

Chapter 1

Data and Statistics

Learning Objectives

1. Obtain an appreciation for the breadth of statistical applications in business and economics.
2. Understand the meaning of the terms elements, variables, and observations as they are used in statistics.
3. Obtain an understanding of the difference between categorical, quantitative, cross-sectional and time series data.
4. Learn about the sources of data for statistical analysis both internal and external to the firm.
5. Be aware of how errors can arise in data.
6. Know the meaning of descriptive statistics and statistical inference.
7. Be able to distinguish between a population and a sample.
8. Understand the role a sample plays in making statistical inferences about the population.
9. Know the meaning of the term data mining.
10. Be aware of ethical guidelines for statistical practice.

Solutions:

1.
 - a. Quantitative
 - b. Categorical
 - c. Categorical
 - d. Quantitative
 - e. Categorical
2.
 - a. The ten elements are the ten cars
 - b. 5 variables: Size, Cylinders, City MPG, Highway MPG, and Fuel
 - c. Categorical variables: Size and Fuel

Quantitative variables: Cylinders, City MPG, and Highway MPG

d.

Variable	Measurement Scale
Size	Ordinal
Cylinders	Ratio
City MPG	Ratio
Highway MPG	Ratio
Fuel	Nominal

3.
 - a. Average mpg for city driving = $182/10 = 18.2$ mpg
 - b. Average mpg for highway driving = $261/10 = 26.1$ mpg

 On average, the miles per gallon for highway driving is $26.1 - 18.2 = 7.9$ mpg greater compared to city driving.
 - c. 3 of 10 or 30% have four cylinder engines
 - d. 6 of 10 or 60% use regular fuel
4.
 - a. Average endowment = $74.6/7 = \$10.657$ billion
 - b. Average percentage admitted = $111/7 = 15.86\%$
 - c. 3 of 7 or 42.9% have NCAA Division III varsity teams
 - d. 3 of 7 or 42.9% have a City: Midsize campus setting
5.
 - a. The seven elements are the seven schools shown
 - b. 5 variables: State, Campus Setting, Endowment, Applicants Admitted, and NCAA Division
 - c. Categorical variables: State, Campus Setting, and NCAA Division

Quantitative variables: Endowment and Applicants Admitted

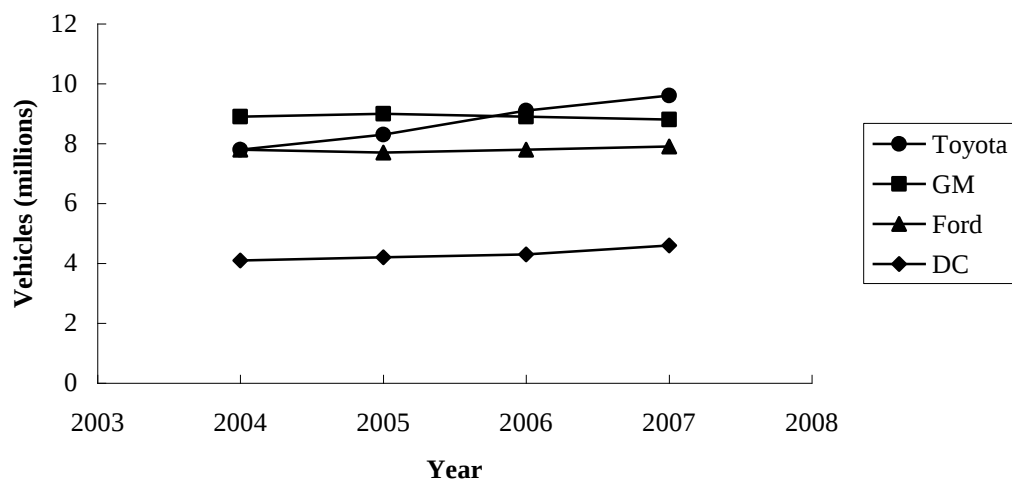
6. Statistics can be referred to as numerical facts. In a broader sense, statistics is the field of study dealing with the collection, analysis, presentation and interpretation of data.
7.
 - a. Quantitative; ratio scale of measurement
 - b. Categorical; nominal scale of measurement
 - c. Categorical; ordinal scale of measurement since the responses can be ordered from earliest (high school) to latest (retirement)
 - d. Quantitative; ratio scale of measurement
 - e. Categorical; nominal scale of measurement
8.
 - a. Categorical
 - b. 30 of 71; 42.3%
9.
 - a. Quantitative; ratio
 - b. Categorical; nominal
 - c. Categorical; ordinal
 - d. Quantitative; ratio
 - e. Categorical; ordinal. The response to this question was recorded as a numerical value from 1 to 10. While the data are numerical, they are not quantitative. The numerical values from 1 to 10 represent categories that *order* the overall rating somewhere between unacceptable and truly exceptional. The data may be ordered by response category with a higher number category indicating a higher overall rating.

While we prefer the categorical; ordinal answer above, at times statisticians may *make the assumption* that the numerical responses are equal-interval measures on a quantitative scale from 1 to 10. When this assumption is made, the data may be considered quantitative with an interval scale of measurement. In this case, additional statistical computations such as the average overall rating become helpful in summarizing the data.

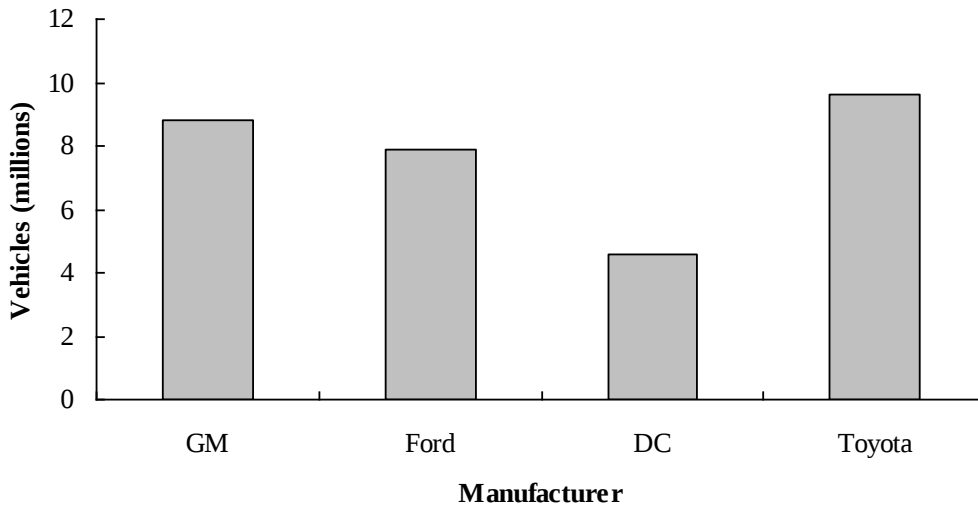
10.
 - a. 1015
 - b. Categorical
 - c. Percentages
 - d. $.10(1015) = 101.5$

101 or 102 respondents said the Federal Bank is doing a good job.

11.
 - a. The population is all visitors coming to the state of Hawaii.
 - b. Since airline flights carry the vast majority of visitors to the state, the use of questionnaires for passengers during incoming flights is a good way to reach this population. The questionnaire actually appears on the back of a mandatory plants and animals declaration form that passengers must complete during the incoming flight. A large percentage of passengers complete the visitor information questionnaire.
 - c. Questions 1 and 4 provide quantitative data indicating the number of visits and the number of days in Hawaii. Questions 2 and 3 provide categorical data indicating the categories of reason for the trip and where the visitor plans to stay.
12.
 - a. Although the data are recorded as numbers, the numbers are codes for the ratings of Fair (1), Average (2), Good (3) and Excellent (4). Thus the variables are categorical with each data value corresponding to a rating category for the variable.
 - b. The data may also be ranked in order of the quality. A higher number indicates a higher rating on a scale from Fair (1) to Excellent (4). Since the data can be ranked or ordered, the scale of measurement is ordinal.
13.
 - a. Federal spending measured in trillions of dollars
 - b. Quantitative
 - c. Time series
 - d. Federal spending has increased over time
14. The answer to this exercise depends upon the time series of the average price per gallon of conventional regular gasoline since April 2009. Contact the website www.eia.doe.gov to obtain the more recent time series data. In the spring of 2009, the average price per gallon was once again increasing. A continuation of the usual summer peak in gasoline prices was anticipated.
15. a.



- b. According to the CSM data, Toyota surpasses General Motors as the biggest auto manufacturer in the world in 2006. In 2006, Toyota manufactured approximately $(9.1 - 8.9) = .2$ million or 200,000 more vehicles than General Motors. The gap is expected to widen to 800,000 vehicles in 2007. General Motors is the only manufacturer showing a decline in vehicle production over the four year period.
- c. The following is a bar chart of cross-sectional data as it shows the number of vehicles manufactured in 2007.

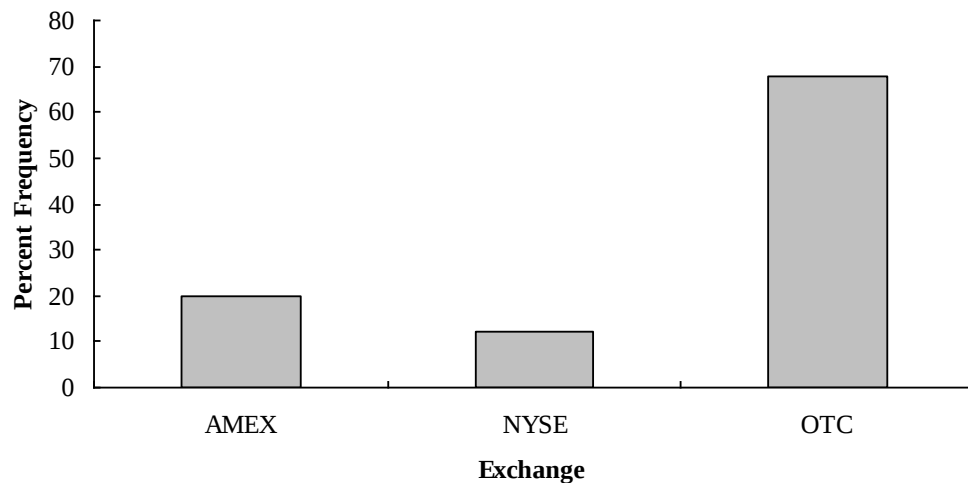


16.
 - a. Quantitative – number of new drugs approved
 - b. Time series from 1996 to 2003
 - c. 18
 - d. 2002; 16 new drugs
 - e. Over the eight-year period, the number of new drugs approved by the FDA declined. From approximately 50 new drugs approved in 1996, the most recent years are showing only 16 to 18 new drugs approved.
17.
 - a. All subscribers of Business Week in North America at the time the survey was conducted.
 - b. Quantitative
 - c. Categorical (yes or no)
 - d. Crossectional - all the data relate to the same time.
 - e. Using the sample results, we could infer or estimate 59% of the population of subscribers have an annual income of \$75,000 or more and 50% of the population of subscribers have an American Express credit card.

18. a. Nielsen is attempting to measure the popularity of each television program by showing the percentage of households that are watching the program.
- b. All households with televisions in the United States.
- c. A census of the population is impossible. A sample provides timely information in that the ratings and share data can be obtained weekly. In addition, the sample saves data collection costs.
- d. The cancellation or renewal of television programs, advertising cost rates for the television programs and the scheduling of television programs are often based on the Nielsen information.
19. a. 43% of managers were bullish or very bullish.
- 21% of managers expected health care to be the leading industry over the next 12 months.
- b. We estimate the average 12-month return estimate for the population of investment managers to be 11.2%.
- c. We estimate the average over the population of investment managers to be 2.5 years.
20. a. The population consists of all customers of the chain's stores in Charlotte, North Carolina.
- b. Some of the ways that could used to collect the data are as follows:
- Customers entering or leaving the store could be surveyed
 - A survey could be mailed to customers who have a shopper's club card for the stores
 - Customers could be given a printed survey when they check out
 - Customers could be given a coupon that asks them to complete a brief on-line survey; if they do, they will receive a 5% discount on their next shopping trip.
21. a. $\frac{155}{430} = .36$ or 36%
- b. 44% of 430 = $.44(430) = 189$ business travelers
- c. Categorical data with categories online travel site, travel agent, direct with airline/hotel, other.
22. a. The two populations are the population of women whose mothers took the drug DES during pregnancy and the population of women whose mothers did not take the drug DES during pregnancy.
- b. It was a survey.
- c. $63 / 3.980 = 15.8$ women out of each 1000 developed tissue abnormalities.
- d. The article reported "twice" as many abnormalities in the women whose mothers had taken DES during pregnancy. Thus, a rough estimate would be $15.8/2 = 7.9$ abnormalities per 1000 women whose mothers had *not* taken DES during pregnancy.
- e. In many situations, disease occurrences are rare and affect only a small portion of the population. Large samples are needed to collect data on a reasonable number of cases where the disease exists.

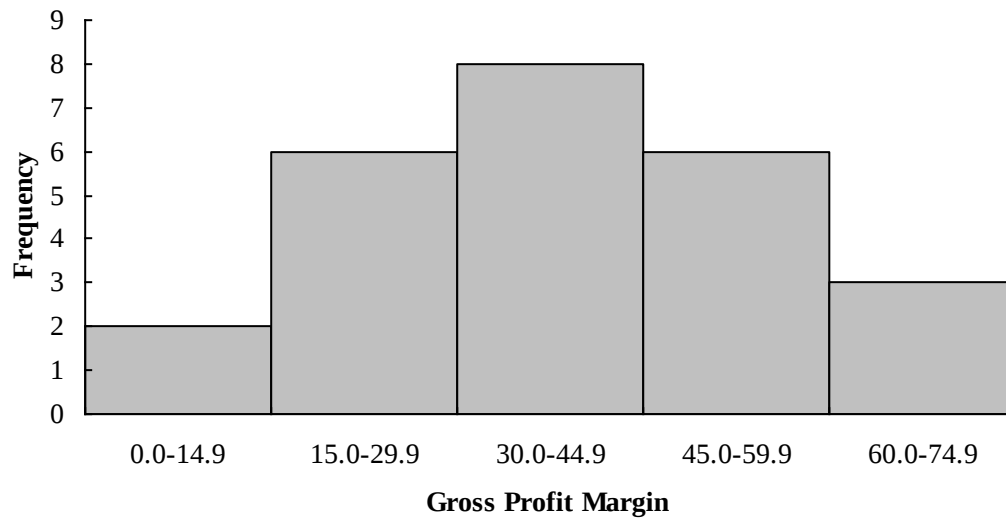
23. Internal data on salaries of other employees can be obtained from the personnel department. External data might be obtained from the Department of Labor or industry associations.
24. a. There are five variables: Exchange, Ticker Symbol, Market Cap, Price/Earnings Ratio and Gross Profit Margin.
- b. Categorical variables: Exchange and Ticker Symbol
- Quantitative variables: Market Cap, Price/Earnings Ratio, Gross Profit Margin
- c. Exchange variable:

Exchange	Frequency	Percent Frequency
AMEX	5	(5/25) 20%
NYSE	3	(3/25) 12%
OTC	<u>17</u>	(17/25) 68%
	25	



- d. Gross Profit Margin variable:

Gross Profit Margin	Frequency
0.0 – 14.9	2
15.0 – 29.9	6
30.0 – 44.9	8
45.0 – 59.9	6
60.0 – 74.9	3



- e. Sum the Price/Earnings Ratio data for all 25 companies.

$$\text{Sum} = 505.4$$

$$\text{Average Price/Earnings Ratio} = \text{Sum}/25 = 505.4/25 = 20.2$$

25. a. This is a statistically correct descriptive statistic for the sample.
- b. An incorrect generalization since the data was not collected for the entire population.
- c. An acceptable statistical inference based on the use of the word “estimate.”
- d. While this statement is true for the sample, it is not a justifiable conclusion for the entire population.
- e. This statement is not statistically supportable. While it is true for the particular sample observed, it is entirely possible and even very likely that at least some students will be outside the 65 to 90 range of grades.