**Statistical Terms**

bias – a systematic favoritism in the data collected such that the collected data is not truly representative of the entire population

confidence interval – the interval in hypothesis testing based on a calculated margin of error. If the margin of error is 2%, the confidence interval is plus or minus 2%.

confidence level – The predetermined level in hypothesis testing that indicates that the sample obtained is representative of the population being studied. So a 95% confidence level means that if you take repeated samples from the same population, that the sample obtained will lie within the specified confidence intervals 95% of the time.

correlation and causation – correlation specifies a relationship that is either direct or inverse. It can be strong or weak, or non-existent. Strong correlation does not imply causation. Do not confuse the two.

data – information collected from a study or measurement

data set – a collection of data from a measurement

descriptive statistics –one branch of statistics that describes data as distinguished from inferential statistics, which makes inferences about a population from a sample of that population

distribution – a list or graph of all of the possible values that a variable can take on in the population and how often those values occur

estimation – used because taking a census of an entire population would be too costly or time consuming.

margin of error – a calculated statistic from a sample that denotes the accuracy by which the sample represents the population at a given confidence level.

confidence interval – the interval in hypothesis testing based on a calculated margin of error. If the margin of error is 2%, the confidence interval is plus or minus 2%.

experiments – a study that imposes a certain amount of control on the study’s subjects and their environment. Purpose of the experiments is to pinpoint a cause and effect relationship between two or variables.

treatment group vs. control group – treatment group are those that receive the treatment that supposedly has an effect on the outcome; the control group consists of those who did not receive the treatment.

placebo – a fake treatment, or lack of treatment, often given to the members of the control group

blind and double blind –a blind experiment is one in which the participants do not know whether they are in the treatment or control group; a double blind study is one in which neither the participants nor the researchers know who is in the treatment or control group

hypothesis testing – a statistical procedure that’s designed to test a claim. The claim is usually being made about a population parameter (e.g. mean).

inferential statistics – that branch of statistics that seeks to draw inferences about populations of data from samples taken from that population.

law of averages – a rule of probability. It says that, in the long term, results will average out to their expected value, but in the short term, no one knows what will happen.

mean/average – a descriptive statistical measure that adds all the values of a set of data and then divides by the number of values.

median – the middle score in a data set. With an even number of data, the average between the two middle values.

mode – the most frequently occurring score. A distribution can have one mode (uni-modal) or two modes (bi-modal) or more.

normal distribution (or bell-shaped curve) – a distribution of data, from smallest to largest, in which most of the data are centered around the average in a big lump and as you move farther out on either side of the mean, you find fewer and fewer data points.

percentile – a percentile reported for a given score is the percentage of values in the data set that fall below that certain score

population – a group items, objects, data, that you want to observe.

probability vs. odds –a probability is a measurement of the likelihood of an event happening. Odds are slightly different. As an example, suppose that the probability that a race horse will win the race is 1 out of 10, or 1 divided by 10 or .10. What are the horse’s odds of winning? They are 9 to 1. That’s because odds are actually a ratio of the changes of losing to the chances of winning. The horse has a 9 in 10 chance of losing and a 1 in 10 chance of winning. Take 9/10 over (divided by) 1/10. The tens cancel out leaving 9/1 which in odds lingo is 9 to 1.

p-value – In hypothesis testing, you use evidence to confirm or deny a claim that is made about a population. The p value is used to weigh the strength of the evidence. The p value is a number between 0 and 1 that reflects the strength of the data that are being used to evaluate the null hypothesis. If the p value is small, you have strong evidence against a null hypothesis. At the 95% confidence level, your p value would be .05

random – each item, object in a population has an equal chance of being chosen

sample – a subset of a population, which can represent it, if the sample is random and sufficiently large

standard deviation – a way to measure the amount of variability among the numbers in a data set. It is the standard, or typical, among of deviation (or distance) from the mean, i.e. the average distance from the mean

standard score – represents the number of standard deviations above or below the mean. Also known as a Z score.

statistic – a number that summarizes data collected from a sample

statistical significance –A statistically significant result is one that would have had a very small probability of happening by chance. The p value reflects that probability.

surveys –a measurement tool that is used to gather people’s opinions along with relevant demographic information.

Z score – the number of standard deviations above or below the mean. Also known as a standard score.