Descriptive Statistics

Measures of Central Tendency

1) **What is Descriptive Statistics?**
   The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.
2) **What is measures of central tendency?**
   A single value that describes the characteristics of the entire mass of unwieldy data. Such a value is called the central value.
3) **What is mean?**
   Mean is a sum of observations divided by number of observations.
4) **What is Median?**
   Median is the value of middle most observations in the data when the observations are arranged in increasing or decreasing order of their values.
5) **What is Mode?**
   The observation with maximum frequency or the most repeated observation is called as mode.
6) **What is Weighted Mean?**
   A mean where some values contribute more than others. Weighted means can help with decisions where some things are more important than others.
7) **What is Geometric Mean?**
   Geometric mean is a special type of average where we multiply the numbers together and then take a square root (for two numbers), cube root (for three numbers) etc. Geometric mean is useful when we want to compare things with very different properties.
8) **What is Harmonic Mean?**
   The reciprocal of the arithmetic mean of the reciprocals.

Measures of Dispersion

1) **What is Measures of Dispersion or Variation?**
   Observations are scattered or dispersed from central value. This variation is called as dispersion.
2) **What is Range?**
   The difference between the largest and smallest data values in a set of values for a variables.
3) **What is Quartile?**
   The quartile which divide the total number of observations into 4 equal parts are called quartiles.
4) **What is Interquartile Range (IQR)?**
   IQR is the difference in value between the upper quartile and lower quartile.
5) **What is Variance?**
   A measurement of the spread between numbers in a dataset.
6) **What is Standard Deviation?**
   Standard Deviation is a number used to tell how measurements for a group are spread out from the average. A low Standard Deviation means that most of the numbers are close to the average. A high standard Deviation means that the numbers are more spread out.
7) **What is Standard (z) score?**
   Z-score is a measure of how many standard deviations below or above the population mean.

Measures of Shape Distributions

1) **What is Measures of Shape Distributions?**
   Measures of shape describe the distribution or pattern of the data within a dataset.
2) **What is Skewness?**
   Skewness is asymmetry in a statistical distribution, in which the curve appears skewed either to the left or to the right. -1 to 1 is acceptable
3) **What is Symmetrical Shape?**
   A symmetrical shape is a set of data values in which the mean equals the median value. (E.g. actual amount of soft drink in a one-liter bottle.)
4) **What is Left-Skewed Shape?**
   Left-Skewed shape is a set of data values in which the mean is less than the median value. Also known as negative skew. (E.g. scores on an exam in which most student score between 80 and 100, whereas a few students score between 10 and 70.)
5) **What is Right-Skewed Shape?**
   Right-Skewed shape is a set of data values in which the mean is greater than the median value. Also known as positive skew. (E.g. Prices of new homes, annual family income.)
6) **What is Kurtosis?**
   Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. -2 to +2 acceptable
7) **What is MesoKurtic? (middle)**
   MesoKurtic indicate an excess kurtosis of zero or close to zero. It means that if the data follows a normal distribution, it follow a MesoKurtic distribution.
8) **What is Leptokurtic? (narrow)**
   Leptokurtic indicates a positive excess kurtosis. The leptokurtic shows heavy tails on either side, indicating the large outlier in a distribution.
9) **What is Platykurtic? (like a plate)**
   A Platykurtic indicate a negative excess kurtosis. The Platykurtic shows flat tails, the flat tails indicate the small outliers in a distribution.
10) **What is the solution to solve the skewness and kurtosis?**
    - Using log10 we can normalize data.
    - Using LN or natural logarithm
    - Box-Cox transformation. $(y^{\lambda}-1)/\lambda$
    - Principal Component Analysis.
    - Min-max
    - Z-score
    - Decimal scaling