Unit: 5. STATISTICS

VOCABULARY

bell curve	A symmetrical bell shaped curve that represents the distribution of a set of values. Also called normal distribution.
bias	Bias is the preference for a particular group or subset. In the example above, the bias was toward viewers who like sports. When your sample is not chosen at random, you introduce bias into your study. So, in a certain sense, a biased sample is the opposite of a random sample. Why? In a biased sample, some members of the population are more likely to be chosen than others. In a random sample, every member of the population is equally likely to be chosen. Bias may also occur in surveys. See "survey" below for more information on this type of bias.
block design	breaking a sample group into subgroups
cofounding	when reasonable explanations for an experiment are not eliminated
data set	A set of numbers that are related.
dependent variable	measured response
expected value	the average outcome of an experiment in the long run based on probability
experiment	process of data collection, with unknown outcome; a well-defined procedure under controlled conditions designed to gather data that tests a hypothesis. For example, you want to

	test the hypothesis that flowers grow faster when they are watered every other day as opposed to every day. Then you could water one set of plants every day and another set of plants every other day. What data would you collect? The height of each plant once a day. What conditions would you need to control? You would need to make sure that the groups of plants are treated the same in every other way by getting the same amount of sun and being maintained at the same temperature, etc.
fair game	one in which the expected value is 0
histogram	A graph that represents data in groups. It is similar to a bar graph, but unlike in a bar graph, there is no spacing between the columns that represent the data.
independent variable	factor that is being varied by the researcher
mean	Average of all numbers in a data set. In statistics, you find the mean of your sample. Then you use that mean to approximate the mean of the entire population. If you want the mean of the data to be a good approximation of the mean of the entire population, your need to make sure that you collect a random sample. To find the mean, you find the sum of the data points you collected and divide it by the number of data points in the set. For example, if the data you collected was {3, 3, 6}. Then the mean would be $(3 + 3 + 6) \div 3 = 12 \div 3 = 4$. It's important

	to note that the mean does not have to be a data point in your set. You can generalize the formula as well. If you collect *n* data points $\{d_1, d_2, d_3, \ldots d_n\}$, then the mean of your data set is $\mu = \frac{d_1 + d_2 + d_3 + \ldots + d_n}{n}$. The Greek letter μ—pronounced "mu"—is used because the word "mean" begins with the letter "m." It is important that you divide by the number of data points in your set, *n*.
normal distribution	See also, bell curve. Symmetrical distribution of data.
observational study	examining the elements (like people) as is
outcomes	results
placebo effect	positive response to neutral treatment
population	entire group of elements (like people) from which data is collected; so, in a certain sense, you can think of the population as the "opposite" of the sample because a sample is not the entire population. A sample is chosen subset upon that you use to draw conclusions about the entire population.
random	Random means free from bias. So, if you choose a random sample from a population, every member of the population has an equal chance of being chosen. It should be pointed out that this "true randomness" is an ideal which is difficult to achieve in real world experiments conducted on real world populations. That is why statisticians make every effort to ensure their

	samples chosen as free from bias as is possible, that is, as randomly as possible under real world conditions.
random sample	sample consisting of elements that have an equal chance of being selected
sample	A subset of data taken from a population that is of sufficiently-large size as to be representative of that population as measured by sample statistics, like the sample mean and standard deviation. In most cases, you want a random sample. Why? A random sample has the best chance of accurately representing the entire population.
sample mean	The mean of a collection of samples that are all taken from a population, and which is compared to the population mean to determine if the samples are representative of the population.
sample survey	a survey of a sample of a population that represents characteristics of the population if the sample is sufficiently large and representative of the population.
standard deviation	A value that represents how far apart the figures in a data set are from the average. A low deviation indicates the figures are very close in value to the average. A high deviation indicates the figures are far apart in value from the average.
statistics	study of collecting, organizing, and interpreting numerical data
survey	A survey is the means of collecting a sample. In statistics, a survey is often a question or set of questions. It is critical that

	the questions, like the sample itself, be free from bias. Bias in a question tends to favor one response. As an extreme example of a biased question about colors might be, "Don't you think blue is beautiful and yellow is yucky?" An unbiased form of the question might be, "Which color do you prefer, blue or yellow?" Though careful, statisticians might be concerned about the order in which the colors are offered, so that might further simplify the question to, "What is your favorite color?" This type of question is most free from bias because it does not suggest a response and therefore does not favor a response.
variance	How spread out the values in a data set are from the mean.

Vocab Arcade

Lesson 12

Variation Rules Reviewed