**STATISTICS**

# What is Statistics?

Statistics is one of the most honestly useful math topics you are likely to study. Nearly every kind of occupation and human activity can benefit from an application of statistics. In the most general sense, statistics describes a set of tools and techniques that can be used to describe, organize, and interpret information or data. What are those data? They could be the scores of students on their final exam in history, the speed with which a new drug relieves headaches, the number of complaints received by customers, the free-throw percentage for different college

basketball players, or the average price of going out for pizza on a Friday night. Statistics is a tool that helps us understand the world around us and make better decisions with the information we have available to us. By knowing which products sell better at a particular time of year, for example, a business can make the best use of product placement and advertising. Knowing what time(s) of the day, week, or year are the busiest, a restaurant manager can efficiently schedule her employees so as not to waste labor costs. Coaches that know the statistics on their players can use that information to help them field stronger offensives or defensives.

# Types of Data

Quantitative Data: When a characteristic provide information in which the count, or quantity is most important is called a numerical variable, or quantitative variable.

Qualitative Data: When a characteristic can be neatly placed into well-defined groups, or categories, that do not depend on order, it is called a categorical variable, or qualitative variable.

# Population vs. Sample

We have already defined a population as the total group being studied. Most of the time, it is extremely difficult or very costly to collect all the information about a population.

In the Galapagos island, it would be very difficult and perhaps even destructive to search every square meter of the habitat to be sure that you counted every tortoise.

In an example closer to home, it is very expensive to get accurate and complete information about all the residents of the United States to help effectively address the needs of a changing population. This is why a complete counting, or census, is only attempted every ten years. Because of these problems, it is common to use a smaller, representative group from the population, called a sample.

# Measurement

**Nominal measurement**

A nominal measurement is one in which the values of the variable are names. The names of the different species of Galapagos tortoises are an example of a nominal measurement.

**Ordinal measurement**

An ordinal measurement involves collecting information of which the order is somehow significant. The name of this level is derived from the use of ordinal numbers for ranking (1st , 2nd, 3rd etc.). If we measured the different species of tortoise from the largest population to the smallest, this would be an example of ordinal measurement. In ordinal measurement, the distance between two consecutive values does not have meaning. The 1st and 2nd largest

tortoise populations by species may differ by a few thousand individuals, while the 7th and 8th may only differ by a few hundred.

**Interval measurement**

With interval measurement, the distance between any two values has a specific meaning. An example commonly cited for interval measurement is temperature (either degrees Celsius or degrees Fahrenheit). A change of 1 degree is the same if the temperature goes from 0\_ C to 1\_ C as it is when the temperature goes from 40\_ C to 41\_ C. In addition, there is meaning to the values between the ordinal numbers. That is, a half of a degree has meaning.

**Ratio measurement**

A ratio measurement is the estimation of the ratio between a magnitude of a continuous quantity and a unit magnitude of the same kind. A variable measured at this level not only includes the concepts of order and interval, but also adds the idea of ’nothingness’, or absolute zero. With the temperature scale of the previous example, 0\_ C is really an arbitrarily chosen number (the temperature at which water freezes) and does not represent the absence of temperature. As a result, the ratio between temperatures is relative, and 40\_ C, for example, is not twice as hot as

20\_ C. On the other hand, for the Galapagos tortoises, the idea of a species having a population of 0 individuals is all too real! As a result, the estimates of the populations are measured on a ratio level, and a species with a population of about 3,300 really is approximately three times as large as one with a population near 1,100

.

# Graphs Representation

* A frequency table is useful to organize data into classes according to the number of occurrences, or frequency, of each class.
* Relative frequency shows the percentage of data in each class.
* A histogram is a graphical representation of a frequency table (either actual or relative frequency).
* A frequency polygon is created by plotting the midpoint of each bin at its frequency and connecting the points with line segments. Frequency polygons are useful for viewing

the overall shape of a distribution of data, as well as comparing multiple data sets. For any distribution of data, you should always be able to describe the shape, center, and spread.

* A data set that is mound shaped can be classified as either symmetric or skewed. Distributions that are skewed left have the bulk of the data concentrated on the

higher end of the distribution, and the lower end, or tail, of the distribution is spread out to the left.

* A skewed-right distribution has a large portion of the data concentrated in the lower values of the variable, with the tail spread out to the right.