NAME : Patricia Stanley

STATISTICS ASSIGNMENT - Inferential Statistics

Inferential Statistics

Inferential statistics is a branch of statistics that makes the use of various analytical tools to draw inferences about the population data from sample data. Apart from inferential statistics, descriptive statistics forms another branch of statistics. Inferential statistics help to draw conclusions about the population while descriptive statistics summarizes the features of the data set.

There are two main types of inferential statistics - hypothesis testing and regression analysis.

The samples chosen in inferential statistics need to be representative of the entire population.

What is Inferential Statistics?

Inferential statistics helps to develop a good understanding of the population data by analyzing the samples obtained from it. It helps in making generalizations about the population by using various analytical tests and tools. In order to pick out random samples that will represent the population accurately many sampling techniques are used. Some of the important methods are simple random sampling, stratified sampling, cluster sampling, and systematic sampling techniques.

Inferential Statistics Definition

Inferential statistics can be defined as a field of statistics that uses analytical tools for drawing conclusions about a population by examining random samples. The goal of inferential statistics is to make generalizations about a population. In inferential statistics, a statistic is taken from the sample data (e.g., the sample mean) that used to make inferences about the population parameter (e.g., the population mean).

Types of Inferential Statistics

Inferential statistics can be classified into hypothesis testing and regression analysis. Hypothesis testing also includes the use of confidence intervals to test the parameters of a population. Given below are the different types of inferential statistics.

Inferential Statistics	
Hypothesis Testing	Regression Analysis
Z test	Linear Regression
F test	Nominal Regression
T test	Logistic Regression
ANOVA Test	Ordinal Regression
Wilcoxon Signed Rank Test	
Mann-Whitney U Test	

Hypothesis Testing

Hypothesis testing is a type of inferential statistics that is used to test assumptions and draw conclusions about the population from the available sample data. It involves setting up a null hypothesis and an alternative hypothesis followed by conducting a statistical test of significance. A conclusion is drawn based on the value of the test statistic, the critical value, and the confidence intervals. A hypothesis test can be left-tailed, right-tailed, and two-tailed. Given below are certain important hypothesis tests that are used in inferential statistics.

Z Test: A z test is used on data that follows a normal distribution and has a sample size greater than or equal to 30. It is used to test if the means of the sample and population are equal when the population variance is known. The right tailed hypothesis can be set up as follows:

Null Hypothesis: $H_O : \mu = \mu_O$

Alternate Hypothesis: $H_1 : \mu > \mu_O$

Test Statistic: $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$. $\bar{x}$ is the sample mean, μ is the population mean

σ is the population standard deviation and n is the sample size.

Decision Criteria: If the z statistic > z critical value then reject the null hypothesis.

T Test: A t test is used when the data follows a student t distribution and the sample size is lesser than 30. It is used to compare the sample and population mean when the population variance is unknown. The hypothesis test for inferential statistics is given as follows:

Null Hypothesis: $H_O : \mu = \mu_O$

Alternate Hypothesis: $H_1 : \mu > \mu_O$

Test Statistics: $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

Decision Criteria: If the t statistic > t critical value then reject the null hypothesis.

F Test: An f test is used to check if there is a difference between the variances of two samples or populations. The right tailed f hypothesis test can be set up as follows:

Null Hypothesis: $H_O : \sigma_1^2 = \sigma_2^2$

Alternate Hypothesis: $H_1 : \sigma_1^2 > \sigma_2^2$

Test Statistic: $f = \frac{\sigma_1^2}{\sigma_2^2}$, where σ_1^2 is the variance of the first population and σ_2^2 is the variance of the second population.

Decision Criteria: If the f test statistic > f test critical value then reject the null hypothesis.

Confidence Interval: A confidence interval helps in estimating the parameters of a population. For example, a 95% confidence interval indicates that if a test is conducted 100 times with new samples under the same conditions then the estimate can be expected to lie within the given interval 95 times. Furthermore, a confidence interval is also useful in calculating the critical value in hypothesis testing.

Apart from these tests, other tests used in inferential statistics are the ANOVA test, Wilcoxon signed-rank test, Mann-Whitney U test, Kruskal-Wallis H test, etc.

Regression Analysis

Regression analysis is used to quantify how one variable will change with respect to another variable. There are many types of regressions available such as simple linear, multiple linear, nominal, logistic, and ordinal regression. The most commonly used regression in inferential statistics is linear regression. Linear regression checks the effect of a unit change of the independent variable in the dependent variable. Some important formulas used in inferential statistics for regression analysis are as follows:

Regression Coefficients:

The straight line equation is given as $y = \alpha + \beta x$, where α and β are regression coefficients.

$$\beta = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_1^n (x_i - \bar{x})^2}$$

$$\beta = r_{xy} \frac{\sigma_y}{\sigma_x}$$

$$\alpha = \bar{y} - \beta \bar{x}$$

Here, $\bar{x}$ is the mean, and σ_x is the standard deviation of the first data set. Similarly, $\bar{y}$ is the mean, and σ_y is the standard deviation of the second data set.

Inferential Statistics Examples

Inferential statistics is very useful and cost-effective as it can make inferences about the population without collecting the complete data. Some inferential statistics examples are given below:

- Suppose the mean marks of 100 students in a particular country are known. Using this sample information the mean marks of students in the country can be approximated using inferential statistics.
- Suppose a coach wants to find out how many average cartwheels sophomores at his college can do without stopping. A sample of a few students will be asked to perform cartwheels and the average will be calculated. Inferential statistics will use this data to make a conclusion regarding how many cartwheel sophomores can perform on average.

Inferential Statistics vs Descriptive Statistics

Descriptive and inferential statistics are used to describe data and make generalizations about the population from samples. The table given below lists the differences between inferential statistics and descriptive statistics.

Inferential Statistics	Descriptive Statistics
Inferential statistics are used to make conclusions about the population by using analytical tools on the sample data.	Descriptive statistics are used to quantify the characteristics of the data.
Hypothesis testing and regression analysis are the analytical tools used.	Measures of central tendency and measures of dispersion are the important tools used.
It is used to make inferences about an unknown population	It is used to describe the characteristics of a known sample or population.
Measures of inferential statistics are t-test, z test, linear regression, etc.	Measures of descriptive statistics are variance, range, mean, median, etc.

SUMMARY:

- Inferential statistics makes use of analytical tools to draw statistical conclusions regarding the population data from a sample.
- Hypothesis testing and regression analysis are the types of inferential statistics.
- Sampling techniques are used in inferential statistics to determine representative samples of the entire population.
- Z test, t-test, linear regression are the analytical tools used in inferential statistics.