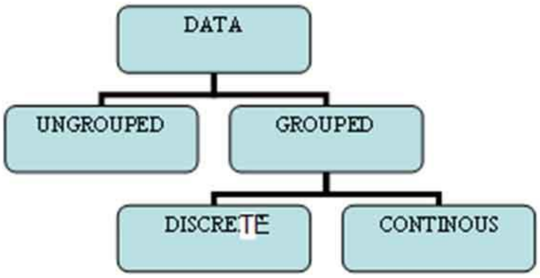
XI: Math
Chapter 15: Statistics

Chapter Notes

Key Concepts

1. Statistics deals with collection presentation, analysis and interpretation of the data.

2. Data can be either ungrouped or grouped. Further, grouped data could be categorized into:
(a) Discrete frequency distribution,
(b) Continuous frequency distribution.



3. Data can be represented in the form of tables or in the form of graphs.

Common graphical forms are: Bar charts,pie diagrams, histograms,

frequency polygons ogives, etc.

4. First order of comparison for the given data is the measures of central tendencies. Commonly used measures are (i) Arithmetic mean (ii) Median(iii) Mode.

5. Arithmetic mean or simply mean is the sum of all observations divided by the number of observations. It cannot be determined graphically. Arithmetic mean is not a suitable measure in case of extreme values in the data.

6. Median is the measure which divides the data in two equal parts. The median is the middle term when the data is sorted.

Incase of odd observations the middle observation is median. In case of even observations the median is the average of the two middle observations.

7. Median can be determined graphically. It does not take into account all the observations.

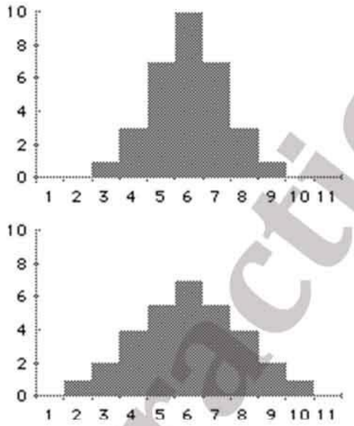
8. The mode is the most frequently occurring observation. For a frequency distribution mode may or may not be defined uniquely.

9. Measures of central tendencies namely mean, median and mode provide us with a single value which is the representative of the entire data. These three measures try to condense the entire data into a single central value

10. Central tendencies indicate the general magnitude of the data.

11. Two frequency distributions may have same central value but still they have different spread or they vary in their variation from central position. So it is important to study how the other observations are scattered around this central position.

12. Two distributions with same mean can have different spread as shown below.



13. Variability or dispersion captures the spread of data. Dispersion helps us to differentiate the data when the measures of central tendency are the same.

14. Like 'measures of central tendency' gives a single value to describe the magnitude of data. **Measures of dispersion** gives a single value to describe variability.

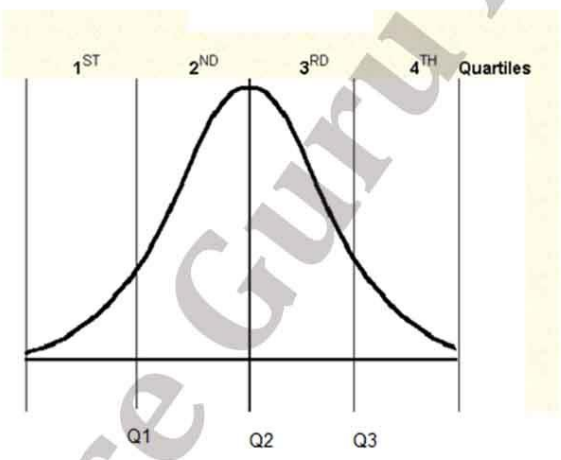
15. The dispersion or scatter of a dataset can be measured from two perspectives:
   (i) Taking the order of the observations into consideration, two measures are
      (a) Range (b) Quartile deviation

(ii)Taking the distance of each observation from the central position, yields two measures, (a) Mean deviation, (b) Variance and Standard deviation

16. **Range** is the difference between the highest and the lowest observation in the given data.
The greater the range is for a data, its observations are far more scattered than the one whose range is smaller.

17. The range at best gives a rough idea of the variability or scatter.

18. Quartile divides the data into 4 parts. There are three quartiles namely Q_1 Q_2 Q_3 and Q_2 is the median only.



19. The quartile deviation is one-half of the difference between the upper quartile and the lower quartile.

20. If x_1, x_2, … x_n are the set of points and point a is the mean of the data. Then the quantity x_i –a is called the deviation of xi from mean a. Then the sum of the deviations from mean is always zero.

21.In order to capture average variation we must get rid of the negative signs of deviations.
There are two remedies

Remedy I: take the Absolute values of the deviations.
Remedy II: take the squares of the deviation.

22. Mean of the absolute deviations about a gives the 'mean deviation about a', where a is the mean. It is denoted as M.D. (a). Therefore, M.D.(a) = Sum of absolute values of deviations from the mean 'a ' divided by the number of observations. Mean deviation can be calculated about median or mode or any other observations.

23. Merits of mean deviation
(1) It utilizes all the observations of the set.
(2) It is least affected by the extreme values.
(3) It is simple to calculate and understand.

24. Mean deviation is the least when calculated about the median.
If the variations between the values is very high, then the median will not be an appropriate central tendency representative.

.

25. **Limitations of Mean Deviation**

i)The foremost weakness of mean deviation is that in its calculations, negative differences are considered positive without any sound reasoning
ii) It is not amenable to algebraic treatment.
(iii) It cannot be calculated in the case of open end(s) classes in the frequency distribution.

26. Measure of variation based on taking the squares of the deviation is called the variance.

27. Let the observations are $x_1, x_2, x_3, .., x_n$
let mean = $\bar{x}$
Squares of deviations: $d_i = (x_i - \bar{x})^2$
Case 1: The sum di is zero. This will imply that all observations are equal to the mean x bar.
Case 2: The sum di is relatively small. This will imply that there is a lower degree of dispersion. And case three
Case 3: The sum di is large. There seems to be a high degree of dispersion.

28. Variance is given by the mean of squared deviations. If variance is small the data points are clustering around mean otherwise they are spread across.

29. Standard deviation is simply expressed as the positive square root of variance of the given data set. Standard deviation of the set of observations

does not change if a non zero constant is added or subtracted from each bservations.
30. Variance takes into account the square of the deviations.
Hence, the unit of variance is in square units of observations.
For standard deviation, its units are the same as that of the observations.
Thats the reason why standard deviation is preferred over variance.

31. Standard deviation can help us compare two sets of observations by describing the variation from the "average" which is the mean.Its widely used in comparing the performance of two data sets. Such as two cricket matches or two stocks.
In Finance it is used to access the risk associated with a particular mutual fund.

32. Merits of Standard deviation
i) It is based on all the observations.
(ii) It is suitable for further mathematical treatments.
(iii) It is less affected by the fluctuations of sampling.

33. A measure of variability which is independent of the units is called as coefficient of variation. Denoted as C.V.
It is given by the ratio of σ the standard deviation and the mean x bar of the data.

34. It is useful for comparing data sets with different units, and wildly varying means. But mean should be non zero. If mean is zero or even if it is close to zero the Coefficient of Variation fails to help.

35. Coefficient of Variation-a dimensionless constant that helps compares the variability of two observations with same or different units.

Key Formulae

1. **Arithmetic mean**
   (a) Raw data

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

   (b) Discrete data

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{1}{N}\sum_{i=1}^{n} x_i f_i$$

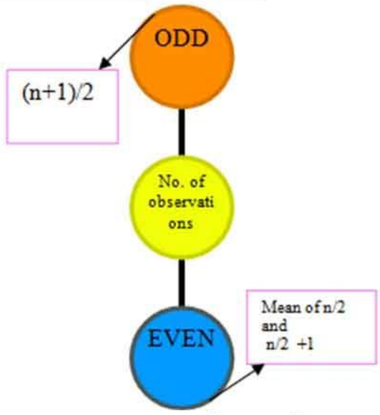
   (c) Step Deviation Method:

$$\bar{x} = a + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times h$$

2. **Median**

(a)

<u>Median Of Ungrouped Data</u>



(b) Median= $l + \left(\dfrac{\frac{n}{2} - cf}{f}\right) \times h$

Where, l= the lower limit of median class.
cf = the cumulative frequency of the class preceding the median class.
f = the frequency of the median class.
h =the class size

3. Mode for a grouped data is given by

Mode= $l + \left(\dfrac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$

l = lower limit of the modal class
h = size o f the class interval
f_1 = frequency of the modal class
f_0 =frequency of the class preceding the modal class
f_2 = frequency of the class succeeding the modal class

3. Mean Deviation about mean $M.D.(\bar{x}) = \dfrac{\sum_{i=1}^{n} f_i \left|x - \right|}{\sum_{i=1}^{n} f_i} = \dfrac{1}{N}\sum_{i=1}^{n} f_i |x_i - \bar{x}|$

4. Mean Deviation about median $M.D.(M) = \dfrac{1}{N}\sum_{i=1}^{n} f_i |x_i - M|$

5. Variance
(a)for ungrouped data

$$\sigma^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

(b) For grouped data

$$\sigma^2 = \frac{\sum_i f_i(x_i - \bar{x})^2}{n}$$

6. Standard Deviations

(a) For ungrouped data $\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

(b) For grouped data $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{2} f_i(x_i - \bar{x})^2}$ where $\bar{x}$ is the mean of the distribution and $N = \sum_{i=1}^{n} f_i$.

(c) Short Cut Method $\sigma = \frac{h}{N}\sqrt{N\sum_{i=1}^{n} f_i y_i^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2}$

7. Coefficient of Variation: $C.V. = \frac{\sigma}{\bar{x}} \times 100, \bar{x} \neq 0,$