

# Glossary terms from module 2

**Addition rule (for mutually exclusive events):** The concept that if the events A and B are mutually exclusive, then the probability of A or B happening is the sum of the probabilities of A and B

**Bayes' theorem:** A math formula for stating that for any two events A and B, the probability of A given B equals the probability of A multiplied by the probability of B given A divided by the probability of B; Also referred to as Bayes' rule

**Bayes' rule:** (Refer to **Bayes' theorem**)

**Bayesian inference:** (Refer to **Bayesian statistics**)

**Bayesian statistics:** A powerful method for analyzing and interpreting data in modern data analytics; Also referred to as Bayesian inference

**Binomial distribution:** A discrete distribution that models the probability of events with only two possible outcomes: success or failure

**Classical probability:** A type of probability based on formal reasoning about events with equally likely outcomes

**Complement of an event:** In statistics, refers to an event not occurring

**Complement rule:** A concept stating that the probability that event A does not occur is one minus the probability of A

**Conditional probability:** Refers to the probability of an event occurring given that another event has already occurred

**Continuous random variable:** A variable that takes all the possible values in some range of numbers

**Dependent events:** The concept that two events are dependent if one event changes the probability of the other event

**Discrete random variable:** A variable that has a countable number of possible values

**Empirical probability:** A type of probability based on experimental or historical data

**Empirical rule:** A concept stating that the values on a normal curve are distributed in a regular pattern, based on their distance from the mean

**False positive:** A test result that indicates something is present when it really is not

**Independent events:** The concept that two events are independent if the occurrence of one event does not change the probability of the other event

**Multiplication rule (for independent events):** The concept that if the events A and B are independent, then the probability of both A and B happening is the probability of A multiplied by the probability of B

**Mutually exclusive:** The concept that two events are mutually exclusive if they cannot occur at the same time

**Normal distribution:** A continuous probability distribution that is symmetrical on both sides of the mean and bell-shaped

**Objective probability:** A type of probability based on statistics, experiments, and mathematical measurements

**Poisson distribution:** A probability distribution that models the probability that a certain number of events will occur during a specific time period

**Posterior probability:** Refers to the updated probability of an event based on new data

**Prior probability:** Refers to the probability of an event before new data is collected

**Probability:** The branch of mathematics that deals with measuring and quantifying uncertainty

**Probability distribution:** A function that describes the likelihood of the possible outcomes of a random event

**Random experiment:** A process whose outcome cannot be predicted with certainty

**Random variable:** A variable that represents the values for the possible outcomes of a random event

**Sample space:** The set of all possible values for a random variable

**Standard deviation:** A statistic that calculates the typical distance of a data point from the mean of a dataset

**Standardization:** The process of putting different variables on the same scale

**Subjective probability:** A type of probability based on personal feelings, experience, or judgment

**Z-score:** A measure of how many standard deviations below or above the population mean a data point is

**Terms and their definitions from the previous module**

# A

**A/B testing:** A way to compare two versions of something to find out which version performs better

# C

**Confidence interval:** A range of values that describes the uncertainty surrounding an estimate

# E

**Econometrics:** A branch of economics that uses statistics to analyze economic problems

# I

**Inferential statistics:** A type of statistics that uses sample data to draw conclusions about a larger population

**Interquartile range:** The distance between the first quartile (Q1) and the third quartile (Q3)

# L

**Literacy rate:** The percentage of the population in a given age group that can read and write

# M

**Mean:** The average value in a dataset

**Measure of central tendency:** A value that represents the center of a dataset

**Measure of dispersion:** A value that represents the spread of a dataset, or the amount of variation in data points

**Measure of position:** A method by which the position of a value in relation to other values in a dataset is determined

**Median:** The middle value in a dataset

**Mode:** The most frequently occurring value in a dataset

## **P**

**Parameter:** A characteristic of a population

**Percentile:** The value below which a percentage of data falls

**Population:** Every possible element that a data professional is interested in measuring

## **Q**

**Quartile:** A value that divides a dataset into four equal parts

## **R**

**Range:** The difference between the largest and smallest value in a dataset

**Representative sample:** A sample that accurately reflects the characteristics of a population

## S

**Sample :** A subset of a population

**Sampling:** The process of selecting a subset of data from a population

**Standard deviation:** A calculation that measures how spread out values are from the mean of a dataset

**Statistic:** A characteristic of a sample

**Statistical significance:** The claim that the results of a test or experiment are not explainable by chance alone

**Statistics:** The study of the collection, analysis, and interpretation of data

**Summary statistics:** A method that summarizes data using a single number

## V

**Variance:** The average of the squared difference of each

data point from the mean