SURVEY ON STATISTICAL TOOLS

Prof. S. N. Gorde[1], Prof. M. K. Kodmelwar[2]
[1,2]TSSM's, Bhivarabai Sawant College of Engineering & Research, Narhe, Pune-41.

Abstract
Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data. In applying statistics to, e.g., a scientific, industrial, or social problem, it is conventional to begin with a statistical population or a statistical model process to be studied. Populations can be diverse topics such as "all people living in a country" or "every atom composing a crystal". Statistics deals with all aspects of data including the planning of data collection in terms of the design of surveys and experiments. See glossary of probability and statistics
Keywords- Mean deviation, standard deviation, coefficient of skewness, central tendency

I. INTRODUCTION

When a census is not feasible, a chosen subset of the population called a sample is studied. Once a sample that is representative of the population is determined, data is collected for the sample members in an observational or experimental setting. Again, descriptive statistics can be used to summarize the sample data. However, the drawing of the sample has been subject to an element of randomness, hence the established numerical descriptors from the sample are also due to uncertainty. To still draw meaningful conclusions about the entire population, **inferential statistics** is needed. It uses patterns in the sample data to draw inferences about the population represented, accounting for randomness. these inferences may take the form of: answering yes/no questions about the data (hypothesis testing), estimating numerical characteristics of the data (estimation), describing associations within the data (correlation) and modeling relationships within the data (for example, using regression analysis). inference can extend to forecasting, prediction and estimation of unobserved values either in or associated with the population being studied; it can include extrapolation and interpolation of time series or spatial data, and can also include data mining.[1][2][3]

II. STATISTIC TOOLS

Following terms are in the statistic

A. Mean deviation- The average can be mean, median or mode. Theoretically median is d best average of choice because sum of deviations from median is minimum, provided signs are ignored. However, practically speaking, arithmetic mean is the most commonly used average for calculating mean deviation and is denoted by the symbol MD.

The mean difference (more correctly, 'difference in means') is a standard statistic that measures the absolute difference between the mean value in two groups in a clinical trial. It estimates the amount by which the experimental intervention changes the outcome on average compared with the control.

$$\mathrm{MD} = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|,$$

B. Standard Deviation **:-** In statistics, the standard deviation (SD, also represented by the Greek letter sigma σ or the Latin letter s) is a measure that is used to quantify the amount of variation or dispersion of a set of data values A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high

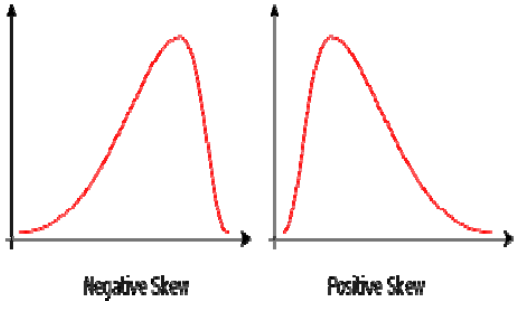
standard deviation indicates that the data points are spread out over a wider range of values.

The standard deviation of a random variable, statistical population, data set, or probability distribution is the square root of its variance. It is algebraically simpler, though in practice less robust, than the average absolute deviation. A useful property of the standard deviation is that, unlike the variance, it is expressed in the same units as the data. There are also other measures of deviation from the norm, including average absolute deviation, which provide different mathematical properties from standard deviation

$$\sigma = \sqrt{\frac{\sum f \cdot (x_i - \mu)^2}{\sum f}}$$

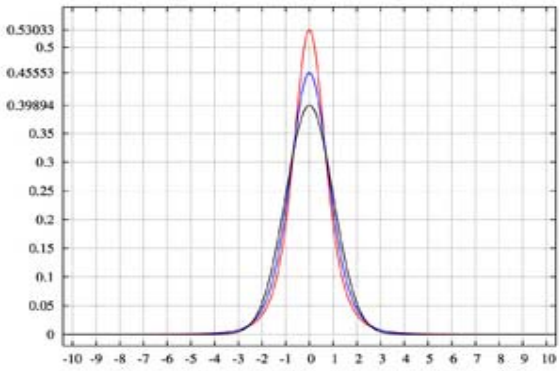
C. Coefficient of skewness :- In probability theory and statistics, **skewness** is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or undefined.[2][3]

The qualitative interpretation of the skew is complicated and unintuitive. Skew does not refer to the direction the curve appears to be leaning; in fact, the opposite is true. For a unimodal distribution, negative skew indicates that the tail on the left side of the probability density function is longer or fatter than the right side – it does not distinguish these two kinds of shape. Conversely, positive skew indicates that the tail on the right side is longer or fatter than the left side. In cases where one tail is long but the other tail is fat, skewness does not obey a simple rule.



D. Coefficient of Kurtoisis:- The **coefficient of kurtosis**, or simply kurtosis, **measures the peakedness of a distribution**. High kurtosis means that values close to the mean are relatively more frequent and extreme values (very far from the mean) are also relatively more frequent. The values in between are relatively less frequent. If you plot a frequency histogram or another chart showing frequency of such distribution, it would have a sharp peak in the middle and fat tails.[1][4]

The kurtosis of any univariate normal distribution is 3. It is common to compare the kurtosis of a distribution to this value. Distributions with kurtosis less than 3 are said to be platykurtic, although this does not imply the distribution is "flat-topped" as sometimes reported. Rather, it means the distribution produces fewer and less extreme outliers than does the normal distribution. An example of a platykurtic distribution is the uniform distribution, which does not produce outliers. Distributions with kurtosis greater than 3 are said to be leptokurtic



III Conclusion

The statistic tools are used for various calculations in the field of engineering science. The paper give idea about the Mean deviation, standard deviation , coefficient of skewness, central tendency for undewrsting & use .

REFERENCES

1. https://en.wikipedia.org/wiki/Kurtosis
2. http://www.macroption.com/coefficient-of-kurtosis/
3. EisenhartC:Expression of the uncertain-Ties of final results. Science968;160:1201-1204
4. Glantz SA: Primer of Biostatistics.New York, McGraw-Hill BookCo, 1981
5. http://www.macroption.com/coefficient-of-kurtosis/
6. http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm