Principles Methods and Applications of Statistics

Posted on May 11, 2020

Statistics is a branch of mathematics that deals with the analysis done with the help of models and representations of any provided data set or any real-life data. It deals with the methods and ways to collect and analyze data, apply a suitable test or model to the data, and extract information by plotting the results. Some of the measures that are used in the subject are mean, variance, regression analysis, skewness, and kurtosis.

Statistics is also described as a process through which a data set is characterized. The dependence of a data set on a small population indicates that interpretations can be developed regarding the population based on the statistical results from the sample. The process comprises collecting the data, evaluating it, and then putting it into a mathematical form.

The applications of statistics are wide-ranging. It is used in areas such as psychology, all kinds of sciences, businesses, manufacturing, and political studies.

There are two types of methods that are utilized by the analysts to analyze the data: descriptive and inferential.

Descriptive Statistics

As the term indicates, this method of analysis is used to describe, display, or summarize the data in a meaningful manner. This technique does not permit us to draw out results from the data or conclude anything from the hypothesis we've made.

Data, if especially large, is difficult to handle in its raw form. Therefore, descriptive statistics help us to describe it appropriately. For instance, if a teacher has 100 results of students' exams, and she wants to know the overall performance, then descriptive statistics will help in doing this. Usually, there are two common kinds of statistics that define the data ("Descriptive Statistics and Interpreting", n.d.):

Central Tendency Measures:

Central tendency measures are the methods used to describe the central position of a frequency distribution for a data set. In this instance, the frequency distribution is merely the distribution and outline of marks gained by the hundred pupils from the bottom to the top. Central position can be described using various statistics, such as mode, median, and mean.

Measures Of Spread:

To summarize the data set, these techniques are used to describe how distributed the results are. For instance, if the mean score of these hundred students is 65 out of 100, it does not imply that all pupils have gained a score of 65. Instead, these marks will be distributed around 65. Some will be less than 65 and others greater. For the description of this distribution or spread, many statistics are available, such as the variance, range, absolute deviation, quartiles, and standard deviation.

While using descriptive statistics, analysts make use of tabulations, using graphs, charts, and tables and then give the statistical summary of the results obtained (Staff, 2018).

Inferential Statistics

We observed that descriptive statistics give us information regarding the data set under consideration. For instance, we could perform a calculation of the standard deviation and mean of the exam scores of a hundred students, and this could give us important knowledge regarding these 100 students. Any group of data under consideration is known as the population. It can be small or large. For example, in the case of 100 students, these 100 pupils will form our population because we are interested in their scores. The application of descriptive statistics rests on the population and its characteristics, such as the mean or standard deviation. These characteristics are known as parameters, which are used to represent the whole population ("Descriptive Statistics and Interpreting", n.d.).

There are situations where we are unable to reach the whole population; for example, if we want to analyze the scores of all students in the UK, it is not possible to investigate the whole population. In such a case, we take a small portion known as the sample from the population, which then represents the whole population. Characteristics of the sample, such as the mean or standard deviation, are known as statistics. The methods that permit us to make use of the samples so that we can make interpretations regarding the population from which the samples are taken are called inferential statistics. Hence, it's important for a sample to accurately represent the population. Sampling is the process by which we can achieve this. Inferential statistics comprises the methods of (1) parameter estimation and (2) testing of statistical hypotheses.

Testing Of Hypotheses

In statistics, the analyst tests an assumption about the parameter of the population. The technique used is dependent on what kind of data set is under consideration and why there is a need to analyze it. The method of hypothesis testing deduces the outcome of a hypothesis based on sample data obtained from a bigger population.

In this process, a statistical sample is tested, with the objective of accepting or rejecting a null hypothesis. It informs the analyst about the validity of the main hypothesis. If it is false, a new hypothesis is formulated and tested, iterating the procedure until the data support a hypothesis.

Testing a statistical hypothesis is done by assessing and investigating a random sample of the population under consideration. This random population sample is used to test 2 different hypotheses, which are the null hypothesis and the alternative hypothesis. The null hypothesis is defined as the hypothesis that is believed to be true. On the other hand, the alternative hypothesis is taken to be untrue.

Stages Of Testing The Hypothesis

Testing of the hypothesis is done using a process that comprises four steps. First, the analyst states the two hypotheses for testing. Formulation of the analysis plan is the second step which tells about the evaluation of the data set. The plan is then carried out and analyzed. Lastly, the results are analyzed, and we either accept or reject the null hypothesis.

The p-value:

The p-value shows the chance of the happening of a given event. This value is taken as a substitute for the points of rejection to make the level of significance smallest at which the rejection of the null hypothesis would take place. The small value of the p-value shows that there isn't any strong proof in favor of the alternative hypothesis.

Selection Of Appropriate Statistical Test

There are various tests used according to the nature of the data set.

The t-Test

This test determines whether the means of two groups are statistically the same or whether there is a noteworthy difference between them.

Wilcoxon-Mann-Whitney Test

In cases where the t-test cannot be applied, the Wilcoxon-Mann-Whitney Test is used, which makes a comparison of the medians instead of the means of the two sets.

Variance Analysis (ANOVA)

It makes a comparison of the means of more than two sets to see if there's any statistically important variation amid them.

Chi-Squared Test

In the case of the categorical variables, the chi-squared test is used. It checks whether the occurrence of a specific categorical variable is correlated with a specific dataset.

Apart from these, there are many other tests that are applied in different situations.

Evaluating Results

After the process of analyzing and testing, appropriate conclusions are drawn from the data set. Seeking the most suitable outcome can be a difficult task as the answers may not be as you require. The following are some points to reflect on while seeking the results:

• The main points of the analysis

• Any other interpretations that can be drawn

• can these outcomes be supported statistically?

• Understanding of the results

• Differences in the results from expectations

References

Creswell, J. W. (1994). Research design: Qualitative & quantitative approaches. Thousand Oaks, CA, US: Sage Publications, Inc.

Staff, I. (2015, October 01). Descriptive Statistics. Retrieved February 16, 2018, from https://www.investopedia.com/terms/d/descriptive_statistics.asp