ASSIGNMENT 2

NEETHU RAJ S

1. What is the difference between inferential statistics and descriptive statistics?

Descriptive statistics is a term given to the analysis of data that helps to describe, show and summarize data in a meaningful way. It is a simple way to describe our data. Descriptive statistics is very important to present our raw data ineffective/meaningful way using numerical calculations or graphs or tables. This type of statistics is applied on already known data.

In inferential statistics predictions are made by taking any group of data in which you are interested. It can be defined as a random sample of data taken from a population to describe and make inference about the population. Any group of data which includes all the data you are interested is known as population. It basically allows you to make predictions by taking a small sample instead of working on whole population.

|  |  |  |
| --- | --- | --- |
| S.No. | Descriptive Statistics | Inferential Statistics |
| 1. | It gives information about raw data which describes the data in some manner. | It makes inference about population using data drawn from the population. |
| 2. | It helps in organizing, analyzing and to present data in a meaningful manner. | It allows us to compare data, make hypothesis and predictions. |
| 3. | It is used to describe a situation. | It is used to explain the chance of occurrence of an event. |
| 4. | It explain already known data and limited to a sample or population having small size. | It attempts to reach the conclusion about the population. |
| 5. | It can be achieved with the help of charts, graphs, tables etc. | It can be achieved by probability. |

2. What is the difference between population and sample in inferential statistics?

A population is the entire group that you want to draw conclusions about. A sample is the specific group that you will collect data from. The size of the sample is always less than the total size of the population.

Population is a complete set and sample is a subset of population. ie Population contains all members of a specified group where as a sample is a subset of population which represents the entire population. In population reports are a true representation of opinion where as in case of sample reports have a margin of error and confidence interval.

3. Most common characteristics used in descriptive statistics?

Descriptive statistics summarizes or describes the characteristics of a data set. Descriptive statistics consists of two basic categories of measures: measures of central tendency and measures of variability (or spread). Measures of central tendency describe the center of a data set. Measures of variability or spread describe the dispersion of data within the set.

Central Tendency: Measures of central tendency focus on the average or middle values of data sets, whereas measures of variability focus on the dispersion of data. These two measures use graphs, tables and general discussions to help people understand the meaning of the analyzed data.

Measures of central tendency describe the center position of a distribution for a data set. A person analyzes the frequency of each data point in the distribution and describes it using the mean, median, or mode, which measures the most common patterns of the analyzed data set.

### Measures of Variability: Measures of variability (or the measures of spread) aid in analyzing how dispersed the distribution is for a set of data. For example, while the measures of central tendency may give a person the average of a data set, it does not describe how the data is distributed within the set. Variance and standard deviation comes under this.

4. How to calculate range and interquartile range?

Range is the difference between the highest and lowest values whereas Interquartile range is the range of the middle half of a distribution. While the [range](https://www.scribbr.com/statistics/range/) gives you the spread of the whole data set, the [interquartile range](https://www.scribbr.com/statistics/interquartile-range/) gives you the spread of the middle half of a data set. The range is the simple measurement of the difference between values in a dataset. To find the range, simply subtract the lowest value from the greatest value, ignoring the others.

Steps to find the interquartile range

1. Order the data from least to greatest.
2. Find the median.
3. Calculate the median of both the lower and upper half of the data.
4. The IQR is the difference between the upper and lower medians.

5. How is the statistical significance of an insight assessed?

Statistical significance is often calculated with statistical hypothesis testing, which tests the validity of a hypothesis by figuring out the probability that your results have happened by chance. Here, a “hypothesis” is an assumption or belief about the relationship between your datasets. The result of a hypothesis test allows us to see whether this assumption holds under scrutiny or not.

A standard hypothesis test relies on two hypotheses.

* Null hypothesis: The default assumption of a statistical test that you’re attempting to disprove (e.g., an increase in cost won’t affect the number of purchases).
* Alternative hypothesis: An alternate theory that contradicts your null hypothesis (e.g., an increase in cost will reduce the number of purchases). This is the hypothesis you hope to prove.