solar System: 0, 0, 1, 2, 63, 61, 27, 13. Arranging these observations in ascending order gives: 0, 0, 1, 2, 13, 27, 61, 63. With eight observations:

The median is $(2 + 13) / 2 = 7.5$.

The first quartile is $(0 + 1) / 2 = 0.5$.

The third quartile is $(27 + 61) / 2 = 44$.

The minimum and maximum are 0 and 63.

Therefore, the five-number summary for this dataset would be: 0, 0.5, 7.5, 44, 63

Qn.4) What is correlation? Give an example with a dataset & graphical representation on jupyter Notebook

Ans) Correlation in statistics refers to the measure of how two variables change together and the strength of their relationship. It helps in understanding patterns, trends, and associations within a dataset, guiding decision-making processes. Visualizing correlation effectively is crucial for data analysis and interpretation.

Example with Dataset & Graphical Representation:

X (Hours of Study) and Y (Exam Scores). We want to explore the correlation between these variables.

Dataset:

X (Hours of Study): 2, 3, 4, 5, 6

Y (Exam Scores): 60, 65, 70, 75, 80

Graphical Representation on Jupyter Notebook:

➤ I am writing code here for visualization on jupyter notebook.

Input:

```
import matplotlib.pyplot as plt

# Dataset
X = [2, 3, 4, 5, 6]
Y = [60, 65, 70, 75, 80]

# Scatter Plot
plt.scatter(X, Y)

plt.xlabel('Hours of Study')
plt.ylabel('Exam Scores')

plt.title('Correlation between Hours of Study and Exam Scores')

plt.show()
```

Output:

