Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

SKU: 9780135780275-SOLUTIONS

INSTRUCTOR'S SOLUTIONS MANUAL

DIACRITECH

STATISTICS: INFORMED DECISIONS USING DATA

SIXTH EDITION

Michael Sullivan, III

Joliet Junior College

Pearson

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

This work is protected by United States copyright laws and is provided solely for the use of instructors in teaching their courses and assessing student learning. Dissemination or sale of any part of this work (including on the World Wide Web) will destroy the integrity of the work and is not permitted. The work and materials from it should never be made available to students except by instructors using the accompanying text in their classes. All recipients of this work are expected to abide by these restrictions and to honor the intended pedagogical purposes and the needs of other instructors who rely on these materials.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Reproduced by Pearson from electronic files supplied by the author.

Copyright © 2021, 2017, 2013 by Pearson Education, Inc. 221 River Street, Hoboken, NJ 07030. All rights reserved.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

ISBN-13: 978-0-13-578017-6
ISBN-10: 0-13-578017-9

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Table of Contents

Preface

Chapter 1 Data Collection
1.1 Introduction to the Practice of Statistics 1
1.2 Observational Studies versus Designed Experiments 4
1.3 Simple Random Sampling 7
1.4 Other Effective Sampling Methods 9
1.5 Bias in Sampling 11
1.6 The Design of Experiments 14
Chapter 1 Review Exercises 24
Chapter 1 Test 28
Case Study: Chrysalises for Cash 30

Chapter 2 Organizing and Summarizing Data
2.1 Organizing Qualitative Data 32
2.2 Organizing Quantitative Data: The Popular Displays 42
2.3 Additional Displays of Quantitative Data 54
2.4 Graphical Misrepresentations of Data 64
Chapter 2 Review Exercises 66
Chapter 2 Test 72
Case Study: The Day the Sky Roared 76

Chapter 3 Numerically Summarizing Data
3.1 Measures of Central Tendency 79
3.2 Measures of Dispersion 86
3.3 Measures of Central Tendency and Dispersion from Grouped Data 103
3.4 Measures of Position and Outliers 113
3.5 The Five-Number Summary and Boxplots 120
Chapter 3 Review Exercises 131
Chapter 3 Test 137
Case Study: Who Was "A Mourner"? 140

Chapter 4 Describing the Relation between Two Variables
4.1 Scatter Diagrams and Correlation 142
4.2 Least-Squares Regression 163
4.3 Diagnostics on the Least-Squares Regression Line 175
4.4 Contingency Tables and Association 189
4.5 Nonlinear Regression: Transformations (online) 201
Chapter 4 Review Exercises 207
Chapter 4 Test 215
Case Study: Thomas Malthus, Population, and Subsistence 219

Chapter 5 Probability
5.1 Probability Rules 221
5.2 The Addition Rule and Complements 230
5.3 Independence and the Multiplication Rule 238
5.4 Conditional Probability and the General Multiplication Rule 242
5.5 Counting Techniques 249
5.6 Simulating Probability Experiments 253
5.7 Putting It Together: Which Method Do I Use? 255
5.8 Bayes's Rule (online) 258

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Chapter 5 Review Exercises 264
Chapter 5 Test 267
Case Study: The Case of the Body in the Bag 269

Chapter 6 Discrete Probability Distributions
6.1 Discrete Random Variables 271
6.2 The Binomial Probability Distribution 281
6.3 The Poisson Probability Distribution 296
6.4 The Hypergeometric Probability Distribution (online) 303
6.5 Combining Random Variables (Online) 307
Chapter 6 Review Exercises 312
Chapter 6 Test 316
Case Study: The Voyage of the St. Andrew 319

Chapter 7 The Normal Probability Distribution
7.1 Properties of the Normal Distribution 321
7.2 Applications of the Normal Distribution 324
7.3 Assessing Normality 344
7.4 The Normal Approximation to the Binomial Probability Distribution 347
Chapter 7 Review Exercises 351
Chapter 7 Test 356
Case Study: A Tale of Blood Chemistry 359

Chapter 8 Sampling Distributions
8.1 Distribution of the Sample Mean 361
8.2 Distribution of the Sample Proportion 374
Chapter 8 Review Exercises 380
Chapter 8 Test 383
Case Study: Sampling Distribution of the Median 385

Chapter 9 Estimating the Value of a Parameter
9.1 Estimating a Population Proportion 391
9.2 Estimating a Population Mean 398
9.3 Estimating a Population Standard Deviation 408
9.4 Putting It Together: Which Method Do I Use? 410
9.5 Estimating with Bootstrapping 416
Chapter 9 Review Exercises 420
Chapter 9 Test 425
Case Study: Fire-Safe Cigarettes 428

Chapter 10 Hypothesis Tests Regarding a Parameter
10.1 The Language of Hypothesis Testing 429
10.2 Hypothesis Tests for a Population Proportion 432
10.2A Using Simulation to Perform Hypothesis Tests on a Population Proportion 443
10.2B Hypothesis Tests for a Population Proportion Using the Normal Model 453
10.3 Hypothesis Tests for a Population Mean 463
10.3A Using Simulation and the Bootstrap to Perform Hypothesis Tests on a Population Mean 473
10.4 Hypothesis Tests for a Population Standard Deviation 480
10.5 Putting It Together: Which Method Do I Use? 483
10.6 The Probability of a Type II Error and the Power of the Test 489
Chapter 10 Review Exercises 494
Chapter 10 Test 499
Case Study: How Old Is Stonehenge? 500

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Chapter 11 Inference on Two Population Parameters
11.1 Inference about Two Population Proportions....503
11.1A Using Randomization Techniques to Compare Two Proportions....518
11.2 Inference about Two Means: Dependent Samples....525
11.2A Using Bootstrapping to Conduct Inference on Two Dependent Means....537
11.3 Inference about Two Means: Independent Samples....544
11.3A Using Randomization Techniques to Compare Two Independent Means....558
11.4 Inference about Two Population Standard Deviations....562
11.5 Putting It Together: Which Method Do I Use?....566
Chapter 11 Review Exercises....581
Chapter 11 Test....587
Case Study: Control in the Design of an Experiment....594

Chapter 12 Inference on Categorical Data
12.1 Goodness-of-Fit Test....596
12.2 Tests for Independence and the Homogeneity of Proportions....607
12.3 Inference about Two Population Proportions: Dependent Samples....629
Chapter 12 Review Exercises....632
Chapter 12 Test....636
Case Study: Feeling Lucky? Well, Are You?....641

Chapter 13 Comparing Three or More Means
13.1 Comparing Three or More Means (One-Way Analysis of Variance)....644
13.2 Post Hoc Tests on One-Way Analysis of Variance....654
13.3 The Randomized Complete Block Design....662
13.4 Two-Way Analysis of Variance....670
Chapter 13 Review Exercises....682
Chapter 13 Test....687
Case Study: Hat Size and Intelligence....691

Chapter 14 Inference on the Least-Squares Regression Model and Multiple Regression
14.1 Testing the Significance of the Least-Squares Regression Model....694
14.1A Using Randomization Techniques on the Slope of the Least-Squares Regression Line....700
14.2 Confidence and Prediction Intervals....705
14.3 Introduction to Multiple Regression....712
14.4 Interaction and Dummy Variables....723
14.5 Polynomial Regression....726
14.6 Building a Regression Model....731
Chapter 14 Review Exercises....735
Chapter 14 Test....739
Case Study: Housing Boom....742

Chapter 15 Nonparametric Statistics
15.1 An Overview of Nonparametric Statistics....748
15.2 Runs Test for Randomness....748
15.3 Inference about Measures of Central Tendency....750
15.4 Inference about the Difference between Two Medians: Dependent Samples....754
15.5 Inference about the Difference between Two Medians: Independent Samples....757
15.6 Spearman's Rank-Correlation Test....762
15.7 Kruskal-Wallis Test....769
Chapter 15 Review Exercises....774
Chapter 15 Test....778
Case Study: Evaluating Alabama's 1891 House Bill 504....780

Solutions to Consumer Reports®....Online

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Preface

This solutions manual accompanies *Statistics: Informed Decisions Using Data, 6e* by Michael Sullivan, III. The Instructor's Solutions Manual contains detailed solutions to all exercises in the text as well as the Consumer Reports® projects and the case studies. The Student's Solutions Manual contains detailed solutions to all odd exercises in the text and all solutions to chapter reviews and tests. A concerted effort has been made to make this manual as user-friendly and error free as possible.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Chapter 1
Data Collection

Section 1.1

1. **(a)** III. Statistics is the science of collecting, organizing, summarizing, and analyzing information in order to draw conclusions and answer questions. It is also about providing a measure of confidence in any conclusions.

   (b) VIII. The population is the entire group of individuals to be studied.

   (c) IV. The sample is a subset of the group of individuals that is being studied.

   (d) VII. The parameter is a numerical summary of a population.

   (e) I. The statistic is the numerical summary of a sample.

   (f) VI. The individual is a person or object that is a member of the group being studied.

   (g) II. Descriptive statistics involves organizing and summarizing data through tables, graphs, and numerical summaries.

   (h) V. Inferential statistics uses methods that take results from a sample and extends them to the population, and measures the reliability of the result.

2. **(a)** V. A discrete variable has either a finite number of possible values or countable number of possible values. The values of these variables typically result from counting.

   (b) IV. Data are information that describe characteristics of an individual.

   (c) VI. A continuous variable has an infinite number of possible values that are not countable. The values of these variables typically result from measurement.

   (d) II. A qualitative variable allows for classification of individuals based on some attribute or characteristic.

   (e) I. A quantitative variable provides numerical measures of individuals. The measures can be added or subtracted, and provide meaningful results.

   (f) III. A variable is the characteristics of the individuals within the population.

3. 18% is a parameter because it describes a population (all of the governors).

4. 72% is a parameter because it describes a population (the entire class).

5. 32% is a statistic because it describes a sample (the high school students surveyed).

6. 13.3% is a statistic because it describes a sample (the 12th graders surveyed).

7. 0.366 is a parameter because it describes a population (all of Ty Cobb's at-bats).

8. 43.92 hours is a parameter because it describes a population (all the men who have walked on the moon).

9. 23% is a statistic because it describes a sample (the 6076 adults studied).

10. 44% is a statistic because it describes a sample (the 100 adults interviewed).

11.	Qualitative	**12.**	Quantitative
13.	Quantitative	**14.**	Qualitative
15.	Quantitative	**16.**	Quantitative
17.	Qualitative	**18.**	Qualitative
19.	Discrete	**20.**	Continuous
21.	Continuous	**22.**	Discrete
23.	Continuous	**24.**	Continuous
25.	Discrete	**26.**	Continuous
27.	Nominal	**28.**	Ordinal
29.	Ratio	**30.**	Interval
31.	Ordinal	**32.**	Nominal
33.	Ratio	**34.**	Interval

35. The population consists of all teenagers 13 to 17 years old who live in the United States. The sample consists of the 1028 teenagers 13 to 17 years old who were contacted by the Gallup Organization.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

36. The population consists of all bottles of Coca-Cola filled by that particular machine on October 15. The sample consists of the 50 bottles of Coca-Cola that were selected by the quality control manager.

37. The population consists of all of the soybean plants in this farmer's crop. The sample consists of the 100 soybean plants that were selected by the farmer.

38. The population consists of all households within the United States. The sample consists of the 50,000 households that are surveyed by the U.S. Census Bureau.

39. The population consists of all women 27 to 44 years of age with hypertension. The sample consists of the 7373 women 27 to 44 years of age with hypertension who were included in the study.

40. The population consists of all full-time students enrolled at this large community college. The sample consists of the 128 full-time students who were surveyed by the administration.

41. Individuals: Alabama, Colorado, Indiana, North Carolina, Wisconsin.
Variables: Minimum age for driver's license (unrestricted); mandatory belt use seating positions, maximum allowable speed limit (rural interstate) in 2011.
Data for minimum age for driver's license: 17, 17, 18, 16, 18;
Data for mandatory belt use seating positions: front, front, all, all, all;
Data for maximum allowable speed limit (rural interstate) 2011: 70, 75, 70, 70, 65 (mph.)
The variable *minimum age for driver's license* is continuous; the variable *mandatory belt use seating positions* is qualitative; the variable *maximum allowable speed limit (rural interstate) 2011* is continuous (although only discrete values are typically chosen for speed limits.)

42. Individuals: 3 Series, 5 Series, 6 Series, 7 Series, X3, Z4 Roadster
Variables: Body Style, Weight (lb), Number of Seats
Data for body style: Coupe, Sedan, Convertible, Sedan, Sport utility, Coupe;
Data for weight: 3362, 4056, 4277, 4564, 4012, 3505 (lb);
Data for number of seats: 4, 5, 4, 5, 5, 2. The variable *body style* is qualitative; the variable *weight* is continuous; the variable *number of seats* is discrete.

43. **(a)** The research objective is to determine if adolescents who smoke have a lower IQ than nonsmokers.

(b) The population is all adolescents aged 18–21. The sample consisted of 20,211 18-year-old Israeli military recruits.

(c) Descriptive statistics: The average IQ of the smokers was 94, and the average IQ of nonsmokers was 101.

(d) The conclusion is that individuals with a lower IQ are more likely to choose to smoke.

44. **(a)** The research objective is to determine if the application of duct tape is as effective as cryotherapy in the treatment of common warts.

(b) The population is all people with warts. The sample consisted of 51 patients with warts.

(c) Descriptive statistics: 85% of patients in group 1 and 60% of patients in group 2 had complete resolution of their warts.

(d) The conclusion is that duct tape is significantly more effective in treating warts than cryotherapy.

45. **(a)** The research objective is to determine the proportion of adult Americans who believe the federal government wastes 51 cents or more of every dollar.

(b) The population is all adult Americans aged 18 years or older.

(c) The sample is the 1017 American adults aged 18 years or older that were surveyed.

(d) Descriptive statistics: Of the 1017 individuals surveyed, 35% indicated that 51 cents or more is wasted.

(e) From this study, one can infer that many Americans believe the federal government wastes much of the money collected in taxes.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

46. **(a)** The research objective is to determine what proportion of adults, aged 18 and over, believe it would be a bad idea to invest $1000 in the stock market.

(b) The population is all adults aged 18 and over living in the United States.

(c) The sample is the 1018 adults aged 18 and over living in the United States who completed the survey.

(d) Descriptive statistics: Of the 1016 adults surveyed, 46% believe it would be a bad idea to invest $1000 in the stock market.

(e) The conclusion is that a little fewer than half of the adults in the United States believe investing $1000 in the stock market is a bad idea.

47. **(a)** State is a *qualitative* variable because it is an individual categorization.

(b) F scale is a *qualitative* variable because each tornado is rated according to a category.

(c) Fatalities is a *quantitative* variable because it is a numerical measure. It is a *discrete* variable because it is countable.

(d) Length is a *quantitative* variable because it is a numerical measure. It is a *continuous* variable because it results from measurement.

48. **(a)** State is a variable measured at the *nominal* level because values of the variable name, label, or categorize. In addition, the naming scheme does not allow for the values of the variable to be arranged in a ranked or specific order.

(b) F scale is a variable measured at the *ordinal* level because the naming scheme allows for the values of the variable to be arranged in a ranked or specific order.

(c) Fatalities is a variable measured at the *ratio* level because the ratio of two values makes sense and a value of zero has meaning.

(d) Length is a variable measured at the ratio level because the ratio of two values makes sense and a value of zero has meaning.

49. *Jersey number* is nominal (the numbers generally indicate a type of position played). However, if the researcher feels that lower caliber players received higher numbers, then *jersey number* would be ordinal since players could be ranked by their number.

50. **(a)** Nominal; the ticket number is categorized as a winner or a loser.

(b) Ordinal; the ticket number gives an indication as to the order of arrival of guests.

(c) Ratio; the implication is that the ticket number gives an indication of the number of people attending the party.

51. **(a)** The research question is to determine if the season of birth affects mood later in life.

(b) The sample consisted of the 400 people the researchers studied.

(c) The season in which you were born (winter, spring, summer, or fall) is a qualitative variable.

(d) According to the article, individuals born in the summer are characterized by rapid, frequent swings between sad and cheerful moods, while those born in the winter are less likely to be irritable.

(e) The conclusion was that the season at birth plays a role in one's temperament.

52. The population is the group to be studied as defined by the research objective. A sample is any subset of the population.

53. Quantitative variables are numerical measures such that meaningful arithmetic operations can be performed on the values of the variable. Qualitative variables describe an attribute or characteristic of the individual that allows researchers to categorize the individual.

The values of a discrete random variable result from counting. The values of a continuous random variable result from a measurement.

54. The four levels of measurement of a variable are nominal, ordinal, interval, and ratio. Examples: Nominal—brand of clothing; Ordinal—size of a car (small, mid-size, large); Interval—temperature (in degrees Celsius); Ratio—number of students in a class (Examples will vary.)

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

55. We say data vary, because when we draw a random sample from a population, we do not know which individuals will be included. If we were to take another random sample, we would have different individuals and therefore different data. This variability affects the results of a statistical analysis because the results would differ if a study is repeated.

56. The process of statistics is to (1) identify the research objective, which means to determine what should be studied and what we hope to learn; (2) collect the data needed to answer the research question, which is typically done by taking a random sample from a population; (3) describe the data, which is done by presenting descriptive statistics; and (4) perform inference in which the results are generalized to a larger population.

57. Age could be considered a discrete random variable. A random variable can be discrete by allowing, for example, only whole numbers to be recorded.

Section 1.2

1. The response variable is the variable of interest in a research study. An explanatory variable is a variable that affects (or explains) the value of the response variable. In research, we want to see how changes in the value of the explanatory variable affect the value of the response variable.

2. **(a)** III. A designed experiment is when a researcher randomly assigns the individuals in a study to groups, intentionally manipulates the value of an explanatory variable, controls other explanatory variables at fixed values, and then records the value of the response variable for each individual.

(b) V. An observational study is when a researcher measures the value of the response variable without attempting to influence the value of either the response or explanatory variables. That is, the researcher observes the behavior of individuals in the study and records the values of the explanatory and response variables.

(c) IV. A lurking variable is an explanatory variable that was not considered in a study, but that affects the value of the response variable in the study. In addition, this variable is typically related to other explanatory variables in the study.

(d) I. Confounding occurs when the effects of two or more explanatory variables are not separated. Therefore, any relation that may exist between an explanatory variable and the response variable may be due to some other variable not accounted for in the study.

(e) II. A confounding variable is an explanatory variable that was considered in a study whose effect cannot be distinguished from a second explanatory variable in the study.

3. **(a)** II. A cohort study follows a group of individuals over a long period of time. Characteristics of the individuals are recorded and some individuals will be exposed to certain factors (not intentionally) and others will not. Because the data are collected over time, cohort studies are prospective.

(b) III. A cross-sectional study collects information about individuals at a specific point in time, or over a short period of time.

(c) I. A case-control study is retrospective, meaning it requires the researcher to look at existing records, or the subject to recall information from the past. Individuals who have certain characteristics are matched with those who don't.

4. An observational study uses data obtained by studying individuals in a sample without trying to manipulate or influence the variable(s) of interest. In a designed experiment, a treatment is applied to the individuals in a sample in order to isolate the effects of the treatment on a response variable. Only an experiment can establish causation between an explanatory variable and a response variable. Observational studies can indicate a relationship, but cannot establish causation.

5. The choice between an observational study and an experiment depends on the circumstances involved. Sometimes there are ethical reasons why an experiment cannot be conducted. Other times the researcher may conduct an observational study first to validate a belief prior to investing a large amount of time and money into a designed experiment. A designed experiment is preferred if ethics, time, and money are not an issue.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

6. A cohort study identifies the individuals to participate and then follows them over a period of time. During this period, information about the individuals is gathered, but there is no attempt to influence the individuals. Cohort studies are superior to case-control studies because cohort studies do not require recall to obtain the data.

7. There is a perceived benefit to obtaining a flu shot, so there are ethical issues in intentionally denying certain seniors access to the treatment.

8. A retrospective study looks at data from the past either through recall or existing records. A prospective study gathers data over time by following the individuals in the study and recording data as they occur.

9. This is an observational study because the researchers merely observed existing data. There was no attempt by the researchers to manipulate or influence the variable(s) of interest.

10. This is an experiment because the researchers intentionally changed the value of the explanatory variable (medication dose) to observe a potential effect on the response variable (cancer growth).

11. This is an experiment because the explanatory variable (teaching method) was intentionally varied to see how it affected the response variable (score on proficiency test).

12. This is an observational study because no attempt was made to influence the variable of interest. Voting choices were merely observed.

13. This is an observational study because the survey only observed preference of Coke or Pepsi. No attempt was made to manipulate or influence the variable of interest.

14. This is an experiment because the researcher intentionally imposed treatments on individuals in a controlled setting.

15. This is an experiment because the explanatory variable (carpal tunnel treatment regimen) was intentionally manipulated in order to observe potential effects on the response variable (level of pain).

16. This is an observational study because the conservation agents merely observed the fish to determine which were carrying parasites. No attempt was made to manipulate or influence any variable of interest.

17. **(a)** This is a cohort study because the researchers observed a group of people over a period of time.

(b) The response variable is whether the individual has heart disease or not. The explanatory variable is whether the individual is happy or not.

(c) There may be confounding due to lurking variables. For example, happy people may be more likely to exercise, which could affect whether they will have heart disease or not.

18. **(a)** This is a cross-sectional study because the researchers collected information about the individuals at a specific point in time.

(b) The response variable is whether the woman has nonmelanoma skin cancer or not. The explanatory variable is the daily amount of caffeinated coffee consumed.

(c) It was necessary to account for these variables to avoid confounding with other variables.

19. **(a)** This is an observational study because the researchers simply administered a questionnaire to obtain their data. No attempt was made to manipulate or influence the variable(s) of interest. This is a cross-sectional study because the researchers are observing participants at a single point in time.

(b) The response variable is body mass index. The explanatory variable is whether a TV is in the bedroom or not.

(c) Answers will vary. Some lurking variables might be the amount of exercise per week and eating habits. Both of these variables can affect the body mass index of an individual.

(d) The researchers attempted to avoid confounding due to other variables by taking into account such variables as "socioeconomic status."

(e) No. Since this was an observational study, we can only say that a television in the bedroom is associated with a higher body mass index.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

20. **(a)** This is an observational study because the researchers merely observed the individuals included in the study. No attempt was made to manipulate or influence any variable of interest.
This is a cohort study because the researchers identified the individuals to be included in the study, then followed them for a period of time (7 years).

(b) The response variable is weight gain. The explanatory variable is whether the individual is married/cohabitating or not.

(c) Answers will vary. Some potential lurking variables are eating habits, exercise routine, and whether the individual has children.

(d) No. Since this is an observational study, we can only say that being married or cohabitating is associated with weight gain.

21. **(a)** This is a cross-sectional study because information was collected at a specific point in time (or over a very short period of time).

(b) The explanatory variable is delivery scenario (caseload midwifery, standard hospital care, or private obstetric care).

(c) The two response variables are (1) cost of delivery, which is quantitative, and (2) type of delivery (vaginal or not), which is qualitative.

22. **(a)** The explanatory variable is web page design; qualitative

(b) The response variables are time on site and amount spent. Both are qualitative.

(c) Answers will vary. A confounding variable might be location. Any differences in spending may be due to location rather than to web page design.

23. Answers will vary. This is a prospective, cohort observational study. The response variable is whether the worker had cancer or not, and the explanatory variable is the amount of electromagnetic field exposure. Some possible lurking variables include eating habits, exercise habits, and other health-related variables such as smoking habits. Genetics (family history) could also be a lurking variable. This was an observational study, and not an experiment, so the study only concludes that high electromagnetic field exposure is associated with higher cancer rates.
The author reminds us that this is an observational study, so there is no direct control over the variables that may affect cancer rates. He also points out that while we should not simply dismiss such reports, we should consider the results in conjunction with results from future studies. The author concludes by mentioning known ways (based on extensive study) of reducing cancer risks that can currently be done in our lives.

24. **(a)** This is a cohort study because a group of individuals was identified to participate in the study, and then they were observed over a period of time.

(b) Because there is a link established between obesity and cell phone use, and a link between obesity and negative health outcomes, if it is determined that cell-phone users are experiencing higher incidence rates of negative health outcomes, it cannot be established that the factor leading to the negative health outcome is due to the cell phone—it may be due to the lurking variable obesity.

25. Because individuals in the early 1900s were pressured to become right-handed, we would see a lower proportion of left-handers who are older in the study. This would make it seem as though left-handers die younger because the older individuals in the study are primarily right-handed.

26. **(a)** The research objective is to determine whether lung cancer is associated with exposure to tobacco smoke within the household.

(b) This is a case-controlled study because there is a group of individuals with a certain characteristic (lung cancer but never smoked) being compared to a similar group without the characteristic (no lung cancer and never smoked). The study is retrospective because lifetime residential histories were compiled and analyzed.

(c) The response variable is whether the individual has lung cancer or not. This is a qualitative variable.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

(d) The explanatory variable is the number of "smoker years." This is a quantitative variable.

(e) Answers will vary. Some possible lurking variables are household income, exercise routine, and exposure to tobacco smoke outside the home.

(f) The conclusion of the study is that approximately 17% of lung cancer cases among nonsmokers can be attributed to high levels of exposure to tobacco smoke during childhood and adolescence. No, we cannot say that exposure to household tobacco smoke causes lung cancer since this is only an observational study. We can, however, conclude that lung cancer is associated with exposure to tobacco smoke in the home.

(g) An experiment involving human subjects is not possible for ethical reasons. Researchers would be able to conduct an experiment using laboratory animals, such as rats.

27. Web scraping can be used to extract data from tables on web pages and then upload the data to a file. Web scraping can also be used to create a data set of words from an online article (that is, fetching unstructured information and transforming it into a structured format through something called parsing and reformatting processes). In addition, web scraping can be used to dynamically call information from websites with links.

28. Answers will vary. Discussions may include the responsibility of host sites to protect private information; the ethics behind using information collected from a site to harm a competitor; the ethics behind putting stress on a host's servers so that the website slows down; the role of social media to scrape its site for "fake news."

Section 1.3

1. The frame is a list of all the individuals in the population.

2. Simple random sampling occurs when every possible sample of size n has an equally likely chance of occurring.

3. Sampling without replacement means that no individual may be selected more than once as a member of the sample.

4. Random sampling is a technique that uses chance to select individuals from a population to be in a sample. It is used because it maximizes the likelihood that the individuals in the sample are representative of the individuals in the population. In convenience sampling, the individuals in the sample are selected in the quickest and easiest way possible (e.g. the first 20 people to enter a store). Convenience samples likely do not represent the population of interest because chance was not used to select the individuals.

5. Answers will vary. We will use one-digit labels and assign the labels across each row (i.e. *Pride and Prejudice* – 0, *The Sun Also Rises* – 1, and so on). In Table I of Appendix A, starting at row 5, column 11, and proceeding downward, we obtain the following labels: 8, 4, 3
In this case, the 3 books in the sample would be *As I Lay Dying*, *A Tale of Two Cities*, and *Crime and Punishment*. Different labeling order, different starting points in Table I in Appendix A, or use of technology will likely yield different samples.

6. Answers will vary. We will use one-digit labels and assign the labels across each row (i.e. *Mady* – 0, *Breanne* – 1, and so on). In Table I of Appendix A, starting at row 11, column 6, and then proceeding downward, we obtain the following labels: 1, 5
In this case, the two captains would be Breanne and Payton. Different labeling order, different starting points in Table I in Appendix A, or use of technology will likely yield different results.

7. **(a)** {616, 630}, {616, 631}, {616, 632}, {616, 645}, {616, 649}, {616, 650}, {630, 631}, {630, 632}, {630, 645}, {630, 649}, {630, 650}, {631, 632}, {631, 645}, {631, 649}, {631, 650}, {632, 645}, {632, 649}, {632, 650}, {645, 649}, {645, 650}, {649, 650}

(b) There is a 1 in 21 chance that the pair of courses will be EPR 630 and EPR 645.

8. **(a)** {1, 2}, {1, 3}, {1, 4}, {1, 5}, {1, 6}, {1, 7}, {2, 3}, {2, 4}, {2, 5}, {2, 6}, {2, 7}, {3, 4}, {3, 5}, {3, 6}, {3, 7}, {4, 5}, {4, 6}, {4, 7}, {5, 6}, {5, 7}, {6, 7}

(b) There is a 1 in 21 chance that the pair *The United Nations* and *Amnesty International* will be selected.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

9. (a) Starting at row 5, column 22, using two-digit numbers, and proceeding downward, we obtain the following values: 83, 94, 67, 84, 38, 22, 96, 24, 36, 36, 58, 34,.... We must disregard 94 and 96 because there are only 87 faculty members in the population. We must also disregard the second 36 because we are sampling without replacement. Thus, the 9 faculty members included in the sample are those numbered 83, 67, 84, 38, 22, 24, 36, 58, and 34

(b) Answers will vary depending on the type of technology used. If using a TI-84 Plus, the sample will be: 4, 20, 52, 5, 24, 87, 67, 86, and 39.

```
47→rand
                47
randInt(1,87)
                 4
                20
                52
                 5
```

```
                 5
                24
                87
                67
                20
                86
                39
```

Note: We must disregard the second 20 because we are sampling without replacement.

10. (a) Starting at row 11, column 32, using four-digit numbers, and proceeding downward, we obtain the following values: 2869, 5518, 6635, 2182, 8906, 0603, 2654, 2686, 0135, 7783, 4080, 6621, 3774, 7887, 0826, 0916, 3188, 0876, 5418, 0037, 3130, 2882, 0662,…. We must disregard 8906, 7783, and 7887 because there are only 7656 students in the population.
Thus, the 20 students included in the sample are those numbered 2869, 5518, 6635, 2182, 0603, 2654, 2686, 0135, 4080, 6621, 3774, 0826, 0916, 3188, 0876, 5418, 0037, 3130, 2882, and 0662.

(b) Answers may vary depending on the type of technology used. If using a TI-84 Plus, the sample will be: 6658, 4118, 9, 4828, 3905, 454, 2825, 2381, 495, 4445, 4455, 5759, 5397, 7066, 3404, 6667, 5074, 3777, 3206, 5216.

```
142→rand
               142
```

```
randInt(1,7656)
              6658
              4118
                 9
              4828
              3905
               454
```

```
              2825
              2381
               495
              4445
              4455
              5759
              5397
```

```
              7066
              3404
              6667
              5074
              3777
              3206
              5216
```

11. (a) Answers will vary depending on the technology used (including a table of random digits). Using a TI-84 Plus graphing calculator with a seed of 17 and the labels provided, our sample would be North Dakota, Nevada, Tennessee, Wisconsin, Minnesota, Maine, New Hampshire, Florida, Missouri, and Mississippi.

```
17→rand
                17
randInt(1,50)
                34
                28
                42
                49
```

```
                34
                23
                19
                29
                 9
                25
                24
```

(b) Repeating part (a) with a seed of 18, our sample would be Michigan, Massachusetts, Arizona, Minnesota, Maine, Nebraska, Georgia, Iowa, Rhode Island, and Indiana.

12. (a) Answers will vary depending on the technology used (including a table of random digits). Using a TI-84 Plus graphing calculator with a seed of 98 and the labels provided, our sample would be Jefferson, Reagan, Madison, Trump, Pierce, Buchanan, Carter, G. W. Bush.

```
98→rand
                98
randInt(1,45,8)
{3 40 4 45 14 1…
```

```
98→rand
                98
randInt(1,45,8)
…45 14 15 39 43}
```

(b) Repeating part (a) with a seed of 99, our sample would be Nixon, Eisenhower, Pierce, Arthur, Trump, Hayes, Clinton, T. Roosevelt.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

13. **(a)** The list provided by the administration serves as the frame. Number each student in the list of registered students, from 1 to 19,935. Generate 25 random numbers, without repetition, between 1 and 19,935 using a random number generator or table. Select the 25 students with these numbers.

    (b) Answers will vary.

14. **(a)** The list provided by the mayor serves as the frame. Number each resident in the list supplied by the mayor, from 1 to 5832. Generate 20 random numbers, without repetition, between 1 and 5832 using a random number generator or table. Select the 20 residents with these numbers.

    (b) Answers will vary.

15. Answers will vary. Members should be numbered 1–32, though other numbering schemes are possible (e.g. 0–31). Using a table of random digits or a random-number generator, four different numbers (labels) should be selected. The names corresponding to these numbers form the sample.

16. Answers will vary. Employees should be numbered 1–29, though other numbering schemes are possible (e.g. 0–28). Using a table of random digits or a random-number generator, four different numbers (labels) should be selected. The names corresponding to these numbers form the sample.

17. Answers will vary.

Section 1.4

1. Stratified random sampling may be appropriate if the population of interest can be divided into groups (or strata) that are homogeneous and nonoverlapping.

2. Systematic sampling does not require a frame.

3. Convenience samples are typically selected in a nonrandom manner. This means the results are not likely to represent the population. Convenience samples may also be self-selected, which will frequently result in small portions of the population being overrepresented.

4. Cluster sample

5. Stratified sample

6. False. In a systematic random sample, every *k*th individual is selected from the population.

7. False. In many cases, other sampling techniques may provide equivalent or more information about the population with less "cost" than simple random sampling.

8. True. When the clusters are heterogeneous, the heterogeneity of each cluster likely resembles the heterogeneity of the population. In such cases, fewer clusters with more individuals from each cluster are preferred.

9. True. Because the individuals in a convenience sample are not selected using chance, it is likely that the sample is not representative of the population.

10. False. With stratified samples, the number of individuals sampled from each strata should be proportional to the size of the strata in the population.

11. Systematic sampling. The quality-control manager is sampling every 8th chip, starting with the 3rd chip.

12. Cluster sampling. The commission tests all members of the selected teams (clusters).

13. Cluster sampling. The airline surveys all passengers on selected flights (clusters).

14. Stratified sampling. The congresswoman samples some individuals from each of three different income brackets (strata).

15. Simple random sampling. Each known user of the product has the same chance of being included in the sample.

16. Convenience sampling. The radio station is relying on voluntary response to obtain the sample data.

17. Cluster sampling. The farmer samples all trees within the selected subsections (clusters).

18. Stratified sampling. The school official takes a sample of students from each of the five classes (strata).

19. Convenience sampling. The research firm is relying on voluntary response to obtain the sample data.

20. Systematic sampling. The presider is sampling every 5th person attending the lecture, starting with the 3rd person.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

21. Stratified sampling. Shawn takes a sample of measurements during each of the four time intervals (strata).

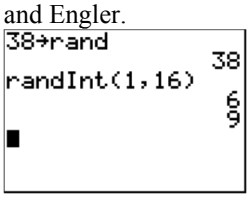
22. Simple random sampling. Each club member has the same chance of being selected for the survey.

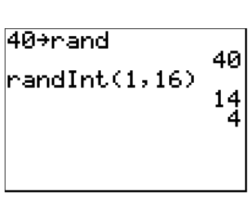
23. The numbers corresponding to the 20 clients selected are 16, $16+25=41$, $41+25=66$, $66+25=91$, $91+25=116$, 141, 166, 191, 216, 241, 266, 291, 316, 341, 366, 391, 416, 441, 466, 491.

24. Since the number of clusters is more than 100, but less than 1000, we assign each cluster a three-digit label between 001 and 795. Starting at row 8, column 38 in Table I of Appendix A, and proceeding downward, the 10 clusters selected are numbered 763, 185, 377, 304, 626, 392, 315, 084, 565, and 508. Note that we discard 822 and 955 in reading the table because we have no clusters with these labels. We also discard the second occurrence of 377 because we cannot select the same cluster twice.

25. Answers will vary. To obtain the sample, number the Democrats 1 to 16 and obtain a simple random sample of size 2. Then number the Republicans 1 to 16 and obtain a simple random sample of size 2. Be sure to use a different starting point in Table I or a different seed for each stratum.

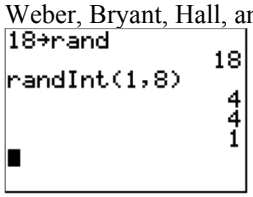
For example, using a TI-84 Plus graphing calculator with a seed of 38 for the Democrats and 40 for the Republicans, the numbers selected would be 6, 9 for the Democrats and 14, 4 for the Republicans. If we had numbered the individuals down each column, the sample would consist of Haydra, Motola, Thompson, and Engler.





26. Answers will vary. To obtain the sample, number the managers 1 to 8 and obtain a simple random sample of size 2. Then number the employees 1 to 21 and obtain a simple random sample of size 4. Be sure to use a different starting point in Table I or a different seed for each stratum.

For example, using a TI-84 Plus graphing calculator with a seed of 18 for the managers and 20 for the employees, the numbers selected would be 4, 1 for the managers and 20, 3, 11, 9 for the employees. If we had numbered the individuals down each column, the sample would consist of Lindsey, Carlisle, Weber, Bryant, Hall, and Gow.



27. **(a)** $\frac{N}{n}=\frac{4502}{50}=90.04$ ® 90; Thus, $k=90$.

(b) Randomly select a number between 1 and 90. Suppose that we select 15. Then the individuals to be surveyed will be the 15th, 105th, 195th, 285th, and so on up to the 4425th employee on the company list.

28. **(a)** $\frac{N}{n}=\frac{945035}{130}=7269.5$ ® 7269; Thus, $k=7269$.

(b) Randomly select a number between 1 and 7269. Suppose that we randomly select 2000. Then we will survey the individuals numbered 2000, 9269, 16,538, and so on up to the individual numbered 939,701.

29. Simple Random Sample:
Number the students from 1 to 1280. Use a table of random digits or a random-number generator to randomly select 128 students to survey.

Stratified Sample:
Since class sizes are similar, we would want to randomly select $\frac{128}{32}=4$ students from each class to be included in the sample.

Cluster Sample:
Since classes are similar in size and makeup, we would want to randomly select $\frac{128}{32}=4$ classes and include all the students from those classes in the sample.

30. No. The clusters were not randomly selected. This would be considered convenience sampling.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

31. Answers will vary. One design would be a stratified random sample, with two strata being commuters and noncommuters, as these two groups each might be fairly homogeneous in their reactions to the proposal.

32. Answers will vary. One design would be a cluster sample, with classes as the clusters. Randomly select clusters and then survey all the students in the selected classes. However, care would need to be taken to make sure that no one was polled twice. Since this would negate some of the ease of cluster sampling, a simple random sample might be the more suitable design.

33. Answers will vary. One design would be a cluster sample, with the clusters being city blocks. Randomly select city blocks and survey every household in the selected blocks.

34. Answers will vary. One appropriate design would be a systematic sample, after doing a random start, clocking the speed of every tenth car, for example.

35. Answers will vary. Since the company already has a list (frame) of 6600 individuals with high cholesterol, a simple random sample would be an appropriate design.

36. Answers will vary. Since a list of all the households in the population exists, a simple random sample is possible. Number the households from 1 to N, then use a table of random digits or a random-number generator to select the sample.

37. **(a)** For a political poll, a good frame would be all registered voters who have voted in the past few elections since they are more likely to vote in upcoming elections.

(b) Because each individual from the frame has the same chance of being selected, there is a possibility that one group may be over- or underrepresented.

(c) By using a stratified sample, the strategist can obtain a simple random sample within each strata (political party) so that the number of individuals in the sample is proportionate to the number of individuals in the population.

38. Random sampling means that the individuals chosen to be in the sample are selected by chance. Random sampling minimizes the chance that one part of the population is over- or underrepresented in the sample. However, it cannot guarantee that the sample will accurately represent the population.

39. Answers will vary.

40. Answers will vary.

Section 1.5

1. A closed question is one in which the respondent must choose from a list of prescribed responses. An open question is one in which the respondent is free to choose his or her own response. Closed questions are easier to analyze, but limit the responses. Open questions allow respondents to state exactly how they feel, but are harder to analyze due to the variety of answers and possible misinterpretation of answers.

2. A certain segment of the population is underrepresented if it is represented in the sample in a lower proportion than its size in the population.

3. **(a)** III. Bias occurs when the results of the sample are not representative of the population.

(b) I. Sampling bias occurs when the techniques used to select individuals for a sample favor one part of the population over another.

(c) IV. Nonresponse bias occurs when the individuals selected to be in the sample who do not respond to the survey have different opinions from those who do respond.

(d) II. Response bias occurs when the answers on a survey do not reflect the true feelings of the respondent.

4. Nonsampling error is the error that results from undercoverage, nonresponse bias, response bias, or data-entry errors. Essentially, it is the error that results from the process of obtaining and recording data. Sampling error is the error that results because a sample is being used to estimate information about a population. Any error that could also occur in a census is considered a nonsampling error.

5. **(a)** Sampling bias. The survey suffers from undercoverage because the first 60 customers are likely not representative of the entire customer population.

(b) Since a complete frame is not possible, systematic random sampling could be used to make the sample more representative of the customer population.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

6. **(a)** Sampling bias. The survey suffers from undercoverage because only homes in the southwest corner have a chance to be interviewed. These homes may have different demographics than those in other parts of the village.

(b) Assuming that households within any given neighborhood have similar household incomes, stratified sampling might be appropriate, with neighborhoods as the strata.

7. **(a)** Response bias. The survey suffers from response bias because the question is poorly worded.

(b) The survey should inform the respondent of the current penalty for selling a gun illegally and the question should be worded as "Do you approve or disapprove of harsher penalties for individuals who sell guns illegally?" The order of "approve" and "disapprove" should be switched from one individual to the next.

8. **(a)** Response bias. The survey suffers from response bias because the wording of the question is ambiguous.

(b) The question might be worded more specifically as "How many hours per night do you sleep, on average?"

9. **(a)** Nonresponse bias. Assuming the survey is written in English, non-English speaking homes will be unable to read the survey. This is likely the reason for the very low response rate.

(b) The survey can be improved by using face-to-face or phone interviews, particularly if the interviewers are multi-lingual.

10. **(a)** Nonresponse bias

(b) The survey can be improved by using face-to-face or phone interviews, or possibly through the use of incentives.

11. **(a)** The survey suffers from sampling bias due to undercoverage and interviewer error. The readers of the magazine may not be representative of all Australian women, and advertisements and images in the magazine could affect the women's view of themselves.

(b) A well-designed sampling plan not in a magazine, such as a cluster sample, could make the sample more representative of the population.

12. **(a)** The survey suffers from sampling bias due to a bad sampling plan (convenience sampling) and possible response bias due to misreported weights on driver's licenses.

(b) The teacher could use cluster sampling or stratified sampling using classes throughout the day. Each student should be weighed to get a current and accurate weight measurement.

13. **(a)** Response bias due to a poorly worded question

(b) The question should be reworded in a more neutral manner. One possible phrasing might be "Do you believe that a marriage can be maintained after an extramarital relation?"

14. **(a)** Sampling bias. The frame is not necessarily representative of all college professors.

(b) To remedy this problem, the publisher could use cluster sampling and obtain a list of faculty from the human resources departments at selected colleges.

15. **(a)** Response bias. Students are unlikely to give honest answers if their teacher is administering the survey.

(b) An impartial party should administer the survey in order to increase the rate of truthful responses.

16. **(a)** Response bias. Residents are unlikely to give honest answers to uniformed police officers if their answer would be seen as negative by the police.

(b) An impartial party should administer the survey in order to increase the rate of truthful responses.

17. No. The survey still suffers from sampling bias due to undercoverage, nonresponse bias, and potentially response bias.

18. The General Social Survey uses random sampling to obtain individuals who take the survey, so the results of their survey are more likely to be representative of the population.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

However, it may suffer from response bias since the survey is conducted by personal interview rather than anonymously on the Internet. The online survey, while potentially obtaining more honest answers, is basically self-selected so may not be representative of the population, particularly if most respondents are clients of the family and wellness center seeking help with health or relationship problems.

19. It is very likely that the order of these two questions will affect the survey results. To alleviate the response bias, either question B could be asked first, or the order of the two questions could be rotated randomly.

20. It is very likely that the order of these two questions will affect the survey results. To alleviate the response bias, the order of the two questions could be rotated randomly. Prohibit is a strong word. People generally do not like to be prohibited from doing things. If the word must be used, it should be offset by the word "allow." The use of the words "prohibit" and "allow" should be rotated within the question.

21. The company is using a reward in the form of the $5.00 payment and an incentive by telling the reader that his or her input will make a difference.

22. The two choices need to be rotated so that any response bias due to the ordering of the questions is minimized.

23. Students should look up the study on the web.

(a) We would expect to find sampling bias due to undercoverage. Individuals who choose not to register to vote may have some characteristics that differ from those who do register.

(b) Yes. Undercoverage also would exist for RDD polls. It is likely the case that the poll is not capturing individuals in a lower socioeconomic class. RBS likely has more of this type of bias because access to cell phones is fairly prevalent today.

(c) RDD had the lower response rate at 6% versus 8% in the RBS survey. This is likely due to the fact that the RBS survey had actual phone numbers to choose from, rather than random digits.

(d) The RDD survey oversampled Republicans 53% to 37%. This could be due to socioeconomic considerations.

24. Today, many individuals no longer have a land-line phone. Therefore, this segment of the population would not be included in any surveys that utilize robocalling. This would be undercoverage. Also, the use of caller ID has likely increased nonresponse bias of phone surveys since individuals may not answer calls from numbers they do not recognize. If individuals with caller ID differ in some way from individuals without caller ID, then phone surveys could also suffer from sampling bias due to undercoverage.

25. It is extremely likely, particularly if households on the do-not-call registry have a trait that is not part of those households that are not on the registry.

26. There is a higher chance that an individual at least 70 years of age will be at home when an interviewer makes contact.

27. Answers will vary. However, the research should include the fact that exit polls tended to undersample non-college-educated whites and oversampled college-educated whites. In this election, non-college educated voters broke for Trump, while college-educated voters were carried by Clinton.

28. – 32. Answers will vary.

33. The *Literary Digest* made an incorrect prediction due to sampling bias (an incorrect frame led to undercoverage) and nonresponse bias (due to the low response rate).

34. Answers will vary. (Gallup incorrectly predicted the outcome of the 1948 election because he quit polling weeks before the election and missed a large number of changing opinions.)

35. **(a)** Answers will vary. Stratified sampling by political affiliation (Democrat, Republican, etc.) could be used to ensure that all affiliations are represented. One question that could be asked is whether or not the person plans to vote in the next election. This would help determine which registered voters are likely to vote.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

(b) Answers will vary. Possible explanations are that presidential election cycles get more news coverage or perhaps people are more interested in voting when they can vote for a president as well as a senator. During non-presidential cycles it is very informative to poll likely registered voters.

(c) Answers will vary. A higher percentage of Democrats in polls versus turnout will lead to overstating the predicted Democrat percentage of Democratic votes.

36. It is difficult for a frame to be completely accurate since populations tend to change over time and there can be a delay in identifying individuals who have joined or left the population.

37. Nonresponse can be addressed by conducting callbacks or offering rewards.

38. Trained, skillful interviewers can illicit responses from individuals and help them give truthful responses.

39. Conducting a presurvey with open questions allows the researchers to use the most popular answers as choices on closed-question surveys.

40. Answers will vary. Phone surveys conducted in the evening may result in reaching more potential respondents; however some of these individuals could be upset by the intrusion.

41. Provided the survey was conducted properly and randomly, a high response rate will provide more representative results. When a survey has a low response rate, only those who are most willing to participate give responses. Their answers may not be representative of the whole population.

42. The order of questions on a survey should be carefully considered, so the responses are not affected by previous questions.

43. The question is ambiguous because it could be interpreted as hours per day, or hours per week, or hours for a particular class. The question could be improved by being more specific, such as, "On average, how many hours do you study each day for your statistics course?"

44. Higher response rates typically suggest that the sample represents the population well. Using rewards can help increase response rates, allowing researchers to better understand the population. There can be disadvantages to offering rewards as incentives. Some people may hurry through the survey, giving superficial answers, just to obtain the reward.

Section 1.6

1. **(a)** An experimental unit is a person, object, or some other well-defined item upon which a treatment is applied.

(b) A treatment is a condition applied to an experimental unit. It can be any combination of the levels of the explanatory variables.

(c) A response variable is a quantitative or qualitative variable that measures a response of interest to the experimenter.

(d) A factor is a variable whose effect on the response variable is of interest to the experimenter. Factors are also called explanatory variables.

(e) A placebo is an innocuous treatment, such as a sugar pill, administered to a subject in a manner indistinguishable from an actual treatment.

(f) Confounding occurs when the effect of two explanatory variables on a response variable cannot be distinguished.

(g) Blinding refers to nondisclosure of the treatment an experimental unit is receiving. There are two types of blinding: single blinding and double blinding.

2. Replication occurs when each treatment is applied to more than one experimental unit.

3. In a single-blind experiment, subjects do not know which treatment they are receiving. In a double-blind experiment, neither the subject nor the researcher(s) in contact with the subjects knows which treatment is received.

4. Completely randomized; matched-pair

5. Blocking

6. True

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

7. **(a)** The research objective of the study was to determine the association between number of times one chews food and food consumption.

(b) The response variable is food consumption; quantitative.

(c) The explanatory variable is chew level (100%, 150%, 200%); qualitative.

(d) The experimental units are the 45 individuals aged 18 to 45 who participated in the study.

(e) Control is used by determining a baseline number of chews before swallowing; same type of food is used in the baseline as in the experiment; same time of day (lunch); age (18 to 45).

(f) Randomization reduces the effect of the order in which the treatments are administered. For example, perhaps the first time through the subjects are more diligent about their chewing than the last time through the study.

8. **(a)** The researchers used an innocuous treatment to account for effects that would result from any treatment being given (i.e. the placebo effect). The placebo is a drug that looks and tastes like topiramate and serves as the baseline against which to compare the results when topiramate is administered.

(b) Being double-blind means that neither the subject nor the researcher in contact with the subjects knows whether the placebo or topiramate is being administered. Using a double-blind procedure is necessary to avoid any intentional or unintentional bias due to knowing which treatment is being given.

(c) The subjects were randomly assigned to the treatment groups (either the placebo or topiramate).

(d) The population is all men and women aged 18 to 65 years diagnosed with alcohol dependence. The sample is the 371 men and women aged 18 to 65 years diagnosed with alcohol dependence who participated in the 14-week trial.

(e) There are two treatments in the study: 300 mg of topiramate or a placebo daily.

(f) The response variable is the percentage of heavy drinking days.

9. **(a)** The response variable is the achievement test scores.

(b) Answers may vary. Some factors are teaching methods, grade level, intelligence, school district, and teacher.
Fixed: grade level, school district, teacher
Set at predetermined levels: teaching method

(c) The treatments are the new teaching method and the traditional method. There are 2 levels of treatment.

(d) The factors that are not controlled are dealt with by random assignment into the two treatment groups.

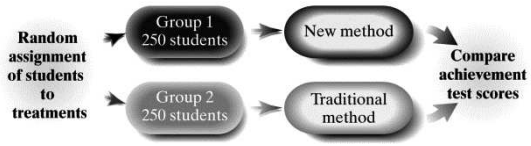
(e) Group 2, using the traditional teaching method, serves as the control group.

(f) This experiment has a completely randomized design.

(g) The subjects are the 500 first-grade students from District 203 recruited for the study.

(h) If students tend to perform worse in classes later in the day (due to being tired or anxious to get out of school), then time of day may be a confounding variable. Suppose group 1 is taught in the morning and group 2 is taught in the afternoon; if group 1 scores better on the achievement test, we won't know whether this is due to the new method of teaching, or due to time of day the course is offered. One solution would be to have a rotating schedule so classes are not always taught at the same time of day.

(e)



Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

10. **(a)** The response variable is the proportion of subjects with a cold.

(b) Answers may vary. Some factors are gender, age, geographic location, overall health, and drug intervention.
Fixed: gender, age, location
Set at predetermined levels: drug intervention

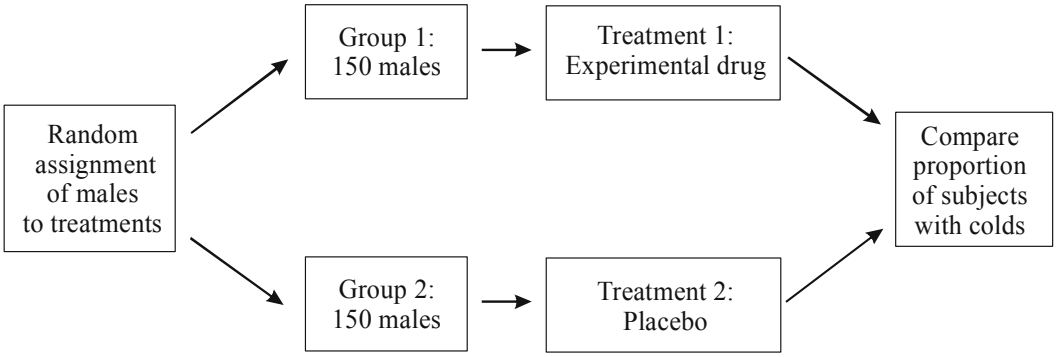
(c) The treatments are the experimental drug and the placebo. There are 2 levels of treatment.

(d) The factors that are not controlled are dealt with by random assignment into the two groups.

(e) This experiment has a completely randomized design.

(f) The subjects are the 300 adult males aged 25 to 29 who have the common cold.

(g)



11. **(a)** This experiment has a matched-pairs design.

(b) The response variable is the level of whiteness.

(c) The explanatory variable or factor is the whitening method. The treatments are Crest Whitestrips Premium in addition to brushing and flossing, and just brushing and flossing alone.

(d) Answers will vary. One other possible factor is diet. Certain foods and tobacco products are more likely to stain teeth. This could impact the level of whiteness.

(e) Answers will vary. One possibility is that using twins helps control for genetic factors such as weak teeth that may affect the results of the study.

12. **(a)** This experiment has a matched-pairs design.

(b) The response variable is the difference in test scores.

(c) The treatment is the mathematics course.

13. **(a)** This experiment has a completely randomized design.

(b) The response variable is the travel time it takes to travel 9.75 meters. It is a quantitative variable.

(c) Priming is the treatment and it is set at two levels—scrambled sentence task using words associated with old age or words not associated with old age.

(d) The subjects are the 30 male and female undergraduates.

(e) The undergraduates did not know to which group they were assigned, and the individual assigning the students did not know to which group the student was assigned.

(f) The conclusion was that the elderly priming condition subjects had a travel time significantly higher than that of the neutral priming condition.

14. **(a)** This experiment has a completely randomized design.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

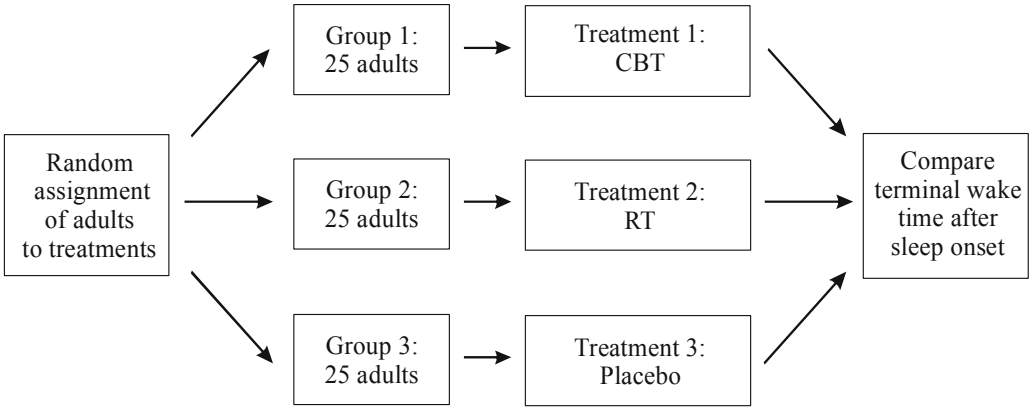
(b) The population being studied is adults with insomnia.

(c) The response variable is the terminal wake time after sleep onset (WASO).

(d) The explanatory variable or factor is the type of intervention. The treatments are cognitive behavioral therapy (CBT), muscle relaxation training (RT), and the placebo.

(e) The experimental units are the 75 adults with insomnia.

(f)



15. **(a)** This experiment has a completely randomized design.

(b) The population being studied is adults over 60 years old and in good health.

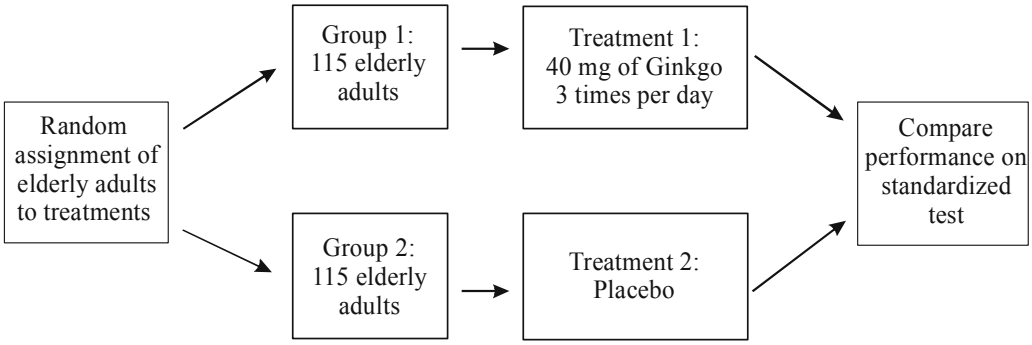
(c) The response variable is the standardized test of learning and memory.

(d) The factor set to predetermined levels (explanatory variable) is the drug. The treatments are 40 milligrams of ginkgo 3 times per day and the matching placebo.

(e) The experimental units are the 98 men and 132 women over 60 years old and in good health.

(f) The control group is the placebo group.

(g)



16. **(a)** This experiment has a completely randomized design.

(b) The population being studied is obese patients.

(c) The response variable is the volume of the stomach. This is a quantitative variable.

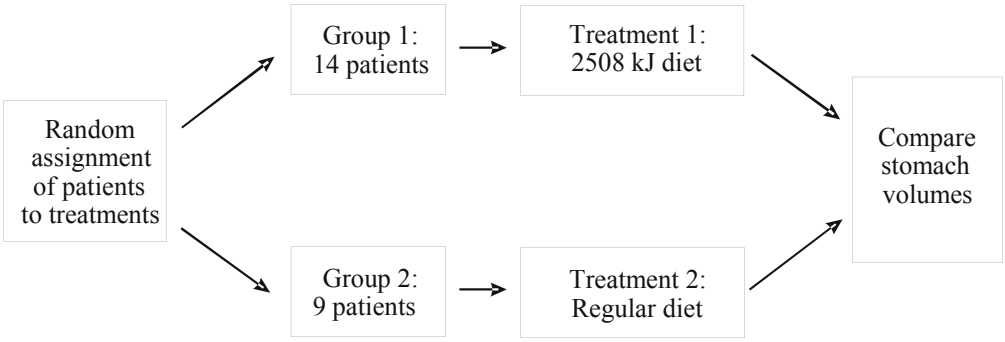
(d) The treatments are the 2508 kJ diet versus the regular diet.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

(e) The experimental units are the 23 obese patients.

(f)



17. **(a)** This experiment has a matched-pairs design.

(b) The population being studied is females with hair loss 20 to 45 years of age.

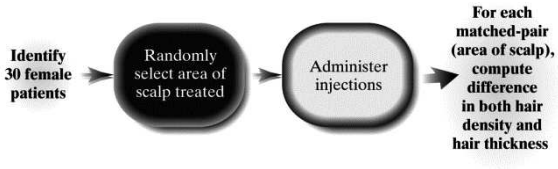
(c) The response variables are hair density and hair diameter.

(d) The treatment is the injection and it is set at two levels: platelet-rich plasma (PRP) injection or saline injection.

(e) The experimental units are the 30 female patients.

(f) Randomization was used to choose the area of the scalp that receives the treatment.

(g)



18. **(a)** This experiment has a matched-pairs design.

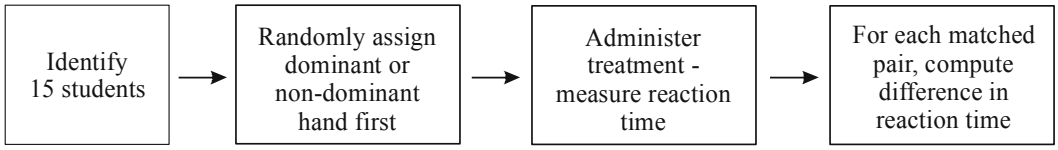
(b) The response variable is the distance the yardstick falls.

(c) The explanatory variable or factor is hand dominance. The treatment is dominant versus non-dominant hand.

(d) The experimental units are the 15 students.

(e) Professor Neil used a coin flip to eliminate bias due to starting on the dominant or non-dominant hand first on each trial.

(f)



19. **(a)** This experiment has a randomized block design.

(b) The response variable is the score on the recall exam.

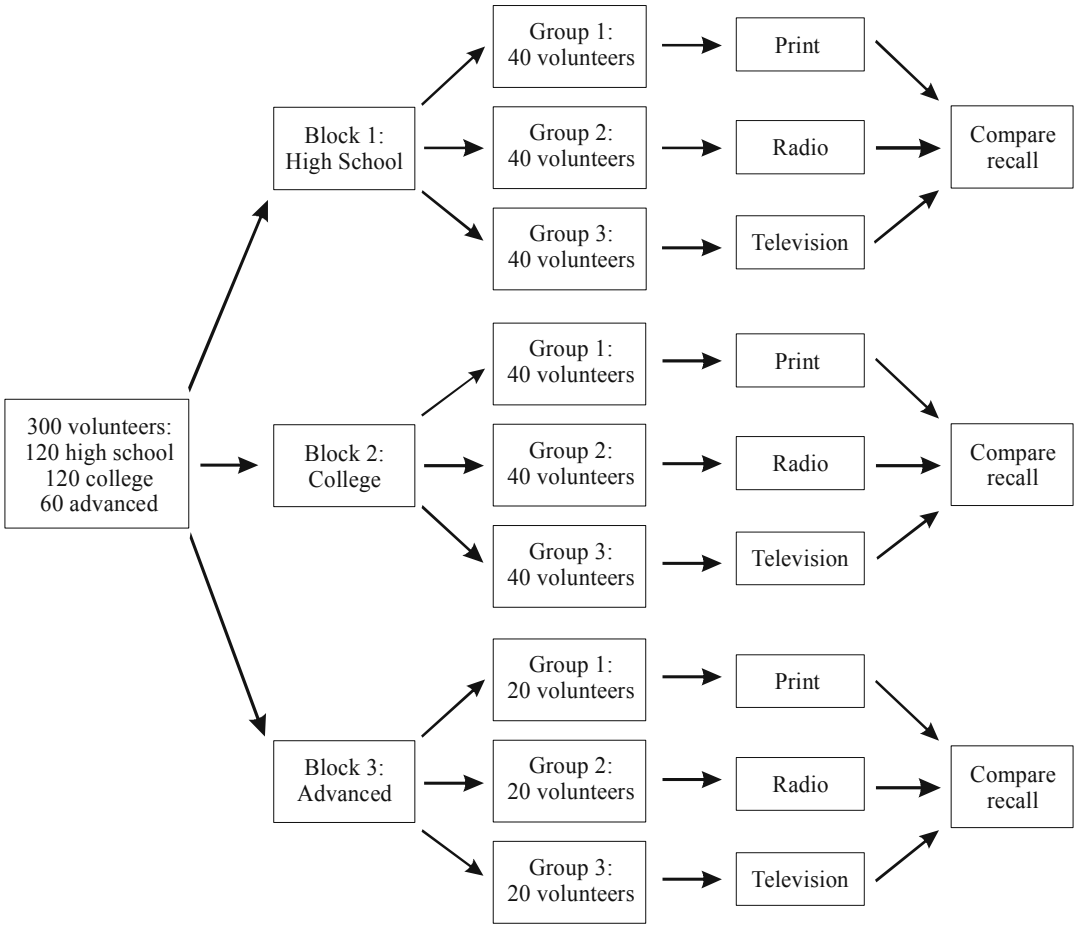
(c) The explanatory variable or factor is the type of advertising. The treatments are print, radio, and television.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

(d) Level of education is the variable that serves as the block.

(e)



20. **(a)** This experiment has a randomized block design.

(b) The response variable is the total number of truancies.

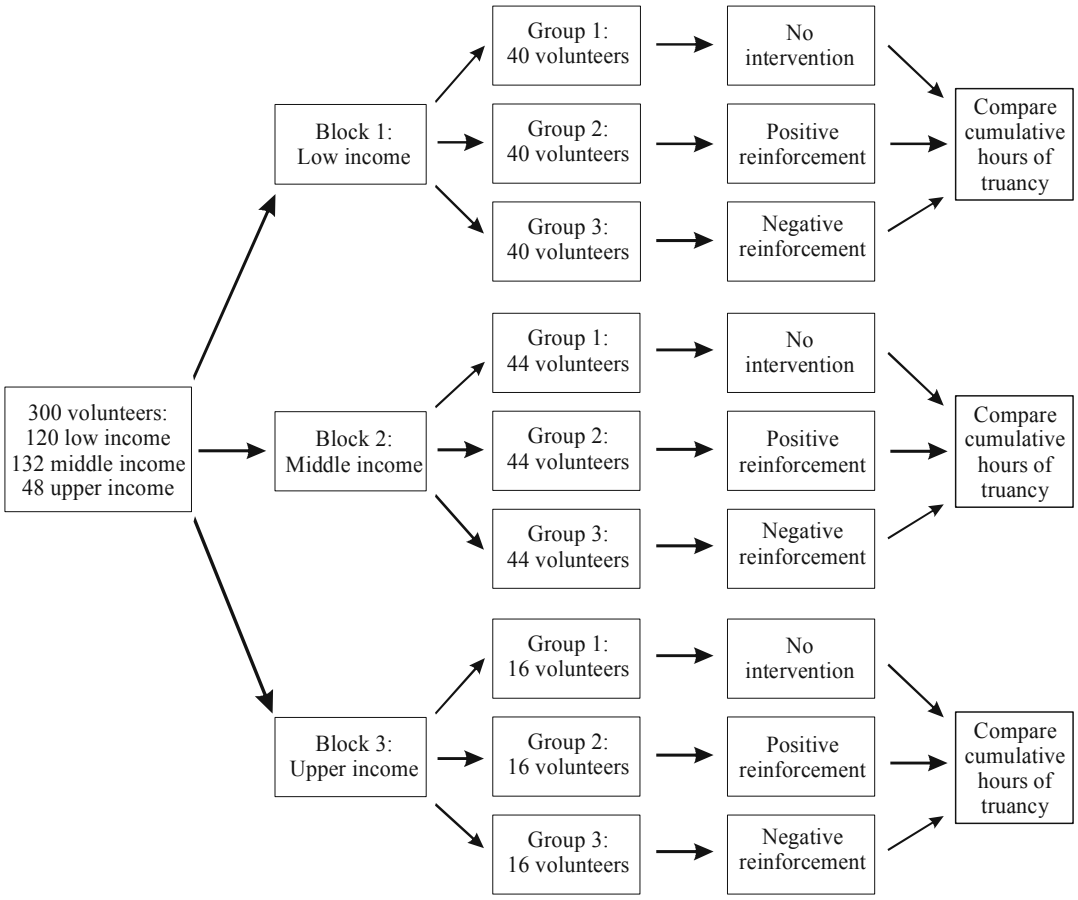
(c) The explanatory variable or factor is the type of intervention. The treatments are no intervention, positive reinforcement, and negative reinforcement.

(d) Income is the variable that serves as the block.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

(e)



21. Answers will vary. Using a TI-84 Plus graphing calculator with a seed of 195, we would pick the volunteers numbered 8, 19, 10, 12, 13, 6, 17, 1, 4, and 7 to go into the experimental group. The rest would go into the control group. If the volunteers were numbered in the order listed, the experimental group would consist of Ann, Kevin, Christina, Eddie, Shannon, Randy, Tom, Wanda, Kim, and Colleen.

22. **(a)** This experiment has a completely randomized design.

(b) Answers will vary. Using a TI-84 Plus graphing calculator with a seed of 223, we would pick the volunteers numbered 6, 18, 13, 3, 19, 14, 8, 1, 17, and 5 to go into group 1.

23. **(a)** This is an observational study because there is no intent to manipulate an explanatory variable or factor. The explanatory variable or factor is whether the individual is a green tea drinker or not, which is qualitative.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

(b) Some lurking variables include diet, exercise, genetics, age, gender, and socioeconomic status.

(c) The experiment is a completely randomized design.

(d) To make this a double-blind experiment, we would need the placebo to look, taste, and smell like green tea. Subjects would not know which treatment is being delivered. In addition, the individuals administering the treatment and measuring the changes in LDL cholesterol would not know the treatment either.

(e) The treatment is the tea, which is set at three levels.

(f) Answers will vary. Other factors you might want to control in this experiment include age, exercise, and diet of the participants.

(g) Randomization could be used by numbering the subjects from 1 to 120. Randomly select 40 subjects and assign them to the placebo group. Then randomly select 40 from the remaining 80 subjects and assign to the one cup of green tea group. The remaining subjects will be assigned to the two cups of green tea group. By randomly assigning the subjects to the treatments, the expectation is that uncontrolled variables (such as genetic history, diet, exercise, etc.) are neutralized (even out).

(h) Exercise is a confounding variable because any change in the LDL cholesterol cannot be attributed to the tea. It may be the exercise that caused the change in LDL cholesterol.

24. **(a)** The research objective is to determine if alerting shoppers about the healthiness of energy-dense snack foods changes the shopping habits of overweight individuals.

(b) The subjects were 42 overweight shoppers.

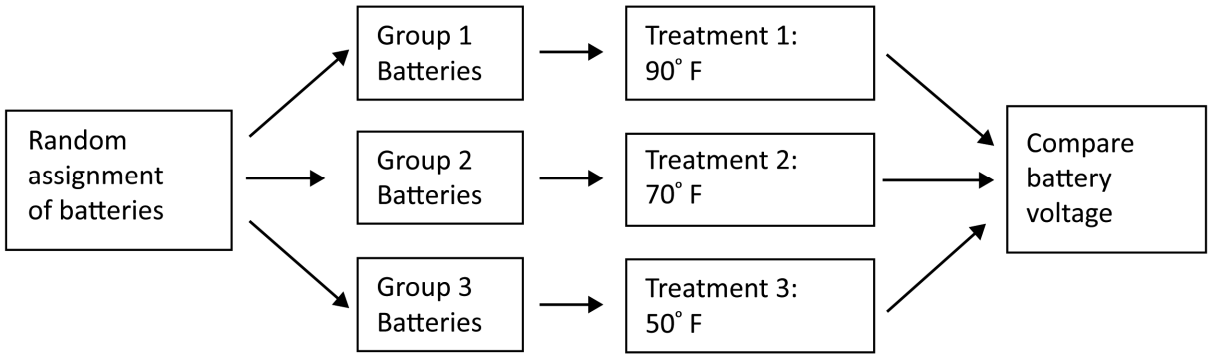
(c) Blinding is not possible because health information is visible.

(d) The explanatory variable is health information or not.

(e) The number of unhealthy snacks purchased is quantitative.

(f) The researchers would not be able to distinguish whether it was the priming or the weight status that played a role in purchase decisions.

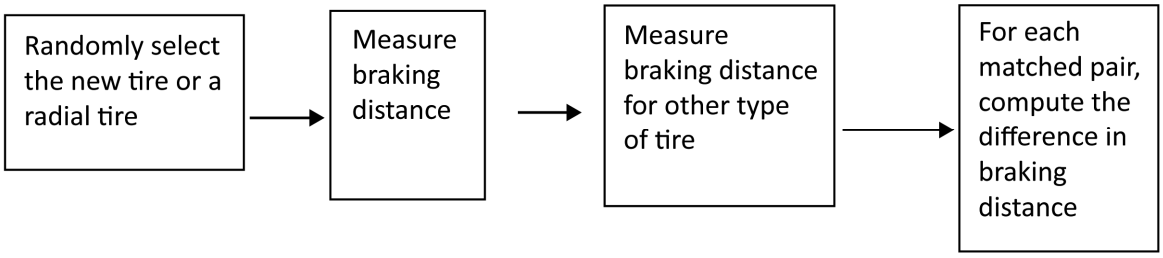
25. Answers will vary. A completely randomized design is probably best.



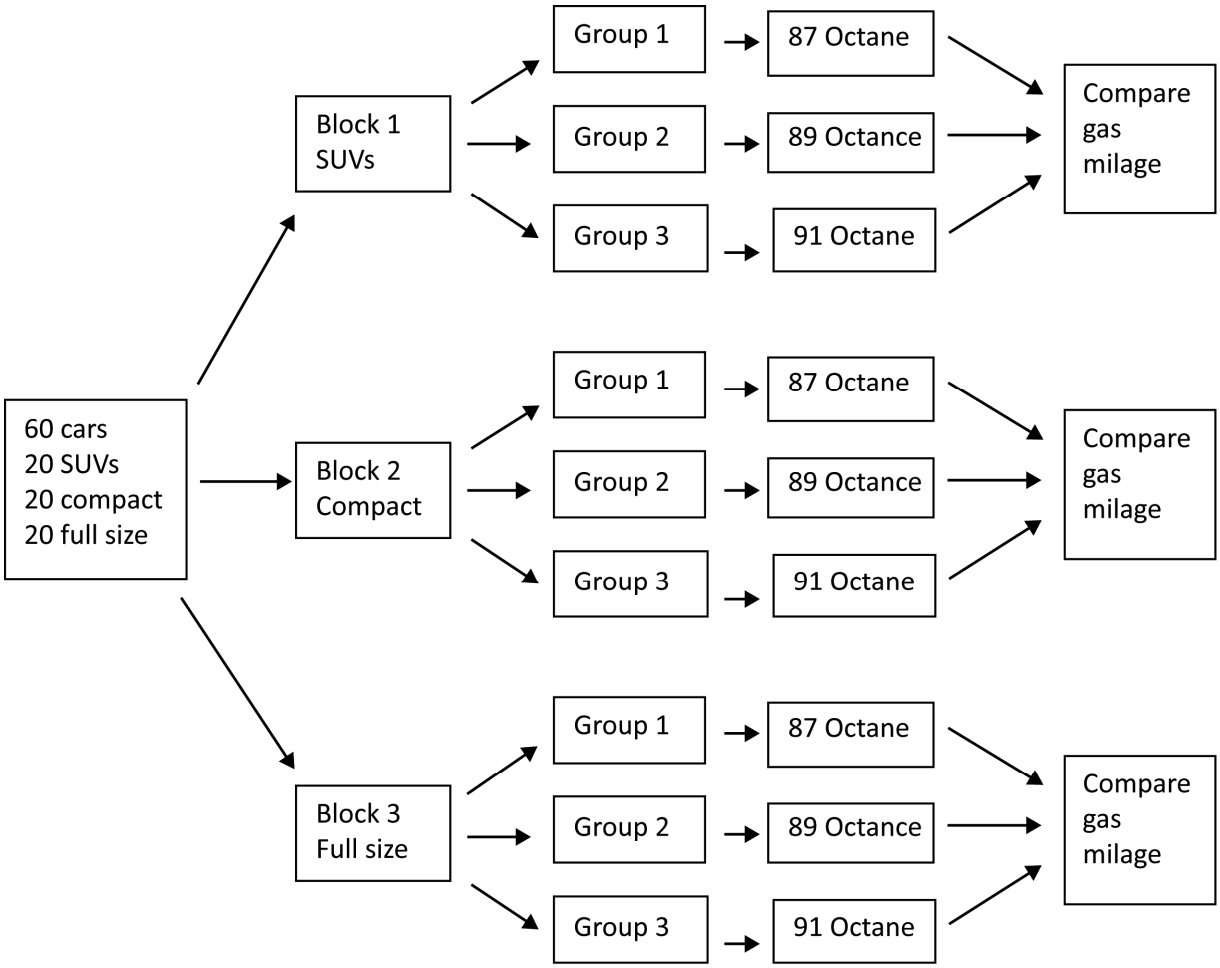
Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

26. Answers will vary. A matched-pairs design matched by car model is likely the best.

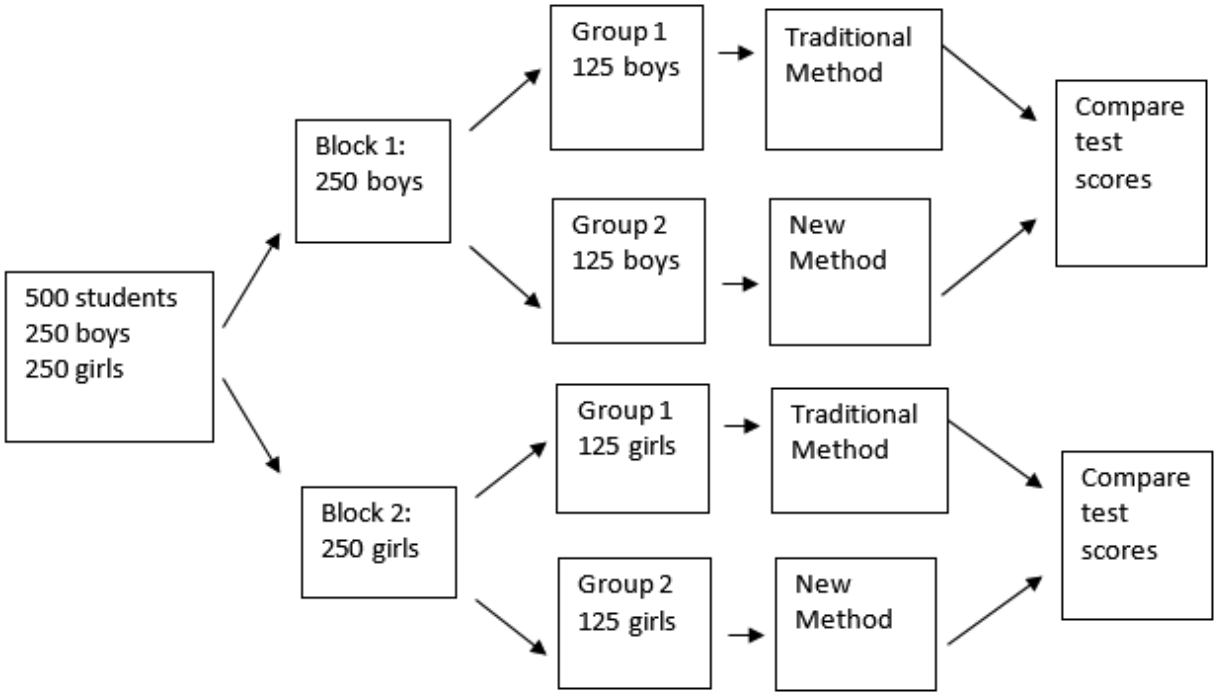


27. Answers will vary. A randomized block design blocked by type of car is likely best.



Copyright © 2021 Pearson Education, Inc.

28. Answers will vary. A randomized block design blocked by gender is likely the best.



29. **(a)** The response variable is blood pressure.

(b) Three factors that have been identified are daily consumption of salt, daily consumption of fruits and vegetables, and the body's ability to process salt.

(c) The daily consumption of salt and the daily consumption of fruits and vegetables can be controlled. The body's ability to process salt cannot be controlled. To deal with variability of the body's ability to process salt, randomize experimental units to each treatment group.

(d) Answers will vary. Three levels of treatment might be a good choice – one level below the recommended daily allowance, one equal to the recommended daily allowance, and one above the recommended daily allowance.

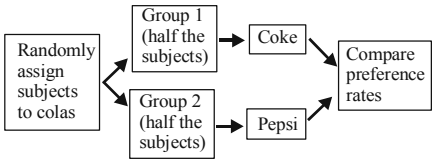
30. Answers will vary.

31. Answers will vary.

32. Answers will vary for the design preference.

<u>Completely Randomized Design</u>

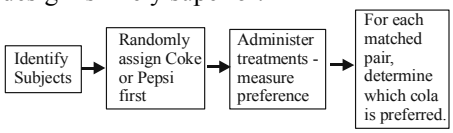
The researcher would randomly assign each subject to either drink Coke or Pepsi. The response variable would be whether the subject likes the soda or not. Preference rates would be compared at the end of the experiment. The subject would be blinded, but the researcher would not. Therefore, this would be a single-blind experiment.



Copyright © 2021 Pearson Education, Inc.

Matched-Pairs Design

The researcher would randomly determine whether each subject drinks Coke first or Pepsi first. To avoid confounding, subjects should eat something bland between drinks to remove any residual taste. The response variable would be either the proportion of subjects who prefer Coke or the proportion of subjects who prefer Pepsi. This would also be a single-blind experiment since the subject would not know which drink was first but the researcher would. The matched-pairs design is likely superior.



33. Answers will vary. Control groups are needed in a designed experiment to serve as a baseline against which other treatments can be compared.

34. **(a)** Answers will vary.

(b) Answers will vary.

35. Answers will vary.

36. In a randomized block design, experimental units are divided into homogeneous groups called *blocks* before being randomly assigned to a treatment within each block. In a stratified random sample, the population is subdivided into homogeneous groups called *strata* before a simple random sample is drawn from each strata. The purpose of blocking is to remove any variability in the response variable that may be attributable to the block.

37. The purpose of randomization is to minimize the effect of factors whose levels cannot be controlled. (Answers will vary.) One way to assign the experimental units to the three groups is to write the numbers 1, 2, and 3 on identical pieces of paper and to draw them out of a "hat" at random for each experimental unit.

38. The six treatments are:
Saturated-fat with drug
Saturated-fat with placebo
Mediterranean with drug
Mediterranean with placebo
NCEP with drug
NCEP with placebo

Chapter 1 Review Exercises

1. **(a)** The response variable is the variable of interest in the study.

(b) A variable is a characteristic of an individual.

(c) A qualitative variable is a variable that allows for classification of individuals based on an attribute or characteristic.

(d) A quantitative variable is a variable that provides numerical measures of individuals. The values of quantitative variables can be added or subtracted and provide meaningful results.

(e) An observational study measures the value of the response variable without attempting to influence the value of the response or explanatory variables.

(f) A study is a designed experiment if a researcher randomly assigns the individuals in a study to groups, intentionally manipulates the value of an explanatory variable and controls other explanatory variables at fixed values, and then records the value of the response variable for each individual.

(g) Confounding occurs when the effects of two or more explanatory variables are not separated. Therefore, any relation that may exist between an explanatory variable and response variable may be due to some other variable or variables not accounted for in the study.

(h) A lurking variable is an explanatory variable that was not considered in the study, but that affects the value of the response variable in the study.

2. The three major types of observational studies are (1) cross-sectional studies, (2) case-control studies, and (3) cohort studies.

Cross-sectional studies collect data at a specific point in time or over a short period of time. Cohort studies are prospective and collect data over a period of time, sometimes over a long period of time. Case-controlled studies are retrospective, looking back in time to collect data either from historical records or from recollection by subjects in the study. Individuals possessing a certain characteristic are matched with those that do not.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

3. The process of statistics refers to the approach used to collect, organize, analyze, and interpret data. The steps are to
(1) identify the research objective,
(2) collect the data needed to answer the research question,
(3) describe the data, and
(4) perform inference.

4. The three types of bias are sampling bias, nonresponse bias, and response bias. Sampling bias occurs when the techniques used to select individuals to be in the sample favor one part of the population over another. Bias in sampling is reduced when a random process is to select the sample. Nonresponse bias occurs when the individuals selected to be in the sample that do not respond to the survey have different opinions from those that do respond. This can be minimized by using callbacks and follow-up visits to increase the response rate. Response bias occurs when the answers on a survey do not reflect the true feelings of the respondent. This can be minimized by using trained interviewers, using carefully worded questions, and rotating question and answer selections.

5. Nonsampling errors are errors that result from undercoverage, nonresponse bias, response bias, and data-entry errors. These errors can occur even in a census. Sampling errors are errors that result from the use of a sample to estimate information about a population. These include random error and errors due to poor sampling plans, and result because samples contain incomplete information regarding a population.

6. The following are steps in conducting an experiment:

(1) *Identify the problem to be solved.*
Give direction and indicates the variables of interest (referred to as the claim).

(2) *Determine the factors that affect the response variable.*
List all variables that may affect the response, both controllable and uncontrollable.

(3) *Determine the number of experimental units.*
Determine the sample size. Use as many as time and money allow.

(4) *Determine the level of each factor.*
Factors can be controlled by fixing their level (e.g. only using men) or setting them at predetermined levels (e.g. different dosages of a new medicine). For factors that cannot be controlled, random assignment of units to treatments helps average out the effects of the uncontrolled factor over all treatments.

(5) *Conduct the experiment.*
Carry out the experiment using an equal number of units for each treatment. Collect and organize the data produced.

(6) *Test the claim.*
Analyze the collected data and draw conclusions.

7. "Number of new automobiles sold at a dealership on a given day" is quantitative because its values are numerical measures on which addition and subtraction can be performed with meaningful results. The variable is discrete because its values result from a count.

8. "Weight in carats of an uncut diamond" is quantitative because its values are numerical measures on which addition and subtraction can be performed with meaningful results. The variable is continuous because its values result from a measurement rather than a count.

9. "Brand name of a pair of running shoes" is qualitative because its values serve only to classify individuals based on a certain characteristic.

10. 73% is a statistic because it describes a sample (the 1011 people age 50 or older who were surveyed).

11. 70% is a parameter because it describes a population (all the passes completed by Cardale Jones in the 2015 Championship Game).

12. Birth year has the *interval* level of measurement since differences between values have meaning, but it lacks a true zero.

13. Marital status has the *nominal* level of measurement since its values merely categorize individuals based on a certain characteristic.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

14. Stock rating has the *ordinal* level of measurement because its values can be placed in rank order, but differences between values have no meaning.

15. Number of siblings has the *ratio* level of measurement because differences between values have meaning and there is a true zero.

16. This is an observational study because no attempt was made to influence the variable of interest. Sexual innuendos and curse words were merely observed.

17. This is an experiment because the researcher intentionally imposed treatments (experimental drug vs. placebo) on individuals in a controlled setting.

18. This was a cohort study because participants were identified to be included in the study and then followed over a period of time with data being collected at regular intervals (every 2 years).

19. This is convenience sampling since the pollster simply asked the first 50 individuals she encountered.

20. This is a cluster sample since the ISP included all the households in the 15 randomly selected city blocks.

21. This is a stratified sample since individuals were randomly selected from each of the three grades.

22. This is a systematic sample since every 40th tractor trailer was tested using a random start with the 12th tractor trailer.

23. **(a)** Sampling bias; undercoverage or nonrepresentative sample due to a poor sampling frame. Cluster sampling or stratified sampling are better alternatives.

(b) Response bias due to interviewer error. A multilingual interviewer could reduce the bias.

(c) Data-entry error due to the incorrect entries. Entries should be checked by a second reader.

24. Answers will vary. Using a TI-84 Plus graphing calculator with a seed of 1990, and numbering the individuals from 1 to 21, we would select individuals numbered 14, 6, 10, 17, and 11. If we numbered the businesses down each column, the businesses selected would be Jiffy Lube, Nancy's Flowers, Norm's Jewelry, Risky Business Security, and Solus, Maria, DDS.

25. Answers will vary. The first step is to select a random starting point among the first 9 bolts produced. Using row 9, column 17 from Table I in Appendix A, he will sample the 3rd bolt produced, then every 9th bolt after that until a sample size of 32 is obtained. In this case, he would sample bolts 3, 12, 21, 30, and so on, until bolt 282.

26. Answers will vary. The goggles could be numbered 00 to 99, then a table of random digits could be used to select the numbers of the goggles to be inspected. Starting with row 12, column 1 of Table 1 in Appendix A and reading down, the selected labels would be 55, 96, 38, 85, 10, 67, 23, 39, 45, 57, 82, 90, and 76.

27. **(a)** To determine the ability of chewing gum to remove stains from teeth

(b) This is an experimental design because the teeth were separated into groups that were assigned different treatments.

(c) Completely randomized design

(d) Percentage of stain removed

(e) Type of stain remover (gum or saliva); Qualitative

(f) The 64 stained bovine incisors

(g) The chewing simulator could impact the percentage of the stain removed.

(h) Gum A and B remove significantly more stain.

28. **(a)** Matched-pairs

(b) Reaction time; Quantitative

(c) Alcohol consumption

(d) Food consumption; caffeine intake

(e) Weight, gender, etc.

(f) To act as a placebo to control for the psychosomatic effects of alcohol

(g) Alcohol delays the reaction time significantly in seniors for low levels of alcohol consumption; healthy seniors that are not regular drinkers.

29. **(a)** This experiment has a randomized block design.

Copyright © 2021 Pearson Education, Inc.

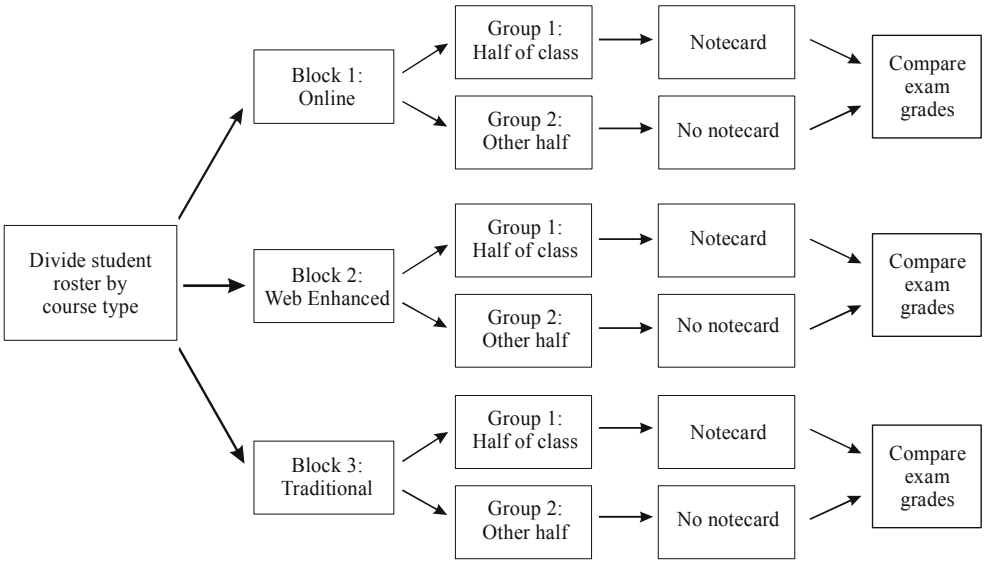
Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

(b) The response variable is the exam grade.

(c) The factor "Notecard use" is set at predetermined levels. The treatments are "with notecard" and "without notecard."

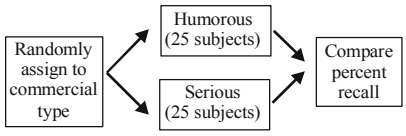
(d) The experimental units are the instructor's statistics students.

(e)

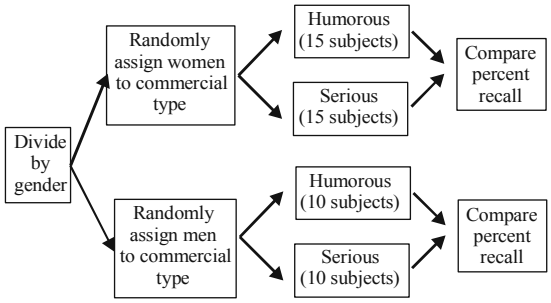


30. Answers will vary. Since there are ten digits (0 – 9), we will let a 0 or 1 indicate that (a) is to be the correct answer, 2 or 3 indicate that (b) is to be the correct answer, and so on. Beginning with row 1, column 8 of Table 1 in Appendix A, and reading downward, we obtain the following:
2, 6, 1, 4, 1, 4, 2, 9, 4, 3, 9, 0, 6, 4, 4, 8, 6, 5, 8, 5
Therefore, the sequence of correct answers would be:
b, d, a, c, a, c, b, e, c, b, e, a, d, c, c, e, d, c, e, c

31. **(a)** Answers will vary. One possible diagram is shown below.



(b) Answers will vary. One possible diagram is shown below.



32. A matched-pairs design is an experimental design where experimental units are matched up so they are related in some way.

In a completely randomized design, the experimental units are randomly assigned to one of the treatments. The value of the response variable is compared for each treatment. In a matched-pairs design, experimental units are matched up on the basis of some common characteristic (such as husband-wife or twins). The differences between the matched units are analyzed.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

33. Answers will vary.

34. Answers will vary.

35. Randomization is meant to even out the effect of those variables that are not controlled for in a designed experiment. Answers to the randomization question may vary; however, each experimental unit must be randomly assigned. For example, a researcher might randomly select 25 experimental units from the 100 units and assign them to treatment #1. Then the researcher could randomly select 25 from the remaining 75 units and assign them to treatment #2, and so on.

Chapter 1 Test

1. Collect information, organize and summarize the information, analyze the information to draw conclusions, provide a measure of confidence in the conclusions drawn from the information collected.

2. The process of statistics refers to the approach used to collect, organize, analyze, and interpret data. The steps are to
(1) identify the research objective,
(2) collect the data needed to answer the research question,
(3) describe the data, and
(4) perform inference.

3. The time to complete the 500-meter race in speed skating is quantitative because its values are numerical measurements on which addition and subtraction have meaningful results. The variable is continuous because its values result from a measurement rather than a count. The variable is at the *ratio* level of measurement because differences between values have meaning and there is a true zero.

4. Video game rating is qualitative because its values classify games based on certain characteristics but arithmetic operations have no meaningful results. The variable is at the *ordinal* level of measurement because its values can be placed in rank order, but differences between values have no meaning.

5. The number of surface imperfections is quantitative because its values are numerical measurements on which addition and subtraction have meaningful results. The variable is discrete because its values result from a count. The variable is at the *ratio* level of measurement because differences between values have meaning and there is a true zero.

6. This is an experiment because the researcher intentionally imposed treatments (brand-name battery versus plain-label battery) on individuals (cameras) in a controlled setting. The response variable is the battery life.

7. This is an observational study because no attempt was made to influence the variable of interest. Fan opinions about the asterisk were merely observed. The response variable is whether or not an asterisk should be placed on Barry Bonds' 756th homerun ball.

8. A *cross-sectional study* collects data at a specific point in time or over a short period of time; a *cohort study* collects data over a period of time, sometimes over a long period of time (prospective); a *case-controlled study* is retrospective, looking back in time to collect data.

9. An experiment involves the researcher actively imposing treatments on experimental units in order to observe any difference between the treatments in terms of effect on the response variable. In an observational study, the researcher observes the individuals in the study without attempting to influence the response variable in any way. Only an experiment will allow a researcher to establish causality.

10. A control group is necessary for a baseline comparison. This accounts for the placebo effect that says that some individuals will respond to any treatment. Comparing other treatments to the control group allows the researcher to identify which, if any, of the other treatments are superior to the current treatment (or no treatment at all). Blinding is important to eliminate bias due to the individual or experimenter knowing which treatment is being applied.

11. The steps in conducting an experiment are to
(1) identify the problem to be solved,
(2) determine the factors that affect the response variable, (3) determine the number of experimental units, (4) determine the level of each factor, (5) conduct the experiment, and
(6) test the claim.

12. Answers will vary. The franchise locations could be numbered 01 to 15 going across. Starting at row 7, column 14 of Table I in Appendix, and working downward, the selected numbers would be 08, 11, 03, and 02. The corresponding locations would be Ballwin, Chesterfield, Fenton, and O'Fallon.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

13. Answers will vary. Using the available lists, obtain a simple random sample from each stratum and combine the results to form the stratified sample. Start at different points in Table I or use different seeds in a random number generator. Using a TI-84 Plus graphing calculator with a seed of 14 for Democrats, 28 for Republicans, and 42 for Independents, the selected numbers would be
Democrats: 3946, 8856, 1398, 5130, 5531, 1703, 1090, and 6369
Republicans: 7271, 8014, 2575, 1150, 1888, 3138, and 2008
Independents: 945, 2855, and 1401

14. Answers will vary. Number the blocks from 1 to 2500 and obtain a simple random sample of size 10. The blocks corresponding to these numbers represent the blocks analyzed. All trees in the selected blocks are included in the sample. Using a TI-84 Plus graphing calculator with a seed of 12, the selected blocks would be numbered 2367, 678, 1761, 1577, 601, 48, 2402, 1158, 1317, and 440.

15. Answers will vary. $\frac{600}{14} \approx 42.86$, so we let $k = 42$. Select a random number between 1 and 42 that represents the first slot machine inspected. Using a TI-84 Plus graphing calculator with a seed of 132, we select machine 18 as the first machine inspected. Starting with machine 18, every 42nd machine thereafter would also be inspected (60, 102, 144, 186, …, 564).

16. In a completely randomized design, the experimental units are randomly assigned to one of the treatments. The value of the response variable is compared for each treatment. In a randomized block design, the experimental units are first divided according to some common characteristic (such as gender). Then each experimental unit within each block is randomly assigned to one treatment. Within each block, the value of the response variable is compared for each treatment, but not between blocks. By blocking, we prevent the effect of the blocked variable from confounding with the treatment.

17. **(a)** Sampling bias due to voluntary response

(b) Nonresponse bias due to the low response rate

(c) Response bias due to poorly worded questions.

(d) Sampling bias due to poor sampling plan (undercoverage)

18. **(a)** This experiment has a matched-pairs design.

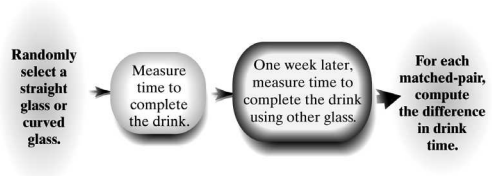
(b) The subjects are the 159 social drinkers who participated in the study.

(c) Treatments are the types of beer glasses (straight glass or curved glass).

(d) The response variable is the time to complete the drink; quantitative.

(e) The type of glass used in the first week is randomly determined. This is to neutralize the effect of drinking out of a specific glass first.

(f)



19. **(a)** This experiment has a completely randomized design.

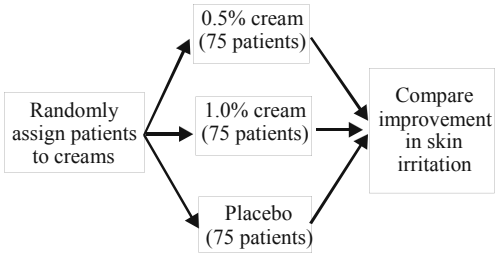
(b) The factor set to predetermined levels is the topical cream concentration. The treatments are 0.5% cream, 1.0% cream, and a placebo (0% cream).

(c) The study is double-blind if neither the subjects, nor the person administering the treatments, are aware of which topical cream is being applied.

(d) The control group is the placebo (0% topical cream).

(e) The experimental units are the 225 patients with skin irritations.

(f)



Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

20. **(a)** This was a cohort study because participants were identified to be included in the study and then followed over a long period of time with data being collected at regular intervals (every 4 years).

(b) The response variable is bone mineral density. The explanatory variable is weekly cola consumption.

(c) The response variable is quantitative because its values are numerical measures on which addition and subtraction can be performed with meaningful results.

(d) The researchers observed values of variables that could potentially impact bone mineral density (besides cola consumption), so their effect could be isolated from the variable of interest.

(e) Answers will vary. Some possible lurking variables that should be accounted for are smoking status, alcohol consumption, physical activity, and calcium intake (form and quantity).

(f) The study concluded that women who consumed at least one cola per day (on average) had a bone mineral density that was significantly lower at the femoral neck than those who consumed less than one cola per day. The study cannot claim that increased cola consumption *causes* lower bone mineral density because it is only an observational study. The researchers can only say that increased cola consumption is *associated* with lower bone mineral density for women.

21. A confounding variable is an explanatory variable that cannot be separated from another explanatory variable. A lurking variable is an explanatory variable that was not considered in the study but affects the response variable in the study.

Case Study: Chrysalises for Cash

Reports will vary. The reports should include the following components:

Step 1: *Identify the problem to be solved*. The entrepreneur wants to determine if there are differences in the quality and emergence time of broods of the black swallowtail butterfly depending on the following factors: (a) early brood season versus late brood season; (b) carrot plants versus parsley plants; and (c) liquid fertilizer versus solid fertilizer.

Step 2: *Determine the explanatory variables that affect the response variable*. Some explanatory variables that may affect the quality and emergence time of broods are the brood season, the type of plant on which the chrysalis grows, fertilizer used for plants, soil mixture, weather, and the level of sun exposure.

Step 3: *Determine the number of experimental units*. In this experiment, a sample of 40 caterpillars/butterflies will be used.

Step 4: *Determine the level of the explanatory variables*:

- **Brood season** – We wish to determine the differences in the number of deformed butterflies and in the emergence times depending on whether the brood is from the early season or the late season. We use a total of 20 caterpillars/butterflies from the early brood season and 20 caterpillars/butterflies from the late brood season.
- **Type of plant** – We wish to determine the differences in the number of deformed butterflies and in the emergence times depending on the type of plant on which the caterpillars are placed. A total of 20 caterpillars are placed on carrot plants and 20 are placed on parsley plants.
- **Fertilizer** – We wish to determine the differences in the number of deformed butterflies and in the emergence times depending on the type of fertilizer used on the plants. A total of 20 chrysalises grow on plants that are fed liquid fertilizer and 20 grow on plants that are fed solid fertilizer.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

- **Soil mixture** – We control the effects of soil by growing all plants in the same mixture.
- **Weather** – We cannot control the weather, but the weather will be the same for each chrysalis grown within the same season. For chrysalises grown in different seasons, we expect the weather might be different and thus part of the reason for potential differences between seasons. Also, we can control the amount of watering that is done.
- **Sunlight exposure** – We cannot control this variable, but the sunlight exposure will be the same for each chrysalis grown within the same season. For chrysalises grown in different seasons, we expect the sunlight exposure might be different and thus part of the reason for potential differences between seasons.

Step 5: *Conduct the experiment.*

(a) We fill eight identical pots with equal amounts of the same soil mixture. We use four of the pots for the early brood season and four of the pots for the late brood season.

For the early brood season, two of the pots grow carrot plants and two grow parsley plants. One carrot plant is fertilized with a liquid fertilizer, one carrot plant is fertilized with a solid fertilizer, one parsley plant is fertilized with the liquid fertilizer, and one parsley plant is fertilized with the solid fertilizer. We place five black swallowtail caterpillars of similar age into each of the four pots.

Similarly, for the late brood season, two of the pots grow carrot plants and two grow parsley plants. One carrot plant is fertilized with a liquid fertilizer, one carrot plant is fertilized with a solid fertilizer, one parsley plant is fertilized with the liquid fertilizer, and one parsley plant is fertilized with the solid fertilizer. We place five black swallowtail caterpillars of similar age into each of the four pots.

(b) We determine the number of deformed butterflies and in the emergence times for the caterpillars/butterflies from each pot.

Step 6: *Test the claim*. We determine whether any differences exist depending on season, plant type, and fertilizer type.

Conclusions:

Early versus late brood season: From the data presented, more deformed butterflies occur in the late season than in the early season. Five deformed butterflies occurred in the late season, while only one occurred in the early season. Also, the emergence time seems to be longer in the early season than in the late season. In the early season, all but one of the 20 emergence times were between 6 and 8 days. In the late season, all 20 of the emergence times were between 2 and 5 days.

Parsley versus carrot plants: From the data presented, the plant type does not seem to affect the number of deformed butterflies that occur. Altogether, three deformed butterflies occur from parsley plants and three deformed butterflies occur from the carrot plants. Likewise, the plant type does not seem to affect the emergence times of the butterflies.

Liquid versus solid fertilizer: From the data presented, the type of fertilizer seems to affect the number of deformed butterflies that occur. Five deformed butterflies occurred when the solid fertilizer was used, while only one occurred when the liquid fertilizer was used. The type of fertilizer does not seem to affect emergence times.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Chapter 2
Organizing and Summarizing Data

Section 2.1

1. Raw data are the data as originally collected, before they have been organized or coded.

2. Number (or count); proportion (or percent)

3. The relative frequencies should add to 1, although rounding may cause the answers to vary slightly.

4. A bar graph is used to illustrate qualitative data. It is a chart in which rectangles are used to illustrate the frequency or relative frequency with which a category appears. A Pareto chart is a bar chart with bars drawn in order of decreasing frequency or relative frequency.

5. **(a)** The most common response was "Boss."

(b) 15% of 1053 is 158.

(c) The graphic cannot be displayed as a pie chart because the percentages do not add up to 100 percent. That is, there is no "whole."

6. **(a)** The largest segment in the pie chart is for "Washing your hands" so the most commonly used approach to beat the flu bug is washing your hands. 61% of respondents selected this as their primary method for beating the flu.

(b) The smallest segment in the pie chart is for "Drinking Orange Juice" so the least used method is drinking orange juice. 2% of respondents selected this as their primary method for beating the flu.

(c) 25% of respondents felt that flu shots were the best way to beat the flu.

7. **(a)** The highest bar corresponds to the position OF (outfield), so OF is the position with the most MVPs.

(b) The bar for first base (1B) reaches the line for 15. Thus, there were 15 MVPs who played first base.

(c) The bar for first base (1B) is 15 on the vertical axis. The bar for third base (3B) reaches 10. 15 – 10 = 5, so there were 5 more MVPs who played first base than third base.

(d) Each of the three outfield positions should be reported as MVPs, rather than treating the three positions as one position.

8. **(a)** 25,561,000 whites were living in poverty.

(b) $$\frac{11{,}190}{25{,}561+9132+11{,}190+1985} \approx 0.234 = 23.4\%$$

In 2017, about 23.4% of the impoverished in the United States were Hispanic.

(c) This graph should use relative frequencies, rather than frequencies. The graph does not account for the different population size of each ethnic group. Without knowing the population sizes, we cannot determine whether a group is disproportionally impoverished.

9. **(a)** 69% of the respondents believe divorce is morally acceptable.

(b) 23% believe divorce is morally wrong. So, 240 million * 0.23 = 55.2 million adult Americans believe divorce is morally wrong.

(c) This statement is inferential, since it is a generalization based on the observed data.

10. **(a)** 5% of identity theft was loan fraud.

(b) 26% of the identity fraud cases in a recent year involved credit card fraud. So, 10 million * 0.26 = 2.6 million cases of credit card fraud occurred in a recent year.

11. **(a)** The proportion of 18–34 year old respondents who are more likely to buy when made in America is 0.42. For 34–44 year olds, the proportion is 0.61.

Copyright © 2021 Pearson Education, Inc.

INSTRUCTOR'S GUIDE FOR STUDENT ACTIVITY WORKBOOK

HEATHER FOES

Rock Valley College

KATHLEEN ALMY

Rock Valley College

MICHAEL SULLIVAN, III

Joliet Junior College

SULLIVAN STATISTICS SERIES

SIXTH EDITION

Michael Sullivan, III

Joliet Junior College

Pearson

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

This work is protected by United States copyright laws and is provided solely for the use of instructors in teaching their courses and assessing student learning. Dissemination or sale of any part of this work (including on the World Wide Web) will destroy the integrity of the work and is not permitted. The work and materials from it should never be made available to students except by instructors using the accompanying text in their classes. All recipients of this work are expected to abide by these restrictions and to honor the intended pedagogical purposes and the needs of other instructors who rely on these materials.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Reproduced by Pearson from electronic files supplied by the author.

Copyright © 2021, 2017, 2013 by Pearson Education, Inc. 221 River Street, Hoboken, NJ 07030. All rights reserved.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

Pearson

ISBN-13: 978-0-13-582062-9
ISBN-10: 0-13-582062-6

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Table of Contents

Activity Title (section number)	Page Number
Chapter 1:	
Building a Survey in StatCrunch	2
Introducing Statistics Through Quotes (1.1)	3
Categorizing Student Survey Data (1.1)	4
Comparing Sampling Methods (1.4)	5
Designing an Experiment (1.6)	6
Chapter 2:	
Exploring Histograms with StatCrunch (2.2)	9
Sorting Histograms by Shape (2.2)	10
Predicting Distribution Shape (2.2)	11
Constructing and Comparing Graphical Representations (2.2-2.3)	12
Recognizing and Correcting Misleading Graphs (2.4)	13
Chapter 3:	
Understanding Measures of Center (3.1)	15
Comparing Statistics to Parameters (3.1-3.2)	16
Exploring Standard Deviation (3.2)	17
Understanding the Standard Deviation Formula (3.2)	18
Scaling and Shifting Data (3.4)	20
Matching Boxplots and Histograms (3.5)	21
Chapter 4:	
Exploring Properties of the Linear Correlation Coefficient (4.1)	23
Finding a Least-Squares Regression Line (4.2)	24
Examining the Relationship Between Arm Length and Height (4.2)	25
Minimizing the Sum of the Squared Residuals (4.2)	27
What Does It Mean to "Regress to the Mean"? (4.2)	28
Understanding Influential Observations (4.3)	31
Investigating the Coefficient of Determination (4.3)	32
Chapter 5:	
Demonstrating the Law of Large Numbers (5.1)	35
Finding the Probability of Getting Heads (5.1)	36
Interpreting Conditional Probabilities (5.4)	37
Let's Make a Deal (5.4)	38
Calculating the Probability of Winning the Lottery (5.5)	39
Exploring the Duplicate Birthday Problem (5.6)	40
Tennis Anyone? (5.6)	41

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Chapter 6:

Finding the Expected Value of a Game (6.1) 43
Exploring a Binomial Distribution from Multiple Perspectives (6.2) 44
Using Binomial Probabilities in Baseball (6.2) 45
Designing a Game (6.2) 46

Chapter 7:

Constructing Probability Distributions Involving Dice (7.1) 48
Fooled by Randomness (7.1) 50
Home Run Distances (7.2) 51
Grading on a Curve (7.2) 53
Modeling with the Normal Distribution (7.2) 55
Analyzing Standardized Test Scores (7.2) 56
Playing Plinko (7.4) 57

Chapter 8:

Creating a Sampling Distribution for the Mean (8.1) 59
Analyzing the Variability in Sample Means (8.1) 60
Describing the Distribution of the Sample Mean:
Approximately Normal Population (8.1) 61
Simulating IQ Scores (8.1) 63
Sampling from Normal and Non-Normal Populations (8.1) 64
Describing the Distribution of the Sample Mean: Non-Normal Data (8.1) 65
Creating a Sampling Distribution for a Proportion (8.2) 67
Describing the Distribution of the Sample Proportion (8.2) 68
Describing the Distribution of the Sample Proportion Using the
Urn Applet (8.2) 69

Chapter 9:

Exploring the Effects of Confidence Level, Sample Size, and Shape I (9.1) 72
Constructing a Confidence Interval with M&M's (9.1) 73
Constructing a Confidence Interval for Average Temperature (9.2) 74
Exploring the Effects of Confidence Level, Sample Size, and Shape II (9.2) 75
Constructing a Confidence Interval from a Non-Normal Distribution (9.2) 76
Constructing a Confidence Interval for Die Rolls (9.2) 77
The Role of Sample Size in the Construction of Confidence Intervals (9.2) 78
Take Me Out to the Ball Game (9.2) 80
Finding a Bootstrap Confidence Interval (9.5) 81

Chapter 10:

The Logic of Hypothesis Testing (10.1) 83
Sensitivity and Specificity (10.1) 84
Interpreting *P*-Values (10.2) 86
Testing a Claim with Skittles I (10.2) 87
Understanding Type I Error Rates I (10.2) 88
Testing Cola Preferences (10.2) 89

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Testing a Claim with Beads (10.2) 90
Analyzing a Research Article I (10.2) 91
Testing a Claim with Skittles II (10.3) 92
Understanding Type I Error Rates II (10.3) 93
Using Bootstrapping to Test a Claim (10.5) 94
Computing the Power of a Test (10.6) 97

Chapter 11:
Making an Inference about Two Proportions (11.1) 99
Analyzing Rates of Drug Side Effects (11.1) 100
Predicting Basketball Game Outcomes (11.1) 101
Considering the Effects of Grammar (11.1) 103
Analyzing Kissing Data (11.1) 104
Analyzing a Research Article II (11.2) 106
Comparing Arm Span and Height (11.2) 107
Comparing Rental Car Prices (11.2) 108
Using Randomization Test for Independent Means (11.3) 109
Comparing Bull and Bear Markets (11.3) 111
Differentiating Between Practical and Statistical Significance (11.3) 113

Chapter 12:
Performing a Goodness-of-Fit Test (12.1) 115
Testing for Homogeneity of Proportions (12.2) 116

Chapter 13:
Designing a Randomized Complete Block Design (13.3) 119
Performing a Two-Way ANOVA (13.4) 120

Chapter 14:
Testing the Significance of a Regression Model (14.1) 112
Using a Randomization Test for Correlation (14.1) 123

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

The subject of statistics, like many things in life, is best learned through active participation. Your students will better understand statistical concepts and be more able to apply them if they are actively engaged during class. These activities give you numerous options for teaching, reinforcing, and applying statistical understanding throughout a typical first course in statistics.

Teaching statistics presents the familiar challenge of covering too much content in too little time. It is often difficult to find time to do activities with a statistics class and still teach all the material you need to teach. Consider ways that you can incorporate activities into your current approach to teaching statistics without necessarily using more time. Perhaps you can start an activity in class and have students complete it on their own. Perhaps you can replace a lecture with an activity that teaches the same concepts. Perhaps you can do an activity as a demonstration to reduce the time commitment. The activities are designed to be flexible and support a variety of classroom uses.

If you are doing activities that utilize applets, either go to www.statcrunch.com, or www.pearsonhighered.com/sullivanstats to access the applet.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Chapter 1

Data Collection

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Building a Survey Using StatCrunch

In this activity, you will build a survey using StatCrunch. As you progress through topics during the course of the semester, you will develop a report summarizing the results of your survey.

1. Write survey questions that allow you to answer a research question. The survey should include demographic information (gender, age, political affiliation, and so on). The survey should also include questions that you wish to have answered. For ideas on the types of questions you might ask, consult the local newspaper, polling organizations (such as Harris Interactive, Pew Research, and Gallup) for questions that they have asked. This will allow you to compare the results of your survey (and subgroups of individuals within your survey) to the polling organizations. It is important that your survey include both quantitative and qualitative data. Have fun with the types of questions you ask. This is your survey and your project, so be creative.

2. Watch the video "StatCrunch Survey" available at www.sullystats.com/statistics-videos or at https://youtu.be/hf4ZTLMX5D8.

3. Build your survey in StatCrunch and administer it. You should attempt to secure at least 200 responses to your survey, so be sure to utilize social media outlets (Facebook and Twitter) along with other forms of electronic communication (websites and e-mail) to spread the word about your survey. Ask the individuals who take your survey to share the link with others. This will help to develop a more representative sample (than just your family and friends).

4. Summarize the results of your survey as you develop more skills. For example, when you are in Chapter 2, create some nice visualizations of the data. In Chapter 3, determine some summary statistics. Are there any outliers? What are the explanations, if any, for the outliers? Maybe draw side-by-side boxplots by some qualitative variable (such as gender or political affiliation). In Chapter 4, determine if there are any associations between quantitative variables. Qualitative variables?

5. Once you get into Part IV of the text—inferential statistics—conduct some inference based on the results of the survey. Do some estimation of parameters (i.e. confidence intervals). Be sure to verify the model requirements are satisfied. Compare two segments of the sample (Republicans versus Democrats; High School Graduates versus College Graduates). The ability to do these types of comparisons hinges on the type of demographic information you collected in the original survey. So, be sure your survey has this type of data in it.

6. State any final conclusions you have found based on your survey results and data analysis.

Copyright © 2021 Pearson Education, Inc.

Introducing Statistics Through Quotes

Overview:

There are numerous popular quotes about statistics, many of which are already familiar to students. A discussion of famous statistical quotes at the beginning of the semester provides a great opportunity to explain what the subject of statistics is all about and introduce students to some of the terminology used in the subject.

Textbook: Section 1.1

Time: 10 minutes

Notes:

This activity works well on the first day of class since it introduces students to the subject and does not require any prior knowledge. You can have students work in pairs or groups to fill in the quotes that are familiar to them first and then use logic and a process of elimination to complete the rest. If students are unable to finish in the time you would like to allot in class, they can be encouraged to Google the ones they have not finished.

At some point in this activity, refer your students to the definition of statistics (section 1.1):

"Statistics is the science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions. In addition, statistics is about providing a measure of confidence in any conclusions."

There are numerous possible extensions and variations to this activity. Here are a few possibilities:

- Students can be asked to find a statistical quote that was not included in this activity.
- If you have an online discussion board for your course, students can post a quote and other students can determine who said it.
- This can be used as an ice-breaker activity on the first day. You can type the quotes and missing words individually on slips of paper and have students work the room to find the word that completes their quote.
- Students can choose one of these quotes at the end of the semester and write an essay on how it relates to the course.

Answers:

1. uncertainty
2. opinion
3. assumptions
4. numbers
5. measure
6. approximate
7. languages
8. certain
9. make
10. suggestive
11. averages
12. statistic
13. statistically
14. probability
15. lottery
16. accuracy
17. truth
18. illumination
19. conclusions
20. experiment

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Categorizing Student Survey Data

Overview:

This activity gives students an opportunity to apply the terminology from Section 1.1. It works well as an ice-breaker activity, but also provides lots of data for future use in the course.

Textbook: Section 1.1

Time: 10 minutes to collect data, at least 20 minutes to classify data and discuss

Notes:

Give students a time limit for collecting the data before they begin so they will be focused.

Tell students that classifying the data is not black and white and to expect some disagreement over levels of measurement and even discrete versus continuous. The data classification will be challenging for them! This is a great time to discuss the fact that statistics is subjective and they should not expect the answers to be as well-defined as they were in algebra. Encourage them to justify their choices in the chart and let them know that there are certainly boxes with more than one reasonable answer. After students have completed the classification chart, you will want to have a whole-class discussion and have students share and defend their choices.

There are numerous possible uses for this data later in the semester. You might consider loading it into StatCrunch so your students have access to the data later.

Possible uses for the data later in the semester:

- Ask students to represent one of the variables with several different types of graphs.
- Ask students to choose the most appropriate type of graphical display (in their opinion) for each variable.
- Ask students to describe the center and spread of one particular variable.
- Treat the data as a simple random sample and have students test a claim such as college students get an average of 8 hours of sleep per night.
- Treat the data as a simple random sample and have students find a confidence interval for the number of hours of TV college students typically watch each night.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Comparing Sampling Methods

Overview:

This activity gives students a chance to implement some of the sampling methods from section 1.4 in the text. You might consider discussing the different sampling methods the day before this activity or have the students read them in the book in preparation for this activity.

Textbook: Section 1.4

Time: 50 minutes

Notes:

You can give students some input in the question used for this activity. Consider having them look on the Gallup website for homework and find a recently asked question or do this together as a class. It might be interesting for them to compare their class results to the results posted on the Gallup website.

A random number generator will be needed for this activity.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Designing an Experiment

Overview:

The purpose of this activity is to allow students to see why good experiments need control, randomization, and replication. In addition, students recognize how much thought needs to go into an experiment, even for something simple like paper airplanes.

Textbook: Section 1.6

Materials: Newspaper, brown-bag paper, 24-pound printer paper
Paper airplane designs

Time: See note below

Notes:

This activity presents a huge task for students. Although the steps are outlined in the activity, there are still many choices for students to make. You will want to provide enough guidance for the students to be successful while still allowing them to struggle with the enormity of the task and the number of decisions to be made.

You will need to allow parts of several different class periods for this activity. In the first period, you can introduce the activity and guide students through the first four steps. Then they will need time to make the planes. Another period might be used to discuss Step 5 before carrying out the experiment in a third period.

This activity could be used as a semester-long project that is revisited and extended as additional techniques are developed. Students could be assigned groups to complete the activity out of class for Chapter 1. When descriptive statistics are developed, students could revisit the problem, summarize, and visualize the experimental results to assist in analyzing them. When inferential statistics are developed, students could return to the problem and test the difference between the results.

Answers:

Step 1: Research question: Will the paper airplanes fly farthest if they are made of newspaper, brown-bag paper, or printer paper?
Response variable: Length of the flight

Step 2:
1. Who makes the planes (You may want to fix this at a single level. If you have more than one individual making the planes, you will have a randomized block design. We do not care about the impact the person has on the flight distance, but we do want to eliminate variability attributed to the person making the planes.)
2. Who flies the planes
3. Paper choice
4. Wind (fixed by doing experiment inside)

Step 3: Each student could make one from each type of paper, resulting in 3 times as many planes as the number of students in the class. Alternatively, each student could make 3 planes of each type to have more replication.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Step 4: Answers will vary.

Step 5: Answers will vary. Students should consider randomization in this part of the experiment. For example, if the students always fly the printer-paper plane first, the results may be misleading due to the "learning curve" in flying the planes.

Step 6: Answers will vary.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Chapter 2

Organizing and Summarizing Data

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Exploring Histograms with StatCrunch

Overview:

Since histograms will be used frequently in the rest of the course, it is important that students understand how to interpret histograms and what they reveal about the shape of the distribution of a data set. Students also need to realize that the choices they make about class limits and class width will affect the look of the histogram and the impression it gives about the data. Since making multiple histograms by hand is time consuming, technology provides a convenient and efficient way to explore these effects.

Textbook: Section 2.2

Time: 25 minutes

Notes:

This activity requires that students have a data set to work with in StatCrunch. If you did the Student Data Survey in Chapter one and saved the data on StatCrunch, then have the students use the data file for this activity. If you did not use the Student Data Activity or do not have the results on StatCrunch, then you can provide another data set or have students use one that is already available on StatCrunch. If your students have not used StatCrunch before, it would help if you can provide a demonstration for the class to make it easier for students to find the data set and the applet needed for this activity.

This activity should be done after you have taught students how to organize data into a frequency distribution and have at least introduced histograms in Section 2.2.

If you want to use this activity as a group assignment, you could assign a particular variable to each group of four students so that they can compare and discuss their histograms as they make them. Alternatively, you could assign different variables to different students to reduce students' ability to get information from each other and ensure they are each doing their own work.

After students have completed this activity, have a discussion to make sure they realize that there is no single "correct" histogram, but there are some histograms that do a very poor job of representing a given variable.

Answers:

3a. lower class limit

4a. class width

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Sorting Histograms by Shape

Overview:

In this activity, students will sort 12 histograms into categories based on their shapes.

Textbook: Section 2.2

Time: 10 – 15 minutes

Notes:

This activity is best done with students working in pairs or groups so they can discuss the shapes of the graphs.

Encourage students to create categories other than the ones suggested if they think it is warranted. Students will often come up with creative names that make more sense to them than the formal category names. Let students know that sorting histograms is subjective and not everyone will agree. It is important that they can defend their categories and choices, however.

Students are likely to have difficulty categorizing graph #8 with the Height data. Some students might describe it as "twin –peaked" or notice the up-down-up pattern. Other students might think it falls into one of the other existing categories. Ask students why they think the Height histogram has this shape and see if they consider that there is a peak for the most common male height and another for the most common female height.

You might consider copying the histograms and cutting them apart so that a stack of individual histograms could be given to a group of students to sort physically.

Possible answers:

Category Name	Category Description	List of Graphs Included
Bell-shaped	The highest frequency occurs in the middle, and the frequencies fall off to the left and right. The histogram is fairly symmetric.	IQ Scores (#2) Sum of Two Dice (#9) ACT Scores (#12)
Uniform	The frequency of each value of the variable is fairly evenly spread out across the values of the variable.	Random Numbers (#1) Die Roll (#4)
Skewed Right	The tail to the right of the peak is longer than the tail to the left of the peak.	Wait Time (#3) Income ($1000s) (#7) Student Ages (#10) Number of Children (#11)
Skewed Left	The tail to the left of the peak is longer than the tail to the right of the peak.	Weight (lbs) (#5) Grades (#6)
Bimodal	There are two different peaks in the frequencies.	Height (in.) (#8)

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Predicting Distribution Shape

Overview:

This activity gives students a chance to predict the shape of a histogram for data generated from their class. After collecting the data and making histograms, they compare their predicted shape to the actual shape of the distribution.

Textbook: Section 2.2

Time: 50 minutes

Notes:

This activity can be started in class and finished individually by students after class. Have students work in pairs or groups to predict the shape of the various distributions so they can discuss their expectations. Then collect the necessary data from them. There are many ways this can be done. You can pass a chart around the class as students are predicting the shapes and then post the data online for students. You could also poll the class verbally and write the data on the board.

After students have made their predictions and the data has been collected, let them finish the activity on their own. They can draw histograms by hand or use StatCrunch or another technology to generate them.

Possible answers:

1. a. Uniform; We would expect the random integers to be generated with fairly even frequencies.

   b. Bell-shaped; We would expect there to be some pulse rate in the middle that occurs with the greatest frequency and very low or high pulse rates to occur with much smaller frequency.

   c. Skewed Right; There might be many families with only a couple of members, but some families could be quite large, making a long right tail on the histogram.

   d. Depends on the particular class; In many college classes, the age distribution is skewed right since most students are 18 – 22 years old and there are often a few much older students. Some classes might have a fairly bell-shaped or even uniform age distribution.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Constructing and Comparing Graphical Representations

Overview:

In addition to learning how to construct various statistical graphs, students need to learn when and why to make a particular graph. This activity will give them an opportunity to choose the graphs that they think best display a particular data set. They will need to consider advantages and disadvantages to various graphs.

Textbook: Sections 2.2 and 2.3

Time: 15 – 20 minutes

Notes:

This activity can be introduced at the end of a class period. The students can make the graphs on their own for homework, and then the discussion can take place during the following class period.

Assign the same variable to at least four students so that those students can form a discussion group on the second day.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual

Recognizing and Correcting Misleading Graphs

Overview:

There are many bad graphs printed in newspapers, magazines, and online sources. This activity gives students the opportunity to find a graph that interests them and discuss its flaws. This type of analysis and the resulting discussion can help make students better consumers of news and information.

Be sure to discuss the difference between a graph that unintentionally misleads the reader and one that purposefully deceives.

Textbook: Section 2.4

Time: 15 – 20 minutes

Notes:

Consider introducing this activity by showing students some examples of bad graphs in the media and discussing with them some of the common manipulations that lead to misleading and deceptive graphs.

This activity can be done individually and outside of class, if desired.

You might consider introducing the activity at the end of a class period and have students look for a bad graph as a homework assignment. The next class period, they could form groups and decide in their group which graph to use for the assignment.

Copyright © 2021 Pearson Education, Inc.

Statistics Informed Decisions Using Data 6th edition Sullivan Solutions Manual