STUDENT SOLUTIONS MANUAL

NANCY S. BOUDREAU

STATISTICS FOR BUSINESS AND ECONOMICS

SEVENTH EDITION

McCLAVE • BENSON • SINCICH

STUDENT SOLUTIONS MANUAL

NANCY S. BOUDREAU

STATISTICS FOR BUSINESS AND ECONOMICS

SEVENTH EDITION

MCCLAVE • BENSON • SINCICH

PRENTICE HALL, UPPER SADDLE RIVER, NJ 07458

Acquisitions Editor: ***Ann Heath***
Production Editor: ***Mindy De Palma***
Cover Designer: ***Paul Gourhan***
Special Projects Manager: ***Barbara A. Murray***
Supplement Cover Mnaager: ***Paul Gourhan***
Editorial Assistant: ***Mindy Ince McClard***
Production Coordinator: ***Alan Fischer***

Printed in the United States of America

10 9 8 7 6 5 4

ISBN 0-13-625260-5

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada, Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Pearson Education Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil,Ltda., *Rio de Janeiro*

Contents

Preface

This solutions manual is designed to accompany the text, ***Statistics for Business and Economics***, Seventh Edition, by James T. McClave, P. George Benson, and Terry Sincich. It provides answers to most odd-numbered exercises for each chapter in the text. Other methods of solution may also be appropriate; however, the author has presented one that she believes to be most instructive to the beginning Statistics student. The student should first attempt to solve the assigned exercises without help from this manual. Then, if unsuccessful, the solution in the manual will clarify points necessary to the solution. The student who successfully solves an exercise should still refer to the manual's solution. Many points are clarified and expanded upon to provide maximum insight into and benefit from each exercise.

Instructors will also benefit from the use of this manual. It will save time in preparing presentations of the solutions and possibly provide another point of view regarding their meaning.

Some of the exercises are subjective in nature and thus omitted from the Answer Key at the end of ***Statistics for Business and Economics***, Seventh Edition. The subjective decisions regarding these exercises have been made and are explained by the author. Solutions based on these decisions are presented; the solution to this type of exercise is often most instructive. When an alternative interpretation of an exercise may occur, the author has often addressed it and given justification for the approach taken.

I would like to thank Kelly Evans for creating the art work and Brenda Dobson for her assistance and for typing this work.

Nancy S. Boudreau
Bowling Green State University
Bowling Green, Ohio

Statistics, Data, and Statistical Thinking

Chapter 1

1.1 Statistics is a science that deals with the collection, classification, analysis, and interpretation of information or data. It is a meaningful, useful science with a broad, almost limitless scope of applications to business, government, and the physical and social sciences.

1.3 The four elements of a descriptive statistics problem are:

1. The population or sample of interest. This is the collection of all the units upon which the variable is measured.
2. One or more variables that are to be investigated. These are the types of data that are to be collected.
3. Tables, graphs, or numerical summary tools. These are tools used to display the characteristic of the sample or population.
4. Conclusions about the data based on the patterns revealed. These are summaries of what the summary tools revealed about the population or sample.

1.5 The first major method of collecting data is from a published source. These data have already been collected by someone else and is available in a published source. The second method of collecting data is from a designed experiment. These data are collected by a researcher who exerts strict control over the experimental units in a study. These data are measured directly from the experimental units. The third method of collecting data is from a survey. These data are collected by a researcher asking a group of people one or more questions. Again, these data are collected directly from the experimental units or people. The final method of collecting data is observationally. These data are collected directly from experimental units by simply observing the experimental units in their natural environment and recording the values of the desired characteristics.

1.7 A population is a set of existing units such as people, objects, transactions, or events. A variable is a characteristic or property of an individual population unit such as height of a person, time of a reflex, amount of a transaction, etc.

1.9 A representative sample is a sample that exhibits characteristics similar to those possessed by the target population. A representative sample is essential if inferential statistics is to be applied. If a sample does not possess the same characteristics as the target population, then any inferences made using the sample will be unreliable.

1.11 A population is a set of existing units such as people, objects, transactions, or events. A process is a series of actions or operations that transform inputs to outputs. A process produces or generates output over time. Examples of processes are assembly lines, oil refineries, and stock prices.

1.13 The data consisting of the classifications A, B, C, and D are qualitative. These data are nominal and thus are qualitative. After the data are input as 1, 2, 3, and 4, they are still nominal and thus qualitative. The only differences between the two data sets are the names of the categories. The numbers associated with the four groups are meaningless.

1.15 a. The population of interest is all citizens of the United States.

b. The variable of interest is the view of each citizen as to whether the president is doing a good or bad job. It is qualitative.

c. The sample is the 2000 individuals selected for the poll.

d. The inference of interest is to estimate the proportion of all citizens who believe the president is doing a good job.

e. The method of data collection is a survey.

f. It is not very likely that the sample will be representative of the population of all citizens of the United States. By selecting phone numbers at random, the sample will be limited to only those people who have telephones. Also, many people share the same phone number, so each person would not have an equal chance of being contacted. Another possible problem is the time of day the calls are made. If the calls are made in the evening, those people who work in the evening would not be represented.

1.17 a. The depth of tread is a number such as .25 inch, .15 inch, etc. Therefore, it is quantitative.

b. Occupations take on values such as doctor, lawyer, carpenter, etc., which are not numeric. Therefore, it is qualitative.

c. Employment status can take on values such as employed or unemployed, which are not numeric. Therefore, it is qualitative.

d. The time in months can take on values such as 1, 2, 3, etc. Therefore, it is quantitative.

1.19 a. Length of maximum span can take on values such as 15 feet, 50 feet, 75 feet, etc. Therefore, it is quantitative.

b. The number of vehicle lanes can take on values such as 2, 4, etc. Therefore, it is quantitative.

c. The answer to this item is "yes" or "no," which are not numeric. Therefore, it is qualitative.

d. Average daily traffic could take on values such as 150 vehicles, 3,579 vehicles, 53,295 vehicles, etc. Therefore, it is quantitative.

e. Condition can take on values "good," "fair," or "poor," which are not numeric. Therefore, it is qualitative.

f. The length of the bypass or detour could take on values such as 1 mile, 4 miles, etc. Therefore, it is quantitative.

g. Route type can take on values "interstate," U.S.," "state," "county," or "city," which are not numeric. Therefore, it is qualitative.

1.21 a. The population from which the sample was selected is the set of all department store executives.

b. There are two variables measured by the authors. They are job-satisfaction and Machiavellian rating for each of the executives.

c. The sample is the set of 218 department store executives who completed the questionnaire.

d. The method of data collection is a survey.

e. The inference made by the authors is that those executives with higher job-satisfaction scores are likely to have a lower 'mach' rating.

1.23 a. Some possible questions are:

1. In your opinion, why has the banking industry consolidated in the past few years? Check all that apply.

   a. Too many small banks with not enough capital.
   b. A result of the Savings and Loan scandals.
   c. To eliminate duplicated resources in the upper management positions.
   d. To provide more efficient service to the customers.
   e. To provide a more complete list of financial opportunities for the customers.
   f. Other. Please list.

2. Using a scale from 1 to 5, where 1 means strongly disagree and 5 means strongly agree, indicate your agreement to the following statement: "The trend of consolidation in the banking industry will continue in the next five years."

   1 strongly disagree 2 disagree 3 no opinion 4 agree 5 strongly agree

b. The population of interest is the set of all bank presidents in the United States.

c. It would be extremely difficult and costly to obtain information from all 10,000 bank presidents. Thus, it would be more efficient to sample just 200 bank presidents. However, by sending the questionnaires to only 200 bank presidents, one risks getting the results from a sample which is not representative of the population. The sample must be chosen in such a way that the results will be representative of the entire population of bank presidents in order to be of any use.

1.25 a. The population of interest is the collection of all major U.S. firms.

b. The variable of interest is whether the firm offers job-sharing to its employees or not.

c. The sample is the set of 1,035 firms selected.

d. The government might want to estimate the proportion of all firms that offer job-sharing to their employees.

1.27 I. Qualitative; the possible responses are "yes" or "no," which are nonnumerical.

II. Quantitative; age is measured on a numerical scale, such as 15, 32, etc.

III. Quantitative; the rating is measured on a numerical scale from 1 to 10, where the higher the rating the more helpful the *Tutorial* instructions.

IV. Qualitative; the possible responses are "laser printer" or "another type of printer," which are nonnumerical.

V. Qualitative; the speeds can be classified as "slower," "unchanged," or "faster," which are nonnumerical.

VI. Quantitative; the number of people in a household who have used Windows 3.0 at least once is measured on a numerical scale, such as 0, 1, 2, etc.

1.29 a. The process being studied is the distribution of pipes, valves, and fittings to the refining, chemical, and petrochemical industries by Wallace Company of Houston.

b. The variables of interest are the speed of the deliveries, the accuracy of the invoices, and the quality of the packaging of the products.

c. The sampling plan was to monitor a subset of current customers by sending out a questionnaire twice a year and asking the customers to rate the speed of the deliveries, the accuracy of the invoices, and the quality of the packaging minutes. The sample is the total numbers of questionnaires received.

d. The Wallace Company's immediate interest is learning about the delivery process of its distribution of pipes, valves, and fittings. To do this, it is measuring the speed of deliveries, the accuracy of the invoices, and the quality of its packaging from the sample of its customers to make an inference about the delivery process to all customers. In particular, it might use the mean speed of its deliveries to the sampled customers to estimate the mean speed of its deliveries to all its customers. It might use the mean accuracy of its invoices to the sampled customers to estimate the mean accuracy of its deliveries to all its customers. It might use the mean rating of the quality of its packaging to the sampled customers to estimate the mean rating of the quality of its packaging to all its customers.

e. Several factors might affect the reliability of the inferences. One factor is the set of customers selected to receive the survey. If this set is not representative of all the customers, the wrong inferences could be made. Also, the set of customers returning the surveys may not be representative of all its customers. Again, this could influence the reliability of the inferences made.

Methods for Describing Sets of Data

Chapter 2

2.1 First, we find the frequency of the grade A. The sum of the frequencies for all five grades must be 200. Therefore, subtract the sum of the frequencies of the other four grades from 200. The frequency for grade A is:

$$200 - (36 + 90 + 30 + 28) = 200 - 184 = 16$$

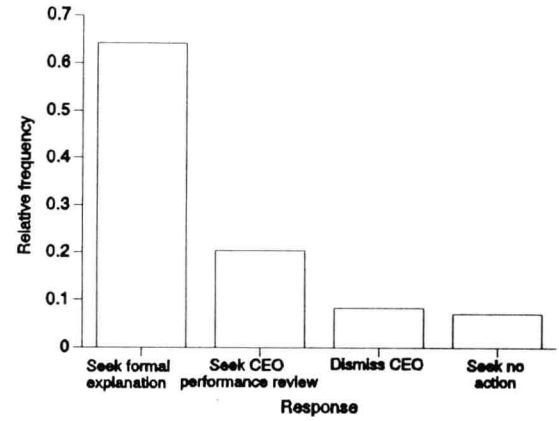
To find the relative frequency for each grade, divide the frequency by the total sample size, 200. The relative frequency for the grade B is 36/200 = .18. The rest of the relative frequencies are found in a similar manner and appear in the table:

Grade on Statistics Exam	Frequency	Relative Frequency
A: 90–100	16	.08
B: 80– 89	36	.18
C: 65– 79	90	.45
D: 50– 64	30	.15
F: Below 50	28	.14
Total	200	1.00

2.3 a. We must first compute the relative frequency for each response. To find the relative frequency, we divide the frequency by the total sample size, 240. For the first category, the relative frequency is 154/240 = .642. The rest of the relative frequencies are found in a similar manner and are shown in the table.

Response	Number of Investors	Relative Frequency
Seek formal explanation	154	.642
Seek CEO performance review	49	.204
Dismiss CEO	20	.083
Seek no action	17	.071
TOTAL	240	1.000

b. The relative frequency graph is:



c. If the chief executive officer and the board of directors differed on company strategy, almost 2/3 of the large investors would seek formal explanation (.642). Approximately 20% (.204) would seek CEO performance review. Very few would dismiss the CEO (.083) or would seek no action (.071).

2.5 a. The variable measured by Performark is the length of time it took for each advertiser to respond back.

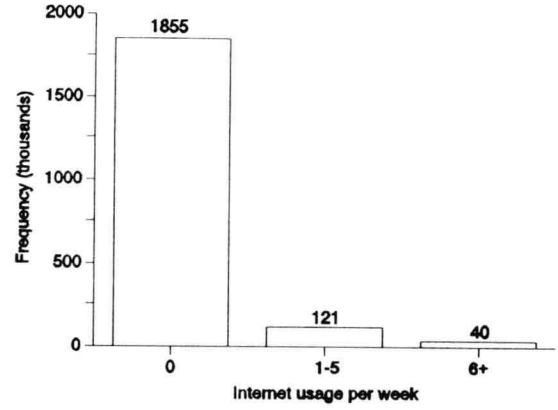
b. The pie chart is:

c. Twenty-one percent of .21 × 17,000 = 3,570 of the advertisers never respond to the sales lead.

d. The information from the pie chart does not indicate how effective the "bingo cards" are. It just indicates how long it takes advertisers to respond, if at all.

2.7 a. The variable measured in the survey was the length of time small businesses used the Internet per week.

b. A bar graph of the data is:



c. The proportion of the 2,016 small businesses that use the Internet on a weekly basis is $(121 + 40)/2{,}016 = 161/2{,}016 = .08$.

2.9 a. From the bar chart, there are approximately 195 transgenic plant trials approved for herbicide tolerance over this period.

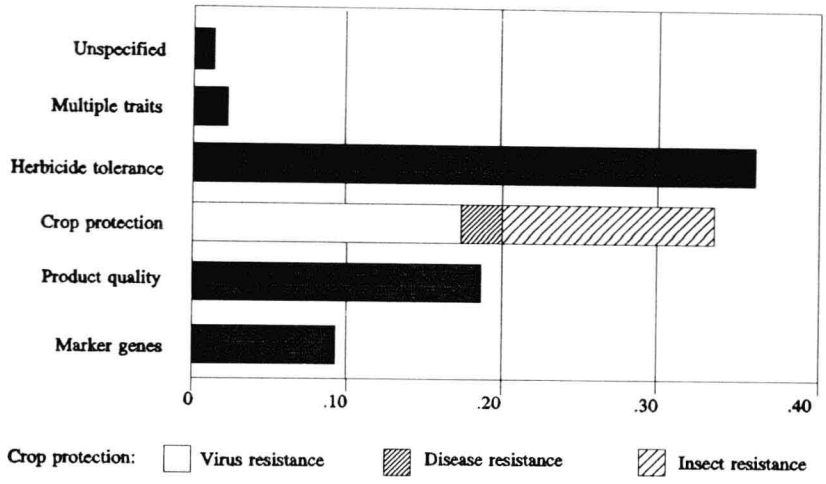
b. From the bar chart, there are approximately 90 transgenic plant trials approved for developing virus resistant crops over this period.

c. To modify the bar graph to show relative frequencies, we must first find the total number of transgenic plant trials over the period. From the bar graph, the following are the approximate frequencies:

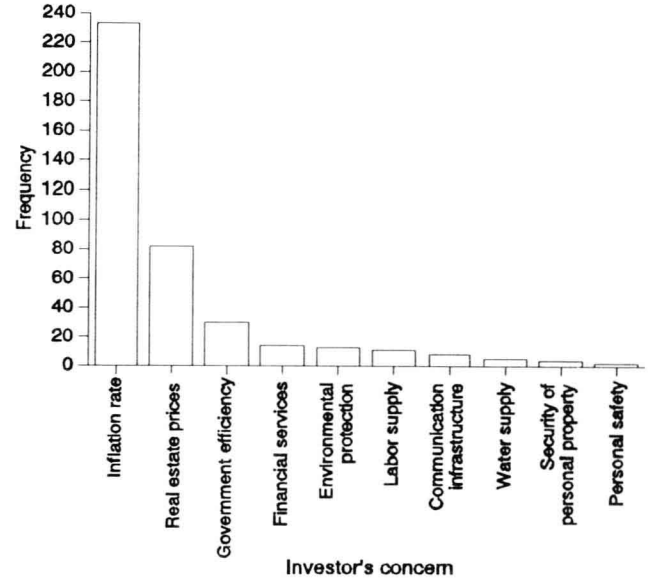
Trait	Frequency	Relative Frequency
Unspecified	9	9/550 = .016
Multiple Traits	13	13/550 = .024
Herbicide Tolerance	195	195/550 = .355
Crop Production	180	180/550 = .327
Product Quality	103	103/550 = .187
Marker Genes	50	50/550 = .091
Totals	550	1.000

To find the relative frequencies, divide the frequency by the total sample size, 550. The relative frequencies appear in the table above.

The bar chart showing the relative frequencies is:



2.11 a. The Pareto diagram is:



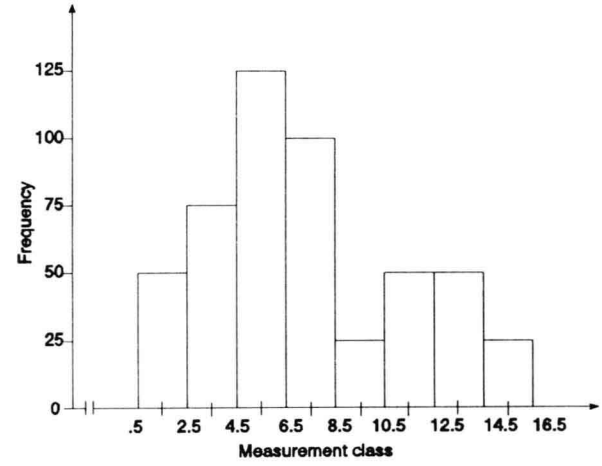
b. The environmental factor of most concern is "Inflation rate" with 233/402 = .58 or 58% of the investors indicating this as their most serious concern. The second most serious concern was "Real Estate prices." Over 20% ((82/402) × 100% = 20.4%) of the investors chose this concern. Each of the other categories were chosen by less than 10% of the investors.

c. Two factors out of 10 represent 20% of the factors. The two factors are "Inflation rate" and "Real estate prices." These two factors represent ((233 + 82)/402 = .78) 78% of the investors. This is very close to 80%.

2.13 To find the number of measurements for each measurement class, multiply the relative frequency by the total number of observations, $n = 500$. The frequency table is:

Measurement Class	Relative Frequency	Frequency
.5 – 2.5	.10	500(.10) = 50
2.5 – 4.5	.15	500(.15) = 75
4.5 – 6.5	.25	500(.25) = 125
6.5 – 8.5	.20	500(.20) = 100
8.5 – 10.5	.05	500(.05) = 25
10.5 – 12.5	.10	500(.10) = 50
12.5 – 14.5	.10	500(.10) = 50
14.5 – 16.5	.05	500(.05) = 25
		500

The frequency histogram is:



2.15 a. This is a frequency histogram because the number of observations is graphed for each interval rather than the relative frequency.

b. There are 14 measurement classes.

c. There are 49 measurements in the data set.

2.17 a. Almost half (14) of the bid prices were between \$99.50 and \$102.50. Seventy percent (21) of the bid prices were between \$96.50 and \$105.50. Only one bid price was greater than \$105.50.

b. The total number of bonds with bid prices greater than \$96.50 is 3 + 14 + 4 + 1 = 22. The proportion of the total is 22/33 = .733.

2.19 a. Using MINITAB, the stem-and-leaf display is:

```
Stem-and-Leaf of PENALTY          N = 38
Leaf Unit = 10000

(28)     0 (0)0(1)(1)111(2)(2)222222(3)(3)33334444899
 10      1 (0)0239
  5      2
  5      3 0
  4      4 0
  3      5
  3      6
  3      7
  3      8 5
  2      9 3
  1     10 0
```

b. See the circled leaves in part a.

c. Most of the penalties imposed for Clean Air Act violations are relatively small compared to the penalties imposed for other violations. All but one of the penalties for Clean Air Act violations are below the median penalty imposed.

2.21 a. Using MINITAB, the three frequency histograms are as follows (the same starting point and class interval were used for each):

```
Histogram of C1       N = 25

Tenth Performance

Midpoint   Count
    4.00       0
    8.00       0
   12.00       1  *
   16.00       5  *****
   20.00      10  **********
   24.00       6  ******
   28.00       0
   32.00       2  **
   36.00       0
   40.00       1  *

Histogram of C2       N = 25

Thirtieth Performance

Midpoint   Count
    4.00       1  *
    8.00       9  *********
   12.00      12  ************
   16.00       2  **
   20.00       1  *

Histogram of C3       N = 25

Fiftieth Performance

Midpoint   Count
    4.00       3  ***
    8.00      15  ***************
   12.00       4  ****
   16.00       2  **
   20.00       1  *
```

b. The histogram for the tenth performance shows a much greater spread of the observations than the other two histograms. The thirtieth performance histogram shows a shift to the left—implying shorter completion times than for the tenth performance. In addition, the fiftieth performance histogram shows an additional shift to the left compared to that for the

thirtieth performance. However, the last shift is not as great as the first shift. This agrees with statements made in the problem.

2.23 a. Most of the observations are below 3000, while two observations are in the high 3000 area and five are 4000 or more. Since SPSS is using a stem which is the 1000's place rounded down to an even number, it is not always possible to determine the actual value of the five largest rates. For example, "8 . 1" could represent a number near 8100 or 9100. The leaves are rounded to the nearest 100. The largest value in the data set is 11,968.23, which is rounded to a representation of "12 . 0".

b. In the second display, the stem is the 1000's place and the leaf is the 100's place. The data seem to be clustering in the 0 to 3000 range with some extreme values.

2.25 a. Using MINITAB, histograms of the two data sets are:

```
Histogram of C1       N = 20

Manufacturing Companies

Midpoint   Count
     0.0       0
    10.0       6  ******
    20.0       6  ******
    30.0       2  **
    40.0       4  ****
    50.0       0
    60.0       0
    70.0       2  **
```

```
Histogram of C2     N = 20

Holding Firms

Midpoint      Count
     0.0          3  ***
    10.0         11  ***********
    20.0          5  *****
    30.0          0
    40.0          0
    50.0          1  *
    60.0          0
    70.0          0
```

b. The P/E ratios of firms in the manufacturing business tend to be higher than the P/E ratios of firms in the holding business. From the histograms, eight of the 20 manufacturing firms had P/E ratios greater than 30, while only one of the 20 holding firms had a P/E ratio greater than 30.

2.27 a. $\sum x = 3 + 8 + 4 + 5 + 3 + 4 + 6 = 33$

b. $\sum x^2 = 3^2 + 8^2 + 4^2 + 5^2 + 3^2 + 4^2 + 6^2 = 175$

c. $\sum (x - 5)^2 = (3 - 5)^2 + (8 - 5)^2 + (4 - 5)^2 + (5 - 5)^2 + (3 - 5)^2 + (4 - 5)^2 + (6 - 5)^2 = 20$

d. $\sum (x - 2)^2 = (3 - 2)^2 + (8 - 2)^2 + (4 - 2)^2 + (5 - 2)^2 + (3 - 2)^2 + (4 - 2)^2 + (6 - 2)^2 = 71$

e. $(\sum x)^2 = (3 + 8 + 4 + 5 + 3 + 4 + 6)^2 = 33^2 = 1089$

2.29 a. $\sum x = 6 + 0 + (-2) + (-1) + 3 = 6$

b. $\sum x^2 = 6^2 + 0^2 + (-2)^2 + (-1)^2 + 3^2 = 50$

c. $\sum x^2 - \frac{(\sum x)^2}{5} = 50 - \frac{6^2}{5} = 50 - 7.2 = 42.8$

2.31 Assume the data are a sample. The sample mean is:

$$\bar{x} = \frac{\sum x}{n} = \frac{3.2 + 2.5 + 2.1 + 3.7 + 2.8 + 2.0}{6} = \frac{16.3}{6} = 2.717$$

The median is the average of the middle two numbers when the data are arranged in order (since $n = 6$ is even). The data arranged in order are: 2.0, 2.1, 2.5, 2.8, 3.2, 3.7. The middle two numbers are 2.5 and 2.8. The median is:

$$\frac{2.5 + 2.8}{2} = \frac{5.3}{2} = 2.65$$

2.33 The mean and median of a symmetric data set are equal to each other. The mean is larger than the median when the data set is skewed to the right. The mean is less than the median when the data set is skewed to the left. Thus, by comparing the mean and median, one can determine whether the data set is symmetric, skewed right, or skewed left.

2.35 a. $\bar{x} = \frac{\sum x}{n} = \frac{7 + \cdots + 4}{6} + \frac{15}{6} = 2.5$

Median $= \frac{3 + 3}{2} = 3$ (mean of 3rd and 4th numbers, after ordering)

Mode = 3

b. $\bar{x} = \frac{\sum x}{n} = \frac{2 + \cdots + 4}{13} = \frac{40}{13} = 3.08$

Median = 3 (7th number, after ordering)

Mode = 3

c. $\bar{x} = \frac{\sum x}{n} = \frac{51 + \cdots + 37}{10} = \frac{496}{10} = 49.6$

Median $= \frac{48 + 50}{2} = 49$ (mean of 5th and 6th numbers, after ordering)

Mode = 50

2.37 a. Assume that the data represent a sample.

For the Health Care firms:

$$\bar{x} = \frac{\sum x}{n} = \frac{81{,}613 + 61{,}564 + \cdots + 20{,}441}{10} = \frac{389{,}569}{10} = 38{,}956.9$$

The data are arranged in order. Since n is even, the median is the average of the middle two numbers:

$$\text{Median} = \frac{33{,}437 + 33{,}072}{2} = \frac{66{,}509}{2} = 33{,}254.5$$

For the Banks:

$$\bar{x} = \frac{\sum x}{n} = \frac{33{,}329 + 26{,}181 + \cdots + 12{,}877}{10} = \frac{184{,}811}{10} = 18{,}481.1$$