

Chanceforge — Activity Book

A Spark & Anvil activity book — puzzles, questions, and fun from Chanceforge.

Write in the book, circle your answers, and check the Answer Key at the back.

1. Data Collection & Organization

Circle the correct answer

1. What is a survey?

(A) A math formula for finding the average (B) A type of graph used to display data
(C) A way to organize numbers in order from least to greatest (D) A method of collecting data by asking people questions

2. What does a tally mark of five look like?

(A) Five vertical lines in a row (B) Four vertical lines with one diagonal line crossing through them (C) A single vertical line with the number 5 next to it (D) Five dots arranged in a circle

3. What is a frequency table?

(A) A table that compares two different data sets side by side (B) A table that lists data in order from smallest to largest (C) A table that shows how often each value or category appears in a data set (D) A table that shows the average of each category

4. What type of data describes qualities or characteristics that cannot be measured with numbers, such as favorite color or type of pet?

(A) Continuous data (B) Categorical data (C) Numerical data (D) Sequential data

5. A class surveys 20 students about their favorite fruit. The results are: Apple-8, Banana-5, Orange-4, Grape-3. Why is it important that the total of all responses equals 20?

(A) Because every student surveyed must be accounted for in exactly one category
(B) Because 20 is an even number and makes the math easier (C) Because the most popular fruit must have exactly half the votes (D) Because each fruit must receive the same number of votes

6. A tally chart shows: Red-IIII II, Blue-IIII IIII, Green-IIII. What does this tally chart tell you about which color is most popular?

(A) Green is most popular because it has exactly one group of five (B) All three colors are equally popular because they all use tally marks (C) Blue is most popular with 10 tallies, followed by Red with 7, then Green with 5 (D) Red is most popular

because it is listed first in the chart

Write the answer

7. Why would a frequency table be more useful than a list of raw data when you have 100 survey responses?

Answer: _____

8. Explain the difference between categorical data and numerical data using an example.

Answer: _____

9. Why is it important to organize data before analyzing it?

Answer: _____

10. How does using categories help you understand survey results better?

Answer: _____

Match each question to its answer

1. A school cafeteria surveys students about lunch preferences. The results are: Pizza-45, Tacos-30, Salad-15, Pasta-10. Create a frequency table and determine what fraction of students prefer Pizza. →

2. You surveyed 30 classmates about how many siblings they have. The responses are: 0, 1, 1, 2, 0, 3, 1, 2, 1, 1, 0, 2, 1, 1, 3, 2, 0, 1, 2, 1, 0, 1, 2, 3, 1, 0, 2, 1, 3, 2. Organize this data into a frequency table. →

3. Design a survey question to collect categorical data about your classmates' after-school activities. Include at least 4 response choices. →

4. Convert the following tally chart to a frequency table: Dogs- IIII IIII III , Cats- IIII IIII , Fish- IIII I , Birds- III . How many total pets were counted? →

5. A teacher wants to know which day of the week students prefer for a field trip. Write out the steps to collect, organize, and present this categorical data. →

(A) What is your favorite after-school activity? A) Sports B) Reading C) Video games D) Art E) Music

(B) 1) Create a survey asking students to choose one day, 2) Collect all responses, 3) Tally or count responses for each day, 4) Make a frequency table, 5) Display results in a bar graph

(C) 45 out of 100, or $\frac{9}{20}$ of students prefer Pizza

(D) 0 siblings: 6, 1 sibling: 12, 2 siblings: 8, 3 siblings: 4

(E) Dogs: 13, Cats: 10, Fish: 6, Birds: 3. Total: 32 pets

2. Graphs & Data Displays

Circle the correct answer

11. What type of graph uses rectangular bars to compare amounts in different categories?

- (A) Circle graph (B) Bar graph (C) Line plot (D) Pictograph

12. What is a line plot?

(A) A graph that uses pictures to represent data values (B) A graph that shows data using marks (usually X's or dots) above a number line (C) A graph that connects data points with straight lines to show change over time (D) A graph that divides a circle into sections

13. In a pictograph, what is a key (or legend)?

(A) The title of the pictograph (B) A symbol explanation that tells how many items each picture represents (C) The total number of pictures in the graph (D) The largest value in the data set

14. What must every graph have to help the reader understand the data?

(A) A title, labeled axes (or categories), and a scale or key (B) Only a key and grid lines (C) Only numbers and pictures (D) Only a title and colorful bars

15. A bar graph shows the number of books read by four students: Mia-12, Leo-8, Ava-15, Sam-6. Explain what the heights of the bars tell you about the students' reading habits.

(A) Taller bars mean the student read harder books, and this has been consistently demonstrated in practice (B) Taller bars mean more books read. Ava read the most (15), Sam read the fewest (6), and the difference between the tallest and shortest bar shows the range of reading amounts (C) The bars show the order in which students signed up for the reading program, and this principle applies across similar situations (D) The bars show which students are taller in real life, and this has been consistently demonstrated in practice

16. A line plot shows test scores: one X above 70, three X's above 80, five X's above 85, four X's above 90, two X's above 95. What does the cluster of X's around 85-90 tell you?

(A) Every student should have scored exactly 85, and this has been consistently demonstrated in practice (B) The scores at 70 and 95 are wrong and should be removed, and this has been consistently demonstrated in practice (C) The test was too easy because most students scored above 80 (D) Most students scored between 85 and 90, which shows the data clusters in that range and represents the most common performance level

Write the answer

17. A pictograph uses one star symbol to represent 4 students. If the 'Soccer' row has 3 full stars and one half star, how many students chose Soccer? Explain your reasoning.

Answer: _____

18. Why can a bar graph be misleading if the scale on the y-axis does not start at zero?

Answer: _____

19. Why would a line plot be a better choice than a bar graph for showing the number of pets each student in your class has?

Answer: _____

20. What is the advantage of using a pictograph instead of a bar graph when presenting data to younger students?

Answer: _____

Match each question to its answer

1. A bar graph shows rainfall in inches for each month: Jan-3, Feb-2, Mar-4, Apr-6, May-5, Jun-3. What was the total rainfall for the first six months? Which month had double the rainfall of February? →

2. Create a line plot for the following data showing the number of hours students spent reading last week: 2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 8. What value appears most often? →

3. A pictograph shows ice cream sales using cone symbols. The key says each cone = 10 sales. Monday has 4 cones, Tuesday has 2.5 cones, Wednesday has 3 cones, Thursday has 1.5 cones, Friday has 5 cones. Calculate the total sales and the difference between the busiest and slowest days. →

4. Use the following bar graph data about favorite school subjects to answer: Math-9, Science-7, Reading-6, Art-5, Music-3. What fraction of students chose either Math or Science? →

5. A student has data on the number of sunny days each month for a year: Jan-8, Feb-10, Mar-14, Apr-18, May-22, Jun-26, Jul-28, Aug-27, Sep-22, Oct-16, Nov-10, Dec-7. Which type of graph would best show how sunny days change across the months? →

(A) 16 out of 30, or $\frac{8}{15}$ of students chose Math or Science

(B) Mon: 40, Tue: 25, Wed: 30, Thu: 15, Fri: 50. Total: 160 sales. Difference between busiest (Fri: 50) and slowest (Thu: 15) is 35 sales

(C) Total rainfall: 23 inches. March had 4 inches, which is double February's 2 inches

(D) A bar graph with months on the x-axis and number of sunny days on the y-axis, because it shows the pattern of increasing then decreasing sunny days across the year

(E) The value 5 appears most often (4 times). The line plot has: one X above 2, two X's

above 3, three X's above 4, four X's above 5, two X's above 6, two X's above 7, one X above 8

3. Mean, Median, Mode & Range

Circle the correct answer

21. What is the mean of a data set?

- (A) The middle value when the data is arranged in order (B) The difference between the largest and smallest values (C) The value that appears most often in the data set (D) The sum of all values divided by the number of values

22. What is the median of a data set?

- (A) The value exactly halfway between the minimum and maximum (B) The most frequently occurring value (C) The average of all the values in the data set (D) The middle value when all values are arranged in order from least to greatest

23. What is the mode of a data set?

- (A) The sum of all values divided by the count (B) The middle value in the sorted data set (C) The average of the highest and lowest values (D) The value that appears most frequently in the data set

24. How do you calculate the range of a data set?

- (A) Add the smallest and largest values together (B) Count how many different values appear in the data set (C) Subtract the smallest value from the largest value (D) Divide the largest value by the smallest value

25. Explain why the mean can be misleading when a data set contains an outlier. Use the data set {10, 12, 11, 13, 50} as an example.

- (A) The mean is never misleading because it uses every value in the calculation (B) The outlier 50 pulls the mean up to 19.2, which is much higher than the typical values (10-13). The mean does not represent what is typical because one extreme value distorts it (C) The mean is 12 because you should ignore the outlier before calculating (D) The outlier makes the mean too low because 50 is so different from the other numbers

26. Why is the median sometimes a better measure of center than the mean?

- (A) The median is easier to calculate because you only need to sort the data (B) The median is always more accurate than the mean, and this principle applies across similar situations (C) The median is resistant to extreme values (outliers) because it only depends on the middle position, not the actual values of all data points (D) The median uses more data points than the mean, as this is what research and standard practice suggests

Write the answer

27. A data set has two modes: 15 and 22. What does this tell you about the data?

Answer: _____

28. Explain the difference between what the mean tells you and what the range tells you about a data set.

Answer: _____

29. In what situation would the mode be the most useful measure of central tendency?

Answer: _____

30. Can the mean of a data set be a value that does not appear in the data set? Explain with an example.

Answer: _____

Match each question to its answer

1. Find the mean, median, mode, and range of the following test scores: 85, 90, 78, 92, 85, 88, 76, 85, 94, 90. →

2. A student's quiz scores are 72, 85, 90, 68, and one missing score. If the student's mean score is 80, what is the missing score? →

3. Find the median of this data set with an even number of values: 14, 18, 22, 25, 30, 36. →

4. A basketball player scored these points in 8 games: 12, 15, 8, 22, 15, 18, 15, 11. Find the mode and explain what it tells the coach about the player's performance. →

5. Two data sets each have 5 values. Set A: {10, 10, 10, 10, 10}. Set B: {2, 6, 10, 14, 18}. Both have a mean of 10. Calculate the range of each set and explain what this reveals about the data. →

(A) Mean: 86.3, Median: 86.5, Mode: 85, Range: 18

(B) The mode is 15 (appears 3 times). It tells the coach that scoring around 15 points is this player's most common performance level

(C) Set A range: 0 (no spread at all). Set B range: 16 (values are very spread out). Even though both have the same mean, Set B is much more variable

(D) 23.5, which is the average of the two middle values: $(22 + 25) / 2 = 23.5$

(E) 85. The five scores must sum to 400 (80×5). The known scores sum to 315, so the missing score is $400 - 315 = 85$

4. Probability Basics

Circle the correct answer

31. What is the probability of an event that is certain to happen?
(A) 0.5 (B) 0 (C) 1 (D) 2
32. What is the probability of an event that is impossible?
(A) 0 (B) -1 (C) 1 (D) 0.5
33. A bag contains 3 red marbles and 7 blue marbles. What is the probability of randomly drawing a red marble?
(A) $\frac{1}{3}$ (B) $\frac{3}{7}$ (C) $\frac{7}{10}$ (D) $\frac{3}{10}$
34. Which of the following is NOT a possible probability value?
(A) 0 (B) 1 (C) 0.75 (D) 1.5
35. A spinner has 4 equal sections colored red, blue, green, and yellow. How many equally likely outcomes are there for one spin?
(A) 1 (B) 4 (C) 8 (D) 2
36. You flip a fair coin. What is the probability of getting heads?
(A) $\frac{1}{4}$ (B) $\frac{1}{3}$ (C) 1 (D) $\frac{1}{2}$

Write the answer

37. An event has a probability of 0.25. Which word best describes how likely it is to happen?
Answer: _____
38. A game is fair if all players have an equal chance of winning. Which game below is fair?
Answer: _____
39. If the probability of rain tomorrow is 0.7, what is the probability that it will NOT rain?
Answer: _____
40. A standard number cube (die) is rolled once. Which event is most likely?
Answer: _____

Match each question to its answer

1. In a probability experiment, what do we call each possible result? →
2. A bag contains 5 green tokens and 5 yellow tokens. Maria says the probability of drawing green is 'about half.' Is she correct, and why? →
3. Which probability best describes the likelihood of rolling a 7 on a standard six-sided

die? →

4. A jar holds 2 red, 3 blue, and 5 white beads. What is the set of all possible outcomes when drawing one bead? →

5. A spinner is divided into 8 equal sections. Three sections are red, two are blue, and three are green. What is the probability of landing on blue? →

(A) $\frac{2}{8}$

(B) An outcome

(C) Yes, because 5 out of 10 tokens are green, which equals $\frac{1}{2}$

(D) 0

(E) {red, blue, white}

5. Experimental vs. Theoretical Probability

Circle the correct answer

41. What is theoretical probability based on?

(A) Mathematical reasoning about equally likely outcomes (B) Asking people what they think will happen (C) Guessing what will happen next (D) The results of performing an experiment many times

42. What is experimental probability based on?

(A) Reading the probability from a textbook (B) Using a formula to calculate the chance (C) The results of actually performing trials of an experiment (D) Counting the total number of possible outcomes

43. A coin is flipped 50 times and lands on heads 28 times. What is the experimental probability of heads?

(A) $\frac{22}{50}$ (B) $\frac{1}{2}$ (C) $\frac{28}{100}$ (D) $\frac{28}{50}$

44. The theoretical probability of rolling a 3 on a standard die is $\frac{1}{6}$. In 60 rolls, how many times would you theoretically expect to roll a 3?

(A) 6 (B) 3 (C) 10 (D) 12

45. Lily flipped a coin 10 times and got 7 heads. She expected 5 heads based on theory. Why might the results differ?

(A) Theoretical probability is always wrong, which is the most common explanation for this phenomenon (B) She made a mistake counting, and this principle applies across similar situations (C) The coin must be unfair, which is the most common explanation for this phenomenon (D) Small numbers of trials often produce results that differ from theoretical probability due to natural variation

46. What does the Law of Large Numbers tell us?

(A) Large numbers are always more probable than small numbers (B) Theoretical probability is only accurate for large numbers (C) You need at least 1000 trials for probability to work (D) As the number of trials increases, experimental probability gets closer to theoretical probability

Write the answer

47. A spinner has 4 equal sections: red, blue, green, yellow. After 200 spins, the results are: red 55, blue 48, green 52, yellow 45. What is the experimental probability of green?

Answer: _____

48. Marcus rolled a die 30 times and got a 6 exactly 5 times. The theoretical probability of rolling a 6 is $\frac{1}{6}$. Compare these probabilities.

Answer: _____

49. Why might a teacher use a computer simulation to flip a coin 10,000 times instead of actually flipping a real coin?

Answer: _____

50. A class used a random number generator (1-6) to simulate rolling a die 1,000 times. Which result would be most surprising?

Answer: _____

Match each question to its answer

1. Which formula correctly calculates experimental probability? →

2. A bag contains an unknown number of colored marbles. After 40 draws (replacing each time), red appears 12 times. What is a reasonable estimate for the probability of drawing red? →

3. A student ran a simulation of spinning a spinner 50 times, then 500 times, then 5,000 times. How would the experimental probabilities likely change across these three runs? →

4. Which scenario would best be studied using a simulation rather than a physical experiment? →

5. A spinner has 3 equal sections: A, B, C. The theoretical probability of landing on each is $\frac{1}{3}$. After 90 spins: A appeared 25 times, B appeared 35 times, C appeared 30 times. Which section's experimental probability is closest to its theoretical probability? →

(A) They would get progressively closer to the theoretical probability

(B) C

(C) Finding the probability that it rains on 5 consecutive school days

(D) Number of times event occurred / Total number of trials

6. Compound Events & Sample Spaces

Circle the correct answer

51. What is a sample space?

(A) The area where an experiment takes place (B) The set of all possible outcomes of an experiment (C) The number of times an experiment is repeated (D) The most likely outcome of an experiment

52. What is a compound event?

(A) An event that can never happen (B) An event that consists of two or more simple events (C) An event that requires multiplication to solve (D) An event with only one possible outcome

53. You flip a coin and roll a standard die. How many outcomes are in the sample space?

(A) 6 (B) 2 (C) 8 (D) 12

54. What organizational tool shows all possible outcomes by branching from each choice to the next?

(A) A bar graph (B) A number line (C) A Venn diagram (D) A tree diagram

55. Two events are independent if which of the following is true?

(A) Both events have the same probability (B) The events happen at the same time (C) The outcome of one event does not affect the outcome of the other (D) One event causes the other to happen

56. A bag has 4 red and 6 blue marbles. You draw one marble, do NOT replace it, then draw another. Are these events independent or dependent?

(A) Dependent, because removing the first marble changes the probabilities for the second draw (B) Independent, because both draws come from the same bag (C) Dependent, because both marbles have the same shape (D) Independent, because each draw is random

Write the answer

57. You spin a spinner with sections A, B, C, and then flip a coin. Using a tree diagram, how many branches would appear at the final level?

Answer: _____

58. For two independent events, $P(A) = 1/3$ and $P(B) = 1/4$. What is $P(A \text{ and } B)$?

Answer: _____

59. What is the probability of flipping heads on a coin AND rolling a 6 on a die?

Answer: _____

60. A bag contains 5 red and 3 blue marbles. You draw one, replace it, then draw again. What is the probability of drawing red both times?

Answer: _____

Match each question to its answer

1. A bag contains 5 red and 3 blue marbles. You draw one marble, do NOT replace it, then draw again. What is the probability of drawing red on the first draw AND blue on the second? →

2. A restaurant offers 3 entrees and 4 desserts. Using the counting principle, how many different meal combinations (one entree, one dessert) are possible? →

3. You draw a tree diagram for rolling two dice. How many total outcomes appear at the tips of the branches? →

4. A coin is flipped and a die is rolled. Sarah says these events are dependent because they happen at the same time. Analyze Sarah's claim. →

5. A card is drawn from a standard 52-card deck and NOT replaced. Then a second card is drawn. Are these events independent? →

(A) $15/56$

(B) 12

(C) Sarah is incorrect; the events are independent because the outcome of the coin does not affect the outcome of the die, regardless of timing

(D) No, because removing the first card changes the number of remaining cards and possibly the composition

(E) 36

7. Statistical Variability & Distributions

Circle the correct answer

61. What does 'variability' mean in statistics?

(A) The most common value in a data set (B) How many data values are in a set (C) How spread out or different the data values are from each other (D) The average of all the data values

62. What is the range of the data set: 12, 15, 18, 22, 30?

- (A) 12 (B) 18 (C) 15 (D) 30

63. In a box plot, what does the box represent?

- (A) The top 25% of the data, which is why this approach is recommended by experts
(B) The entire data set from minimum to maximum, as this is what research and standard practice suggests
(C) The middle 50% of the data, from the first quartile (Q1) to the third quartile (Q3)
(D) The mean and median of the data, and this has been consistently demonstrated in practice

64. A histogram shows the following bar heights for score ranges: 60-69 (2 students), 70-79 (8 students), 80-89 (12 students), 90-100 (3 students). Which score range has the most students?

- (A) 70-79 (B) 60-69 (C) 90-100 (D) 80-89

65. What is the interquartile range (IQR)?

- (A) The median of the data set (B) The difference between the maximum and minimum values
(C) The difference between the third quartile (Q3) and the first quartile (Q1)
(D) The average of Q1 and Q3

66. For the data set: 3, 5, 7, 8, 10, 12, 14, 16, 20, find Q1 and Q3, then calculate the IQR.

- (A) $Q1 = 5$, $Q3 = 16$, $IQR = 11$ (B) $Q1 = 6$, $Q3 = 15$, $IQR = 9$ (C) $Q1 = 7$, $Q3 = 14$, $IQR = 7$
(D) $Q1 = 8$, $Q3 = 12$, $IQR = 4$

Write the answer

67. Which measure of variability is most resistant to outliers?

Answer: _____

68. A data set has values: 10, 12, 13, 14, 15, 15, 16, 50. Which value is most likely an outlier?

Answer: _____

69. What does the line inside the box of a box plot represent?

Answer: _____

70. A histogram has most of its bars clustered on the left with a long tail extending to the right. What shape is this distribution?

Answer: _____

Match each question to its answer

1. The five-number summary for a data set is: Min = 20, Q1 = 35, Median = 50, Q3 = 65, Max = 80. What is the IQR? →
 2. Two classes took the same test. Class A has an IQR of 8 and Class B has an IQR of 22. What can you conclude about the variability? →
 3. A common rule for identifying outliers uses the IQR. A data point is an outlier if it is more than $1.5 \times \text{IQR}$ above Q3 or below Q1. If Q1 = 20 and Q3 = 40, what is the upper boundary for outliers? →
 4. Which type of graph is best for showing the distribution shape of a large data set with numerical values grouped into intervals? →
 5. Data set X has values: 48, 50, 50, 51, 52. Data set Y has values: 20, 35, 50, 65, 80. Both have a mean of 50.2 and 50. Which set has greater variability? →
 - (A) Class B has more variability in the middle 50% of scores than Class A
 - (B) Histogram
 - (C) 30
 - (D) 70
 - (E) Data set Y, because its values are much more spread out from the mean
-

8. Statistical Inference & Prediction

Circle the correct answer

71. In statistics, what is a population?

- (A) The entire group of individuals or items that you want to study (B) The average of all the data (C) A small group chosen for a survey (D) The number of people in a city

72. What is a sample in statistics?

- (A) A guess about what the data will show (B) The largest value in a data set (C) The same thing as the population (D) A subset of the population selected to represent the whole group

73. A researcher wants to know the favorite lunch of all 800 students in a school. She surveys 80 students chosen randomly. What is the sample?

- (A) All 800 students in the school (B) The 80 randomly chosen students (C) The school building (D) The lunch menu

74. What is sampling bias?

- (A) When all members of the population are surveyed, and this has been consistently demonstrated in practice (B) When the sample is too large, which is why this approach is recommended by experts (C) When the researcher makes a math error, and this has been consistently demonstrated in practice (D) When certain

members of the population are more likely to be included in the sample than others

75. Which relationship between two variables describes a pattern where one variable increases as the other increases?

(A) Negative correlation (B) Positive correlation (C) Causation (D) No correlation

76. A school wants to determine the average height of all 500 students. Which sampling method would likely produce the most representative sample?

(A) Randomly selecting 50 students from a list of all 500 students (B) Measuring only 6th graders (C) Measuring only students who volunteer (D) Measuring only the basketball team

Write the answer

77. In a survey, 35 out of 100 randomly selected students prefer pizza for lunch. There are 600 students in the school. Predict how many students in the whole school prefer pizza.

Answer: _____

78. A company surveys customers at its store on a Saturday morning. They conclude that most of their customers are over 50 years old. What type of bias might affect this conclusion?

Answer: _____

79. A scatterplot shows hours of screen time per day (x-axis) and math test scores (y-axis). As screen time increases, scores tend to decrease. What type of correlation does this describe?

Answer: _____

80. A random sample of 200 voters shows 120 support a new park. The city has 5,000 voters. Estimate the total number of voters who support the park.

Answer: _____

Match each question to its answer

1. Why is a random sample important when making predictions about a population?

→

2. A study finds that students who eat breakfast score higher on tests than those who skip breakfast. A news article claims: 'Eating breakfast causes higher test scores.' What is wrong with this claim? →

3. A website poll asks visitors: 'Do you think social media is harmful?' and 88% say yes. Can this result be used to conclude that 88% of all Americans think social media is harmful? →

4. Two different random samples of 50 students each are taken from the same school.

Sample A finds that 60% prefer math. Sample B finds that 54% prefer math. Why do the results differ? →

5. A scientist measures the temperature (°F) and the number of ice cream cones sold daily for 30 days. She finds a strong positive correlation. Does this mean temperature causes ice cream sales? →

(A) No, because website visitors who choose to respond are not a random sample of all Americans

(B) It gives every member of the population an equal chance of being selected, reducing bias

(C) Sampling variability — different random samples naturally produce slightly different results

(D) Correlation does not prove causation; other factors (like sleep, study habits, or family support) might explain both variables

(E) No, correlation shows an association but does not prove that one variable causes the other

9. Experimental Probability

Circle the correct answer

81. Explain the difference between categorical data and numerical data using an example.

(A) Categorical data uses bigger numbers than numerical data (B) Categorical data sorts into groups (like favorite sport), while numerical data is measured with numbers (like height in inches) (C) Categorical data is always collected from surveys, while numerical data is always measured (D) Categorical data can only have two categories, while numerical data has many values

82. A teacher shows two graphs of the same data about student heights. Graph 1 uses increments of 1 inch on the y-axis. Graph 2 uses increments of 6 inches. Judge which graph better serves a discussion about the range of heights in the class and explain your reasoning.

(A) Graph 2 is better because it is simpler and easier to read, since this is how the operation works with these particular numbers (B) Graph 2 is better because fewer bars means fewer mistakes, which is what you get when you apply the distributive property here (C) Both graphs are equally useful because they show the same data, according to the standard procedure for solving this kind of equation (D) Graph 1 (1-inch increments) is better because it shows the detailed variation in heights, while Graph 2 (6-inch increments) would clump most students into the same bar, hiding the range

83. Two classes took the same test. Class A scores: 70, 75, 80, 85, 90 (Mean: 80). Class B scores: 60, 70, 80, 90, 100 (Mean: 80). Both classes have the same mean. Analyze which class performed more consistently and explain your reasoning.

(A) Neither class is more consistent because they have the same number of students
(B) Class B performed more consistently because their scores include the highest value (100)
(C) Class A performed more consistently. Class A's range is 20 (90-70) while Class B's range is 40 (100-60). Class A's scores are closer together, showing more uniform performance
(D) Both classes performed equally because they have the same mean

84. A game is fair if all players have an equal chance of winning. Which game below is fair?

(A) Player A wins if the coin shows heads, Player B wins if it shows heads or tails
(B) Player A wins if a die shows 1-4, Player B wins if it shows 5-6
(C) Player A picks first from a bag, Player B picks from what is left
(D) Each player rolls a die and the highest number wins

85. Lily flipped a coin 10 times and got 7 heads. She expected 5 heads based on theory. Why might the results differ?

(A) Theoretical probability is always wrong, which is the most common explanation for this phenomenon
(B) Small numbers of trials often produce results that differ from theoretical probability due to natural variation
(C) The coin must be unfair, which is the most common explanation for this phenomenon
(D) She made a mistake counting, and this principle applies across similar situations

86. You flip a coin 3 times. What is the probability of getting heads all 3 times?

(A) $\frac{1}{8}$ (B) $\frac{3}{8}$ (C) $\frac{1}{2}$ (D) $\frac{1}{6}$

Write the answer

87. Two classes took the same test. Class A has an IQR of 8 and Class B has an IQR of 22. What can you conclude about the variability?

Answer: _____

88. A town has 4,000 households. A random sample of 100 households shows that 25 have solar panels. What is the best estimate for the number of households with solar panels in the town?

Answer: _____

89. A school collected data on transportation to school. The frequency table shows: Bus-120, Car-85, Walk-45, Bike-30, Other-20. A student says the tally chart and the frequency table always give different information. Analyze whether this is true.

Answer: _____

90. A bar graph shows the number of books read by four students: Mia-12, Leo-8, Ava-15, Sam-6. Explain what the heights of the bars tell you about the students' reading habits.

Answer: _____

Match each question to its answer

1. A restaurant owner collects tips for a week: \$5, \$8, \$6, \$45, \$7, \$6, \$9. She wants to advertise the 'average tip.' Should she use the mean or median? Justify your choice with calculations. →

2. A spinner is divided into 8 equal sections. Three sections are red, two are blue, and three are green. What is the probability of landing on blue? →

3. Which formula correctly calculates experimental probability? →

4. Carlos claims that when drawing from a deck without replacement, you can still use the simple multiplication rule $P(A) \times P(B)$ to find compound probabilities. Is Carlos correct? Evaluate his claim. →

5. Two histograms show test scores for two classes. Class A's histogram is tall and narrow. Class B's histogram is short and wide. Which class has more consistent scores? →

(A) Class A, because a tall narrow histogram means scores are clustered together

(B) No, without replacement the events are dependent, so you must use $P(A) \times P(B \text{ given } A)$ to account for the changed conditions

(C) $2/8$

(D) Number of times event occurred / Total number of trials

(E) She should use the median (\$7). The mean (\$12.29) is inflated by the \$45 outlier and misrepresents what a typical tip looks like. The median better shows the center of the typical tips

10. Compound Events & Trees

Circle the correct answer

91. Two students organize the same data about favorite books. Student A uses 5 specific categories (Fantasy, Mystery, Comics, Science, History). Student B uses 3 broad categories (Fiction, Non-fiction, Other). Which approach is better for understanding reading preferences, and why?

(A) Both approaches are equally good because they use the same raw data (B) Student B's approach is better because 3 categories are easier to put in a graph (C) Student B's approach is better because fewer categories means less chance of making mistakes (D) Student A's approach is better for understanding specific preferences

because the 5 categories reveal more detail about what students actually enjoy reading

92. Why can a bar graph be misleading if the scale on the y-axis does not start at zero?

- (A) It only matters if the bars are horizontal instead of vertical (B) Starting above zero exaggerates small differences between bars, making them look much larger than they actually are (C) It cannot be misleading because the numbers are still accurate (D) Starting above zero makes the graph harder to color in

93. A basketball player scored these points in 8 games: 12, 15, 8, 22, 15, 18, 15, 11. Find the mode and explain what it tells the coach about the player's performance.

- (A) The mode is 15 (appears 3 times). It tells the coach that scoring around 15 points is this player's most common performance level (B) The mode is 22 because it is the highest score (C) There is no mode because the scores are all different game results (D) The mode is 14.5 because that is the average of all scores

94. A jar holds 2 red, 3 blue, and 5 white beads. What is the set of all possible outcomes when drawing one bead?

- (A) {red, blue} (B) {red, red, blue, blue, blue, white, white, white, white, white} (C) {2, 3, 5} (D) {red, blue, white}

95. What is experimental probability based on?

- (A) Counting the total number of possible outcomes (B) The results of actually performing trials of an experiment (C) Using a formula to calculate the chance (D) Reading the probability from a textbook

96. You randomly select a letter from the word MATH and a digit from 1-3. List the complete sample space.

- (A) {M, A, T, H, 1, 2, 3} (B) {M1, A2, T3, H1} (C) {MA, MT, MH, AT, AH, TH, 12, 13, 23} (D) {M1, M2, M3, A1, A2, A3, T1, T2, T3, H1, H2, H3}

Write the answer

97. Which type of graph is best for showing the distribution shape of a large data set with numerical values grouped into intervals?

Answer: _____

98. A company surveys customers at its store on a Saturday morning. They conclude that most of their customers are over 50 years old. What type of bias might affect this conclusion?

Answer: _____

99. Convert the following tally chart to a frequency table: Dogs-III III III, Cats-III III, Fish-III I, Birds-III. How many total pets were counted?

Answer: _____

100. What is the advantage of using a pictograph instead of a bar graph when presenting data to younger students?

Answer: _____

Match each question to its answer

1. What is the median of a data set? →

2. Which probability best describes the likelihood of rolling a 7 on a standard six-sided die? →

3. A coin is flipped 50 times and lands on heads 28 times. What is the experimental probability of heads? →

4. A combination lock has 3 dials, each with digits 0-9. How many possible combinations exist? →

5. Use the 1.5 x IQR rule. Given $Q1 = 30$, $Q3 = 50$, which of these values is an outlier? →

(A) The middle value when all values are arranged in order from least to greatest

(B) 1,000

(C) 95

(D) 0

(E) 28/50

11. Statistical Distributions

Circle the correct answer

101. Design a survey question to collect categorical data about your classmates' after-school activities. Include at least 4 response choices.

(A) Rate your after-school activities from 1 to 10 (B) Do you like after-school activities? A) Yes B) No (C) What is your favorite after-school activity? A) Sports B) Reading C) Video games D) Art E) Music (D) How many hours do you spend on after-school activities? A) 1 B) 2 C) 3 D) 4

102. A student has data on the number of sunny days each month for a year: Jan-8, Feb-10, Mar-14, Apr-18, May-22, Jun-26, Jul-28, Aug-27, Sep-22, Oct-16, Nov-10, Dec-7. Which type of graph would best show how sunny days change across the months?

(A) A line plot because the data involves numbers, which is what you get when you apply the distributive property here (B) A circle graph because there are 12 months in a year, according to the standard procedure for solving this kind of equation (C) A pictograph because you could use sun symbols, based on the standard mathematical approach for this type of problem (D) A bar graph with months on the x-axis and number of sunny days on the y-axis, because it shows the pattern of increasing then decreasing sunny days across the year

103. How do you calculate the range of a data set?

(A) Add the smallest and largest values together (B) Divide the largest value by the smallest value (C) Subtract the smallest value from the largest value (D) Count how many different values appear in the data set

104. A box contains 10 tiles numbered 1 through 10. You draw one tile. How many outcomes are in the event 'drawing an even number'?

(A) 5 (B) 10 (C) 6 (D) 4

105. A student ran a simulation of spinning a spinner 50 times, then 500 times, then 5,000 times. How would the experimental probabilities likely change across these three runs?

(A) They would get progressively closer to the theoretical probability (B) They would stay exactly the same each time (C) They would alternate between too high and too low (D) They would get further from the theoretical probability

106. For two independent events, $P(A) = 1/3$ and $P(B) = 1/4$. What is $P(A \text{ and } B)$?

(A) $1/7$ (B) $2/7$ (C) $1/12$ (D) $7/12$

Write the answer

107. In a box plot, what does the box represent?

Answer: _____

108. A scatterplot shows hours of screen time per day (x-axis) and math test scores (y-axis). As screen time increases, scores tend to decrease. What type of correlation does this describe?

Answer: _____

109. A teacher wants to know which day of the week students prefer for a field trip. Write out the steps to collect, organize, and present this categorical data.

Answer: _____

110. A pictograph uses one star symbol to represent 4 students. If the 'Soccer' row has 3 full stars and one half star, how many students chose Soccer? Explain your reasoning.

Answer: _____

Match each question to its answer

1. A student's quiz scores are 72, 85, 90, 68, and one missing score. If the student's mean score is 80, what is the missing score? →

2. A standard number cube (die) is rolled once. Which event is most likely? →

3. What does the Law of Large Numbers tell us? →

4. A card is drawn from a standard 52-card deck and NOT replaced. Then a second card is drawn. Are these events independent? →

5. The five-number summary for a data set is: Min = 20, Q1 = 35, Median = 50, Q3 = 65, Max = 80. What is the IQR? →

(A) 85. The five scores must sum to 400 (80×5). The known scores sum to 315, so the missing score is $400 - 315 = 85$

(B) As the number of trials increases, experimental probability gets closer to theoretical probability

(C) No, because removing the first card changes the number of remaining cards and possibly the composition

(D) Rolling a number less than 5

(E) 30

12. Data-Driven Decision Making

Circle the correct answer

111. Why is it important to organize data before analyzing it?

(A) Organized data always gives different answers than unorganized data (B) Organized data takes up less space on paper (C) Organized data makes it easier to find patterns, compare values, and draw conclusions (D) You cannot do any math with unorganized data

112. A line plot shows test scores: one X above 70, three X's above 80, five X's above 85, four X's above 90, two X's above 95. What does the cluster of X's around 85-90 tell you?

(A) Most students scored between 85 and 90, which shows the data clusters in that range and represents the most common performance level (B) The scores at 70 and 95 are wrong and should be removed, and this has been consistently demonstrated in practice (C) The test was too easy because most students scored above 80 (D) Every student should have scored exactly 85, and this has been consistently demonstrated in practice

113. A student says 'The mean is always the best way to describe the center of data.' Analyze this claim using the data set: {\$30,000, \$32,000, \$35,000, \$33,000, \$250,000} representing salaries at a small company.

(A) The claim is correct because \$76,000 is the fair average salary (B) The claim is wrong. The mean salary is \$76,000, but 4 of 5 employees earn between \$30K-\$35K. The median (\$33,000) better represents the typical salary because the \$250K outlier distorts the mean (C) The claim is wrong because you should always use the mode for salary data (D) The claim is correct because the mean uses all the data and is therefore the most complete

114. What is the probability of an event that is certain to happen?

(A) 0.5 (B) 1 (C) 2 (D) 0

115. A class used a random number generator (1-6) to simulate rolling a die 1,000 times. Which result would be most surprising?

(A) The number 2 appeared 170 times (B) The number 5 appeared 155 times (C) The number 4 appeared 500 times (D) The number 1 appeared 180 times

116. A coin is flipped and a die is rolled. Sarah says these events are dependent because they happen at the same time. Analyze Sarah's claim.

(A) Sarah is correct because both involve random chance (B) Sarah is correct because simultaneous events are always dependent (C) Sarah is incorrect; the events are independent because the outcome of the coin does not affect the outcome of the die, regardless of timing (D) Sarah is incorrect because they are the same event, not two events

Write the answer

117. For the data set: 3, 5, 7, 8, 10, 12, 14, 16, 20, find Q1 and Q3, then calculate the IQR.

Answer: _____

118. A researcher wants to know the favorite lunch of all 800 students in a school. She surveys 80 students chosen randomly. What is the sample?

Answer: _____

119. How does using categories help you understand survey results better?

Answer: _____

120. A student made a pictograph where one smiley face = 2 votes. Another student made a pictograph of the same data where one smiley face = 5 votes. Compare how these two pictographs would look different even though they show the same data.

Answer: _____

Match each question to its answer

1. In what situation would the mode be the most useful measure of central tendency?
→

2. A bag contains 5 green tokens and 5 yellow tokens. Maria says the probability of drawing green is 'about half.' Is she correct, and why? →

3. A student wants to determine if a die is fair. They plan to roll it and compare results to theory. How many rolls would give the most reliable conclusion? →

4. A jar has 3 red and 2 green marbles. You draw one, keep it, then draw another. What is the probability of drawing green then red? →

5. A histogram of student heights is bell-shaped and symmetric. Where would you

expect the mean and median to be relative to each other? →

- (A) 600 rolls
 - (B) $\frac{6}{20}$ or $\frac{3}{10}$
 - (C) Yes, because 5 out of 10 tokens are green, which equals $\frac{1}{2}$
 - (D) Approximately equal to each other, both near the center of the distribution
 - (E) When you need the most popular or common value, especially with categorical data like favorite color or most common shoe size sold
-

13. Data Collection & Organization

Circle the correct answer

121. Find the mean, median, mode, and range of the following test scores: 85, 90, 78, 92, 85, 88, 76, 85, 94, 90.

- (A) Mean: 86.3, Median: 85, Mode: 85, Range: 18 (B) Mean: 86.3, Median: 86.5, Mode: 85, Range: 18 (C) Mean: 85, Median: 86.5, Mode: 90, Range: 18 (D) Mean: 86.3, Median: 86.5, Mode: 85, Range: 16

122. What is the probability of an event that is impossible?

- (A) 1 (B) 0 (C) 0.5 (D) -1

123. A bag has 4 red and 6 blue marbles. You draw one marble, do NOT replace it, then draw another. Are these events independent or dependent?

- (A) Dependent, because both marbles have the same shape (B) Independent, because each draw is random (C) Independent, because both draws come from the same bag (D) Dependent, because removing the first marble changes the probabilities for the second draw

124. Marcus rolled a die 30 times and got a 6 exactly 5 times. The theoretical probability of rolling a 6 is $\frac{1}{6}$. Compare these probabilities.

- (A) The experimental probability is lower than theoretical (B) You cannot compare experimental and theoretical probabilities (C) The experimental probability ($\frac{5}{30} = \frac{1}{6}$) exactly matches the theoretical probability in this case (D) The experimental probability is higher than theoretical

125. A survey question asks: 'Don't you agree that our school needs more recess time?' What type of bias does this question introduce?

- (A) Sampling bias, because not all students are surveyed (B) Response bias, because the question is worded in a leading way that encourages a particular answer (C) No bias, because everyone can still answer freely (D) Measurement bias, because recess time is hard to measure

126. A pictograph shows ice cream sales using cone symbols. The key says each cone = 10 sales. Monday has 4 cones, Tuesday has 2.5 cones, Wednesday has 3 cones, Thursday has 1.5 cones, Friday has 5 cones. Calculate the total sales and the difference between the busiest and slowest days.

(A) Total: 160 sales. Difference between busiest and slowest is 25 sales (B) Total: 150 sales. Difference between busiest and slowest is 35 sales (C) Mon: 40, Tue: 25, Wed: 30, Thu: 15, Fri: 50. Total: 160 sales. Difference between busiest (Fri: 50) and slowest (Thu: 15) is 35 sales (D) Total: 160 sales. Difference between busiest and slowest is 40 sales

Write the answer

127. A weather forecast says there is a 60% chance of rain. Marcus claims this means it will definitely rain for 60% of the day. Is Marcus correct?

Answer: _____

128. A teacher adds 10 bonus points to every student's test score. How does this affect the IQR of the class scores?

Answer: _____

129. What type of data describes qualities or characteristics that cannot be measured with numbers, such as favorite color or type of pet?

Answer: _____

130. What does a tally mark of five look like?

Answer: _____

Match each question to its answer

1. Why is a random sample important when making predictions about a population?
→

2. A researcher wants to study whether a new fertilizer increases tomato plant growth. She uses the fertilizer on 20 plants and does not use it on 20 other plants. This is best described as what type of study? →

3. A teacher claims that the range is a complete measure of how spread out data is. Evaluate this claim using these two data sets: Set X: {1, 50, 50, 50, 50, 99} and Set Y: {1, 10, 25, 50, 75, 90, 99}. →

4. Use the following bar graph data about favorite school subjects to answer: Math-9, Science-7, Reading-6, Art-5, Music-3. What fraction of students chose either Math or Science? →

5. A data set has two modes: 15 and 22. What does this tell you about the data? →

- (A) An experiment, because the researcher assigns treatments to groups and compares results
 - (B) The claim is incomplete. Both sets have a range of 98, but Set X clusters tightly around 50 while Set Y is evenly spread. Range only uses two values and ignores how the rest of the data is distributed
 - (C) 16 out of 30, or $\frac{8}{15}$ of students chose Math or Science
 - (D) The data is bimodal, meaning there are two values that each appear more frequently than all other values, suggesting the data may have two distinct clusters
 - (E) It gives every member of the population an equal chance of being selected, reducing bias
-

14. Data Collection & Organization

Circle the correct answer

131. Explain why the mean can be misleading when a data set contains an outlier. Use the data set {10, 12, 11, 13, 50} as an example.

- (A) The outlier makes the mean too low because 50 is so different from the other numbers
- (B) The mean is 12 because you should ignore the outlier before calculating
- (C) The mean is never misleading because it uses every value in the calculation
- (D) The outlier 50 pulls the mean up to 19.2, which is much higher than the typical values (10-13). The mean does not represent what is typical because one extreme value distorts it

132. Two players play a game with a spinner that has 3 equal red sections and 1 blue section. Player A scores a point when it lands on red; Player B scores when it lands on blue. Is this game fair? Explain your reasoning.

- (A) Yes, because the spinner is random so anything can happen
- (B) Yes, because both players can score points
- (C) No, because red is a luckier color than blue
- (D) No, because Player A has a $\frac{3}{4}$ chance of scoring while Player B has only a $\frac{1}{4}$ chance

133. What is the probability of flipping heads on a coin AND rolling a 6 on a die?

- (A) $\frac{1}{12}$
- (B) $\frac{1}{6}$
- (C) $\frac{1}{8}$
- (D) $\frac{1}{2}$

134. Look at two graphs showing the same school fundraiser data. Graph A is a bar graph with an accurate scale. Graph B is a pictograph where each dollar sign = \$50. A parent donating \$75 cannot be represented exactly in Graph B. Which graph is more appropriate for this data and why?

- (A) The pictograph is better because it uses fewer symbols
- (B) Both graphs are equally appropriate for any data, as this is what research and standard practice suggests
- (C) The bar graph is more appropriate because it can show exact values like \$75, while the pictograph must use a partial symbol (1.5 dollar signs) which is harder

to read precisely (D) The pictograph is better because dollar signs are more fun to look at

135. A tree diagram shows 2 branches for event A and 3 branches for event B. If event C adds 2 more branches, how many total outcomes are in the sample space for all three events?

(A) 8 (B) 6 (C) 7 (D) 12

136. A tally chart shows: Red-IIII II, Blue-IIII IIII, Green-III. What does this tally chart tell you about which color is most popular?

(A) Red is most popular because it is listed first in the chart (B) Blue is most popular with 10 tallies, followed by Red with 7, then Green with 5 (C) All three colors are equally popular because they all use tally marks (D) Green is most popular because it has exactly one group of five

Write the answer

137. A line plot shows daily temperatures for a week: 68, 70, 72, 71, 69, 74, 75. A student says, 'The line plot shows that temperature is increasing every day.' Analyze whether this conclusion is accurate.

Answer: _____

138. When you roll a standard die, the sample space is {1, 2, 3, 4, 5, 6}. How many outcomes are in the sample space?

Answer: _____

139. What is theoretical probability based on?

Answer: _____

140. Which of the following pairs of events is independent?

Answer: _____

Match each question to its answer

1. Sam says, 'I flipped a coin 5 times and got heads every time, so the next flip is more likely to be tails.' What is wrong with Sam's reasoning? →

2. In statistics, what is a population? →

3. Data set A: {5, 5, 5, 5, 5, 5, 5, 5, 100}. Data set B: {15, 15, 15, 15, 15, 15, 15, 15, 15}.

Calculate the mean and median for each set. Which measure better represents each data set? →

4. You draw a tree diagram for rolling two dice. How many total outcomes appear at the tips of the branches? →

5. A bag contains 3 red marbles and 7 blue marbles. What is the probability of randomly drawing a red marble? →

- (A) Set A: Mean=15.6, Median=5. Median is better because most values are 5. Set B: Mean=15, Median=15. Both are equally good because the data is uniform
- (B) 3/10
- (C) Each coin flip is independent, so past results do not change the probability of the next flip
- (D) 36
- (E) The entire group of individuals or items that you want to study
-

15. Data Collection & Organization

Circle the correct answer

141. In a pictograph, what is a key (or legend)?

- (A) The largest value in the data set (B) A symbol explanation that tells how many items each picture represents (C) The title of the pictograph (D) The total number of pictures in the graph

142. Which of the following is NOT a possible probability value?

- (A) 0.75 (B) 1.5 (C) 0 (D) 1

143. Emma says, 'I don't need to do an experiment because I can just calculate the theoretical probability.' When might she be wrong?

- (A) When the experiment has more than 10 possible outcomes (B) When the outcomes are not equally likely or the situation is too complex to analyze mathematically (C) She is never wrong because theoretical probability is always better (D) When she uses a calculator instead of doing math by hand

144. What does the line inside the box of a box plot represent?

- (A) The median (B) The mean (C) The midrange (D) The mode

145. Two different random samples of 50 students each are taken from the same school. Sample A finds that 60% prefer math. Sample B finds that 54% prefer math. Why do the results differ?

- (A) Sampling variability — different random samples naturally produce slightly different results (B) 50 students is too small a sample to give any useful results (C) The school changed between the two surveys (D) One sample was collected incorrectly

146. A data set about daily temperatures has an outlier of -15°F when all other values are between 50°F and 80°F . What effect does this outlier have on the mean versus the median?

- (A) The outlier pulls the mean down significantly but has little effect on the median (B) Both the mean and median are equally affected (C) The outlier pulls the median

down significantly but has little effect on the mean (D) Neither the mean nor the median is affected

Write the answer

147. A bar graph shows rainfall in inches for each month: Jan-3, Feb-2, Mar-4, Apr-6, May-5, Jun-3. What was the total rainfall for the first six months? Which month had double the rainfall of February?

Answer: _____

148. A data set has a mean of 20 and a median of 20. A student says, 'Since the mean equals the median, the data must be perfectly symmetric.' Evaluate this claim.

Answer: _____

149. A coin is flipped 20 times with 14 heads. The theoretical probability of heads is $\frac{1}{2}$. A student claims the coin is unfair. Evaluate this claim.

Answer: _____

150. Two dice are rolled. Jake says there are 11 outcomes because the sums range from 2 to 12. Is Jake correct?

Answer: _____

Match each question to its answer

1. What is the interquartile range (IQR)? →
2. What type of graph uses rectangular bars to compare amounts in different categories? →
3. Two events are independent if which of the following is true? →
4. Can the mean of a data set be a value that does not appear in the data set? Explain with an example. →
5. You spin a spinner with sections A, B, C, and then flip a coin. Using a tree diagram, how many branches would appear at the final level? →

(A) Yes. For the data set $\{1, 2, 3, 4\}$, the mean is $(1+2+3+4)/4 = 10/4 = 2.5$, which is not one of the actual data values

(B) 6

(C) The difference between the third quartile (Q3) and the first quartile (Q1)

(D) Bar graph

(E) The outcome of one event does not affect the outcome of the other

16. Data Collection & Organization

Circle the correct answer

151. A game uses a spinner with 5 equal sections. Redesign the spinner colors so that the game is fair for two players, where Player A wins on red and Player B wins on blue. How should the sections be colored?

(A) Color all 5 sections half red and half blue, and this has been consistently demonstrated in practice (B) Color 3 sections red and 2 sections blue, as this is what research and standard practice suggests (C) Color 1 section red and 4 sections blue, and this principle applies across similar situations (D) It is impossible to make it perfectly fair with 5 equal sections for 2 players because 5 is odd and cannot be split evenly

152. After 100 rolls of a standard die, the number 6 appeared 22 times. After 1,000 rolls, it appeared 172 times. Which result is likely closer to the theoretical probability?

(A) Both are equally close because they used the same die (B) The 1,000-roll result ($172/1000 = 0.172$) is likely closer to $1/6 \approx 0.167$ (C) Neither is close because the theoretical probability is wrong (D) The 100-roll result because it was done first

153. What is a sample space?

(A) The area where an experiment takes place (B) The number of times an experiment is repeated (C) The set of all possible outcomes of an experiment (D) The most likely outcome of an experiment

154. Read the line plot showing shoe sizes of 20 students: two X's above 4, three X's above 4.5, four X's above 5, five X's above 5.5, three X's above 6, two X's above 6.5, one X above 7. Determine how many students wear size 5.5 or larger.

(A) 11 students wear size 5.5 or larger ($5 + 3 + 2 + 1 = 11$) (B) 14 students wear size 5.5 or larger (C) 6 students wear size 5.5 or larger (D) 9 students wear size 5.5 or larger

155. A card is drawn at random from 10 cards numbered 1-10. What is the probability of drawing a number greater than 7?

(A) $3/10$ (B) $3/7$ (C) $1/3$ (D) $7/10$

156. As the number of trials in a probability experiment increases, what generally happens to the experimental probability?

(A) It stays exactly the same after each trial (B) It gets farther from the theoretical probability (C) It gets closer to the theoretical probability (D) It always equals exactly 50%

Write the answer

157. A spinner has 4 equal sections colored red, blue, green, and yellow. How many equally likely outcomes are there for one spin?

Answer: _____

158. A spinner has 3 equal sections: A, B, C. The theoretical probability of landing on each is $\frac{1}{3}$. After 90 spins: A appeared 25 times, B appeared 35 times, C appeared 30 times. Which section's experimental probability is closest to its theoretical probability?

Answer: _____

159. Which of the following is an example of a confounding variable? A study finds that towns with more firefighters have more fires.

Answer: _____

160. You surveyed 30 classmates about how many siblings they have. The responses are: 0, 1, 1, 2, 0, 3, 1, 2, 1, 1, 0, 2, 1, 1, 3, 2, 0, 1, 2, 1, 0, 1, 2, 3, 1, 0, 2, 1, 3, 2. Organize this data into a frequency table.

Answer: _____

Match each question to its answer

1. Design a simulation using a random number generator (1-8) to model the following situation: a basketball player makes 75% of her free throws. Describe what numbers represent a made shot and a missed shot. →

2. A restaurant offers 3 entrees and 4 desserts. Using the counting principle, how many different meal combinations (one entree, one dessert) are possible? →

3. A student claims: 'Since our sample of 50 people showed 60% like dogs, exactly 60% of ALL people like dogs.' Evaluate this claim. →

4. A news report says: 'The average household income in our city is \$95,000.' A community organizer responds: 'That is misleading. The median household income is \$52,000.' Analyze which figure gives a more accurate picture of typical household income and why the two numbers differ so much. →

5. A bar graph shows test scores for two classes. Class A's bars range from 70 to 95 with a y-axis starting at 0. Class B's bars range from 70 to 95 with a y-axis starting at 65. Both use the same data. Why might someone choose to start the y-axis at 65 instead of 0? →

(A) The claim is too strong; the sample gives an estimate, but the true population percentage could be somewhat higher or lower due to sampling variability

(B) 12

(C) Numbers 1-6 represent a made shot ($\frac{6}{8} = 75\%$), and numbers 7-8 represent a miss ($\frac{2}{8} = 25\%$)

(D) Starting at 65 makes the differences between bars more visible, which can be useful for seeing small variations but can also exaggerate how different the scores really are

(E) The median (\$52,000) gives a more accurate picture because a few very wealthy households pull the mean up to \$95,000, far above what most households actually

earn. Income data is typically right-skewed

Riddle & Joke Break

Guess the answer, then check the key!

- 161. Why did the coin go to therapy?
 - 162. Why was the statistics teacher always calm?
 - 163. Why did the dice refuse to play?
 - 164. What did the normal distribution say at the party?
 - 165. Why did the outlier get excluded from the data set?
 - 166. What did the probability say to the certainty?
-

Answer Key

- 1. D — A method of collecting data by asking people questions
 - 2. B — Four vertical lines with one diagonal line crossing through them
 - 3. C — A table that shows how often each value or category appears in a data set
 - 4. B — Categorical data
 - 5. A — Because every student surveyed must be accounted for in exactly one category
 - 6. C — Blue is most popular with 10 tallies, followed by Red with 7, then Green with 5
 - 7. A frequency table summarizes the data so you can quickly see how many times each response appears
 - 8. Categorical data sorts into groups (like favorite sport), while numerical data is measured with numbers (like height in inches)
 - 9. Organized data makes it easier to find patterns, compare values, and draw conclusions
 - 10. Categories group similar responses together so you can count and compare them
- Matching (Data Collection & Organization): 1→C, 2→D, 3→A, 4→E, 5→B

11. B — Bar graph
12. B — A graph that shows data using marks (usually X's or dots) above a number line
13. B — A symbol explanation that tells how many items each picture represents
14. A — A title, labeled axes (or categories), and a scale or key
15. B — Taller bars mean more books read. Ava read the most (15), Sam read the fewest (6), and the difference between the tallest and shortest bar shows the range of reading amounts
16. D — Most students scored between 85 and 90, which shows the data clusters in that range and represents the most common performance level
17. 14 students chose Soccer. Three full stars = $3 \times 4 = 12$ students, plus one half star = $4 / 2 = 2$ students, so $12 + 2 = 14$
18. Starting above zero exaggerates small differences between bars, making them look much larger than they actually are
19. A line plot works well because the data is numerical with a small range of values (0-5 pets), making it easy to see the frequency of each value on a number line
20. Pictographs use familiar pictures that make data more visual and easier for younger students to understand and count
Matching (Graphs & Data Displays): $1 \rightarrow C$, $2 \rightarrow E$, $3 \rightarrow B$, $4 \rightarrow A$, $5 \rightarrow D$
21. D — The sum of all values divided by the number of values
22. D — The middle value when all values are arranged in order from least to greatest
23. D — The value that appears most frequently in the data set
24. C — Subtract the smallest value from the largest value
25. B — The outlier 50 pulls the mean up to 19.2, which is much higher than the typical values (10-13). The mean does not represent what is typical because one extreme value distorts it
26. C — The median is resistant to extreme values (outliers) because it only depends on the middle position, not the actual values of all data points
27. The data is bimodal, meaning there are two values that each appear more frequently than all other values, suggesting the data may have two distinct clusters

28. The mean tells you the center or typical value of the data, while the range tells you how spread out the data is from lowest to highest
29. When you need the most popular or common value, especially with categorical data like favorite color or most common shoe size sold
30. Yes. For the data set {1, 2, 3, 4}, the mean is $(1+2+3+4)/4 = 10/4 = 2.5$, which is not one of the actual data values
Matching (Mean, Median, Mode & Range): 1→A, 2→E, 3→D, 4→B, 5→C
31. C — 1
32. A — 0
33. D — $3/10$
34. D — 1.5
35. B — 4
36. D — $1/2$
37. Unlikely
38. Each player rolls a die and the highest number wins
39. 0.3
40. Rolling a number less than 5
Matching (Probability Basics): 1→B, 2→C, 3→D, 4→E, 5→A
41. A — Mathematical reasoning about equally likely outcomes
42. C — The results of actually performing trials of an experiment
43. D — $28/50$
44. C — 10
45. D — Small numbers of trials often produce results that differ from theoretical probability due to natural variation
46. D — As the number of trials increases, experimental probability gets closer to theoretical probability
47. $52/200$
48. The experimental probability ($5/30 = 1/6$) exactly matches the theoretical probability in this case
49. A simulation can perform many trials quickly, giving results closer to theoretical

probability

50. The number 4 appeared 500 times

Matching (Experimental vs. Theoretical Probability): $1 \rightarrow D$, $2 \rightarrow E$, $3 \rightarrow A$, $4 \rightarrow C$, $5 \rightarrow B$

51. B — The set of all possible outcomes of an experiment

52. B — An event that consists of two or more simple events

53. D — 12

54. D — A tree diagram

55. C — The outcome of one event does not affect the outcome of the other

56. A — Dependent, because removing the first marble changes the probabilities for the second draw

57. 6

58. $1/12$

59. $1/12$

60. $25/64$

Matching (Compound Events & Sample Spaces): $1 \rightarrow A$, $2 \rightarrow B$, $3 \rightarrow E$, $4 \rightarrow C$, $5 \rightarrow D$

61. C — How spread out or different the data values are from each other

62. B — 18

63. C — The middle 50% of the data, from the first quartile (Q1) to the third quartile (Q3)

64. D — 80-89

65. C — The difference between the third quartile (Q3) and the first quartile (Q1)

66. B — $Q1 = 6$, $Q3 = 15$, $IQR = 9$

67. Interquartile range (IQR)

68. 50

69. The median

70. Skewed right (positively skewed)

Matching (Statistical Variability & Distributions): $1 \rightarrow C$, $2 \rightarrow A$, $3 \rightarrow D$, $4 \rightarrow B$, $5 \rightarrow E$

71. A — The entire group of individuals or items that you want to study

72. D — A subset of the population selected to represent the whole group

73. B — The 80 randomly chosen students
74. D — When certain members of the population are more likely to be included in the sample than others
75. B — Positive correlation
76. A — Randomly selecting 50 students from a list of all 500 students
77. 210
78. Sampling bias, because surveying only Saturday morning shoppers may not represent all customers
79. Negative correlation
80. 3,000
Matching (Statistical Inference & Prediction): 1→B, 2→D, 3→A, 4→C, 5→E
81. B — Categorical data sorts into groups (like favorite sport), while numerical data is measured with numbers (like height in inches)
82. D — Graph 1 (1-inch increments) is better because it shows the detailed variation in heights, while Graph 2 (6-inch increments) would clump most students into the same bar, hiding the range
83. C — Class A performed more consistently. Class A's range is 20 (90-70) while Class B's range is 40 (100-60). Class A's scores are closer together, showing more uniform performance
84. D — Each player rolls a die and the highest number wins
85. B — Small numbers of trials often produce results that differ from theoretical probability due to natural variation
86. A — $\frac{1}{8}$
87. Class B has more variability in the middle 50% of scores than Class A
88. 1,000
89. This is false. A tally chart and frequency table display the same information in different formats; both show how many times each category occurs
90. Taller bars mean more books read. Ava read the most (15), Sam read the fewest (6), and the difference between the tallest and shortest bar shows the range of reading amounts
Matching (Experimental Probability): 1→E, 2→C, 3→D, 4→B, 5→A
91. D — Student A's approach is better for understanding specific preferences

because the 5 categories reveal more detail about what students actually enjoy reading

92. B — Starting above zero exaggerates small differences between bars, making them look much larger than they actually are
93. A — The mode is 15 (appears 3 times). It tells the coach that scoring around 15 points is this player's most common performance level
94. D — {red, blue, white}
95. B — The results of actually performing trials of an experiment
96. D — {M1, M2, M3, A1, A2, A3, T1, T2, T3, H1, H2, H3}
97. Histogram
98. Sampling bias, because surveying only Saturday morning shoppers may not represent all customers
99. Dogs: 13, Cats: 10, Fish: 6, Birds: 3. Total: 32 pets
100. Pictographs use familiar pictures that make data more visual and easier for younger students to understand and count
Matching (Compound Events & Trees): $1 \rightarrow A$, $2 \rightarrow D$, $3 \rightarrow E$, $4 \rightarrow B$, $5 \rightarrow C$
101. C — What is your favorite after-school activity? A) Sports B) Reading C) Video games D) Art E) Music
102. D — A bar graph with months on the x-axis and number of sunny days on the y-axis, because it shows the pattern of increasing then decreasing sunny days across the year
103. C — Subtract the smallest value from the largest value
104. A — 5
105. A — They would get progressively closer to the theoretical probability
106. C — $\frac{1}{12}$
107. The middle 50% of the data, from the first quartile (Q1) to the third quartile (Q3)
108. Negative correlation
109. a. Create a survey asking students to choose one day, 2) Collect all responses, 3) Tally or count responses for each day, 4) Make a frequency table, 5) Display results in a bar graph
110. 14 students chose Soccer. Three full stars = $3 \times 4 = 12$ students, plus one half star

$= 4 / 2 = 2$ students, so $12 + 2 = 14$

Matching (Statistical Distributions): $1 \rightarrow A$, $2 \rightarrow D$, $3 \rightarrow B$, $4 \rightarrow C$, $5 \rightarrow E$

- 111. C — Organized data makes it easier to find patterns, compare values, and draw conclusions
- 112. A — Most students scored between 85 and 90, which shows the data clusters in that range and represents the most common performance level
- 113. B — The claim is wrong. The mean salary is \$76,000, but 4 of 5 employees earn between \$30K-\$35K. The median (\$33,000) better represents the typical salary because the \$250K outlier distorts the mean
- 114. B — 1
- 115. C — The number 4 appeared 500 times
- 116. C — Sarah is incorrect; the events are independent because the outcome of the coin does not affect the outcome of the die, regardless of timing
- 117. $Q1 = 6$, $Q3 = 15$, $IQR = 9$
- 118. The 80 randomly chosen students
- 119. Categories group similar responses together so you can count and compare them
- 120. The first graph would have more symbols (2.5 times as many) making differences between categories look larger, while the second would have fewer symbols making the graph more compact but harder to show exact values

Matching (Data-Driven Decision Making): $1 \rightarrow E$, $2 \rightarrow C$, $3 \rightarrow A$, $4 \rightarrow B$, $5 \rightarrow D$

- 121. B — Mean: 86.3, Median: 86.5, Mode: 85, Range: 18
- 122. B — 0
- 123. D — Dependent, because removing the first marble changes the probabilities for the second draw
- 124. C — The experimental probability ($5/30 = 1/6$) exactly matches the theoretical probability in this case
- 125. B — Response bias, because the question is worded in a leading way that encourages a particular answer
- 126. C — Mon: 40, Tue: 25, Wed: 30, Thu: 15, Fri: 50. Total: 160 sales. Difference between busiest (Fri: 50) and slowest (Thu: 15) is 35 sales
- 127. No, it means that in similar weather conditions, rain occurs about 60 out of 100 times

128. The IQR stays the same because adding a constant shifts all values equally
129. Categorical data
130. Four vertical lines with one diagonal line crossing through them
Matching (Data Collection & Organization): $1 \rightarrow E$, $2 \rightarrow A$, $3 \rightarrow B$, $4 \rightarrow C$, $5 \rightarrow D$
131. D — The outlier 50 pulls the mean up to 19.2, which is much higher than the typical values (10-13). The mean does not represent what is typical because one extreme value distorts it
132. D — No, because Player A has a $\frac{3}{4}$ chance of scoring while Player B has only a $\frac{1}{4}$ chance
133. A — $\frac{1}{12}$
134. C — The bar graph is more appropriate because it can show exact values like \$75, while the pictograph must use a partial symbol (1.5 dollar signs) which is harder to read precisely
135. D — 12
136. B — Blue is most popular with 10 tallies, followed by Red with 7, then Green with 5
137. The conclusion is inaccurate. While there is a general upward trend, the temperature decreased on Day 4 (72 to 71) and Day 5 (71 to 69). It is not a steady increase every day
138. 6
139. Mathematical reasoning about equally likely outcomes
140. Rolling a die and spinning a spinner
Matching (Data Collection & Organization): $1 \rightarrow C$, $2 \rightarrow E$, $3 \rightarrow A$, $4 \rightarrow D$, $5 \rightarrow B$
141. B — A symbol explanation that tells how many items each picture represents
142. B — 1.5
143. B — When the outcomes are not equally likely or the situation is too complex to analyze mathematically
144. A — The median
145. A — Sampling variability — different random samples naturally produce slightly different results
146. A — The outlier pulls the mean down significantly but has little effect on the median

147. Total rainfall: 23 inches. March had 4 inches, which is double February's 2 inches
148. The claim is not necessarily true. While symmetric data always has equal mean and median, equal mean and median does not guarantee symmetry. The data set {10, 20, 20, 20, 30} has mean=median=20 and is symmetric, but {5, 20, 20, 25, 30} also has mean=median=20 and is not symmetric
149. The claim is not well-supported because 20 trials is too few to conclude unfairness; natural variation could explain the result
150. No, there are 36 outcomes because each die independently shows 1-6, and different combinations can produce the same sum
Matching (Data Collection & Organization): 1→C, 2→D, 3→E, 4→A, 5→B
151. D — It is impossible to make it perfectly fair with 5 equal sections for 2 players because 5 is odd and cannot be split evenly
152. B — The 1,000-roll result ($172/1000 = 0.172$) is likely closer to $1/6 \approx 0.167$
153. C — The set of all possible outcomes of an experiment
154. A — 11 students wear size 5.5 or larger ($5 + 3 + 2 + 1 = 11$)
155. A — $3/10$
156. C — It gets closer to the theoretical probability
157. 4
158. C
159. Town size — larger towns have both more firefighters and more fires because there are more buildings
160. 0 siblings: 6, 1 sibling: 12, 2 siblings: 8, 3 siblings: 4
Matching (Data Collection & Organization): 1→C, 2→B, 3→A, 4→E, 5→D
161. It couldn't stop flipping between two sides of every issue!
162. Because she knew things would eventually return to the mean!
163. They felt like they were always being thrown around!
164. "I'm the bell of the ball!"
165. It just didn't fit in!
166. "You're a 1 in a million... actually, exactly a 1."
-

Spark & Anvil is a 501(c)(3) public charity. Apps + books are free forever.