# Statistics Assignment 1

## Q1: What exactly is the difference between descriptive and inferential statistics?

Ans:

Descriptive statistics, as its name suggests, describes or provides an insight on what sort of data we have under consideration. It describes the data in terms of descriptives like mean, median, mode, standard deviation, variance in the data, population size, type of data variables, their measures like if they are nominal or ordinal data etc. To basically understand the underlying data and to organize it, we use the descriptive statistics.

Inferential statistics is for the purpose of finding data patterns, comparisons and drawing conclusions by testing and analyzing data using different statistical tests like z-test, t-test, analysis of variance, significance testing etc. We draw conclusions based on the outcomes of the different tests for the purpose of making further decisions. In other words, we convert data to useful information by statistically analyzing it.

## Q2: I’m not sure what is the difference between a sample and a population?

Ans:

**Population** is the whole dataset of the domain under consideration. It is like the superset of the data.

E.g.

1. All schools in a country
2. All students in a school
3. All students in a class

Here, in a. we are talking of schools in the country, so the population data domain is the country and all the schools of the country will be its population data.

Similarly, in b. we are talking about students in a particular school, so the domain under consideration is school and population data is all the students in the school.

Same is for c. as in the above cases.

**Sample** is the subset of the population data that represents the entire population. We extract data from population data over some criteria. That data is then supposed to represent the entire population data. This is called the sample. Since the population data can be huge and will require enormous resources to process the entire data, we use the sample.

## Q3: What distinguishes descriptive statistics from other types of statistics?

Ans:

Descriptive statistics gives us the information about the data like if it is categorical or non-categorical, or continuous/discrete data, what is the distribution like, what are the measure scales, how much it is varied around its average or mid-point etc. It also helps in finding the outliers – the data at the extremes of the distribution causing difficulty in understanding the general data distribution. Some of the descriptive statistics are:

* Mean or the average
* Median or the mid-point data value
* Mode or the most occurring data point(s)
* Frequency or counts of a particular data point in the distribution
* Distribution break points like percentiles or centiles, quartiles, minimum, maximum
* Lower fence/Upper fence for detecting outliers
* Data order or ranking
* Variance and standard deviation

## Q4: What is the difference between quantitative and qualitative data?

Ans:

Data as generally observed, is basically of two types:

* Quantitative
* Qualitative

**Quantitative data** is basically numeric data upon which the operations like addition subtraction etc. can be performed. Quantitative data means that data can be quantified. It can be measured as a discrete value or a continuous value based on the kind of data it is. E.g. Number of eggs in a basket – this is a discrete quantitative as we can quantify it with a specific number. For the continuous data, e.g. can be Height of a person. The height can be any real number so it is a continuous type of quantitative data. Continuous data is usually measured on an interval scale.

**Qualitative data** is non-numeric data that basically specifies the a feature or identity. For e.g. Name of the school, Gender etc. Such type of data cannot be used in mathematical operations as it is not numeric type. Such data is qualitative. It can be ordered or non-ordered.

## Q5: What is the definition of a percentile?

Ans:

Percentile a.k.a centile for a given data-value ‘k’, is the percentage of data that lie below it, in the given ordered data distribution. Denoted as P(k). Formula to calculate the percentile is:

P(k) = (# of values before k)/(# of total data values or N) \* 100, where x is the data-value under consideration

e.g. A = {4,4,5,7,8,25,25,67,80,100}, if x=67 for which we want to find the percentile, we will calculate it as: P(x) = 7 / 10 \* 100 = 70%