

## Glossary

**Analysis of Variance** | also referred to as ANOVA, is a method of testing whether or not the means of three or more populations are equal. The method is applicable if: (1) all populations of interest are normally distributed. (2) the populations have equal standard deviations. (3) samples (not necessarily of the same size) are randomly and independently selected from each population. (4) The test statistic for analysis of variance is the  $F$ -ratio. [OpenStax]

**Average** | a number that describes the central tendency of the data; there are a number of specialized averages, including the arithmetic mean, weighted mean, median, mode, and geometric mean. [OpenStax]

**Bernoulli Trials** | an experiment with the following characteristics: (1) There are only two possible outcomes called "success" and "failure" for each trial. (2) The probability  $p$  of a success is the same for any trial (so the probability  $q = 1 - p$  of a failure is the same for any trial). [OpenStax]

**Binomial Distribution** | a discrete random variable (RV) that arises from Bernoulli trials. There are a fixed number,  $n$ , of independent trials. "Independent" means that the result of any trial (for example, trial 1) does not affect the results of the following trials, and all trials are conducted under the same conditions. Under these circumstances the binomial RV  $X$  is defined as the number of successes in  $n$  trials. The notation is:  $X \sim B(n, p)$   $\mu = np$  and the standard deviation is  $\sigma = \sqrt{npq}$ . The probability of exactly  $x$  successes in  $n$  trials is  $P(X = x) = \binom{n}{x} p^x q^{n-x}$ . [OpenStax]

**Binomial Experiment** | a statistical experiment that satisfies the following three conditions: (1) There are a fixed number of trials,  $n$ . (2) There are only two possible outcomes, called "success" and "failure," for each trial. The letter  $p$  denotes the probability of a success on one trial, and  $q$  denotes the probability of a failure on one trial. (3) The  $n$  trials are independent and are repeated using identical conditions. [OpenStax]

**Binomial Probability Distribution** | a discrete random variable (RV) that arises from Bernoulli trials; there are a fixed number,  $n$ , of independent trials. "Independent" means that the result of any trial (for example, trial one) does not affect the results of the following trials, and all trials are conducted under the same conditions. Under these circumstances the binomial RV  $X$  is defined as the number of successes in  $n$  trials. The notation is:  $X \sim B(n, p)$ . The mean is  $\mu = np$  and the standard deviation is  $\sigma = \sqrt{npq}$ . The probability of exactly  $x$  successes in  $n$  trials is  $P(X = x) = \binom{n}{x} p^x q^{n-x}$ . [OpenStax]

**Blinding** | not telling participants which treatment a subject is receiving [OpenStax]

**Box plot** | a graph that gives a quick picture of the middle 50% of the data [OpenStax]

**Categorical Variable** | variables that take on values that are names or labels [OpenStax]

**Central Limit Theorem** | Given a random variable (RV) with known mean  $\mu$  and known standard deviation  $\sigma$ , we are sampling with size  $n$ , and we are interested in two new RVs: the sample mean,  $\bar{X}$ , and the sample sum,  $\sum X$ . If the size ( $n$ ) of the sample is sufficiently large, then  $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$  and  $\sum X \sim N(n\mu, \sqrt{n}\sigma)$ . If the size ( $n$ ) of the sample is sufficiently large, then the distribution of the sample means and the distribution of the sample sums will approximate a normal distributions regardless of the shape of the population. The mean of the sample means will equal the population mean, and the mean of the sample sums will equal  $n$  times the population mean. The standard deviation of the distribution of the sample means,  $\frac{\sigma}{\sqrt{n}}$ , is called the standard error of the mean. [OpenStax]

**Central Limit Theorem** | Given a random variable (RV) with known mean  $\mu$  and known standard deviation  $\sigma$ . We are sampling with size  $n$  and we are interested in two new RVs - the sample mean,  $\bar{X}$ , and the sample sum,  $\sum X$ . If the size  $n$  of the sample is sufficiently large, then  $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$  and  $\sum X \sim N\left(n\mu, \sqrt{n}\sigma\right)$ . If the size  $n$  of the sample is sufficiently large, then the distribution of the sample means and the distribution of the sample sums will approximate a normal distribution regardless of the shape of the population. The mean of the sample means will equal the population mean and the mean of the sample sums will equal  $n$  times the population mean. The standard deviation of the distribution of the sample means,  $\frac{\sigma}{\sqrt{n}}$ , is called the standard error of the mean. [OpenStax]

**Cluster Sampling** | a method for selecting a random sample and dividing the population into groups (clusters); use simple random sampling to select a set of clusters. Every individual in the chosen clusters is included in the sample. [OpenStax]

**Coefficient of Correlation** | a measure developed by Karl Pearson (early 1900s) that gives the strength of association between the independent variable and the dependent variable; the formula is:  $r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$  where  $n$  is the number of data points. The coefficient cannot be more than 1 or less than  $-1$ . The closer the coefficient is to  $\pm 1$ , the stronger the evidence of a significant linear relationship between  $x$  and  $y$ . [OpenStax]

**Conditional Probability** | the likelihood that an event will occur given that another event has already occurred [OpenStax]

**Confidence Interval (CI)** | an interval estimate for an unknown population parameter. This depends on: (1) The desired confidence level. (2) Information that is known about the distribution (for example, known standard deviation). (3) The sample and its size. [OpenStax]

**Confidence Level (CL)** | the percent expression for the probability that the confidence interval contains the true population parameter; for example, if the  $CL = 90\%$ , then in 90 out of 100 samples the interval estimate will enclose the true population parameter. [OpenStax]

**contingency table** | the method of displaying a frequency distribution as a table with rows and columns to show how two variables may be dependent (contingent) upon each other; the table provides an easy way to calculate conditional probabilities. [OpenStax]

**Continuous Random Variable** | a random variable (RV) whose outcomes are measured; the height of trees in the forest is a continuous RV. [OpenStax]

**Control Group** | a group in a randomized experiment that receives an inactive treatment but is otherwise managed exactly as the other groups [OpenStax]

**Convenience Sampling** | a nonrandom method of selecting a sample; this method selects individuals that are easily accessible and may result in biased data. [OpenStax]

**Cumulative Relative Frequency** | The term applies to an ordered set of observations from smallest to largest. The cumulative relative frequency is the sum of the relative frequencies for all values that are less than or equal to the given value. [OpenStax]

**Data** | a set of observations (a set of possible outcomes); most data can be put into two groups: qualitative (an attribute whose value is indicated by a label) or quantitative (an attribute whose value is indicated by a number). Quantitative data can be separated into two subgroups: discrete and continuous. Data is discrete if it is the result of counting (such as the number of students of a given ethnic group in a class or the number of books on a shelf). Data is continuous if it is the result of measuring (such as distance traveled or weight of luggage) [OpenStax]

**decay parameter** | The decay parameter describes the rate at which probabilities decay to zero for increasing values of  $x$ . It is the value  $m$  in the probability density function  $f(x) = me^{(-mx)}$  of an exponential random variable. It is also equal to  $m = \frac{1}{\mu}$ , where  $\mu$  is the mean of the random variable. [OpenStax]

**Degrees of Freedom (df)** | the number of objects in a sample that are free to vary. [OpenStax]

**Dependent Events** | If two events are NOT independent, then we say that they are dependent. [OpenStax]

**Discrete Random Variable** | a random variable (RV) whose outcomes are counted [OpenStax]

**Double-blinding** | the act of blinding both the subjects of an experiment and the researchers who work with the subjects [OpenStax]

**Equally Likely** | Each outcome of an experiment has the same probability. [OpenStax]

**Error Bound for a Population Mean (EBM)** | the margin of error; depends on the confidence level, sample size, and known or estimated population standard deviation. [OpenStax]

**Error Bound for a Population Proportion (EBP)** | the margin of error; depends on the confidence level, the sample size, and the estimated (from the sample) proportion of successes. [OpenStax]

**Event** | a subset of the set of all outcomes of an experiment; the set of all outcomes of an experiment is called a sample space and is usually denoted by  $S$ . An event is an arbitrary subset in  $S$ . It can contain one outcome, two outcomes, no outcomes (empty subset), the entire sample space, and the like. Standard notations for events are capital letters such as  $A$ ,  $B$ ,  $C$ , and so on. [OpenStax]

**Expected Value** | expected arithmetic average when an experiment is repeated many times; also called the mean. Notations:  $\mu$ . For a discrete random variable (RV) with probability distribution function  $P(x)$ , the definition can also be written in the form  $\mu = \sum \{xP(x)\}$ . [OpenStax]

**Experiment** | a planned activity carried out under controlled conditions [OpenStax]

**Experimental Unit** | any individual or object to be measured [OpenStax]

**Explanatory Variable** | the independent variable in an experiment; the value controlled by researchers [OpenStax]

**Exponential Distribution** | a continuous random variable (RV) that appears when we are interested in the intervals of time between some random events, for example, the length of time between emergency arrivals at a hospital; the notation is  $X \sim \text{Exp}(m)$ . The mean is  $\mu = \frac{1}{m}$  and the standard deviation is  $\sigma = \frac{1}{m}$ . The probability density function is  $f(x) = me^{-mx}$ ,  $x \geq 0$  and the cumulative distribution function is  $P(X \leq x) = 1 - e^{-mx}$ . [OpenStax]

**First Quartile** | the value that is the median of the of the lower half of the ordered data set [OpenStax]

**Frequency** | the number of times a value of the data occurs [OpenStax]

**Frequency Polygon** | looks like a line graph but uses intervals to display ranges of large amounts of data [OpenStax]

**Frequency Table** | a data representation in which grouped data is displayed along with the corresponding frequencies [OpenStax]

**Geometric Distribution** | a discrete random variable (RV) that arises from the Bernoulli trials; the trials are repeated until the first success. The geometric variable  $X$  is defined as the number of trials until the first success. Notation:  $X \sim G(p)$ . The mean is  $\mu = \frac{1}{p}$  and the standard deviation is  $\sigma = \sqrt{\frac{1}{p} \left( \frac{1-p}{p} \right)}$ . The probability of exactly  $x$  failures before the first success is given by the formula:  $P(X = x) = p(1-p)^{x-1}$ . [OpenStax]

**Geometric Experiment** | a statistical experiment with the following properties: (1) There are one or more Bernoulli trials with all failures except the last one, which is a success. (2) In theory, the number of trials could go on forever. There must be at least one trial. (3) The probability,  $p$ , of a success and the probability,  $q$ , of a failure do not change from trial to trial [OpenStax]

**Hypergeometric Experiment** | a statistical experiment with the following properties: (1) You take samples from two groups. (2) You are concerned with a group of interest, called the first group. (3) You sample without replacement from the combined groups. (4) Each pick is not independent, since sampling is without replacement. (5) You are not dealing with Bernoulli Trials. [OpenStax]

**Hypergeometric Probability** | a discrete random variable (RV) that is characterized by: (1) A fixed number of trials. (2) The probability of success is not the same from trial to trial. We sample from two groups of items when we are interested in only one group.  $X$  is defined as the number of successes out of the total number of items chosen. Notation:  $X \sim H(r, b, n)$ , where  $r$  = the number of items in the group of interest,  $b$  = the number of items in the group not of interest, and  $n$  = the number of items chosen. [OpenStax]

**Hypothesis** | a statement about the value of a population parameter, in case of two hypotheses, the statement assumed to be true is called the null hypothesis (notation  $H_0$ ) and the contradictory statement is called the alternative hypothesis (notation  $H_a$ ). [OpenStax]

**Hypothesis Testing** | Based on sample evidence, a procedure for determining whether the hypothesis stated is a reasonable statement and should not be rejected, or is unreasonable and should be rejected. [OpenStax]

**Independent Events** | The occurrence of one event has no effect on the probability of the occurrence of another event. Events  $A$  and  $B$  are independent if one of the following is true: (1)  $P(A|B) = P(A)$ , (2)  $P(B|A) = P(B)$ , (3)  $P(A \text{ AND } B) = P(A)P(B)$  [OpenStax]

**Inferential Statistics** | also called statistical inference or inductive statistics; this facet of statistics deals with estimating a population parameter based on a sample statistic. For example, if four out of the 100 calculators sampled are defective we might infer that four percent of the production is defective. [OpenStax]

**Informed Consent** | Any human subject in a research study must be cognizant of any risks or costs associated with the study. The subject has the right to know the nature of the treatments included in the study, their potential risks, and their potential benefits. Consent must be given freely by an informed, fit participant. [OpenStax]

**Institutional Review Board** | a committee tasked with oversight of research programs that involve human subjects [OpenStax]

**Interval** | also called a class interval; an interval represents a range of data and is used when displaying large data sets [OpenStax]

**Level of Significance of the Test** | probability of a Type I error (reject the null hypothesis when it is true). Notation:  $\alpha$ . In hypothesis testing, the Level of Significance is called the preconceived  $\alpha$  or the preset  $\alpha$ . [OpenStax]

**Lurking Variable** | a variable that has an effect on a study even though it is neither an explanatory variable nor a response variable [OpenStax]

**Mean** | a number that measures the central tendency; a common name for mean is "average." The term "mean" is a shortened form of "arithmetic mean." By definition, the mean for a sample (denoted by  $\bar{x}$ ) is  $\bar{x} = \frac{\sum x}{n}$  where  $\sum x$  is the sum of all values in the sample and  $n$  is the number of values in the sample. The mean for a population (denoted by  $\mu$ ) is  $\mu = \frac{\sum x}{N}$  where  $\sum x$  is the sum of all values in the population and  $N$  is the number of values in the population. [OpenStax]

**Mean of a Probability Distribution** | the long-term average of many trials of a statistical experiment [OpenStax]

**Median** | a number that separates ordered data into halves; half the values are the same number or smaller than the median and half the values are the same number or larger than the median. The median may or may not be part of the data. [OpenStax]

**memoryless property** | For an exponential random variable  $X$ , the memoryless property is the statement that knowledge of what has occurred in the past has no effect on future probabilities. This means that the probability that  $X$  exceeds  $x + k$ , given that it has exceeded  $x$ , is the same as the probability that  $X$  would exceed  $k$  if we had no knowledge about it. In symbols we say that  $P(X > x + k | X > x) = P(X > k)$  [OpenStax]

**Midpoint** | the mean of an interval in a frequency table [OpenStax]

**Mode** | the value that appears most frequently in a set of data [OpenStax]

**Mutually Exclusive** | Two events are mutually exclusive if the probability that they both happen at the same time is zero. If events  $A$  and  $B$  are mutually exclusive, then  $P(A \text{ AND } B) = 0$ . [OpenStax]

**Nonsampling Error** | an issue that affects the reliability of sampling data other than natural variation; it includes a variety of human errors including poor study design, biased sampling methods, inaccurate information provided by study participants, data entry errors, and poor analysis. [OpenStax]

**Normal Distribution** | a continuous random variable (RV) with pdf  $f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ , where  $\mu$  is the mean of the distribution and  $\sigma$  is the standard deviation; notation:  $X \sim N(\mu, \sigma)$ . If  $\mu = 0$  and  $\sigma = 1$ , the RV is called a standard normal distribution. [OpenStax]

**Numerical Variable** | variables that take on values that are indicated by numbers [OpenStax]

**One-Way ANOVA** | a method of testing whether or not the means of three or more populations are equal; the method is applicable if: (1) all populations of interest are normally distributed. (2) the populations have equal standard deviations. (3) samples (not necessarily of the same size) are randomly and independently selected from each population. (4) The test statistic for analysis of variance is the F-ratio. [OpenStax]

**Outcome** | a particular result of an experiment [OpenStax]

**Outlier** | an observation that does not fit the rest of the data [OpenStax]

**p-value** | the probability that an event will happen purely by chance assuming the null hypothesis is true. The smaller the p-value, the stronger the evidence is against the null hypothesis. [OpenStax]

**Paired Data Set** | two data sets that have a one to one relationship so that: (1) both data sets are the same size, and (2) each data point in one data set is matched with exactly one point from the other set. [OpenStax]

**Parameter** | a number that is used to represent a population characteristic and that generally cannot be determined easily [OpenStax]

**Parameter** | a numerical characteristic of a population [OpenStax]

**Placebo** | an inactive treatment that has no real effect on the explanatory variable [OpenStax]

**Point Estimate** | a single number computed from a sample and used to estimate a population parameter [OpenStax]

**Poisson distribution** | If there is a known average of  $\lambda$  events occurring per unit time, and these events are independent of each other, then the number of events  $X$  occurring in one unit of time has the Poisson distribution. The probability of  $k$  events occurring in one unit time is equal to  $P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$ . [OpenStax]

**Poisson Probability Distribution** | a discrete random variable (RV) that counts the number of times a certain event will occur in a specific interval; characteristics of the variable: (1) The probability that the event occurs in a given interval is the same for all intervals. (2) The events occur with a known mean and independently of the time since the last event. The distribution is defined by the mean  $\mu$  of the event in the interval. Notation:  $X \sim P(\mu)$ . The mean is  $\mu = np$ . The standard deviation is  $\sigma = \sqrt{\mu}$ . The probability of having exactly  $x$  successes in  $r$  trials is  $P(X = x) = \frac{e^{-\mu} \mu^x}{x!}$ . The Poisson distribution is often used to approximate the binomial distribution, when  $n$  is “large” and  $p$  is “small” (a general rule is that  $n$  should be greater than or equal to 20 and  $p$  should be less than or equal to 0.05). [OpenStax]

**Pooled Proportion** | estimate of the common value of  $p_1$  and  $p_2$ . [OpenStax]

**Population** | all individuals, objects, or measurements whose properties are being studied [OpenStax]

**Probability** | a number between zero and one, inclusive, that gives the likelihood that a specific event will occur [OpenStax]

**Probability** | a number between zero and one, inclusive, that gives the likelihood that a specific event will occur; the foundation of statistics is given by the following 3 axioms (by A.N. Kolmogorov, 1930's): Let  $S$  denote the sample space and  $A$  and  $B$  are two events in  $S$ . Then: (1)  $0 \leq P(\text{event}) \leq 1$ , (2) If  $\text{event}(A)$  and  $\text{event}(B)$  are any two mutually exclusive events, then  $P(\text{event}(A \text{ OR } B)) = P(\text{event}(A)) + P(\text{event}(B))$  and (3)  $P(\text{event}(S)) = 1$ . [OpenStax]

**Probability Distribution Function (PDF)** | a mathematical description of a discrete random variable (RV), given either in the form of an equation (formula) or in the form of a table listing all the possible outcomes of an experiment and the probability associated with each outcome. [OpenStax]

**Proportion** | the number of successes divided by the total number in the sample [OpenStax]

**Qualitative Data** | See Data. [OpenStax]

**Quantitative Data** | See Data. [OpenStax]

**Random Assignment** | the act of organizing experimental units into treatment groups using random methods [OpenStax]

**Random Sampling** | a method of selecting a sample that gives every member of the population an equal chance of being selected. [OpenStax]

**Random Variable (RV)** | a characteristic of interest in a population being studied; common notation for variables are upper case Latin letters  $X$ ,  $Y$ ,  $Z$ ,...; common notation for a specific value from the domain (set of all possible values of a variable) are lower case Latin letters  $x$ ,  $y$ , and  $z$ . For example, if  $X$  is the number of children in a family, then  $x$  represents a specific integer 0, 1, 2, 3,... Variables in statistics differ from variables in intermediate algebra in the two following ways. (1) The domain of the random variable (RV) is not necessarily a numerical set; the domain may be expressed in words; for example, if  $X$  = hair color then the domain is {black, blond, gray, green, orange}. (2) We can tell what specific value  $x$  the random variable  $X$  takes only after performing the experiment [OpenStax]

**Relative Frequency** | the ratio of the number of times a value of the data occurs in the set of all outcomes to the number of all outcomes to the total number of outcomes [OpenStax]

**Representative Sample** | a subset of the population that has the same characteristics as the population [OpenStax]

**Response Variable** | the dependent variable in an experiment; the value that is measured for change at the end of an experiment [OpenStax]

**Sample** | a subset of the population studied [OpenStax]

**Sample Space** | the set of all possible outcomes of an experiment [OpenStax]

**Sampling Bias** | not all members of the population are equally likely to be selected [OpenStax]

**Sampling Distribution** | Given simple random samples of size  $n$  from a given population with a measured characteristic such as mean, proportion, or standard deviation for each sample, the probability distribution of all the measured characteristics is called a sampling distribution. [OpenStax]

**Sampling Error** | the natural variation that results from selecting a sample to represent a larger population; this variation decreases as the sample size increases, so selecting larger samples reduces sampling error. [OpenStax]

**Sampling with Replacement** | Once a member of the population is selected for inclusion in a sample, that member is returned to the population for the selection of the next individual. [OpenStax]

**Sampling without Replacement** | A member of the population may be chosen for inclusion in a sample only once. If chosen, the member is not returned to the population before the next selection. [OpenStax]

**Simple Random Sampling** | a straightforward method for selecting a random sample; give each member of the population a number. Use a random number generator to select a set of labels. These randomly selected labels identify the members of your sample. [OpenStax]

**Skewed** | used to describe data that is not symmetrical; when the right side of a graph looks “chopped off” compared the left side, we say it is “skewed to the left.” When the left side of the graph looks “chopped off” compared to the right side, we say the data is “skewed to the right.” Alternatively: when the lower values of the data are more spread out, we say the data are skewed to the left. When the greater values are more spread out, the data are skewed to the right. [OpenStax]

**Standard Deviation** | a number that is equal to the square root of the variance and measures how far data values are from their mean; notation:  $s$  for sample standard deviation and  $\sigma$  for population standard deviation. [OpenStax]

**Standard Deviation of a Probability Distribution** | a number that measures how far the outcomes of a statistical experiment are from the mean of the distribution [OpenStax]

**Standard Error of the Mean** | the standard deviation of the distribution of the sample means, or  $\frac{\sigma}{\sqrt{n}}$ . [OpenStax]

**Standard Normal Distribution** | a continuous random variable (RV)  $X \sim N(0, 1)$ ; when  $X$  follows the standard normal distribution, it is often noted as  $Z \sim N(0, 1)$ . [OpenStax]

**Statistic** | a numerical characteristic of the sample; a statistic estimates the corresponding population parameter. [OpenStax]

**Stratified Sampling** | a method for selecting a random sample used to ensure that subgroups of the population are represented adequately; divide the population into groups (strata). Use simple random sampling to identify a proportionate number of individuals from each stratum. [OpenStax]

**Student's  $t$ -Distribution** | investigated and reported by William S. Gossett in 1908 and published under the pseudonym Student. The major characteristics of the random variable (RV) are: (1) It is continuous and assumes any real values. (2) The pdf is symmetrical about its mean of zero. However, it is more spread out and flatter at the apex than the normal distribution. (3) It approaches the standard normal distribution as  $n$  gets larger. (4) There is a “family” of  $t$ -distributions: every representative of the family is completely defined by the number of degrees of freedom which is one less than the number of data items. [OpenStax]

**Systematic Sampling** | a method for selecting a random sample; list the members of the population. Use simple random sampling to select a starting point in the population. Let  $k$  = (number of individuals in the population)/(number of individuals needed in the sample). Choose every  $k$ th individual in the list starting with the one that was randomly selected. If necessary, return to the beginning of the population list to complete your sample. [OpenStax]

**The AND Event** | An outcome is in the event  $\text{event}(A \text{ AND } B)$  if the outcome is in both  $\text{event}(A)$  AND  $\text{event}(B)$  at the same time. [OpenStax]

**The Complement Event** | The complement of event  $\text{event}(A)$  consists of all outcomes that are NOT in  $\text{event}(A)$ . [OpenStax]

**The Conditional Probability of A GIVEN B** |  $P(\text{event}(A|B))$  is the probability that event  $\text{event}(A)$  will occur given that the event  $\text{event}(B)$  has already occurred. [OpenStax]

**The Conditional Probability of One Event Given Another Event** |  $P(A|B)$  is the probability that event  $A$  will occur given that the event  $B$  has already occurred. [OpenStax]

**The Law of Large Numbers** | As the number of trials in a probability experiment increases, the difference between the theoretical probability of an event and the relative frequency probability approaches zero. [OpenStax]

**The Or Event** | An outcome is in the event  $\text{event}(A \text{ OR } B)$  if the outcome is in  $\text{event}(A)$  or is in  $\text{event}(B)$  or is in both  $\text{event}(A)$  and  $\text{event}(B)$ . [OpenStax]

**The OR of Two Events** | An outcome is in the event  $A \text{ OR } B$  if the outcome is in  $A$ , is in  $B$ , or is in both  $A$  and  $B$ . [OpenStax]

**Treatments** | different values or components of the explanatory variable applied in an experiment [OpenStax]

**Tree Diagram** | the useful visual representation of a sample space and events in the form of a “tree” with branches marked by possible outcomes together with associated probabilities (frequencies, relative frequencies) [OpenStax]

**Type 1 Error** | The decision is to reject the null hypothesis when, in fact, the null hypothesis is true. [OpenStax]

**Type 2 Error** | The decision is not to reject the null hypothesis when, in fact, the null hypothesis is false. [OpenStax]

**Uniform Distribution** | a continuous random variable (RV) that has equally likely outcomes over the domain,  $a < x < b$ ; it is often referred as the rectangular distribution because the graph of the pdf has the form of a rectangle. Notation:  $X \sim U(a, b)$ . The mean is  $\mu = \frac{a+b}{2}$  and the standard deviation is  $\sigma = \sqrt{\frac{(b-a)^2}{12}}$ . The probability density function is  $f(x) = \frac{1}{b-a}$  for  $a < x < b$  or  $a \leq x \leq b$ . The cumulative distribution is  $P(X \leq x) = \frac{x-a}{b-a}$ . [OpenStax]

**Uniform Distribution** | a continuous random variable (RV) that has equally likely outcomes over the domain,  $(a < x < b)$ ; often referred as the Rectangular Distribution because the graph of the pdf has the form of a rectangle. Notation:  $X \sim U(a, b)$ . The mean is  $\mu = \frac{a+b}{2}$  and the standard deviation is  $\sigma = \sqrt{\frac{(b-a)^2}{12}}$ . The probability density function is  $f(x) = \frac{1}{b-a}$  for  $a < x < b$  or  $a \leq x \leq b$ . The cumulative distribution is  $P(X \leq x) = \frac{x-a}{b-a}$ . [OpenStax]

**Variable** | a characteristic of interest for each person or object in a population [OpenStax]

**Variable (Random Variable)** | a characteristic of interest in a population being studied. Common notation for variables are upper-case Latin letters  $X, Y, Z, \dots$ . Common notation for a specific value from the domain (set of all possible values of a variable) are lower-case Latin letters  $x, y, z, \dots$ . For example, if  $X$  is the number of children in a family, then  $x$  represents a specific integer 0, 1, 2, 3, .... Variables in statistics differ from variables in intermediate algebra in the two following ways. (1) The domain of the random variable (RV) is not necessarily a numerical set; the domain may be expressed in words; for example, if  $X$  = hair color, then the domain is {black, blond, gray, green, orange}. (2) We can tell what specific value  $x$  of the random variable  $X$  takes only after performing the experiment. [OpenStax]

**Variance** | mean of the squared deviations from the mean; the square of the standard deviation. For a set of data, a deviation can be represented as  $x - \bar{x}$  where  $x$  is a value of the data and  $\bar{x}$  is the sample mean. The sample variance is equal to the sum of the squares of the deviations divided by the difference of the sample size and one. [OpenStax]

**Venn Diagram** | the visual representation of a sample space and events in the form of circles or ovals showing their intersections [OpenStax]

**z-score** | the linear transformation of the form  $z = \frac{x - \mu}{\sigma}$ ; if this transformation is applied to any normal distribution  $X \sim N(\mu, \sigma)$  the result is the standard normal distribution  $Z \sim N(0,1)$ . If this transformation is applied to any specific value  $x$  of the RV with mean  $\mu$  and standard deviation  $\sigma$ , the result is called the z-score of  $x$ . The z-score allows us to compare data that are normally distributed but scaled differently. [OpenStax]