1,001 Practice Problems

Statistics

FOR DUMMIES®

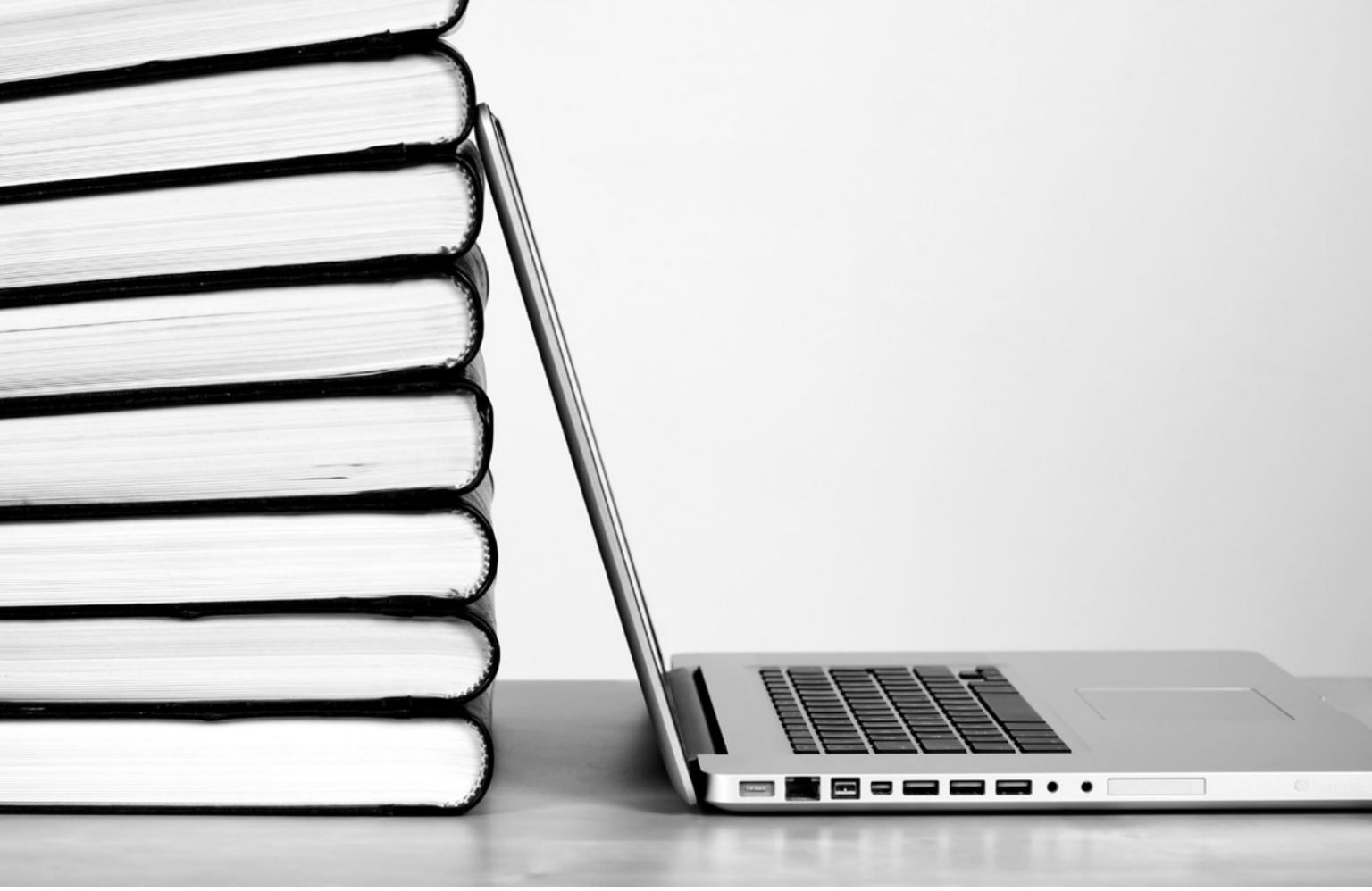
A Wiley Brand

Get:

Free one-year access to practice problems online:

- All 1,001 practice problems online — from easy to hard
- Track your progress, see where you need more help, and create customized problem tests
- Study what, where, and when you want

- 1,001 practice problems covering all of the topics in your statistics course
- Detailed, step-by-step answers and explanations for every problem



This book comes with access to more content online.

Create custom quizzes from hundreds
of practice problems!

Register your book or ebook at
www.dummies.com/go/getaccess.

Select your product, and then follow the prompts
to validate your purchase.

You'll receive an email with your PIN and instructions.

1,001 Statistics Practice Problems

FOR DUMMIES®

A Wiley Brand

FOR DUMMIES®

A Wiley Brand

1,001 Statistics Practice Problems For Dummies®

Published by: **John Wiley & Sons, Inc.,** 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Published simultaneously in Canada

Library of Congress Control Number: 2013954236

ISBN 978-1-118-77604-9 (pbk); ISBN 978-1-118-77605-6 (ebk); ISBN 978-1-118-77616-2 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents at a Glance

Table of Contents

Introduction

One thousand and one practice problems for statistics! That's probably more than a professor would assign you in one semester (we hope!). And it's more than you'd ever want to tackle in one sitting (and we don't recommend you try). So why so many practice problems, and why this book?

Many textbooks are pretty thin on exercises, and even those that do contain a fair number of problems can't focus on all aspects of each topic. With so many problems available in this book, you get to choose how many problems you want to work on. And the way these problems are organized helps you quickly find and dig into problems on particular topics you need to study at the time. Whether you're into the normal distribution, hypothesis tests, the slope of a regression line, or histograms, it's all here and easy to find.

Then there's the entertainment factor. What better way to draw a crowd than to invite people over for a statistics practice problems marathon!

What You'll Find

This book contains 1,001 statistics problems divided into 17 chapters, organized by the major statistical topics in a first-semester introductory course. The problems basically take on three levels:

- **Statistical literacy:** Understanding the basic concepts of the topic, including terms and notation
- **Reasoning:** Applying the ideas within a context
- **Thinking:** Putting ideas and concepts together to solve more difficult problems

In addition to providing plenty of problems to work on in each chapter, this book also provides worked-out solutions with detailed explanations, so you aren't left high and dry if you get a wrong answer. So you can rest assured that when you work for 30 minutes on a problem, get an answer of 1.25, and go to the back of the book to see that the correct answer is actually 1,218.31, you'll find a detailed explanation to help you figure out what went wrong with your calculations.

How This Workbook Is Organized

This book is divided into two main parts: the questions and the answers.

Part 1: The Questions

The questions in this book center on the following areas:

- **Descriptive statistics and graphs:** After you collect and review data, your first job is to make sense of it. The way to do that is two-fold: (1) Organize the data in a visual manner so you can see it, and (2) crank out some numbers that describe it in a basic way.
- **Random variables:** A *random variable* is a characteristic of interest that varies in a random fashion. Each type of random variable has its own pattern in which the data falls (or is expected to fall), along with its own mean and standard deviation for the data. The pattern of a random variable is called its *distribution*.

  The random variables in this book include the binomial, the normal (or *Z*), and the *t*. For each random variable, you practice identifying its characteristics, seeing what its pattern (distribution) looks like, determining its mean and standard deviation, and, most commonly, finding probabilities and percentiles for it.
- **Inference:** This term can seem complex (and word on the street says it is), but inference basically just means taking the information from your data (your sample) and using it to draw conclusions about the group you're interested in (your population).

  The two basic types of statistical inferences are confidence intervals and hypothesis testing:

  - You use *confidence intervals* when you want to make an estimate regarding the population — for example, "What percentage of all kindergarteners in the United States are obese?"
  - You use a *hypothesis test* when someone has a supposed value regarding the population, and you're putting it to the test. For example, a researcher claims that 14 percent of today's kindergarteners are obese, but you question whether it's really that high.

  The underpinnings needed for both types of inference are margin of error, standard error, sampling distributions, and the central limit theorem. All of them play a major role in statistics and can be somewhat complex, so make sure you spend time on these elements as a backdrop for confidence intervals and hypothesis tests.
- **Relationships:** One of the most important and common uses of statistics is to look for relationships between two random variables. If variables are categorical (such as gender), you explore relationships by using two-way tables containing rows and columns, and you examine relationships by looking at and comparing percentages among and within groups. If both variables are numerical, you explore relationships graphically by using scatter plots, quantify them by using correlation, and use them to make predictions (one variable predicting the other) by using regression. Studying relationships helps you get at the essence of how statistics is applied in the real world.
- **Surveys:** Before you analyze data in all the ways mentioned in this list, you have to collect the data. Surveys are one of the most common means of data collection; the main ideas of surveys to address with practice are planning a survey, selecting a representative sample of individuals to survey, and carrying out the survey properly. The main goal in all of these areas is to avoid *bias* (systematic favoritism). Many types of bias exist, and in this book, you practice identifying and seeing ways to minimize them.

Part II: The Answers

This part provides detailed answers to every question in this book. You see how to set up and work through each problem and how to interpret the answer.

Beyond the Book

This book of 1,001 practice problems will keep you busy with pencil and paper for a while, but like the infomercials say, "Wait! There's more!" Your book purchase also comes with a free, one-year subscription to all 1,001 practice problems online. Track your progress and view personalized reports that show where you need to study the most and what you're pretty comfortable with.

What you'll find online

The online practice that comes free with this book offers you the same 1,001 questions and answers that are available here, presented in a multiple-choice format. Multiple-choice questions force you to zoom in on the details that can make or break your correct solution to the problem. Sometimes one of the possible wrong answers will catch you in the act of making a certain error. But that's great because after you identify a particular error (often a common error that many others make as well), you'll know not to fall into that trap again.

The beauty of the online problems is that you can customize your online practice — that is, you can select the types of problems and the number of problems you want to work on. The online program keeps track of your performance so you can focus on the areas where you need the biggest boost.

You can access this online tool by using a PIN code, as described in the next section. Keep in mind that you can create only one login with your PIN. Once the PIN is used, it's no longer valid and is nontransferable. So you can't share your PIN with others after you've established your login credentials. In other words, the problems are yours and only yours!

How to register

To gain access to the online practice, all you have to do is register. Just follow these simple steps:

1. **Register your book or ebook at Dummies.com to get your PIN. Go to `www.dummies.com/go/getaccess`.**
2. **Select your product from the dropdown list on that page.**
3. **Follow the prompts to validate your product, and then check your email for a confirmation message that includes your PIN and instructions for logging in.**

If you do not receive this email within two hours, please check your spam folder before contacting us through our Technical Support website at `http://support.wiley.com` or by phone at 877-762-2974.

Now you're ready to go! You can come back to the practice material as often as you want — simply log on with the username and password you created during your initial login. No need to enter the access code a second time.

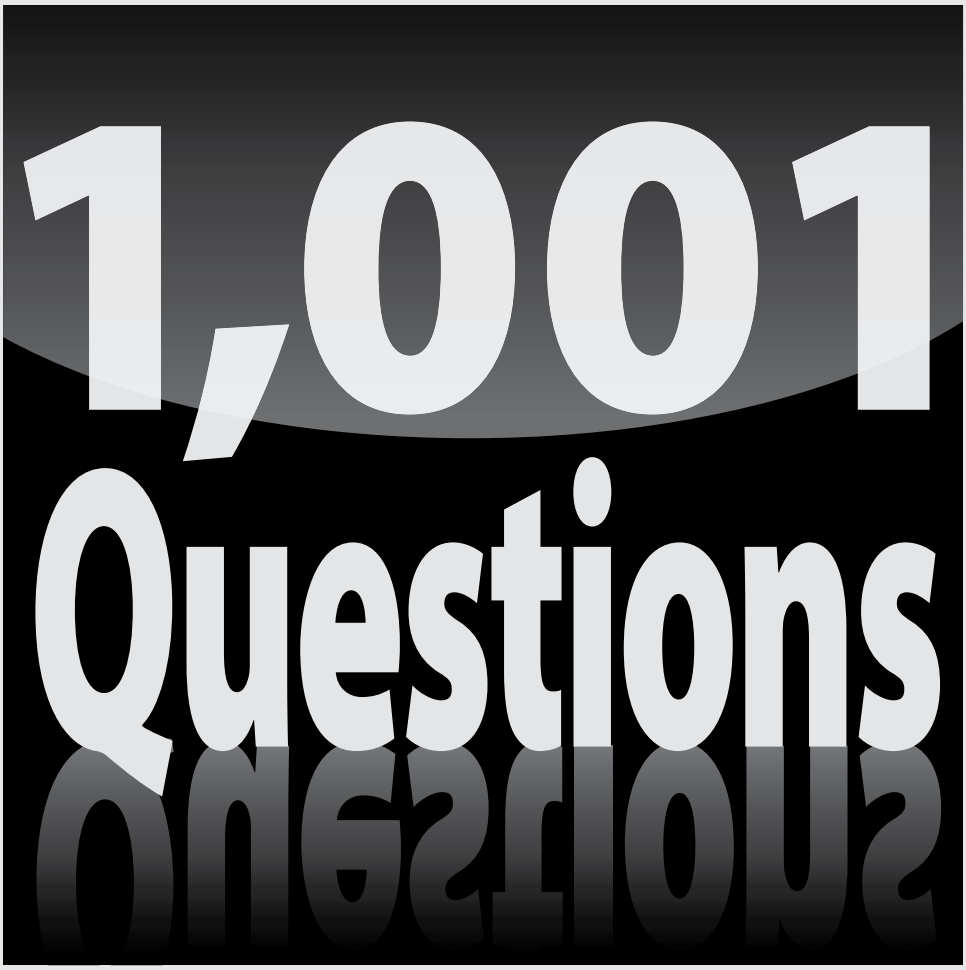
Your registration is good for one year from the day you activate your PIN.

Where to Go for Additional Help

The written solutions for the problems in this book are designed to show you what you need to do to get the correct answer to those particular problems. Although a bit of background information is injected at times, the solutions aren't meant to teach the material outright. Solutions to the problems on a given topic contain the normal statistical language, symbols, and formulas that are inherent to the topic, with the assumption that you're familiar with them.

If you're ever confused about why a problem is done a certain way, or you want more info to fill in between the lines, or you just feel like you need to go back and refresh your memory on some of the topics, several *For Dummies* books are available as a reference, including *Statistics For Dummies*, *Statistics Essentials For Dummies*, and *Statistics Workbook For Dummies*, all written by Deborah J. Rumsey, PhD, and published by Wiley.

Part I
The Questions



Visit www.dummies.com for great (and free!) Dummies content online.

In this part . . .

Statistics can give anyone problems. Terms, notation, formulas — where do you start? You start by practicing problems that hone the right skills. This book gives you practice — 1,001 problems worth of practice, to be exact. Working problems like these helps you figure out what you do and don't understand about setting up, working out, and interpreting your answers to statistics problems. Here's the breakdown in a nutshell:

- Warm up with statistical vocabulary, descriptive statistics, and graphs (Chapters 1 through 3).
- Work with random variables, including the binomial, normal, and *t*-distributions (Chapters 4 through 6).
- Decipher sampling distributions and margin of error and build confidence intervals for one- and two-population means and proportions (Chapters 7 through 10).
- Master the general concepts of hypothesis testing and perform tests for one- and two-population means and proportions (Chapters 11 through 13).
- Get behind the scenes on collecting good data and spotting bad data in surveys (Chapter 14).
- Explore relationships between two quantitative variables, using correlation and simple linear regression (Chapters 15 and 16).
- Look for relationships between two categorical variables, using two-way tables and independence (Chapter 17).

Chapter 1

Basic Vocabulary

Everything's got its own lingo, and statistics is no exception. The trick is to get a handle on the lingo right from the get-go, so when it comes time to work problems, you'll pick up on cues from the wording and get going in the right direction. You can also use the terms to search quickly in the table of contents or the index of this book to find the problems you need to dive into in a flash. It's like with anything else: As soon as you understand what the language means, you immediately start feeling more comfortable.

The Problems You'll Work On

In this chapter, you get a bird's-eye view of some of the most common terms used in statistics and, perhaps more importantly, the context in which they're used. Here's an overview:

- The big four: population, sample, parameter, and statistic
- The statistics terms you'll calculate, such as the mean, median, standard deviation, *z*-score, and percentile
- Types of data, graphs, and distributions
- Data analysis terms, such as confidence intervals, margin of error, and hypothesis tests

What to Watch Out For

Pay particular attention to the following:

- Pick out the big four in every situation; they'll follow you wherever you go.
- Really get the idea of a distribution; it's one of the most confusing ideas in statistics, yet it's used over and over — so nail it now to avoid getting hammered later.
- Focus not only on the terms for the statistics and analyses you'll calculate but also on their interpretation, especially in the context of a problem.

Picking Out the Population, Sample, Parameter, and Statistic

1–4 *You're interested in knowing what percent of all households in a large city have a single woman as the head of the household. To estimate this percentage, you conduct a survey with 200 households and determine how many of these 200 are headed by a single woman.*

1. In this example, what is the population?

2. In this example, what is the sample?

3. In this example, what is the parameter?

4. In this example, what is the statistic?

Distinguishing Quantitative and Categorical Variables

5–6 *Answer the problems about quantitative and categorical variables.*

5. Which of the following is an example of a quantitative variable (also known as a numerical variable)?

(A) the color of an automobile

(B) a person's state of residence

(C) a person's zip code

(D) a person's height, recorded in inches

(E) Choices (C) and (D)

6. Which of the following is an example of a categorical variable (also known as a qualitative variable)?

(A) years of schooling completed

(B) college major

(C) high-school graduate or not

(D) annual income (in dollars)

(E) Choices (B) and (C)

Getting a Handle on Bias, Variables, and the Mean

7–11 *You're interested in the percentage of female versus male shoppers at a department store. So one Saturday morning, you place data collectors at each of the store's four entrances for three hours, and you have them record how many men and women enter the store during that time.*

7. Why can collecting data at the store on one Saturday morning for three hours cause bias in the data?

(A) It assumes that Saturday shoppers represent the whole population of people who shop at the store during the week.

(B) It assumes that the same percentage of female shoppers shop on Saturday mornings as any other time or day of the week.

(C) Perhaps couples are more likely to shop together on Saturday mornings than during the rest of the week, bringing the percentage of males and females closer than during other times of the week.

(D) The subjects in the study weren't selected at random.

(E) All of these choices are true.

8. Because a variable is a characteristic of each individual on which data is collected, which of the following are variables in this study?

(A) the day you chose to collect data

(B) the store you chose to observe

(C) the gender of each shopper who comes in during the time period

(D) the number of men entering the store during the time period

(E) Choices (C) and (D)

9. In this study, _____ is a categorical variable, and _____ is a quantitative variable.

10. Which chart or graph would be appropriate to display the proportion of males versus females among the shoppers?

(A) a bar graph

(B) a time plot

(C) a pie chart

(D) Choices (A) and (C)

(E) Choices (A), (B), and (C)

11. How would you calculate the mean number of shoppers per hour?

Understanding Different Statistics and Data Analysis Terms

12–17 *Answer the problems about different statistics and data analysis terms.*

12. Which of the following data sets has a median of 3?

(A) 3, 3, 3, 3, 3

(B) 2, 5, 3, 1, 1

(C) 1, 2, 3, 4, 5

(D) 1, 2, 4, 4, 4

(E) Choices (A) and (C)

13. Susan scores at the 90th percentile on a math exam. What does this mean?

14. You took a survey of 100 people and found that 60% of them like chocolate and 40% don't. Which of the following gives the distribution of the "chocolate versus no chocolate" variable?

(A) a table of the results

(B) a pie chart of the results

(C) a bar graph of the results

(D) a sentence describing the results

(E) all of the above

15. Suppose that the results of an exam tell you your z-score is 0.70. What does this tell you about how well you did on the exam?

16. A national poll reports that 65% of Americans sampled approve of the president, with a margin of error of 6 percentage points. What does this mean?

17. If you want to estimate the percentage of all Americans who plan to vacation for two weeks or more this summer, what statistical technique should you use to find a range of plausible values for the true percentage?

Using Statistical Techniques

18–19 *You read a report that 60% of high-school graduates participated in sports during their high-school years.*

18. You believe that the percentage of high-school graduates who played sports is higher than what was reported. What type of statistical technique do you use to see whether you're right?

19. You believe that the percentage of high-school graduates who played sports in high school is higher than what's in the report. If you do a hypothesis test to challenge the report, which of these p-values would you be happiest to get?

(A) $p = 0.95$

(B) $p = 0.50$

(C) $p = 1$

(D) $p = 0.05$

(E) $p = 0.001$

Working with the Standard Deviation

20 *Solve the problem about standard deviation.*

20. Which data set has the highest standard deviation (without doing calculations)?

(A) 1, 2, 3, 4

(B) 1, 1, 1, 4

(C) 1, 1, 4, 4

(D) 4, 4, 4, 4

(E) 1, 2, 2, 4

Chapter 2

Descriptive Statistics

Descriptive statistics are statistics that describe data. You've got the staple ingredients, such as the mean, median, and standard deviation, and then the concepts and graphs that build on them, such as percentiles, the five-number summary, and the box plot. Your first job in analyzing data is to identify, understand, and calculate these descriptive statistics. Then you need to interpret the results, which means to see and describe their importance in the context of the problem.

The Problems You'll Work On

The problems in this chapter focus on the following big ideas:

- Calculating, interpreting, and comparing basic statistics, such as mean and median, and standard deviation and variance
- Using the mean and standard deviation to give ranges for bell-shaped data
- Measuring where a certain value stands in a data set by using percentiles
- Creating a set of five numbers (using percentiles) that can reveal some aspects of the shape, center, and variation in a data set

What to Watch Out For

Pay particular attention to the following:

- Be sure you identify which descriptive statistic or set of descriptive statistics is needed for a particular problem.
- After you understand the terminology and calculations for these descriptive statistics, step back and look at the results — make comparisons, see if they make sense, and find the story they tell.
- Remember that a percentile isn't a percent, even though they sound the same! When used together, remember that a percentile is a cutoff value in the data set, while a percentage is the amount of data that lies below that cutoff value.
- Be aware of the units of any descriptive statistic you calculate (for example, dollars, feet, or miles per gallon). Some descriptive statistics are in the same units as the data, and some aren't.

Understanding the Mean and the Median

21–32 *Solve the following problems about means and medians.*

21. To the nearest tenth, what is the mean of the following data set? 14, 14, 15, 16, 28, 28, 32, 35, 37, 38

22. To the nearest tenth, what is the mean of the following data set? 15, 25, 35, 45, 50, 60, 70, 72, 100

23. To the nearest tenth, what is the mean of the following data set? 0.8, 1.8, 2.3, 4.5, 4.8, 16.1, 22.3

24. To the nearest thousandth, what is the mean of the following data set? 0.003, 0.045, 0.58, 0.687, 1.25, 10.38, 11.252, 12.001

25. To the nearest tenth, what is the median of the following data set? 6, 12, 22, 18, 16, 4, 20, 5, 15

26. To the nearest tenth, what is the median of the following data set? 18, 21, 17, 18, 16, 15.5, 12, 17, 10, 21, 17

27. To the nearest tenth, what is the median of the following data set? 14, 2, 21, 7, 30, 10, 1, 15, 6, 8

28. To the nearest hundredth, what is the median of the following data set? 25.2, 0.25, 8.2, 1.22, 0.001, 0.1, 6.85, 13.2

29. Compare the mean and median of a data set that has a distribution that is skewed right.

30. Compare the mean and the median of a data set that has a distribution that is skewed left.

31. Compare the mean and the median of a data set that has a symmetrical distribution.

32. Which measure of center is most resistant to (or least affected by) outliers?

Surveying Standard Deviation and Variance

33–48 *Solve the following problems about standard deviation and variance.*

33. What does the standard deviation measure?

34. According to the 68-95-99.7 rule, or the empirical rule, if a data set has a normal distribution, approximately what percentage of data will be within one standard deviation of the mean?

35. A realtor tells you that the average cost of houses in a town is $176,000. You want to know how much the prices of the houses may vary from this average. What measurement do you need?

(A) standard deviation
(B) interquartile range
(C) variance
(D) percentile
(E) Choice (A) or (C)

36. What measure(s) of variation is/are sensitive to outliers?

(A) margin of error
(B) interquartile range
(C) standard deviation
(D) Choices (A) and (B)
(E) Choices (A) and (C)

37. You take a random sample of ten car owners and ask them, "To the nearest year, how old is your current car?" Their responses are as follows: 0 years, 1 year, 2 years, 4 years, 8 years, 3 years, 10 years, 17 years, 2 years, 7 years. To the nearest year, what is the standard deviation of this sample?

38. A sample is taken of the ages in years of 12 people who attend a movie. The results are as follows: 12 years, 10 years, 16 years, 22 years, 24 years, 18 years, 30 years, 32 years, 19 years, 20 years, 35 years, 26 years. To the nearest year, what is the standard deviation for this sample?

39. A large math class takes a midterm exam worth a total of 100 points. Following is a random sample of 20 students' scores from the class:

Score of 98 points: 2 students
Score of 95 points: 1 student
Score of 92 points: 3 students
Score of 88 points: 4 students
Score of 87 points: 2 students
Score of 85 points: 2 students
Score of 81 points: 1 student
Score of 78 points: 2 students
Score of 73 points: 1 student
Score of 72 points: 1 student
Score of 65 points: 1 student

To the nearest tenth of a point, what is the standard deviation of the exam scores for the students in this sample?

40. A manufacturer of jet engines measures a turbine part to the nearest 0.001 centimeters. A sample of parts has the following data set: 5.001, 5.002, 5.005, 5.000, 5.010, 5.009, 5.003, 5.002, 5.001, 5.000. What is the standard deviation for this sample?

41. Two companies pay their employees the same average salary of $42,000 per year. The salary data in Ace Corp. has a standard deviation of $10,000, whereas Magna Company salary data has a standard deviation of $30,000. What, if anything, does this mean?

42. In which of the following situations would having a small standard deviation be most important?

(A) determining the variation in the wealth of retired people

(B) measuring the variation in circuitry components when manufacturing computer chips

(C) comparing the population of cities in different areas of the country

(D) comparing the amount of time it takes to complete education courses on the Internet

(E) measuring the variation in the production of different varieties of apple trees

43. Suppose that you're comparing the means and standard deviations for the daily high temperatures for two cities during the months of November through March.

Sunshine City: $\mu = 46°\text{F}; \sigma = 18°\text{F}$

Lake Town: $\mu = 42°\text{F}; \sigma = 8°\text{F}$

What's the best analysis for comparing the temperatures in the two cities?

44. Everyone at a company is given a year-end bonus of $2,000. How will this affect the standard deviation of the annual salaries in the company that year?

45. Calculate the sample variance and the standard deviation for the following measurements of weights of apples: 7 oz, 6 oz, 5 oz, 6 oz, 9 oz. Express your answers in the proper units of measurement and round to the nearest tenth.

46. Calculate the sample variance and the standard deviation for the following measurements of assembly time required to build an MP3 player: 15 min, 16 min, 18 min, 10 min, 9 min. Express your answers in the proper units of measurement and round to the nearest whole number.

47. Calculate the standard deviation for these speeds of city traffic: 10 km/hr, 15 km/hr, 35 km/hr, 40 km/hr, 30 km/hr. Express your answers in the proper units of measurement and round to the nearest whole number.

48. Which of the following data sets has the same standard deviation as the data set with the numbers 1, 2, 3, 4, 5? (Do this problem without any calculations!)

(A) Data Set 1: 6, 7, 8, 9, 10

(B) Data Set 2: –2, –1, 0, 1, 2

(C) Data Set 3: 0.1, 0.2, 0.3, 0.4, 0.5

(D) Choices (A) and (B)

(E) None of the data sets gives the same standard deviation as the data set 1, 2, 3, 4, 5.

Employing the Empirical Rule

49–56 *Use the empirical rule to solve the following problems.*

49. According to the empirical rule (or the 68-95-99.7 rule), if a population has a normal distribution, approximately what percentage of values is within one standard deviation of the mean?

50. According to the empirical rule (or the 68-95-99.7 rule), if a population has a normal distribution, approximately what percentage of values is within two standard deviations of the mean?

51. If the average age of retirement for the entire population in a country is 64 years and the distribution is normal with a standard deviation of 3.5 years, what is the approximate age range in which 95% of people retire?

52. Last year's graduates from an engineering college, who entered jobs as engineers, had a mean first-year income of $48,000 with a standard deviation of $7,000. The distribution of salary levels is normal. What is the approximate percentage of first-year engineers that made more than $55,000?

53. What is a necessary condition for using the empirical rule (or 68-95-99.7 rule)?

54. What measures of data need to be known to use the empirical (68-95-99.7) rule?

55. The quality control specialists of a microscope manufacturing company test the lens for every microscope to make sure the dimensions are correct. In one month, 600 lenses are tested. The mean thickness is 2 millimeters. The standard deviation is 0.000025 millimeters. The distribution is normal. The company rejects any lens that is more than two standard deviations from the mean. Approximately how many lenses out of the 600 would be rejected?

56. Biologists gather data on a sample of fish in a large lake. They capture, measure the length of, and release 1,000 fish. They find that the standard deviation is 5 centimeters, and the mean is 25 centimeters. They also notice that the shape of the distribution (according to a histogram) is very much skewed to the left (which means that some fish are smaller than most of the others). Approximately what percentage of fish in the lake is likely to have a length within one standard deviation of the mean?

Measuring Relative Standing with Percentiles

57–64 *Solve the following problems about percentiles.*

57. What statistic reports the relative standing of a value in a set of data?

58. What is the statistical name for the 50th percentile?

59. Your score on a test is at the 85th percentile. What does this mean?

60. Suppose that in a class of 60 students, the final exam scores have an approximately normal distribution, with a mean of 70 points and a standard deviation of 5 points. Bob's score places him in the 90th percentile among students on this exam. What must be true about Bob's score?

61. On a multiple-choice test, your actual score was 82%, which was reported to be at the 70th percentile. What is the meaning of your test results?

62. Seven students got the following exam scores (percent correct) on a science exam: 0%, 40%, 50%, 65%, 75%, 90%, 100%. Which of these exam scores is at the 50th percentile?

63. Students scored the following grades on a statistics test: 80, 80, 82, 84, 85, 86, 88, 90, 91, 92, 92, 94, 96, 98, 100. Calculate the score that represents the 80th percentile.

64. Some of the students in a class are comparing their grades on a recent test. Mary says she almost scored in the 95th percentile. Lisa says she scored at the 84th percentile. Jose says he scored at the 88th percentile. Paul says he almost scored in the 70th percentile. Bill says he scored at the 95th percentile. Rank the five students from highest to lowest in their grades.