

CHAPTER 1 – AN INTRODUCTION TO BUSINESS STATISTICS AND ANALYTICS

§1.1, 1.2 CONCEPTS

1.1 Any characteristic of a population element is called a variable.

Quantitative: we record numeric measurements that represent quantities.

Qualitative: we record which of several categories the element falls into.

LO1-1, LO1-2

1.2

- a. Quantitative; dollar amounts correspond to values on the real number line.
- b. Quantitative; net profit is a dollar amount.
- c. Qualitative; which stock exchange is a category.
- d. Quantitative; national debt is a dollar amount.
- e. Qualitative; which type of medium is a category.

LO1-2

1.3

- (1) Cross-sectional data are collected at approximately the same point in time whereas time series data are collected over different time periods.
- (2) The numbers of cars sold in 2017 by 10 different sales people are cross-sectional data.
- (3) The numbers of cars sold by a particular sales person for the years 2013 – 2017 are time series data.

LO1-3

1.4

- (1) The response variable is whether or not the person has lung cancer.
- (2) The factors are age, sex, occupation, and number of cigarettes smoked per day.
- (3) This is an observational study.

LO1-5

1.5 A data warehouse is a central repository of an organization's data where the data can be retrieved, managed, and analyzed. Big data refers to the massive amounts of data, often collected in real time, that sometimes need quick preliminary analysis for effective business decision making.

LO1-6

§1.1, 1.2 METHODS AND APPLICATIONS

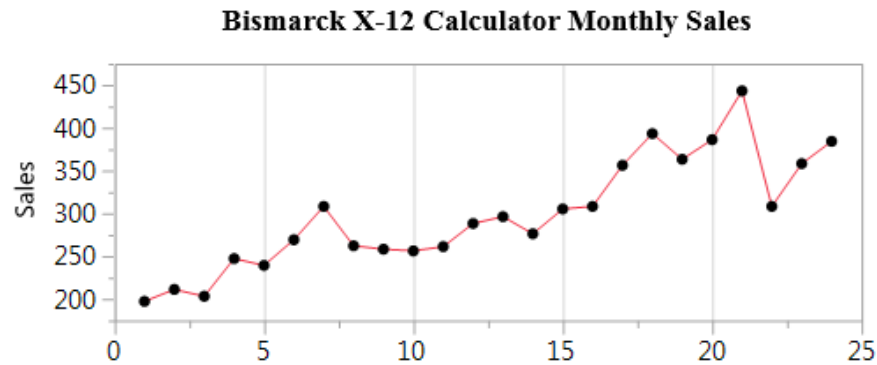
1.6 \$398,000 for a Ruby model on a treed lot

LO1-1

1.7 \$494,000 for a Diamond model on a lake lot; \$447,000 for a Ruby model on a lake lot

LO1-1

1.8



This chart shows that sales are increasing over time.

LO1-4

§1.3, 1.4 CONCEPTS

1.9

- (1) A *population* is the set of all elements about which we wish to draw conclusions.
- (2) You might study the population of all purchasers of a particular laundry detergent.
- (3) A census is the examination of all of the population measurements. A *sample* is a subset of the elements in a population.

LO1-7

1.10

- a. Descriptive statistics is the science of describing the important aspects of a set of measurements.
- b. *Statistical inference* is the science of using a sample of measurements to make generalizations about the important aspects of a population of measurements.
- c. A *random sample* is a subset of size n chosen from a population in such a way that every possible set of elements of size n has the same chance of being chosen. Briefly, the sample is chosen fairly, with no favoritism or prejudice.
- d. A *process* is a sequence of operations that takes input(s) and generates output(s).

LO1-8, LO1-9

- 1.11** When we choose a sample of size n without replacement, all n elements selected are different. However, when selecting with replacement, we might choose some elements multiple times. We tend to get a more complete picture of the population when we sample without replacement.

LO1-9

§1.3, 1.4 METHODS AND APPLICATIONS

- 1.12** We would select companies 3, 8, 9, 14, and 7, so our random sample would contain Coca-Cola, Coca-Cola Enterprises, Reynolds American, Pepsi Bottling Group, and Sara Lee.

LO1-9

1.13

- a. We would select registrations 33,276; 3,427; 8,178; 51,259; 60,268; 58,586; 9,998; 14,346; 24,200; and 7,351.
- b. Most of the 73,219 scores should fall between 36 and 48, the most extreme scores in the sample. Since 46 of the 65 sample values are 42 or higher, we estimate that approximately $46/65 = 70.77\%$ of all scores would be at least 42.

LO1-9

1.14

- a. 5:47 P.M.
- b. We would estimate that the wait times of most customers would fall between 0.4 and 11.6 minutes, the most extreme times in the sample. Since 60 of the 100 sample wait times are less than 6 minutes, we estimate that $60/100 = 60\%$ of all customers would wait less than 6 minutes.

LO1-9

- 1.15** No. This is a voluntary response sample and thus is probably not representative of the population of all television viewers.

LO1-9

- 1.16** We estimate that most breaking strengths will be between 41.7 lbs. and 63.8 lbs., the smallest and largest observed values.

LO1-9

§1.5 CONCEPTS

- 1.17** *Predictive analytics* are supervised learning techniques since there is a particular response variable we are trying to predict.

LO1-10

- 1.18** *Descriptive analytics* are unsupervised learning techniques since there is not a particular response variable we are trying to predict.

LO1-10

- 1.19** *Data mining* is the process of discovering useful knowledge in extremely large data sets.

LO1-10

- 1.20** *Prescriptive analytics are techniques that combine external and internal constraints with results from descriptive and predictive analytics.*

LO1-10

§1.6 CONCEPTS

- 1.21** *A ratio variable is a quantitative variable measured on a scale such that ratios of values of the variables are meaningful and there is an inherently defined zero value.*

An interval variable is a quantitative variable such that ratios of values of the variable are not meaningful and there is not an inherently defined zero value.

LO1-11

- 1.22** *An ordinal variable is a qualitative variable such that there is a meaningful ordering, or ranking, of the categories.*

A nominative variable is a qualitative variable such that there is no meaningful ordering, or ranking, of the categories.

LO1-11

§1.6 METHODS AND APPLICATIONS

- 1.23** Letter Grades: *Ordinal* – each grade from A to F indicates an increasingly lower grade.
Door Choices: *Nominative* – each door is the same except for the number given. For example, Door 1 is not better or worse or higher or lower than Door 3.
TV Classifications: *Ordinal* – each category from TV-G to TV-MA indicates programming appropriate for increasingly older viewers.
PC Ownership: *Nominative* – no ordering of categories.
Restaurant Ratings: *Ordinal* – each rating from 5-star to 1-star indicates an increasingly lower rating.
Filing Status: *Nominative* – no ordering of categories.

LO1-11

- 1.24** PC OS: *Nominative* – no ordering of categories.
Movie Classifications: *Ordinal* – each category from G to X indicates a movie appropriate for increasingly older audiences.
Education Level: *Ordinal* – each ranking from Elementary to Graduate School indicates an increasingly higher educational level.
Football Rankings: *Ordinal* – each ranking from 1 to 10 indicates a team with an increasingly lower ranking (not as good).
Stock Exchanges: *Nominative* – no ordering of categories.
Zip Codes: *Nominative* – no ordering of categories. For example, zip code 45056 is not lower than zip code 90015.

LO1-11

§1.7 CONCEPTS

- 1.25** When the population consists of two or more groups that differ with respect to the variable of interest.

Strata are non-overlapping groups of similar units, and should be chosen so that the units in each stratum are similar on some characteristic (often a categorical variable).

LO1-12

- 1.26** Cluster sampling is often used when selecting a sample from a large geographical region. The name derives from the fact that, at each stage, we “cluster” elements into subpopulations.

LO1-12

- 1.27** First divide 1853 by 100 (since n is 100) and round down to 18. We randomly select one company from the first 18 (in a list of all the companies). From the company selected we simply count down the list of companies by 18 to get to the next company to select. We continue this process until we have reached a sample size of 100.

LO1-12

- 1.28** A stratified random sample is selected by dividing the population into some number of strata, and then randomly sampling inside each stratum.

Potential strata: (1) students who live off campus and (2) students who live on campus.

LO1-12

- 1.29** (1) List all cities with population $> 10,000$. (2) Randomly select a number of such cities. (3) Within each selected city, randomly select a number of city blocks. (4) Within each selected city block, take a random sample of individuals.

LO1-12

§1.8 CONCEPTS

1.30

- a. Phone survey: inexpensive, but impersonal with a low response rate
Mail survey: inexpensive, but impersonal with a low response rate
Mall survey: more expensive, personal interview, questions more easily explained, higher response rate
- b. Dichotomous: yes/no, easily analyzed but very limited in range or depth of response
Multiple choice: more range or depth of response
Open ended: much more difficult to analyze, but significantly more nuance to responses

LO1-13

1.31

- a. *Undercoverage* occurs when some elements in the population are left out of the process of choosing the sample.

- b. *Nonresponse* occurs when data cannot be obtained from an element selected in a sample.
- c. *Response bias* occurs when respondents are reluctant to answer honestly or when the questions are slanted to influence responses.

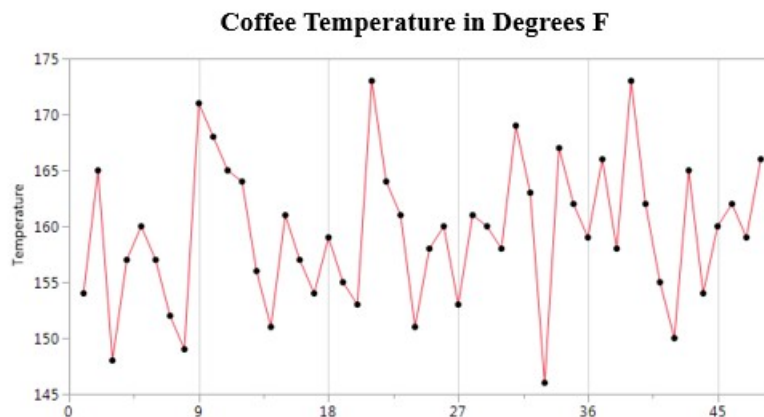
LO1-13

- 1.32 The sample may be *biased* because it is not stated that the recipients of the survey were chosen at random. In addition there may be errors of *undercoverage* and *nonresponse*. Since the return of a survey is *voluntary*, it is possible that people having strong opinions are more likely to respond than those whose opinions are more moderate.

LO1-13

SUPPLEMENTARY EXERCISES

1.33



Basing the limits on the minimum and maximum temperatures observed, the lower limit is 146°F and the upper limit is 173°F.

LO1-4, LO1-9

- 1.34 The time series plot shows that on Mondays through Thursdays, a higher percentage of people wait longer than one minute to be seated compared to Fridays through Sundays. A potential solution is to staff at a higher level on Mondays through Thursdays.

LO1-4

INTERNET EXERCISE

- 1.35 Analyses will vary.