**INTRODUCTION TO STATISTICS**

**1.Difference Between Inferential Statistics and Descriptive Statistics:**

|  |  |
| --- | --- |
| **INFERENTIAL STATISTICS** | **DESCRIPTIVE STATISTICS** |
| 1. It makes use of analytical tools to draw conclusion about the population data from the sample data obtained from it. 2. **Example**: the mean mark of 100 students in a country are known then the mean mark of students in a country can be approximated 3. **Hypothesis testing** and **regression analysis** are the analytical tools used. 4. It is used to make inference about an unknown population. | 1.It summarizes the features of the dataset.  2.**Example**: in a movie theatre, how many of them liked the movie and not liked and then put the findings on a pie chart.  3.**Measures of central tendency** and **measures of dispersion** are the tools used.  4.It is used to find the characteristics of a known sample or population. |

**2 Differentiate Population and Sample:**

|  |  |
| --- | --- |
| **POPULATION** | **SAMPLE** |
| * Population is defined as the entire dataset. * It can be group of individuals, events, objects, etc., * Example: all the residents in a country is a population set. * The characteristics of a population is **parameter.** | * Sample is the unbiased subset of population that best represents the whole data. * It is a subset of larger population that contains the characteristics of that population. * Example: the residents who live above the poverty line is a sample. * The characteristics of a sample is **statistics.** |

**3. Define Hypothesis and Differentiate Null Hypothesis and Alternative Hypothesis:**

* Hypothesis is a **testable statement that relates two or more variables**.
* **Example:** a teacher assumes 60% of the college students are from lower middle-class families.

|  |  |
| --- | --- |
| **NULL HYPOTHESIS** | **ALTERNATIVE HYPOTHESIS** |
| * It states that there is no difference between the characteristics of a population. * **H0** is the symbol used here. * **Example:** the sanitizer manufacturer claims that its product kills 95% of germs on average. | * It defines there is a statistically important relationship between two variables. * **H1** or **Ha** is the symbol used here. * **Example**: if the sanitizer kills less than 95% of germs is alternative hypothesis. |

**4.Central Limit Theorem**:

Central limit theorem is a statistical theory which states that a random sample of size n has a finite variance, the mean of samples will be normally distributed and the mean of samples would be approximately equal to the mean of whole population.

**5.Differentiate Type1 Error and Type2 Error:**

|  |  |
| --- | --- |
| **TYPE1 ERROR** | **TYPE2 ERROR** |
| * It occurs when the null hypothesis of an event or experiment is true, but still, it is rejected. * It is called **false positive** (i.e., an event shows that a condition is present when it is absent). * **Example**: authentication users are considered as hackers. | * It occurs when null hypothesis is false but fails to be refused. * It is called **false negative** (i.e., an event that is real which is rejected and observed as a miss). * Example: hackers are considered as authentication users. |

**6.LINEAR REGRESSION:**

Linear regression is used to predict the value of a variable based on the value of another variable.

The variable you want to predict is called dependent variable and the one you are using to predict the value of another variable is called independent variable.

**Example:** the height of the person linearly related to the weight of the person.

**7**. **ASSUMPTIONS REQUIRED FOR LINEAR REGRESSION**:

* Linear relationship
* Multivariate normally
* No or little multicollinearity
* No auto-correlation
* Homoscedasticity

**8.** **HOW IS THE STATISTICAL SIGNIFICANCE OF AN INSIGHT ASSESSED?**

To assess the statistical significance, use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Then calculating the p-value which is likelihood of getting the test’s observed findings if the null hypothesis is true. Then selecting the threshold of significance (alpha) and rejecting the null hypothesis if the p-value is smaller than alpha value.

**9.** **MEAN:**

Mean is defined as sum of all observations divided by total number of observations.

For example: if the data is 1, 2, 4, 5 and the mean is 3.

**10.STANDARD DEVIATION:**

Standard deviation is the measure of amount of variation or the dispersion of a set of values.

**11.** **CORRELATION:**

It is a statistical method to assess a possible linear association between two continuous variables.

**12.** **COVARIANCE**:

It is a measure of the relationship between two random variables and to extend, they change together.

**13.WHERE IS INFERENTIAL STATISTICS USED?**

Inferential statistics is used **for comparing the parameters of two or more samples and makes generalizations about the larger population based on these samples.**

**14 WHAT IS ONE SAMPLE T TEST?**

The one-sample t-test is used **when we want to know whether our sample comes from a particular population but we do not have full population information available to us**.

For instance, we may want to know if a particular sample of college students is similar to or different from college students in general.

**15. WHAT IS THE RELATIONSHIP BETWEEN STANDARD VARIANCE AND STANDARD DEVIATION?**

Variance is the square of standard deviation.

**16. WHAT IS ONE WAY ANOVA TEST?**

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

One-way ANOVA Null Hypothesis