МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
"ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"

# STATISTICS: PRACTICUM

**Exercises on Statistics Course**
**for students of bachelor level in 073 “Management”**
**and 072 “Finance, Banking and Insurance”**

# СТАТИСТИКА: ПРАКТИКУМ

Методичні вказівки до практичних занять з курсу «Статистика»
для студентів спеціальностей
073 «Менеджмент» та 072 «Фінанси, банківська справа та страхування»

До друку дозволяю Мігущенко Р.П.

Харків
НТУ "ХПІ"
2018

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
"ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"

# STATISTICS: PRACTICUM

**Exercises on Statistics Course**
**for students of bachelor level in 073 "Management"**
**and 072 "Finance, Banking and Insurance"**

# СТАТИСТИКА: ПРАКТИКУМ

Методичні вказівки до практичних занять з курсу «Статистика»
для студентів спеціальностей
073 «Менеджмент» та 072 «Фінанси, банківська справа та страхування»


Затверджено
редакційно-видавничою
радою університету,
протокол № 2 від 24.05.2018 р.



Харків
НТУ "ХПІ"
2018

Методичні вказівки до практичних занять з курсу «Статистика» для студентів спеціальностей 073 «Менеджмент» та 072 «Фінанси, банківська справа та страхування» / Уклад. Н.В Ширяєва. – Харків: НТУ «ХПІ», 2018. – 129 с.

Укладач: Н.В. Ширяєва

Рецензент: О.Б. Білоцерківський, к.т.н, доц. НТУ "ХПІ"

Кафедра зовнішньоекономічної діяльності та фінансів

## CONTENT

# ВСТУП

Статистика здійснює збирання, обробку та аналіз даних про масові соціально-економічні явища, які характеризують всі сторони життя та діяльності населення, виявляє взаємозв'язки різних сторін в економіці, вивчає динаміку її розвитку та прийняття ефективних управлінських рішень на всіх рівнях. Важливу роль при цьому відіграє розв’язання задач.

Методичні вказівки до практичних занять містять основи теми, які розглядаються під час вивчення курсу «Статистика», включаючи групування статистичних даних, середні величини, статистичні розподіли, вибіркове спостереження, кореляційно-регрессійний аналіз, оцінювання, ряди динаміки, індекси та їх використання в економіко-статистичних дослідженнях. Кожна тема має типові приклади з розв’язаннями за матеріалом, що вивчається, а також питання та задачі для розв’язку (деякі завдання розв’язуються за допомогою комп’ютеру).

Дані методичні вказівки не замінюють підручники зі статистики. Теоретичні основи викладаються у стислому вигляді. Даються тільки ті відомості, які необхідні безпосередньо для розв’язання завдань. В якості підручників можна використати [1 – 7].

Призначено для студентів спеціальностей 073 «Менеджмент» та 072 «Фінанси, банківська справа та страхування»

# PREFACE

Our world grows in complexity. People often forced to make decisions on one or another problem in the presence of considerable uncertainty. Yet solutions to these problems are essential to our well-being and even our ultimate survival. We are also continually pressured by distressing economic problems such as raising inflation, a cumbersome tax system, and excessive swings in the business cycle. Our entire social and economic fabric is threatened by environmental pollution, a burdensome public debt, an ever-increasing crime rate, and unpredictable interest rates.

Statistics provides collection, processing and analysis of massive social and economic phenomena that characterize all aspects of life and activities of the population, shows the relationship of various aspects of the economy, studying the dynamics of development and effective management decision making at all levels.

The main goal of these exercises is to provide students with methodology of solving economical problems using statistics. The text is written for nonstatisticians and can be used by students in all disciplines. Examples are taken from a wide variety of disciplines that emphasize the concepts and are to the point.

# 1 DESCRIBING DATA SETS. ORGANIZATION AND GROUPING

## 1.1 Solved problems

**Example 1**. A data set has 150 observations. How many classes should be in the frequency table? (**Answer**: $2^c \geq n$, about eight.)

**Example** 2. A data set contains 100 observations, the largest of which is 212 and the smallest of which is 42. Suppose you desire a frequency table with seven classes. (a) What is the class interval? (b) What is the midpoint of the first class if its lower boundary is set at 40?

**Answer:** (a) $CI = \dfrac{212 - 42}{7} = 24.29$. You may decide to round up to 25 for the sake of convenience. (b) If the first class is set with boundaries of 40 and 64, the midpoint is 52, i.e. (40 + 64.3) / 2 = 52.15.

**Example 3. Middle East Uses Oil Revenues to Buy the United States**. In many financial circles there was considerable concern over the acquisitions of Western assets by Middle East oil sheiks during the 1980s. In 1998, *U.S. News& World Report* described the purchase by Saudi Arabia of half-interest in Texaco's refinery and service station operations. Kuwait bought 22 percent of giant British Petroleum for $ 5.5 billion. By 1987, six Persian Gulf nations – Saudi Arabia, Kuwait, Bahrain, Oman, Qatar, ant the United Arab Emirates – owned $ 330 billion in Western assets.

The US Treasury Department collected data on the various acquisitions by several Arab countries during the 1980s. The figures shown in Table 2.1 are in billions of dollars, and represent 42 single purchases of Western interests by Arab governments.

Table 2.1 – Single Purchases of Western Interests by Arab Governments

| | | | | | | |
|---|---|---|---|---|---|---|
| 5.5 | 1.5 | 2.7 | 3.9 | 2.8 | 8.9 | 6.2 |
| 1.9 | 5.3 | 4.7 | 6.1 | 8.3 | 3.5 | 4.7 |
| 2.8 | 4.1 | 3.2 | 5.3 | 1.9 | 4.2 | 5.3 |
| 7.3 | 10.1 | 4.8 | 2.1 | 2.7 | 8.2 | 6.2 |
| 10.9 | 3.9 | 9.2 | 5.1 | 4.2 | 9.7 | 3.9 |
| 8.7 | 1.8 | 1.5 | 9.3 | 5.7 | 8.2 | 7.3 |

Statisticians at the Treasury Department set up a frequency table to facilitate the interpretation of the data.

The *number of classes* is

$$2^c \geq 42 \text{ where } c \approx 6.$$

The *class interval* is

$$CI = \frac{\text{Largest value - Smallest value}}{\text{Number of desired classes}} = \frac{10.9 - 1.5}{6} = 1.6 \approx 1.5.$$

Rounding the interval down to 1.5 for convenience, we could construct Table 2.2. This table requires seven classes in order to tally all observations. If we wished to strictly retain six classes, the interval would have to be increased slightly. Such flexibility is permissible.

Using tools from Chapter 2, one can provide a clear and definite characterization of the data. The message they convey can easily be determined.

Table 2.2 – Western Assets Purchased by Persian Gulf Nations

| **Class** | **Frequency** |
|---|---|
| 1.5 and under 3.0 | 10 |
| 3.0 and under 4.5 | 8 |
| 4.5 and under 6.0 | 9 |
| 6.0 and under 7.5 | 5 |
| 7.5 and under 9.0 | 5 |
| 9.0 and under 10.5 | 4 |
| 10.5 and under 12.0 | 1 |

Table 2.3 – Less-Than Cumulative Frequency Distribution

| **Class** | **Frequency** | **Cumulative Frequency** |
|---|---|---|
| Less than 1.5 | 0 | 0 |
| Less than 3.0 | 10 | 10 |
| Less than 4.5 | 8 | 18 |
| Less than 6.0 | 9 | 27 |
| Less than 7.5 | 5 | 32 |
| Less than 9.0 | 5 | 37 |
| Less than 10.5 | 4 | 41 |
| Less than 12.0 | 1 | 42 |
| Total | 42 | |

Table 2.4 – More - Than Cumulative Frequency Distribution

| Class | Frequency | Cumulative Frequency |
|---|---|---|
| 1.5 or more | 10 | 42 |
| 3.0 or more | 8 | 32 |
| 4.5 or more | 9 | 24 |
| 6.0 or more | 5 | 15 |
| 7.5 or more | 5 | 10 |
| 9.0 or more | 4 | 5 |
| 10.5 or more | 1 | 1 |
| 12.0 or more | 0 | 0 |
| Total | 42 | |

Table 2.8 – Relative Frequency Distribution

| Class | Frequency | Relative Frequency |
|---|---|---|
| 1.5 up to 3.0 | 10 | 10 / 42 = 23.8 % |
| 3.0 up to 4.5 | 8 | 8 / 42 = 19.0 % |
| 4.5 up to 6.0 | 9 | 9 / 42 = 21.4 % |
| 6.0 up to 7.5 | 5 | 5 / 42 = 11.9 % |
| 7.5 up to 9.0 | 5 | 5 / 42 = 11.9 % |
| 9.0 up to 10.5 | 4 | 4 / 42 = 9.5 % |
| 10.5 up to 12.0 | 1 | 1 / 42 = 2.4 % |
| Total | 42 | 100 % |

## 1.2 Conceptual Questions

**1.** Why is it necessary to organize data in some systematic manner after they are collected? Why not just leave them in their raw form in order to preserve their integrity and not violate their true meaning?

**2.** Define and give examples of the methods of organizing data listed below. What are the advantages of each?

*a.* frequency distribution.

*b.* cumulative frequency distribution.

*c.* relative frequency distribution.

*d.* Pie chart.

*e.* Histogram.

**3.** What is the relationship between a frequency polygon and an ogive?

**4.** How will the class limits of a frequency distribution be affected if you are working with continuous data as opposed to discrete data?

**5.** Distinguish between less-than and more-than cumulative frequency distributions. Present a brief example of each.

**6.** Briefly discuss the rules that must be followed in setting class intervals and boundaries for a frequency distribution. What consideration(s) must be given if you are working with continuous data?

### 1.3 Problems

**7.** A particular data set has 130 observations. Approximately how many classes should the frequency distribution contain?

**8.** if the data set in Problem 7 has 53 as its lowest value and 180 as its highest value, what class boundaries would you advice? Construct the entire set of class boundaries for *(a)* discrete data and *(b)* continuous data. Are the intervals that you selected the only possibilities? How might they vary?

**9.** Bill Bissey, vice president of Bank One in Indianapolis, has control over the approval of loans for local business development. Over the past five years the largest loan was $ 1.2 million, and the smallest was $ 10,000. He wishes to construct a frequency table with 10 classes. What would the boundaries of the classes be? What would the class interval be?

**10.** Mr. Bissey also maintains records on personal savings accounts. Of the 40 new accounts opened last month, the current balances are in $

| | | | |
|---|---|---|---|
| 179.80 | 890.00 | 712.10 | 415.00 |
| 112.17 | 1200.00 | 293.00 | 602.02 |
| 1150.00 | 1482.00 | 579.00 | 1312.52 |
| 100.00 | 695.15 | 287.00 | 1175.00 |
| 1009.10 | 952.51 | 1112.52 | 783.00 |
| 1212.43 | 510.52 | 1394.05 | 1390.00 |
| 470.53 | 783.00 | 1101.00 | 666.66 |
| 780.00 | 793.10 | 501.01 | 1555.10 |
| 352.00 | 937.01 | 711.11 | 1422.03 |
| 1595.10 | 217.00 | 1202.00 | 1273.01 |

Construct a frequency table with seven classes. Are you working with continuous data or discrete data? Explain.

**11.** Using the data from Problem 10 construct and interpret a relative frequency table, a cumulative frequency table, and a cumulative relative frequency table.

**12.** Using the data from Problem 10, construct a histogram and a frequency polygon.

**13.** Using the data from Problem 10, construct a bar chart showing the percentages in classes.

**14.** Profits and losses for the 50 largest firms (by sales) on the Fortune 500 list for 1992 are given below in millions of dollars. The lowest value is a loss of 4.453 million (–4453) and the highest is a gain of 5.600 million. Construct a frequency table with the appropriate number of classes.

| | | | | |
|---:|---:|---:|---:|---:|
| –4,453.00 | 1,567.00 | 454.00 | 258.00 | 939.00 |
| 5,600.00 | 1,773.00 | 709.00 | 535.0 | 460.00 |
| –2,258.00 | 1,484.00 | –578.00 | 1,461.00 | –387.00 |
| –2,827.00 | 20.00 | 368.00 | 601.00 | –404.00 |
| 2,636.00 | –1,021.00 | 755.00 | –273.00 | 63.00 |
| 1,920.00 | 1,080.00 | –732.00 | 1,681.00 | 308.00 |
| 3,006.00 | 17.00 | –617.00 | –142.00 | 73.00 |
| 1,403.00 | 311.00 | 1,154.00 | 454.00 | 97.00 |
| 1,294.00 | 942.00 | –1,086.00 | 2,056.00 | 505.00 |
| –795.00 | 423.00 | 184.00 | 97.00 | 1,293.00 |

**15.** Using the data from Problem 14, construct a

*a*. Cumulative frequency distribution.

*b*. Relative frequency distribution.

*c*. Cumulative relative frequency distribution.

**16.** Fortune magazine recently provided a listing of the incomes of chief executive officers of some of the largest US corporations. The annual salaries, in thousands of dollars, were

| | | | | | |
|---:|---:|---:|---:|---:|---:|
| 1,380 | 1,380 | 889 | 1,030 | 1,050 | 1,180 |
| 997 | 1,201 | 920 | 783 | 709 | 883 |
| 842 | 900 | 1,000 | 950 | 970 | 990 |
| 1,273 | 753 | 1,350 | 1,290 | 815 | 1,250 |
| 712 | 850 | 797 | 1,153 | 1,080 | 1,300 |

*a.* Construct the stem-and-leaf design.

*b.* Develop a frequency distribution. Use seven classes. In what way could the stem-and-leaf design be of help in determining class boundaries?

*c.* Construct the corresponding histogram.

*d.* Illustrate and interpret the relative frequency distribution.

*e.* Construct more-than and less-than cumulative frequency distributions.

**17.** As a private economic consultant, you find it necessary to faithfully read *The Wall Street Journal* to remain current in your professional field. A recent report in *WSJ* showed the following data for percentages of executives in 42 top US corporations suffering from drug abuse problems.

| | | | | | | |
|---|---|---|---|---|---|---|
| 5.9 | 8.8 | 14.3 | 8.3 | 9.1 | 5.1 | 15.3 |
| 17.5 | 17.3 | 15.0 | 9.3 | 9.9 | 7.0 | 16.7 |
| 10.3 | 11.5 | 17.0 | 8.5 | 7.2 | 13.7 | 16.3 |
| 12.7 | 8.7 | 6.5 | 6.8 | 13.4 | 5.5 | 15.2 |
| 8.4 | 9.8 | 7.3 | 10.0 | 11.0 | 13.2 | 16.3 |
| 9.1 | 12.3 | 8.5 | 16.0 | 10.2 | 11.7 | 14.2 |

*a.* Construct a stem-and-leaf design.

*b.* Construct a corresponding histogram.

*c.* Construct the frequency distribution and find the class midpoints.

*d.* Construct the relative frequency distribution.

*e.* Construct more-than cumulative frequency distributions and its corresponding ogive.

*f.* Construct the frequency polygon.

**18.** The January 4, 1993, issue of *Business Week* reported the annual percentage changes in several commodity prices. Construct a bar chart showing these changes.

| | |
|---|---|
| Gold | –8.8 % |
| Steel crap | –0.5 % |
| Copper | 0.6 % |
| Aluminum | 8.1 % |
| Wheat | –5.0 % |

**19.** The *PAC 10 Conference Profile* reported the number of players from each conference drafted into the National Football League in 1993 as

| League | Number of players |
|---|---|
| Big 10 | 20 |
| Big 8 | 16 |
| SEC | 30 |
| PAC 10 | 31 |
| ACC | 27 |

*a*. Construct a bar chart using these data.

*b*. Construct a pie chart.

**20.** These data regarding where Americans spend their vacations were recently published in Travel and Leisure; city 31 %; ocean 26 %; lake 5 %; mountains 10 %; state and national parks 6 %; small rural town 22 %. As director of the tourist board for your state, it is your job to present these data in a pie chart.

**21.** In June of 1993, the investment firm of Smith Barney released data on the price per share in US dollars for certain country funds. Construct a stacked bar chart from these data shown below.

| Country | Price per share |
|---|---|
| Malaysia | $ 16.13 |
| Chile | 29.50 |
| Mexico | 22.88 |
| Jakarta | 6.88 |

**22.** In a recent issue of U.S.News&World Report, changes in Italy's GNP from 1981 to 1988 were reported as 0.17 %; –0.52 %; –0.20 %; 2.8 %; 2.3 %; 2.8 %; 2.74 %; 2.7 %.

**23**. The May 18, 1993 issue of the *WSJ* listed the annual sales in millions of dollars of some of the more popular candies. Construct a bar chart and a pie chart.

| | |
|---|---|
| Snickers | $ 36.3 |
| Butterfinger | 20.3 |
| Kit Kat | 22.0 |
| M&M Plain | 18.2 |
| M&M Peanut | 16.5 |

**24.** The US Department of Commerce has reported net US international investment – the level of US investments abroad minus foreign investment in the United States – of the following amounts (in billions of dollars):

| | | | | | |
|---|---|---|---|---|---|
| 1980 | 106.3 | 1984 | 3.6 | 1988 | –577.5 |
| 1981 | 141.1 | 1985 | –111.9 | 1989 | –736.5 |
| 1982 | 197.0 | 1986 | –263.6 | 1990 | –886.5 |
| 1983 | 89.6 | 1987 | –420.5 | 1991 | –1,000.0 |

Construct a bar graph to illustrate these figures, and interpret its meaning.

**25.** Big Al, the local loan shark, currently has 120 outstanding accounts payable. Big Al's accountant informs him that of the 25 accounts in the $ 0 to $ 4.999 range, 10 are due, 5 are overdue, and the rest are delinquent, placing the debtor in danger of being visited by Big Al's enforcer. Of the 37 in the $ 5,000 to $ 9,999 range, 15 are due, 10 are overdue and the rest are delinquent. There are 39 in the $ 10,000 to $ 14,999 that show 11 are due, 10 are overdue and the rest are delinquent. On the remaining accounts, in the $ 15,000 and up range, 5 are due, 7 are overdue, and the rest are delinquent. Big Al wants to see a contingency table for these accounts. Interpret their significance by citing a few of the statistics you fell are most important and revealing.

**26.** Construct a frequency table for the advertising budgets of top US corporations. These figures are in thousands of dollars.

| | | | | | |
|---|---|---|---|---|---|
| 51.02 | 79.82 | 82.22 | 92.29 | 60.00 | 72.14 |
| 63.09 | 99.99 | 77.77 | 59.99 | 60.00 | 65.72 |
| 83.21 | 71.11 | 73.99 | 89.89 | 89.99 | 89.49 |
| 69.99 | 50.00 | 58.73 | 79.99 | 80.00 | 90.00 |
| 81.00 | 57.03 | 70.00 | 61.11 | 80.00 | 69.99 |

Put five classes in the table, with intervals of $ 10, and begin with $ 50 as the lower boundary of the first class.

**27.** Construct a stem-and-leaf design for the data in Problem 26.

**28**. These closing prices of randomly selected stocks listed on the NYSE were taken from *The WSJ*:

| | | | | | |
|---|---|---|---|---|---|
| 20 1/8 | 18 7/8 | 25 3/8 | 18 3/4 | 18 1/8 | 22 5/8 |
| 19 3/8 | 19 3/8 | 19 5/8 | 21 1/8 | 21 1/2 | 21 7/8 |
| 25 3/4 | 22 1/8 | 17 1/2 | 18 7/8 | 24 5/8 | 22 1/2 |
| 17 1/4 | 18 1/2 | 25 7/8 | 21 7/8 | 20 1/8 | 20 1/4 |

*a.* How many classes should the frequency table contain?

*b.* What is the proper class interval?

*c.* Construct the frequency table.

**29.** Monthly sales for Luggit Hardware last year were (in hundreds of dollars)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Jan | 42 | April | 51 | July | 63 | Oct | 58 |
| Feb | 57 | May | 53 | Aug | 79 | Nov | 62 |
| March | 49 | June | 54 | Sept | 69 | Dec | 70 |

Construct a line chart for these data.

**30.** The results of a quality control study on the number of "bugs" in 80 software projects at NASA's Goddard Space Center was reported in the December 4, 1992, issue of Business Week. The average defects over a 15-year period were

| Year | Defects |
|---|---|
| 1976 | 8.5 |
| 1978 | 8.0 |
| 1980 | 7.5 |
| 1982 | 7.0 |
| 1984 | 6.4 |
| 1986 | 6.0 |
| 1988 | 5.6 |
| 1990 | 5.0 |

Construct a line chart from these data. What can you conclude from the chart?

**32.** On April 2, 1993, the US Bureau of Census reported the level of federal benefits programs in billions of dollars for 1991 and 1992 as

| **Federal benefits programs** | **1991** | **1992** |
|---|---|---|
| Social security | $ 281.9 | 263.9 |
| AFDC | 15.5 | 13.5 |
| SSI | 17.9 | 15.3 |
| Food stamps | 20.0 | 17.0 |
| Medicare | 147.9 | 117.9 |

Using these data, construct a

*a.* Bar chart.

*b.* Pie chart.

**33.** The Dow Jones averages for 30 industrial stocks reported the following values in June of 1988. Construct a high-low-close chart based on these data.

| **Date** | **High** | **Low** | **Close** |
|---|---|---|---|
| June 9 | 2119.31 | 2081.79 | 2093.35 |
| June 10 | 2123.58 | 2084.82 | 2101.71 |
| June 13 | 2144.15 | 2084.64 | 2099.40 |
| June 14 | 2148.12 | 2111.13 | 2124.47 |

**34.** Each spring, *Fortune* magazine reports the 500 largest US industrial corporations ranked on the basis of sales. In 1992, the largest firm, General Motors, had sales of $ 123,780.1 million, and the 500th largest firm was Guilford Mills, which had sales of $ 528.8 million. Determine how many classes should be formed if you wished to construct a frequency table.

**35.** What is the class interval for the frequency table from Problem 34?

**36.** Refer to Problem 34. It the lower boundary of the first class was set at $ 529 million, what values would the remaining boundaries take? Illustrate by constructing the entire table.

**37.** The observations listed are times (in minutes) that 30 students took to complete their first statistics test.

| | | | | | |
|---|---|---|---|---|---|
| 42.3 | 39.2 | 58.9 | 42.7 | 38.9 | 39.2 |
| 70.0 | 67.7 | 45.5 | 69.1 | 52.4 | 68.3 |
| 37.2 | 52.6 | 53.3 | 55.5 | 68.3 | 52.5 |

| | | | | | |
|---|---|---|---|---|---|
| 69.2 | 63.2 | 61.9 | 63.9 | 61.2 | 64.9 |
| 41.9 | 39.2 | 45.7 | 41.7 | 70.1 | 69.8 |

Determine the appropriate number of classes.

**38.** What should the class interval be for Problem 37?

**39.** Construct the frequency table based on data in Problem 37. Tally the observations and record the frequency in each class.

**40.** Construct a more-than and less-than cumulative frequency distribution from the data in Problem 37.

**41**. Construct a relative frequency distribution from Problem 37.

**42.** Construct a histogram, frequency polygon and a more-than ogive for Problem 37.

**43**. The US Travel Data Center for *Better Homes and Gardens* reported the following numbers of people who took vacations last summer based on the structure of family members. Numbers are in millions of adults.

| Vacation type | Adults with Children | Adults w/o Children |
|---|---|---|
| Weekend trips | 59 | 45 |
| Nonweekend short trips | 30 | 21 |
| Nonweekend long trips | 39 | 31 |

Construct a single bar chart illustrating these values.

**44.** Present the information in Problem 47 with a stacked bar chart.

**45**. The Bureau of Labor Statistics in the US Department of Labor produced a report in a 1988 edition of Newsweek that revealed certain economics trends. It showed that in 1969 the percentages of families in the lower, middle, and upper social classes were 32 percent, 58 percent, and 10 percent, respectively. In 1986, the figures were 29 percent, 53 percent, and 18 percent. Construct pie charts for these data. What conclusion concerning the mobility between social classes does the chart suggest?

**46**. A January 1993 issue of *Business Week* listed the price-earnings ratios for several firms in the leisure-time industry. Summarize these ratios in a pie chart.

| | |
|---|---|
| Walt Disney | 28 |
| Circus Circus | 25 |
| Carnival Cruise Lines | 16 |
| Harley-davidson | 25 |
| Musicland stores | 15 |

**47.** Unemployment rates over the last 15 months were reported by the US Department of Labor as 4.3; 4.7; 5.4; 5.5; 5.5; 5.6; 5.7; 5.3; 5.7; 6.1; 6.5; 7.2; 7.0; 7.1; 6.8 percent. Construct stem-and-leaf design.

**48.** As a class project, a junior marketing student surveyed 20 local businesses in Peoria, Illinois, concerning their preferences for a new product. Their responses were recorded as a “1” if they liked the product, a “2” if they disliked it, and “3” if they had no opinion. Annual sales levels for the stores were also recorded as a

| | |
|---|---|
| 1 | If sales were less than $ 50,000 |
| 2 | If sales were $ 50,000 but less than $ 100,000 |
| 3 | If sales were $ 100,000 but less than $ 200,000 |
| 4 | If $ 200,000 or more |

Construct a contingency table based on the data shown here. What conclusions can you reach?

| **Opinion** | **Sales** | **Opinion** | **Sales** |
|---|---|---|---|
| 1.00 | 4.00 | 3.00 | 1.00 |
| 1.00 | 4.00 | 2.00 | 1.00 |
| 3.00 | 3.00 | 3.00 | 2.00 |
| 1.00 | 4.00 | 3.00 | 4.00 |
| 3.00 | 1.00 | 1.00 | 4.00 |
| 3.00 | 1.00 | 1.00 | 4.00 |
| 3.00 | 1.00 | 1.00 | 4.00 |
| 1.00 | 2.00 | 3.00 | 4.00 |
| 2.00 | 3.00 | 2.00 | 4.00 |
| 1.00 | 4.00 | 3.00 | 1.00 |

### 1.4 Empirical Exercises

**1.** Survey 40 people (20 males and 20 females) on how they feel about the economic policies of the Ukraine (or other country) president. Use the following categories: excellent, above average, average, below average, and poor. Construct a histogram showing the difference in opinion between males and females.

**2.** From the financial pages of The Wall Street Journal, or any newspaper reporting the activities of the New York Stock Exchange, collect closing prices on 50 to 100 randomly selected stocks. Report the results in a frequency table, showing the cumulative frequency, relative frequency, and cumulative relative frequency for each class. Justify your choice of class interval and the number of classes. Interpret our results. Assuming the stocks you selected were representative of the trading on the NYSE that day, what conclusion can you draw? Construct a histogram based on the data.

### 1.5 Computer Exercises

**1.** Data for the profits (in thousands of dollars) and the earnings per share for 32 firms listed on the New York Stock Exchange are shown here. Using your computer package, create a frequency table, a histogram, and a stem-and-leaf. Comment on the results. What information can be gathered?

| Profit | EPS | Profit | EPS | Profit | EPS | Profit | EPS |
|---|---|---|---|---|---|---|---|
| 4556. | 0.56 | 7653.00 | 2.30 | 5247..00 | 0.67 | 6453.00 | 5.10 |
| 6543.00 | 1.20 | 7424.00 | 0.65 | 6543.00 | 7.20 | 1636.00 | 5.40 |
| 7653.00 | 2.40 | 645.00 | 3.20 | 6554.00 | 3.20 | 6553.00 | 0.54 |
| 76.00 | 3.20 | 5434.00 | 2.70 | 6745.00 | 1.50 | 7635.00 | 1.80 |
| 6345.00 | 6.40 | 6345.00 | 3.20 | 984.00 | 4.20 | 1734.00 | 2.90 |
| 635.00 | 4.20 | 7624.00 | 2.90 | 663.00 | 2.50 | 2514.00 | 0.87 |
| 6455.00 | 0.98 | 6543.00 | 1.50 | 8734.00 | 1.60 | 763.00 | 2.40 |
| 534.00 | 1.40 | 234.00 | 0.20 | 546.00 | 2.50 | 2974.00 | 1.30 |

# 2 MEASURES OF CENTRAL TENDENCY AND DISPERSION

## 2.1 Solved problems

**Example 1.** Denny is rightfully worried about his grade in statistics. He scored the following grades on the five tests given this semester: 63, 59, 71, 41, 32. His professor has warned Denny that any grade below 60 is failing. Calculate and interpret Denny's mean grade.

**Solution:** Denny's rather disappointing average can be calculated as

$$\mu = \frac{\sum X_i}{N} = \frac{63+59+71+41+32}{5} = \frac{266}{5} = 53.2.$$

**Interpretation:** Since all five grades were used in the calculation, 53.2 is the mean of the population and is therefore a parameter. Obviously, he's in trouble!

**Example 2.** The CEO for White-Knuckle Airlines wishes to determine the average growth rate in revenue based on the figures in the Table 5.1. If the average growth rate is less that the industry average of 10 percent, a new advertising company will be undertaken.

Table 3.1 – Revenues for White-Knuckle Airlines

| **Year** | **Revenue, $** | **Percentage of Previous Year** |
|---|---|---|
| 1992 | 50,000 | – |
| 1993 | 55,000 | 55 / 50 = 1.10 |
| 1994 | 66,000 | 66 / 55 = 1.20 |
| 1995 | 60,000 | 60 / 66 = 0.91 |
| 1996 | 78,000 | 78 / 60 = 1.30 |

**Solution**. It is first necessary to determine what percentage each year's revenue is of the previous year. This is found by dividing revenues in 1993 by those in 1992. The result 1.10 reveals that 1993 revenues are 110 percent of revenues in 1992. Taking the *geometric mean* of these percentages gives

$$GM = \sqrt[4]{1.1 * 1.2 * 0.91 * 1.3} = 1.1179.$$

Subtracting “1” in order to convert to an average annual increase yields 0.1179, or an 11.79 percent average increase over the five-year-period.

The *simple arithmetic average* is

$$\overline{X} = \frac{1.1 + 1.2 + 0.91 + 1.3}{4} = \frac{4.51}{4} = 1.1275.$$

or a 12.75 percent average change. Division by “4” was because there were four changes over the five-year-period.

However, if an average increase of 12.75 percent based on the simple arithmetic average is applied to the series starting with \$ 50,000, the results are

\$ 50,000 * 1.1275 = \$ 56,375;
\$ 56,375 * 1.1275 = \$ 63,563;
\$ 63,563 * 1.1275 = \$ 71,667;
\$ 71,667 * 1.1275 = \$ 80,805.

Since \$ 80,805 exceeds the \$ 78,000 White-Knuckle Airlines actually earned, the 12.75 percent increase is obviously too high. On the other hand, if the geometric mean growth of 11.79 percent is used, we get

\$ 50,000 * 1.1179 = \$ 55,895;
\$ 55,895 * 1.1179 = \$ 62,485;
\$ 62,485 * 1.1179 = \$ 69,852;
\$ 69,852 * 1.1179 = \$ 78,088 ≈ \$ 78,000.

This gives a value of \$ 78,088, which is much closer to the actual revenue figure of \$ 78,000.

**Interpretation:** The geometric mean represents the average change over time. Since the growth rate exceeds the industry average of 10 %, the new advertising campaign will not be undertaking.

**Example 3.** In a recent issue of *Fortune,* 30 companies out of the 500 largest US corporations reported revenues in millions of dollars as shown in the frequency table.

Table 3.2 – Revenues of 30 US Corporations (in millions of dollars)

| **Class** | ***f*** | ***M*** | ***F*** | ***fM*** |
|---|---|---|---|---|
| 20 up to 30 | 5 | 25 | 5 | 125 |
| 30 up to 40 | 2 | 35 | 7 | 70 |
| 40 up to 50 | 4 | 45 | 11 | 180 |
| 50 up to 60 | 8 | 55 | 19 | 440 |
| 60 up to 70 | 7 | 65 | | 455 |
| 70 up to 80 | 2 | 75 | | 150 |
| 80 up to 90 | 1 | 85 | | 85 |
| 90 up to 100 | 1 | 95 | | 95 |
| | 30 | – | | 1,600 |

As an economic analyst for a large Wall Street brokerage firm, you must report to your office manager before the end of the day's trading on the New York Stock Exchange the following statistics for these revenue figures: *(a)* the mean, *(b)* the median, and (c) the mode.

**Solution:**

a) the mean: $\overline{X}_{grouped} = \frac{\sum X_i f_i}{\sum f_i} = \frac{\sum M f_i}{\sum f_i} = \frac{1{,}600}{30} = \$53.33\,(\text{million});$

b) the median:

$$Median = L_{md} + \left[\frac{n/2 - F}{f_{md}}\right] \mathbf{C} = 50 + \left[\frac{15-11}{8}\right](10) = \$55\,(\text{million});$$

c) the mode: $Mode = L_{mo} + \left[\frac{D_a}{D_b + D_a}\right] \mathbf{C} = 50 + \frac{4}{5}(10) = \$58\,(\text{million}).$

**Interpretation:** The values serve as estimates of central tendencies for the revenue data. Your manager can now gain insight into the nature of the earnings of these firms. Although all three are representative of the center point, your manager may prefer one of them.

**Example 4. Statistical Application.** Establishing company policy to deal with sick leave is an important issue facing many business firms. According to an article in *Industry World*, DuPont Chemical recently had to devise a policy which would offer its workers ample time off from work in the event of illness, yet not prove to be overly burdensome to the firm. It was felt that ' 'too little was known about the frequency and duration of illnesses of DuPont employees to develop an adequate plan."

A committee made up of labor and managerial personnel collected data on the number of workers who requested sick days and on the duration of their sick leave. There were over 2,000 employees in the study. The data were “placed into a manageable form” (most likely, a frequency table) from which it was possible to determine such things as the average length of time a worker was absent, the average number of workers on sick leave at any time, as well as other relevant factors to be considered when deciding on a proper sick leave policy.

**Example 5.** Mr. Boggs wishes to determine the stability of the price of a particular stock. He decides to base his judgment regarding stability on the standard deviation of the stock's daily closing price. Checking the financial pages, Boggs learns that the stock has been traded on the exchange for quite some time and there are many closing prices dating back several months. Rather than using all these prices, Boggs decides to simplify his arithmetic and select a random sample of $n$ = 1 days. (Although 7 is probably too small a sample, it will serve our purposes for the moment.) He notes the closing prices of \$ 87, \$ 120, \$ 54, \$ 92, \$ 73, \$ 80, and \$ 63.

**Solution**: Then, $\overline{X}$ = \$ 81.29, and

$$s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1} = \frac{\begin{array}{c}(87-81.29)^2 + (120-81.29)^2 + (54-81.29)^2 + \\ + (92-81.29)^2 + (73-81.29)^2 + (80-81.29)^2 + (63-81.29)^2\end{array}}{7-1},$$

$s^2$ = 465.9 dollars squared,

$$s = \sqrt{465.9} = 21.58 \text{ dollars.}$$

**Interpretation:** Boggs has estimated the mean closing price of the stock to be $ 81.29, with a tendency to vary above and below that price by $ 21.58. A further explanation of the use and interpretation of the standard deviation is offered later in this chapter. However, keep in mind that Boggs can always interpret the standard deviation of $ 21.58 as a measure of the tendency of the closing prices to fluctuate around their mean of $ 81.29.

**Example 6.** Recall that, using data from P&P's passenger list, we calculated $\overline{X}$ =78.3, $s$ = 10.8, and the *Median* = 77.9. Given this information, the CEO for P&P can plainly see that the data are skewed right, since the mean exceeds the median. In addition, he also wants a measure of the extent or degree of skewness.

**Solution:** We have

$$P = \frac{3(78.3 - 77.9)}{10.8} = 0.11.$$

**Interpretation:** Since $P$ > 0, the data for P&P are, as presumed, skewed right. The extent to which they are skewed is reflected in the value of the Pearsonian coefficient. If we were to graph the data, they would appear as in Figure 1.

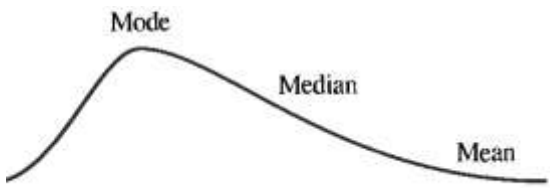



Figure 1 – Distribution of People's Weights skewed right

**Example 7.** In Example 3.5, Markus Boggs computed the average price of a certain stock to be $ 81.29, with a standard deviation of $ 21.58. Assume that a second stock has recorded the following closing prices: $ 147, $ 120, $ 115, $ 110, $ 100, $ 73, and $ 105. He wishes to compare the variation in the prices of these two stocks.

**Solution:** Markus calculates the standard deviation of the second stock as follows:

| X | $X^2$ |
|---|---|
| 147 | 21,609 |
| 120 | 14,400 |
| 115 | 13,225 |
| 110 | 12,100 |
| 100 | 10,000 |
| 73 | 5,329 |
| 105 | 11,025 |
| 770 | 87,688 |

$$\overline{X} = \frac{770}{7} = \$110;$$

$$s^2 = \frac{\sum X^2 - n\overline{X}^2}{n-1} = \frac{87{,}668 - 7(110)^2}{6} = 498;$$

$$s = \sqrt{498} = \$22.30 > \$21.58.$$

Since the second stock has a higher standard deviation, Markus might be inclined to conclude that it exhibits a greater degree of dispersion. However, that decision would belie the true nature of the two data sets. He should notice that the second data set also has a larger mean and that, in general, he is working with larger numbers in the second data set. No wonder the standard deviation is larger. A more accurate depiction of the comparable degree of dispersion can be formed by relating each standard deviation to its mean. This produces the coefficient of variation:

$$CV = \frac{s}{\overline{X}}(100).$$

Then CV for first and second stocks are calculated as following

$$CV_1 = \frac{21.58}{81.29}(100) = 26.55\ \%,$$

$$CV_2 = \frac{22.53}{110}(100) = 20.27\ \%.$$

**Interpretation:** Despite the larger standard deviation of the second stock, its prices, relative to their mean, deviate less than do the prices of the first stock relative to their mean. Markus can therefore conclude that the price of the second stock is less volatile than the first in a relative sense.

Furthermore, since the coefficient of variation is expressed as a percentage, it can be used to compare measures of dispersion expressed in different units. It's possible to compare the variation in sales revenue measured in dollars with labor inputs measured in numbers of employees with the coefficient of variation.

## 2.3 Exercises

**1**. The Snowflake markets ski boots in San Luis Obispo, California. Of the last 100 pairs sold, 4 were size 9, 33 were size 9 1/2, 26 were size 10, 29 were size 10 1/2, and 8 were size 13. Comment on the use of the mean, median, and mode as measures of central tendency and the use of each in making decisions regarding sizes to hold in inventory. Calculate each measure.

**2.** Are the observations in Problem 9 normal, skewed right, or skewed left? Explain. Draw the distribution, label the axes, and show the placement of the three measures of central tendency.

**3.** The football coach for State University recorded the yardage gained (or lost) on the first down by his offense in last Saturday's game. The values were

7; 12; –5; 4; 0; –7; 2; –3.

Calculate the mean, median, and mode.

**4**. As interest rates fell in early 1993, a sample of mortgage rates for 15-year mortgages at local lending institutions in Peoria. Illinois, was found to be

7.1 %; 7.3 %; 7.0 %; 6.9 %; 6.6 %; 6.9 %; 6.5 %; 7.3 %; 6.85 %

*a.* Calculate and interpret the mean, median, and mode.

*b.* Are these data skewed left, or right, or are they normally distributed? Calculate the Pearsonian coefficient as a measure of the skewness.

*c.* Calculate and interpret the variance and the standard deviation.

**5.** A survey of lending institutions in an urban center near Peoria (see previous problem) revealed mortgage rates of

7.1 %; 7.3 %; 6.3 %; 6.7 %; 6.8 %; 6.85 %; 7.5 %.

*a.* Are mortgage rates higher in Peoria or the other urban center?

*b.* Which city seems to have the most consistent rates among institutions?

*c.* Calculate and interpret the Pearsonian coefficient of skewness.

**6.** Alan Munday manufactures a paint sealant for automobiles in the Denver area. He uses four different chemicals in the production process. To make his product, Munday must use 2 gallons of calcimine which costs $ 2.50 per gallon, 1/2 gallon of kalsolite at $ 1.25 per gallon, 1 gallon of binder costing $ 0.75 per gallon, and 3 gallons of drying oil at $ 2.00 per gallon. Calculate the cost of a gallon of the sealant.

**7.** Monthly sales in hundreds of dollars at Fast Freddie's Used Car Emporium were

| | | | |
|---|---|---|---|
| January | 69 | April | 74 |
| February | 73 | May | 79 |
| March | 78 | June | 82 |

Fast Freddie hopes to achieve sales of 87 by August. Will he do so if sales continue to increase at the same rate?

**8.** The May 31, 1993, issue of *Business Week* reported that the number of transactions in billions performed at the nation's ATM banking facilities were

| | | | |
|---|---|---|---|
| 1987 | 3.9 | 1990 | 4.5 |
| 1988 | 4.1 | 1991 | 6.5 |
| 1989 | 4.3 | 1992 | 6.5 |

The banking industry intends to prepare for 8 billion transactions by 1994. Will that be sufficient to handle the level of activity that you predict for that year?

**9.** As owner of a prospering advertising agency in Chicago, George Kay earns $ 110,000 a year. His seven newest employees make $ 15,000, $ 21,000, $ 18,500, $ 17,900, $ 21,200, $ 15,900, and $ 22,500. Which measure of central tendency do you think is the best indication of the average of all eight incomes? Calculate the average.

**10.** The Noah Fence Company sells four types of fencing to local suburbanites. Grade A costs Noah $ 5.00 per running foot to install, grade B costs $ 3.50, grade C costs $ 2.50, and grade D costs $ 2.00. Yesterday, Noah installed

100 yards of A, 150 yards of B, 75 yards of C, and 200 yards of D. What was Noah's average installation cost per foot?

**11.** A recent account of investment opportunities surveyed by *Fortune* magazine revealed that typical investments in stocks yielded a 7.2 percent return, money held in bonds provided an 8.1 percent yield, funds placed in real estate generated a 3.7 percent yield, and assets held in the form of cash yielded a 5.5 percent return in a savings or money market account. John Wise, according to *Fortune*, holds 40 percent of his wealth in stocks, 40 percent in bonds, and the rest in cash. Lars Holton maintains a portfolio consisting of 34 percent stocks, 10 percent real estate, 16 percent cash, and 40 percent bonds. Which investor has the higher average rate of return?

**12.** A sample of weekly sales receipts for Pig-In-A-Poke Bar-B-Q are, in hundreds of dollars,

43.3; 54.2; 34.8; 42.9; 49.2; 29.5; and 28.6.

An advertising program designed to even out sales is implemented. A subsequent sample of sales proves to be

45.5; 39.5; 35.7; 36.7; 42.6; 42.14.

Did the advertising campaign achieve its goal of smoothing weekly sales?

**13.** From the money he won in the state lottery, Lucky Louie purchased one $ 10,000 corporate bond, a $ 5,000 federal security, and deposited $ 7,000 and $ 8,000, respectively, in two money market accounts. The four investments yield returns of 12, 7, 8, and 9 percent, respectively. What is Lucky's average rate of return?

**14.** *Time* magazine reported that Jim Maxey, a former journalist and part-time private eye in California, runs a computerized on-line networking system for hackers who wish to meet other computer jocks. Maxey's records show that clients spend an average of 7.8 hours per week on the system. There are currently 453 clients. Assume the standard deviation in lengths of time is 0.6 hours. At least how many clients should Maxey plan will use the system for between 6.6 hours and 9 hours?

**15.** Answer the previous problem if it is assumed that times are normally distributed. Draw a graph illustrating your responses.

**16.** Bill Karl purchased 20 shares of stock at $ 15 each, 50 shares at $ 20, 100 shares at $ 30, and 75 shares at $ 35.

*a.* What is the total amount of his investment?

*b.* What is the average price per share?

**17.** The ages of 50 of the nation's CEOs of top corporations reported in the May 24, 1993, issue of *Forbes* produced the following frequency table.

*a.* Calculate and interpret the mean, median, and mode.

*b.* Calculate and interpret the variance and the standard deviation.

| **Ages** | **Frequency** |
|---|---|
| 50 and under 55 | 8 |
| 55 and under 60 | 13 |
| 60 and under 65 | 15 |
| 65 and under 70 | 10 |
| 70 and under 75 | 3 |
| 75 nd under 80 | 1 |

**18.** The same issue of *Forbes* (see the previous problem) also provided data on salaries in thousands of dollars. The following frequency table resulted:

| **Salary (in $ 1,000's)** | **Frequency** |
|---|---|
| 90 and under 440 | 9 |
| 440 and under 790 | 11 |
| 790 and under 1140 | 10 |
| 1140 and under 1490 | 8 |
| 1490 and under 1840 | 4 |
| 1840 and under 2190 | 3 |
| 2190 and under 2540 | 5 |

*a.* Calculate the mean, median, and mode. Interpret your answers.

*b.* Are salaries as dispersed as ages in the previous problem?

**19.** Janna Vice uses two different machines to produce paper chutes for Kodak copiers. A sample of the chutes from the first machine measured 12.2, 11.9, 11.8, 12.1, 11.9, 12.4, 11.3, and 12.3 inches. Chutes made with the second machine measured 12.2, 11.9, 11.5, 12.1, 12.2, 11.9, and 11.8 inches. Janna must use the machine with the greater consistency in sizes of the chutes. Which machine should she use?

**20.** Helen Highwater is director of personnel services for a large bank in Chicago. She must hire a secretary based on typing efficiency. One candidate for the job typed a manuscript six times with the following number of mistakes: 5,6, 2, 1,2, 0. Another candidate typed the same manuscript six times with 3, 4, 5, 3, 4, and 5 mistakes. Which candidate should Helen hire?

**21.** Scores on the first two sets of statistics tests you took were normally distributed and reported means of 90 for Test A and 50 for Test B. Would you hope for a high or a low standard deviation for Test A? Would you want a high or low standard deviation for Test B if you feel that you did well on the test? Why? Draw a graph illustrating the logic in your responses.

**22.** The following sample data have been obtained for the number of daily customers at Rosie's Flower Shoppe:

34, 45, 23, 34, 26, 32, 31, 41.

Calculate the variance, the standard deviation, and the interquartile range.

**23.** The following is a sample of the earnings per share, in dollars, for stocks listed on the New York Stock Exchange:

1.12, 1.43, 2.17. –1.19, 2.87, –1.49.

*a.* Calculate the variance and the standard deviation.

*b.* Interpret the standard deviation.

**24.** A sample of the daily number of newspapers sold by the *Kansas City Star* (in hundreds) is shown for the last 50 days. Construct a frequency table with five classes and compute

*a.* the mean.

*b.* the median.

*c.* the variance and the standard deviation.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 615 | 778 | 500 | 712 | 742 | 615 | 663 | 783 | 715 | 555 |
| 635 | 597 | 573 | 753 | 687 | 579 | 689 | 643 | 701 | 595 |
| 643 | 783 | 599 | 654 | 792 | 537 | 673 | 678 | 743 | 687 |
| 693 | 711 | 749 | 553 | 543 | 640 | 683 | 772 | 559 | 642 |
| 743 | 782 | 692 | 593 | 718 | 622 | 590 | 672 | 778 | 735 |

**25.** Thirty-five "typical" households were surveyed regarding their estimate of the increase in their taxes from President Clinton's proposed economic package in the summer of 1993. Their responses were reported in a local paper and are shown here in $ :

| | | | | | | |
|---|---|---|---|---|---|---|
| 812 | 735 | 739 | 489 | 1015 | 945 | 875 |
| 689 | 589 | 499 | 1010 | 658 | 499 | 908 |
| 956 | 872 | 521 | 512 | 758 | 658 | 825 |
| 1000 | 965 | 865 | 498 | 1009 | 687 | 851 |
| 625 | 729 | 954 | 687 | 745 | 953 | 857 |

*a.* Construct a frequency table with the proper number of classes and an interval of $ 85. Set the lower boundary of the first class at 485.

*b.* Compute and interpret the mean, median, and mode, the variance, and the standard deviation.

**26**. For the following sample of 24 daily observations for the number of miles, rounded to the nearest one-tenth, driven by Ronnie Roadrunner on his job as a traveling salesman,

| | | | | | |
|---|---|---|---|---|---|
| 100.3 | 117.7 | 101.1 | 127.3 | 121.9 | 119.1 |
| 122.7 | 98.9 | 125.7 | 130.7 | 122.3 | 105.5 |
| 93.4 | 115.3 | 98.1 | 120.1 | 97.2 | 99.4 |
| 112.0 | 97.3 | 95.3 | 101.3 | 129.7 | 109.1 |

Construct a frequency table with six classes and calculate the

*a.* mean.

*b.* median.

*c.* variance and standard deviation.

**27.** The number of hours worked each week by Ronnie over the past two months are

52 48 37 54 48 15 42 12

Assuming these are sample data, calculate the

*a.* mean.

*b.* median.

*c.* mode.

*d.* Which is probably a better measure of the center point?

**28.** Using Ronnie's work hours from the previous problem, calculate and interpret the

*a.* range.

*b.* variance.

*c.* standard deviation.

*d.* mean absolute deviation. What is the drawback of MAD?

*e.* first quartile.

*f.* 25th percentile.

*g.* interquartile range.

**29.** Data on the frequency of absenteeism in a plant owned by Novell Electronics are shown in the table. Management retains your statistical expertise to analyze their employees' behavior as compared with national norms. Studies show that the average number of days per year that employees are absent across the nation for similar plants is about 27. Help Novell evaluate and discern the meaning of the figures. Calculate the mean, median, and mode, and comment.

| **Days Absent per Year** | **Frequency** |
|---|---|
| 0 – 10 | 4 |
| 11 – 21 | 10 |
| 22 – 32 | 12 |
| 33 – 43 | 15 |
| 44 – 54 | 10 |
| 55 – 65 | 5 |
| 66 and over | 4 |
| Total | 60 |

**30.** The disc jockeys on KAYS claim they play more songs each hour than their crosstown rivals on KROC. Over the last 24 hours, data on the number of songs played for both stations were collected and tabulated. Use the data to prepare a report comparing the two stations. Your finished report is to be submitted to the Federal Communications Commission and contain references to measures of central tendency and measures of dispersion.

| Number of Hits per Hour | KAYS | KROC |
|---|---|---|
| 5 – 10 | 2 | 4 |
| 11 – 16 | 4 | 5 |
| 17 – 22 | 6 | 7 |
| 23 – 28 | 8 | 5 |
| 29 – 34 | 2 | 2 |
| 35 – 40 | 2 | 1 |

**31.** Quarterly losses by 50 firms were placed in this frequency table.

| Losses (in $ 1,000's) | Frequency (number of firms) |
|---|---|
| 50 and under 100 | 2 |
| 100 and under 150 | 8 |
| 150 and under 200 | 15 |
| 200 and under 250 | 10 |
| 250 and under 300 | 12 |
| 300 and under 350 | 1 |
| 350 and under 400 | 2 |

*a.* Calculate and interpret the mean, median, and mode.

*b.* Calculate and interpret the variance and the standard deviation.

**32.** The June 1993 issue of *Prevention* magazine measured the reduction in stress levels of people who used a treadmill for 30 minutes a day at least four times each week. These reductions are shown here in a frequency table.

| Reduction in Stress Level | Frequency (number of exercises) |
|---|---|
| 10 and under 15 | 3 |
| 15 and under 20 | 4 |
| 20 and under 25 | 7 |
| 25 and under 30 | 10 |
| 30 and under 35 | 15 |
| 35 and under 40 | 8 |
| 40 and under 45 | 5 |

*a.* Can the average person expect to decrease their stress level by 25 points'?

*b.* What variation in stress reduction might exist from one person to the next?

**33.** A report in a recent issue of *Fortune* magazine described the increased use of fax machines in daily business activity. It stated that certain financial offices tend to fax documents on the average of “15 to 20 times each day.” As office director, Pam Whiting is concerned about the rising costs associated with the fax machine in her office. The results of a sample of the number of times her office faxes documents each day over a 66-day period are tabulated.

*a.* Calculate the mean, median, and mode, and comment on whether Pam has cause for such concern.

*b.* Compute the variance and standard deviation to obtain a measure of the dispersion in the daily use of the fax.

| **Class (number of documents)** | **Frequency (number of days)** |
|---|---|
| 0 and under 5 | 8 |
| 5 and under 10 | 7 |
| 10 and under 15 | 12 |
| 15 and under 20 | 10 |
| 20 and under 25 | 19 |
| 25 and under 30 | 15 |

**34.** *The Wall Street Journal* described a dispute between management and the local labor union regarding the efficiency and productivity of the workers. Management argued that it was taking the employees more than 20 minutes to complete a certain job task. If 85 employees are timed, yielding the results tabulated, is management correct based on this sample? Compute all three measures of central tendency.

| **Class (number of minutes)** | **Frequency (number of employees)** |
|---|---|
| 5 and under 7 | 2 |

| | |
|---|---|
| 7 and under 9 | 8 |
| 9 and under 11 | 10 |
| 11 and under 13 | 15 |
| 13 and under 15 | 17 |
| 15 and under 17 | 14 |
| 17 and under 19 | 7 |
| 19 and under 21 | 9 |
| 21 and under 23 | 3 |

**35.** Management from the previous problem is also worried that employees' performance is too erratic; there is too much variation in the amount of time it takes the workers to complete the task. Identify and compute the statistic that would address management's concern.

**36.** Using the data from the two previous problems, compute the interquartile range.

**37.** *Forbes* magazine presented an article about a British financial and investment consortium that was buying up racehorses in Lexington, Kentucky. Concern within the consortium focused on the depressed nature of racetracks around the country, and the fact that betting at racetracks is down to a 20-year low of only $ 13.9 billion. The winnings of 88 thoroughbreds are classified here. The consortium wants an average of $ 100,000 for each horse it owns.

*a.* Do these data suggest the consortium is obtaining that goal? Calculate all three measures of central tendency, and comment.

*b.* Calculate the proper measure of variation in the horses' earnings.

| **Winnings ($ 1,000's)** | **Frequency (number of horses)** |
|---|---|
| 10 and under 30 | 8 |
| 30 and under 50 | 7 |
| 50 and under 70 | 12 |
| 70 and under 90 | 15 |
| 90 and under 110 | 18 |
| 110 and under 130 | 19 |
| 130 and under 150 | 9 |

**38.** Closing prices for 42 randomly selected stocks on the New York Stock Exchange are classified in the table. If you wanted to reflect an "up" market showing the highest possible average closing prices, would you use the mean, the median, or the mode?

| **Class (closing prices)** | **Frequency (number of stocks)** |
|---|---|
| $ 0 and under 10 | 2 |
| 10 and under 20 | 0 |
| 20 and under 30 | 4 |
| 30 and under 40 | 7 |
| 40 and under 50 | 9 |
| 50 and under 60 | 7 |
| 60 and under 70 | 6 |
| 70 and under 80 | 4 |
| 80 and under 90 | 2 |
| 90 and under 100 | 1 |

**39.** Using the data from the previous problem, calculate the variance and the standard deviation.

**40.** Given the following nine tests scores for Professor Pundit's economics class, compute the Pearsonian coefficient of skewness. Assume these represent sample data.

80 83 87 85 90 86 84 82 88

**41.** Unionists at a Ford Motor Company plant in Toledo argue that, in violation of the labor agreement, production line workers average a lower hourly wage with greater variability than do office workers. A sample of $n = 10$ is taken from each class of workers, yielding the following values. Do they support the unionists' charge?

| **Workers** | **Production Workers** | **Office Workers** |
|---|---|---|
| 1 | 12.15 | 15.12 |
| 2 | 18.17 | 18.42 |
| 3 | 19.42 | 17.12 |
| 4 | 15.17 | 16.92 |

| | | |
|---|---|---|
| 5 | 18.63 | 18.15 |
| 6 | 16.42 | 15.81 |
| 7 | 15.49 | 19.12 |
| 8 | 18.73 | 19.15 |
| 9 | 19.12 | 18.73 |
| 10 | 18.36 | 19.66 |

**42.** Two competing brands of running shoes were tested for wear. Each reported the following number of hours of use before significant wear was detected.

| **Brand A** | **Brand B** |
|---|---|
| 97 | 78 |
| 83 | 56 |
| 75 | 87 |
| 82 | 54 |
| 98 | 89 |
| 65 | 65 |
| 75 | |

*a.* Which shoe seems to exhibit the longest wear?

*b.* Which shoe seems to have a quality control program which produces the most consistency in their wear?

**43.** Employees at the Kansas City plant average 42.2 units of output per hour with a standard deviation of 10.2 units. The same values for the Denver plant are 41.7 and 12.7.

*a.* Which employees are more consistent in their production pattern?

*b.* Which might have the more productive employees if output is normally distributed?

**44.** In the Fall issue of *Review of Business,* Professors Gregorowicz and Moberly at Auburn University, Montgomery, examined net income per farm in the American agricultural sector. Assume that their data formed the frequency table shown here.

| **Net Income (in $ 1,000's)** | **Frequency (number of farms)** |
|---|---|
| 0 and under 5 | 5 |

| | |
|---|---|
| 5 and under 10 | 12 |
| 10 and under 15 | 10 |
| 15 and under 20 | 15 |
| 20 and under 25 | 17 |
| 25 and under 30 | 12 |
| 30 and under 35 | 8 |
| 35 and under 40 | 10 |

*a.* Estimate the mean, median, and modal incomes of all US farms.

*b.* Estimate the variance and the standard deviation for the population of all US farms.

*c.* If farms with the lowest 10 percent incomes are to receive government subsidies, how low must a firm's income be before it is entitled to a subsidy?

*d.* If farms in the upper 10 percent income range are to pay a surcharge, how high can income be without causing a surcharge?

**45.** An investment broker finds two promising stocks. The first yields an average return of 10 percent with a standard deviation of 1.2 percent. The second yields a 20 percent average rate of return with a standard deviation of 5 percent. Using the coefficient of variation as a measure of risk, she advises her more conservative client to invest in the first. Would you agree? Explain.

**46.** Manly Bankford works as a stockbroker for E. F. Hutton. His records show that the rates of return (in percent) on two securities for 10 selected months were

| | | | | | |
|---|---|---|---|---|---|
| Security 1 | 5.6 | 7.2 | 6.3 | 6.3 | 7.1 |
| | 8.2 | 7.9 | 5.3 | 6.2 | 6.2 |
| Security 2 | 7.5 | 7.3 | 6.2 | 8.3 | 8.2 |
| | 8.0 | 8.1 | 7.3 | 5.9 | 5.3 |

*a.* Which security might be better for those clients interested in a higher return?

*b.* Which security should Manly advise to those clients who prefer less risk?

**47.** The price-earning ratios for 30 different stocks trading on the New York Stock Exchange (NYSE) are shown here.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.8 | 5.2 | 7.6 | 5.7 | 6.2 | 6.6 | 7.5 | 8.0 | 9.0 | 7.7 |
| 3.7 | 7.3 | 6.7 | 7.7 | 8.2 | 9.2 | 8.3 | 7.3 | 8.2 | 6.5 |
| 5.4 | 9.3 | 10.0 | 7.3 | 8.2 | 9.7 | 8.4 | 4.7 | 7.4 | 8.3 |

*a.* Calculate the mean and standard deviation.

*b.* According to Chebyshev's Theorem, at least how many price-earnings ratios lie within two standard deviations of the mean?

*c.* How many actually do lie within two standard deviations of the mean?

**48.** The local mechanic at Vinney's Auto Shop and Charm School tells you that the repairs on your car will cost $ 714.12. Industry data show that the average bill for repairs similar to yours is $ 615, with a standard deviation of **$ 3**1. What might you conclude about Vinney' s rates if you assume that repairs are normally distributed?

**49.** Given here is the frequency table of monthly sales in dollars for skydiving equipment in the southern California market (figures are in hundreds).

| **Class (in $ 100's)** | **Number of Months** |
|---|---|
| 5 and under 10 | 5 |
| 10 and under 15 | 7 |
| 15 and under 20 | 9 |
| 20 and under 25 | 10 |
| 25 and under 30 | 8 |
| 30 and under 35 | 3 |
| 35 and under 40 | 2 |

*a.* You are chief statistician for the Bounce Twice Parachute Company, and your manager requests a breakdown on the frequency of sales. He is interested in that value below which at most 60 percent of the observations fell along with a complete quartile breakdown.

*b.* In addition, you feel that it would be useful to determine the values of the 10th and 90th percentiles.

**50.** Recently, *Fortune* printed an article on baby-boomers who at a relatively young age had become CEOs for major US corporations. This information is summarized here. Compute the mean, median, and modal ages of the corporate moguls.

| CEO | Corporation | Age |
|---|---|---|
| William Beasley III | Lone Star Technologies | 41 |
| Christie Hefner | Playboy Enterprises | 35 |
| Joe "Rod" Canion | Compaq Computer | 43 |
| Jeffrey H. Coors | Adolph Coors Company | 43 |
| Richard Sharp | Circuit City Stores | 41 |
| Jim Manzi | Lotus Development | 36 |

**51.** A supervisor at an assembly plant received the following efficiency ratings over the 12 months

56, 69, 48, 75, 65, 72, 81, 43, 61, 42, 36, 52

*a.* If she wishes to create the most favorable impression, should she report the mean, median, or modal rating in her annual self-evaluation?

*b.* How consistent have her ratings been?

**52.** The accompanying table displays incomes of stockbrokers who work in the Paine- Webber office in San Jose, California. Compute the mean, median, and modal income levels.

| Class (in $ 100's) | Frequency |
|---|---|
| 40 and under 60 | 3 |
| 60 and under 80 | 5 |
| 80 and under 100 | 10 |
| 100 and under 120 | 12 |
| 120 and under 140 | 17 |
| 140 and under 160 | 21 |
| 160 and under 180 | 19 |
| 180 and under 200 | 10 |
| 200 and under 220 | 5 |
| 220 and under 240 | 8 |

**53**. Grade point averages for a sample of 50 students at Ohio State University are shown here. Determine the number of classes using the proper class interval for a frequency table. Construct the table and compute the mean, median, and mode.

| | | | | |
|---|---|---|---|---|
| 3.72 | 1.97 | 2.22 | 0.59 | 4.00 |
| 1.47 | 4.00 | 3.99 | 1.83 | 1.75 |
| 2.77 | 3.01 | 197 | 2.87 | 2 52 |
| 3.01 | 1.29 | 2.57 | 3.31 | 0.97 |
| 0.57 | 2.01 | 3.40 | 0.59 | 3.34 |
| 1.52 | 3.59 | 1.09 | 1.55 | 3 20 |
| 3.44 | 3.88 | 0 59 | 2.57 | 3.32 |
| 3.54 | 3.01 | 2.89 | 2.79 | 2.23 |
| 3.45 | 1.11 | 0.88 | 3.88 | 1.19 |
| 2.09 | 109 | 1 55 | 3.42 | 2.02 |

**54**. *Fortune* reported the earnings of the 500 largest US industrial corporations based on total sales in 1988. The earnings of the 34 largest are shown here. Figures are in billions ol dollars.

| | | |
|---|---|---|
| General Motors 102 | Shell Oil 21 | McDonnell Douglas 13 |
| Exxon 76 | Amoco 20 | Rockwell 12 |
| Ford Motors 72 | United Technologies 17 | Allied Signal 12 |
| IBM 54 | Occidental Oil 17 | PepsiCo 12 |
| Mobil Oil 51 | Procter & Gamble 17 | Lockheed 11 |
| GE 39 | Atlantic Richfield 16 | Kraft 11 |
| Texaco 34 | Nabisco 16 | Phillips 11 |
| AT&T 34 | Boeing 15 | Westinghouse 11 |
| Du Pont 30 | Tenneco 15 | Xerox 10 |
| Chrysler 26 | BP America 15 | Goodyear 10 |
| Chevron 26 | Dow 13 | |
| Philip Morris 22 | Kodak 13 | |

Compute the

*a.* range.

*b.* three quartiles.

*c.* first and last deciles.

**55.** Provide a list of firms in the interquartile range given the data in the previous problem

**56.** From the two previous problems, calculate and interpret the variance and standard] deviation of the top 20 firms in the Fortune 500 list assuming that the list constitutes al

*a.* population.

*b.* sample.

**57.** A data set of size $n$ = 5,000 not known to be normal has a mean of 500 and a standard deviation of 25. At least how many observations are within the range 450 to 550?

**58.** A data set that is known to be normal contains $n$ = 10,000 observations and has a mean of 52.5 and a standard deviation of 5.5.

*a.* How many observations are within the interval 47 and 58?

*b.* How many observations are within the interval of 41.5 and 63.5?

*c.* How many observations are either below 36 or above 69?

**59.** If the distribution of data is normal in the previous problem, what percentage of the observations

*a.* lie between the mean and one standard deviation above the mean? Draw a graph to illustrate.

*b.* lie between the mean and one standard deviation below the mean? Draw a graph to illustrate.

*c.* are either more than one standard deviation above the mean or less than one standard deviation below the mean? Draw a graph to illustrate.

**60.** A sample of $n$ = 1,000 of 36-ounce boxes of Kellogg's Rice Krispies is selected off the production line. The net weights of the contents are normally distributed and average 37 ounces with a standard deviation of 2 ounces. The CEO for Kellogg's insists that at least 75 percent of all boxes coming off the production line are within 0.5 ounce of the net weight specified on the box (36 ounces in this case). Does this sample tend to support the CEO's statement? Draw a graph to illustrate.

### 2.3 Empirical Exercises

**1.** Select 100 stocks at random from the financial pages of the daily newspaper. Construct a frequency table for the closing prices of these stocks, and compute the mean and standard deviation for these prices.

**2.** From Standard & Poor's or Moody's manuals obtain data for the number of shares of common stock outstanding for 30 to 100 firms. Prepare a frequency distribution table for these data. Calculate the mean, median, and modal number of shares outstanding.

**3.** Using the data from Problem 69, calculate and interpret the variance and the standard deviation. How do the three quartiles compare with the mean, median, and mode calculated in Problem 69?

Collect net income figures for the 10 largest and 10 smallest firms on *Fortune's* 500 list. You can find the list in an April issue, probably Number 9 for the current year. Compare the amount of dispersion in net income for the largest firms with that exhibited by the smallest firms. What are your conclusions regarding this comparison, and how might you theorize an explanation for your findings?

## 2.5 Mini-Case Applications

**1.** John White is a stockbroker for Paine-Webber. One of his biggest clients expresses an interest in investing approximately $ 100,000 in the market. Three stocks interest him, but he wishes to invest his money in only one of them. Since he is risk-averse, he wants to invest in the stock that has the lowest risk (lowest variability in returns). As his broker, John is given the following information to help his client make the right decision.

| Year | Returns | | |
|---|---|---|---|
| | **Stock A (%)** | **Stock B (%)** | **Stock C (%)** |
| 1983 | 15 | 17 | 19 |
| 1984 | 28 | 20 | 13 |
| 1985 | 17 | 20 | 16 |
| 1986 | 15 | 19 | 22 |
| 1987 | 25 | 15 | 25 |

In which stock should he invest?

**2.** George Kay lives in Peoria, where recent increases in property tax assessments have many residents alarmed. In an attempt to fight city hall, Mr. Kay and several of his neighbors collect data on tax assessments for several cities throughout the Midwest. They find that the mean assessment is $ 512 with a standard deviation of $ 56. It appears that the data are normally distributed.

Mr. Kay became especially upset to learn that his tax assessment for the coming year is $ 652. The mayor of Peoria promises that he will reduce the average Peoria tax burden. His goal is to reduce taxes in the city so that the average amount paid does not exceed one standard deviation above the mean, based on the Midwest data in Mr. Kay's study.

Mr. Kay retains a professional statistician to seek the answers to some questions. Of importance to Mr. Kay is the amount by which his taxes must come down if they are to meet the pledge offered by the mayor. Furthermore, should he be alarmed by the current tax assessment? What factors other than those presented here must be considered?

**3.** The quantitative analysis group for Paine-Webber conducted a survey of 50 of their offices throughout the world. They wanted to analyze office performance in terms of the number of clients each was able to generate, and the ability of each office to retain client loyalty.

Arble Jackson has just been hired by the group as a new statistical analyst. He is responsible for preparing the data for presentation to upper management. The board of directors for Paine-Webber is interested in prioritizing financial assistance to local offices.

Those offices with an exemplary record of performance in terms of obtaining and retaining clients will be eligible to receive expansion funds from the home office.

Jackson's job, therefore, is to prepare and analyze the following data for the 50 offices, with the intent to establish standards to which each local office's performance can be compared. Again, the board is concerned primarily with client production and retention.

**Number of Clients for 50 Paine-Webber Offices**

| | | | | |
|---|---|---|---|---|
| 117 | 122 | 98 | 97 | 155 |
| 92 | 212 | 115 | 213 | 250 |
| 83 | 293 | 265 | 217 | 220 |
| 150 | 173 | 342 | 178 | 150 |
| 212 | 315 | 162 | 183 | 163 |

| | | | | |
|---|---|---|---|---|
| 342 | 392 | 352 | 211 | 214 |
| 279 | 215 | 291 | 143 | 184 |
| 283 | 143 | 173 | 194 | 243 |
| 307 | 357 | 233 | 165 | 219 |
| 99 | 293 | 315 | 183 | 210 |

Teltronics manufactures components for communications systems used by NASA in its space flights. Some concern has arisen regarding productivity. Weekly production levels in 1,000s of units for its two major plants in Chicago and Memphis are shown below.

You have just been hired as a production analyst and, using these data, you must make certain decisions regarding the output patterns of these two plants. Your immediate superior is primarily interested in the average level of output and whether that level is maintained at a relatively consistent rate. How would you respond to your superior's request for the information he desires?

| **Chicago** | **Memphis** | **Chicago** | **Memphis** | **Chicago** | **Memphis** | **Chicago** | **Memphis** |
|---|---|---|---|---|---|---|---|
| 34 | 39 | 65 | 45 | 65 | 74 | 64 | 35 |
| 37 | 42 | 64 | 37 | 46 | 27 | | |
| 54 | 43 | 64 | 72 | 38 | 53 | | |
| 59 | 52 | 82 | 45 | 65 | 54 | | |

## 3 PROBABILITY AND PROBABILITY DISTRIBUTIONS

### 3.1 Solved problems

**Example 1.** John Dough, Megatycoon, is considering a purchase of 500 shares of ABC, Inc. on the New York Stock Exchange. He wishes to determine the probability that (1) its price will decline in today's trading, (2) its price will rise in today's trading, and (3) its price will remain unchanged. As his highly paid statistician, you must determine these probabilities.

**Solution:** You quickly recognize this as a standard case of relative frequency probability. By inspecting the records from NYSE trading over the past 100 days, you learn that the price declined 20 of those days, rose 50 of those days, and remained unchanged for the rest.

The probability of the event the price will decrease, $\mathbf{P(D),}$ in today's trading is

$$P(D) = \frac{\text{Number of times event has occurred}}{\text{Total number of observations}} = \frac{\text{Number of times the price has declined}}{\text{Total number of observations}} =$$

$$= \frac{20}{100} = 0.2 \ \ or \ \ 20 \ \%.$$

The probability of that the price will rise, $\mathbf{P(R),}$ is

$$P(R) = \frac{50}{100} = 0.5 \ \ or \ 50 \ \%.$$

The probability of that the price will remain unchanged, $\mathbf{P(U),}$ is

$$P(U) = \frac{30}{100} = 0.3 \ \ or \ 30 \ \%.$$

**Interpretation:** The probability that the price will fall is 20%, and the probability that id will rise is 50%. There is a 1 out of 5 chance that it will fall and a 1 out of 2 chance] that it will rise. Notice that, in accordance with the second property of probability, the] probabilities of all three events sum to one since one of the three events must occur.

**Example 2.** Craps is a game of chance played with two dice. The rules of at least one version of the game (there are several variations) state that you immediately win if you roll craps, a 7 or 11, on the first roll. If you roll anything other than craps, you must roll that number again (which is called your *mark* or *point*) before rolling either a 7 or 11. If you roll a 7 or 11 before rolling your mark, you lose.

*a.* What is the probability of winning a game of craps on the first roll?

*b.* If your mark is a 6, are you more likely to win or lose the game?

**Solution:** The sample space of all possible outcomes is the sum of the two dice. These outcomes and the ways in which they can occur are shown here.

**Outcome of First Die**

| **Outcome of Second Die** | | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|---|
| | **1** | 2 | 3 | 4 | 5 | 6 | 7 |
| | **2** | 3 | 4 | 5 | 6 | 7 | 8 |
| | **3** | 4 | 5 | 6 | 7 | 8 | 9 |
| | **4** | 5 | 6 | 7 | 8 | 9 | 10 |
| | **5** | 6 | 7 | 8 | 9 | 10 | 11 |
| | **6** | 7 | 8 | 9 | 10 | 11 | 12 |

*a.* there are 36 possible outcomes. Only eight of these produce a 7 or 11, which result in a win. Thus,

$$\mathrm{P(win\ on\ first\ roll)} = \frac{8}{36};$$

*b.* If you roll a 6, the probability of repeating that roll is

$$\mathrm{P}(6) = \frac{5}{36}.$$

**Interpretation:** Since 8/36 is more than 5/36, you are more likely to lose by rolling craps than you are to win by rolling your mark of 6.

**Example 3.** Consider the following case. Of 200 people in a survey, 120 are men and 80 are women. Of the 120 men, 40 own stock in a US corporation, while of the 80 women, 20 are stockholders. Thus, we have

120 men, 40 of whom own stock;

80 women, 20 of whom own stock.

Determine the probability of selecting one person at random who is a man and owns stock. This task will require the use of conditional probability.

$$P(M\ and\ S) = P(M) * P(S \mid M) = \frac{120}{200} * \frac{40}{120} = \frac{40}{200}.$$

The $P(M)$ is 120/200 since 120 of the 200 people are men. The $P(S \mid M)$ is 40/120 because it asks for the probability of stock ownership given that we have only the 120 men.

The statement $P(S \mid M)$ must be used because the probability of owning stock depends on gender: a different (higher) proportion of men own stock than do women. If, by pure coincidence, 60 of the 120 men and 40 of the 80 women owned stock, then, regardless of gender, $P(S \mid M) – P(S \mid W) – P(S) = 0.5$. Gender and ownership of stock would be independent events, and $P(S)$ could have been used in place of $P(S \mid M)$.

In a more general sense, if two events, $A$ and $B$. are independent, then

$$P(A) = P(A|B).$$

**Example 4.** Assume that we wish to find the probability of drawing either an ace or a heart in a single draw from a deck of cards. That is, we seek $P(A\ or\ H)$. (There are 13 hearts in a deck of 52 cards.) Note that "Ace" and "Heart" are not mutually exclusive. They both occur if the ace of hearts is drawn. Then, using Formula

$$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B),$$

As the result we have

$$P(A\ or\ H) = P(A) + P(H) - P(A\ and\ H) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}.$$

The joint event A and H occurs only if the ace of hearts is drawn, the probability of which is 1 /52. It is necessary to subtract $P(A$ and $H)$ because the probability of drawing the ace of hearts was included twice – once when we computed the probability of all 4 aces, and once when we computed the probability of all 13 hearts. Since there is only one ace of hearts, we need to deduct the probability of the joint event in order to avoid **double counting** – counting the ace of hearts twice.

**Example 5.** Consider a case in which the two events are mutually exclusive. Assume we want to find the probability of drawing either a heart or a diamond. These events are mutually exclusive (you cannot draw both a heart and a diamond on the same draw) and the joint probability of H and D is zero. Thus, Formula

$$P(A\ or\ B) = P(A) + P(B).$$

yields

$$P(H\ or\ D) = P(H) + P(D) = \frac{13}{52} + \frac{13}{52} = \frac{26}{52} = \frac{1}{2}.$$

This should come as no surprise since one-half of the cards are red.

**Example 6.** What is the probability of picking a card and getting a face card or a 10?

$$P(A \text{ or } B) = P(A) + P(B) = 12/52 + 4/52 = 16/52.$$

**Example 7.** What is the probability of rolling a die and getting a 3 or an odd number?

$$P(A \text{ or } B) = P(A) + P(B) – P(A \text{ and } B) = 1/6 + 3/6 – 1/6 = 3/6.$$

**Example 8.** *Business Week* described Procter & Gamble's efforts to compare Tide, their best selling laundry detergent, to Surf, produced by Lever Brothers. P&G surveyed shoppers in a large mall near Cincinnati, Ohio, to determine their preferences. For notational purposes, assume that $T_1$ means that the first shopper preferred Tide and $S_2$ means that the second shopper preferred Surf. Industry reports reveal that Tide has 20 percent of the market and Surf enjoys 5 percent. Assuming these market shares translate into probabilities, P&G executives want to know, if two shoppers are selected at random, what is the probability that

*a.* Both prefer Tide: P($T_1$, and $T_2$).

*b.* Both prefer Surf: P($S_1$, and $S_2$).

*c.* The first prefers Tide and the second prefers Surf: P($T_1$, and $S_2$).

*d.* The first prefers Surf and the second prefers Tide: P($S_1$, and $T_2$).

*e.* One shopper prefers one detergent and the second prefers the other: P($T_1$, and $S_2$,) or P($S_1$, and $T_2$).

**Solution:** Since we may assume that the preference of one shopper has no effect or influence on the preference of the other, the two events are independent. Using Formula

$$P(A \text{ } and \text{ } B) = P(A) \times P(B),$$

we have

*a.* $P(T_1 \text{ } and \text{ } T_2) = P(T_1) \times P(T_2) = 0.20 \times 0.20 = 0.04;$

*b.* $P(S_1 \text{ } and \text{ } S_2) = P(S_1) \times P(S_2) = 0.05 \times 0.05 = 0.0025;$

*c.* $P(T_1 \text{ } and \text{ } S_2) = P(T_1) \times P(S_2) = 0.20 \times 0.05 = 0.01;$

*d.* $P(S_1 \text{ } and \text{ } T_2) = P(S_1) \times P(T_2) = 0.05 \times 0.20 = 0.01.$

*e.* Each detergent could be preferred by one shopper if ($S_1$, and $T_2$) or ($T_1$, and $S_2$). Since the event can happen the first way *or* the second way, we must add the two probabilities.

$$\text{P(each is preferred)} = (S_1, \text{ and } T_2) + (T_1, \text{ and } S_2) =$$
$$= 0.05 \text{ x } 0.20 + 0.20 \text{ x } 0.05 = 0.01 + 0.01 = 0.02.$$

**Interpretation:** Since the joint events are independent, the probability of the second event does not require consideration of the first event, and the element of conditional probability is not needed.

**Example 9. A Probability Distribution.** The investment firm of Loosit and Lye employs 20 investment analysts. Every morning each analyst is assigned up to five stocks to evaluate. The assignments made this morning were

| **Outcome ($x_i$) (number of stocks)** | **Frequency of $x_i$ (number of analysts)** |
|---|---|
| 1 | 4 |
| 2 | 2 |
| 3 | 3 |
| 4 | 5 |
| 5 | 6 |
| **Total** | **20** |

*a.* Mr. Loosit wishes to develop a probability distribution for the random variable of the number of stocks assigned to the analysts this morning.

**Solution:** The probability distribution must show each possible outcome, the number of stocks assigned to each broker, $x_h$ and the probability associated with each outcome. The value the random variable can take is some number 1 through 5. Thus, the probability distribution appears as

| **Number of Stocks ($x_i$)** | **Frequency** | **$P(X = x_i)$** |
|---|---|---|
| 1 | 4 | 4 / 20 = 0.20 |
| 2 | 2 | 2 / 20 = 0.10 |
| 3 | 3 | 3 / 20 = 0.15 |
| 4 | 5 | 5 / 20 = 0.25 |
| 5 | 6 | 6 / 20 = 0.30 |
| **Total** | – | **1.00** |

Or, the probability distribution might be expressed as

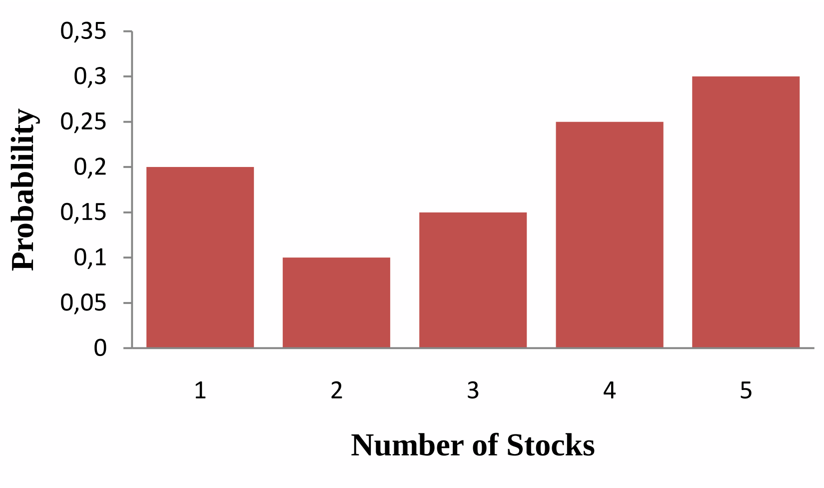



Figure 3.1 – Probability Distribution for Example 9, Chapter 3

**Interpretation:** If any analyst is picked at random, the probability that he or she will have five stocks to evaluate today is greater than any other single number. Only 10 percent of the analysts have two stocks to analyze before the day is over.

*b*. Mr. Lye determines the mean and variance for the probability distribution of the number of stocks assigned each analyst in the firm.

**Solution:** The mean becomes

$$\mu = \sum [(x_i)\, P(x_i)] = (1 \times 4/20) + (2 \times 2/20) + (3 \times 3/20) + (4 \times 5/20) +$$
$$+(5 \times 6/20) = 3.35\; stocks.$$

Using Formula

$$\sigma^2 = \sum [(x_i - \mu)^2\, P(x_i)],$$

or

$$\sigma^2 = \sum [(x_i)^2\, P(x_i)] - (\mu)^2.$$

to calculate the variance yields

$$\sigma^2 = (1 - 3.35)^2 \left(\frac{4}{20}\right) + (2 - 3.35)^2 \left(\frac{2}{20}\right) + (3 - 3.35)^2 \left(\frac{3}{20}\right) +$$
$$+(4 - 3.35)^2 \left(\frac{5}{20}\right) + (5 - 3.35)^2 \left(\frac{6}{20}\right) = 2.23.$$

**Interpretation:** The brokers are assigned an average of 3.35 stocks to evaluate and analyze. The variance of 2.23 carries the usual interpretation. It is a measure of the dispersion around the mean of 3.35.

**Example 10. Mathematical Expectation**

Suppose someone offers you the bet of flipping a coin three times. If you get all heads, you win \$ 20, while two heads (and one tail) wins you \$ 10. If you get one head (and two tails), you lose \$ 12. and all tails causes you to lose \$ 20. Should you take this bet?

The probability of **three heads** is 1/8.

$$H_1 \; and \; H_2 \; and \; H_3 = 1/2 \times 1/2 \times 1/2 = 1/8.$$

The probability of **two heads** is 3/8.

$$\begin{aligned} &H_1 \; and \; H_2 \; and \; T_3 && = 1/2 \times 1/2 \times 1/2 && = 1/8; \\ &H_1 \; and \; T_2 \; and \; H_3 && = 1/2 \times 1/2 \times 1/2 && = 1/8; \\ &T_1 \; and \; H_2 \; and \; H_3 && = 1/2 \times 1/2 \times 1/2 && = 1/8; \\ & && && = 3/8. \end{aligned}$$

The probability of **one head** is 3/8.

$$\begin{aligned} &H_1 \; and \; T_2 \; and \; T_3 && = 1/2 \times 1/2 \times 1/2 && = 1/8; \\ &T_1 \; and \; H_2 \; and \; T_3 && = 1/2 \times 1/2 \times 1/2 && = 1/8; \\ &T_1 \; and \; T_2 \; and \; H_3 && = 1/2 \times 1/2 \times 1/2 && = 1/8; \\ & && && = 3/8. \end{aligned}$$

The probability of **no heads** is 1/8.

$$T_1 \; and \; T_2 \; and \; T_3 = 1/2 \times 1/2 \times 1/2 = 1/8.$$

| **Event** | **Probability [P($x_i$)]** | **Payoff ($x_i$)** |
|---|---|---|
| 3 heads | 1/8 | \$ 20 |
| 2 heads | 3/8 | 10 |
| 1 head | 3/8 | –12 |
| 0 heads | 1/8 | –20 |

The expected value of the payoff, that is, the amount you might expect to win, is

$$E(X) = (20) * (1/8) + (10) * (3/8) - (12) * (3/8) - (20) * (1/8) =$$
$$= 2.5 + 3.75 - 4.5 - 2.5 = -0.75.$$

This game has a negative expected value. Does this mean if you play the game, you must lose? No. You can win as much as $ 20. It means that if you play the game many times, you will, on the average, lose 75 cents for each time you have played. The expected value is the average amount you will win (or lose in this case) if you play the game often enough. Given the rules of this game, if you play a large number of times (theoretically, an infinite number of times) you will suffer a net loss.

**Example 11. A Binomial Distribution.** Consider the following situation. A credit manager for American Express has found that $\pi = 10$ percent of their card users do not pay the full amount of indebtedness during any given month. She wants to determine the probability that of $n = 20$ accounts randomly selected, $x = 5$ of them are not paid. This can be written as $P(X = 5| n = 20, \pi = 0.10)$, which is read as "the probability of five successes given that there are 20 trials and the probability of a success on any one trial is 10 percent."

Does this set of circumstances fit the conditions for a binomial distribution?

There are only two possible outcomes. Either the account (1) is, or (2) is not paid in full. We can assign the identity of a success to an account that is not paid, and a failure to one that is.

The probability of a success remains constant at $\pi = 0.10$ from one trial (account) to the next.

The probability one person pays off his or her account is independent of whether someone else pays.

Last, there are many different accounts available for examination.

Clearly, the binomial distribution applies, and we can use it to solve the credit manager's problem.

The probability that 5 accounts out of the 20 sampled remain unpaid can be calculated by using Formula

$$P(x) = \frac{n!}{x!\,(n-x)!}\pi^x(1-\pi)^{n-x} = \ {}_nC_x(\pi)^x(1-\pi)^{n-x}.$$

Where $n = 20$, $X = 5$, and $\pi = 0.10$, we have

$${}_{20}C_5(0.10)^5(0.90)^{20-5} = (15504) * (0.00001) * (0.2058911) = 0.0319.$$

If the probability that any one account is not paid in full is $\pi = 0.10$, then there is a 3.19 percent chance that exactly 5 of 20 accounts selected at random will retain a positive balance.

**Example 12. Jacques the Jock.** Records show that the firm's computer is up and running 80 percent of the time. Jacques, the firm's resident computer jock, examines the computer nine times. What is the probability that it is functional three times?

**Solution:** This problem calls for $n = 9$, $X = 3$, and $\pi = 0.80$. That is, $P(X = 3 \mid n = 9, \pi = 0.80)$. However, since Binomial Distrbution table has values for $\pi$ only up to 0.50, the probability of $x = 6$ successes (9 – 3) if $\pi = 0.20$ (1.00 – 0.80). Thus, $P(X = 6 \mid n = 9, \pi = 0.20)$ is 0.0028. There is less than a 1 % chance that the computer will be running exactly three of the nine times Jacques inspects it.

A brief sketch such as that shown here can prove useful. Since we seek $P(X = 3 \mid n = 9, \pi = 0.80)$ we can construct two ordered arrays, one from 0 to 9 for $\pi = 0.8$ and one from 9 to 0 for $\pi = 0.2$.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | **3** | 4 | 5 | 6 | 7 | 8 | 9 | ($\pi = 0.8$) |
| 9 | 8 | 7 | **6** | 5 | 4 | 3 | 2 | 1 | 0 | ($\pi = 0.2$) |

The probability of three successes with $\pi = 0.8$ as shown in the upper array cannot be found in the table. However, the probability of three out of 9 successes with $\pi = 0.8$ is, as seen in the lower array, the same as the probability of six out of nine failures each with a probability of 1 – 0.8 = 0.2. That is, if you got three successes (at $\pi$ = 0.8), you must have gotten six failures (at $\pi$ = 0.2). We can then merely find $P(X = 6 \mid n = 9, \pi = 0.20) = 0.0028$.

**Interpretation:** When the value for $\pi$ exceeds 0.50, it is necessary to adjust the values for $X$ and $\pi$.

**Example 13. Using Poisson probability function.** Suppose we are interested in the probability that exactly five customers will arrive during the next hour (or any given hour) of business. Simple observation over the past 80 hours has shown that 800 customers have entered our business. Thus, $\mu$ = 10 per hour. Using Formula

$$P(x) = \frac{\mu^x \times e^{-\mu}}{x!}.$$

We get

$$P(5) = \frac{10^5 \times 2.71828^{-10}}{5!} = 0.0378.$$

**Interpretation:** There is a 3.78 % chance that exactly five customers will enter the store during the next hour.

**3.2 Conceptual Questions**

1. Clearly distinguish in your own words among the features of the three approaches to probability. Under what circumstances would each type be more appropriate than the others?

2. Which approach to probability would be most appropriate under each of the following conditions?

*a.* Rolling a die and noting the outcome.

*b.* Trying to determine if the price of gold will rise or fall.

*c.* Determining the likelihood that you will pass this course.

*d.* Estimating the likelihood that your company's sales will increase next year given the government's economic forecast.

*e.* *A* professional gambler or odds-maker establishing the odds on the winner of a prizefight.

3. What is the sample space assigned to the experiment of drawing a card from a deck and noting the

*a.* suit?

*b.* color?

4. What is the sample space in the experiment of observing the behavior of the price of a certain stock on the NYSE?

5. What is the relationship between probability and odds? How do professional gamblers ensure that they will win?

6. Define each of the following and provide business-related examples:

*a.* Complementary events.

*b.* Collectively exhaustive events.

*c.* A union of events.

*d.* An intersection of events.

**3.3 Problems**

**1**. What is the probability that

*a.* you will be elected president of the United States in the next election?

*b.* somebody will be elected president in the next election?

**2.** Odds that the chairman of a major firm will be replaced are set at 3:1. What is the probability that he will be replaced?

**3.** If the probability a certain boxer will win a prizefight is 80 percent,

*a.* what betting odds will be set by the odds-makers?

*b.* if you bet \$ 12 on the boxer and he wins, how much do you get?

*c.* if you bet \$ 1 he will lose, how much will you win if he does lose?

**4.** What is the probability of

*a.* rolling a die and getting a 3?

*b.* rolling a die and getting an odd number?

*c.* drawing a heart from a deck?

*d.* drawing a jack from a deck?

5. Quick-Think, Inc., offers two types of computers: (1) the standard model, and (2) the deluxe model. Each can be purchased with an optional hard drive. Construct a Venn diagram to show

*a.* the standard models and the deluxe models.

*b.* the standard models and those models with a hard drive.

*c.* the standard models, the deluxe models, and those models without a hard drive.

6. What is the probability of

*a.* rolling a die and getting either a 6 or an odd number and drawing a heart from a deck of cards?

*b.* rolling a die and getting either a 6 or an even number and drawing a heart from a deck?

*c.* flipping a coin twice and getting two heads and drawing a king from a deck?

**7.** Why is the probability of rolling a die and getting a 6 or an even number equal to the probability of getting an even number, while the probability of drawing a king or a spade is not equal to the probability of drawing a spade? That is, why is $P$(6 or even) = $P$(even) while $P$(king or spade) ≠ P(spade)?

**8.** Prove $P$(6 or even) = $P$(even).

**9.** As reported in the June 1993 issue of *Datamation*, the sales division of Apple Computers examined sales results of its sales staff by sampling the revenues generated by staff personnel who had received three different types of training. Assume that these sample data were as shown here, and recorded the number of people in each type of training and their relative sales levels. For example, five people received training level 1 and had above-average sales.

| **Revenue** | **Training level** | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| Above Average | 5 | 10 | 8 |
| Average | 12 | 5 | 7 |
| Below Average | 10 | 3 | 5 |

What is the probability that an individual selected at random

*a.* had above average sales?

*b.* received training level 3?

*c.* had above average sales and received training level 3?

*d.* had below average sales and training level 2?

*e.* had average sales or level 3?

*f* had above average sales or level 2?

*g.* did not have average sales?

*h.* did not have average sales or level 1?

**10.** Construct a probability table from the data in the previous problem.

**11.** Using the table from the previous problem, show calculations of parts $(a)$, $(e)$, and $(h)$ of Problem 15.

**12.** Output levels for Bell Electronics of three different relay switches at plants in three different cities are shown below along with the defect rates for each switch type.

| City | Output | Defects |
|---|---|---|
| **Seattle** | | |
| Type A | 40 | 10 |
| Type B | 20 | 5 |
| Type C | 7 | 6 |
| **Topeka** | | |
| Type A | 12 | 5 |
| Type B | 15 | 10 |
| Type C | 10 | 2 |
| **Omaha** | | |
| Type A | 10 | 5 |
| Type B | 10 | 2 |
| Type C | 10 | 4 |

If a switch is chosen at random for inspection by Bell's quality control circle, what is the probability that it is

*a.* from Seattle and defective?

*b.* from Topeka and not defective?

*c.* type A and defective?

*d.* type B and not defective?

*e.* from Omaha or defective?

*f.* from Topeka or not defective?

**13.** Dell Publishing has 75 different book titles classified by type and cost as follows:

| **Type** | **Cost** | | |
|---|---|---|---|
| | **$ 10** | **$ 15** | **$ 20** |
| Fiction | 10 | 8 | 3 |
| Biography | 12 | 10 | 9 |
| Historical | 4 | 17 | 2 |

Find the probability that a book selected at random is

*a.* either fiction or costs \$ 10.

*b.* both historical and costs \$ 20.

*c.* historical and costs either \$ 10 or \$ 15.

*d.* is fiction and costs less than \$ 20.

*e.* is biographical or costs \$ 15.

*f* is biographical or more than \$ 10.

**14.** In the summer of 1993, the federal government's General Accounting Office examined the tendency for Congressional support of various government agencies to affect the ability of those agencies to meet budget constraints. Of 63 agencies studied, 42 were supported in principle by Congress. Of these 42, 5 were under budget and the rest were over budget. Ten were not supported by Congress. Only one of these was under budget and the rest exceeded budget guidelines. For the remaining 11, it was not clear from Congressional debate if Congress supported them, and all of them were over budget. Find the probability that for an agency selected at random,

*a.* it was supported and under budget.

*b.* it was not supported and under budget.

*c.* support was unknown and it was not under budget.

*d.* support was unknown and it was under budget.

*e.* it was not supported or over budget.

*f.* it was not supported and was over budget.

*g.* it was supported and under budget.

*h.* it was supported or under budget.

**15.** Construct the probability table from the data in the previous problem.

**16.** Construct the probability tree from the data in Problem 20.

**17.** The management department at State University has access to three fax machines. The probability each is out of service is 20/100, 25/100, and 30/100, respectively. Assuming independence, find the probability that

*a.* the first and the second are out of service.

*b.* the first and the third are out of service.

c. all are out of service.

*d.* none are out of service.

*e.* one is out of service.

*f.* two are out of service.

*g.* two or more are out of service.

**18.** Given the data in the previous problem,

*a.* construct a probability tree.

*b.* prove the events are independent while in Problem 20 the events are dependent.

**19.** The Federal Savings and Loan Insurance Corporation (FSLIC) has found that 7 percent of the heads of households who have a home mortgage with a savings and loan association are currently unemployed. It has also found that 12 percent of all borrowers default on their loans and that 60 percent of all unemployed persons default. What is the probability that a borrower

*a.* is unemployed and will default?

*b.* is unemployed or will default?

**20.** In 1993, the National Health Institute surveyed over 200 young executives regarding their exercise habits. The NHI found that 60 percent jogged, 25 percent swam, and 12 percent did both on a regular basis. Find the percentage of executives

*a.* who jog or swim.

*b.* who, given they swim, also jog.

*c.* who swim, given they jog.

**21.** Mark buys three different stocks. The probability the first will rise in value is 1/3, the probability the second will rise is 3/4, and the probability the third will rise is 1/10. Determine the probability that

*a.* all will rise in value.

*b.* none will rise.

*c.* one will rise.

*d.* two will rise.

*e.* at least two will rise.

*f.* at least one will rise.

**22.** From a list of 12 growth stocks and 17 income stocks, an analyst must pick 2 for his client. What is the probability that *a.* both are growth stocks[9]

*b.* the first is a growth stock and the second is an income stock?

*c.* one is a growth stock and the other is an income stock?

**23.** A local construction company found that only 20 percent of all jobs were completed on time, while 30 percent of all jobs suffered cost over-runs. In addition, cost over-runs occurred 75 percent of the time that a job was completed on time. The owner of the firm wants to find the probability that a job suffers a cost over-run

*a.* and is done on time.

*b.* or is done on time.

**24.** From the previous problem, how can you prove that cost over-runs and the probability that a job is done on time are not mutually exclusive events?

**25.** Given the information from Problem 29. What is the probability that if there is a cost overrun, the job will be done on time?

**26.** *Fortune* magazine found that 10 percent of those workers in upper-level corporate executive positions were women, and that 3 percent of those in the upper level were women with MBA degrees. The board of directors of a large corporation, whose executive profile fits this description, wishes to select one of their executive women at random. How likely is it that they will select an MBA?

**27.** Refer to Problem 32. It is also known that 25 percent of corporate executives hold MB As. What is the probability of selecting a woman from among their MBAs?

**28.** Larry Lactose and his three children, Luci, Loni, and Lenny, walk by a gumball machine containing gum of three different colors. Five of the pieces are blue, four are green, and three are red. Each child wants a piece of gum, but Larry knows an argument will pursue unless each child gets a different color. If the gum costs 5 cents each, what is the minimum Larry can spend and get three pieces of different colors, and what is the probability that he will do so?

**29.** Ten units of output are selected from the production line. Three of these 10 are defective. If 5 are to be drawn from the 10, what is the probability that 2 are defective?

**30.** Thirty percent of the units produced by a firm are defective.

*a.* If 5 are drawn at random, what is the probability that 2 are defective?

*b.* Why does this answer differ slightly from that obtained in Problem 35? The numbers are essentially the same.

### 3.4 Empirical Exercises

**1.** Flip a coin 100 times (or 5 coins 20 times to reduce the amount of time involved). According to the classical approach to probability, how many heads should you get? How many did you get? Explain any difference that might result.

**2.** What would happen to the proportion of heads if you flipped the coin 200 times? Explain.

### 3.5 Mini-Case Applications

1. Rita Davis is manager of a plant in Toledo that manufactures glass products. She has recently been concerned about sagging productivity levels. The home office in Corning, Pennsylvania, has instructed her to take actions to reverse the slide in worker output. In this effort Ms. Davis has contacted Emerson Company, an educational reinforcement firm which specializes in the training and retraining of production line workers.

Emerson has outlined two training programs for Davis's workers. However, Davis is unable to determine which plan would be more beneficial to her firm. Their first plan would cost Davis $ 5,000 to administer, while the second would require an outlay of $ 7,500. In her effort to select the more useful training program, Davis has learned that of all workers trained by Emerson, 80 percent showed improved productivity. This still does not answer the question as to which plan she should choose. She also is able to determine that, given a worker showed improvement, there was a 60 percent chance he took the first training program and an 80 percent chance he took the second. Also, if no improvement was shown, there was a 20 percent chance he took the first training program and a 50 percent chance he took the second. Davis also figures that each 10 percent improvement in productivity will increase plant revenues by $ 1,000. On this basis, Davis hires you as a statistical consultant to assist them in determining which training plan is preferable. How do you respond?

2. A recent issue of *Fortune* magazine addressed the attempts by Dow Chemical to identify those employees with superior managerial potential. It seems

that Dow administers a test to those employees who might have executive capabilities. Past evidence shows that 70 percent of the people are of managerial quality. Kyle Rogers, who is the director of personnel relations for the southern branch of Dow, has determined that the test correctly identifies a good manager only 80 percent of the time, and 10 percent of the time it incorrectly identifies someone as worthy of consideration for a management position who later proves to lack any managerial talent. Rogers recently administered the test to two people. One passed the test, and one did not. Rogers wonders how to interpret these results given that the test is not perfect. He had always felt that the person who failed the test had exhibited the qualities Dow seeks in its managers, and he therefore feels that, despite the test results, this person should be promoted and given a broader range of responsibilities. Rogers wishes to determine the probability that each of the people tested does possess management skills.

# 4 SAMPLING DISTRIBUTIONS: AN INTRODUCTION TO INFERENTIAL STATISTICS

## 4.1. Solved Problems

**Example 1. There's No Place Like Nome.** Although Nome, Alaska is located on the Seaward peninsula very close to Siberia, the local chamber of commerce proudly displays its summertime temperatures. Alaska Today, a travel magazine devoted to touring throughout our northernmost state, recently published the high temperatures on June 1 for the past five years. They were 68, 73, 65, 80, and 72, for a mean of $\mu$ = 71.6. Assume you wish to estimate this mean by taking a sample of size $n$ = 3. Construct the entire sampling distribution for this process. What observations might you make?

**Solution.** There are ${}_5C_3$ = 10 possible samples you could take.

| **Sample number** | **Sample elements $X_i$** | **Sample mean $\bar{X}$** | **Sample number** | **Sample elements $X_i$** | **Sample mean $\bar{X}$** |
|---|---|---|---|---|---|
| 1 | 68, 73, 65 | 68.67 | 6 | 68, 80, 72 | 73.33 |
| 2 | 68, 73, 80 | 73.67 | 7 | 73, 65, 80 | 72.67 |
| 3 | 68, 73, 72 | 71.00 | 8 | 73, 65, 72 | 70.00 |
| 4 | 68, 65, 80 | 71.00 | 9 | 73, 80, 72 | 75.00 |
| 5 | 68, 65, 72 | 68.33 | 10 | 65, 80, 72 | 72.33 |

The sampling distribution is

| $\bar{X}$ | $P(\bar{X})$ |
|---|---|
| 68.67 | 1/10 |
| 73.67 | 1/10 |
| 71.00 | 2/10 |
| 68.33 | 1/10 |
| 73.33 | 1/10 |
| 72.67 | 1/10 |
| 70.00 | 1/10 |
| 75.00 | 1/10 |
| 72.33 | 1/10 |
| Total | 1.00 |

**Interpretation**. Since your mean was $\mu$ = 71.6, five of the 10 possible samples will produce a mean, $\bar{X}$, in excess of 71.6, while the remaining five samples underestimate $\mu$.

**Example 2. The Standard Error.** Your professor promises to give you an A if you can compute the standard error of the sampling distribution from Example 1.

$$\mu = \bar{\bar{X}} = \frac{\sum \bar{X}}{K} = 71.6;$$

$$\sigma_{\bar{X}}^2 == \frac{\sum(\bar{X} - \bar{\bar{X}})^2}{K} = 4.31 \text{ degrees squared;}$$

$$\sigma_{\bar{X}} == \sqrt{4.31} = 2.08 \text{ degrees.}$$

**Interpretation.** All 10 possible sample means averaged 71.6. That is, $\bar{\bar{X}} = \mu = 71.6$. There was a tendency for the sample means to vary from that 71.6 by 2.08 degrees.

**Example 3**. Consider the years of work experience for four employees of 10, 7, 12, and 8. Given a sample size of $n$ = 2, calculate ($a$) the sampling distribution, and ($b$) the grand mean.

**Answer:**

($a$)

| X | $\bar{X}$ |
|---|---|
| 10, 7 | 8.5 |
| 10, 12 | 11.0 |
| 10, 8 | 9.0 |
| 7, 12 | 9.5 |
| 7, 8 | 7.5 |
| 12, 8 | 10.0 |

($b$) $\bar{\bar{X}} = \mu = 55.5/6 = 9.25.$

**Example 4. Using the FPC.** The US Bureau of Census wishes to estimate the birthrates per 100,000 people in the nation's 100 largest cities. It is known that the standard deviation in the birthrates for these 100 urban centers is 12 births per 100,000 people. (*a*) Calculate the variance and standard error of the sampling distribution of $n$ = 8 cities. (*b*) Calculate the variance and standard error of the sampling distribution of $n$ = 15 cities.

**Solution.**

*(a)* Since $n$ is less than 10 percent of $N$, the FPC is not required. The variance and standard error are

$$\sigma_{\overline{X}}^2 == \frac{\sigma^2}{n} = \frac{12^2}{8} = 18, \ \sigma_{\overline{X}} == \sqrt{18} = 4.24.$$

*(b)* Since $n$ is more than 10 percent of $N$, the FPC is required. The variance and standard error are

$$\sigma_{\overline{X}}^2 == \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right] = \frac{12^2}{8}\left[\frac{100-15}{100-1}\right] = 8.24, \ \sigma_{\overline{X}} == \sqrt{8.24} = 2.87.$$

**Interpretation**. The larger sample has a smaller standard error and will tend to result in less sampling error in estimating the birthrates in the 100 cities.

**Example 5.** From a population of 200 observations, a sample of $n$ = 50 selected. Calculate the standard error if the population standard deviation equals 22.

**Answer.** 2.7

**Example 6.** The Telecom recorded telephone messages for its customers. These messages averaged 150 seconds with a standard deviation of 15 seconds. Telecom wished to determine the probability that the mean duration of a sample of $n$ = 35 phone calls is between 150 and 155 seconds, that is $P(150 < \overline{X} < 155)$.

**Solution**. Since $n > 1$ and a sample was taken, Formula (7.8) must be used. Then

$$Z = \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{155-150}{15/\sqrt{35}} = 1.97 \text{ or an area of 0.4756.}$$

Thus,

$$P(150 < \overline{X} < 155) = P(0 < Z < 1.97) = 0.4756.$$

See it in Figure 7.3. So the probability that a sample of 35 calls will have duration within the range of 150 and 155 seconds is 47.56 %. Such a quite big percentage is because the sampling distribution is less dispersed than the original population, i.e. the dispersion of original population is bigger than dispersion of the sampling distribution ($\sigma > \sigma/\sqrt{n}$).

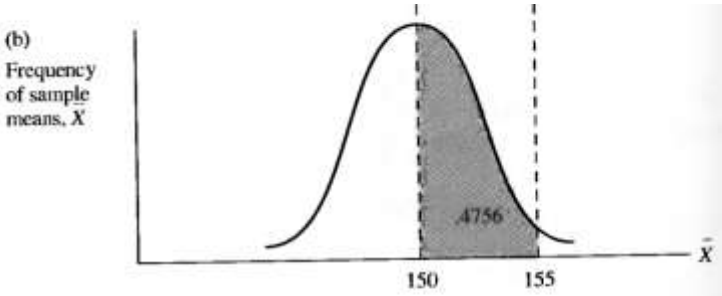



Figure 4.1 – The Mean of a Sample Observations

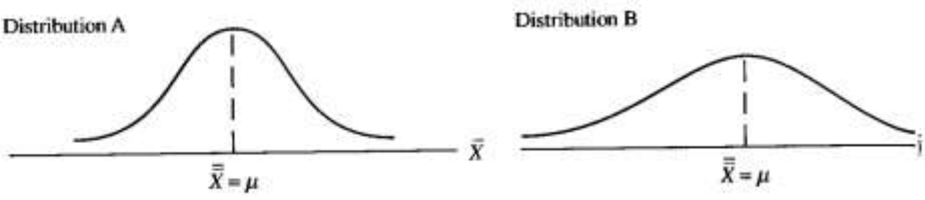
**Interpretation.** On the basis of this information, Telecom can make more intelligent decisions regarding the need for new equipment.

### 4.2 Conceptual Questions

1. If a sample is taken in which $n$ < 30, what problems might we have in working with it?

2. If several samples of a given size are taken from a population, what would influence the variability of those sample means? What happens to that variability as $n$ increases?

3. From a single population, two sampling distributions are formed by taking all possible samples of a given size to get sampling distribution A, and all possible samples of a different size to get sampling distribution B. these distributions are graphed here. Which distribution contains the larger sample size? How can you tell?



**4.3 Problems**

**1. Sampling distribution for Telecom (Example 6, Chapter 4)**

Telecom plans to install new equipment that would improve the efficiency of their operations. However, before they can decide if such an investment would be cost-effective, they must determine the probability that the mean of a sample of $n = 35$

*a.* Is between 145 and 150.

*b.* Is greater than 145.

*c.* Is less than 155.

*d.* Is between 145 and 155.

*e.* Is greater than 155.

**2. The Stock Market Panic.** According to a story in Business Week, Yargo Panic (pronounced PAN-eesh) fled Hungary to start BSE Pharmaceutical Inc. in California. Congressional hearings on AIDS research uncovered the fact that BSE had encouraged doctors to use a drug not approved by the FDA. BSE's stock fell dramatically, but Panic assured investors its daily price had averaged $ 22.10, with a standard deviation of $ 9.80. A sample of $n = 50$ days revealed a mean price of $ 20.40. What might this tell you regarding BSE's market performance?

**3. Buying the American Way.** The standard deviation of the purchases of consumers at a particular store is $ 18. A random sample of 100 customers is selected. (*a*) What is the standard error of the sampling distribution? (*b*) What is the probability the sample mean exceeds the population mean by more than $ 5?

**4. The Probability Error.** Batex manufactures telephone dialing systems for Bell Laboratories that are to average 3.2 inches in length, with a standard deviation of 1.1 inches. In examining a sample on $n = 64$ systems, Batex does not want the sampling error to exceed 0.20 inch. If the probability that the sampling error exceeds this tolerable limit is greater than 5 %, Batex must revise its quality control procedures. Given the data here, should Batex take such corrective action?

**5.** From a population with a mean of 53 and a standard deviation of 18, a sample of $n = 49$ is taken. What is the probability the sample mean will exceed 55?

**6.** Temperatures in Tallahassee average 87 degrees, with a standard deviation of 37 degrees. If a sample of $n = 100$ days is selected, what is the probability the sample error will exceed 5 degrees?

**7.** A population of weekly sales (in \$ 1,000's) at Blazing Salads, a vegetarian restaurant in Chicago, is 27, 32, 17, 21, and 32.

*a*. Calculate and interpret $\sigma$.

*b*. Set $n$ = 2 and develop the sampling distribution.

*c*. Calculate and interpret $\sigma_{\overline{X}}$. How does it relate to $\sigma$?

*d*. Calculate and interpret $\mu$.

*e*. Calculate and interpret $\overline{\overline{X}}$. How does it compare with $\mu$?

**8.** Using the data in Problem 4, set $n$ = 3 and

*a*. Develop the sampling distribution.

*b*. Calculate $\overline{\overline{X}}$ and $\sigma_{\overline{X}}$. How do they differ from the values in Problem 4, and why?

**9.** A sample of $n$ = 50 is taken from a large population with a mean of 12.2 and a standard deviation of 4.1. What are the mean and standard error of the sampling distribution of sample means?

**10.** A sampling distribution with n = 10 has a mean of $\overline{\overline{X}} = 105$ and a standard error of 10.2. What are the mean and standard deviation of the population?

**11.** From a large population with mean 75 and standard deviation of 10, samples of $n$ = 40 are taken. Illustrate graphically the sampling distribution. What is its mean and standard error? If samples of $n$ = 80 are taken, how would the sampling distribution compare with that obtained when n = 40? Graphically compare the two distributions of sample means.

**12.** Customers at the Madison Hair Garden, a beauty shop in Madison, Connecticut, average 40.7 per day, with a standard deviation of 12.9. If a sample of 100 days is taken, what is the probability that the mean number of customers will exceed 43?

**13.** Jim Sears manufactures farm equipment. His work requires the use of steel bars which must have a mean length of at least 50 inches. The bars can be purchased from a supplier in Kansas City whose bars average only 47 inches, with a standard deviation of 3.6 inches. If Sears is to buy 81 bars, should he use the supplier in Kansas City or Dallas?

**14.** According to *Business Week*, the average year of experience of airline pilots is 25.2. Assume a standard deviation of 12 years. This year you must make 36 business flights. You hope the average length of experience for the pilots of your flights is over 30. How likely is it that $\overline{X} > 30$?

**15.** The mean deposits at first of America Bank in Peoria are \$ 7,012, with a standard deviation of \$ 532, and they are normally distributed.

*a*. If one deposit is selected at random, what is the probability that it exceeds \$ 6,911?

*b*. If a sample of $n$ = 35 deposits is randomly selected, what is the probability that the mean will exceed \$ 6,911?

*c*. Why does taking a sample decrease your answer? Draw graphs, one directly above the other, to illustrate.

**16.** *The Wall Street Journal* described the problem PepsiCo, makers of soft drinks, had with ensuring that the proper amount of fluid was in each bottle. The quality control department decided to examine the bottling plant in Cleveland to ensure that production standards were being met, by taking a sample of containers, designed to hold 67.7 ounces, to determine if $\overline{X}$ was close to 67.7 ounces as stated on the container. The manager of the plant felt certain that he was, on the average, filling the containers to the correct amount. He therefore wants to maximize the probability that $\overline{X}$ is close to 67.7. Should he insist that a very large sample be taken, or would it be to his benefit to make the sample smaller?

**17.** The mean expenditure for Christmas gifts for a family of four was \$ 612 in 1988, according to USA Today. Assuming a standard deviation of \$ 150 and a normal distribution

*a*. What is the probability a single family of four spent over \$ 600 for Christmas?

*b*. What is the probability 100 families spent over \$ 600 on the average?

*c*. What is the probability 200 families spent over \$ 600 on the average?

### 4.4 Mini-Case Application

During the past school year it has come to light that the county school districts in Kentucky are deficient in their use of funds. A state official claims that the county school boards have spent an average of \$ 110,500 more that allotted by the state government. Some counties have been particularly wasteful, while others have tried to minimize excess expenditures. The standard deviation in overspending is said to be $ 35,000.

The governor orders a study of the financial condition for all school boards in the state. An official sympathetic to the county system hopes to take a sample of 50 counties for the pilot study. The methodology to be employed in the study will be perfected and then applied to the counties. The official is basing her proposal in the hope of taking a sample obtaining a mean of no more than $ 90,000 in overspending. However, the sample mean is $ 145,700. This causes considerable question regarding the estimate that all systems average $ 110,500 in excess spending. Given this uncertainty about the population mean, the governor begins to question the likelihood that any sample of size 50 could be obtained in which the sample error was less than $ 20,000.

How might the principles of a normal distribution be used to aid the governor in his quest for answers? Present your findings in a statistical abstract.

## 5. ESTIMATION

### 5.1 Solved Problems

**Example 1. Estimating Market Values.** The director of finance for a Chicago-based manufacturer wants to analyze the market value of business firms of a similar size. Market value is defined as the number of common shares outstanding, times the share price as listed on an organized exchange.

The results of this analysis will impact on the firm's decision to issue additional shares of stock. A sample of 600 firms reveals a mean market value of \$ 850 million. Assuming the population standard deviation is \$ 200 million, construct a 95 percent confidence interval for the mean market value of all firms of this size.

**Solution**: The point estimate is $\overline{X}$ = \$ 850 million. Since the level of confidence is 95 percent, the figure illustrates that the appropriate area is 0.95/2 = 0.4750.

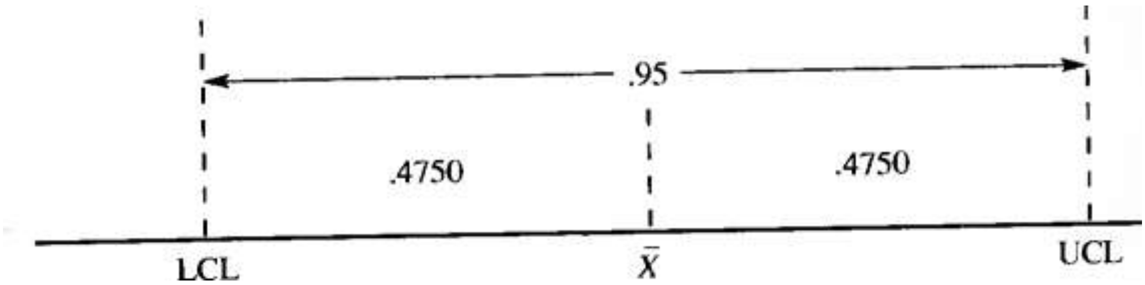



Figure 5.1 – Confidence interval for Example 1

A *Z*–value of 1.96 is called for.

$$C.I.\text{ for } \mu = \overline{X} \pm Z\sigma_{\overline{X}};$$

$$= 850 \pm (1.96)\ \frac{200}{\sqrt{600}} =$$

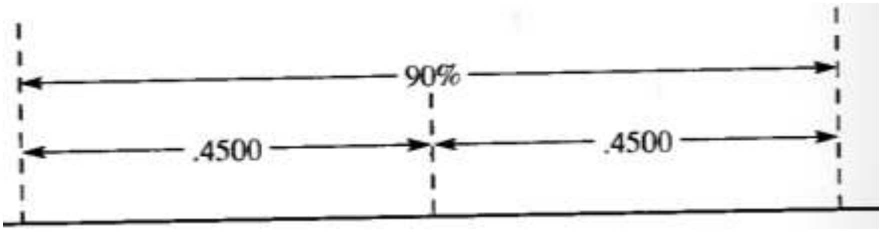
$$= 850 \pm 16.00;$$

$$\$\ 834.00 < \mu < \$\ 866.00.$$

**Interpretation:** The director can be 95 percent confident that the mean market value is, between \$ 834 million and \$ 866 million, and can now proceed with the decision to issue additional stock.

**Example 2. Mean Completion Time: A Quality Control Case.** You have just graduated with a degree in business and have obtained a position with a large manufacturing firm. The director of marketing has asked you to estimate the mean time required to complete a particular unit of the manufacturing process. A sample of 600 units yields a mean of 7.2 days. Since the population standard deviation is unknown, the sample standard deviation of s = 1.9 days must be used. Calculate and interpret the 90 percent interval for the mean completion time for the manufacturing process. If this mean time is estimated to be in excess of 7 days, anew process will be implemented to reduce production costs.

**Solution.**




LCL $\overline{X} = 7.2$ UCL

Figure 5.2 – Confidence interval for Example 2

As shown in the figure, when the 90 percent confidence level is divided by 2, the appropriate area is 0.4500. Again, this division is carried out solely because the table we use gives only those areas from the mean to some value above it (UCL) or some value below it (LCL). The table will not give the Z–value for the 90 percent of the area from the LCL to the UCL.

As you look in the Z–table for 0.4500, you find that the closest values are 0.4495 (Z = 1.64) and 0.4505 (Z = 1.65). The most conservative approach is to use 1.65, which is the practice we will follow here.

$$C.I. \text{ for } \overline{X} \pm Z S_{\overline{X}} =$$

$$= 7.2 \pm (1.65)\ \frac{1.9}{\sqrt{600}} =$$

$$= 7.2 \pm 0.128;$$

$$7.072 \leq \mu \leq 7.328 \text{ days.}$$

**Interpretation**: You can now inform the director that you are 90 percent confident that the mean time required to complete the manufacturing phase is between 7.072 days and 7.328 days. Based on this estimate, the new procedure should now be adopted.

As we noted earlier, a higher level of confidence requires a wider, less precise interval. This next example computes a 99 percent confidence interval for the mean completion time for the production process, and compares the results to those obtained from the 90 percent confidence interval in the previous example.

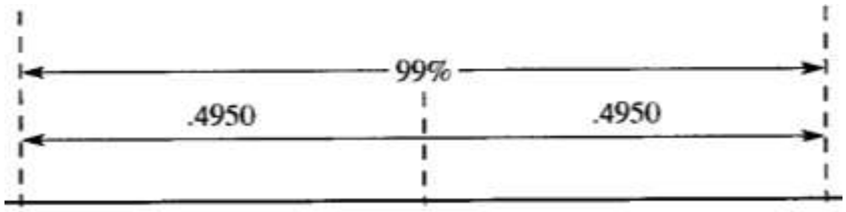
**Example 3. A Comparison of Intervals.** The director of finance now wants to be 99 percent confident that the interval contains the true parameter for the mean completion time for the process under examination. He therefore instructs you to construct a 99 percent interval for $\mu$.

**Solution:**

$$C.I. \text{ for } \mu = \overline{X} \pm Z S_{\overline{X}}$$

$$= 7.2 \pm Z\ \frac{1.9}{\sqrt{600}}.$$

When 0.99 is divided by 2, the resulting value of 0.4950 yields a Z-value of 2.58. Thus,




LCL $\qquad \overline{X} = 7.2 \qquad$ UCL

Figure 5.3 – Confidence interval for Example 3

$$C.I. \text{ for } \mu = 7.2 \pm (2.58)\ \frac{1.9}{\sqrt{600}} =$$

$$= 7.2 \pm 0.20;$$

$$7.00 < \mu < 7.40.$$

**Interpretation:** The 99 percent confidence interval of 7.00 to 7.40 is indeed wider than the 90 percent confidence interval of 7.072 to 7.328 found in the previous example.

Again, this happens because given our sample size of 600, if we desire a greater level of confidence that the interval contains *p*, the interval must be wider. This is reflected in the figure.

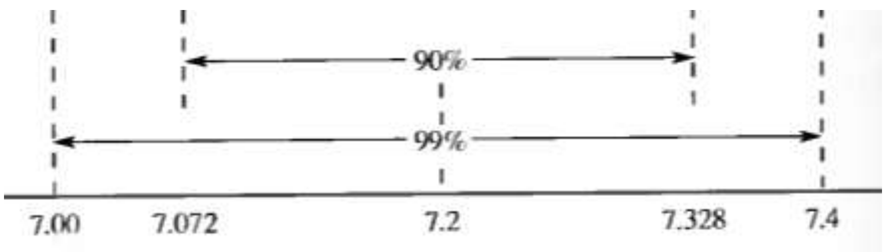



Figure 5.3 – Confidence intervals Comparison for Example 3

**Example 4**. This example demonstrates how the *Federal Bureau of Investigation* used confidence intervals in its study on employee thefts. A recent report by the FBI revealed how data were obtained on the number and extent of thefts by employees in US businesses. Surveillance equipment was installed in a sample of workplaces. A confidence interval for the mean loss due to employee pilferage estimated that loss to be between 7.1 and 10.3 billion dollars annually.

**Example 5. Sample Size for the population mean $\mu$. «It's Snow Problem".** The owner of a small ski resort in southern Wisconsin is considering the purchase of a snowmaking machine to assist Mother Nature in providing a proper base for ski enthusiasts. If the average snowfall seems insufficient, he feels that the machine should soon pay for itself. He plans to estimate the average snowfall in the area, bin has no idea how large his sample should be. He only knows that he wants to be 99 percent confident of his findings and that the error should not exceed 1 inch. The owner promises you season lift tickets if you can help him.

**Solution:** You begin with a large (n > 30) pilot sample that produces a standard deviation of 3.5 inches. Thus,

$$n = \frac{Z^2(s)^2}{(error)^2} = \frac{(2.58)^2(3.5)^2}{(1)^2} = 81.5.$$

or 82 snowfalls over the past several years.

**Interpretation:** You can now collect the data on 82 past snowfalls that will be used to estimate the average snowfall. With this information the owner can determine if Mother Nature needs help. More importantly, you can spend the rest of the winter skiing for free.

**Example 6. Sample Size for intervals of population proportion π.** The city council is planning an ordinance prohibiting smoking in all public buildings including restaurants, taverns, and theaters. Only private housing will be exempt. However, before such an ordinance is brought before the council, this august body wishes to estimate the proportion of residents that supports such a plan. Lacking any statistical skills, the council hires you as a consultant. Your first step will be to determine the necessary sample size. You are told that your error should not exceed 2 percent and that you must be 95 percent confident of your results.

**Solution**: Since no pilot survey was previously taken, you must temporarily set π at 0.5 for the purpose of resolving the sample size.

$$n = \frac{Z^2\pi(1-\pi)}{(error)^2};$$

$$n = \frac{(1.96)^2(0.5)(0.5)}{(.02)^2} = 2.401 \text{ citizens.}$$

**Interpretation:** With the data supplied by 2,401 people you can proceed with your estimate of the proportion of all residents who might favor the ordinance. The council can then make its determination regarding the city-wide smoking policy.

### 5.2 Conceptual Questions

**1.** A 95 % confidence interval is constructed, yielding a lower confidence limit of 62 and an upper confidence limit of 69. Can you conclude from this that is a 95 % probability the parameter is between 62 and 69? Explain.

**2.** Jose has a thriving business in Acapulco selling authentic plastic Inca relics to American tourists. He selects n = 60 days to estimate his daily profit. However, Jose does not know whether the population of daily profits is normally distributed, and is uncertain how to proceed. What should he do?

**3.** The supervisor instructs Mary Martin to construct a 99 % confidence interval for weekly output. Upon receipt of the results, the supervisor then tells Freddie Frost to repeat the effort using the same sample size. Freddie's results differ from Mary's. The supervisor is convinced that either Mary or Freddie is wrong. They must persuade the supervisor otherwise. How can they do this?

**4.** Martha calculates a 95 % confidence interval for the mean number of customers who are delinquent in paying their bills. However, the resulting interval is too wide and lacking in precision. If she doesn't want to increase the α – value, what she can do?

### 5.3 Problems

**1**. If the sample mean is 42.5, construct a 90 percent confidence interval for the population mean if $n = 81$ and $\sigma = 15.3$. Answer: $39.70 \leq \mu \leq 45.31$.

**2**. Given the information in problem 1, construct a 99 percent confidence interval. Why is it different? Answer: $38.11 \leq \mu \leq 46.89$. The interval must be wider since the confidence is greater.

**3**. A sample of 100 observations reveals a mean of 16.3. What is the 95 percent confidence interval if $\sigma = 3.7$? Answer: $15.57 \leq \mu \leq 17.03$.

**4**. Construct the 95 percent confidence interval given the information in problem 3 if ($\sigma - 5$). Explain the difference. Answer: $15.18 \leq \mu \leq 17.42$. The interval must be wider since the data are more dispersed.

**5.** A sample of 20 people drove an average of 27.3 miles to and from work each day, with a standard deviation of 8.3 miles. What is the 90 percent confidence interval for the mean driving distance for the population? Assume distances are normally distributed. Answer: $24.09 \le \mu \le 30.51$.

**6.** Seventeen employees at ABC, Inc. averaged 8.3 sick days last year, with a standard deviation of 1.2 days. Assuming sick days are normally distributed, what is the 98 percent confidence interval for sick days for all of ABC's employees? Answer: $7.55 \le \mu \le 9.05$.

**7.** If the population standard deviation in problem 2 were 1.2 days, what is the 98 percent confidence interval? Answer: $7.62 \le \mu \le 8.98$ by the Z-test since $\sigma$ is known.

**8.** A motel chain wants to determine a confidence interval for the daily mean occupancy of its rooms. It wants an interval of five rooms and a 99 percent confidence level. How large should the sample be if $\sigma^2 = 40$ rooms per day? Answer: $n = 43$ days

**9.** What would happen to the required number of days if the population was less variable? Say, $\sigma^2 = 30$? Answer: n decreases; $n = 32$.

**10** A banker wants to estimate the percentage of people who respond favorably to a new TV advertising campaign. The interval width is to be 5 percent, with a 90 percent level of confidence. Determine *n*. A pilot sample estimated it at 0.85. Answer: 556 people.

**11.Estimating market values**. The director of finance of Chicago – based manufacturer wants to analyze the market value of business firms of a similar size. Market value is defined as the number of common shares outstanding, times the share price as listed on an organized exchange. The results of this analysis will impact on the firm's decision to issue additional shares of stock. A sample of 600 firms reveals a mean market value of $ 850 million. Assuming the population standard deviation is $ 200 million, construct a 95 % confidence interval for the mean market value of all firms of this size.

**12.Mean Completion Time: A Quality Control Case.** You have just graduated with a degree in business and have a position with a large manufacturing firm. The director of marketing has asked you to estimate the mean time required to complete a particular unit of the manufacturing process. A sample of 600 units yields a mean of 7.2. days. Since the population standard deviation is unknown, the sample standard deviation of $s$ = 1.9 days must be used. Calculate and interpret the 90 % interval for the mean completion time for the manufacturing process. If this mean time is estimated to be in excess of 7 days, a new process will be implemented to reduce production costs.

**13.** Consider the following problem taken from a news story in *The Wall Street Journal.* A construction firm was charged with inflating the expense vouchers it files for construction contracts with the federal government. The contract states that a certain type of job should average $ 1,150. In the interest of time, the directors of only 12 government agencies were called on to enter court testimony regarding the firm's vouchers. If a mean of $ 1,275 and a standard deviation of $ 235 are discovered from testimony, would a 95 % confidence interval support the firm's legal case? Assume voucher amounts are normal.

**14. Headhunters in Paradise**. Executive search firms specialize in helping corporations locate and secure top management talent. Called "headhunters", these firms are responsible for the placement of many of the nation's top CEOs. Business week recently reported on the "efforts by headhunters to place executives in a heavenly corporate setting." A source was quoted in the story as saying that "one out of every four CEO's is an outsider – an executive with less than five years at the company he runs." If, in a sample of 350 US corporations, 77 have outsider CEOs, would a 99 % confidence interval support the quote?

**15. Speed kills.** A February 1989 issue of *Fortune* details efforts by companies to increase the rate at which they can develop, produce, and market their products. A survey of 50 companies by Kaiser Associates, a Vienna, Virginia, consulting firm, found that almost all firms placed emphasis on "time-based strategy", as the new approach is called. The attraction to TBS, as one CEO put it, comes from the fact that "speed kills the competition."

**16.** General Electric became concerned about the length of time it took to deliver custom-made circuit breaker panels. The Akron, Ohio, plant felt it averaged about three weeks to deliver a panel after receiving the order. If the last 100 orders averaged 3.4 weeks with a standard deviation of 1.1 weeks, is the estimate of 3 weeks confirmed at the 98 % level?

**17.** In a sample of $n$ = 12 design projects, AT&T took an average of 2.3 years to design and develop a new phone. The design times for the 12 projects had a standard deviation of 1.5 years. If we assume a normal distribution in design times, what is the 90 % confidence interval for mean times?

**18.** As a quality control expert, you want to estimate the mean thickness of optical lenses produced by your firm. A sample of 120 lenses reveals a mean of 0.52 mm. the population standard deviation is known to be 0.17 mm. you feel that you can risk a probability of error of only 1 %. Construct the appropriate confidence interval.

**19.** How would the previous problem change if $\sigma$ was unknown and the sample deviation was 0.17 mm? Calculate the interval.

**20.** Japan's Genichi Taguchi developed the concept of the quality loss function (QLF), which ties the cost of quality directly to the variance in the production process. IBM and several other American companies have embraced Taguchi's teachings in the effort to reduce cost without suffering a drop in product quality. A sample of cost measures for IBM's new coprocessor has yielded a mean of $ 212.10 with a standard deviation of $ 57.10.

**21.** Construct and interpret a 99 % confidence interval if the sample size was 500 coprocessors.

**22.** Construct and interpret a 99 % C.I. if a sample size was 100.

**23.** Explain why the second interval is wider if both carry a 99 percent level of confidence.

**24.** In a random sample of 100 men in the United States, 55 of them were married. Determine an approximate 95 percent confidence interval estimate for the true proportion of men who are married.

**25.** In a random sample of 100 persons, 77 percent of them said that when they pray, they pray for world peace. Determine a 90 percent confidence interval estimate for the true proportion of people who pray for world peace.

**26.** A statistician wishes to estimate, with 99 % confidence, the proportion of people who trust DNA testing. A previous study showed that 91 % of those who were surveyed trusted DNA testing. The statistician wishes to be accurate to within 3 % of the true proportion. What is the minimum sample size necessary for the statistician to carry out the analysis?

**27.** A random sample of 100 public school teachers in a particular state has a mean salary of $ 31,578. It is known from past history that the standard deviation of the salaries for the teachers in the state is $ 4,415. Construct a 99 % confidence interval estimate for the true mean salary for public school teachers for the given state.

**28.** The president of a large community college wishes to estimate the average distance commuting students' travel to the campus. A sample of 64 students was randomly selected and yielded a mean of 35 ml and a standard deviation of 5 ml. Construct a 95 % confidence interval estimate for the true mean distance commuting students travel to the campus.

**29.** What sample size should be selected to estimate the mean age of workers in a large factory to within ±1 year at a 95 % confidence level if the standard deviation for the ages is 3.5 years?

**30.** A study was conducted to determine the type of car owned. A sample of 150 females revealed that 50 of them owned a foreign car.

a) Construct a 90 percent CI for the proportion of females who own foreign cars.

b) Interpret in your own words what the confidence interval in (a) means.

c) If the information given in this problem can be considered as a preliminary study, compute the sample size necessary to construct a 95 percent confidence interval for the true proportion of females who own a foreign car with a maximum error of ±2 percentage points.

**31.** A random sample of 16 public school teachers in a particular state has a mean salary of $ 33,000 with a standard deviation of $ 1,000. Construct a 99 % confidence interval estimate for the true mean salary for public school teachers for the given state.

**32.** The president of a small community college wishes to estimate the average distance commuting students' travel to the campus. A sample of 12 students was randomly selected and yielded the following distances in miles: 27, 35, 33, 30, 39, 25, 38, 22, 27, 37, 33, 40. Construct a 95 % confidence interval estimate for the true mean distance commuting students travel to the campus.

## 5.4 Computer Exercise

**1.** Given these $n$ = 40 observations for the weekly output at a manufacturing plant, construct the 99 % interval. Interpret your results.

| Week | Output | Week | Output |
|---|---|---|---|
| 1 | 23 | 21 | 43 |
| 2 | 32 | 22 | 42 |
| 3 | 43 | 23 | 23 |
| 4 | 43 | 24 | 32 |
| 5 | 47 | 25 | 56 |
| 6 | 45 | 26 | 48 |
| 7 | 65 | 27 | 23 |
| 8 | 34 | 28 | 43 |
| 9 | 64 | 29 | 45 |
| 10 | 45 | 30 | 56 |
| 11 | 65 | 31 | 39 |
| 12 | 38 | 32 | 29 |
| 13 | 27 | 33 | 28 |
| 14 | 64 | 34 | 45 |
| 15 | 74 | 35 | 28 |
| 16 | 58 | 36 | 43 |
| 17 | 59 | 37 | 31 |
| 18 | 49 | 38 | 51 |
| 19 | 38 | 39 | 43 |
| 20 | 47 | 40 | 34 |

**2.** From the list of Fortune 500 companies (which can be found in one of the March or April issues) select a random sample of firms, using Table A of random numbers. Calculate a confidence interval for the mean level of sales or some other variable reported in the list. Interpret your findings.

## 5.5. Empirical Exercises

**1.** Select a random sample of at least 30 from among the students on your campus, or in your dorm, sorority, fraternity, or similar identifiable population. Designate some characteristic of interest such as grade point average, number of dates per week or month, height, age, or other variable that can be easily ascertained. Collect the data. Select a value for α and calculate the confidence interval. Submit your results along with their interpretation and the raw data you collected.

**2.** Select a random sample of at least 50 people. Identify a binary characteristic that has only two possible outcomes, such as married or single, in-state or out-of-state student. Collect the data. Select a value for $\alpha$ and calculate the confidence interval for the proportion of successes. Explain why the characteristic of age or height would not work for this exercise?

# 6 SIMPLE REGRESSION

## 6.1. Solved problems

**Example 1.** The management of Hop Scotch Airlines, the world's smallest air carrier, assumes a direct relationship between *advertising expenditures* and the *number of passengers* who choose to fly Hop Scotch. To determine if this relationship does exist and, if so, what its exact nature might be, the statisticians employed by Hop Scotch set out to use OLS procedures **to determine the regression model**.

Monthly values for advertising expenditures and number of passengers are collected for the $n$ = 15 most recent month. The data are shown in table. "*Passengers*" is labeled as the Y-variable since it is assumed to depend on advertising.

Determine the regression model and interpret the results.

| **Observation (month)** | **Advertising (in \$ 1,000's) (X)** | **Passengers (in \$ 1,000's) (Y)** | **Observation (month)** | **Advertising (in \$ 1,000's) (X)** | **Passengers (in \$ 1,000's) (Y)** |
|---|---|---|---|---|---|
| 1 | 10 | 15 | 9 | 19 | 24 |
| 2 | 12 | 17 | 10 | 10 | 17 |
| 3 | 8 | 13 | 11 | 11 | 16 |
| 4 | 17 | 23 | 12 | 13 | 18 |
| 5 | 10 | 16 | 13 | 16 | 23 |
| 6 | 15 | 21 | 14 | 10 | 15 |
| 7 | 10 | 14 | 15 | 12 | 16 |
| 8 | 14 | 20 | **Total** | **187** | **268** |

**Solution:**

a) Applying OLS method

$$SS_x=\sum X^2 - \frac{(\sum X)^2}{n} = 2{,}469 - \frac{(187)^2}{15} = 137.73333;$$

$$SS_y=\sum Y^2 - \frac{(\sum Y)^2}{n} = 4{,}960 - \frac{(268)^2}{15} = 171.733333;$$

$$SS_{xy}=\sum XY - \frac{(\sum X)(\sum Y)}{n} = 3{,}490 - \frac{(187)(268)}{15} = 148.93333;$$

b) In order to make decisions regarding allocations for the advertising budget, the accounting department for Hop Scotch Airlines must determine the nature of the relationship between advertising expenditures and the number of passengers. The senior accountant recognizes that regression analysis would be of invaluable assistance. Coefficients for simple regression model can be determined as

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{148.93333}{137.73333} = 1.013166 \text{ or } 1.08;$$

$$b_0 = \bar{Y} - b_1\bar{X} = 17.86667 - (1.08)(12.46667) = 4.3865 \text{ or } 4.4.$$

The regression equation is therefore

$$\hat{Y} = 4.40 + 1.08X \,.$$

**Interpretation:** the model tells us that if, for example, \$ 10,000 is spent on advertising ($X$ = 10), then

$$\hat{Y} = 4.40 + 1.08(10) = 15.2.$$

By multiplying 15.2 by 1,000 since Y-values were originally expressed in thousands, we predict on the basis of our model that 15,200 “brave souls” will choose to fly Hop Scotch when \$ 10,000 is spent on advertising.

The meaning of $b_1$ (the regression coefficient): it indicates by how much $Y$ will change for every one-unit change in the $X$-variable. In the case of Example 1, since $b_1$ = 1.08, for every additional \$ 1,000 (which is one unit since $X$ is measured in thousands) that Hop Scotch spends on advertising, 1,080 more passengers will choose the friendly skies of Hop Scotch. Again, this value of 1,080 requires that coefficient $b_1$ is multiplied by \$ 1,000, since passengers was also expressed in thousands.

If advertising is increased by one unit to \$ 11,000, the estimate of total passengers becomes

$$\hat{Y} = 4.40 + 1.08(11) = 16.28 \text{ or } 16{,}280 \text{ passengers.}$$

That means, if X is increased from 10 to 11, the predicted number of passengers is 16,280. This is exactly 1,080 more than the 15,200 passengers predicted to fly if X = 10. Such information is useful in determining if an increase in the advertising budget is justified.

## 6.2 Conceptual Questions

1. What is meant by "minimizing the sum of the errors squared" in your model for the trucking firm in Problem 1?
2. In what way might autocorrelation and heteroscedasticity present a problem in your regression model?
3. What is the difference between regression and correlation?
4. Identify the dependent and independent variables in each case:
   a. Time spent working on a term paper and the grade received.
   b. Height of a son and height of a father.
   c. A woman's age and the cost of her life insurance.
   d. Price of a product and the number of units purchased by an individual.
   e. Demand for a product and the number of consumers in the market.

## 6.3 Problems

### 1. The relationship between Income and Flights

The CEO feels that the number of flights people take depends in some manner upon their income. Flights are therefore seen as the dependent variable, while income assumes the role of the independent variable. Simple regression analysis can provide the CEO with the precise knowledge he desires regarding the relationship between these two variables.

| Passenger | Flights | Income | Passenger | Flights | Income |
|---|---|---|---|---|---|
| 1 | 5 | 30 | 6 | 8 | 54 |
| 2 | 4 | 27 | 7 | 9 | 42 |
| 3 | 7 | 38 | 8 | 11 | 63 |
| 4 | 10 | 48 | 9 | 8 | 52 |
| 5 | 11 | 59 | 10 | 9 | 47 |

**2. A Keynesian Consumption Function.** In his famous 1936 book, *A General Theory of Employment, Interest and Money*, the noted British economist John Maynard Keynes proposed a theoretical relationship between income and personal consumption expenditures. Keynes argued that as income went up, consumption would rise by a smaller amount. This theoretical relationship has been empirically tested many times since 1936. Milton Friedman, former professor of economics at the University of Chicago, and winner of the Nobel Prize in economics, collected extensive data on income ($I$) and consumption ($C$) in the US over a long period of time. Shown in a table below are 10 observations on annual levels of consumption and income used by Friedman in his study. Using these data, derive a consumption function under the assumption that there is exists a linear relationship between ($C$) and ($I$). Figures are in billions of dollars.

| Year | Income | Consumption | Year | Income | Consumption |
|---|---|---|---|---|---|
| 1 | 284.8 | 191.0 | 6 | 398.0 | 254.4 |
| 2 | 328.4 | 206.3 | 7 | 419.2 | 266.7 |
| 3 | 345.5 | 216.7 | 8 | 441.1 | 281.4 |
| 4 | 364.6 | 230.0 | 9 | 447.3 | 290.1 |
| 5 | 364.8 | 236.5 | 10 | 483.7 | 311.2 |

**3. Lee Iacocca's Financial Inquiry.** Lee Iacocca, chairman and CEO of Chrysler Corporation, expressed a concern regarding the company's high cost structure following the acquisition of AMC, and what he called "skimpy profits in the face of rising sales". In 1988 he ordered company executives to undertake a concerned study of Chrysler's cost structure as it related to reported sales. The data on K-car production shown here were collected. Company analyst used them to construct a regression model depicting the manner in which costs depended on production and subsequent sales volume. Costs were therefore taken as the depended variable. Figures for costs are in units of \$ 100,000, and values for sales are in \$ 1,000,000. The data are monthly values for the company as a whole. Construct and interpret a simple regression model. Use $\alpha$ = 5 % for calculating *C.I.* for $\beta_1$.

| Month | Costs | Sales | Month | Costs | Sales |
|---|---|---|---|---|---|
| 1 | 15.8 | 23 | 6 | 13.5 | 19 |

| | | | | | |
|---|---|---|---|---|---|
| 2 | 12.3 | 18 | 7 | 13.7 | 20 |
| 3 | 14.5 | 21 | 8 | 15.9 | 22 |
| 4 | 15.7 | 23 | 9 | 13.7 | 19 |
| 5 | 12.7 | 18 | 10 | 14.3 | 21 |

**4.** Consider the following data set showing copper and zinc prices in dollar per kilogram.

| **# (i)** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| **Zinc** | 1.64 | 1.66 | 1.64 | 1.65 | 1.64 | 1.61 | 1.62 | 1.59 | 1.57 |
| **Copper** | 9.71 | 9.54 | 9.43 | 9.41 | 9.41 | 9.40 | 9.30 | 9.18 | 9.00 |

Fitting the regression line to the data yields

$$a_0 = 0.0790,\ a_1 = 5.7269,\ R^2=0.716,\ \bar{x} = 1.6233,\ \bar{y} = 9.3756,$$

$$SSE = 0.0937,\ \sum x^2 = 23.7241,\ \sum y^2 = 791.4392, \sum xy = 137.0181.$$

And the following *p*-values with a two-sided alternative

| $H_0$ | $a_0 = 0$ | $a_1 = 0$ |
|---|---|---|
| ***p*-value** | 0.9725 | 0.0040 |

a. Interpret $a_0$ , $a_1$, the *p*-values (use $\alpha = 0.05$) and $R^2$.

b. Determine a 90 % confidence interval for the slope.

c. Predict the copper price for a zinc price of $x = 1.60$ and calculate a 95 % prediction.

**5.** Data for the consumption of beef product s reported by the agricultural division at Florida State University are shown for 10 Florida counties. Figures are for May 1988 and are on a per-capita basis.

| **Country** | **Consumption(pounds)** | **Price(per pound)** |
|---|---|---|
| Dade | 6,5 | 3,19 |
| Taylor | 6,7 | 2,99 |
| Broward | 6,4 | 3,22 |
| Leon | 6,4 | 3,34 |

| | | |
|---|---|---|
| Duval | 6,9 | 2,85 |
| Alachua | 7,0 | 2,73 |
| Dixie | 6,5 | 3,04 |
| Okaloosa | 6,5 | 3,09 |
| Manatee | 6,3 | 2,88 |
| Orange | 6,4 | 2,91 |

The governor of Florida requests that state economists in Tallahassee estimate the linear demand curve for beef products.

a. Which is the dependent variable? (hint: A demand curve can he expressed with either price (*P*) or quantity(*Q*) in the dependent role. However, when, as in this case, the analysis seems to view the issue from the standpoint of the consumer rather than the producer, it is customary to argue that *Q* is a function of *P*).

b. Estimate the linear demand cure, using OLS

c. Interpret the results

d. What economic principle dictates that the regression coefficient should carry a negative sign?

**6.** Does studying really pay off? To answer this question, a curious student in a statistics class asked 10 students how many hours they studied for the most recent test and the grade they received. The data are recorded here.

a. Based on the coefficient of the regression model (or more aptly, its sign) what do you conclude?

b. If you study one more hour, what will happen to your grade according the model?

| **Grade** | **Hours** |
|---|---|
| 89 | 25 |
| 92 | 26 |
| 32 | 12 |
| 92 | 32 |
| 90 | 29 |
| 30 | 10 |
| 87 | 21 |
| 88 | 27 |
| 34 | 15 |
| 30 | 18 |

**7.** To reduce crimes, the president has budgeted more money to put more police on our city streets. What information does the regression model offer based on these data for the number of police on patrol and the daily number of reported crimes? Use the formulas that illustrate that the OLS model is indeed based on deviations from the mean by calculating.

| Police | Number of Reported Crimes |
|---|---|
| 13 | 8 |
| 15 | 9 |
| 23 | 12 |
| 25 | 18 |
| 15 | 8 |
| 10 | 6 |
| 9 | 5 |
| 20 | 10 |

**8.** As a safety feature, Gulf Leisure limits the horsepower in the jet skis it rents to tourists in Tampa, Florida. The intent is to prevent inexperienced drivers from going too fast. Do these data suggest that controlling horsepower will accomplish this goal based on the regression coefficient? Fully interpret the results.

| Speed(mph) | Horsepower |
|---|---|
| 50 | 35 |
| 35 | 20 |
| 45 | 35 |
| 47 | 40 |
| 60 | 50 |
| 65 | 60 |
| 72 | 65 |
| 37 | 30 |

**9.** Aunt Bea wants to get more yield from her Big Boy tomato plants this summer by increasing the number of times she uses fertilizer. Based on the data shown here, does the coefficient for the regression model suggest this is possible? Use the formulas that illustrate that the OLS model is indeed based on deviations from the mean by calculating.

| Use of Fertilizer | Yield(pounds) |
|---|---|
| 4.00 | 12.00 |
| 9.00 | 20.00 |
| 5.00 | 15.00 |
| 8.00 | 17.00 |
| 2.00 | 7.00 |

**10.** Twelve school districts in the Chicago area were interested in whether tensing properly tax rates could be associated with the number of pupils in a classroom in the local schools. Does this seem to be the case based on the data shown here?

| Tax assessment Rates | Pupils per Class |
|---|---|
| 1.20 | 32 |
| 1.20 | 36 |
| 1.10 | 25 |
| 1.30 | 20 |
| 1.10 | 39 |
| 1.20 | 42 |
| 1.30 | 25 |
| 1.30 | 21 |
| 1.20 | 35 |
| 1.40 | 16 |
| 1.40 | 39 |
| 1.30 | 27 |

a. If it is thought that more pupils require higher taxes, which is the dependent variable? Calculate and interpret the regression model. Do larger classes seem to be associated with higher taxes?

b. Calculate: and interpret the coefficient of determination and the correlation coefficient. Does it seem this model is useful?

c. Calculate and interpret the standard error of the estimate.

**11.** Based on figures released by the Internal Revenue Service, a national group of citizens has expressed concern that the budget for the IRS has not been used effectively. The IRS argued that an increase in the number of taxpayers filing returns explains the budget problems. Relevant data are provided here.

| Tax returns | IRS Budget(in billions of dollars) |
|---|---|
| 116 | 6.7 |
| 116 | 6.2 |
| 118 | 5.4 |
| 118 | 5.9 |
| 120 | 3.7 |
| 117 | 5.9 |
| 118 | 4.7 |
| 121 | 4.2 |

a. Construct the regression model. Does the IRS argument seem plausible?

b. Calculate and interpret the coefficient of determination.

c. Calculate and interpret the standard error of the estimate.

**12.** It was recently reported in Financial Weekly that E.F. Hutton was interested in the relationship between a person's income and the amount of money they had invested in the stock market. Fifty individuals were randomly selected on the presumption that an individual's investments in the stock market are influenced by his or her income. Using regression and correlation analysis, do these data suggest such a relationship? (Data are in thousands of dollars)

a Identify The dependent and independent variables.

b. calculate and interpret the results of the regression model.

c. are the data cross-sectional or time-series?

$$\sum x = 9{,}385; \quad \sum y = 988.1; \quad \sum x^2 = 3{,}025{,}553; \quad \sum y^2 = 32{,}224.51;$$
$$\sum xy = 303{,}471.3.$$

**13.** Given the six values for *Y* and *X* shown in table

| | | | | | | |
|---|---|---|---|---|---|---|
| *Y* | 1 | 2.5 | 4 | 3 | 3.2 | 5.2 |
| *X* | 5 | 3 | 2.5 | 3 | 3.2 | 3 |

a. Determine the regression model

b. What happens to $Y$ if $X$ increases by 1 unit?

c. What is $Y$ if $X = 0$?

**14.** Given the seven values for $Y$ and $X$ shown in table

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $Y$ | 9 | 8 | 5.3 | 5 | 6.1 | 7.1 | 8 |
| $X$ | 3 | 1.5 | 1 | 1 | 2.5 | 3 | 4 |

a. Determine the regression model

b. What value would you predict for $Y$ if $X = 4$?

c. What is $Y$ if $X = 0$?

**15.** The student government at the local university is trying to determine if the admission price to the game room in the student center has an impact on the number of students who use the facilities. The cost of admission and the number of students who enter the room are recorded for 12 successive Friday nights and shown here. Construct and interpret the regression model. Use $\alpha = 10$ % for calculating *C.I.* for $\beta_1$.

| Number of tickets | Price | Number of tickets | Price |
|---|---|---|---|
| 95 | 1,25 | 98 | 1 |
| 83 | 1,5 | 85 | 1,5 |
| 75 | 1,75 | 75 | 2 |
| 72 | 2 | 65 | 2,5 |
| 69 | 2,1 | 98 | 1,1 |
| 101 | 1 | 86 | 1,5 |

**16.** Given the seven values for $Y$ and $X$ shown in table

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $Y$ | 8.2 | 7 | 9 | 10 | 8 | 7.2 | 9.5 |
| $X$ | 10 | 6.5 | 8 | 12 | 6 | 7 | 11 |

a. Determine the regression model

b. What happens to $Y$ if $X$ increases by 1 unit?

c. What is $Y$ if $X = 0$?

# 7 TIME-SERIES ANALYSIS AND FORECASTING

## 7.1 Solved problems

**Example 1. A Smooth Rhodes**

a) Ralph Rhodes wishes to use smoothing techniques to average out and forecast levels of capital investments his firm has made over the past several years. He calculated both three-year and four-year moving averages. The four-year MA, since it contains an even number of terms, must subsequently be centered.

| Year | Investment ($ 1,000) (Y) | Three-Term MA | Four-Term MA | Centered Four-Term MA |
|---|---|---|---|---|
| 1985 | 73.2 | | | |
| 1986 | 68.1 | 71.37 | 72.50 | |
| 1987 | 72.8 | 72.27 | 72.15 | 72.33 |
| 1988 | 75.9 | 73.50 | 72.45 | 72.30 |
| 1989 | 71.8 | 72.33 | 71.25 | 71.85 |
| 1990 | 69.3 | 69.70 | 69.15 | 70.20 |
| 1991 | 68,0 | 68.27 | 68.68 | 68.91 |
| 1992 | 67.5 | 68.47 | 69.65 | 69.16 |
| 1993 | 69.9 | 70.20 | 71.48 | 70.56 |
| 1994 | 73.2 | 72.80 | 72.83 | 72.15 |
| 1995 | 75.3 | 73.80 | | |
| 1996 | 72.9 | | | |

Using the three-term MA, 73.8 is the estimate of the long-run average around which all observations to fall and, as such, is the forecast for any future time period. The four-term MA produces an estimate of 72.15.

b) Ralph also decides to forecast using exponential smoothing. Smoothing constants of α=0.2 and α=0.7 are employed. What forecasts for 1997 would Ralph get, and which is probably more accurate?

If $\alpha = 0.2$,

$$F_{t+1} = \alpha A_1 + (1-\alpha)F_t,$$
$$F_{1986} = (0.2)(73.2) + (0.8)(73.2) = 73.,2$$
$$F_{1987} = (0.2)(68.1) + (0.8)(73.2) = 72.18,$$
$$F_{1997} = (0.2)(72.9) + (0.8)(71.82) = 72.04.$$

I $\alpha = 0.7$,

$$F_{1986} = (0.7)(73.2) + (0.3)(73.2) = 73.2,$$
$$F_{1987} = (0.7)(68.1) + (0.3)(73.2) = 69.63,$$
$$F_{1988} = (0.7)(72.8) + (0.3)(69.63) = 71.85,$$
$$F_{1997} = (0.7)(72.9) + (0.3)(74.32) = 73.33.$$

| **Year** | **Actual Investment** | **Forecast (α = 0.2)** | **Error** | **Forecast (α = 0.7)** | **Error** |
|---|---|---|---|---|---|
| 1985 | 73.2 | | | | |
| 1986 | 68.1 | 73.2 | 5.10 | 73.2 | 5.10 |
| 1987 | 72.8 | 72.18 | –0.62 | 69.63 | –3.17 |
| 1988 | 75.9 | 72.30 | –3.60 | 71.85 | 4.05 |
| 1989 | 71.8 | 73.02 | 1.22 | 74.68 | 2.88 |
| 1990 | 69.3 | 72.78 | 3.48 | 72.67 | 3.37 |
| 1991 | 68,0 | 72.08 | 4.08 | 70.31 | 2.31 |
| 1992 | 67.5 | 71.27 | 3.77 | 68.69 | 1.19 |
| 1993 | 69.9 | 70.51 | 0.61 | 67.86 | –2.04 |
| 1994 | 73.2 | 70.39 | –2.81 | 69.29 | –3.91 |
| 1995 | 75.3 | 70.95 | –4.35 | 72.03 | –3.27 |
| 1996 | 72.9 | 71.82 | –1.08 | 74.32 | 1.42 |
| | | 72.04 | | 73.33 | |

The more accurate, reliable forecast would be the one producing the smaller MSE, for α=0.2,

$$MSE = \frac{\sum (F_t - A_t)^2}{n-1} = \frac{(5.10)^2 + (-0.62)^2 + ... + (-1.08)^2}{12-1} = 10.20.$$

For α=0.7,

$$MSE = \frac{(5.10)^2 + (-3.17)^2 + ... + (1.42)^2}{12-1} = 10.10.$$

The forecast values using $\alpha = 0.7$ appear more accurate on the average.

**Example 2. A Dark Rainbow**

For the past several years, business conditions for Rainbow Enterprises who have been rather black. The CEO has collected quarterly totals of the number of employees who have been laid off over the past four years.

a) The CEO would like to forecast the number of layoffs for the first and second quarters of 1995, using linear trend analysis.

| Time | Layoffs(Y) | t(X) | XY | $X^2$ |
|---|---|---|---|---|
| 1991-I | 25 | 1 | 25 | 1 |
| II | 27 | 2 | 54 | 4 |
| III | 32 | 3 | 96 | 9 |
| IV | 29 | 4 | 116 | 16 |
| 1992-I | 28 | 5 | 140 | 25 |
| II | 32 | 6 | 192 | 36 |
| III | 34 | 7 | 238 | 49 |
| IV | 38 | 8 | 304 | 64 |
| 1993-I | 35 | 9 | 315 | 81 |
| II | 37 | 10 | 370 | 100 |
| III | 37 | 11 | 407 | 121 |
| IV | 39 | 12 | 468 | 144 |
| 1994-I | 38 | 13 | 494 | 169 |
| II | 42 | 14 | 588 | 196 |
| III | 44 | 15 | 660 | 225 |
| IV | 45 | 16 | 720 | 256 |
| **Total** | **562** | **136** | **5187** | **1496** |
| **Average** | **35.1** | **8.5** | | |

$$SSx = \sum X^2 - \frac{(\sum X)^2}{n} = 1496 - \frac{(136)^2}{16} = 340;$$
$$SSxy = \sum XY - \frac{(\sum X)(\sum Y)}{n} = 5187 - \frac{(136)(562)}{16} = 410;$$
$$b_1 = \frac{410}{340} = 1.206;$$
$$b_0 = \overline{Y} - b_1\overline{X} = 35.13 - 1.206(8.5) = 24.88.$$

For the first quarter of 1995,

$$\hat{Y}_t = 24.88 + 1.206(17) = 45.38.$$

For the second quarter of 1995,

$$\hat{Y}_t = 24.88 + 1.206(18) = 46.59.$$

b) The CEO now wants to develop the seasonal indexes for the number of layoffs.

| Time | Layoffs(Y) | Centered MA | Ratio to MA |
|---|---|---|---|
| 1991-I | 25 | | |
| II | 27 | | |
| III | 32 | 28.625 | 1.1179 |
| IV | 29 | 29.625 | 0.9789 |
| 1992-I | 28 | 30.500 | 0.9180 |
| II | 32 | 31.875 | 1.0039 |
| III | 34 | 33.875 | 1.0037 |
| IV | 38 | 35.375 | 1.0742 |
| 1993-I | 35 | 36.375 | 0.9622 |
| II | 37 | 36.875 | 1.0034 |
| III | 37 | 37.375 | 0.9900 |
| IV | 39 | 38.375 | 1.0163 |
| 1994-I | 38 | 39.875 | 0.9530 |
| II | 42 | 41.500 | 1.0120 |
| III | 44 | | |
| IV | 45 | | |
| | 562 | | |

The four-term (since quarterly data are used) MA is calculated and centered, followed by the ratio to MA. The mean ratio to MA is then determined for each quarter. Since the means sum to 4.0111, the normalization ratio is 4/4.0111 = 0.997. The seasonal indexes are obtained by multiplying each ratio to MA by 0.997.

| | 1991 | 1992 | 1993 | 1994 | Mean | Seasonal Indexes |
|---|---|---|---|---|---|---|
| I | | 0.9180 | 0.9622 | 0.9530 | 0.9444 | 0.9416 |
| II | | 1.0039 | 1.0034 | 1.0120 | 1.0064 | 1.0034 |
| III | 1.1179 | 1.0037 | 0.9900 | | 1.0372 | 1.0341 |
| IV | 0.9789 | 1.0742 | 1.0163 | | 1.0231 | 1.0200 |
| | | | | | 4.0111 | 3.9991 $\approx$ 4 |

c) The CEO for Rainbow wants to determine layoffs if the seasonal factors are eliminated. Deseasonalized levels of layoffs for 1991-I and 1991-II are, respectively

$$\frac{25}{0.9416} = 26.55 \text{ employees,}$$

and

$$\frac{27}{1.0034} = 26.91 \text{employees.}$$

d) Rainbow executives think that general movements in the business cycle influence their need to lay off employees. They decide to calculate the cyclical components for each time period.

| **(1) Time Period** | **(2) Layoffs** | **(3) Trend Projection** | **(4) Seasonal Index** | **(5) Statistical Norm (3)*(4)** | **(6) Cyclical-Irregula [(2)/(5)]*100** | **(7) Cyclical Component 4-term MA** |
|---|---|---|---|---|---|---|
| 1991-I | 25 | 26.08 | 0.9416 | 24.56 | 101.80 | |
| II | 27 | 27.29 | 1.0034 | 27.38 | 99.61 | |
| III | 32 | 28.49 | 1.0341 | 29.46 | 108.62 | 100.5 |
| IV | 39 | 29.70 | 1.0200 | 30.29 | 95.73 | 99.61 |
| 1992-I | 28 | 30.90 | 0.9416 | 29.10 | 96.24 | 98.73 |
| II | 32 | 32.11 | 1.0034 | 32.22 | 99.32 | 99.01 |
| III | 34 | 33.32 | 1.0341 | 34.46 | 98.68 | 101.51 |
| IV | 38 | 34.52 | 1.0200 | 35.21 | 107.92 | 102.55 |
| 1993-I | 35 | 35.73 | 0.9416 | 33.64 | 104.03 | 102.01 |
| II | 37 | 36.93 | 1.0034 | 37.06 | 99.85 | 100.06 |
| III | 37 | 38.14 | 1.0341 | 39.44 | 93.81 | 98.15 |
| IV | 39 | 39.35 | 1.0200 | 40.14 | 97.17 | 97.64 |
| 1994-I | 38 | 40.55 | 0.9416 | 38.18 | 99.52 | 98.34 |
| II | 42 | 41.76 | 1.0034 | 41.90 | 100.23 | 99.33 |
| III | 44 | 42.96 | 1.0341 | 44.42 | 99.04 | |
| IV | 45 | 44.17 | 1.0200 | 45.05 | 99.88 | |

They take a four-term MA of the cyclical-irregular values to produce just the cyclical component in column (7). For 1991-III, the layoff of 32 employees represents 100.5 percent of the trend.

e) If layoffs in 1995–I are 46, what might Rainbow expect total layoffs for 1995 to be? Since the first quarter typically represents a period in which layoffs are only 94.16 percent of the average for the full year, quarterly layoffs based on 46 for 1995–I , would be

$$\frac{46}{0.9416} = 48.85.$$

For the whole year, layoffs would total (48.85) * (4) = 195 employees.

f) In a final effort to control the number of necessary layoffs, Rainbow executives wish to obtain deseasonalized figures for each time period. They obtain these by dividing the actual number of layoffs by the appropriate seasonal (quarterly) index. A partial listing of the results is shown.

| **Year-Quarter** | **Layoffs** | **Seasonal Index** | **Deseasonalized Layoffs** |
|---|---|---|---|
| 1991-I | 25 | 0.9416 | 26.55 |
| II | 27 | 1.0034 | 26.91 |
| III | 32 | 1.0341 | 30.94 |
| IV | 29 | 1.0200 | 28.43 |
| 1992-I | 28 | 0.9416 | 29.7 |

The deseasonalized values represent the number of layoffs when seasonal forces have been eliminated.

### 7.2. Conceptual Questions

**1.** As a new business manager for Rocky's Bar and Grill, you wish to examine the trend in sales for the last several months. However, receipts behave so erratically that no pattern seems evident. How might you manipulate these data so as to detect a pattern?

**2.** Given the conditions in the previous problem, if you were to engage in exponential smoothing, should you use a small value for $\alpha$ or a large value? Explain.

**3.** Sales at Acme, Inc., are thought to fluctuate around a long-term mean. Sometimes they exceed that mean; at other times sales are below that long-run average. What is the best method of forecasting those sales?

**4.** Define:

- Secular trend.
- Cyclical variations.
- Irregular variations
- Seasonal variations.

**5.** Which components of a time series are represented by a 12-month (or four-quarter) moving average?

**6.** How can you best remove the T and C components from a time series?

**7.** The seasonal index in January for a particular series is greater than 1. Will the seasonally adjusted value be greater than or less than the original value? Explain.

**8.** What does it mean to say that the seasonal index is greater than 1? Less than 1?

**9.** What components of a time series are contained in the ratio to moving average?

## 7.2 Problems

**1.** Car-R-US has recorded sales (in$ 1,000s) over the last three years.

| **Month** | **1993** | **1994** | **1995** |
|---|---|---|---|
| January | 17,2 | 18,1 | 16,3 |
| February | 18,7 | 19,2 | 17,3 |
| March | 19,7 | 20,3 | 18,5 |
| April | 20,2 | 21,5 | 20,3 |
| May | 21,7 | 22,0 | 21,0 |
| June | 23,1 | 24,7 | 25,0 |
| July | 24,2 | 23,9 | 22,7 |
| August | 25,7 | 26,2 | 25,0 |
| September | 21,2 | 22,0 | 21,9 |
| October | 19,3 | 18,0 | 17,3 |
| November | 22,7 | 19,7 | 21,2 |
| December | 19,3 | 17,3 | 16,2 |

a) Plot the data. Does there appear to be any trend in the data? Any cyclical or seasonal variation?

b) Compute a 12-month moving average. Which component(s) do these values reflect?

**2** Calculate the seasonal indexes for each month using the data for Cars-R-US from the previous problem.

**3** In Problem 10, what are the seasonally adjusted sales figures for the first six months of 1995? How do you interpret them?

**4** In Problem 10, what are the deseasonalized values for the last six months of 1995? How would you interpret them?

**5** *Business Monthly* recently reported the dollar value of "perks" received by business executives over the past several years. These data do not include that portion of the executives' healthcare paid by the employer, and are adjusted for inflation. Use linear trend analysis to predict the value for the year 2000. How well does the model explain the trend in perk levels?

| Year | **Perk** | **Year** | **Perk** |
|---|---|---|---|
| 1980 | $ 3,200 | 1989 | $ 4,280 |
| 1981 | 3,640 | 1990 | 4,450 |
| 1982 | 3,850 | 1991 | 4,500 |
| 1983 | 3,700 | 1992 | 4,490 |
| 1984 | 3,920 | 1993 | 4,560 |
| 1985 | 3,880 | 1994 | 4,680 |
| 1986 | 3,950 | 1995 | 4,790 |
| 1987 | 4,100 | | |
| 1988 | 4,150 | | |

**6** According to *Business Week*, Taiwan's market share of the world's microchip industry on a quarterly basis has been as shown here in percentages:

| | **1989** | **1990** | **1991** |
|---|---|---|---|
| I | 1.1 | 1.4 | 2.5 |
| II | 1.3 | 1.8 | 2.8 |
| III | 1.6 | 2.1 | 3.1 |
| IV | 1.5 | 2.6 | 3.1 |

a) Use a four-period moving average to remove seasonal variations. Calculate the seasonal indexes.

b) Compute the deseasonalized market shares.

**7** Inventories for Bake-O-Donuts for the past two years were

| **Month** | **1994** | **1995** |
|---|---|---|
| January | $ 87 | $ 95 |
| February | 93 | 102 |
| March | 102 | 112 |
| April | 112 | 115 |
| May | 93 | 99 |
| June | 82 | 90 |
| July | 80 | 83 |
| August | 73 | 79 |
| September | 93 | 84 |
| October | 102 | 89 |
| November | 115 | 92 |
| December | 112 | 91 |

a) Use a 12-period moving average to remove seasonal variations.

b) Calculate the seasonal indexes.

c) What are the seasonally adjusted inventory levels?

**8** Use a two-period moving average to forecast the earnings per share (EPS) in January 1995 for Goober's Salvage Yard. What would the forecast be for July of 1995?

| **1994** | **EPS** |
|---|---|
| January | 0.65 |
| February | 0.72 |
| March | 0.68 |
| April | 0.51 |
| May | 0.62 |
| June | 0.73 |
| July | 0.74 |
| August | 0.78 |
| September | 0.60 |
| October | 0.57 |
| November | 0.63 |
| December | 0.65 |

**9** *Fortune* reported Dynatech's quarterly profits (in millions of dollars) as shown here:

| | 1989 | 1990 | 1991 |
|---|---|---|---|
| I | 4.2 | 5.9 | 6.5 |
| II | 3,8 | 5.2 | 7.0 |
| III | 4.9 | 6.3 | 7.5 |
| IV | 5.8 | 6.8 | 8.0 |

a) Graph the data. Does a cyclical component appear? Does a trend appear?

b) The CEO for Dynatech wants to achieve profits of 10.5 by the fourth quarter of 1992. Use trend analysis to determine if the goal might be reached.

c) Compute the seasonal indexes for all four quarters and the deseasonalized values.

**10** Using the data from the previous problem, isolate the cyclical component.

**11** Mopeds, Inc., is concerned about slumping sales. If monthly sales fall below $ 9,000 the Northeast regional office must be closed down. According to the figures shown here, is that likely to occur within the next five months? Figures are in thousands.

| **1994** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| J | F | M | A | M | J | J | A | S | O | N | D |
| 18 | 17.3 | 16.9 | 18.1 | 16.8 | 16.3 | 15.1 | 14.5 | 14 | 14.5 | 14 | 13.1 |
| **1995** | | | | | | | | | | | |
| 13.9 | 13.1 | 12.8 | 12.4 | 11.8 | 11.9 | 11.7 | 11.5 | 11.1 | 11.2 | 11.2 | 11.1 |

**12** Using the data from the previous problem, calculate the seasonal indexes.

**13** Using the data for Mopeds, Ins., what is the strength of the relationship between sales and time? Plot the trend line against the actual data.

**14** From the regression model you calculated in the problem for Mopeds, Inc., what is the average monthly change in sales?

**15** Shareholder wealth measured in millions of dollars for a major international corporation is shown here for a three-year-period. Calculate the seasonal indexes and the deseasonalized values.

| | **1992** | **1993** | **1994** |
|---|---|---|---|
| January | 42.10 | 47.50 | 48.00 |
| February | 43.00 | 45.20 | 47.20 |
| March | 45.00 | 48.50 | 45.10 |
| April | 41.20 | 49.00 | 45.60 |
| May | 46.00 | 51.00 | 48.60 |
| June | 46.90 | 52.00 | 48.00 |
| July | 47.20 | 51.20 | 46.50 |
| August | 47.90 | 51.40 | 41.20 |
| September | 47.80 | 54.00 | 42.50 |
| October | 48.00 | 52.20 | 45.40 |
| November | 48.20 | 57.00 | 48.20 |
| December | 47.20 | 53.00 | 49.50 |

**16** John Wolf feels that exponential smoothing with an $\alpha$ -value of 0.8 can best forecast September inventories of his medical supply firm. His brother and business partner thinks an $\alpha$ of 0.4 should be used. What is the forecast in each case, and who is correct based on these values for inventories per month?

| **Inventories** | J | F | M | A | M | J | J | A |
|---|---|---|---|---|---|---|---|---|
| **($ 100)** | 41 | 48 | 37 | 32 | 45 | 43 | 49 | 38 |

**17** An economist for the president's task force on health care claims that costs have risen dramatically, as shown by these data. The projection for the year 2000 is $ 1,500 billion.

| **Year** | **Cost in Billions** |
|---|---|
| 1950 | $ 13 |
| 1960 | 27 |
| 1970 | 75 |
| 1980 | 248 |
| 1990 | 510 |

a) Graph the data. Does it appear a linear model will forecast accurately? If no, what type of logarithmic transformation is necessary?

b) Develop the trend model and project costs for the year 2000.

**18.** By about what percentage are health costs increasing each decade according to the previous problem?

**19.** What is the strength of the relationship between time and health care costs based on the logarithmic model developed in Problem 28?

**20.** Calculate a linear model using the earlier data for health care costs. How does it compare in its ability to explain costs compared with the logarithmic model? Plot the original data along with the predicted values based on both models. Comment.

**21** CNN News reported that employee thefts have been on the rise. Figures for Livingston Plastics are shown here. Livingston is considering as a security system, which should reduce thefts to virtually zero by August 1992. The security system will cost approximately $ 1,500 per month. Basing your decision on potential savings by August 1992, should the security system be implemented?

| **1991** | **Thefts ($ 100)** |
|---|---|
| January | |
| February | 10.1 |
| March | 12.3 |
| April | 11.7 |
| May | 12.2 |
| June | 13.7 |
| July | 13.5 |
| August | 14.1 |
| September | 14.9 |
| October | 15.6 |
| November | 15.1 |
| December | 16.2 |
| **1992** | |
| January | 16.5 |
| February | 16.5 |
| March | 17.1 |

**22** How well does the linear model in the previous problem describe the trend in employee thefts?

**23** The chief economist for the US Department of Commerce reported these seasonally adjusted values for the consumption of durable goods for 1989. Values are in billions of dollars.

| **J** | **F** | **M** | **A** | **M** | **J** | **J** | **A** | **S** | **O** | **N** | **D** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 129 | 124 | 126 | 131 | 123 | 125 | 122 | 127 | 126 | 125 | 131 | 134 |

a) Calculate the linear trend model. Does it appear to be a good fit based on the $r^2$-value?

b) Plot the original data and the expected values based on the model. Comment on the results.

**24** Three Finger Louis, the town's only butcher, is concerned about the volume of customers' bad debt he must write off as uncollectible each month. Dollar amounts in hundreds are shown here for the past three years.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1991** | | | | | | | | | | | |
| 14.1 | 13.7 | 12.1 | 13.1 | 13.5 | 9.1 | 7.2 | 6.1 | 8.7 | 10.1 | 11.8 | 12.2 |
| **1992** | | | | | | | | | | | |
| 15.2 | 14.1 | 13.2 | 13.9 | 14.0 | 9.5 | 7.2 | 6.5 | 9.1 | 11.5 | 12.2 | 13.4 |
| **1993** | | | | | | | | | | | |
| 13.7 | 12.5 | 11.8 | 12.0 | 13.0 | 8.7 | 6.3 | 6.0 | 8.2 | 9.8 | 10.9 | 11.8 |

a) Plot the data. Does a seasonal factor seem to exist? (Consider a "season" to be one month)

b) Use 12-month moving average to smooth out the seasonal variation.

c) Calculate seasonal indexes.

d) Deseasonalize the data.

e) Plot the original data and the decentralized data.

**25** Payroll taxes for ABL, Inc., have shown the progression recorded here in hundreds of dollars.

| J | F | M | A | M | J | J | A |
|---|---|---|---|---|---|---|---|
| 20.1 | 30.15 | 45.23 | 67.84 | 101.76 | 152.63 | 228.95 | 343.43 |

a) Plot the data.

b) Compute the trend model.

c) How well does the model explain taxes?

d) What are projected taxes in December?

**26** What is the instantaneous rate of growth in taxes in the previous problem?

**27** Packer Industries is concerned that sales may fall below $ 100,000 in December. Using these data in thousands of dollars, what is your projection? Plot the data first.

| Jan | Feb | Mar | Apr | May | June | July | Aug |
|---|---|---|---|---|---|---|---|
| 42.7 | 57.3 | 68.3 | 76.8 | 84 | 88.1 | 90 | 90.1 |

**28** *US News & World Report* stated that projections by the US Department of Commerce for median earnings of full-time workers were

| Year | Earnings ($ 1,000) |
|---|---|
| 1990 | 24.28 |
| 1995 | 30.26 |
| 2000 | 37.71 |
| 2005 | 47.00 |
| 2010 | 58.56 |
| 2015 | 73.00 |
| 2020 | 90.94 |
| 2025 | 113.33 |
| 2030 | 171.23 |
| 2035 | 176.00 |
| 2040 | 219.33 |

a) Plot the data.

b) Compute the trend model.

c) What is the projection for the year 2050?

**29** How well does the model in the previous problem explain earnings?

**30** Starting in July 1992, workers at the local plant unionized. Hourly wages prior to that time are shown here along with the trend in bonuses following unionization. Use trend analysis to determine if wages would be any different in December 1993 had the plant not unionized. Graph the data first.

| Time Period (1992) | Hourly Wages |
|---|---|
| January | 10.12 |
| February | 10.57 |
| March | 11.05 |
| April | 11.65 |
| May | 12.01 |
| June | 12.50 |
| July | 16.25 |
| August | 20.00 |
| September | 26.50 |
| October | 34.70 |
| November | 45.00 |
| December | 58.10 |

**31** Milles Products recorded the profits in the table.

a) Use exponential smoothing to forecast future profits. First, set $\alpha = 0.2$, then 0.9.

b) Which $\alpha$ -value produces a more reliable estimate?

c) How could you have known this beforehand?

| **Week** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| **Profits ($ 1,000)** | 10 | 25 | 3 | 15 | 2 | 27 | 5 |

**32** *The Journal of Real Estate Appraisal and Economics* reported the rates of return on national real estate investments trusts (REIT) for 1991 to be

| | | | | | |
|---|---|---|---|---|---|
| Jan | 4.2 | May | 3.9 | Sep | 4.1 |
| Feb | 3.7 | June | 3.9 | Oct | 4.0 |
| March | 4.5 | July | 4.0 | Nov | 3.9 |
| April | 4.4 | Aug | 3.8 | Dec | 4.0 |

a) Forecast January 1992 with a three-period moving average.

b) Forecast January 1992 using exponential smoothing. Set $\alpha$ at 0.8.

**33** *Fortune* magazine reported the vacancy rates for office space in Houston as shown here.

| | **1988** | **1989** | **1990** |
|---|---|---|---|
| Jan | 22.2% | 18.8% | 17.5% |
| Feb | 21.7 | 18.3 | 17.7 |
| Mar | 20.2 | 18.6 | 17.1 |
| Apr | 21.5 | 18.5 | 17.1 |
| May | 20.1 | 18.0 | 17.0 |
| June | 19.7 | 18.6 | 17.3 |
| July | 20.5 | 17.5 | 17.5 |
| Aug | 19.2 | 17.9 | 17.4 |
| Sep | 19.1 | 17.8 | 17.9 |
| Oct | 19.1 | 17.8 | 17.8 |
| Nov | 20.5 | 17.8 | 17.1 |
| Dec | 19.0 | 17.5 | 17.4 |

a) Harry uses a three-period moving average to forecast for April 1991. What answer does he get? Is it a good idea to use a moving average technique in this case? Explain.

b) Deseasonalize the data and calculate the values.

**34** Willie Wiener sells hot dogs in Central Park. His quarterly revenues are as shown here in hundreds of dollars.

| | 1989 | 1990 | 1991 |
|---|---|---|---|
| I | 47 | 59 | 61 |
| II | 52 | 64 | 64 |
| III | 57 | 63 | 69 |
| IV | 56 | 60 | 72 |

Develop the trend model for Willie. Forecast his revenue for March of 1992.

**35** Using the data from the previous problem, calculate the cyclical component.

### 7.4 Computer Exercises

**1.** Access the file TREND. It contains 18 months of data for unemployment rates in the Midwest. Develop the trend model and forecast the unemployment rate foe month 20.

**2.** Access the file SMOOTH. It contains data for monthly revenues of McNeese, Inc., a small manufacturing firm in Chicago. Using a four-period moving average, forecast revenues in the next month. Compare this result to the forecast you get using exponential smoothing, setting $\alpha$ at 0.2 and at 0.9.

### 7.5 MN-Case Applications

1. In 1990, Senator Daniel Moynihan proposed changes in the payroll tax. Shown here are the median earnings of full-time workers, along with the taxes paid under the existing system and those required by the Moynihan plan.

In considering the relative impact of the two tax systems, congressional committees must examine the level of taxes workers would be required to pay well into the future. Provide the necessary estimates of incomes and taxes, and discuss the relative merits of the two tax systems. Which tax system calls for a higher rate of growth in taxes paid? What is the rate at which income is growing?

Senator Moynihan's plan would lower payroll tax rates for the next 25 years but boost them substantially after that.

| Year | Median Income | Payroll Tax, 1990 | Moynihan Plan | Difference |
|---|---|---|---|---|
| 1990 | $ 24,280 | $ 1,587 | $ 1,823 | $ 34 |
| 1995 | 30,258 | 2,315 | 1,962 | 333 |
| 2000 | 37,708 | 2,885 | 2,470 | 415 |
| 2005 | 46,991 | 3,595 | 3,078 | 517 |
| 2010 | 58,560 | 4,480 | 3,836 | 644 |
| 2015 | 72,977 | 5,583 | 5,583 | 0 |
| 2020 | 90,943 | 6,957 | 7,594 | –637 |
| 2025 | 113,331 | 8,670 | 10,483 | –1,813 |
| 2030 | 141,231 | 10,804 | 13,064 | –2,260 |
| 2035 | 175,999 | 13,464 | 16,280 | –2,816 |
| 2040 | 219,327 | 16,779 | 20,280 | –3,509 |

**Note:** Earnings are estimates for an employee working at least 35 hours per week, 50 weeks a year, with an annual 4.5 percent increase. Under existing law, wages up to a mandated cutoff (or wage base) are now taxed at a rate of 7.65 percent. Estimates are in current, not constant, dollars. USN&WR-Basic data: US Dept. of Commerce. USN&WR estimates.

# 8 INDEX NUMBERS

## 8.1 Solved Problems

### Example 1 Price Indexes for Gasoline

Monthly prices for a gallon of gasoline are shown here. Using March as the base period, calculate the price indexes. What was the percentage increase from March to May and from May to June? What is the percentage point increase from May to June?

| Jan | Feb | Mar | Apr | May | June | July |
|---|---|---|---|---|---|---|
| 1.79 | 1.82 | 1.96 | 2.01 | 2.10 | 2.25 | 2.15 |

**Solution:**

| **Price Indexes (March = 100)** | | | |
|---|---|---|---|
| **Month** | **Index** | **Month** | **Index** |
| Jan | $\frac{1.79}{1.96}(100) = 91.3$ | May | $\frac{2.10}{1.96}(100) = 107.1$ |
| Feb | $\frac{1.82}{1.96}(100) = 92.9$ | June | $\frac{2.25}{1.96}(100) = 114.8$ |
| March | $\frac{1.96}{1.96}(100) = 100$ | July | $\frac{2.15}{1.96}(100) = 109.7$ |
| April | $\frac{2.01}{1.96}(100) = 102.6$ | | |

The percentage increase from March to May is

$$\frac{107.1 - 100.0}{100/0} = 7.1\ \%.$$

From May to June, the percentage increase is

$$\frac{114.8 - 107.1}{107.1} = 7.2\ \%.$$

And the percentage point increase is

$$114.8 - 107.1 = 7.7\ \%.$$

**Interpretation:** The base period will always report an index of 100. Periods in which the values are less than the base year will have an index less than 100, and periods with values in excess of that in the base year will have an index above 100.

If more than one time period is used as the base, the average of the values for those periods becomes the base. If Nipp and Tuck were to use the years 1993–1994 as the base period, the base value is ($ 3.00 + $ 3.30)/2 = 3.15. The index numbers for beef are then

$$1993 = (3 / 3.15) * (100) = 95.2;$$
$$1994 = (3.3 / 3.15) * (100) = 104.8;$$
$$1995 = (4.5 / 3.15) * (100) = 142.9.$$

**Example 2 Hair Today**

The Dippy Do Hair Salon is considering price adjustments in its services. Harriet Follicle, manager of Dippy Doo, wants to calculate Laspeyres and Paasche indexes, using these data for prices and the number of services rendered. January is taken as the base period.

| | **Price ($)** | | | **Quantity** | | |
|---|---|---|---|---|---|---|
| | **Jan** | **Feb** | **Mar** | **Jan** | **Feb** | **Mar** |
| Shampoo | 10 | 12.00 | 16.50 | 20 | 22 | 25 |
| Trim | 8 | 10.50 | 9.50 | 25 | 20 | 25 |
| Style | 12 | 13.50 | 14.00 | 30 | 31 | 33 |

**Solution**: The Laspeyres is based on the following table:

| | **Price** | | | | $P_R \times Q_{Jan}$ | | |
|---|---|---|---|---|---|---|---|
| | **Jan** | **Feb** | **Mar** | **Quantity in Jan** | **Jan** | **Feb** | **Mar** |
| Shampoo | $ 10 | $ 12.00 | $ 16.50 | 20 | 200 | 240 | 330 |
| Trim | 8 | 10.50 | 9.50 | 25 | 200 | 262.5 | 237.5 |
| Style | 12 | 13.50 | 14.00 | 30 | 360 | 405 | 420 |
| | | | | | 760 | 907.5 | 987.5 |

$$L_{Jan} = \frac{\sum(P_{Jan} \times Q_{Jan})}{\sum(P_{Jan} \times Q_{Jan})} \times 100 = \frac{760}{760}(100) = 100;$$

$$=$$

$$L_{Feb} = \frac{\sum(P_{Feb} \times Q_{Jan})}{\sum(P_{Jan} \times Q_{Jan})} \times 100 = \frac{907.5}{760}(100) = 119.4;$$

$$L_{Mar} = \frac{\sum(P_{Mar} \times Q_{Jan})}{\sum(P_{Jan} \times Q_{Jan})} \times 100 = \frac{987.5}{760}(100) = 129.4.$$

The Paasche index requires another set of calculations.

| | Jan | | Feb | | Mar | |
|---|---|---|---|---|---|---|
| | *P* | *Q* | *P* | *Q* | P | *Q* |
| Shampoo | 10 | 20 | 12.00 | *22* | 16.50 | 25 |
| Trim | 8 | *25* | 10.50 | 20 | 9.50 | 25 |
| Style | 12 | 30 | 13.50 | 31 | 14.00 | 33 |

| | | Price×Quantity | | |
|---|---|---|---|---|
| $P_{Jan}\,Q_{Jan}$ | $P_{Feb}Q_{Feb}$ | $P_{Mar}P_{Mar}$ | $P_{Jan}P_{Feb}$ | $P_{jan}Q_{Mar}$ |
| 200 | 264 | 412.5 | 220 | 250 |
| 200 | 210 | 237.5 | 160 | 200 |
| 360 | 418.5 | 462 | 372 | 396 |
| 760 | 892.5 | 1,112 | 752 | 846 |

$$P_{Jan} = \frac{\sum(P_{Jan} \times Q_{Jan})}{\sum(P_{Jan} \times Q_{Jan})} \times 100 = \frac{760}{760}(100) = 100;$$

$$P_{Feb} = \frac{\sum(P_{Feb} \times Q_{Feb})}{\sum(P_{Jan} \times Q_{Feb})} \times 100 = \frac{892.5}{752}(100) = 118.7;$$

$$P_{Mar} = \frac{\sum(P_{Mar} \times Q_{Mar})}{\sum(P_{Jan} \times Q_{Mar})} \times 100 = \frac{1{,}112}{846}(100) = 131.4.$$

**Interpretation:** The two indexes produce different results. They are based on different weighting systems. However, it is clear that an increase in prices by Dippy Doo is unwise. Prices have risen by 29.9 percent according to the Laspeyres index in only three months.

**Example 3.** The Laspeyres and Paasche Indexes Your firm manufactures three grades of lubricant. The prices and quantities sold for each are as follows:

| Grade | Prices ($) | | | Quantities | | |
|---|---|---|---|---|---|---|
| | **Oct** | **Nov** | **Dec** | **Oct** | **Nov** | **Dec** |
| A | 3.00 | 3.30 | 4.50 | 250 | 320 | 350 |
| B | 2.00 | 2.20 | 2.10 | 150 | 200 | 225 |
| C | 4.00 | 4.50 | 3.64 | 80 | 90 | 70 |

a. Calculate the Laspeyres index with October as the base.

| $P_{oct}Q_{oct}$ | $P_{Nov}Q_{Oct}$ | $P_{Dec}Q_{Oct}$ |
|---|---|---|
| 750 | 825 | 1,125.0 |
| 300 | 330 | 315.0 |
| 320 | 360 | 291.2 |
| 1,370 | 1,515 | 1,731.2 |

$$L = \frac{\sum(P_R \times Q_B)}{\sum(P_B \times Q_B)}(100).$$

October:

$$L_{Oct} = \frac{1{,}370}{1370}(100) = 100.$$

November:

$$L_{Nov} = \frac{1{,}515}{1370}(100) = 110.58.$$

December:

$$L_{Dec} = \frac{1{,}731.2}{1{,}370}(100) = 126.4.$$

***b.*** Calculate the Paasche index with October as the base.

| $P_{Oct} \times Q_{Oct}$ | $P_{Nov} \times Q_{Nov}$ | $P_{Dec} \times Q_{Dec}$ | $P_{Oct} \times Q_{Nov}$ | $P_{Oct} \times Q_{Dec}$ |
|---|---|---|---|---|
| 750 | 1,056 | 1,575.0 | 960 | 1,050 |
| 300 | 440 | 472.5 | 400 | 450 |
| 320 | 405 | 254.8 | 360 | 280 |
| 1,370 | 1,901 | 2,302.3 | 1,720 | 1,780 |

$$P = \frac{\sum(P_R \times Q_B)}{\sum(P_B \times Q_B)}(100).$$

October:

$$P_{Oct} = \frac{1,370}{1,370}(100) = 100.$$

November:

$$P_{Now} = \frac{1,901}{1,720}(100) = 110.52.$$

December:

$$P_{Dec} = \frac{2,302.3}{1,780}(100) = 129.34.$$

**Example 3. Statistical Application A.** Regarding a recent labor dispute over pay levels, an article in The *Wall Street Journal* illustrated how price indexes can be used to adjust wages. A local chapter of the United Mine Workers had agreed to a specific wage increase, but the miners were still concerned that inflation might erode their hard-earned wage gains. They refused to enter the mines until some assurance was provided that their incomes would be protected from rising prices. The miners finally agreed to a deal whereby their wages would increase by l percent for every 3 percent increase in the CPI. This was in addition to other wage increases labor had already gained. Statistical Application A shows how price indexes can be important to all who work for a living.

**Example 4. Statistical Application B.** The local newspaper in Huntington, West Virginia, carried a story about a water company that had requested a rate hike from the state Public Utilities Commission (PUC). The government denied the water company’s request for a rate hike, arguing that they had just raised rates to consumers the past year. The water company acknowledged the recent hike in rates, but argued that inflationary pressures made another one mandatory. The company used price indexes to show that the real cost of water had actually gone down when measured against overall inflationary rates. On this basis, the government reversed its earlier decision and granted the water company its request for another rate hike.

**8.2 Problems**

**1. Stock Price Indexes** As an economic analyst for your firm, you must examine the behavior of the price of your stock sold on the New York Stock Exchange. The average daily prices for each month are shown for the common stock and the preferred stock of your firm.

| Month | Common Stock Price | Preferred Stock Price |
|---|---|---|
| January | 5.12 | 10.15 |
| February | 6.14 | 11.12 |
| March | 7.52 | 12.11 |
| April | 7.22 | 15.01 |
| May | 8.15 | 14.12 |
| June | 8.90 | 15.17 |
| July | 8.70 | 15.90 |

a. Develop a simple price index for each stock, with March as the base period.

b. Develop a composite index for both stocks (March = 100).

c. Develop a simple price index for the common stock, with March – May as the base. The base value is

$$(7.52 + 7.22 + 8.15) / 3 = 7.63.$$

2. **A Gross Report by the President** According to the 1991 *Economic Report of the President***,** gross national product in billions of current dollars is as shown here. Use the CPI-U to obtain real (or constant) GNP in 1982 – 84 dollars.

| Year | Money GNP | CPI (82 – 84 = 100) |
|---|---|---|
| 1985 | 4,010.3 | 107.6 |
| 1986 | 4,231.6 | 109.6 |
| 1987 | 4,524.3 | 113.6 |
| 1988 | 4,880.6 | 1183 |
| 1989 | 5,348 | 121.3 |

## 8.2. Conceptual Questions

**1.** You have just been hired as a marketing specialist for the Jensu Japanese Knife Company. Your boss wants you to explain how index numbers can be used to aid in marketing decisions. How do you respond? What is an index number? What is the difference between a simple index, a composite index, and a weighted composite index?

**2.** While developing the index numbers in the previous problem, you are asked to explain the difference between money (current) income and real (constant) income.

**3.** You have just been promoted to chief accountant for the Doit and Quick Accounting firm in Crab Apple Cove, Maine. Your new job duties require that you deflate a time series. What does it mean to deflate a time series, and what is the purpose in doing so?

**4.** The local chamber of commerce wants you to explain the relationship between income and changes in prices. Specifically, you must explain the following: If prices were falling, what would happen to real income if money income remained constant, went up, or went down?

**5.** The chamber of commerce also wants you to answer the following question: If prices increase more rapidly than money income, what happens to real income? How would you respond?

**6.** The statistical analyst for your firm is on maternity leave. The manager of the finance division wants you to develop an index for the firm's sales over the past several years. Under what conditions might you use a Paasche index? a Laspeyres index? What is the difference between them? What are their advantages?

**7.** Your boss wants to develop a weighted price index, but he isn't certain which type to use.

a. If he wants to ensure a more meaningful comparison over time, which one would you advise?

b. Why did you answer part (a) the way you did?

c. If he wants to reflect current consumption patterns, which index would you advise?

d. Why did you answer part (c) the way you did?

## 8.3 Problems

**1**. Consumer data supplied by the U.S. Department of Commerce for the summer of 1996 revealed the following:

| | Unit | Prices | | |
|---|---|---|---|---|
| | | 1994 | 1995 | 1996 |
| Beef | 1 pound | $ 3.12 | $ 3.89 | $ 3.92 |
| Milk | 1 gallon | 2.10 | 2.42 | 2.51 |
| Chicken | 1 pound | 1.95 | 2.10 | 2.12 |
| Bread | 1 loaf | 0.99 | 0.89 | 1.12 |

Compute and interpret the simple index for each commodity using 1994 as the base period

**2.** Given the data from the previous problem, compute

a. The percentage increase in the price of each product for

1994 to 1995.

1994 to 1996.

1995 to 1996.

b. The percentage point increase for each product for

1994 to 1995.

1994 to 1996.

1995 to 1996.

**3.** Below are costs for a one-day stay in the hospital. Use 1993 as the base year and compute the simple index. Interpret the index obtained for 1990.

| 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|---|---|---|---|---|---|
| $ 356 | $ 408 | $ 512 | $ 589 | $ 656 | $ 689 |

4. Sammy Studd wants to purchase an entire new wardrobe of athletic wear for the summer. He has collected the data seen here and wonders how prices have changed over the past three years. Compute and interpret the composite price index for all four goods using 1994 as the base year.

| | 1994 | 1995 | 1996 |
|---|---|---|---|
| Shoes | $ 89.90 | $ 115.12 | $ 125.00 |
| Sweats | 52.50 | 65.50 | 75.50 |
| Shorts | 25.75 | 35.95 | 45.90 |
| Socks | 12.10 | 10.00 | 9.50 |

**5.** Prices for a new line of toy dolls by The Krazy Kid Kollection are shown here. Using 1994 – 1995 as the base period, calculate a simple price index for all three toys.

| | 1994 | 1995 | 1996 |
|---|---|---|---|
| Killer Joe | $ 17.90 | $ 21.50 | $ 25.00 |
| Pyro Phil | 15.00 | 25.00 | 29.95 |
| Maniac Mark | 10.00 | 11.00 | 1200 |

**6.** Bell electronic wishes to analyze price changes for three of its products over the past three years. The necessary data are given here.

| Product | | Prices | | | Quantities | |
|---|---|---|---|---|---|---|
| | 1993 | 1994 | 1995 | 1993 | 1994 | 1995 |
| A | $ 10.00 | $ 15.50 | $ 20.00 | 150 | 170 | 160 |
| B | 3.00 | 5.00 | 7.50 | 55 | 68 | 120 |
| C | 69.00 | 75.00 | 75.00 | 100 | 90 | 85 |

Compute and interpret the Laspeyres and I'aasche indexes using 1993 as the base period.

**7.** Using the data from the previous problem, compute the Fisher index.

**8.** Mr. Mom has kept this record of his food costs over the past five years.

| | 1987 | 1988 | 1989 | 1990 | 1991 |
|---|---|---|---|---|---|
| Kumquats | 40 | 52 | 47 | 0 | 10 |
| Bean sprouts | 20 | 32 | 52 | 50 | 45 |
| Guava | 17 | 18 | 23 | 17 | 20 |

- Develop a simple price index for kumquats, with 1988 as the base.
- Determine the composite index with 1988 as the base.
- Calculate the average of relative indexes (1988 = 100).

**9.** Assume Mr. Mom bought the quantities shown:

| | 1987 | 1988 | 1989 | 1990 | 1991 |
|---|---|---|---|---|---|
| Kumquats | 40 | 52 | 47 | 0 | 10 |
| Bean sprouts | 20 | 32 | 52 | 50 | 45 |

| | | | | | |
|---|---|---|---|---|---|
| Guava | 17 | 18 | 23 | 17 | 20 |

a. Calculate the weighted composite index which uses fixed-base weights (1988 =100).

b. Calculate the weighted composite index using weights from the reference period.

**10.** Based on the Paasche index from the previous problem, what was the percentage increase in prices from

a. 1988 to 1989?

b. 1988 to 1990?

c. 1989 to 1990?

**11**. In the table are indexes for the CPI for selected years with different bases.

a. Splice them.

b. Rebase the spliced series to (1985 = 100).

| **Year** | **CPI (1967 = 100)** | **CPI (1982–84 = 100)** |
|---|---|---|
| 1980 | 246.8 | |
| 1981 | 272.4 | |
| 1982 | 289.1 | |
| 1983 | 298.4 | |
| 1984 | 311.1 | 103.9 |
| 1985 | | 107 6 |
| 1986 | | 109.6 |
| 1987 | | 113.6 |

**12.** Here are values for GNP in current dollars taken from the *Economic Report of the President*. Use the data in Table 18–12 to find GNP in constant (real) dollars. Interpret the deflated series. (Data are in billions of dollars.)

| **Year** | **GNP** | **Year** | **GNP** | **Year** | **GNP** |
|---|---|---|---|---|---|
| 1975 | 1,598.4 | 1980 | 2,732.0 | 1985 | 4,014 9 |
| 1976 | 1,782.8 | 1981 | 3,052.6 | 1986 | 4,231.6 |
| 1977 | 1,990.5 | 1982 | 3,166.0 | 1987 | 4,524.3 |
| 1978 | 2,249.7 | 1983 | 3,405.7 | 1988 | 4,880 6 |

| | | | | | |
|---|---|---|---|---|---|
| 1979 | 2,508.2 | 1984 | 3,772.2 | 1989 | 5,348.2 |

**13.** Jamie's money income over the past several years is shown here. His income has risen over $ 4,000, and he feels good about that. Should he?

| Year | Money Income | Year | Money Income |
|---|---|---|---|
| 1984 | $ 42,111 | 1987 | $ 44,670 |
| 1985 | 42,953 | 1988 | 45,573 |
| 1986 | 43 815 | 1989 | 46,540 |

**14.** According to the *Statistical Abstract of the United States*, monthly salaries for new accounting graduates were as follows:

| | Cost per Usage | | | Usage Frequency | | |
|---|---|---|---|---|---|---|
| | 1989 | 1990 | 1991 | 1989 | 1990 | 1991 |
| Studio costs | $ 120 | $ 145 | $ 165 | 30 | 35 | 37 |
| Recording equipment | 420 | 530 | 620 | 40 | 43 | 46 |
| Backup singers | 300 | 250 | 200 | 50 | 63 | 72 |

Sam O'Donnell, the director of statistical procedures, must calculate a Laspeyres index and a Paasche index, using 1989 as the base, and then determine the rate at which costs have risen each year under both indexes, as well as the rate of inflation over all three years.

**15.** From Problem 30, which index is probably a better measure of Columbia's rise in costs? Why? Support your choice.

**16.** Indexed living costs shown here were taken from the *Monthly Labor Review*. What are the average of relatives indexes for all three years? (1982 = 100)

| Region | 1989 | 1990 | 1991 |
|---|---|---|---|
| East | 126.1 | 132.8 | 139.7 |
| South | 117.3 | 121.2 | 124.6 |
| Midwest | 112.7 | 119.3 | 121.2 |
| West | 119.2 | 123.4 | 125.2 |

**17.** From the data in Problem 32, which region has the lowest average of relatives index over the three-year period?

**18.** In 1989, Bob's starting salary was $ 28,800. In 1984, Bob's sister, Martha, earned a starting salary of $ 25,000. Bob belittled his sister's low starting salary. What might Martha's response be?

**19.** The federal government's 1990 fiscal expenditure on national defense was $ 296.3 billion. President Bush budgeted $ 303.3 billion for 1991, and thereby caused a protest from those who objected to the increase. Based on the 1990 CPI of 127.2 taken from Table 18–12 and a 1991 projected CPI of 131.2, the president's advisors argued that defense spending decreased. How is this so?

**20.** Red Wing Enterprises markets sweaters of wool, cotton, and cashmere. Prices and volumes are shown here. Using 1989 as the base year, develop Laspeyres and Paasche indexes.

| | **Prices** | | | **Quantities** | | |
|---|---|---|---|---|---|---|
| | **Wool** | **Cotton** | **Cashmere** | **Wool** | **Cotton** | **Cashmere** |
| 1989 | 45 | 37 | 85 | 14 | 18 | 12 |
| 1990 | 52 | 41 | 105 | 18 | 22 | 10 |
| 1991 | 54 | 39 | 90 | 22 | 25 | 18 |
| 1992 | 56 | 44 | 110 | 25 | 20 | 19 |

**21.** Just Pizza bought the amounts of ingredients at the prices shown in the table. Janet Jackson, manager of Just Pizza, is worried about rising prices. Develop Laspeyres and Paasche indexes for her, using January as the base.

| | | **Price/Pound** | | | **Pounds Used (100's)** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Jan** | **Feb** | **Mar** | **Apr** | **Jan** | **Feb** | **Mar** | **Apr** |
| Cheese | 2.10 | 2.15 | 2.20 | 2.25 | 10 | 12 | 15 | 12 |
| Pepperoni | 1.18 | 1.20 | 1.25 | 1.31 | 8 | 10 | 8 | 10 |
| Sausage | 1.25 | 1.31 | 1.35 | 1.42 | 7 | 6 | 7 | 7 |

**22.** Using Janet's data from the previous problem, does the Paasche index show that the rate at which prices are going up is increasing or decreasing?

**23.** In a recent issue of *Business Week*, Manuel Johnson, then vice-chairman of the Federal Reserve System, commented on the advisability of maintaining the stock of monetary reserves at a constant real level. The money stock (MI) in 1987 was $ 742 billion. What level is needed in 1990 to achieve the same real stock of money reserves?

**8.4 Mini-Case Applications**

1. Caterpillar, Inc., maintains worldwide sales forces in New York and its headquarters in Peoria, Illinois. Recent personnel changes, coupled with drastic conditions in the agricultural economy in the late 1980s, created significant problems for Caterpillar.

Competition developed between the New York and Peoria offices to hold down labor costs. It was decided that each office should receive the same money per employee in real terms based on local price indexes.

In 1989, company records showed that Peoria was to receive, in thousands of dollars, prior to the necessary cutback, 42.7, 42.0, 38.9, and 45.7 during the years 1989 – 1992. The price indexes for central Illinois during those years were 104.1, 107.2, 107.9 and 110.3. Indexes for New York were 127.1; 131.4; 137.2 and 141.2.

How much money must Caterpillar allocate for New York each year to ensure both locations got equal labor expenditures measured in real terms?

## LIST OF LITERATURE


1. Попов І. І. Теорія статистики. Практикум : навч. посіб. / І. І. Попов. – К. : КНТЕУ, 2006. –290 с.

2. Практикум по теории статистики: учеб. пособие / под ред. Р. А. Шмойловой. – М.: Финансы и статистика, 2002. –416с.

3. Сигел Э. Практическая бизнес-статистика : пер. с анг / Э. Сигел. – М. : Вильямс, 2002. –1021 с.

4. Webster Allen L. Applied Statistics for Business and Economics. Second Edition / Allen L. Webster. – The USA: Richard D.Irwin, Inc., 1995. – 1068 p.

5. Feller W. An Introduction to probability Theory and its Applications. Volume 1. Third Edition / W. Feller. – The USA: John Wiley&Sons, Inc. 1968. – 527 p.

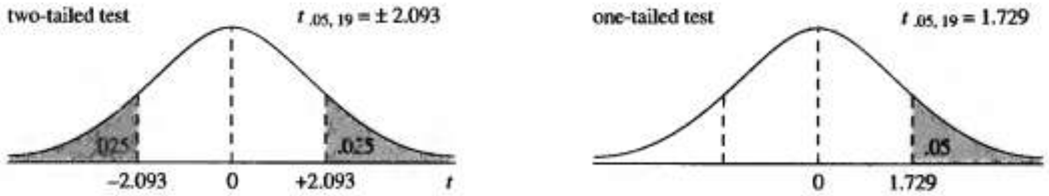
6. Fernandes M. Statistics for Business and Economics / M. Fernandes. Marcelo Fernandes & Ventus Publishing ApS. 2009. – 150 p.

7. Jaisingh Lloyd R. Statistics for Utterly Confused / Lloyd R. Jaisingh. McGraw–Hill. 2000. – 337 p.

## APPENDIX 1

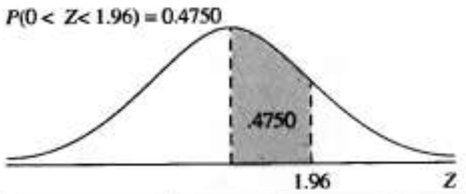
## DISTRIBUTION TABLES

Table A.1 – The $t$ Distribution




| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.900 | 0.700 | 0.500 | 0.300 | 0.200 | 0.100 | 0.050 | 0.020 | 0.010 | α value | two tailed-test |
| | 0.100 | 0.300 | 0.500 | 0.700 | 0.800 | 0.900 | 0.950 | 0.980 | 0.990 | CL | |
| | 0.450 | 0.350 | 0.250 | 0.150 | 0.100 | 0.050 | 0.025 | 0.010 | 0.005 | α value | one tailed-test |
| | 0.550 | 0.650 | 0.750 | 0.850 | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 | CL | |
| d.f. | Values of $t$ | | | | | | | | | | |
| 1 | 0.158 | 0.510 | 1.000 | 1.963 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | | |
| 2 | 0.142 | 0.445 | 0.816 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | | |
| 3 | 0.137 | 0.424 | 0.765 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | | |
| 4 | 0.134 | 0.414 | 0.741 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | | |
| 5 | 0.132 | 0.408 | 0.727 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | | |
| 6 | 0.131 | 0.404 | 0.718 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | | |
| 7 | 0.130 | 0.402 | 0.711 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | | |
| 8 | 0.130 | 0.399 | 0.706 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | | |
| 9 | 0.129 | 0.398 | 0.703 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | | |
| 10 | 0.129 | 0.397 | 0.700 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | | |
| 11 | 0.129 | 0.396 | 0.697 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | | |
| 12 | 0.128 | 0.395 | 0.695 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | | |
| 13 | 0.128 | 0.394 | 0.694 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | | |
| 14 | 0.128 | 0.393 | 0.692 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | | |
| 15 | 0.128 | 0.393 | 0.691 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | | |
| 16 | 0.128 | 0.392 | 0.690 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | | |
| 17 | 0.128 | 0.392 | 0.689 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | | |
| 18 | 0.127 | 0.392 | 0.688 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | | |
| 19 | 0.127 | 0.391 | 0.688 | 1.066 | 1.328 | **1.729** | **2.093** | 2.539 | 2.861 | | |
| 20 | 0.127 | 0.391 | 0.687 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | | |
| 21 | 0.127 | 0.391 | 0.686 | 1.063 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | | |
| 22 | 0.127 | 0.390 | 0.686 | 1.061 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | | |
| 23 | 0.127 | 0.390 | 0.685 | 1.060 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | | |
| 24 | 0.127 | 0.390 | 0.685 | 1.059 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | | |
| 25 | 0.127 | 0.390 | 0.684 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | | |
| 26 | 0.127 | 0.390 | 0.684 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | | |
| 27 | 0.127 | 0.389 | 0.684 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | | |
| 28 | 0.127 | 0.389 | 0.683 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | | |
| 29 | 0.127 | 0.389 | 0.683 | 1.055 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | | |
| 30 | 0.127 | 0.389 | 0.683 | 1.055 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | | |
| 40 | 0.126 | 0.388 | 0.681 | 1.050 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | | |
| 60 | 0.126 | 0.387 | 0.679 | 1.045 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | | |
| 120 | 0.126 | 0.386 | 0.677 | 1.041 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | | |
| ∞ | 0.126 | 0.385 | 0.674 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | | |

Table A.2 – Z-test (The Standard Normal Distribution)




| Z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | **0.4750** | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |
| 2.5 | 0.4938 | 0.4940 | 0.4941 | 0.4943 | 0.4945 | 0.4946 | 0.4948 | 0.4949 | 0.4951 | 0.4952 |
| 2.6 | 0.4953 | 0.4955 | 0.4956 | 0.4957 | 0.4959 | 0.4960 | 0.4961 | 0.4962 | 0.4963 | 0.4964 |
| 2.7 | 0.4965 | 0.4966 | 0.4967 | 0.4968 | 0.4969 | 0.4970 | 0.4971 | 0.4972 | 0.4973 | 0.4974 |
| 2.8 | 0.4974 | 0.4975 | 0.4976 | 0.4977 | 0.4977 | 0.4978 | 0.4979 | 0.4979 | 0.4980 | 0.4981 |
| 2.9 | 0.4981 | 0.4982 | 0.4982 | 0.4983 | 0.4984 | 0.4984 | 0.4985 | 0.4985 | 0.4986 | 0.4986 |
| 3.0 | 0.4987 | 0.4987 | 0.4987 | 0.4988 | 0.4988 | 0.4989 | 0.4989 | 0.4989 | 0.4990 | 0.4990 |
| 3.1 | 0.4990 | 0.4991 | 0.4991 | 0.4991 | 0.4992 | 0.4992 | 0.4992 | 0.4992 | 0.4993 | 0.4993 |
| 3.2 | 0.4993 | 0.4993 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4995 | 0.4995 | 0.4995 |
| 3.3 | 0.4995 | 0.4995 | 0.4995 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4997 |
| 3.4 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4998 |
| 3.5 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 |
| 3.6 | 0.4998 | 0.4998 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.7 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.8 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.9 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |

Навчально-методичне видання

**СТАТИСТИКА: ПРАКТИКУМ**

Методичні вказівки до практичних занять з курсу «Статистика»
для студентів спеціальностей
073 «Менеджмент» та 072 «Фінанси, банківська справа та страхування»

Укладач: ШИРЯЄВА Наталя Володимирівна

Роботу до видання рекомендував проф. В. А. Міщенко
В авторській редакції

План 2018 р. поз. 336
Підписано до друку 24.05.18.
Гарнітура  Таймс. Ум. друк. арк. 5,38.

Видавничий центр НТУ "ХПІ".
Свідоцтво про державну реєстрацію ДК № 9458 від 21.08.2017 р.
61002, Харків 2, вул. Кірпічова, 2

Самостійне електронне видання