Basic Business Statistics 15th edition Berenson Solutions Manual

SKU: 9780138069438-SOLUTIONS

Chapter 1

1.1 (a) The menu items represent a categorical variable. Each menu item represents a separate category.
(b) The variable that contains the prices is a numerical variable.
(c) One menu item and the price of that item would be one instance or occurrence of data.

1.2 (a) Business size represents a categorical variable because each size represents a particular category.
(b) The measurement scale is ordinal, because of the different sizes.

1.3 The variable speed trials is a continuous numerical variable because time can have any value from 0 to any reasonable unit of time.

1.4 (a) The telephone number assigned to the smartphone is a categorical variable.
(b) The data usage for a current month (in GB) is a numerical variable that is continuous because any value within a range of values can occur.
(c) The length (in minutes and seconds) of the last voice call made using the smartphone is a numerical variable that is continuous because time can have any value from 0 to any reasonable unit of time.
(d) The number of apps installed on the smartphone is a numerical variable that is discrete because the outcome is a count.
(e) Whether a device protection plan exists is a categorical variable because the answer can be only yes or no.

1.5 (a) numerical, ratio
(b) numerical, ratio
(c) categorical, nominal
(d) categorical, nominal
(e) numerical, ratio

1.6 (a) numerical, continuous
(b) categorical
(c) numerical, discrete
(d) categorical
(e) categorical

1.7 (a) numerical, ratio scale, continuous
(b) numerical, ratio scale
(c) numerical, ratio scale, discrete

1.8 (a) numerical, continuous
(b) numerical, discrete
(c) numerical, continuous
(d) categorical
(e) categorical

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

1.9 (a) Income may be considered discrete if we "count" our money. It may be considered continuous if we "measure" our money; we are only limited by the way a country's monetary system treats its currency.

(b) The first format would provide more information because it includes a ratio value while the second measure would only include a range of values for each choice category.

1.10 The variable test score would be numerical, and presumably, in the range of 0 through 100. If fractional credit for an answer is possible, the variable would need to be continuous and not discrete.

1.11 (a) The population is "members of the retailer's rewards program from the metropolitan area." A systematic or random sample could be taken of members from the rewards program from the metropolitan area. The director might wish to collect both numerical and categorical data.

(b) Three categorical questions might be occupation, marital status, type of clothing. Numerical questions might be age, average monthly hours shopping for clothing, income.

1.12 (a) 0001
(b) 0040
(c) 0902

1.13 (a) Sample without replacement: Start at row 29. Read from left to right in 3-digit sequences and continue unfinished sequences from end of row to beginning of next row.
Row 29: 124 783 762 299 659 310 658 361 369 889 588 692 957
Rows 29-30: 157
Row 30: 175 555 646 541 142 547 704 570 342 672 937 837
Rows 30-31: 929
Row 31: 161 611 075 801 030 783 159 309 132 762 671 073 000
Row 32: 780 257 353 914 621 390 444 745 003 197 127 874 770
Rows 32-33: 927
Row 33: 587 672 288 014 510 175 128 228 668 765 530 493
Rows 33-34: 251
Row 34: 669 020 427 042 516 447 773 709 739 459 239 668 263
Row 35: 701 835 806 565 489 318 338 209 316 747 103 865 929
Row 35-36: 390
Row 36: 730 353 851 567 999 742 508 667 802 875 573 672
Rows 36-37: 571
Row 37: 093 493 242 134 312 459 002 770 485 820 090 658 595
Row 38: 824 623 016
Note: All sequences above 127 and all repeating sequences are discarded.

(b) Use the same technique as in part (a).
Note: All sequences above 127 are discarded. There were no repeating sequences.

1.14 A simple random sample would be less practical for personal interviews because of travel costs, unless interviewees are paid to attend a central interviewing location.

1.15 This is a probability sample because the selection is based on chance. It is not a simple random sample because A is more likely to be selected than B or C.

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

1.16 Here all members of the population are equally likely to be selected and the sample selection mechanism is based on chance. But selection of two elements is not independent; for example if *A* is in the sample, we know that *B* is also, and that *C* and *D* are not.

1.17 (a) Since a complete roster of registered students exists, a simple random sample of 200 students could be taken. If student satisfaction with the quality of campus life randomly fluctuates across the student body, a systematic 1-in-20 sample could also be taken from the population frame. If student satisfaction with the quality of life may differ by status and by experience/class level, a stratified sample using eight strata, full-time freshmen through full-time seniors and part-time freshmen through part-time seniors, could be selected. If student satisfaction with the quality of life is thought to fluctuate as much within clusters as between them, a cluster sample could be taken.

(b) A simple random sample is one of the simplest to select. The population frame is the registrar's file of 3,000 student names.

(c) A systematic sample is easier to select by hand from the registrar's records than a simple random sample, since an initial person at random is selected and then every 20th person thereafter would be sampled. The systematic sample would have the additional benefit that the alphabetic distribution of sampled students' names would be more comparable to the alphabetic distribution of student names in the campus population.

(d) If rosters by status and class designations are readily available, a stratified sample should be taken. Since student satisfaction with the quality of life may indeed differ by status and class level, the use of a stratified sampling design will not only ensure all strata are represented in the sample, it will also generate a more representative sample and produce estimates of the population parameter that have greater precision.

(e) If all 3,000 registered students reside in one of 10 on-campus residence halls which fully integrate students by status and by class, a cluster sample should be taken. A cluster could be defined as an entire study house, and the students of a single randomly selected study house could be sampled. Since each study house has 300 students, a systematic sample of 150 students can then be selected from the chosen cluster of 300 students. Alternately, a cluster could be defined as a floor of one of the 10 study houses. Suppose there are six floors in each dormitory with 50 students on each floor. Three floors could be randomly sampled to produce the required 150 student sample. Selection of an entire study house may make distribution and collection of the survey easier to accomplish. In contrast, if there is some variable other than status or class that differs across study houses, sampling by floor may produce a more representative sample.

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

1.18 (a)

Row 16: 2323 6737 5131 8888 1718 0654 6832 4647 6510 4877
Row 17: 4579 4269 2615 1308 2455 7830 5550 5852 5514 7182
Row 18: 0989 3205 0514 2256 8514 4642 7567 8896 2977 8822
Row 19: 5438 2745 9891 4991 4523 6847 9276 8646 1628 3554
Row 20: 9475 0899 2337 0892 0048 8033 6945 9826 9403 6858
Row 21: 7029 7341 3553 1403 3340 4205 0823 4144 1048 2949
Row 22: 8515 7479 5432 9792 6575 5760 0408 8112 2507 3742
Row 23: 1110 0023 4012 8607 4697 9664 4894 3928 7072 5815
Row 24: 3687 1507 7530 5925 7143 1738 1688 5625 8533 5041
Row 25: 2391 3483 5763 3081 6090 5169 0546

Note: All sequences above 5000 are discarded. There were no repeating sequences.

(b)

0089 0189 0289 0389 0489 0589 0689 0789 0889 0989
1089 1189 1289 1389 1489 1589 1689 1789 1889 1989
2089 2189 2289 2389 2489 2589 2689 2789 2889 2989
3089 3189 3289 3389 3489 3589 3689 3789 3889 3989
4089 4189 4289 4389 4489 4589 4689 4789 4889 4989

(c) With the single exception of invoice 0989, the invoices selected in the simple random sample are not the same as those selected in the systematic sample. It would be highly unlikely that a random process would select the same units as a systematic process.

1.19 (a) A stratified sample should be taken so that each of the three strata will be proportionately represented.

(b) The number of observations in each of the three strata out of the total of 100 should reflect the proportion of the three categories in the customer database. For example, 500/1000 = 50% so 50% of 100 = 50 customers should be selected from the potential customers; similarly, 300/1000 = 30% so 30 customers should be selected from those who have purchased once, and 200/1000 = 20% so 20 customers from the repeat buyers.

(c) It is not simple random sampling because, unlike the simple random sampling, it ensures proportionate representation across the entire population.

1.20 Before accepting the results of a survey of college students, you might want to know, for example:
Who funded the survey? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What questions were asked? Were the questions clear, accurate, unbiased, valid? What operational definition of *immediately and effortlessly* was used? What was the response rate?

1.21 (a) Possible coverage error: Only employees in a specific division of the company were sampled.

(b) Possible nonresponse error: No attempt is made to contact nonrespondents to urge them to complete the evaluation of job satisfaction.

(c) Possible sampling error: The sample statistics obtained from the sample will not be equal to the parameters of interest in the population.

(d) Possible measurement error: Ambiguous wording in questions asked on the questionnaire.

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

1.22 The results are based on a survey of bank executives. If the frame is supposed to be banking institutions, how is the population defined? There is no information about the response rate, so there is an undefined nonresponse error.

1.23 Before accepting the results of the survey, you might want to know, for example:
Who funded the survey? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What questions were asked? Were they clear, accurate, unbiased, valid? What was the response rate? What was the margin of error? What was the sample size? What frame was used?

1.24 Before accepting the results of the survey, you might want to know, for example:
Who funded the survey? Why was it conducted? What was the population from which the sample was selected? What sampling design was used? What mode of response was used: a personal interview, a telephone interview, or a mail survey? Were interviewers trained? Were survey questions field-tested? What questions were asked? Were the questions clear, accurate, unbiased, valid? What was the response rate? What was the margin of error? What was the sample size? What frame was used?

1.25 Only the second value, 2.7GB, contains units.

1.26 (a) Invalid values include Appel, Samsun, APPLE, Apple iPhone, and mOTOROLA.
(b) Appel should be Apple. Samsun should be Samsung. APPLE should be Apple. Apple iPhone should be Apple. mOTOROLA should be Motorola.

1.27 (a) Employee ID has data integration error. Payment Method should do a domain check. Age should mark missing value and do domain check and should check outliers. Division should do a domain check and look for data integration errors. Department has data integration errors and should check domain.
(b) In Employee ID the first and eleventh value are identical, as well as the sixteenth value "EPM16" should be "EMP16". In Payment Method, corrections to values "Male", "F", and "EMP12" should be corrected as well as identifying the missing ninth value. In Age, "k20" should be "20". Also, "221" should be corrected as well as identifying the missing seventeenth value. In Division, "EAST" should be "East", "N" should be "North", "N" should be "North", and "Nroth" should be "North". In Department "Customer Rel." should be "Customer Relations", "Brand Mgt." should be "Brand Mgt", "Operati0ns" should be "Operations", "Human capital" should be "Human Capital", and "FIN" should be "Finance".

1.28 (a) Fund Number contains the wrong data. Category should do a domain check. 5-Yr Return should format the number as a percentage. 10-Yr Retrun should mark missing value. Net Expense Ratio needs a domain check and should check outliers. Rating should do a domain check and check outliers. Assets needs a domain check and should mark missing value.
(b) The last three values in New Expense Ratio; the eleventh value in Assets.
(c) For both New Expense Ratio and Assets, a maximum value could be defined.

1.29 Data cleaning improves the quality of the data while data wrangling changes the organization of the data.

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

1.30 No, because the categories do not seem to be mutually exclusive. The categories need specific ranges, such as younger than 21, 21 to 23, 35 to 54, and 55 or older.

1.31 Stack the data by Housekeeping Requested values.

1.32 (a) The times for each of the hotels would be arranged in separate columns.
(b) The hotel names would be in one column and the times would be in a second column.

1.33 For data cleaning, 10-year return and Rating should mark missing values. For data wrangling, the data could be stacked by market cap, fund type, YTD return, 1-year return, 10-year return, life of fund, expense ratio, or fund rating.

1.34 A population contains all the items of interest whereas a sample contains only a portion of the items in the population.

1.35 A statistic is a summary measure describing a sample whereas a parameter is a summary measure describing an entire population.

1.36 Categorical random variables yield categorical responses such as yes or no answers. Numerical random variables yield numerical responses such as your height in inches.

1.37 Discrete random variables produce numerical responses that arise from a counting process. Continuous random variables produce numerical responses that arise from a measuring process.

1.38 Both nominal and ordinal variables are categorical variables but no ranking is implied in nominal variable such as male or female while ranking is implied in ordinal variable such as a student's grade of A, B, C, D and F.

1.39 Both interval and ratio variables are numerical variables in which the difference between measurements is meaningful but an interval variable does not involve a true zero such as standardized exam scores while a ratio variable involves a true zero such as height.

1.40 A list of values defines a domain for a categorical variable, whereas a range defines a domain for a numerical variable.

1.41 Items or individuals in a probability sampling are selected based on known probabilities while items or individuals in a nonprobability samplings are selected without knowing their probabilities of selection.

1.42 Missing values are values that were not collected for a variable. Outliers are values that seem excessively different from most of the other values

1.43 In unstacked arrangements, separate numerical variables are created for each group in the data. For example, you might create a variable for the weights of men and a second variable for the weights of women. In stacked arrangements, a single numerical variable is paired with a categorical variable that represents the categories. For example, all weights would be in one variable, with a categorical variable indicating male or female.

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

1.44 Coverage error is error generated due to an improperly or inappropriately framed population which can result in a sample that may not be representative of the population that one wishes to study. Non-response error is error generated due to members of a chosen sample not being contacted even after repeated attempts so that information that should be provided is missing.

1.45 Sampling error results from the variability of outcomes of different samples. This sample to sample variation is inevitably connected to the sampling process. Measurement error is error that results from either self-reported data or data that is collected in an inconsistent manner by those who are responsible for collecting and summarizing the desired information.

1.46 Microsoft Excel:
This product features a spreadsheet-based interface that allows users to organize, calculate, and organize data. Excel also contains many statistical functions to assist in the description of a dataset. Excel can be used to develop worksheets and workbooks to calculate a variety of statistics including introductory and advanced statistics. Excel also includes interactive tools to create graphs, charts, and pivot tables. Excel can be used to summarize data to better understand a population of interest, compare across groups, predict outcomes, and to develop forecasting models. These capabilities represent those that are generally relevant to the current course.

Excel also includes many other statistical capabilities that can be further explored on the Microsoft Office Excel official website.

1.47 (a) The population of interest include banking executives representing institutions of various sizes and U.S. geographic locations.
(b) The collected sample includes 163 banking executives from institutions of various sizes and U.S. geographic locations.
(c) A parameter of interest is the percentage of the population of banking executives that identify customer experience initiatives as an area where increased spending is expected.
(d) A statistic used to the estimate the parameter in (c) is the percentage of the 163 banking executives included in the sample who identify customer experience initiatives as an area where increased spending is expected. In this case, the statistic is 55%.

1.48 The answers are based on an article titled "U.S. Satisfaction Still Running at Improved Level" and written by Lydia Saad (August 15, 2018). The article is located on the following site: https://news.gallup.com/poll/240911/satisfaction-running-improved-level.aspx?g_source=link_NEWSV9&g_medium=NEWSFEED&g_campaign=item_&g_content=U.S.%2520Satisfaction%2520Still%2520Running%2520at%2520Improved%2520Level

(a) The population of interest includes all individuals aged 18 and older who live within the 50 U.S. states and the District of Columbia.
(b) The collected sample includes a random sample of 1,024 individuals aged 18 and older who live within the 50 U.S. states and the District of Columbia.
(c) A parameter of interest is the percentage of the population of individuals aged 18 and older and live within the 50 U.S. states and the District of Columbia who are satisfied with the direction of the U.S.
(d) A statistic used to the estimate the parameter in (c) is the percentage of the 1,024 individuals included in the sample. In this case, the statistic is 36%.

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

1.49 The answers were based on information obtained from the following site:

(a) The population of interest is U.S. CEOs
(b) The sample included 1,000 U.S. CEOs.
(c) A parameter of interest would be the percentage of CEOs among the population of interest that believe that AI will significantly change the way they will do business in the next five years.
(d) The statistic used to estimate the parameter in (c) is the percentage of CEOs among the 1,000 CEOs included in the sample who believe that AI will significantly change the way they will do business in the next five years. In this case, the statistic is 80% agree with this statement

1.50 (a) One variable collected with the American Community Survey is marital status with the following possible responses: now married, widowed, divorced, separated, and never married.
(b) The variable in (a) represents a categorical variable.
(c) Because the variable in (a) is a categorical, this question is not applicable. If one had chosen age in years from the American Community Survey as the variable, the answer to (c) would be discrete.

1.51 Answers will vary depending on the specific sample survey used. The below answers were based on the sample survey located at: bit.ly/21qjI6F

(a) An example of a categorical variable included in the survey is gender with male or female as possible answers.
(b) An example of a numerical variable included in the survey would be the number of phone calls made or received from or to ones direct supervisor in an average week.

1.52 (a) The population of interest consisted of 10,000 benefited employees of the University of Utah.
(b) The sample consisted of 3,095 employees of the University of Utah.
(c) Gender, marital status, and employment category represent categorical variables. Age in years, education level in years completed, and household income represent numerical variables.

1.53 (a) Key social media platforms used represents a categorical variable. The frequency of social media usage represents a discrete numerical variable. Demographics of key social media platform users represent categorical variables.
(b) 1. Which of the following is your preferred social media platform: YouTube, Facebook, or Twitter?
2. What time of the day do you spend the most amount of time using social media: morning, afternoon, or evening?
3. Please indicate your ethnicity?
4. Which of the following do you most often use to access social media: mobile device, laptop computer, desktop computer, other device?
5. Please indicate whether you are a home owner: Yes or No?
(c) 1. For the past week, how many hours did you spend using social media?
2. Please indicate your current age in years.
3. What was your annual income this past year?
4. Currently, how many friends have you accepted on Facebook?
5. Currently, how many twitter followers do you have?

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

FIRST THINGS FIRST

OBJECTIVES

- Statistics is a way of thinking that can lead to better decision making
- Statistics requires analytics skills and is an important part of your business education
- Recent developments such as the use of business analytics and "big data" have made knowing statistics even more critical
- The DCOSAC framework guides your application of statistics
- The opportunity business analytics represents for business students

OVERVIEW AND KEY CONCEPTS

Think Differently About Statistics

Case illustrates some of the newer applications of statistics and business analytics are affecting business. **Statistics** are the methods that allow you to work with data effectively. Business statistics provides you with a formal basis to summarize and visualize business data, reach conclusions about that data, make reliable predictions about business activities, and improve business processes. The **D**efine **C**ollect **O**rganize **S**ummarize **A**nalyze **C**ommunicate framework helps minimize possible errors of thinking and analysis.

- **D**efine the data needed to study in order to solve a problem or meet an objective.
- **C**ollect the data from appropriate sources.
- **O**rganize and **S**ummarize the data collected.
- **A**nalyze and reach conclusions about the data collected.
- **C**ommunicate the results of the analysis to other stakeholders seeking to enhance a business process or who are tasked with decision making.

Understand that analytical skills are more important than arithmetic skills.
Distinguish between statistic and Statistics.

Key Definitions

- **Data** are the facts about the world that one seeks to study and explore. Data can be **unsummarized** or **summarized**.
- **Big data** are collections of data that cannot be easily browsed or analyzed using traditional methods.
- **Variable** defines a characteristic, or property, of an item or individual that can vary among the occurrences of those items or individuals.
- The methods that primarily help summarize and present data comprise **descriptive statistics**.
- Methods that use data collected from a small group to reach conclusions about a larger group comprise **inferential statistics**.
- **Business analytics** combine traditional statistical methods with methods from management science and information systems to form an interdisciplinary tool that supports fact-based decision making.
- **Logical causality** means that you can plausibly claim something directly causes something else.

Starting Point for Using Software

Understand basic computer operations. Understand how software represents and stores data. Understand conventions used for software instructions in this book. Review Rules for Using Software Properly (Exhibit FTF.1).

Copyright ©2024 Pearson Education, Inc.

Basic Business Statistics 15th edition Berenson Solutions Manual

Basic Business Statistics 15th edition Berenson Solutions Manual