STA-O: The Basics of Statistics

What is Statistics?

Statistics is the science of gathering, organizing, analyzing, and presenting numerical data. Statistics is used to draw conclusions on large amounts of data by measuring probabilities of potential outcomes and extrapolating that data to predict future results.

Statistics is a discipline that estimates data, and as such cannot prove anything. An estimate is an educated guess, and statistics acknowledges a range of error of estimates and uses probability to describe the changes of an educated guess.

Often, statistics is misused in an attempt to skew the view of data, but it is unable to change the actual data – only the perceptive view of it can be altered.

Why do we use statistics?

Any time we need to categorize or characterize a group of things in order to research it, we cannot gather data on every single thing in that group. For example, we wish to find the average rainfall across the south east of the United States. It is unrealistic to measure the rainfall in distances of 5 miles apart; therefore a larger distance is required between rain collectors gathering the data. This limits the sample – therefore we must estimate the true parameters by using statistical methodology.

By statistically analyzing data, underlying patterns that are not otherwise observable can be found – leading to deeper understanding of the system under observation.

Different Types of Data

1. discrete data – non-continuous, usually values that can be counted
2. continuous data – assume an infinite number of values between any two specific values and are obtained via measurement (e.g. fractions, decimals)

Basic Definitions

- *Variable* – a characteristic or attribute that can assume different values – in other words, a value that can change dependent upon the data; random variables are determined by chance
- *Data* – values (measurement or observation) that variables can assume
- *Population* – the total set of measurements, consisting of all things under study; usually the population is not determinable and must be estimated by a sample
- *Sample* – group of measurements or subjects selected from a population, taken as a representation of the population
- *Descriptive Statistics* – the collection, organization, summarization, analysis, and presentation of data

- *Inferential Statistics* – generalize using samples to create populations, perform estimations and hypothesis tests, determine relationships among variables, and make predictions
- *Ordinal Level of Measurement* – classifies data into categories that can be ranked using estimates
- *Interval Level of Measurement* – ranks data and the precise differences between units of measure exist (however, no meaningful zero (or null) value exists)
- *Ratio Level of Measurement* – ranks data in the same way interval measurement does, but has a true zero; in addition, true ratios exist when the same variable is measured on two different members of the population (correlation)

There are different types of statistics! What type do I use? – The answer depends upon:

-What questions you want answered
-The types and quality of data
-If the study involves one or more samples and/or one or more variables
-The type of data being studied

Signifigant Figures:

-Should always maintain significant digits through all calculations to avoid rounding errors
-The last digit should imply precision and is an estimate (due to rounding)
ex. 63.246 can mean any number between 63.245<u>5</u> and 63.246<u>5</u>

Rounding:

-Do not round if the digit following is a number equal to or less than 5
ex. 35.4874 is left alone, as is 58.235
ex. 23.45278 is rounded to 23.4528
-Add one to the number if it is followed by a number greater than 5
ex. 9.23846 is rounded to 9.2385