11. How can we calculate the range of the data?

Ans.

The range is calculated by subtracting the lowest value from the highest value. Larger range means high variability. Smaller range means lower variability.

12. Is the range sensitive to outliers?

Ans.

Yes, range is very sensitive to outliers.

13. What is the meaning of standard deviation?

Ans.

Standard deviation is a measure to calculate the dispersion of the data with relation to the mean. If the standard deviation is low then the data are clustered to the mean. If the standard deviation is high then the data are spread away from mean.

14. What are the scenarios where outliers are kept in the data?

Ans.

Scenario-1:

In Credit card fraud detection the outliers are important for the analysis because here the exceptions are more important than the normal cases.

Scenario-2 :

In finding the best students of the class we may create an analysis on IQ of students, the students with more IQ will be considered. That's why there is a change in the linear regression line.

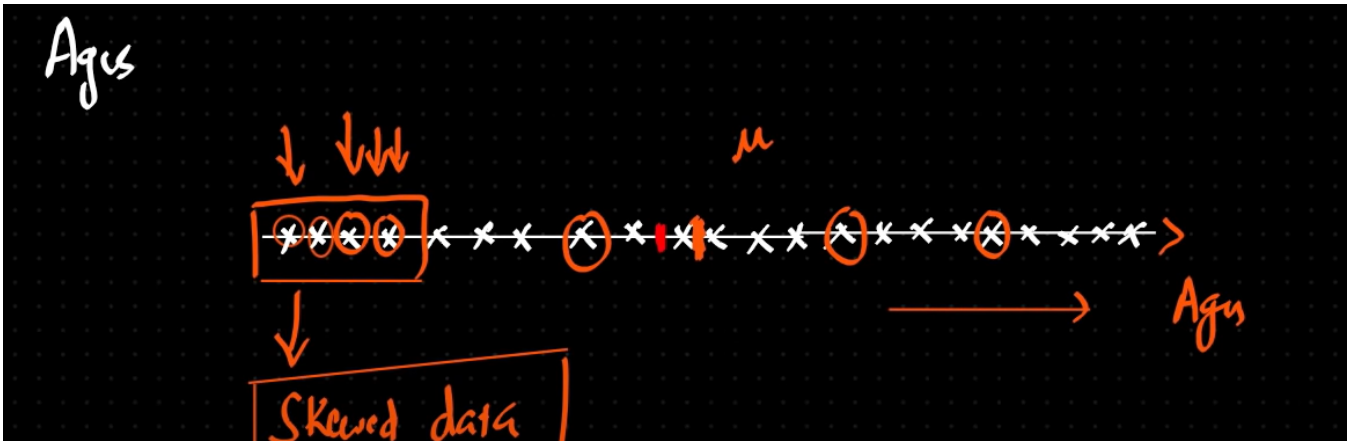
15. What is Bessel's correction?

Ans.

In statistics, Bessel's correction is the use of n-1 instead of n in several formulas, including the sample variance and standard deviation, where n is the number of observations in a sample. This method corrects the bias in the estimation of the population variance. It also partially corrects the bias in the estimation of the population standard deviation, thereby, providing more accurate results.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$



16. What do you understand about a spread out and concentrated curve

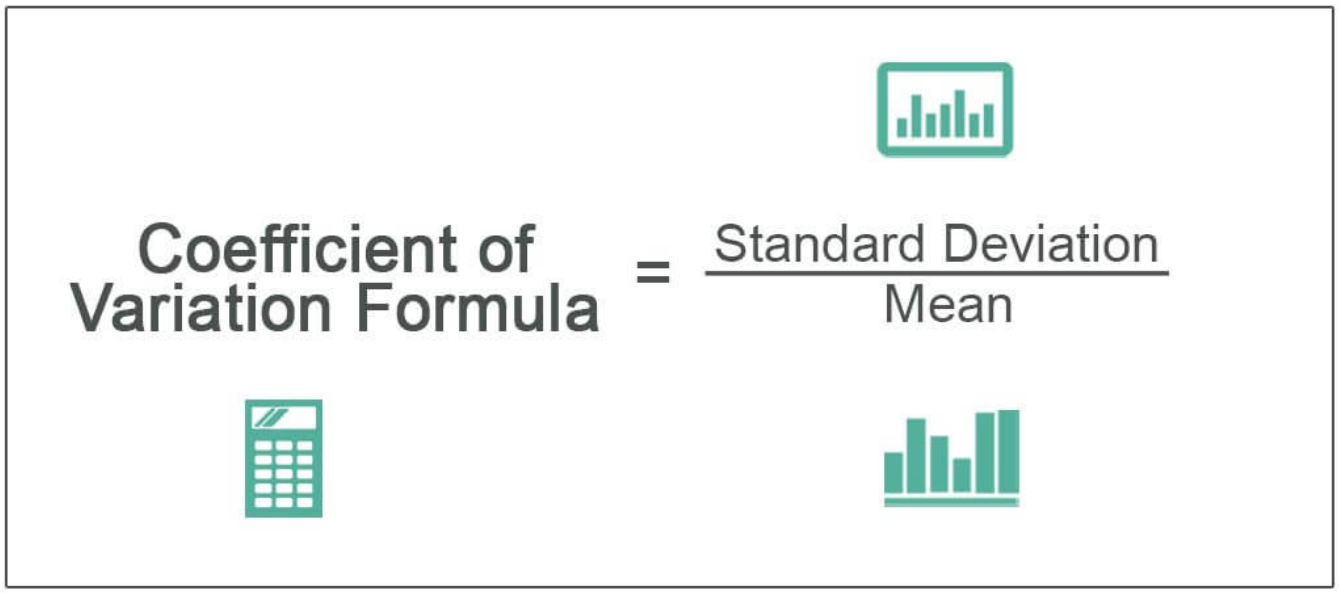
Ans.

Spread out curve shows how far the data ranges from the mean of the centre of the data or how widely spread data are. Concentrated curve shows how closely the data are to the centre of the data.

17. Can you calculate the coefficient of variation?

Ans.

Co-efficient of variation represents the ratio of the standard deviation to the mean. It is a statistic to compare the degree of variation from one data series to another, even if the mean are different from one another.



18. State the case where the median is a better measure when compared to the mean.

Ans.

Median is less affected by the outliers compared to mean. And it is considered in the calculation of measure of central tendency if the dataset are not symmetric.

19. How is missing data handled in statistics?

Ans.

In statistics we handle the missing data by

1. Eliminating the dataset
2. By imputation of mean, median & mode

20. What is meant by mean imputation for missing data? Why is it bad?

Ans.

Mean imputation is where the missing values are imputed by the calculated mean of the dataset. This method can lead to severely biased result if data are MCAR(Missing Completely At Random) and imputation never preserve the relationship among variables.

In [ ]:

```
1
```