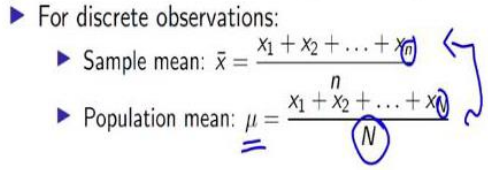
1. As far as the numbers of observations are limited and discrete, you can treat each number as a category to construct the frequency table. In order to graph a frequency table, use bar charts.
2. If the number of observations is more than manageable and continuous, create classes of data (more than 5, but no more than 20), and construct histograms. The class width and the class mark (average of class min and max) is selected as appropriate.
3. Each observation should belong to some class and no observation should belong to more than one class.
4. It's common, but not essential, to choose equal width for all classes.
5. Class interval starts from its left boundary point, and up to, but not including the right boundary point.
6. IN a stem-leaf diagram, each observation is split into two parts, namely a stem consisting of all but the right most digit and a leaf which has the right most digit.
7. In a stem-left diagram, the stems are sorted in ascending order and written in the vertical column. Against each stem, the leaves are sorted in ascending order.
8. Most common descriptive measures of data are Measures of central tendency, and Measures of dispersion.
9. Measures of central tendency include mean, median and mode.
10. Mean is measured by summing up all data points, and dividing by the number of observations. In the case of a sample, number of observations is denoted by ***n*** and in the case of population, number of observations is denoted by ***N***




11. The mean is very sensitive to outliers.
12. Mean is represented as $\sum_{i=1}^{n} f_i x_i$. In the case of continuous observations, xi represents the class-mark and hence an approximation of the actual value.
13. When a constant value is added to each observed value, the mean increases by the same constant value. Thus,

    Let $y_i = x_i + c$ where $c$ is a constant then $\bar{y} = \bar{x} + c$
14. When each observed value is multiplied by a constant value, the mean also gets multiplied by the same constant value. Thus,

    Let $y_i = x_i c$ where $c$ is a constant then $\bar{y} = \bar{x}c$
15. Median of a dataset is the number that divides the bottom 50% with the top 50%. To calculate it, the data set must be ordered first.
16. In order to obtain median of a data-set, arrange all observations in ascending order. Let n be the number of observations. If n is odd, median is the observation numbered (n+1)/2. If n is even, median is the average of observations marked n/2 and (n/2+1).
17. Median is not very sensitive to the outliers, as much mean is.
18. Median need not belong to the data-set. It's a computed value, if you've even number of observations.
19. When a constant value is added to each observed value, the median increases by the same constant value.
20. When each observed value is multiplied by a constant value, the median also gets multiplied by the same constant value.

21. Mode is the frequently occurring value in a data-set. If no value occurs more than once, there's no mode for the data-set.
22. Mode is not affected by the outliers in the data.
23. When a constant value is added to each observed value, the mode increases by the same constant value.
24. When each observed value is multiplied by a constant value, the mode also gets multiplied by the same constant value.
25. Having the same mean, median and mode for two data sets do not guarantee that they are similar. Their dispersions (range, variance, std. deviation) could be different.
26. Range of a dataset is the difference between the maximum and minimum values in it.
27. Range is very sensitive to outliers.
28. Variance measures the deviations of each observation from a central value (typically mean of the dataset)
29. Std. deviation is the square root of variance, in the case of sample and population.
30. Std. deviation is measured in the same units as the original data.
31. Variance of a population and a sample are given by the formulae. This formulae are referred to as the defining formulae.
    - The variance is computed using the following formulae
      - Population variance: $\sigma^2 = \frac{(x_1-\mu)^2+(x_2-\mu)^2+\ldots+(x_N-\mu)^2}{N}$
      - Sample variance: $s^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n-1}$

    Note that in the case of sample variance, the denominator is one less than the total numbers of observations.
32. To find population std. deviation, use the following simpler formula (called computing formula).
    $$\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

If A is the assumed mean, and d = x – A for every x, above equation for population std. deviation can also be written as $\sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$

33. To find sample variance, use the following simpler formula (called computing formula).
    $$s^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}$$

If A is the assumed mean, and d = x – A for every x, above equation for population variance can also be represented using an equation similar to above. Replace X with d.

34. When a constant value is added to each observation in sample/population, the variance doesn't change.
35. When each observed value in the sample/population is multiplied with a constant value, the variance gets multiplied by the square of the same constant value. Thus,
    - Let $y_i = x_i c$ where $c$ is a constant then

      $$new\ variance = c^2 \times old\ variance$$

36. When a constant value is added to each observation in sample/population, the std. deviation doesn't change.
37. When each observed value in the sample/population is multiplied with a constant value, the std. deviation gets multiplied by the same constant value.
38. Variance and std. deviation are also very sensitive to outliers.
39. 100p percentile of a dataset is that datapoint that has the property that at least 100p percent of the data are less than or equal to it and at least 100(1-p) percent of the data are greater than or equal to it. If two datapoints satisfy this condition, then 100p percentile is given by the arithmetic average of these values.
40. Median is the $50^{th}$ percentile.
41. algorithm to compute percentile of a dataset.
    a. Arrange the data in increasing order
    b. If np is not an integer, determine the smallest integer greater than np. The data value in that position is the sample 100p percentile.
    c. If np is an integer, then the average of the value at the positions *np* and *(np + 1)* is the sample *100p* percentile.
42. $25^{th}$ percentile is called first quartile. $50^{th}$ percentile is called median or second quartile. $75^{th}$ percentile is called third quartile.
43. Quartiles breakup a dataset into 4 parts, with 25% of data values less than Q1, 25% between Q1 and median. 25% between median and Q3, and 25% greater than Q3.
44. Five number summary consists of the minimum, Q1, Q2(median), Q3, maximum of the dataset.
45. inter-quartile range (IQR) = Q3 – Q1.
46. IQR is a very important measure of dispersion of a dataset.