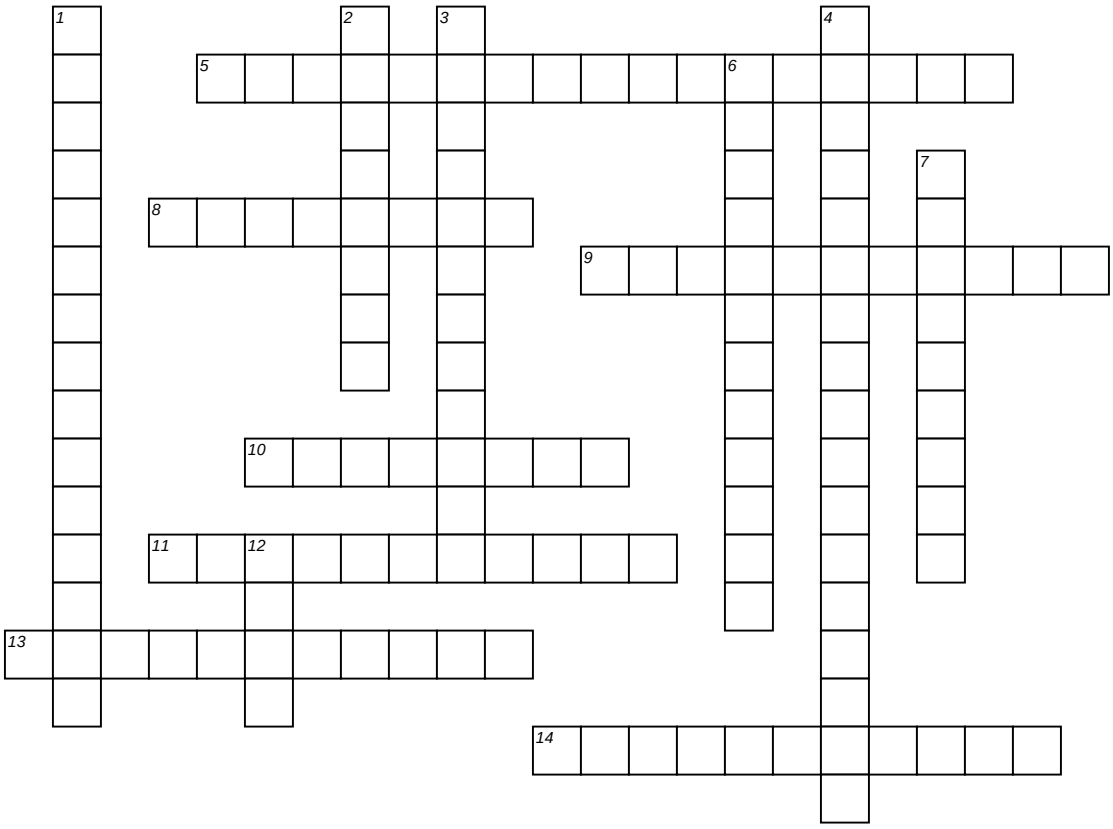
Displaying & Describing Categorical Data

Advanced Placement Statistics



Stats: Modeling the World, Chapters 2-3

ACROSS

5 table that lists the categories of a variable and gives the proportion of observations for each category

8 shows how a "whole" divides into categories by showing a wedge of a circle whose area corresponds to the proportion in each category

9 table that displays counts and, sometimes, percentages of individuals falling into named categories on two or more variables

10 in a contingency table, the distribution of either variable alone

11 relationship between two variables such that the conditional distribution of one variable is the same for each category of the other.

13 the distribution of a variable when considering only a smaller group of individuals

14 variable that describes data using words or numerals as labels

DOWN

1 phenomenon where when averages are taken across different groups, they can appear to contradict the overall averages

2 shows a bar representing the count of each category in a categorical variable

3 the possible values of the variable and the relative frequency of each value

4 shows bars divided proportionally into segments corresponding to the percentage in each group

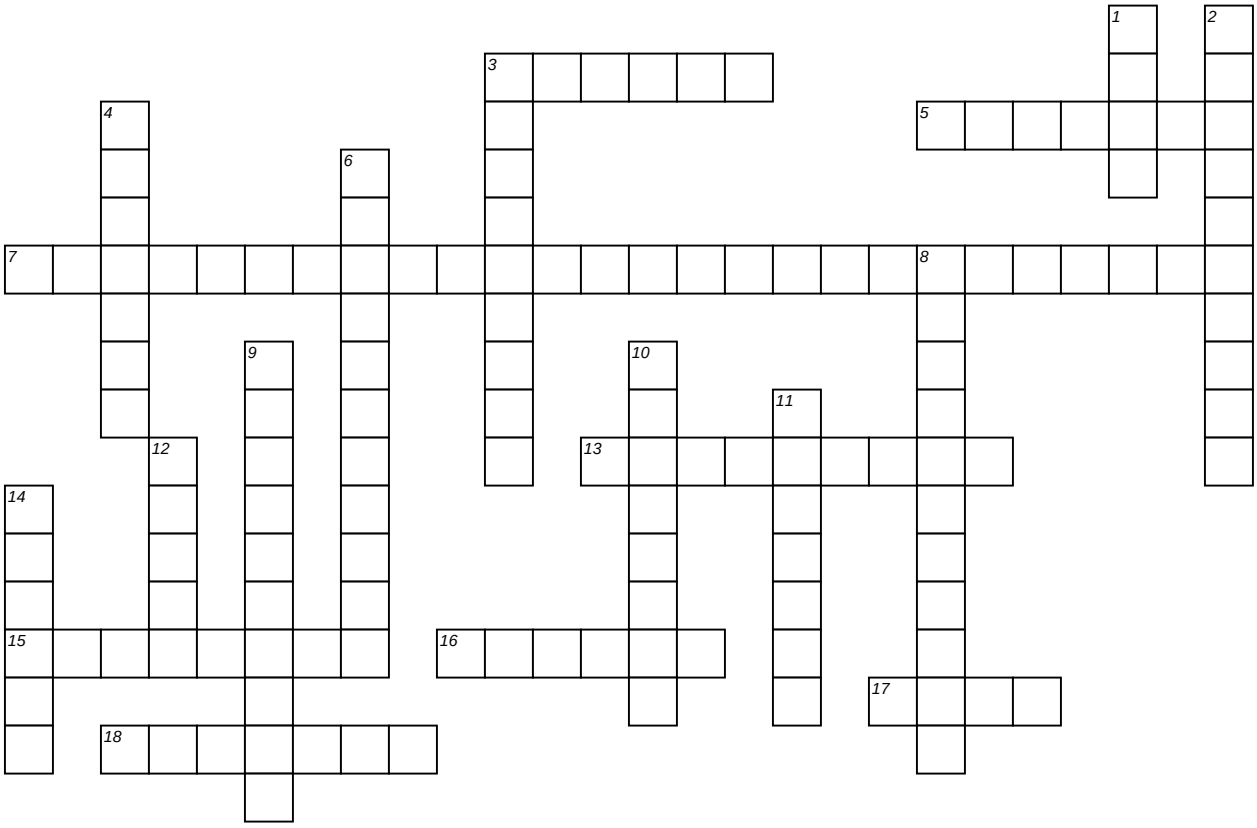
6 variable that describes data using numbers as numerical values

7 table that lists the categories of a variable and gives the count of observations for each category

12 systematically recorded information, together with its context

Displaying Quantitative Data

Advanced Placement Statistics



Stats: Modeling the World, Chapter 4

ACROSS

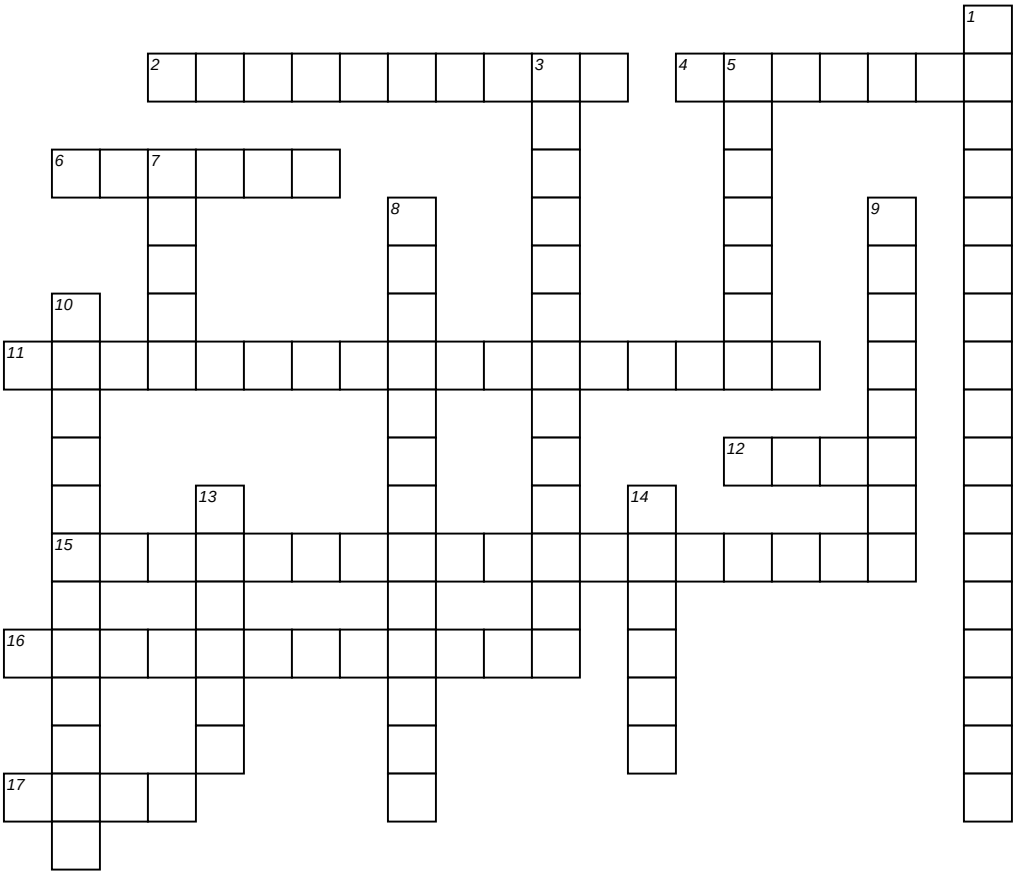
3 when a distribution is not symmetric and one tail stretches out farther than the other
5 distributions with two modes
7 uses adjacent bars to show the distribution of values in a quantitative variable, where each bar represents the proportion of values falling in an interval
13 uses adjacent bars to show the distribution of values in a quantitative variable, where each bar represents the number of values falling in an interval
15 used to display data that change over time
16 a numerical summary of how tightly the values are clustered around the "center"
17 the parts of a distribution that typically trail off on either side
18 a distribution roughly flat in shape

DOWN

1 a hump or high point in the shape of the distribution of a variable
2 distributions with more than two modes
3 shape where the two halves on either side of the center look approximately like mirror images of each other
4 an extreme value that doesn't appear to belong with the rest of the data
6 shape where the longer tail stretches to the right
8 type of display that shows quantitative data values in a way that shows the shape of the distribution in addition to individual data values
9 shape where the longer tail stretches to the left
10 having one mode
11 graphs a dot for each case against a single axis
12 reveals single vs. multiple modes and symmetry vs. skewness
14 a value that summarizes the entire distribution with a single number, a "typical" value

Describing Distributions Numerically

Advanced Placement Statistics



Stats: Modeling the World, Chapter 5

ACROSS

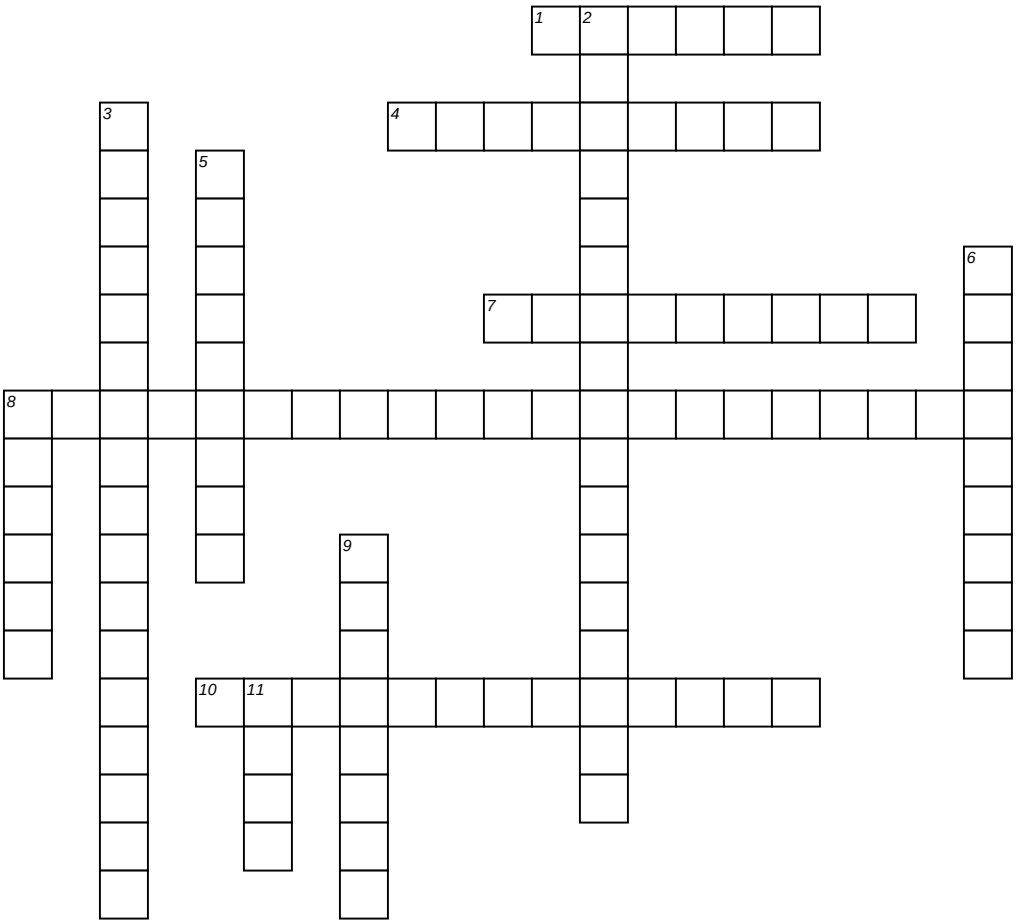
2 the number that falls above a given % of the data
4 displays the 5-number summary as a central box with the whiskers that extend to the non-outlying data values.
6 a numerical summary of how tightly the values are clustered around the "center"
11 minimum, lower quartile, median, upper quartile, maximum
12 found by summing all the data values and dividing by the count
15 the difference between the first and third quartiles
16 variable that describes data using numbers as numerical values
17 the value that occurs most frequently

DOWN

1 the square root of the variance
3 the value with a quarter of the data below it
5 value more than 1.5 times the IQR below Q1 or above Q3
7 difference between the maximum value and the minimum value
8 the value with a quarter of the data above it
9 the sum of the squared deviations from the mean, divided by the count minus one
10 the possible values of the variable and the relative frequency of each value
13 a value that summarizes the entire distribution with a single number, a "typical" value
14 the middle value of a distribution with half the data above and half the data below it

Standard Deviation & the Normal Model

Advanced Placement Statistics



Stats: Modeling the World, Chapter 6

ACROSS

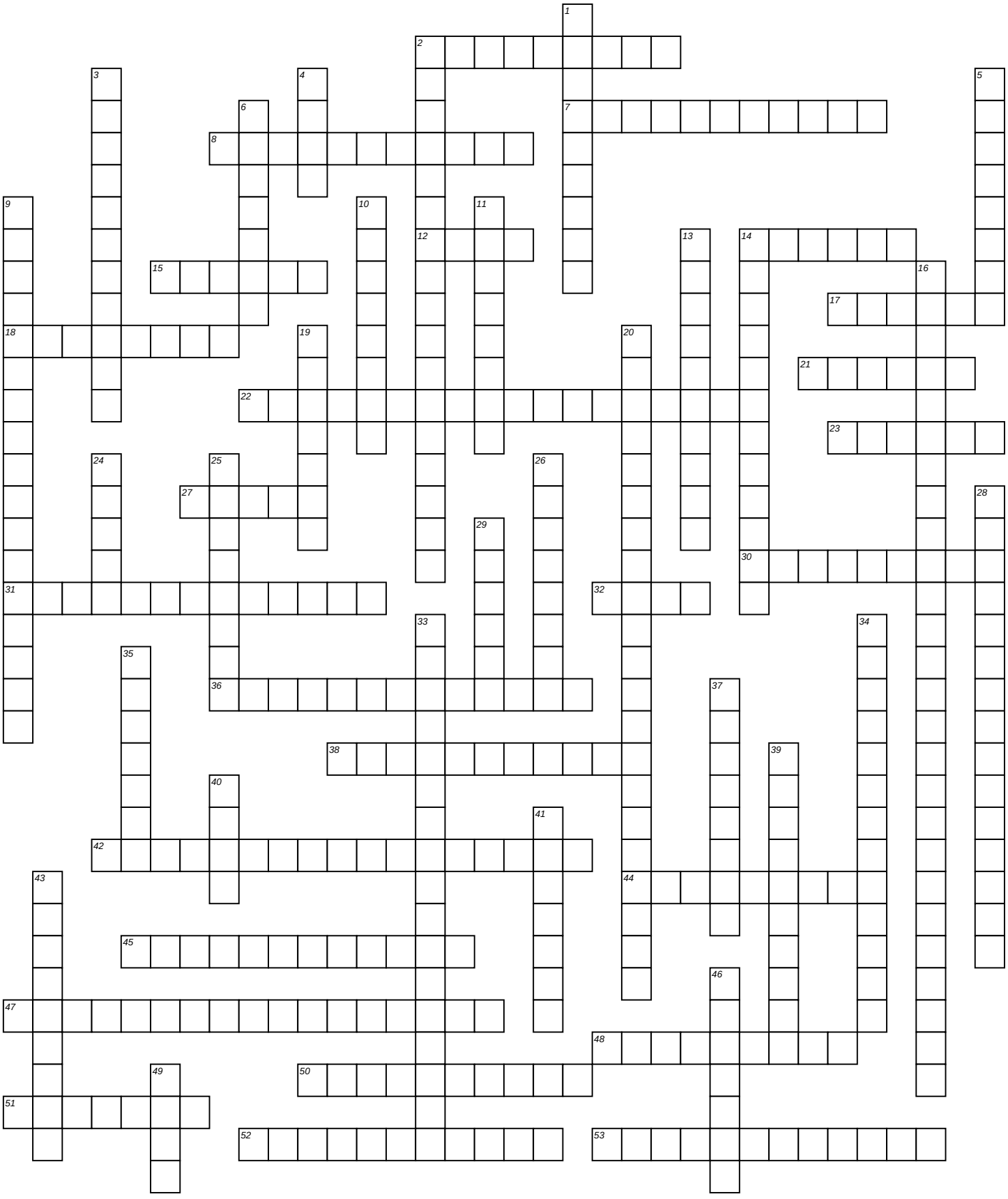
1 tells how many standard deviations a value is from the mean
4 the process of multiplying each value by a constant that multiplies both the measures of position and measures of spread by that constant
7 numerical attribute of a model
8 display to help assess whether a distribution of data is approximately Normal
10 in a Normal Model, about 68% of the values fall within 1 standard deviation of the mean, about 95% within 2 standard deviations, and about 99.7% within 3 standard deviations

DOWN

2 the value found by subtracting the mean and dividing by the standard deviation
3 the square root of the variance
5 type of Normal model with mean 0 and standard deviation 1
6 numerical attribute of a set of data
8 model used for certain unimodal, symmetric distributions
9 the sum of the squared deviations from the mean, divided by the count minus one
11 center of the Normal model

Exploring and Understanding Data

Advanced Placement Statistics



Stats: Modeling the World, Chapters 1-6

ACROSS

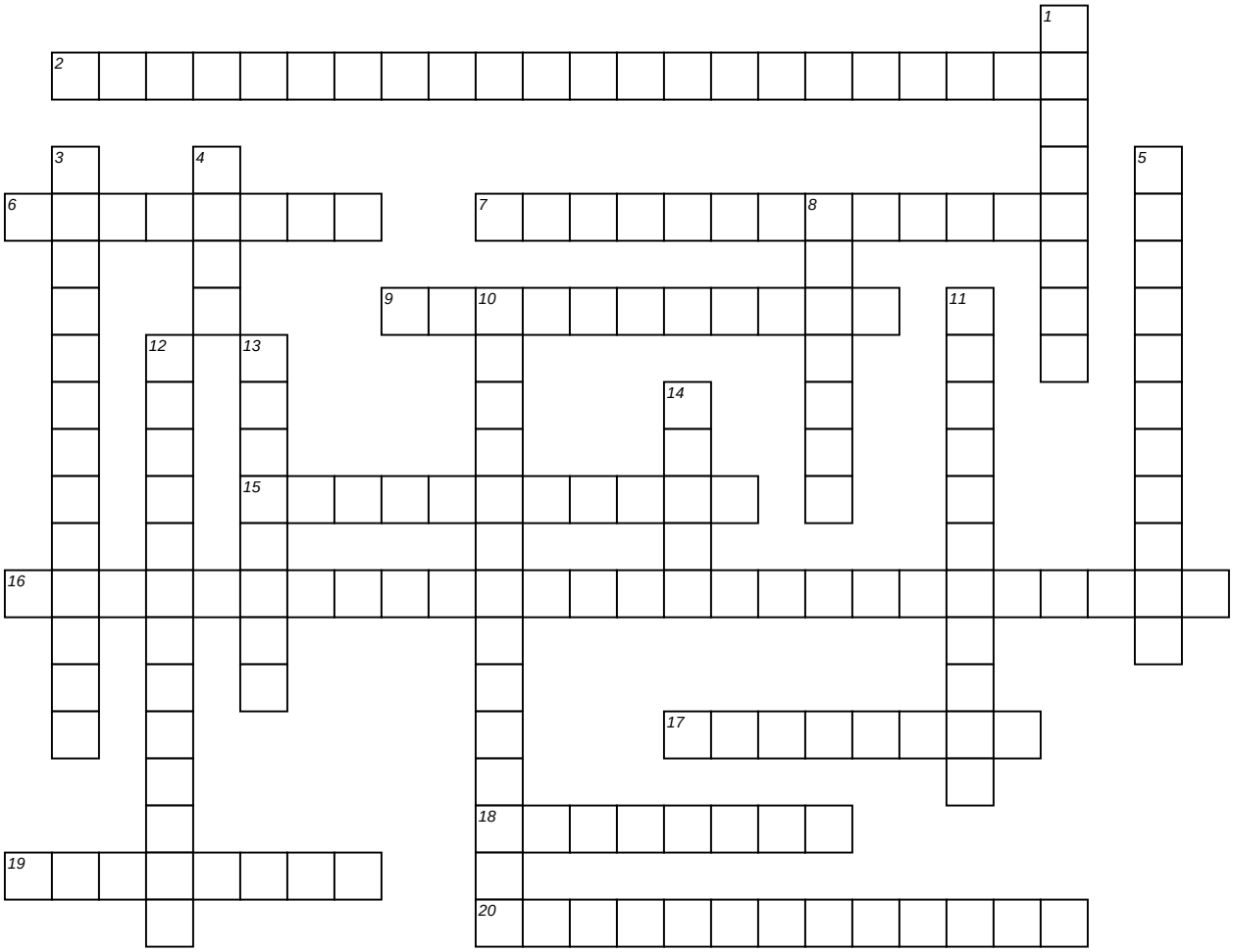
2 table that lists the categories of a variable and gives the count of observations for each category
7 variable that describes data using words or numerals as labels
8 table that displays counts and, sometimes, percentages of individuals falling into named categories on two or more variables
12 found by summing all the data values and dividing by the count
14 when a distribution is not symmetric and one tail stretches out farther than the other
15 tells how many standard deviations a value is from the mean
17 a numerical summary of how tightly the values are clustered around the "center"
18 used to display data that change over time
21 model used for certain unimodal, symmetric distributions
22 the difference between the first and third quartiles
23 the middle value of a distribution with half the data above and half the data below it
27 difference between the maximum value and the minimum value
30 uses adjacent bars to show the distribution of values in a quantitative variable, where each bar represents the number of values falling in an interval
31 the value with a quarter of the data above it
32 a hump or high point in the shape of the distribution of a variable
36 in a Normal Model, about 68% of the values fall within 1 standard deviation of the mean, about 95% within 2 standard deviations, and about 99.7% within 3 standard deviations
38 the distribution of a variable when considering only a smaller group of individuals
42 the value found by subtracting the mean and dividing by the standard deviation
44 numerical attribute of a model
45 variable that describes data using numbers as numerical values
47 shows bars divided proportionally into segments corresponding to the percentage in each group
48 numerical attribute of a set of data
50 distributions with more than two modes
51 distributions with two modes
52 relationship between two variables such that the conditional distribution of one variable is the same for each category of the other.
53 the possible values of the variable and the relative frequency of each value

DOWN

1 the process of multiplying each value by a constant that multiplies both the measures of position and measures of spread by that constant
2 minimum, lower quartile, median, upper quartile, maximum
3 type of display that shows quantitative data values in a way that shows the shape of the distribution in addition to individual data values
4 systematically recorded information, together with its context
5 Normal model with mean 0 and standard deviation 1
6 graphs a dot for each case against a single axis
9 table that lists the categories of a variable and gives the proportion of observations for each category
10 shows how a "whole" divides into categories by showing a wedge of a circle whose area corresponds to the proportion in each category
11 shows a bar representing the count of each category in a categorical variable
13 the number that falls above a given % of the data
14 shape where the longer tail stretches to the right
16 uses adjacent bars to show the distribution of values in a quantitative variable, where each bar represents the proportion of values falling in an interval
19 an extreme value that doesn't appear to belong with the rest of the data
20 display to help assess whether a distribution of data is approximately Normal
24 reveals single vs. multiple modes and symmetry vs. skewness
25 the sum of the squared deviations from the mean, divided by the count minus one
26 in a contingency table, the distribution of either variable alone
28 phenomenon where when averages are taken across different groups, they can appear to contradict the overall averages
29 a value that summarizes the entire distribution with a single number, a "typical" value
33 the square root of the variance
34 the value with a quarter of the data below it
35 displays the 5-number summary as a central box with the whiskers that extend to the non-outlying data values.
37 having one mode
39 shape where the longer tail stretches to the left
40 the value that occurs most frequently
41 value more than 1.5 times the IQR below Q1 or above Q3
43 shape where the two halves on either side of the center look approximately like mirror images of each other
46 a distribution that's roughly flat in shape
49 the parts of a distribution that typically trail off on either side

Exploring Relationships Between Variables

Advanced Placement Statistics



Stats: Modeling the World, Chapters 7-10

ACROSS

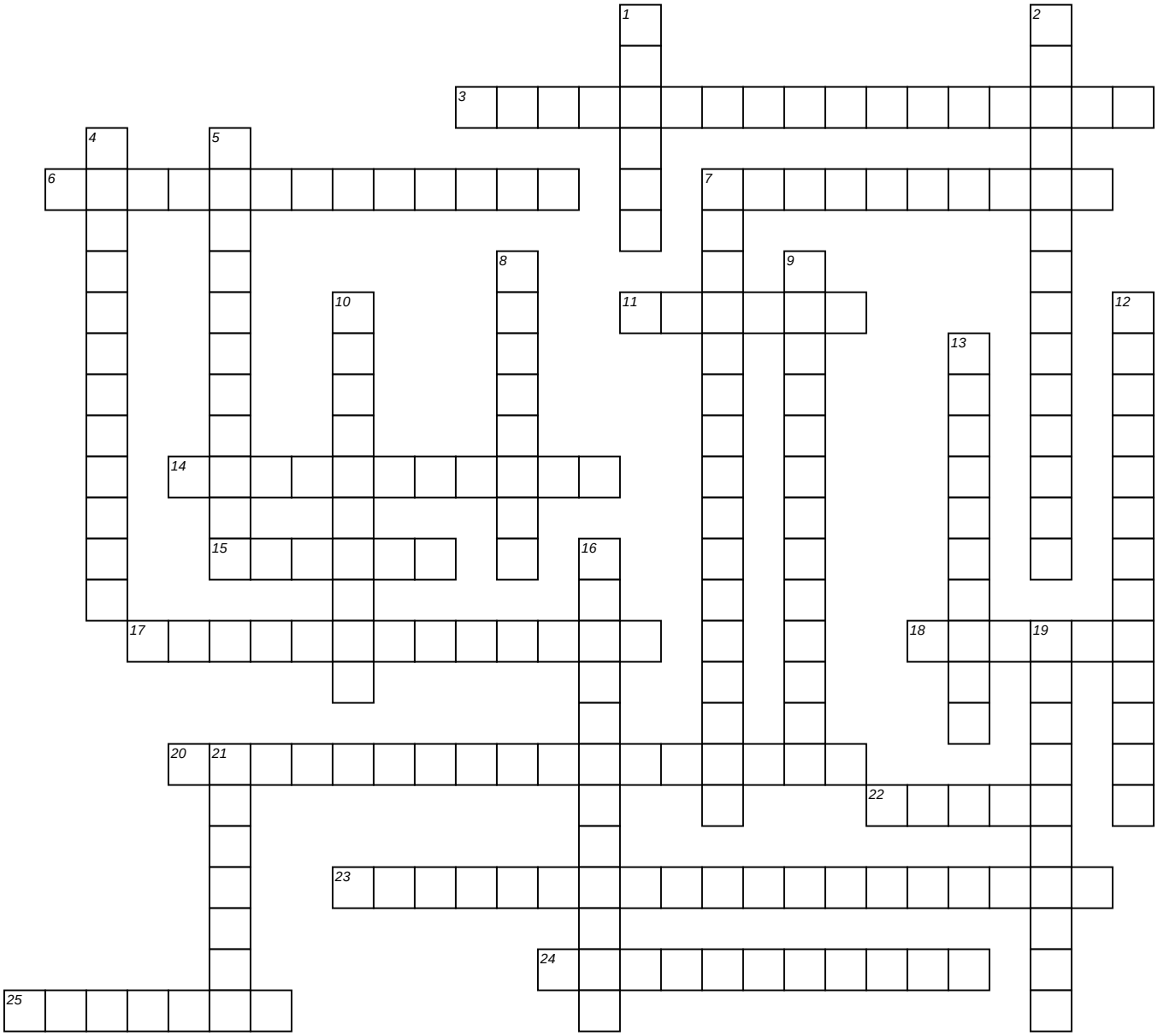
2 numerical measure of the direction and strength of a linear association
6 variable that you hope to predict or explain
7 predicting for values of x within the ones used to find the linear model equation
9 variable that accounts for, explains, predicts, or is otherwise responsible for the y-variable
15 relationship between two quantitative variables
16 overall measure of how successful the regression is in linearly relating y to x
17 the difference between the actual data value and the corresponding value predicted by a model
18 data points whose x-values are far from the mean of x have a high amount of this
19 type of association where as one variable increases, so does the other
20 predicting for values of x far from the ones used to find the linear model equation

DOWN

1 general measure of scatter around the underlying relationship between two quantitative variables
3 point on the scatterplot representing the mean x-value and mean y-value
4 shape of a scatterplot
5 shows the relationship between two quantitative variables measured on the same cases
8 a variable that is not explicitly part of a model but affects the way the variables in the model appear to be related
10 y-hat
11 point that when omitted, results in a very different regression model
12 least squares regression line
13 type of association where an increases in one variable generally correspond to decreases in the other
14 measures the change in the y-value per unit change in x-value

Randomness & Sample Surveys

Advanced Placement Statistics



Stats: Modeling the World, Chapters 11-12

ACROSS

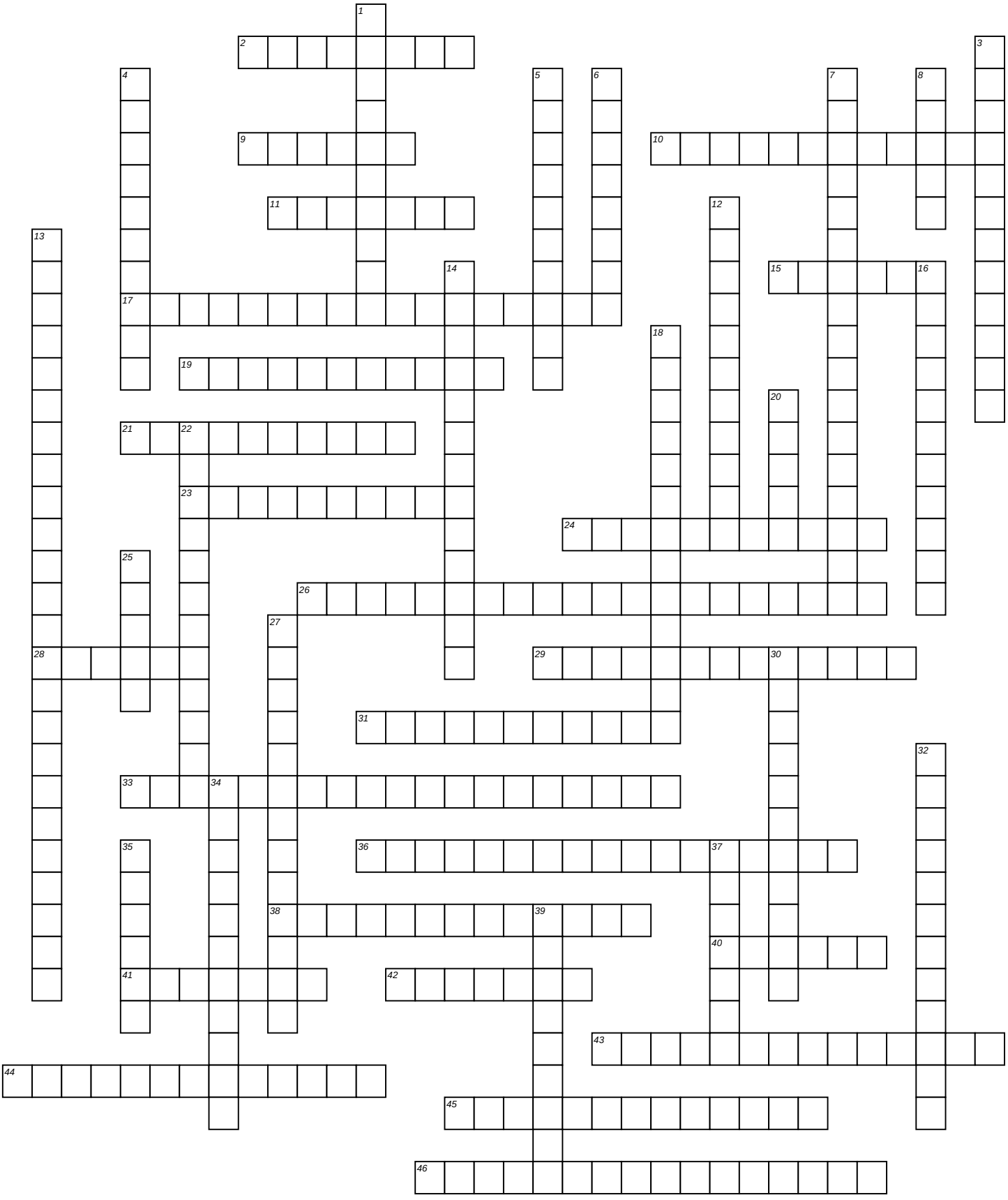
3 type of bias that is problematic because those who volunteer tend to have strong negative opinions
6 process by which each individual is given a fair chance of selection
7 models random events by using random numbers to specify outcomes with relative frequencies that correspond to the true real-world relative frequencies we are trying to model
11 any systematic failure of a sampling method to represent its population
14 type of bias that is problematic because the intended sample is incomplete
15 a subdivision of the population
17 a list of individuals from whom the sample is drawn
18 a (representative) subset of a population, examined in hope of learning about a population
20 sampling design where individuals can choose on their own whether to participate in the sample
22 the sequence of several components representing events that we are pretending will take place
23 the natural tendency of randomly drawn samples to differ, one from another
24 sampling design where individuals are chosen based on who is easily available
25 sampling design in which entire groups are chosen at random

DOWN

1 a sample that consists of the entire population
2 an occurance for which we know what outcomes could happen, but not which particular values will happen
4 a study that asks questions of a sample drawn from some population in the hope of learning something about the entire population
5 similar in makeup
7 sampling design in which the population is divided into several strata, and random samples are then drawn from each stratum
8 type of bias that is problematic because false information may be given
9 not similar in makeup
10 sample drawn by selecting individuals systematically from a sampling frame
12 type of bias that is problematic because some groups are not represented in the sample
13 sampling schemes that combine several sampling methods
16 sampling design in which each set of n elements in the population has an equal chance of selection.
19 the entire group of individuals or instances about whom we hope to learn
21 an individual result of a component of a simulation

Gathering Data

Advanced Placement Statistics



Stats: Modeling the World, Chapters 11-13

ACROSS

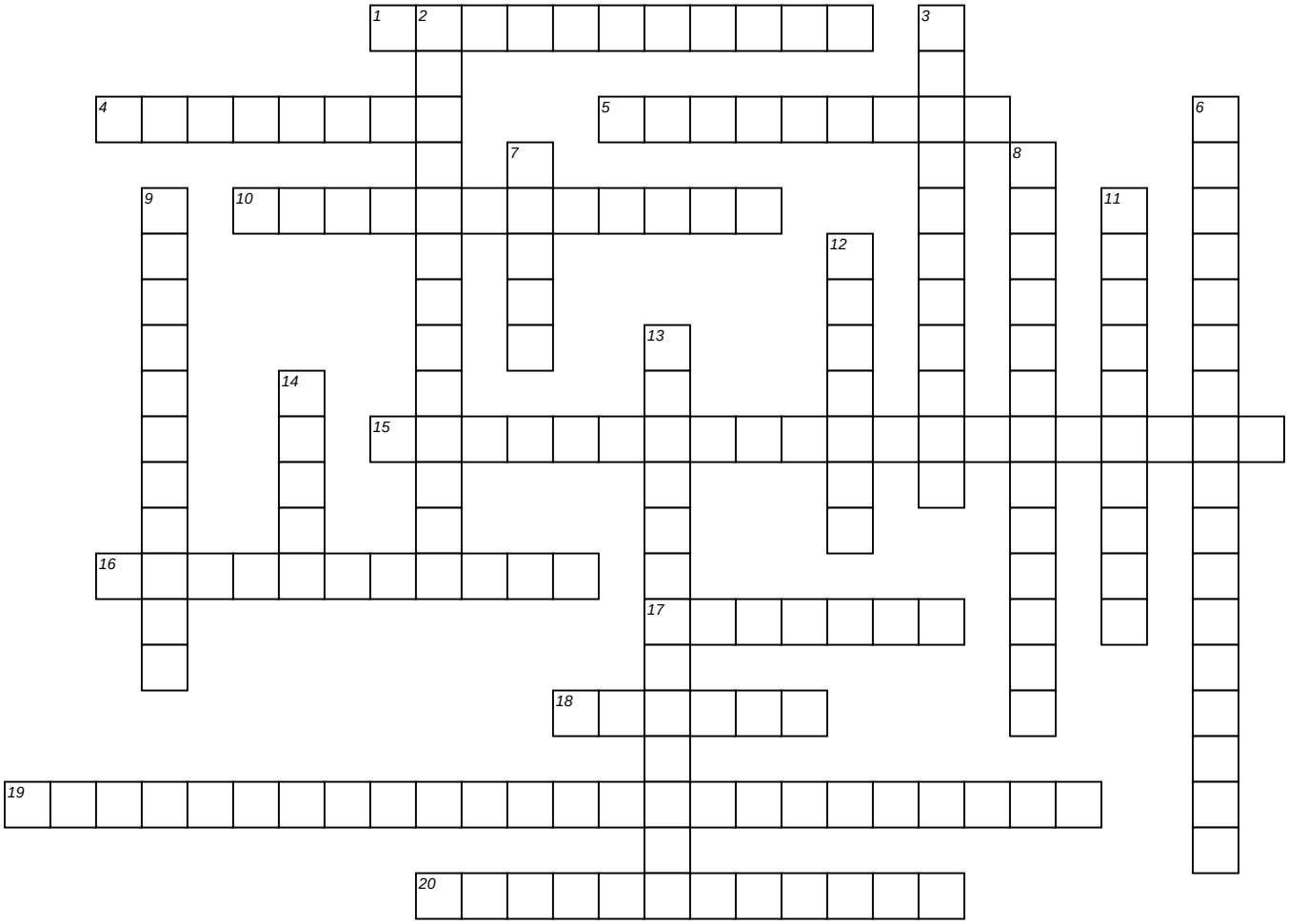
2 type of bias that is problematic because false information may be given
9 a (representative) subset of a population, examined in hope of learning about a population
10 experimental units assigned to a baseline treatment level
11 aspects of the experiment that we know may have an effect on the response, but that are not the factors being studied
15 any systematic failure of a sampling method to represent its population
17 individuals on which an experiment is done
19 sampling design where individuals are chosen based on who is easily available
21 sample drawn by selecting individuals systematically from a sampling frame
23 sampling schemes that combine several sampling methods
24 when either the subjects or the people who have contact with them do not know which treatment a subject has received
26 type of experiment in which all experimental units have an equal chance of receiving any treatment
28 a subdivision of the population
29 observational study in which subjects are selected and then their previous conditions or behaviors are determined
31 similar in makeup
33 the natural tendency of randomly drawn samples to differ, one from another
36 sampling design where individuals can choose on their own whether to participate in the sample
38 process by which each individual is given a fair chance of selection
40 a sample that consists of the entire population
41 an individual result of a component of a simulation
42 sampling design in which entire groups are chosen at random
43 an occurance for which we know what outcomes could happen, but not which particular values will happen
44 type of bias that is problematic because some groups are not represented in the sample
45 study based on data in which no treatments have been assigned to subjects
46 sampling design in which the population is divided into several strata, and random samples are then drawn from each stratum

DOWN

1 the entire group of individuals or instances about whom we hope to learn
3 a study that asks questions of a sample drawn from some population in the hope of learning something about the entire population
4 study in which subjects are randomly assigned to treatments
5 models random events by using random numbers to specify outcomes with relative frequencies that correspond to the true real-world relative frequencies we are trying to model
6 people who are studied
7 type of bias that is problematic because those who volunteer tend to have strong negative opinions
8 when groups of experimental units are similar, they are gathered into these groups
12 type of bias that is problematic because the intended sample is incomplete
13 when an observed difference is too large to believe that it is likely to have occurred naturally
14 the tendency of many human subjects to show a response even when administered a placebo
16 neither the subjects nor the people who have contact with them know which treatment a subject has received
18 not similar in makeup
20 specific values that the experimenter chooses for a factor
22 sampling design in which each set of n elements in the population has an equal chance of selection.
25 the sequence of several components representing events that we are pretending will take place
27 a list of individuals from whom the sample is drawn
30 when the levels of one factor are associated with the levels of another factor so their effects cannot be separated
32 type of study in which subjects who are similar in ways not under study may be grouped together and then compared with each other on the variables of interest
34 observational study in which subjects are followed to observe future outcomes
35 variable whose levels are controlled by the experimenter
37 treatment known to have no effect, administered so that all groups experience the same conditions
39 the process or intervention applied to randomly assigned experimental units

Experiments & Observational Studies

Advanced Placement Statistics



Stats: Modeling the World, Chapter 13

ACROSS

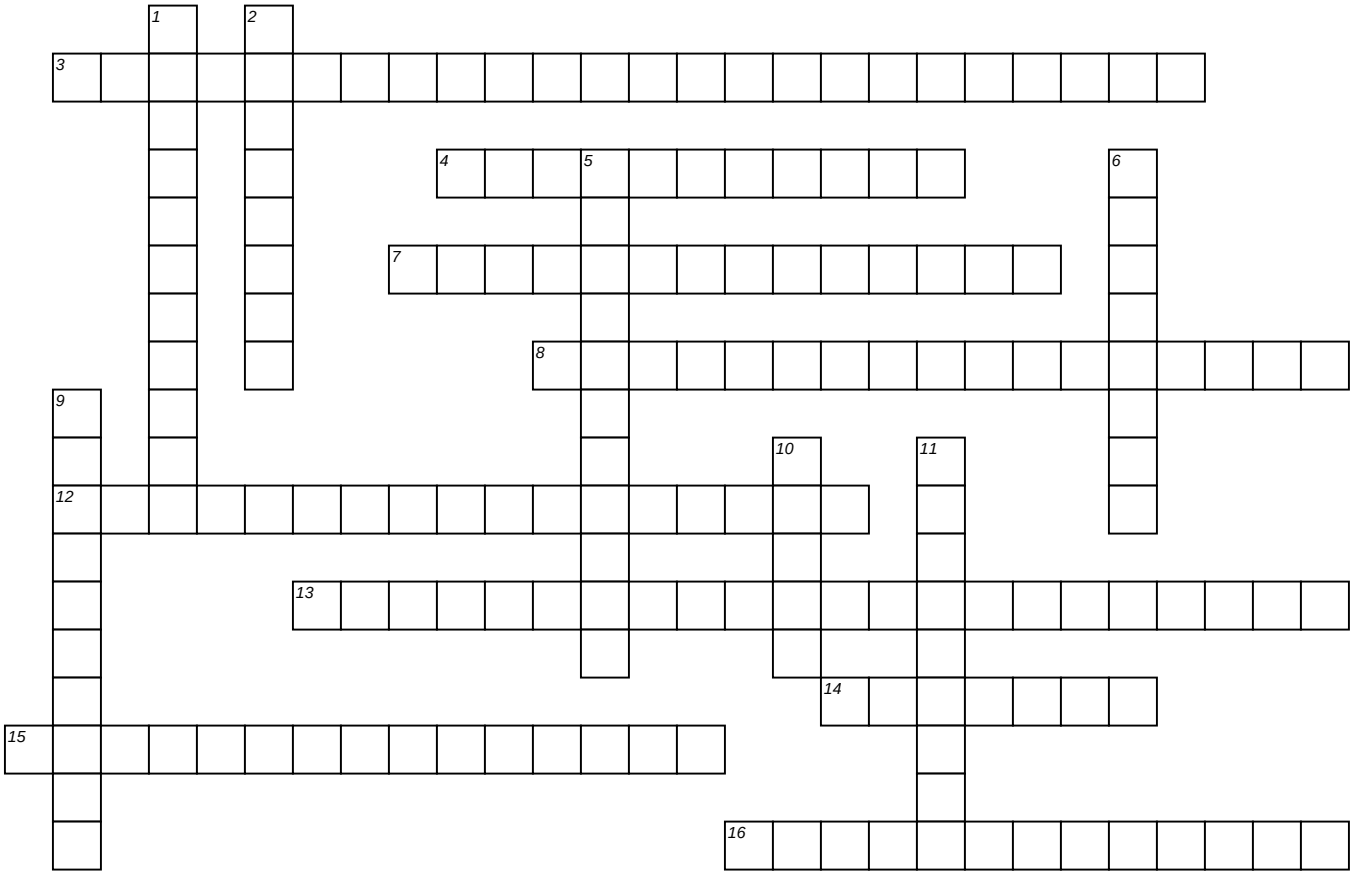
1 neither the subjects nor the people who have contact with them know which treatment a subject has received
4 people who are studied
5 the process or intervention applied to randomly assigned experimental units
10 experimental units assigned to a baseline treatment level
15 type of experiment in which all experimental units have an equal chance of receiving any treatment
16 when either the subjects or the people who have contact with them do not know which treatment a subject has received
17 treatment known to have no effect, administered so that all groups experience the same conditions
18 variable whose levels are controlled by the experimenter
19 when an observed difference is too large to believe that it is likely to have occurred naturally
20 type of study in which subjects who are similar in ways not under study may be grouped together and then compared with each other on the variables of interest

DOWN

2 study based on data in which no treatments have been assigned to subjects
3 when the levels of one factor are associated with the levels of another factor so their effects cannot be separated
6 individuals on which an experiment is done
7 when groups of experimental units are similar, they are gathered into these groups
8 the tendency of many human subjects to show a response even when administered a placebo
9 observational study in which subjects are followed to observe future outcomes
11 study in which subjects are randomly assigned to treatments
12 aspects of the experiment that we know may have an effect on the response, but that are not the factors being studied
13 observational study in which subjects are selected and then their previous conditions or behaviors are determined
14 specific values that the experimenter chooses for a factor

Randomness & Probability

Advanced Placement Statistics



Stats: Modeling the World, Chapters 14-17

ACROSS

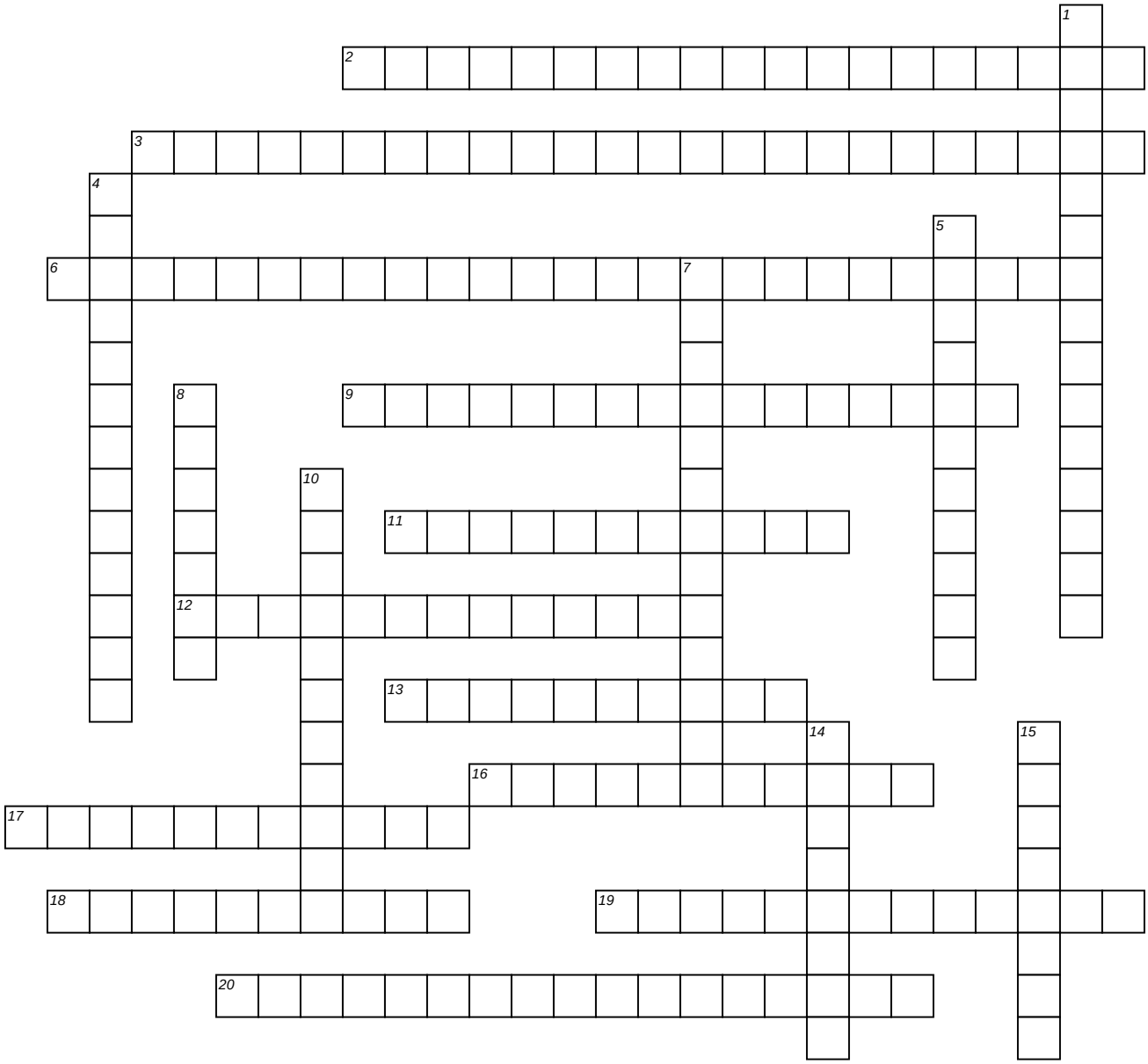
3 random variable that can take any numeric value within a range of values
4 collection of all possible outcome values
7 variable that assumes any of several different values as a result of some random event
8 the long-run relative frequency of repeated independent events settles down the true relative frequency as the number of trials increases
12 disjoint
13 random variable that can take one of a finite number of a distinct outcome
14 value measured, observed, or reported for an individual instance for a trial
15 situation where there are two possible outcomes, the probability of success is constant, and the trials are independent
16 the theoretical long-run average value

DOWN

1 relationship between events if knowing one event occurs does not alter the probability that the other event occurs
2 relationship between two events if they share mo outcomes in common
5 a number between 0 and 1 that reports the likelihood of an event's occurrence
6 probability model appropriate for a random variable that counts the number of successes in a fixed number of Bernoulli trials
9 set of outcomes that are not in the specified event
10 collection of outcomes
11 probability model appropriate for a random variable that counts the number of Bernoulli Trials until the first success.

Collecting and Organizing Data

Advanced Placement Statistics



Stats: Modeling the World, Chapters 11-18

ACROSS

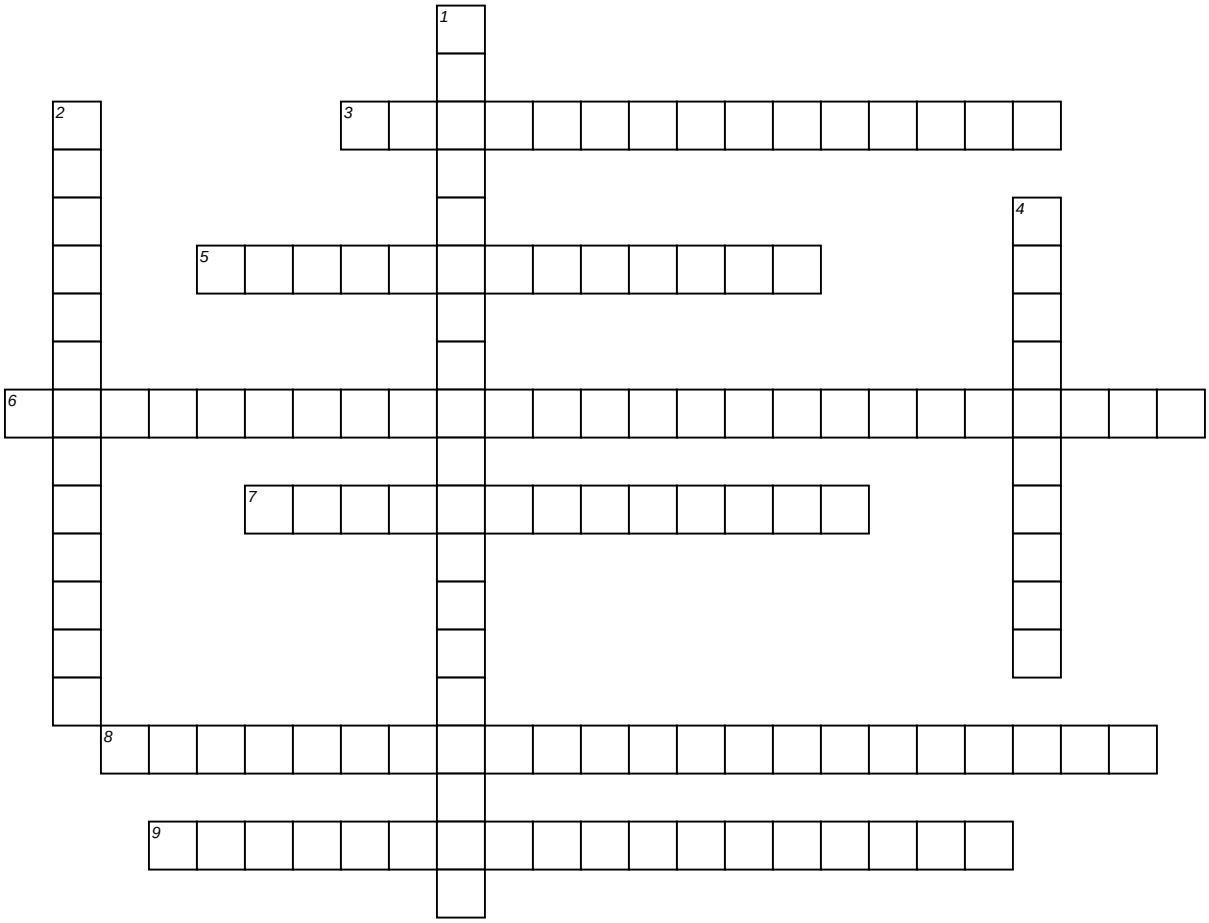
2 states that the sampling distribution model of the sample mean (and proportion) is approximately Normal for large n, regardless of the distribution of the population, as long as the observations are independent
3 when an observed difference is too large to believe that it is likely to have occurred naturally
6 shows the behavior of the statistic over all the possible samples for the same size n
9 sampling design in which the population is divided into several strata, and random samples are then drawn from each stratum
11 type of bias that is problematic because the intended sample is incomplete
12 the theoretical long-run average value
13 sampling schemes that combine several sampling methods
16 neither the subjects nor the people who have contact with them know which treatment a subject has received
17 similar in makeup
18 number of successes out of a whole
19 not similar in makeup
20 type of bias that is problematic because those who volunteer tend to have strong negative opinions

DOWN

1 situation where there are two possible outcomes, the probability of success is constant, and the trials are independent
4 when we estimate the standard deviation of a sampling distribution using statistics found from the data
5 when the levels of one factor are associated with the levels of another factor so their effects cannot be separated
7 type of bias that is problematic because some groups are not represented in the sample
8 sampling design in which entire groups are chosen at random
10 relationship between events if knowing one event occurs does not alter the probability that the other event occurs
14 relationship between two events if they share no outcomes in common
15 type of bias that is problematic because false information may be given

Confidence Intervals for Proportions

Advanced Placement Statistics



Stats: Modeling the World, Chapters 18-19

ACROSS

3 proportion of all samples of size n that will produce a statistic within a corresponding margin of error of the true parameter

5 when we estimate the standard deviation of a sampling distribution using statistics found from the data

6 shows the behavior of the statistic over all the possible samples for the same size n

7 extent of the interval on either side of the observed statistic value

8 confidence interval for the true value of a proportion

9 an interval of values usually of the form estimate ± margin of error

DOWN

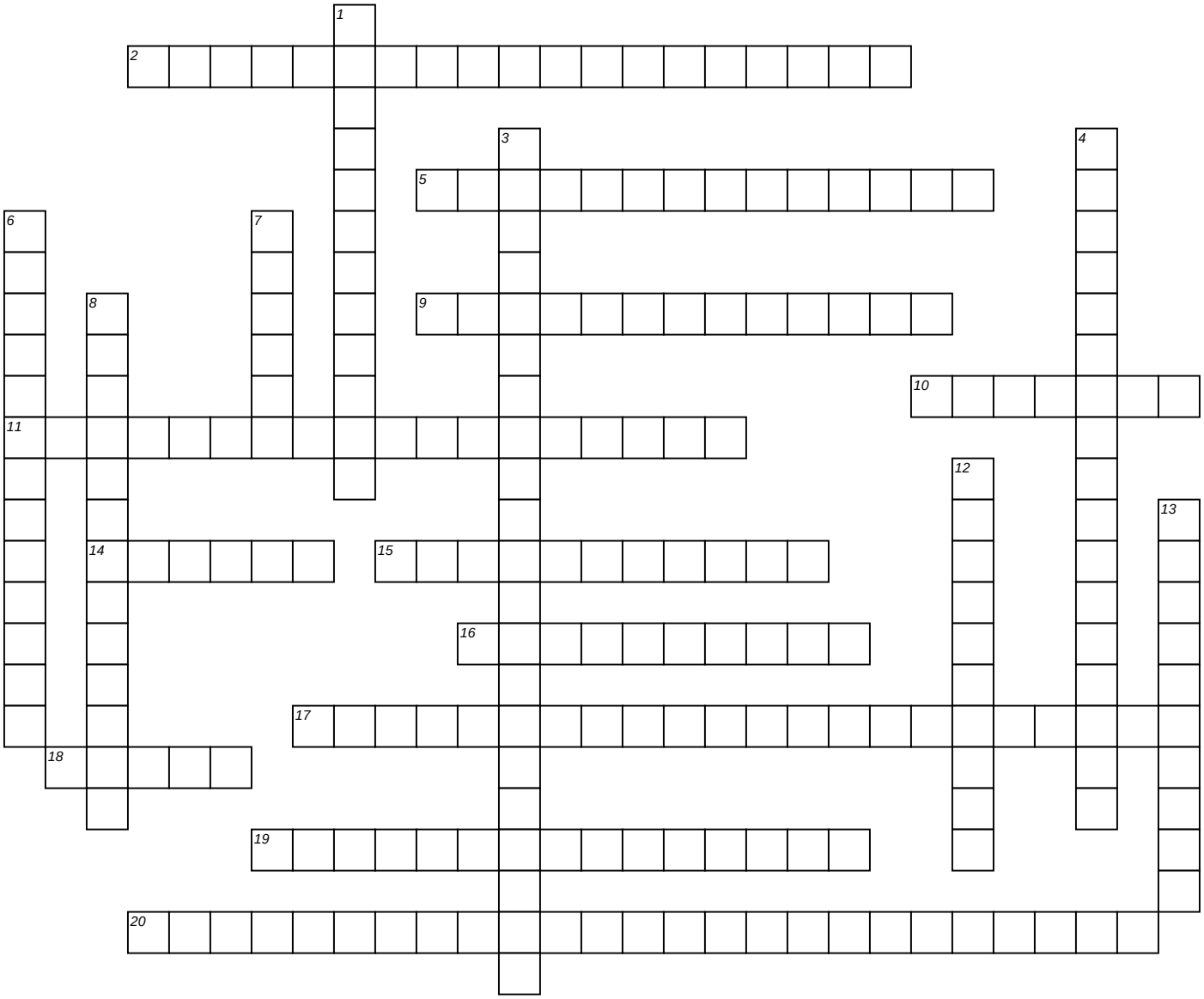
1 states that the sampling distribution model of the sample mean (and proportion) is approximately Normal for large n, regardless of the distribution of the population, as long as the observations are independent

2 number of standard errors to move away from the mean of the sampling distribution to correspond to the specified level of confidence

4 number of successes out of a whole

Inferences About Proportions

Advanced Placement Statistics



Stats: Modeling the World, Chapters 18-22

ACROSS

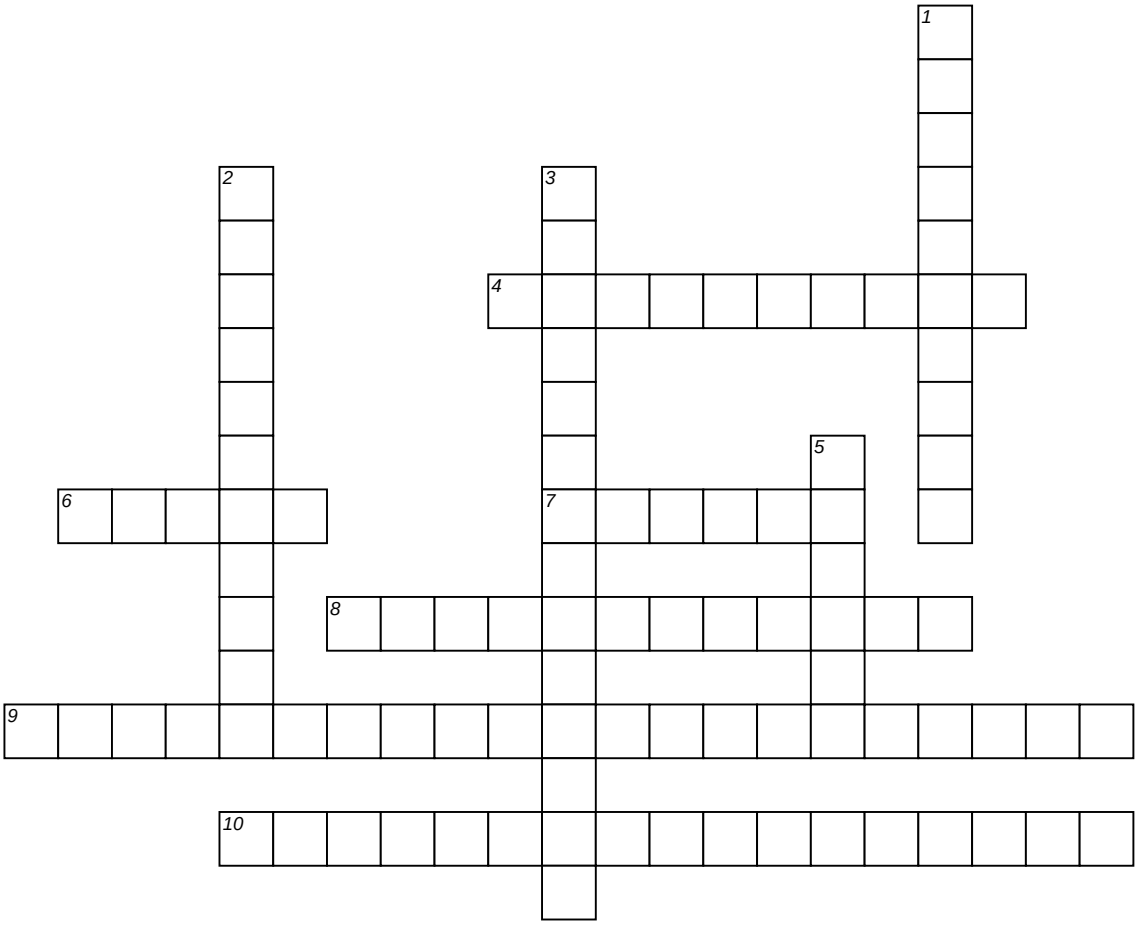
2 states that the sampling distribution model of the sample mean (and proportion) is approximately Normal for large n, regardless of the distribution of the population, as long as the observations are independent
5 claim that specifies a value for some population parameter that can form the basis for assuming a sampling distribution for a test statistic
9 extent of the interval on either side of the observed statistic value
10 combining groups when we have data from different sources that we believe are homogeneous
11 an interval of values usually of the form estimate ± margin of error
14 decision made when the p-value is too small to believe that the statistic could have occurred due to chance variation
15 error of failing to reject a null hypothesis when in fact it is false (also called a "false positive")
16 error of rejecting a null hypothesis when in fact it is true (also called a "false positive")
17 confidence interval for the true value of a proportion
18 probability that a hypothesis test will correctly reject a false null hypothesis
19 proportion of all samples of size n that will produce a statistic within a corresponding margin of error of the true parameter
20 shows the behavior of the statistic over all the possible samples for the same size n

DOWN

1 decision made when the p-value is large enough to believe that the statistic could have occurred due to chance variation
3 proposes what we should conclude if we find the null hypothesis to be unlikely
4 alpha level
6 number of standard errors to move away from the mean of the sampling distribution to correspond to the specified level of confidence
7 the probability of observing a value for a test statistic at least as far from the hypothesized value as the statistic actually observed if the null hypothesis is true
8 when we estimate the standard deviation of a sampling distribution using statistics found from the data
12 number of successes out of a whole
13 threshold P-value that determines when we reject a null hypothesis

Testing Hypotheses about Proportions

Advanced Placement Statistics



Stats: Modeling the World, Chapters 20-21

ACROSS

4 threshold P-value that determines when we reject a null hypothesis

6 probability that a hypothesis test will correctly reject a false null hypothesis

7 the probability of observing a value for a test statistic at least as far from the hypothesized value as the statistic actually observed if the null hypothesis is true

8 decision made when the p-value is large enough to believe that the statistic could have occurred due to chance variation

9 proposes what we should conclude if we find the null hypothesis to be unlikely

10 alpha level

DOWN

1 error of rejecting a null hypothesis when in fact it is true (also called a "false positive")

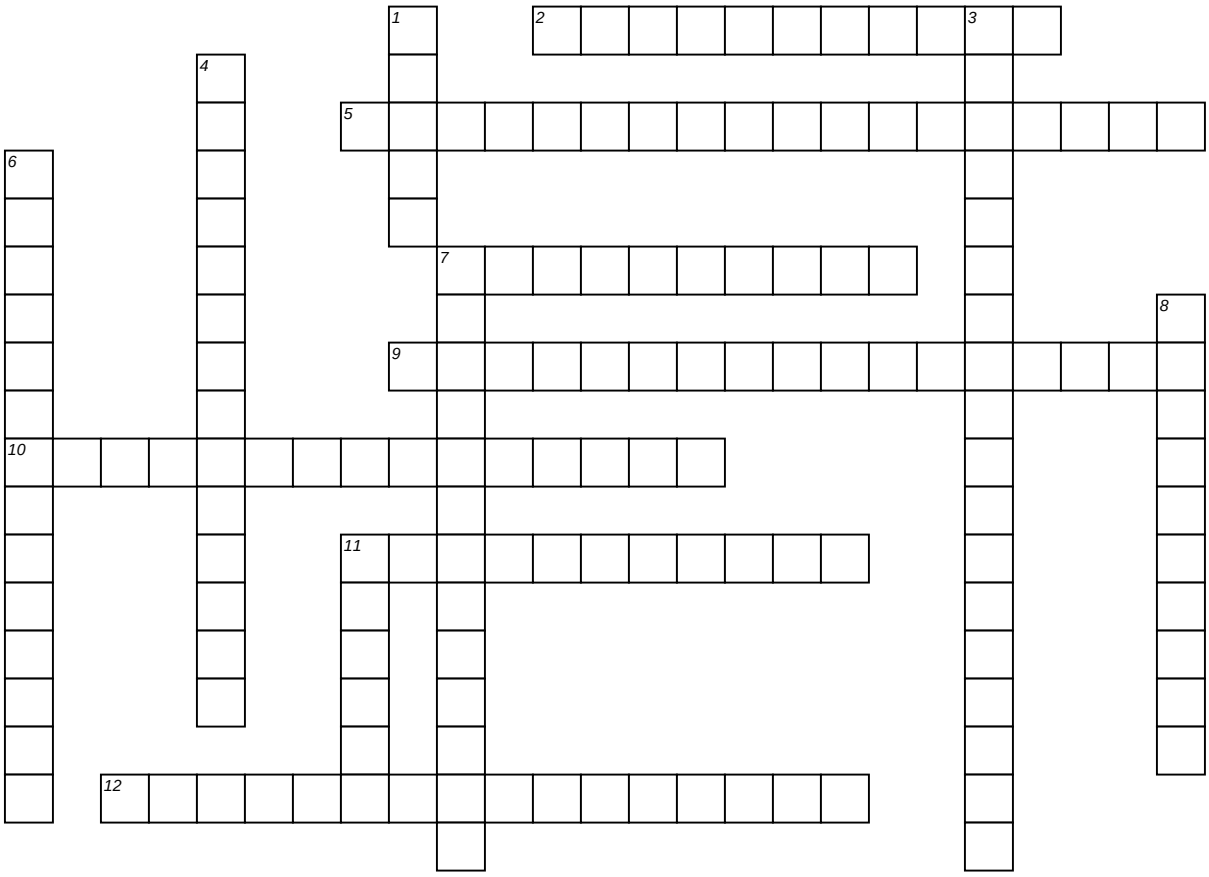
2 error of failing to reject a null hypothesis when in fact it is false (also called a "false positive")

3 claim that specifies a value for some population parameter that can form the basis for assuming a sampling distribution for a test statistic

5 decision made when the p-value is too small to believe that the statistic could have occurred due to chance variation

Inferences About Means

Advanced Placement Statistics



Stats: Modeling the World, Chapters 23-25

ACROSS

2 error of failing to reject a null hypothesis when in fact it is false (also called a "false positive")

5 type of interval used to show a plausible range of values for the true mean difference of two independent populations

7 error of rejecting a null hypothesis when in fact it is true (also called a "false positive")

9 alpha level

10 type of interval used to show a plausible range of values for the true mean difference of two populations that are not independent

11 type of test used to show the true mean difference is less than, greater than, or unequal to a null value, for two populations that are not independent

12 as this increases, t-distributions approach the Normal model

DOWN

1 probability that a hypothesis test will correctly reject a false null hypothesis

3 type of interval used to show a plausible range of values for the true mean of a population

4 type of test used to show that the true mean of one population is less than, greater than, or unequal to the mean of another population

6 type of test used to show that the true mean of a population is less than, greater than, or unequal to a null value

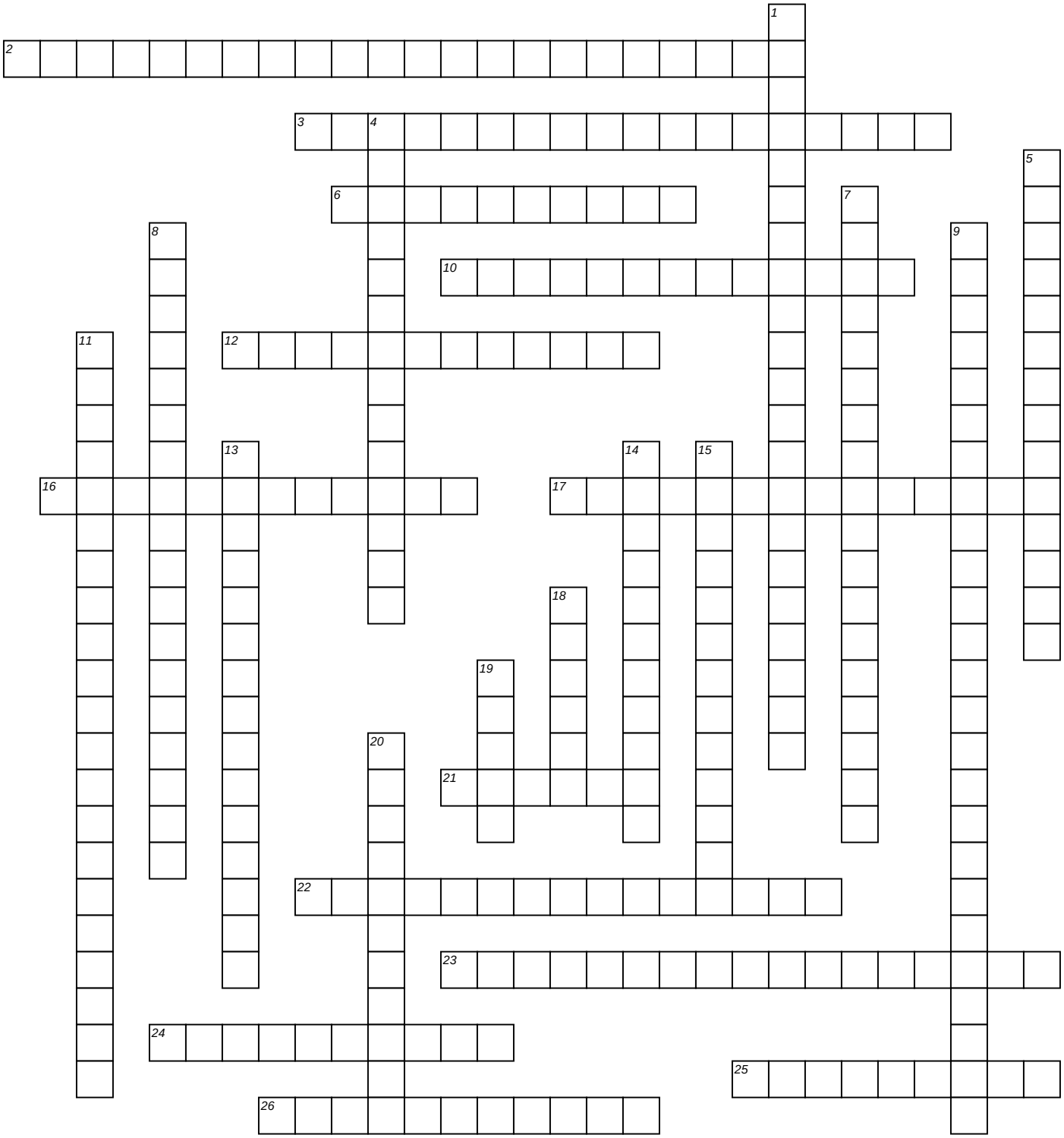
7 distribution that is unimodal, symmetric, and bell-shaped, but generally has fatter tails and a narrower center than the Normal model

8 threshold P-value that determines when we reject a null hypothesis

11 the probability of observing a value for a test statistic at least as far from the hypothesized value as the statistic actually observed if the null hypothesis is true

Inference

Advanced Placement Statistics



Stats: Modeling the World, Chapters 23-27

ACROSS

2 confidence interval for the true value of a proportion
3 an interval of values usually of the form estimate ± margin of error
6 threshold P-value that determines when we reject a null hypothesis
10 distribution that is unimodal, symmetric, and bell-shaped, but generally has fatter tails and a narrower center than the Normal model
12 type of chi-square test used for categorical data collected from a single sample
16 decision made the the p-value is large enough to believe that the statistic could have occurred due to chance variation
17 type of test used to show that the true mean of one population is less than, greater than, or unequal to the mean of another population
21 decision made when the p-value is too small to believe that the statistic could have occurred due to chance variation
22 type of interval used to show a plausible range of values for the true mean difference of two populations that are not independent
23 type of chi-square test used to show the actual distribution of a categorical variable is not the same as the proposed distribution
24 error of rejecting a null hypothesis when in fact it is true (also called a "false positive")
25 type of distribution that is skewed right, but appears more symmetric as the degrees of freedom increase
26 type of test used to show the true mean difference is less than, greater than, or unequal to a null value, for two populations that are not independent

DOWN

1 proposes what we should conclude if we find the null hypothesis to be unlikely
4 claim that specifies a value for some population parameter that can form the basis for assuming a sampling distribution for a test statistic
5 type of test used to show that the true mean of a population is less than, greater than, or unequal to a null value
7 type of interval used to show a plausible range of values for the true mean difference of two independent populations
8 type of interval used to show a plausible range of values for the true mean of a population
9 type of interval used to show a plausible range of values for the true slope for two quantitative variables
11 type of test used to show there is a relationship between two quantitative variables
13 proportion of all samples of size n that will produce a statistic within a corresponding margin of error of the true parameter
14 type of chi-square test used for categorical data collected from two or more groups
15 extent of the interval on either side of the observed statistic value
18 the probability of observing a value for a test statistic at least as far from the hypothesized value as the statistic actually observed if the null hypothesis is true
19 probability that a hypothesis test will correctly reject a false null hypothesis
20 error of failing to reject a null hypothesis when in fact it is false (also called a "false negative")