

# 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. 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.
2. a. The ten elements are the ten tablet computers
- b. 5 variables: Cost (\$), Operating System, Display Size (inches), Battery Life (hours), CPU Manufacturer
- c. Categorical variables: Operating System and CPU Manufacturer

Quantitative variables: Cost (\$), Display Size (inches), and Battery Life (hours)

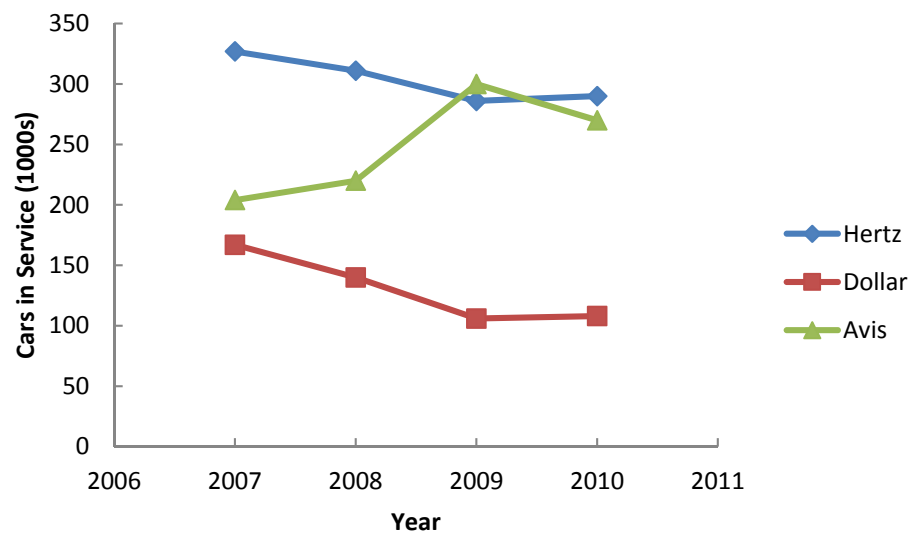
d.

<u>Variable</u>	<u>Measurement Scale</u>
Cost (\$)	Ratio
Operating System	Nominal
Display Size (inches)	Ratio
Battery Life (hours)	Ratio
CPU Manufacturer	Nominal

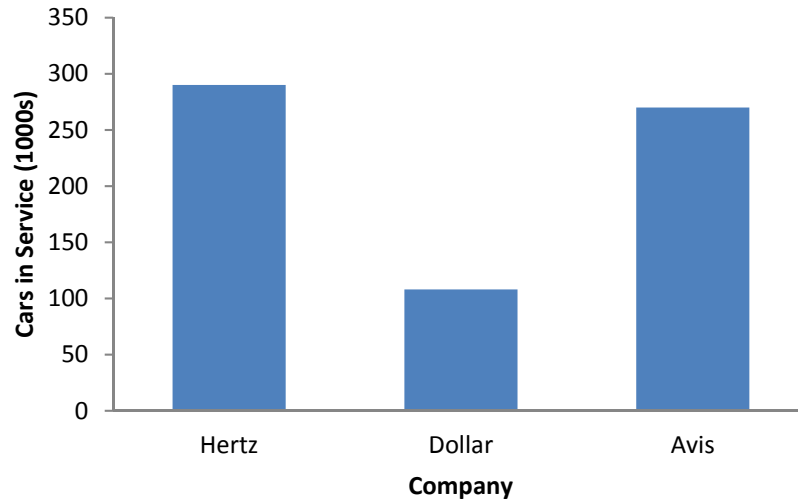
3. a. Average cost =  $5829/10 = \$582.90$
- b. Average cost with a Windows operating system =  $3616/5 = \$723.20$   
Average cost with an Android operating system =  $1714/4 = \$428.5$   
The average cost with a Windows operating system is much higher.
- c. 2 of 10 or 20% use a CPU manufactured by TI OMAP
- d. 4 of 10 or 40% use an Android operating system
4. a. There are eight elements in this data set; each element corresponds to one of the eight models of cordless telephones
- b. Categorical variables: Voice Quality and Handset on Base  
Quantitative variables: Price, Overall Score, and Talk Time
- c. Price – ratio measurement  
Overall Score – interval measurement  
Voice Quality – ordinal measurement  
Handset on Base – nominal measurement  
Talk Time – ratio measurement
5. a. Average Price =  $545/8 = \$68.13$
- b. Average Talk Time =  $71/8 = 8.875$  hours
- c. Percentage rated Excellent: 2 of 8  $2/8 = .25$ , or 25%
- d. Percentage with Handset on Base: 4 of 8  $4/8 = .50$ , or 50%

6.
  - a. Categorical
  - b. Quantitative
  - c. Categorical
  - d. Quantitative
  - e. Quantitative
7.
  - a. Each question has a yes or no categorical response.
  - b. Yes and no are the labels for the customer responses. A nominal scale is being used.
8.
  - a. 762
  - b. Categorical
  - c. Percentages
  - d.  $.67(762) = 510.54$   
510 or 511 respondents said they want the amendment to pass.
9.
  - a. Categorical
  - b. 30 of 71; 42.3%
10.
  - a. Categorical
  - b. Percentages
  - c. 44 of 1080 respondents or approximately 4% strongly agree with allowing drivers of motor vehicles to talk on a hand-held cell phone while driving.
  - d. 165 of the 1080 respondents or 15% of said they somewhat disagree and 741 or 69% said they strongly disagree. Thus, there does not appear to be general support for allowing drivers of motor vehicles to talk on a hand-held cell phone while driving.
11.
  - a. Categorical
  - b.  $295 + 672 + 51 = 1018$
  - c.  $(295/1018)100 = 29\%$
  - d. Support against;  $(672/1018)100 = 66\%$  said they would vote against the law.
12.
  - 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.
13. a. Google revenue in billions of dollars  
b. Quantitative  
c. Time series  
d. Google revenue is increasing over time.
14. a. The graph of the time series follows:



- b. In 2007 and 2008 Hertz was the clear market share leader. In 2009 and 2010 Hertz and Avis have approximately the same market share. The market share for Dollar appears to be declining.
- c. The bar chart for 2010 is shown below.



This chart is based on cross-sectional data.

15.
  - a. Quantitative – number of recreational boating accidents
  - b. Time series
  - c. July; 1100
  - d. 2.9%; Yes, because most recreational boating takes place during the summer months.
  - e. The bar graph follows the shape of a bell curve.
  
16. The answers to this exercise depend on updating the time series of the average price per gallon of conventional regular gasoline as shown in Figure 1.1. Contact the website [www.eia.doe.gov](http://www.eia.doe.gov) to obtain the most recent time series data. The answers should focus on the most recent changes or trend in the average price per gallon.
  
17. 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.
  
18.
  - a.  $684/1021$ ; or approximately 67%
  - b. 612
  - c. Categorical
  
19.
  - 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.

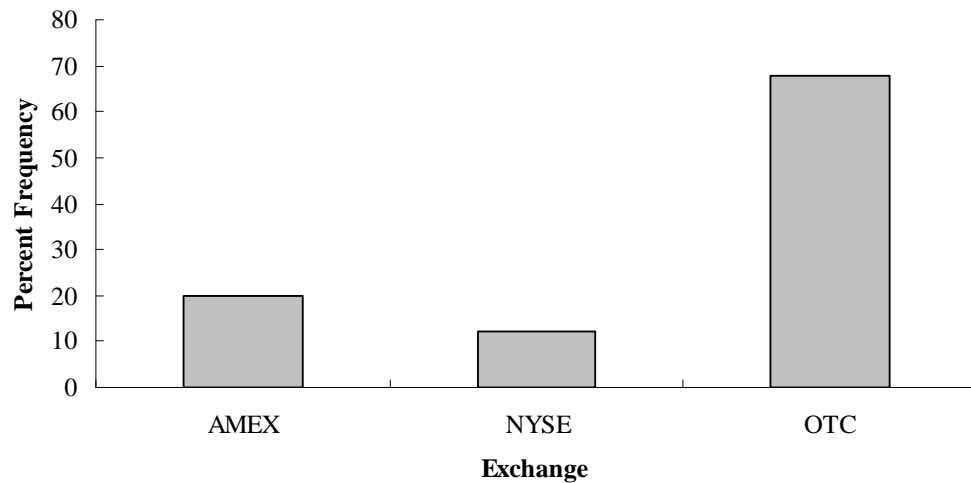
20. a. 43% of managers were bullish or very bullish.  
21% of managers expected health care to be the leading industry over the next 12 months.
- b. We estimate the average 12-month return estimate for the population of investment managers to be 11.2%.
- c. We estimate the average over the population of investment managers to be 2.5 years.
21. 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.
22. a. The population consists of all clients that currently have a home listed for sale with the agency or have hired the agency to help them locate a new home.
- b. Some of the ways that could be used to collect the data are as follows:
- A survey could be mailed to each of the agency’s clients.
  - Each client could be sent an email with a survey attached.
  - The next time one of the firm’s agents meets with a client they could conduct a personal interview to obtain the data.
23. a. This finding is applicable to the population of all American adults.
- b. This finding is applicable to the population of American adults that own a cellphone and/or a tablet computer.
- c. They conducted a sample survey. It would be way too costly to survey all American adults or all American adults who own cellphones and/or tablet computers. As we will see later in the text, very good results can be obtained using a sample survey.
- d. These results should be quite interesting to restaurant owners. It suggests that it would be worthwhile for them to have a website and to consider advertising through an internet search company, such as Google.
24. 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. It is entirely possible and even very likely that at least some students will be outside the 65 to 90 range of grades.
25. 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

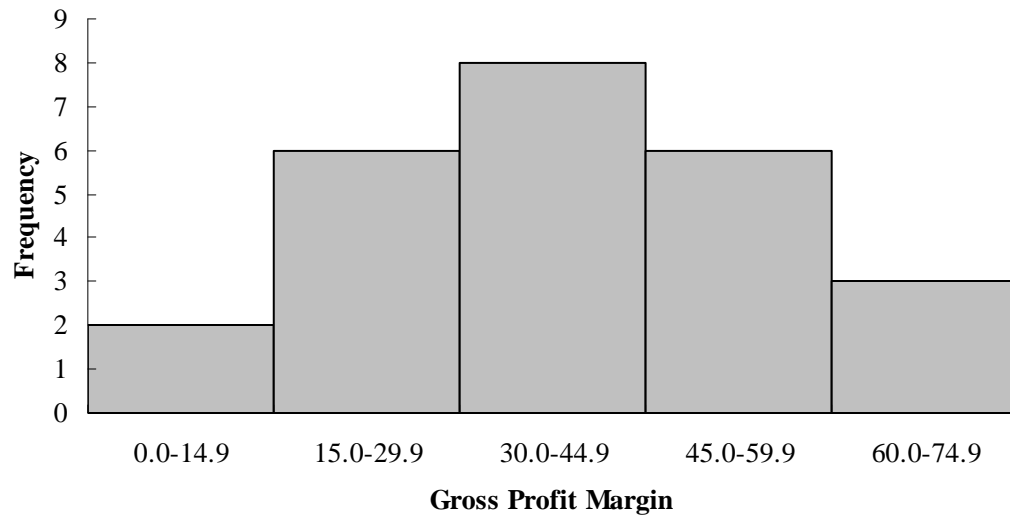
- c. Exchange variable:

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.

$$\text{Sum} = 505.4$$

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