Tables, Statistics, and Probability

LEARNING OBJECTIVES

After completing this chapter, you will be able to:

- Draw inferences about data presented in a variety of graphical formats
- Find an unknown value given the average
- Calculate mean, median, mode, and range
- Describe standard deviation and margin of error
- Determine whether a survey is valid or biased
- Draw inferences about surveys and data samples
- Calculate probabilities based on data sets

50/600 SmartPoints® (Medium Yield)

How Much Do You Know?

Directions: Try the questions that follow. Show your work so that you can compare your solutions to the ones found in the Check Your Work section immediately after this question set. The "Category" heading in the explanation for each question gives the title of the lesson that covers how to solve it. If you answered the question(s) for a given lesson correctly, and if your scratchwork looks like ours, you may be able to move quickly through that lesson. If you answered incorrectly or used a different approach, you may want to take your time on that lesson.

Questions 1 and 2 refer to the following information.

The amount of glucose, or sugar, in a person's blood is the primary indicator of diabetes. When a person without diabetes fasts (doesn't eat) for eight hours prior to taking a blood sugar test, that person's glucose level will be below 100 milligrams per deciliter. An individual is considered at risk for diabetes, but is not diagnosed as diabetic, when fasting glucose levels are between 100 and 125. If the level is above 125, the person is considered to have diabetes. The following table shows the ages and glucose levels of a group of study participants.

Study Results				
Age Group	<100 mg/dL	100–125 mg/dL	>125 mg/dL	Total
18–25	9	22	17	48
26–35	16	48	34	98
36–45	19	35	40	94
Older than 45	12	27	21	60
Total	56	132	112	300

1. According to the data, which age group had the smallest percentage of people with a healthy blood sugar level?

   A) 18–25
   B) 26–35
   C) 36–45
   D) Older than 45

2. Based on the table, if a single participant is selected at random from all the participants, what is the probability that he or she will be at risk for diabetes and be at least 36 years old?

   A) $\frac{7}{60}$

   B) $\frac{11}{25}$

   C) $\frac{31}{77}$

   D) $\frac{31}{150}$

History Majors Declared at College X

Year	Number of History Majors
2010	225
2011	287
2012	162
2013	240
2014	s

3.   The table above shows the number of history majors declared each year at a certain college from 2010 to 2014. If the median number of history majors declared for the five years was 225, what is the greatest possible value of s ?

A)   161

B)   225

C)   239

D)   288

4.   A writers association sponsored a nationwide convention attended by 1,650 nonfiction writers. Before the convention, the association surveyed 150 of the writers who were planning to attend (chosen at random) about their lunch preferences. Thirty-eight said they preferred salads, 23 preferred pizza, 59 preferred sandwiches, and 30 preferred grilled chicken. Based on the results of this survey, how many of the writers attending the convention can be expected to want sandwiches for lunch?

5.   A researcher conducted a poll to determine how many people in a city of 100,000 residents enjoy the taste of nondairy milks such as coconut milk or almond milk. The researcher polled 800 city residents who are allergic to dairy. Of those polled, 72 percent responded that they enjoyed the taste of nondairy milks.

Which of the following indicates why the survey results would not allow for a reliable conclusion about the taste preferences of the city's residents?

A)   The researcher did not ask people if they prefer the taste of dairy milk to nondairy milks.

B)   The survey sample is not representative of the city's residents.

C)   The population of the city is too large to get a reliable survey sample.

D)   The survey sample likely consisted of only adults and did not consider the opinion of children.

Cookies Baked

	Chocolate Chip	Oatmeal Raisin	Total
With Nuts		40	
Without Nuts			104
Total			186

6.   A baker makes 186 cookies. Some are chocolate chip and some are oatmeal raisin, and both kinds are made with and without nuts, as shown in the table above. Because they are more popular, the baker makes $\frac{2}{3}$ of the cookies chocolate chip. If a chocolate chip cookie is chosen at random, what is the probability that it will have nuts?

A) $\frac{21}{93}$

B) $\frac{21}{62}$

C) $\frac{41}{93}$

D) $\frac{21}{41}$

Check Your Work

1. B

Difficulty: Medium

Category: Tables and Graphs

Getting to the Answer: To calculate the percentage of people in each age group with a healthy blood sugar level (<100 mg/dL), divide the number of people in that age group with a healthy blood sugar level by the total number of participants in that same age group and multiply by 100%. Choice **(B)** is correct because $\frac{16}{98} \times 100\% \approx 0.1633 \times 100\% = 16.33\%$, which is a lower percentage than in the other age groups (18.75% for 18–25; 20.21% for 36–45; and 20% for Older than 45).

2. D

Difficulty: Medium

Category: Probability

Getting to the Answer: This question requires careful reading of the table. The first criterion is fairly straight-forward—you're looking for a participant with a blood sugar level in the 100–125 mg/dL range, so focus on that column in the table. The second criterion is a bit trickier— "at least 36 years old" means you'll need to use the values in both the row for 36–45 and the row for Older than 45. Within the 100–125 mg/dL range, there are 35 in the 36–45 age group and 27 in the Older than 45 age group, resulting in a total of $35 + 27 = 62$ out of 300 total participants overall. The probability of randomly selecting one participant who fits the criteria, therefore, is $\frac{62}{300}$, which reduces to $\frac{31}{150}$, or **(D)**.

3. B

Difficulty: Medium

Category: Statistics

Getting to the Answer: The median is the middle number in a series of numbers. Arrange the number of history majors from least to greatest, making sure that 225 is in the middle. Use *s* to balance out the number of history majors on either side of 225. Because there are already two numbers above the median (240 and 287), there must be two numbers below the median, 162 and *s*:

s, 162, 225, 240, 287

or

162, *s*, 225, 240, 287

Because *s* could be on either side of 162, it could be anything less than or equal to 225. Its greatest possible value is therefore 225, which is **(B)**.

4. 649

Difficulty: Medium

Category: Surveys and Data Samples

Getting to the Answer: The question indicates that of the 150 survey respondents, 59 wanted sandwiches. To find the number of the writers attending the convention who can be expected to want sandwiches for lunch, set up a proportion using the number of survey respondents who want sandwiches, the total number of survey respondents, and the total number of writers attending the convention: $\frac{59}{150} = \frac{x}{1,650}$. Cross-multiply to find that $x = 649$. Grid in **649**.

5. B

Difficulty: Medium

Category: Surveys and Data Samples

Getting to the Answer: Determine who is being surveyed and what results are intended. In this question, the intended result is to determine whether people in a large city like nondairy milks. To obtain a representative sample of the population, the survey should randomly select individuals from the population. However, the researcher polled only people who have a dairy allergy. Those people may have acquired tastes and may not be representative of the general population of the city, making **(B)** the correct answer.

6. B

Difficulty: Medium

Category: Probability

Getting to the Answer: The table is not complete, so your first step is to fill in the missing values. Start with what you know and work from there. It may not be necessary to complete the entire table, so stop when you have enough information to answer the question.

There are 186 cookies total and 104 are without nuts, which means $186 - 104 = 82$ have nuts. Because the table already indicates that 40 of those cookies are oatmeal raisin, $82 - 40 = 42$ are chocolate chip. Recall that $\frac{2}{3}$ of the total number of cookies are chocolate chip, which means there are $\frac{2}{3} \times 186 = 124$ chocolate chip cookies, so you can fill this number in the "Total" row of the chocolate chip column. You do not need to fill in any more of the table because the question asks only about chocolate chip cookies with nuts. There are 124 chocolate chip cookies total and 42 of them have nuts, so the probability of randomly choosing one with nuts is $\frac{42}{124}$, or $\frac{21}{62}$, which is **(B)**.
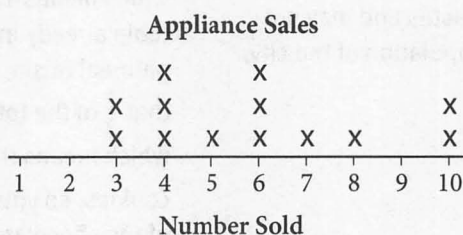
Tables and Graphs

LEARNING OBJECTIVES

After this lesson, you will be able to:

- Draw inferences about data presented in a variety of graphical formats
- Find an unknown value given the average

To solve a question like this:

Appliance Sales

```
        X       X
    X   X       X               X
    X   X   X   X   X   X        X
  ─────────────────────────────────
    1   2   3   4   5   6   7   8   9   10
```
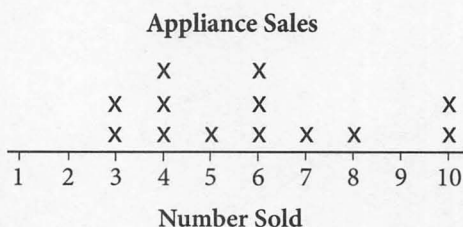
Number Sold

An appliance salesperson sets a goal to sell an average of 6 appliances per day for the first two weeks of his new job. The dot plot shows the number he sold each day during the first 13 days. What is the minimum number of appliances he must sell on the 14th day in order to reach his goal?

A) 5

B) 6

C) 7

D) 8

You need to know this:

The SAT uses some straightforward methods of representing data sets that you are certainly already familiar with. You likely don't need to review, for example, how to look up information in a table or read a bar chart. There are, however, some less common types of plots that show up from time to time that can be confusing at first glance. Graphics you may see on test day include the following:

- **Tables, bar charts, and line graphs** show up all the time in the Math sections (and in the Reading and Writing and Language sections, too). They shouldn't be difficult to interpret, but it's helpful to keep in mind that the test maker often includes more information than you actually need. It's important to consider what the question asks for so that you find only the information that you need.

- **Frequency tables and dot plots** are ways of representing how many times a data point appears within a data set. The sample problem presents its data as a dot plot:

Appliance Sales

```
        X       X
    X   X       X               X
    X   X   X   X   X   X        X
  ─────────────────────────────────
    1   2   3   4   5   6   7   8   9   10
```

Number Sold

Each "X" represents one instance in the data set of each "number sold." So, for example, there were two different days on which this person sold 3 appliances, three different days on which this person sold 4 appliances, and so on. The data could just as easily be written as a data set: {3, 3, 4, 4, 4, 5, 6, 6, 6, 7, 8, 10, 10}. Or it could be placed in a frequency table:

Number Sold	Frequency
1	0
2	0
3	2
4	3
5	1
6	3
7	1
8	1
9	0
10	2

- **Histograms** look a lot like bar charts and can be read in the same way, but they are similar to frequency tables and dot plots in that they show how many times a certain value shows up in a data set for a variable. The histogram for the appliances data set would look like this:
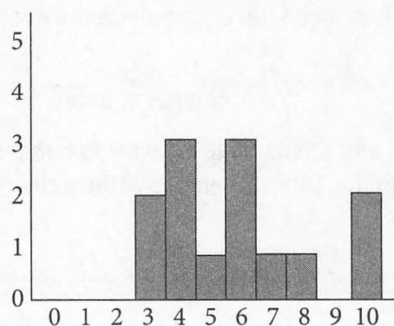
Notice that the histogram is basically the same as the dot plot for this data set. Histograms are better for representing larger data sets for which individual dots would be difficult to count.

You need to do this:

- When presented with a question that uses a graph or table to present information, first inspect the format of the graph or table. What kind of graph or table is it? What information is presented on each axis? What information do you need to find in order to answer the question?

- Find the information you need from the table or graph and then use the information for any calculation the question might require, such as taking the average, finding the median, or thinking about standard deviation.

- Use the average formula, $\text{average} = \dfrac{\text{sum}}{\text{number of items}}$, to find unknowns. For example, if you know that the average of 5 terms is 7, and you know that 4 of the terms are 3, 6, 8, and 9, you can call the last term x and plug into the equation, then solve for x:

$$7 = \frac{3+6+8+9+x}{5}$$
$$35 = 26 + x$$
$$x = 9$$

Explanation:

This question gives you an average and asks for a missing value, which is a kind of calculation that shows up in word problems all the time. First, set up a general equation for the average:

$$\text{Average} = \frac{\text{sum}}{\text{number of items}}$$

The scenario takes place over 14 days, and the average is given as 6 items per day. Let a represent the unknown number of appliances sold on the 14th day and then fill in the number of appliances sold the previous days from the dot plot:

$$6 = \frac{3+3+4+4+4+5+6+6+6+7+8+10+10+a}{14}$$

Multiply both sides by 14 to get rid of the fraction and simplify the addition on the right before isolating a:

$$84 = 3+3+4+4+4+5+6+6+6+7+8+10+10+a$$
$$84 = 76 + a$$
$$a = 8$$

Choice **(D)** is correct.

Try on Your Own

Directions: Take as much time as you need on these questions. Work carefully and methodically. There will be an opportunity for timed practice at the end of the chapter.

Questions 1 and 2 refer to the following information.

	Bob's Bookshop	Clara's Bookshop	Derek's Bookshop	Evelyn's Bookshop	Total
Monday	14	7	15	12	48
Tuesday	8	13	15	13	49
Wednesday	10	13	12	14	49
Thursday	8	15	14	10	47
Friday	13	7	10	9	39
Total	53	55	66	58	232

HINT: For Q1, fraction $= \dfrac{\text{part}}{\text{whole}}$. Which *part* is the question asking for? Out of which *whole*?

1. Which of the four bookshops made the greatest fraction of its total sales on Tuesday?

   A) Bob's Bookshop

   B) Clara's Bookshop

   C) Derek's Bookshop

   D) Evelyn's Bookshop

2. What fraction of all the books sold on Monday, Wednesday, and Friday were sold at Derek's Bookshop and Evelyn's Bookshop?

   A) $\dfrac{9}{29}$

   B) $\dfrac{11}{32}$

   C) $\dfrac{9}{17}$

   D) $\dfrac{18}{31}$

Questions 3 and 4 refer to the following information.

Numerous health studies have found that people who eat breakfast are generally healthier and weigh less than people who skip this meal. The following table shows the results of a study related to this topic.

Breakfast Study Results

	Breakfast ≤ 1 Time per Week	Breakfast 2–4 Times per Week	Breakfast 5–7 Times per Week	Total
Within Healthy Weight Range	6	15	36	57
Outside Healthy Weight Range	38	27	9	74
Total	44	42	45	131

3. What percent of the participants who were outside a healthy weight range ate breakfast one or fewer times per week?

   A) 29.01%

   B) 51.35%

   C) 56.49%

   D) 86.36%

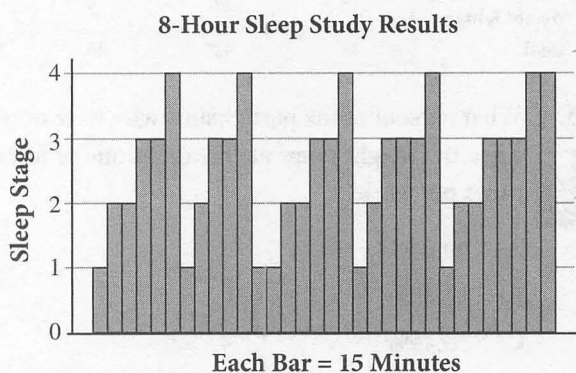HINT: For Q4, which group in the study is of interest to this company?

4. A large company that provides breakfast for all its employees wants to determine how many of them are likely to be within a healthy weight range, given that all the employees take advantage of the free breakfast all five weekdays. If the company has 3,000 employees, and assuming the participants in the study are a good representative sample, about how many of the employees are likely to be within a healthy weight range?

   A) 825

   B) 1,030

   C) 1,900

   D) 2,400

Questions 5 and 6 refer to the following information.

When people sleep, they experience various types of brain activity. Scientists have classified these types of activity into four sleep stages: 1, 2, 3, and 4 (also known as REM). Stage 3 is the only stage considered to be deep sleep. Suppose a person went to a sleep clinic to have his or her sleeping brainwaves analyzed. A technician monitored the person's brainwaves in 15-minute intervals, for 8 continuous hours, and categorized them into one of the four stages. The bar graph below shows the results of the one-night study.

8-Hour Sleep Study Results

Each Bar = 15 Minutes

5. Based on the graph, how many minutes did the patient spend in non-deep sleep over the course of the entire night?

6. After the completion of the one-night study, the patient was monitored an additional four nights. Over the total number of nights that the patient spent at the clinic, he spent an average of 180 minutes in stage 3 sleep per night. If the patient spent an average of 175 minutes in stage 3 sleep on the second, third, and fourth nights, how many minutes did he spend in stage 3 sleep on the last night?
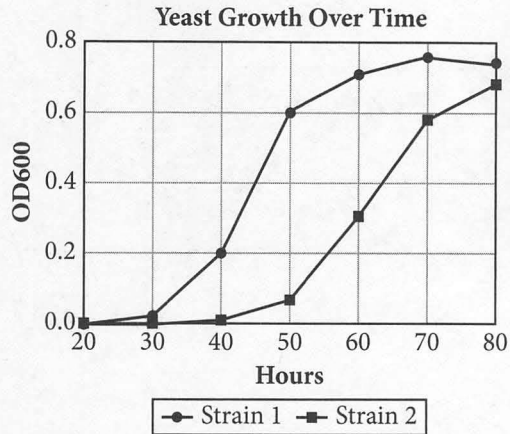
Yeast Growth Over Time

7. A microbiologist is comparing the growth rates of
    two different yeast strains. She indirectly measures
    the number of yeast cells by recording the optical
    density (OD600) of each strain every 10 hours. The
    measurements are presented in the graph above.
    Based on the data, which of the following is NOT a
    true statement?

    A) Between hours 30 and 80, Strain 1 had a
        higher OD600 reading than Strain 2.

    B) The growth rate of Strain 2 was less than the
        growth rate of Strain 1 until hour 50, at which
        point Strain 1's growth rate became the lesser
        one.

    C) Between hours 50 and 70, Strain 2's OD600
        reading increased by approximately 0.03 every
        hour.

    D) The growth rate of Strain 1 was greater than
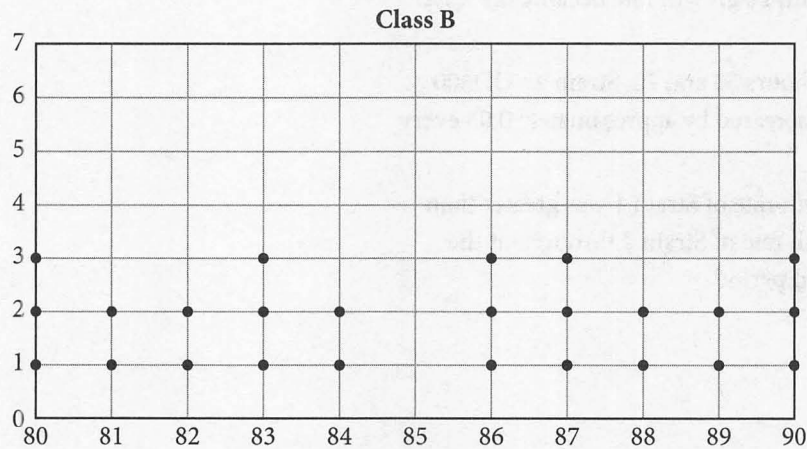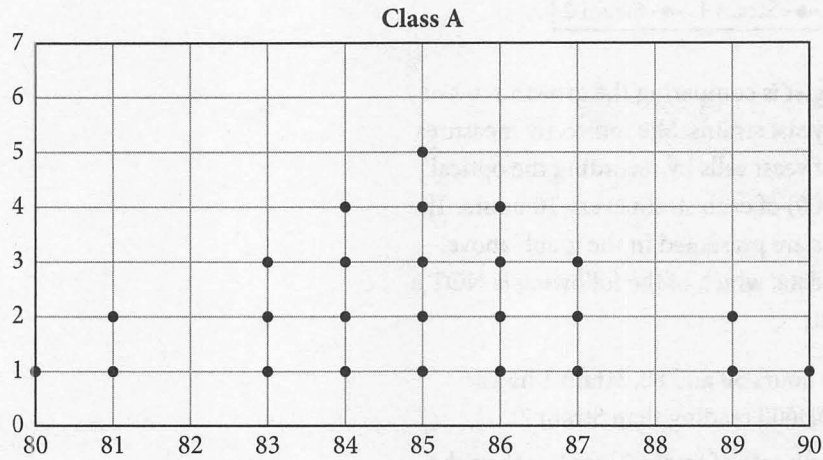   the growth rate of Strain 2 throughout the
        monitored period.

Statistics

LEARNING OBJECTIVES

After this lesson, you will be able to:

- Calculate mean, median, mode, and range
- Describe standard deviation and margin of error

To answer a question like this:

Class A

Class B

Two classes of 25 students each took an identical exam. Their percent correct scores are shown in the dot plots above. If M_A and S_A are the median and standard deviation, respectively, of class A, and M_B and S_B are the median and standard deviation, respectively, of class B, then which of the following statements is true?

A) $M_A < M_B$ and $S_A < S_B$

B) $M_A > M_B$ and $S_A < S_B$

C) $M_A > M_B$ and $S_A > S_B$

D) $M_A < M_B$ and $S_A > S_B$

You need to know this:

Suppose a nurse took a patient's pulse at different times of day and found it to be 75, 78, 71, 71, and 68. Here are six fundamental statistics figures you can determine for this data set:

- **Mean (also called arithmetic mean or average):** The sum of the values divided by the number of values. For this data set, the mean pulse is $\frac{75+78+71+71+68}{5} = \frac{363}{5} = 72.6$.

- **Median:** The value that is in the middle of the set *when the values are arranged in ascending order*. The pulse values in ascending order are 68, 71, 71, 75, and 78. The middle term is the third term, making the median 71. (If the list consists of an even number of values, the median is the average of the middle two values.)

- **Mode:** The value that occurs most frequently. The value that appears more than any other is 71, which appears twice (while all other numbers appear only once), so it is the mode. If more than one value appears the most often, that's okay; a set of data can have multiple modes. For example, if the nurse took the patient's pulse a sixth time and it was 68, then both 71 and 68 would be modes for this data set.

- **Range:** The difference between the highest and lowest values. In this data set, the lowest and highest values are 68 and 78, respectively, so the range is $78 - 68 = 10$.

- **Standard deviation:** A measure of how far a typical data point is from the mean. A low standard deviation means most values in the set are fairly close to the mean; a high standard deviation means there is much more spread in the data set. On the SAT, *you will need to know what standard deviation is and what it tells you about a set of data, but you won't have to calculate it.*

- **Margin of error:** A description of the maximum expected difference between a true statistics measure (for example, the mean or median) for a data pool and that same statistics measure for a random sample from the data pool. A lower margin of error is achieved by increasing the size of the random sample. As with standard deviation, *you will need to know what a margin of error is on the SAT, but you won't be asked to calculate one.*

You need to do this:

- To compare two standard deviations, look at how spread out the data set is. The more clustered the data, the lower the standard deviation.

- To find the median, arrange *all* values in order. In a dot plot or frequency distribution table, that means finding the group with the middle value.

Explanation:

Start with the standard deviation. The scores in class A are more clustered around the mean, so the standard deviation for class A will be smaller than that for class B, where the scores are more spread out. Eliminate (B) and (C).

To calculate the medians of the two classes, you need to find the middle value in each data set. Each class has 25 students, so the middle score will be the 13th term. Count from the left of each dot plot to find that the 13th score for class A is 85 and for class B is 86. So the median for class B is greater. **(A)** is correct.

Try on Your Own

Directions: Take as much time as you need on these questions. Work carefully and methodically. There will be an opportunity for timed practice at the end of the chapter.

Questions 8 and 9 refer to the following information.

An anthropologist chose 250 citizens at random from each of two European countries and separated them into groups based on how many languages they spoke. The results are shown in the table below.

Number of Languages	Country A	Country B
1	55	70
2	80	30
3	50	20
4	40	70
5	25	60

There are a total of 550,000 citizens in country A and 1.3 million citizens in country B.

8. What is the median number of languages spoken by the sample of citizens from country B?

   A) 1
   B) 2
   C) 3
   D) 4

9. Based on the data in the table, which of the following statements most accurately reflects the estimated total number of citizens who speak multiple languages in both countries?

   A) Half as many citizens in both countries combined speak exactly five languages as citizens who speak exactly four languages.

   B) Twice as many citizens in both countries combined speak exactly five languages as citizens who speak exactly four languages.

   C) The number of citizens in both countries combined who speak exactly two languages is less than the number of citizens who speak exactly four languages.

   D) The number of citizens in both countries combined who speak exactly two languages is greater than the number of citizens who speak exactly four languages.

	Charles	Gautam	Brin
Run 1	8.3	8.5	8.4
Run 2	7.7	8.0	8.0
Run 3	7.1	8.5	7.5
Run 4	6.6	7.8	9.0
Run 5	8.0	8.1	7.5
Run 6	6.6	7.5	7.2
Mean Score	7.38	8.07	7.93
Standard Deviation	0.73	0.39	0.67

10. Charles, Gautam, and Brin participated in a snowboarding competition. The scores for each of their six qualifying runs are shown in the table above. According to the data, which of the following is a valid conclusion?

   A) Charles had the smallest mean score, so his performance was the least consistent.

   B) Gautam had the smallest standard deviation, so his performance was the most consistent.

   C) Charles had the largest standard deviation, so his performance was the most consistent.

   D) Brin had the highest score on any one run, so her performance was the most consistent.

Ages of Used Cars in Dealer Inventory

Age (Model Years)	Number of Cars
1	3
2	5
3	18
4	17
5	11
6	6
7	2

11. The table above shows the distribution of the ages (in model years) of the cars in a certain dealer's inventory. Which of the following correctly lists the mean, median, and mode of the ages of the cars in ascending order?

   A) Mean, Median, Mode

   B) Median, Mode, Mean

   C) Mode, Mean, Median

   D) Mode, Median, Mean

Math

HINT: For Q12, start with the most definite information. What do you know, given the average of the first 12 days?

12. A company produces an equal number of a certain product each working day in the month. The daily reject rates for the first 12 days of this month ranged from 0.0 percent to 1.5 percent with an average of 0.4 percent. If there are 22 working days this month and the company's monthly reject goal is 0.5 percent or less, what is the maximum reject rate, as a percent, for the 13th day that would still enable the company to attain its goal for the month? (Round your answer to the nearest tenth and ignore the percent sign when gridding your response.)

Number of Persons in 65 Households

Number of Persons in Household
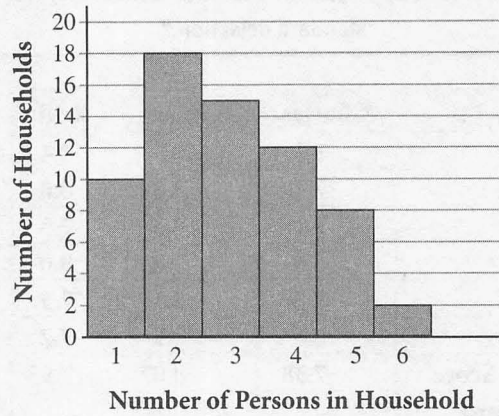
13. Based on the graph above, how many households have a number of persons greater than the median number of persons?

A) 2

B) 10

C) 22

D) 37

Surveys and Data Samples

LEARNING OBJECTIVES

After this lesson, you will be able to:

● Determine whether a survey is valid or biased

● Draw inferences about surveys and data samples

To answer a question like this:

A book club wanted to determine the average number of books read each year by residents of a certain town, so it conducted a survey of 100 patrons of the town's public library. The average number of books read per year by these 100 patrons was 51.5. Which of the following statements must be true based on this information?

A) The survey is biased due to a poor choice of sampling method.

B) The survey is not biased and will likely produce a correct estimate of the number of books read annually by the town's residents.

C) The average number of books read annually by all the town's residents is 51.5.

D) The average number of books read per town resident per year cannot be determined from such a small sample.

You need to know this:

You will see occasional questions on the SAT Math sections that do not test any calculation or even your ability to interpret numerical data. Instead, these questions test your ability to draw logical conclusions about surveys and data sampling methods.

Answering these questions correctly hinges on your ability to tell whether a data sample is **representative** of the larger population. A representative sample is a small group that shares key characteristics with a larger group you are trying to draw conclusions about.

A sample that is selected truly at random is generally representative of the larger group. For example, a scientist who wants to learn the average height of the penguins in a colony of 200 might measure the heights of a random sample of only 20 penguins. As long as the 20 penguins are selected at random, their average height will approximate the average height of the birds in the entire colony.

On the other hand, a sample that is not selected at random may not be representative and may lead to a biased conclusion. For instance, imagine that a small town uses volunteer firefighters and that a stipulation for becoming a volunteer firefighter is living within a mile of the fire station. If you wanted to know what percent of households in the town include at least one volunteer firefighter, you would need to survey a random sample of households from the entire town, not just a sample of households within a mile of the fire station. A sample of households within a mile of the fire station would be a biased sample and would lead to an erroneous conclusion (namely, that the percent of households in the town that include at least one volunteer firefighter is higher than it actually is).

Math

You need to do this:

- Check whether the data sample represents the larger population. If it doesn't, the survey is biased.
- In questions that ask you to draw a conclusion from a random (unbiased) sample, look for the answer choice for which the representative sample accurately reflects the larger population. For example, in a question asking for a conclusion based on a sample of librarians, the correct answer will match the sample to a larger population of librarians, not to a population of, say, accountants.

Explanation:

The sample in this question includes 100 public library patrons. This is not a randomly selected sample. It's a good bet that frequent readers of books will be overrepresented at a public library. Thus, the survey is biased, so **(A)** is correct.

Try on Your Own

Directions: Take as much time as you need on these questions. Work carefully and methodically. There will be an opportunity for timed practice at the end of the chapter.

HINT: For Q14, who is in the survey group? Who is in the larger population? Are these groups different? If so, the survey is likely biased.

---

14. A railroad company is planning to build a new station along one of its busiest lines into the downtown area where many commuters work. The company chooses a town where it plans to build the new station. To assess the opinion of the town's residents, the company surveys a sample of 200 residents who commute to the downtown area for work. Over 80 percent of those surveyed are in favor of building the new station.

    Which of the following is true about the survey's reliability?

    A) It is unreliable because the survey sample is not representative of the entire town.

    B) It is unreliable because the survey sample is too small.

    C) It is reliable because nobody in the survey sample works for the railroad company.

    D) It is reliable because the survey sample excludes people who do not ride the train.

15. A bottled water company conducts a survey to find out how many bottles of water people consume per day. If a representative and random sample of 500 people is chosen from a population estimated to be 50,000, which of the following accurately describes how the mean of the sample data relates to the estimated mean of the entire population?

    A) The mean of the sample data is equal to the estimated mean of the population.

    B) The mean of the sample data cannot be used to estimate the mean of the much larger population.

    C) The mean of the sample data should be multiplied by 100 to get the estimated mean of the population.

    D) The mean of the sample data should be multiplied by 1,000 to get the estimated mean of the population.

16. A department store manager wants to determine why customers return the products they buy. The manager surveyed randomly selected customers and asked them to explain why they were returning their products. This sample included 70 customers who were returning dinnerware, of whom 80 percent indicated that at least one piece of dinnerware was chipped or broken.

    Which of the following conclusions is best supported by the sample data?

    A) Most of the products returned to the store contain chipped or broken pieces.

    B) Dinnerware products are more likely to contain chipped or broken pieces than other products.

    C) Most customers returning dinnerware returned products containing chipped or broken pieces.

    D) At least 80 percent of the products sold at the store contain chipped or broken pieces.

17. The owner of a miniature golf course wants to determine what color golf ball is most popular at the course. The owner asked 150 randomly surveyed children what color they prefer. Approximately 60 percent of them said they prefer red, while approximately 30 percent of them said blue.

    This data best supports which of the following conclusions?

    A) Most people prefer a red golf ball when playing miniature golf.

    B) Red golf balls are used twice as often for miniature golf as blue golf balls.

    C) Most children at the miniature golf course prefer a red golf ball.

    D) Approximately 10 percent of miniature golf players prefer a white golf ball.

HINT: For Q18, find the result of the sample in the chart, and then apply that result to the larger group.

A candy company sells jelly beans in five colors: black, green, orange, red, and yellow. The company sells boxes of jelly beans, each of which contains 20 individual bags. Each individual bag contains 75 jelly beans. A customer purchased 5 boxes of jelly beans and selected one bag at random from each box. The customer counted the number of each color in each bag. The results are shown in the chart below.

Color	Bag 1	Bag 2	Bag 3	Bag 4	Bag 5
Black	10	12	8	11	9
Green	13	11	13	12	12
Orange	22	20	21	21	21
Red	20	21	23	21	22
Yellow	10	11	10	10	11

18. Which of the following is the closest approximation of the total number of green jelly beans in the customer's purchase?

    A) 60

    B) 240

    C) 1,200

    D) 4,500

Probability

To answer a question like this:

Number of Cyclists in Regional Race, by Age and Town

Town	Age (years)					Total
	15 to 18	19 to 25	26 to 34	35 to 46	47 and Older	
Pine Falls	9	52	31	26	29	147
Greenville	14	38	42	53	30	177
Salem	5	17	18	13	10	63
Fairview	19	41	32	34	27	153
Total	47	148	123	126	96	540

The table above shows the number of participants in a regional bicycle race, categorized by town and age group. Based on the table, if a cyclist from Fairview is chosen at random, which of the following is closest to the probability that the cyclist was 35 or older at the time of the race?

A) 0.40

B) 0.18

C) 0.11

D) 0.05

You need to know this:

Probability is a fraction or decimal between 0 and 1 comparing the number of desired outcomes to the number of total possible outcomes. A probability of 0 means that an event will not occur; a probability of 1 means that it definitely will occur. The formula is as follows:

$$\text{Probability} = \frac{\text{number of desired outcomes}}{\text{number of total possible outcomes}}$$

For instance, if you roll a six-sided die, each side showing a different number from 1 to 6, the probability of rolling a number higher than 4 is $\frac{2}{6} = \frac{1}{3}$, because there are two numbers higher than 4 (5 and 6) and six numbers total (1, 2, 3, 4, 5, and 6).

To find the probability that an event will *not* happen, subtract the probability that the event will happen from 1. Continuing the previous example, the probability of *not* rolling a number higher than 4 would be:

$$1 - \frac{1}{3} = \frac{2}{3}$$

The SAT tends to test probability in the context of data tables. Using a table, you can find the probability that a randomly selected data value (be it a person, object, etc.) will fit a certain profile. For example, the following table summarizing a survey on water preference might be followed by a question asking for the probability that a person randomly selected for a follow-up survey falls into a given category.

	Tap	Carbonated	Bottled	Total
Female	325	267	295	887
Male	304	210	289	803
Total	629	477	584	1,690

If the question asked for the probability of randomly selecting a female who prefers tap water from all the participants of the original survey, you would calculate it using the same general formula as before:

$$\frac{\text{\# female, tap}}{\text{\# total}} = \frac{325}{1,690} = \frac{5}{26} \approx 0.192.$$

If the question asked for the probability of randomly selecting a female for the follow-up survey, given that the chosen participant prefers tap water, the setup is a little different. This time, the number of possible outcomes is the total participants *who prefer tap water*, which is 629, not the grand total of 1,690. The calculation is now $\frac{\text{\# female, tap}}{\text{\# total, tap}} = \frac{325}{629} \approx 0.517.$

Conversely, if you needed to find the probability of selecting someone who prefers tap water for the follow-up survey, given that the chosen participant is female, the new number of possible outcomes would be the female participant total (887). The calculation becomes $\frac{\text{\# female, tap}}{\text{\# total, females}} = \frac{325}{887} \approx 0.366.$

You need to do this:

- Determine the number of desired and total possible outcomes by looking at the table.
- Read the question carefully when determining the number of possible outcomes: do you need the entire set or a subset?

Explanation:

The number of desired outcomes is the number of cyclists from Fairview who are 35 or older. That means you need to add the "35 to 46" and "47 and Older" categories: $34 + 27 = 61$. The number of possible outcomes is the total number of cyclists from Fairview. The number is given in the totals column: 153. Plug these numbers into the probability formula and divide:

$$\text{Probability} = \frac{\text{\# Fairview, 35 and Older}}{\text{\# Fairview, Total}} = \frac{61}{153} \approx 0.40$$

The correct answer is **(A)**.

Try on Your Own

Directions: Take as much time as you need on these questions. Work carefully and methodically. There will be an opportunity for timed practice at the end of the chapter.

	Marked Defective	Not Marked Defective	Total
Defective Bearing	392	57	449
Non-defective Bearing	168	49,383	49,551
Total	560	49,440	50,000

19. A manufacturing plant produces 50,000 bearings per week. Of these, 449 will be defective. The manager of the plant installs a new quality control device that is designed to detect defective bearings and mark them with a laser. The device is allowed to run for a week and the results are tallied as shown in the table above. According to these results, to the nearest percent, what is the probability that a part that is marked defective will actually be defective?

A) 30%

B) 43%

C) 70%

D) 87%

The table below shows the distribution of four species of fish at a hatchery that has approximately 6,000 fish.

Species	Percent of Total
Carp	50
Salmon	25
Tilapia	15
Tuna	10

A biologist randomly tests 5 percent of each species of fish for mercury content. Her findings are shown in the following table.

Mercury Content Test Results

Species	Number of Fish with Dangerous Mercury Levels
Carp	11
Salmon	6
Tilapia	5
Tuna	8

20. Based on the biologist's findings, if a single salmon is randomly selected from those that were tested, what is the probability that this particular fish would have a dangerous mercury level?

A) 0.001

B) 0.004

C) 0.02

D) 0.08

Math

Type of Engineer	Specialization		Total
	Robotics	AV	
Mechanical	198	245	443
Electrical	149	176	325
Total	347	421	768

21. In a research study, a group of mechanical and electrical engineers indicated their specialization preference between robotics and autonomous vehicles (AV). The results are shown in the table above. What is the probability that a randomly selected engineer will be a mechanical engineer specializing in autonomous vehicles?

A) 0.229

B) 0.319

C) 0.553

D) 0.582

HINT: For Q22, how many groups have *at least* 8 days vacation?

Yearly Paid Vacation Days at Excor Manufacturing

	0–7	8–14	14–30	Total
Hourly	79	183	38	300
Salaried	8	27	65	100
Total	87	210	103	400

22. The human resources department at Excor Manufacturing decided to collect data on the paid vacation days accrued by hourly and salaried employees. The table above shows the results of the data collection. If an employee has at least 8 paid vacation days, what is the probability that the person is a salaried employee?

A) $\frac{92}{313}$

B) $\frac{221}{300}$

C) $\frac{313}{400}$

D) $\frac{92}{100}$

Engine Type	Fuel Economy (miles per gallon)	
	0–45 mpg	45+ mpg
Hybrid		
Internal Combustion (IC)		
Total	53	258

23. The daily engine production goals of an automobile manufacturer are summarized in the incomplete table above. The factory produced six times as many hybrid engines that achieve 45+ miles per gallon as it did hybrid engines that achieve 0–45 miles per gallon, and the factory produced four times as many internal combustion (IC) engines that achieve 45+ miles per gallon as it did IC engines that achieve 0–45 miles per gallon. If the factory produced 53 engines that achieve 0–45 miles per gallon and 258 engines that achieve 45+ miles per gallon, which of the following is the approximate probability that a 45+ miles per gallon engine selected at random is IC?

A) 0.566

B) 0.535

C) 0.465

D) 0.386

On Test Day

The SAT tests the concept of average fairly heavily. The average formula will serve you well on questions that ask about a sum of values or the average of a set of values, but for questions that give you the average and ask for a missing value in the data set, there is an alternative that can be faster: the balance approach.

The balance approach is based on the the idea that if you know what the average is, you can find the totals on both sides of the average and then add the missing value that makes both sides balance out. This approach is especially helpful if the values are large and closely spaced. Imagine that a question gives you the set {976, 980, 964, 987, x} and tells you that the average is 970. You would reason as follows: 976 is 6 over the average, 980 is 10 over, 964 is 6 under, and 987 is 17 over. That's a total of $6 + 10 - 6 + 17 = 27$ over, so x needs to be 27 under the average, or $970 - 27 = 943$.

Try solving the question below both ways, using first the average formula and then the balance approach. If you find the latter to be fast and intuitive, add it to your test-day arsenal.

Height Change in Inches							
Plant Type	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9
Zinnia	3	2	1	4	2	1	1
Sunflower	3		8	6	7	2	5
Marigold	1	1	3	2	4	4	3

24. The table above summarizes the height change (inches) over a seven-week period of three different plants grown by Ms. Walker's biology class. If the mean height change for the sunflower plant over the seven-week period is 5 inches, what was the height change in week 4?

The correct answer and both ways of solving can be found at the end of the chapter.

How Much Have You Learned?

Directions: For testlike practice, give yourself 15 minutes to complete this question set. Be sure to study the explanations, even for questions you got correct. They can be found at the end of this chapter.

**Question 25 refers to the following information.**

The table below summarizes the results of a survey about favorite leisure activities for a group of high school students. Assume that every student has a favorite leisure activity and that each student could select only one favorite.

	Freshmen	Sophomores	Juniors	Seniors	Total
Sports	144	122	134	115	515
Video Games	126	140	152	148	566
Music	120	117	153	148	538
Reading	110	114	63	98	385
Total	500	493	502	509	2,004

25. The research group that conducted the survey wants to select one sophomore at random for a follow-up survey. What is the probability that the student selected will prefer a type of leisure activity other than video games?

A) $\dfrac{140}{493}$

B) $\dfrac{140}{2,004}$

C) $\dfrac{353}{493}$

D) $\dfrac{353}{2,004}$

Questions 26 and 27 refer to the following information.

Attendance at Museum X: Age Distribution
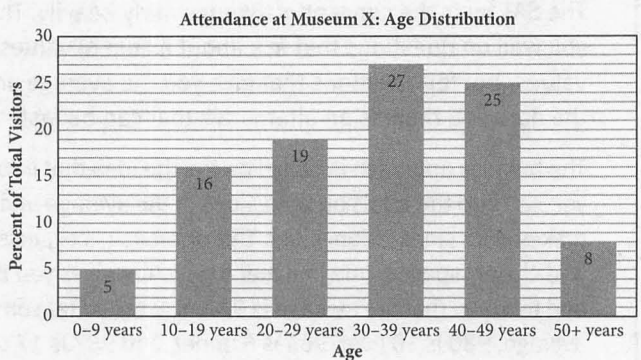
The bar graph above shows the age distribution of visitors to museum X in 2014. Visitors aged 0–9 years get into museum X for free, visitors aged 50 and older pay $5 for admission, and everyone else pays $10.

26. If 553 people aged 20 years and older visited museum X in 2014, then approximately how many people visited museum X in 2014?

27. Assuming, as before, that 553 people aged 20 years and older visited museum X in 2014, how much revenue did museum X collect from tickets sold to people aged 40 and older in 2014? (Ignore the dollar sign when gridding your response.)

28. According to the table, what percent of all the board games sold by the boutique have a "bad" average customer rating? Round to the nearest tenth of a percent and ignore the percent sign when entering your answer.

Questions 28 and 29 refer to the following information.

	1	2	3	4	5	Total
Strategy	5	17	24	10	5	61
Trivia	3	12	28	8	3	54
Role-playing	3	10	30	14	2	59

A small boutique sells board games online. The boutique specializes in strategy, trivia, and role-playing games. Any customer who purchases a game is invited to rate the game on a scale of 1 to 5. A rating of 1 or 2 is considered "bad," a rating of 3 is considered "average," and a rating of 4 or 5 is considered "good." The table above shows the distribution of average customer ratings of the games sold. For example, 24 of the strategy games sold have an average customer rating of 3.

29. The boutique decides to stop selling 50 percent of the games that have a "bad" average customer rating to make room for promising new stock. Assuming no significant changes in ratings in the foreseeable future, what should the difference be between the percentages of games with a "bad" average customer rating before and after the games are removed? Round to the nearest tenth of a percent and ignore the percent sign when entering your answer.

Math

30. Fit and Fab, a membership-only gym, is hoping to open a new branch in a small city in Pennsylvania that currently has no fitness centers. According to the gym management's research, approximately 12,600 residents live within driving distance of the gym. Fit and Fab sends out surveys to a sample of 300 randomly selected residents in this area (all of whom respond) and finds that 40 residents say they would visit a gym if one were located in their area. Based on past survey research, Fit and Fab estimates that approximately 30 percent of these respondents would actually join the gym if one were opened in the area. Based on this information and the results of the sample survey, about how many residents should Fit and Fab expect to join its new branch?

A) 134

B) 504

C) 1,680

D) 3,780
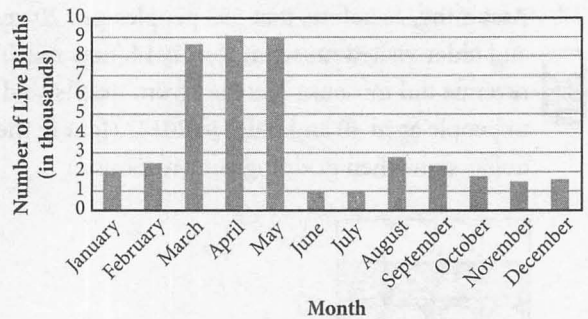
31. Most animals give birth during a general time of year. This is because animals naturally breed so that their young will be born at the time of year when there will be an adequate food supply. The bar graph shows the number of live births in California of a particular jackrabbit species, the black-tailed jackrabbit, over the course of year X. Based on the data, which of the following would be an appropriate conclusion?

A) In general, rabbits give birth during March, April, and May.

B) In general, rabbits give birth during June, July, and August.

C) In general, black-tailed jackrabbits in California give birth during March, April, and May.

D) In general, black-tailed jackrabbits in California give birth during June, July, and August.

32. Soil contains a wide variety of nutrients, including nitrogen, phosphorous, potassium, magnesium, sulfur, and iron. A fertilizer company conducted an experimental study to determine which of five additives is most effective in helping soil retain nutrients. If, after application of the additives, the fertilizer company tested only for the soil nutrients nitrogen and potassium, which of the following is a valid conclusion?

   A) The additive that is found to be the most effective will work for all nutrients in the soil.

   B) The additive that is found to be the most effective will work only for nitrogen and potassium.

   C) The study is biased and therefore not significantly relevant to determining which additive is most effective.

   D) The study will be able to produce results concerning only the effects of the additives on nitrogen and potassium.

33. The average (arithmetic mean) of $p + 12t - 5$, $q + 16t + 4$, and $r - 7t + 25$ is $10t + 34$. In terms of t, what is the average of p, q, and r?

   A) $3t + 26$
   B) $4t + 16$
   C) $6t + 16$
   D) $8t - 37$

34. As part of its market research, a company sent out a survey to see how much consumers would be willing to pay for a certain product. The survey distinguished between a store brand version of the product and a brand name version, and people participating in the survey received questions about only one of the versions. A summary of the survey results is shown in the following bar graph.

Consumer Price Points
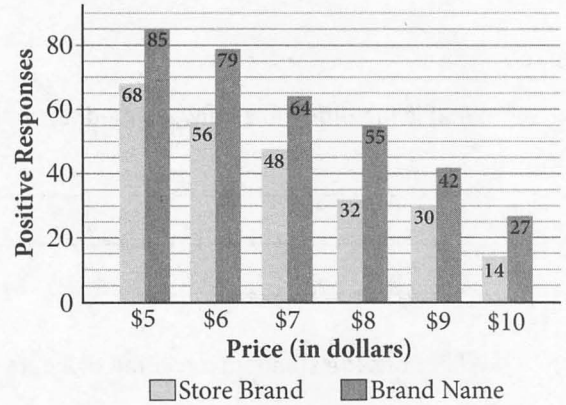
If a consumer is chosen at random from the 600 respondents, what is the probability that the consumer is willing to pay at least $8 for the product?

Reflect

Directions: Take a few minutes to recall what you've learned and what you've been practicing in this chapter. Consider the following questions, jot down your best answer for each one, and then compare your reflections to the expert responses on the following page. Use your level of confidence to determine what to do next.

What are some common ways the SAT may present data?

What is the difference between median, mode, and range?

What does the standard deviation of a data set tell you?

When can you generalize the results of a survey of a small group to a larger group?

What are two ways to calculate the probability of a single event?

Expert Responses

What are some common ways the SAT may present data?

The SAT may present data in tables, bar charts, line graphs, dot plots, and histograms.

What is the difference between median, mode, and range?

The median of a set is the middle value, whereas the mode is the most common value. The range of a set is the distance between the smallest value and the largest one.

What does the standard deviation of a data set tell you?

A data set's standard deviation reflects how far apart the numbers are from each other. The standard deviation of a set whose numbers are all the same—for example, {5, 5, 5, 5}—is 0. The greater the distance between the numbers, the greater the standard deviation.

When can you generalize the results of a survey of a small group to a larger group?

A survey can be generalized to a larger population if the data sample is representative. To be representative, the data sample needs to be drawn at random from the larger population.

What are two ways to calculate the probability of a single event?

One way is to use the basic probability formula:

$$\text{Probability} = \frac{\text{number of desired outcomes}}{\text{number of total outcomes}}$$

Alternatively, the probability that an event happens is 1 minus the probability that it doesn't happen.

Next Steps

If you answered most questions correctly in the "How Much Have You Learned?" section, and if your responses to the Reflect questions were similar to those of the SAT expert, then consider tables, statistics, and probability an area of strength and move on to the next chapter. Come back to this topic periodically to prevent yourself from getting rusty.

If you don't yet feel confident, review those parts of this chapter that you have not yet mastered, then try the questions you missed again. In particular, make sure that you understand the six terms explained in the Statistics lesson and the probability formulas explained in the Probability lesson. As always, be sure to review the explanations closely. Finally, **go online** (www.kaptest.com/moreonline) for additional practice on the highest yield topics in this chapter.

Answers and Explanations

1. B

Difficulty: Easy

Getting to the Answer: The trickiest part of this question is understanding what is being asked. You need to find the shop that had the most Tuesday sales *as a fraction of its total sales*, so focus only on those rows in the table. For each shop, divide the number of books it sold on Tuesday by the number of books it sold all week. Use your calculator to speed up this step.

Bob's Bookshop: $\dfrac{\text{Tuesday total}}{\text{weekly total}} = \dfrac{8}{53} \approx 0.1509$

Clara's Bookshop: $\dfrac{\text{Tuesday total}}{\text{weekly total}} = \dfrac{13}{55} \approx 0.2364$

Derek's Bookshop: $\dfrac{\text{Tuesday total}}{\text{weekly total}} = \dfrac{15}{66} \approx 0.2273$

Evelyn's Bookshop: $\dfrac{\text{Tuesday total}}{\text{weekly total}} = \dfrac{13}{58} \approx 0.2241$

The greatest portion of Tuesday sales belongs to Clara's Bookshop, so **(B)** is correct.

2. C

Difficulty: Medium

Getting to the Answer: Add the number of books sold by Derek and Evelyn on Monday, Wednesday, and Friday; then divide the result by the total number of books sold on those days.

Derek's and Evelyn's M/W/F sales:
$(15 + 12) + (12 + 14) + (10 + 9) = 72$

Total M/W/F sales: $48 + 49 + 39 = 136$

Divide Derek's and Evelyn's M/W/F sales by total M/W/F sales to get $\dfrac{72}{136}$, which simplifies to $\dfrac{9}{17}$. Choice **(C)** is correct.

3. B

Difficulty: Easy

Getting to the Answer: The question asks only about participants who were outside a healthy weight range, so focus on this row: 38 out of the 74 participants who were outside a healthy weight range ate breakfast one or fewer times per week. This expressed as a percent is $\dfrac{38}{74} \times 100\% = 0.51351 \times 100\% = 51.35\%$, which matches **(B)**.

4. D

Difficulty: Medium

Getting to the Answer: The question asks about employees who eat breakfast every weekday, so focus on the "5–7 times per week" column in the table. Assuming the participants in the study were a good representative sample, 36 out of 45, or 80%, of the 3,000 employees are likely to be within a healthy weight range. Multiply $0.8 \times 3,000$ to arrive at 2,400, which is **(D)**.

5. 300

Difficulty: Easy

Getting to the Answer: Read the graph carefully, including the key at the bottom indicating that each bar represents 15 minutes. The question states that only stage 3 is considered deep sleep, and the question asks how much time was spent in non-deep sleep. You could count all of the bars that don't represent stage 3, but it would be faster to count the bars that do and then subtract that number from the total. There are 12 bars that represent stage 3, which means the person spent $12 \times 15 = 180$ minutes in deep sleep. The study was for 8 hours, or 480 minutes, so the person spent $480 - 180 =$ **300** minutes in non-deep sleep.

6. 195

Difficulty: Medium

Strategic Advice: In multi-part Math question sets, the second question often uses information that you had to calculate in the first question. Keep track of your computations and reuse information so that you don't waste time repeating calculations.

Getting to the Answer: Set up the average formula and start filling in the values to find the missing night's stage 3 sleep. You're given two of the three quantities needed for the average formula: the total number of nights is 5 and the average over all the nights is 180 minutes. You know from your work in the previous question that the patient spent 180 minutes in stage 3 sleep on the first night. The following 3 nights, as given in the questions stem, average out to 175 minutes each, so even though you don't know their precise values, you can represent them as 175 three times. Use a variable, *x*, for the unknown number of minutes in stage 3 sleep on the fifth night and solve:

$$180 = \frac{180 + 175 + 175 + 175 + x}{5}$$

$$900 = 705 + x$$

$$x = 195$$

Grid in **195**.

7.   D

Difficulty: Medium

Getting to the Answer: Compare each statement to the line graph one at a time, eliminating true statements as you work. Start with (A): at every reading after 20 hours, Strain 1 has a higher OD600 level than Strain 2, so this statement is true. Eliminate (A). Choice (B) states that Strain 2's growth rate (slope) overtook Strain 1's at hour 50, which is consistent with the line graph; eliminate it. It looks as though (C) requires time-consuming calculations, so skip it for now. Choice (D) states that Strain 1's growth rate was greater than Strain 2's over the entire period. This statement contradicts what you already confirmed in (B), which makes **(D)** false and, therefore, correct. There's no need to check (C).

8.   D

Difficulty: Medium

Strategic Advice: The median is the middle value when all of the values are in numerical order, so you'll need to find the total number of values in the set and figure out which one is the middle value.

Getting to the Answer: The note above the chart says that the total number of people who were surveyed in country B is 250. Since it is an even set of values, the median will be the average of the 125th and 126th values. To get to those values, add the number of citizens who speak one or two languages: $70 + 30 = 100$. The hundredth value falls within the group that speaks two languages. Keep going because this group does not include the 125th and 126th values. Add the citizens who speak three languages: $100 + 20 = 120$. Still not quite there, so add the citizens who speak four languages: $120 + 70 = 190$. This means that the 125th and 126th values are both 4, so the median is 4. **(D)** is correct.

9.   C

Difficulty: Hard

Getting to the Answer: The answer choices compare the number of citizens who speak different numbers of languages. (A) and (B) compare the number of

citizens who speak exactly five languages to the number of citizens who speak exactly four languages. Notice that in both countries, the fraction that speaks exactly five languages is smaller than the fraction that speaks exactly four languages, so (B) can be eliminated.

To evaluate (A), you need to calculate the number of citizens who speak five languages and the number of citizens who speak four languages. Take the population in each country and multiply it by the fraction of citizens surveyed who speak the specified number of languages (the denominator of the fraction will be the total number surveyed, 250):

Country A: $(550,000)\left(\frac{25}{250}\right) = (550,000)\left(\frac{1}{10}\right) = 55,000$

Country B: $(1,300,000)\left(\frac{60}{250}\right) = (1,300,000)\left(\frac{6}{25}\right) = 312,000$

This means that $55,000 + 312,000 = 367,000$ citizens speak five languages. Next, calculate how many citizens speak exactly four languages:

Country A: $(550,000)\left(\frac{40}{250}\right) = (550,000)\left(\frac{4}{25}\right) = 88,000$

Country B: $(1,300,000)\left(\frac{70}{250}\right) = (1,300,000)\left(\frac{7}{25}\right) = 364,000$

This means that $88,000 + 364,000$, or $452,000$, citizens speak exactly four languages. Because $312,000$ is not half of $364,000$, eliminate (A).

Use logic to decide between (C) and (D). Notice that the fraction of citizens who speak four languages in country B, which has a larger population, is more than twice the fraction of citizens in country B who speak two languages—while in country A, which has a smaller population, the fraction of citizens who speak two languages is exactly twice the fraction of citizens who speak four languages. Country B's larger population means that overall, more citizens of both countries combined will speak four languages than two languages. Choice **(C)** is correct.

For the record, here are the calculations for **(C)** and (D):

Country A: $(550,000)\left(\frac{80}{250}\right) = (550,000)\left(\frac{8}{25}\right) = 176,000$

Country B: $(1,300,000)\left(\frac{30}{250}\right) = (1,300,000)\left(\frac{3}{25}\right) = 156,000$

The number of citizens who speak two languages in both countries is $176,000 + 156,000$, or $332,000$, which is fewer than the $452,000$ who speak four languages.

10. B

Difficulty: Easy

Getting to the Answer: Consider the definitions of mean and standard deviation: mean is a measure of center, while standard deviation is a measure of spread. The closer the data points for a given snowboarder are to the mean, the more consistent that snowboarder's performance, so the explanation should involve standard deviation. Based on this, you can eliminate (A) and (D). Greater consistency means lower standard deviation (and vice versa); the only choice that reflects this—and correctly represents the data in the table—is **(B)**.

11. C

Difficulty: Medium

Getting to the Answer: You'll have to determine the values of all three measurements so that you can place them in ascending order. The *mode* is 3 because there are 18 cars of that age, which is the most of any age. The total number of cars is $3 + 5 + 18 + 17 + 11 + 6 + 2 = 62$. Since this is an even number, the *median* age will be the average of the 31st and 32nd values. The are $3 + 5 + 18 = 26$ cars that are 1, 2, and 3 years old and 17 that are 4 years old. Thus, the 27th through 43rd ($26 + 17 = 43$) values are 4, and that is the median.

To find the *mean*, multiply each value by its frequency, total those values, and divide by 62. So, $1 \times 3 = 3$, $2 \times 5 = 10$, $3 \times 18 = 54$, $4 \times 17 = 68$, $5 \times 11 = 55$, $6 \times 6 = 36$, $7 \times 2 = 14$, and $3 + 10 + 54 + 68 + 55 + 36 + 14 = 240$. Divide 240 by 62 to get approximately 3.87. The ascending order of the three values is mode (3), mean (3.87), and median (4), so **(C)** is correct.

12. 6.2

Difficulty: Hard

Getting to the Answer: Read this question carefully, since it is rather unusual. In order to meet a goal of a 0.5% reject rate for the month, use the average formula,

$$0.5\% = \frac{\text{maximum allowable sum of the daily reject rates}}{22}.$$

Thus, the sum of the daily reject rates for a 0.5% average is $22 \times 0.5\% = 11.0\%$. Since 12 days have already passed with an average reject rate of 0.4%, the sum of the daily rates so far is $12 \times 0.4\% = 4.8\%$. So, the sum of the daily rates for the next 10 days cannot exceed

$11.0\% - 4.8\% = 6.2\%$ if the monthly average is to be 0.5% or less.

The question doesn't ask for the total or average of the next 10 days, however. Instead, it asks for the maximum reject rate on the next single day that could still conceivably allow the company to meet its monthly goal. If the other 9 remaining days all had a 0.0% reject rate, then the 10-day total would be the reject rate for that 13th day. This is 6.2%, so grid in **6.2**.

13. C

Difficulty: Medium

Getting to the Answer: There are 65 data points, so the median will be the middle data point, or the 33rd data point once the data are listed in order. Count the number of data points from the end. There are 10 households with 1 person and 18 households with 2 persons, so that adds up to 28 households. Since there are 15 households with 3 persons, the 33rd household will fall in that group, so the median household size is 3 persons. Add the number of households with more than 3 persons: $12 + 8 + 2 = 22$. **(C)** is correct.

14. A

Difficulty: Easy

Getting to the Answer: Any sample used to determine a general opinion needs to be representative and unbiased. The railroad company fails to meet that requirement, surveying only people who commute to work and who would probably benefit from the station. This potentially leaves out a large portion of the population who may not share the commuters' favorable opinion. The use of a biased sample group makes the survey unreliable and not representative, which makes **(A)** the correct answer.

15. A

Difficulty: Medium

Getting to the Answer: As long as a sample is representative, without bias, and relatively large, the mean and median of the sample data will be the same as the expected mean and median of the population from which the sample was taken. The sample here meets all of these requirements. So the mean of the sample equals the estimated mean of the general population, and **(A)** is the correct answer.

16. C

Difficulty: Medium

Getting to the Answer: As the customers were selected at random, it is reasonable to assume that the survey results will be representative of what is true for customers in general. However, the data provided refers only to people who bought dinnerware. So, an inference can be drawn only about dinnerware returns. Based on the 80% of surveyed customers who returned dinnerware items because of damage, it is reasonable to infer that this statistic will be similar for all customers who return dinnerware. That makes **(C)** the correct answer. Choices (A), (B), and (D) are incorrect because they are not confined to dinnerware.

17. C

Difficulty: Medium

Getting to the Answer: To make a reliable inference from a survey, the survey sample needs to be representative, unbiased, and relatively large. In this case, the miniature golf course owner surveyed only children who played at that course. Thus, any inference drawn from the data must be about such children. Since 60% of the surveyed children prefer a red golf ball, it is reasonable to infer that a similar percentage of total children at that golf course would prefer red golf balls. Thus, **(C)** is the correct answer.

18. C

Difficulty: Hard

Getting to the Answer: The customer purchased 5 boxes, each of which contains 20 bags, which means the customer bought a total of $5 \times 20 = 100$ bags. In each bag tested, there were between 11 and 13 green jelly beans. As the bags were chosen at random, it's reasonable to expect that the results will be consistent for all 100 bags. With 11–13 beans per bag, the total number of green jelly beans will likely be between $11 \times 100 = 1,100$ and $13 \times 100 = 1,300$. Right in the middle of that range is 1,200, making **(C)** the correct answer.

Note that if you chose (A) 60, you might have been thinking of the total green jelly beans in just the 5 randomly selected bags. Similarly, if you chose (B) 240, you calculated the approximate number in just one box of 20 bags. Incorrect answer choices often try to anticipate minor mistakes you might make in your calculations. Be sure to confirm that you answered the question being asked.

19. C

Difficulty: Medium

Getting to the Answer: The number of desired outcomes is 392 (marked bearings that are defective). The number of total possible outcomes is 560 (all the bearings that are marked defective). Thus, the probability that a bearing marked defective is in fact defective is $\frac{392}{560} \times 100\% = 0.70 \times 100\% = 70\%$. **(C)** is correct.

20. D

Difficulty: Hard

Getting to the Answer: The probability that one randomly selected salmon from those that were tested would have a dangerous level of mercury is equal to the number of salmon that had dangerous mercury levels divided by the total number of salmon that were tested. This means you need only two numbers to answer this question. One of those numbers is in the second table—6 salmon had dangerous mercury levels. Finding the other number is the tricky part. Use information from the question stem and the first table. The biologist tested 5% of the total number of each species of fish, and 25% of the 6,000 fish are salmon. So, the biologist tested 5% of 25% of 6,000 fish. Multiply to find that $0.05 \times 0.25 \times 6,000 = 75$ salmon were tested. This means the probability is $\frac{6}{75} = 0.08$, which matches **(D)**.

21. B

Difficulty: Medium

Strategic Advice: Recognizing which value goes in the denominator, whether it is the entire total or the total of a subgroup, is essential for probability questions that are based on data in a table.

Getting to the Answer: The question indicates that the random selection is from all the engineers, or the entire total of 768. The specific engineer to be selected is a mechanical engineer who specializes in autonomous vehicles, and the table indicates that there are 245 such engineers. Therefore, the probability of selecting a mechanical engineer specializing in autonomous vehicles from all the engineers is $\frac{245}{768}$, which is ≈ 0.319. **(B)** is correct.

Note that the incorrect answer choices often reflect common misunderstandings and simple table-reading errors. For example, (C) and (D) both use the wrong total and (A) is the probability of choosing an electrical engineer specializing in autonomous vehicles.

22. A

Difficulty: Medium

Strategic Advice: Be on the lookout for "at least" language. It will usually require adding data from multiple rows or columns.

Getting to the Answer: The probability of choosing an employee with "at least" 8 paid vacation days who is salaried is the number of salaried employees with 8 or more paid vacation days divided by the total number of employees with 8 or more paid vacation days. Find the number of salaried employees with 8 or more paid vacation days by adding the salaried employees with 8–14 paid vacation days, 27, and the salaried employees with 15–30 paid vacation days, 65. That means there are $27 + 65$, or 92, salaried employees with 8 or more paid vacation days. The total number of employees with 8 or more paid vacation days is 210 (the total number of employees with 8–14 paid vacation days) plus 103 (the total number of employees with 15–30 paid vacation days), or 313. The probability is $\frac{92}{313}$. **(A)** is correct.

23. C

Difficulty: Hard

Strategic Advice: If totals and the relationships between the data are the only information provided, write out a system of equations.

Getting to the Answer: The question gives the relationships between unknown values in the table, so fill them in accordingly. For hybrid engines, let h be the number that are rated 0–45 miles per gallon. There are 6 times as many hybrid engines that achieve 45+ miles per gallon, so that is $6h$. Similarly, for IC, let c be the number of IC engines that are 0–45 miles per gallon. Because the factory produced four times as many IC engines that achieve 45+ miles per gallon, fill in that blank with $4c$:

	Fuel Economy (miles per gallon)	
Engine Type	**0–45 mpg**	**45+ mpg**
Hybrid	h	$6h$
Internal Combustion (IC)	c	$4c$
Total	53	258

Write a system of equations based on the two miles per gallon columns:

$$h + c = 53$$
$$6h + 4c = 258$$

Solve the first equation for h and substitute the result into the second equation to solve for c:

$$h = 53 - c$$
$$6(53 - c) + 4c = 258$$
$$318 - 6c + 4c = 258$$
$$-2c = -60$$
$$c = 30$$

If c is 30, then there are 4×30, or 120, IC engines that achieve 45+ miles per gallon. The probability of choosing one of those 120 engines out all 258 engines that achieve 45+ miles per gallon is $\frac{120}{258}$, or 0.465. **(C)** is correct. Notice that there is no need to actually calculate h to answer this question.

24. 4

Difficulty: Medium

Category: Statistics

Strategic Advice: When the goal is to find a missing value in a set of data and the average is given, consider using the balance approach. We'll demonstrate both approaches starting with the average formula.

Getting to the Answer: The question is about the height change for the sunflower plant, so ignore the data for the other plants. The given height changes for the other six weeks are: 3, 8, 6, 7, 2, and 5. The average change in height for the sunflower plant is given as 5. If you call the missing value x, plugging the known values into the average formula results in the following:

$$\frac{3 + x + 8 + 6 + 7 + 2 + 5}{7} = 5$$

$$\frac{3 + x + 8 + 6 + 7 + 2 + 5}{7} = 5$$

$$\frac{31 + x}{7} = 5$$

$$31 + x = 35$$

$$x = 4$$

Grid in **4** as the correct answer.

Alternatively, to use the balance approach, write down how much each value is above or below the average of 5. For example, the Week 3 value of 3 is 2 below the average: $3 - 5 = -2$.

Week 3	Week 5	Week 6	Week 7	Week 8	Week 9
3: −2	8: +3	6: +1	7: +2	2: −3	5: +0

Now observe that, excluding Week 4, the values are $-2 + 3 + 1 + 2 - 3 = 1$. Without Week 4, the total is 1 more than what you'd expect based on the average. So for the values to balance out to the average, the Week 4 value must be 1 less than the average of 5, or $5 - 1 = 4$. Grid in **4**.

25. C

Difficulty: Medium

Category: Probability

Getting to the Answer: The research firm is choosing one student from among the sophomores, so you need the total number of sophomores, 493, in the denominator of the probability formula, not the total number of students. There are $493 - 140 = 353$ sophomores who prefer activities other than video games, so the probability of choosing a sophomore who doesn't prefer video games is $\frac{353}{493}$. Choice **(C)** is correct.

26. 700

Difficulty: Medium

Category: Tables and Graphs

Getting to the Answer: Identify the pieces of the graph you need and then convert from the percent to the total. You know from the graph that 21% (5% + 16%) of the visitors are *not* aged 20 or older. This means that $100\% - 21\% = 79\%$ were 20+. Given that the number of visitors aged 20+ is 553, use the three-part percent formula to calculate the total:

$$\text{total} \times 79\% = 553$$
$$\text{total} = \frac{553}{0.79} = 700$$

Grid in **700**.

27. 2030

Difficulty: Medium

Category: Tables and Graphs

Getting to the Answer: From the previous question, you know that the total number of visitors in 2014 was 700. According to the bar graph, 25% of all the visitors were aged 40–49, and 8% were aged 50+. This means that $0.25 \times 700 = 175$ attendees were aged 40–49, and $0.08 \times 700 = 56$ were aged 50+. The visitors aged 40–49 paid $175 \times \$10 = \$1{,}750$ for their tickets, and the visitors aged 50+ paid $56 \times \$5 = \$280$ for their tickets. The total revenue for the two groups was therefore $\$1{,}750 + \$280 = \$2{,}030$. Grid in **2030**.

28. 28.7

Difficulty: Medium

Category: Tables and Graphs

Getting to the Answer: To be considered "bad," a game must have a rating of 1 or 2. Begin by counting the number of "bad" games. There are $5 + 3 + 3 = 11$ games with a rating of 1 and $17 + 12 + 10 = 39$ games with a rating of 2. That's a total of $11 + 39 = 50$ games. Divide this by the total number of games and multiply by 100%: $\frac{50}{61 + 54 + 59} \times 100\% \approx 28.7\%$. Grid in **28.7**.

29. 12

Difficulty: Medium

Category: Tables and Graphs

Getting to the Answer: You know from the previous question that 50 games are "bad." Reducing this number by 50% is the same as halving it, meaning there will be 25 remaining "bad" games after the removal. Subtract this from the original total game count ($61 + 54 + 59 = 174$) to get the new total, $174 - 25 = 149$. Divide the new "bad" count by this total, and then multiply by 100% as you did before: $\frac{25}{149} \times 100\% \approx 16.7785\%$. Subtracting the new percentage from the old one (rounded to a minimum of four decimal places just to be safe) gives $28.7356\% - 16.7785\% = 11.9571\%$. This rounds to 12.0%. Grid in **12**.

30. B

Difficulty: Medium

Category: Surveys and Data Samples

Getting to the Answer: According to the sample survey, $\frac{40}{300}$ *say* they would join the gym. But the gym estimates that only 30% of these respondents would *actually* join, so multiply 40 by 30% to find that the gym can expect $\frac{12}{300} = 0.04 = 4\%$ of the respondents to join. Multiply this by the total number of residents: $12{,}600 \times 0.04 = 504$ residents, so **(B)** is correct.

31. C

Difficulty: Easy

Category: Tables and Graphs

Getting to the Answer: The question states that the data collected was about black-tailed jackrabbits in California, so any conclusion drawn can be generalized only to that particular species of rabbit in California, not to all rabbits generally. According to the data, the California jackrabbit gives birth mostly during March, April, and May, so **(C)** is correct.

32. D

Difficulty: Medium

Category: Surveys and Data Samples

Getting to the Answer: The question indicates that there are a wide variety of nutrients in the soil, but the study only tests for nitrogen and potassium, so the sample is limited. You can eliminate (A) and (B) because all nutrients were not included in the sample, so you can't say anything about them one way or the other. The additives may or may not help the soil retain other types of nutrients, and you certainly don't know which of the five additives would produce the best results. You can eliminate (C) because the question doesn't tell you anything about the data collection methods, so you can't determine whether the study was biased. This means that the study will only be able to produce results concerning the effects of the additives on soil retention of nitrogen and potassium. Thus, **(D)** is correct.

33. A

Difficulty: Hard

Category: Statistics

Getting to the Answer: The number of variables may look daunting, but the question is just asking you to find an average. The average of a set of terms is equal to the sum of the terms divided by the number of terms; thus, the average of p, q, and r is $\frac{p + q + r}{3}$.

Use the average formula to set up an equation using the rest of the information in the question. Even though you are given expressions with lots of variables, the average formula is still the sum of the terms divided by the number of terms. Notice that the question states that the average of these expressions is $10t + 34$, which translates to "$= 10t + 34$." So you can set up an equation like this:

$$\frac{(p + 12t - 5) + (q + 16t + 4) + (r - 7t + 25)}{3} = 10t + 34$$

Because the question asks for the average of p, q, and r in terms of t, solve the equation above for $\frac{p + q + r}{3}$:

$$\frac{p + 12t - 5 + q + 16t + 4 + r - 7t + 25}{3} = 10t + 34$$

$$\frac{p + q + r + 21t + 24}{3} = 10t + 34$$

$$\frac{p + q + r}{3} + \frac{21t}{3} + \frac{24}{3} = 10t + 34$$

$$\frac{p + q + r}{3} = 10t + 34 - 7t - 8$$

$$\frac{p + q + r}{3} = 3t + 26$$

(A) is correct.

34. 1/3 or .333

Difficulty: Medium

Category: Probability

Getting to the Answer: First, find the number of respondents willing to pay at least $8 (which means $8 or more). Be careful—the question doesn't specify store brand or brand name, so use both versions of the product:

$$32 + 55 + 30 + 42 + 14 + 27 = 200$$

Now, find the total number of people in the survey. Again, the question doesn't specify store brand or brand name. The question stem states that there is a total of 600 respondents. This means the probability that a randomly chosen respondent is willing to pay at least $8 is $\frac{200}{600}$, or $\frac{1}{3}$. Grid in **1/3** or **.333**.