Student Solutions Manual to Accompany

INTRODUCTORY STATISTICS

Third Edition

Sheldon M. Ross

Prepared by
Lloyd R. Jaisingh
Mathematics and Computer Science Department
Morehead State University

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Student Solutions Manual to Accompany

INTRODUCTORY STATISTICS

Third Edition

Sheldon M. Ross

Prepared by

Lloyd R. Jaisingh

Mathematics and Computer Science Department

Morehead State University

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
Elsevier, The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

ISBN: 978-0-12-384668-6

For information on all Academic Press publications
visit our Web site at *www.elsevierdirect.com*

Table of Contents

This page intentionally left blank

Chapter 1 INTRODUCTION TO STATISTICS

Review Problems

2. From the graph, as it relates to the consumption of buttermilk, there was a gradual decrease until about 1958 when the consumption started an upward trend. The consumption peaked around 1990 and then remained approximately constant.

With regards to the consumption of whole milk, there was a gradual increase until around 1941 when there was a sharp increase until around 1944 when the consumption began to decrease.

In addition, around 1986, the consumption of buttermilk surpassed the consumption of whole milk.

4. If the researcher scrutinizes the volunteers and then decides on the groupings, he/she may allow his/her own bias to play a role. For example, he/she may accidentally put the "healthy" patients in a certain group. As such, this approach is not advisable.

6. Answer (c) is correct, assuming that the voter registration list matches the voting population. The other four choices do not reflect a random representation of the voting population.

8. No. This will not lead to a representative sample since most of the obituaries will likely be from well-known or affluent people from New York City and its vicinity.

10. The shopping mall. Shopping malls are usually frequented by a broader cross section of the population; hence, the sample obtained from the shopping mall will be more representative.

12. (a) No. The article fails to consider the proportion of pedestrians not killed who normally wear light-colored or dark-colored clothing. Confounding variables may exist, such as a person's age, gender, and safety conscience.

(b) The data regarding these confounding variables need to be known, for both pedestrians killed and pedestrians not killed.

14. Since 2% of London's population died in 1658, this implies that 2 out of every 100 died or 1 out of every 50 died. Thus, according to Graunt's method, the estimated population of London for January 1, 1658 is

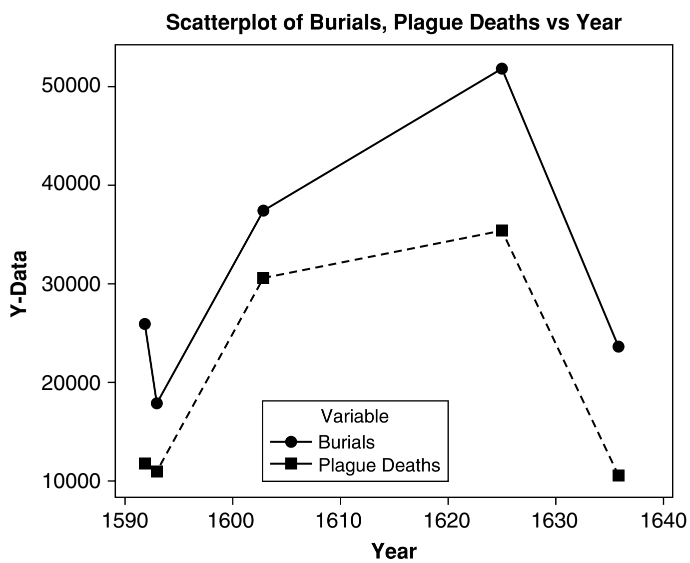
$$12{,}246 \times 50/1 = 612{,}300 \text{ people.}$$

16. The table below shows data for the total deaths in England for five different years. Also indicated is the number of deaths due to the plague and the proportion of deaths by the plague.

Year	*Burials*	*Plague Deaths*	*% Plague Deaths*
1592	25,886	11,503	44.44
1593	17,844	10,662	59.75
1603	37,294	30,561	81.95
1625	51,758	35,417	68.43
1636	23,359	10,400	44.52

If "severe" is defined in terms of the proportion of deaths, then the year 1603 was the most severe in terms of death by the plague.

If "severe" is defined in terms of the number of deaths, then the year 1625 was the most severe in terms of death by the plague. See the following graph.



18. Answer will vary from student to student.

Chapter 2 DESCRIBING DATA SETS

2.2 FREQUENCY TABLES AND GRAPHS

Problems

2. (a) Number of weeks in which there were ***at least*** 2 bikes sold

$$= 7 + 10 + 8 + 5 + 2 + 1 = 33.$$

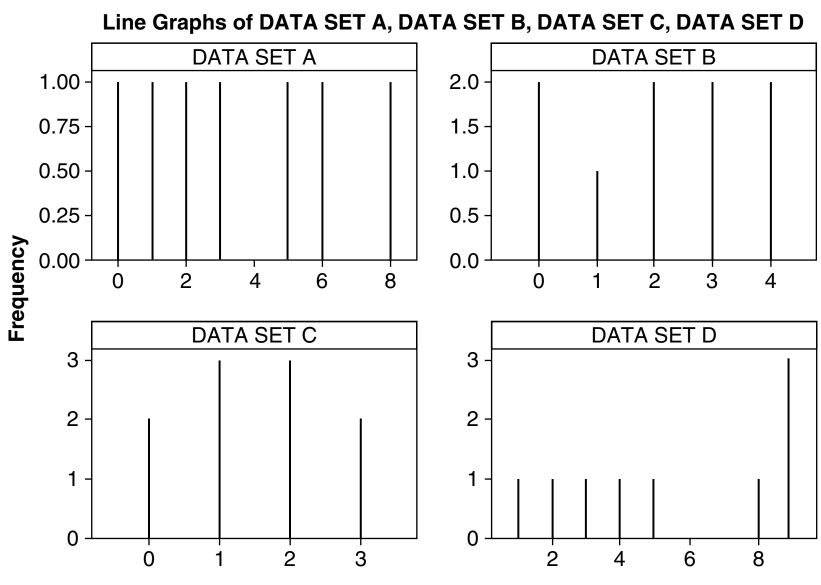
(a) Number of weeks in which there were ***at least*** 5 bikes sold

$$= 5 + 2 + 1 = 8.$$

(c) Number of weeks in which an ***even number*** of bikes were sold

$$= 3 + 7 + 8 + 2 = 20.$$

4. One way to determine whether the data sets are symmetrical, approximately symmetrical, or not at all symmetrical is to construct a projection graph for each data set and analyze.

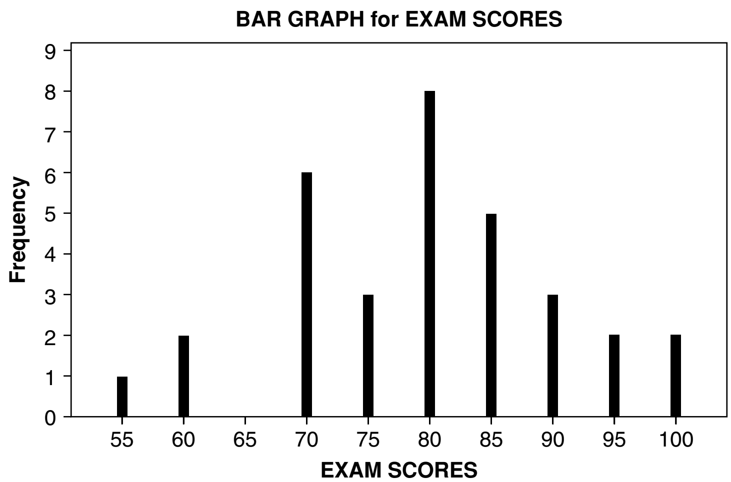


From these projection graphs, and based on the definition of symmetry in the text, it appears that only data sets B and C can be considered symmetrical. Note that data set A might be considered approximately symmetrical, while data set D is not at all symmetrical.

6. Below shows the frequency table for the *statistics scores* for the 32 students. In the frequency table, the frequency is denoted by f.

Score	55	60	70	75	80	85	90	95	100
f	1	2	6	3	8	5	3	2	2

Below is the bar graph of the ***frequencies*** versus ***exam scores*** for the data set in part (a).



8. (a) The relative frequency table for the ***time (months) to tumor progression*** data is shown below. The relative frequency is denoted by f/n, expressed as a percentage (%).

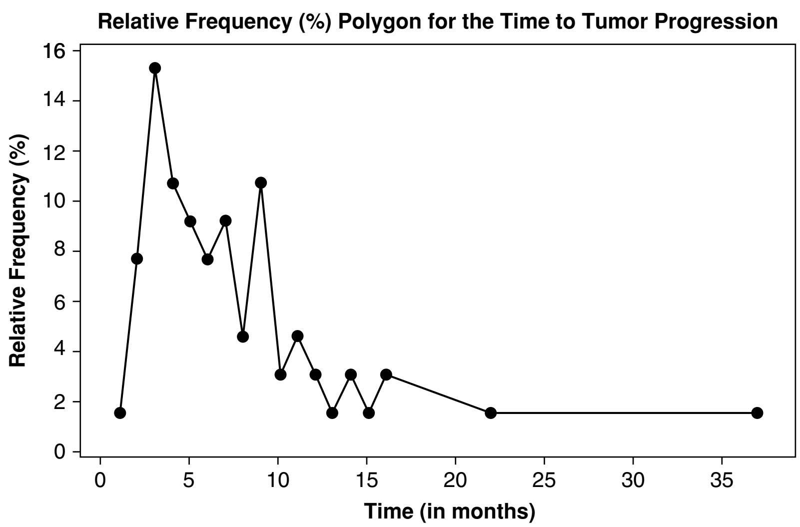
Time to Tumor Progression, months	***Frequency, f***	***Relative Frequency, f/n (%)***
1	1	1.54
2	5	7.69
3	10	15.38
4	7	10.77
5	6	9.23
6	5	7.69
7	6	9.23
8	3	4.62
9	7	10.77
10	2	3.08
11	3	4.62
12	2	3.08

(*Continued*)

(*Continued*)

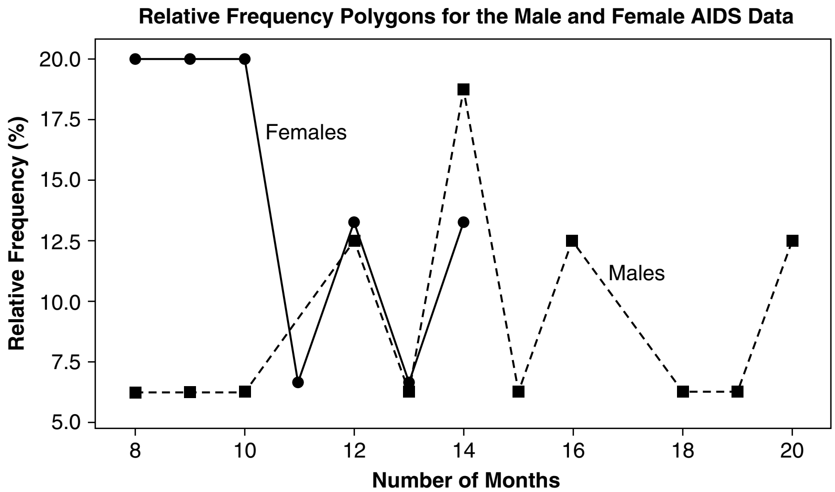
Time to Tumor Progression, months	*Frequency, f*	*Relative Frequency, f/n (%)*
13	1	1.54
14	2	3.08
15	1	1.54
16	2	3.08
22	1	1.54
37	1	1.54

(b) Below is the relative frequency polygon graph for the ***time to tumor progression*** data. The relative frequency is expressed as a percentage (%).



(c) The data set is not symmetrical. Because of the apparent outlier point of 37 months, the less extreme outlier of 22, and the peak at 3 months, the data set may be classified as skewed to the right.

10. The relative frequency polygons for the ***number of months from diagnosis to death for the sample of male and female AIDS patients*** are shown below.

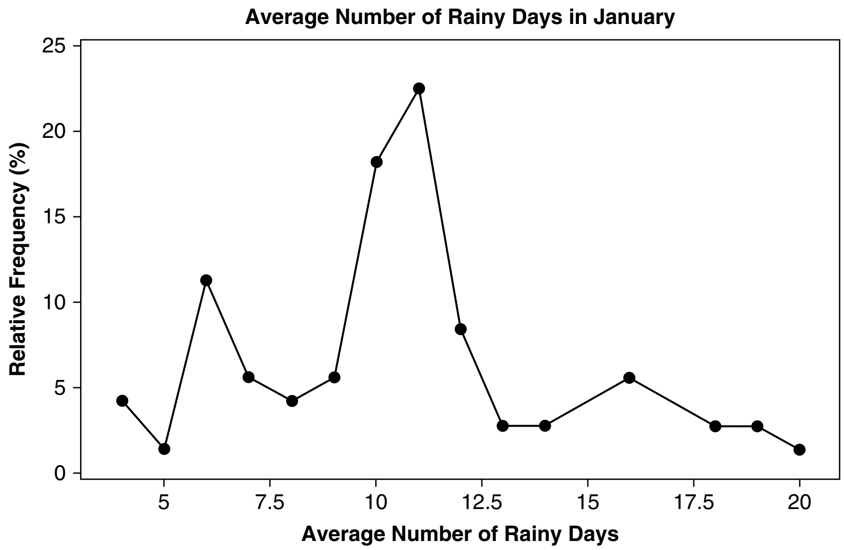


From the polygons, one can see that the male AIDS patients tend to live longer than the female AIDS patients from the time of diagnosis until death.

12. The table below shows the relative frequency table for the data relating to the ***average number of rainy days in January for the different cities***. The relative frequency is expressed as a percentage (%).

Average Number of Rainy Days (January)	*Frequency, f*	*Relative Frequency, f/n (%)*
4	3	4.23
5	1	1.41
6	8	11.27
7	4	5.63
8	3	4.23
9	4	5.63
10	13	18.31
11	16	22.54
12	6	8.45
13	2	2.82
14	2	2.82
16	4	5.63
18	2	2.82
19	2	2.82
20	1	1.41

The relative frequency polygon for the ***average number of rainy days in January for the different cities*** is shown below.



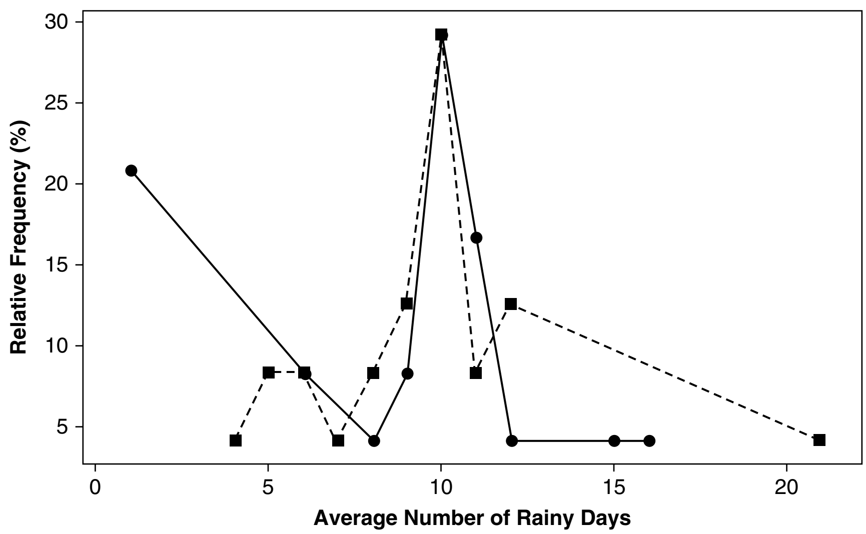
14. The tables below show the relative frequency tables for the data values relating to the ***average number of rainy days in June and December for the first 24 cities***. The relative frequency is expressed as a percentage (%).

Avg. No. of Rainy Days	1	6	8	9	10	11	12	15	16
Frequency, f	5	2	1	2	7	4	1	1	1
f/n (%) (June)	20.83	8.33	4.17	8.33	29.17	16.67	4.17	4.17	4.17

Avg. No. of Rainy Days	4	5	6	7	8	9	10	11	12	21
Frequency, f	1	2	2	1	2	3	7	2	3	1
f/n (%) (December)	4.17	8.33	8.33	4.17	8.33	12.50	29.17	8.33	12.50	4.17

The graph below displays the relative frequency polygons for the data values relating to the ***average number of rainy days in June and December for the first 24 cities***. The relative frequency is expressed as a percentage (%).

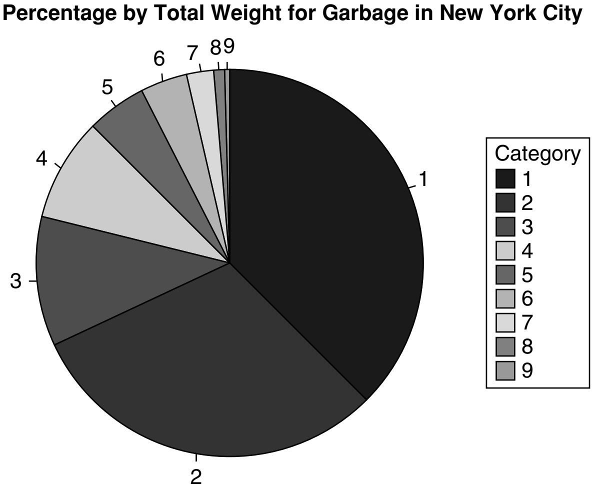
Relative Frequency Polygons for June and December (First 24 Values)



16. Below is the pie chart for the various percentages (by total weight) for the various ***types of garbage from New York City***. This information was taken from the NY Times.

***Note*:** In the chart, category 1 ≡ *organic material (37.3%)*, 2 ≡ *paper (30.8%)*, 3 ≡ *bulk (10.9%)*, 4 ≡ *plastic (8.5%)*, 5 ≡ *glass (5%)*, 6 ≡ *metal (4%)*, 7 ≡ *inorganic (2.2%)*, 8 ≡ *aluminum (0.9%)*, and 9 ≡ *hazardous waste (0.4%)*.

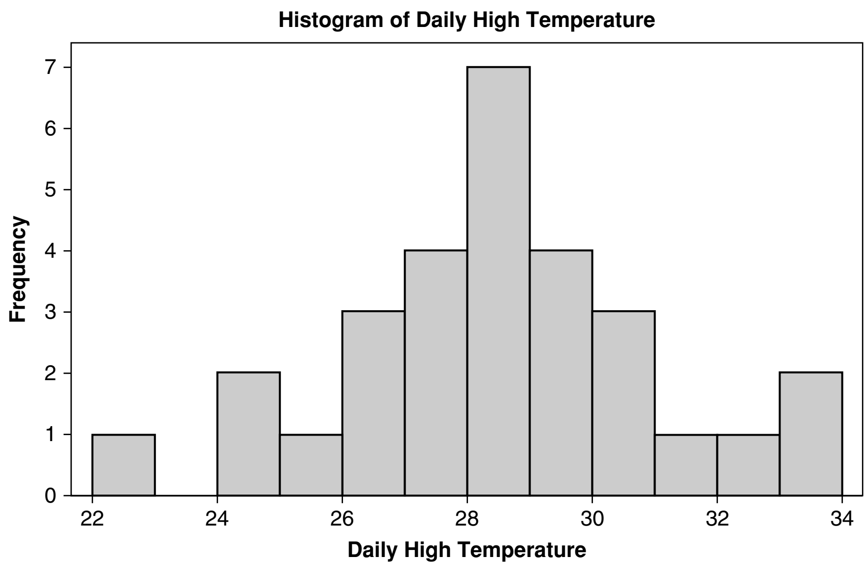
Pie Chart of Percentage by Total Weight for Garbage in New York City



2.3 GROUPED DATA AND HISTOGRAMS

Problems

2. (a) The frequency histogram for the ***daily high temperature (in Celsius) for July 4 in San Francisco*** over a sequence of 30 years is shown below.



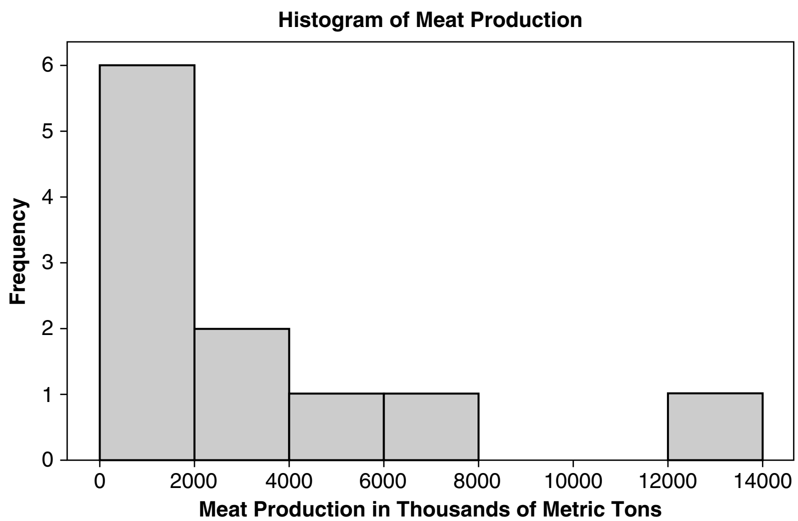
(b) A "typical" July 4 temperature in San Francisco would lie between 26 and 31 degrees Celsius.

(c) The data set is approximately symmetrical. Also, the modal temperature is around 28 or 29 degrees Celsius and is near the middle of the data set.

4. Based on the information, we can have the table starting at class interval 6–9 and ending at class interval 27–30, or you can start at class interval 3–6 and end at class interval 24–27. The solution presented here starts at class interval 6–9 and ends at class interval 27–30. Below is the complete table. The *numbers* that were filled in are typed in *bold*.

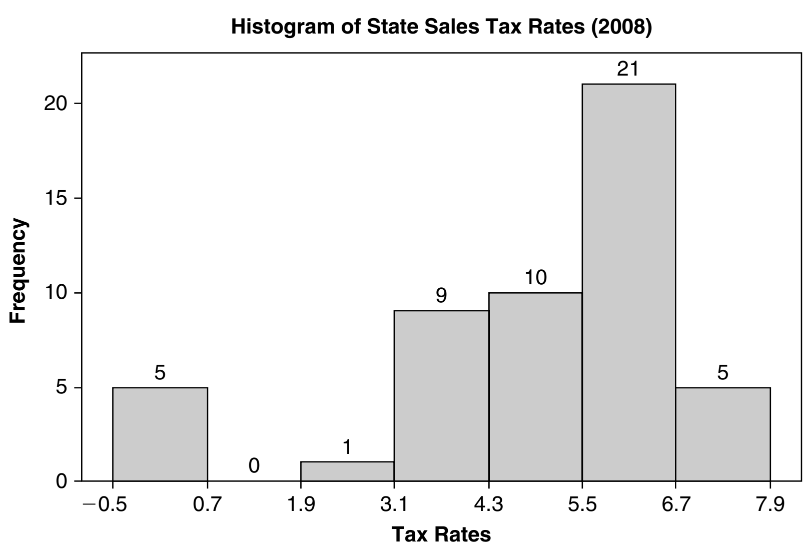
Class Intervals	*Frequency*	*Relative Frequency*
6–9	**10**	0.05
9–12	14	**0.07**
12–15	18	**0.09**
15–18	38	**0.19**
18–21	**20**	0.10
21–24	42	**0.21**
24–27	11	**0.055**
27–30	**47**	**0.235**

6. (a) The frequency histogram for ***2002 meat production*** is shown below.

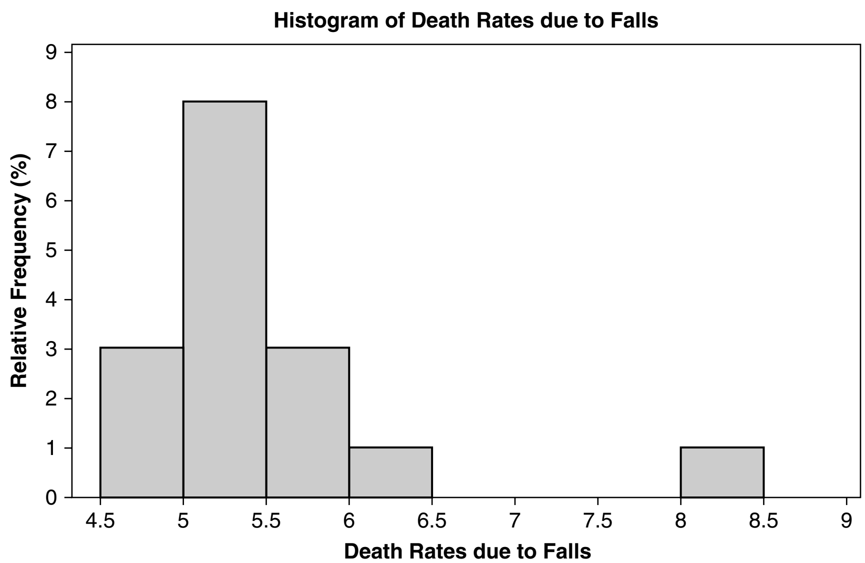


(b) There is an outlier in the data set. This outlying point is the value for the meat production in the United States, which was 12,424 thousands of metric tons. This outlying point is shown by the box to the far right in the frequency histogram in part (a).

8. The frequency histogram of the 2008 state sales tax rates of the 50 states and the District of Columbia along with the frequency counts in each class (interval). Since no specification is given as to the number of classes to use, the histogram is presented with 7 classes. Observe that the left most bar corresponds to the 5 states with zero sales tax rates.



10. The relative frequency histogram for the ***death rates due to falls*** is shown below. This graph was constructed with 8 class intervals with a width of 0.5, and the relative frequency was expressed as a percentage.



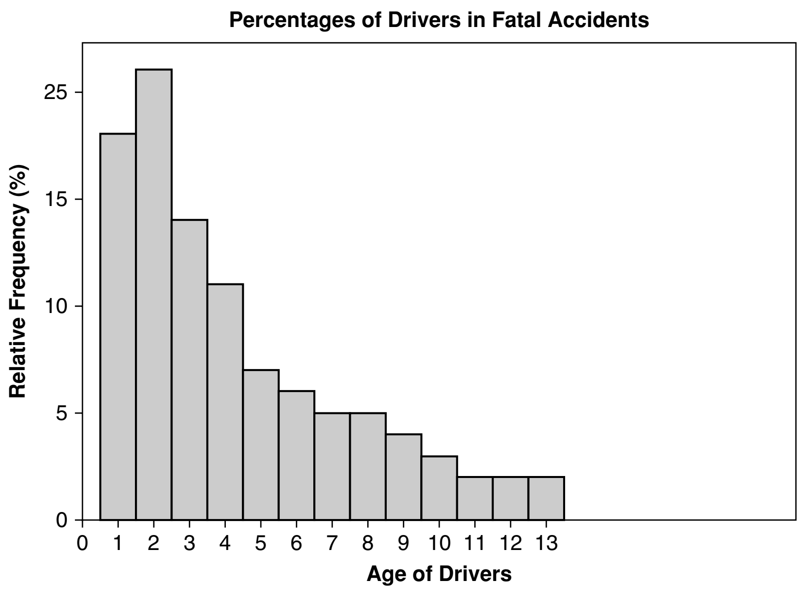
12. No. Based on the information given in the table for accidental (motor vehicle) death rates, it seems that the death rates are decreasing over time. Also, note that the same conclusion would be drawn if the annual accidental death rates based on **all** principal types of accidental deaths were compared (i.e., not just for motor vehicles).

14. (a) The relative frequency histogram for the ***age of drivers*** is shown below. The relative frequency is expressed as a percentage, and the interval width is 5.

***Note*:** The age intervals from the table in the text were coded such that age group 15–20 was represented as 1 along the horizontal axis, age group 20–25 was represented as 2, . . . , and the age class of over 75 was represented as 13 along the horizontal axis.

(b) The relative frequency histogram for the ***age of drivers*** who were killed in car accidents is shown below. The relative frequency is expressed as a percentage.

***Note*:** The age intervals from the table in the text were coded such that age group 15–20 was represented as 1 along the horizontal axis, age group 20–25 was represented as 2, . . . , and the age class over 75 was represented as 13 along the horizontal axis.



(c) The age group that accounts for the largest number of *fatal accidents* is from 20 to 25 years old. (Observe the histogram in part (b)).

(d) The age group from 15 to 20 years should *pay the highest insurance premiums* since it appears that the largest proportion of fatalities occur in this age group. (Rationale: the ratio $18/9 = 2$ is greater than any other ratio for the table.)

2.4 STEM-AND-LEAF PLOTS

Problems

***Note*: These stem-and-leaf plots can be displayed differently than what is presented here since one can use different stem-and-leaf classes.**

2. A stem-and-leaf plot for the **maximal marginal 2008 tax rates for a variety of states** is shown below.

0	.00, .00
3	.00, .40
4	.54, .63
5	.00, .00, .30, .50, .95
6	.00, .00, .00, .45
7	.00, .80
8	.25, .50, .98
9	.30

4. (a) The stem-and-leaf plot for the ***reaction times*** is shown below. It uses an increment of 1. That is, values are listed from 0.0 to 0.9, 1.0 to 1.9, 2.0 to 2.9, and 3.0 to 3.9.

0	.4, .5, .6, .6, .7, .7, .7, .9, .9	(9)
1	.1, .1, .2, .2, .3, .3, .3, .5, .5, .5, .6, .6, .7, 8, .8, .8, .8	(17)
2	.0, .1, .1, .2, .3, .4, .4, .5, .5, .6, .6, .6, .8, .8, .8, .9, .9	(17)
3	.0, .1, .2, .2, .3	(5)

(b) The stem-and-leaf plot for the ***reaction times*** is shown below. It uses an increment of 0.5. That is, values are listed from 0.0 to 0.4, 0.5 to 0.9, 1.0 to 1.4, etc.

0	.4	(1)
0	.5, .6, .6, .7, .7, .7, .9, .9	(8)
1	.1, .1, .2, .2, .3, .3, .3	(7)
1	.5, .5, .5, .6, .6, .7, .8, .8, 8, .8	(10)
2	.0, .1, .1, .2, .3, .4, .4	(7)
2	.5, .5, .6, .6, .6, .8, .8, .8, .9, .9	(10)
3	.0, .1, .2, .2, .3	(5)

(c) The stem-and-leaf plot for the 48 reaction times in part (b) seems to be more informative. We can readily see that an equal number of values lie between 1.0 and 1.4 and between 2.0 and 2.4. Also, an equal number of values lie between 1.5 and 1.9 and between 2.5 and 2.9. In addition, we can see that the majority of values lie between 1 and 2.9 (34 in all).

(d) 1.8. This value could be chosen since it is the middle (median) reaction time. Also, another possible answer is 1.9 or 2. This value could be chosen since there are an equal number of values between 1.0 and 1.9 and between 2.0 and 2.9. So one could report the average between 1 and 2.9 as a typical reaction time.

6. (a) A stem-and-leaf plot for the ***volatility of the 32 companies*** is shown below.

0.1	7, 9	(2)
0.2	0, 2, 4	(3)
0.2	6, 6, 6, 8, 8, 9	(6)
0.3	0, 0, 1, 1, 2, 3, 3, 3	(8)
0.3	5, 5, 7, 7, 7, 8	(6)
0.4	2, 4	(2)
0.4	5, 8	(2)
0.5	0, 1	(2)
0.5		(0)
0.6	3	(1)

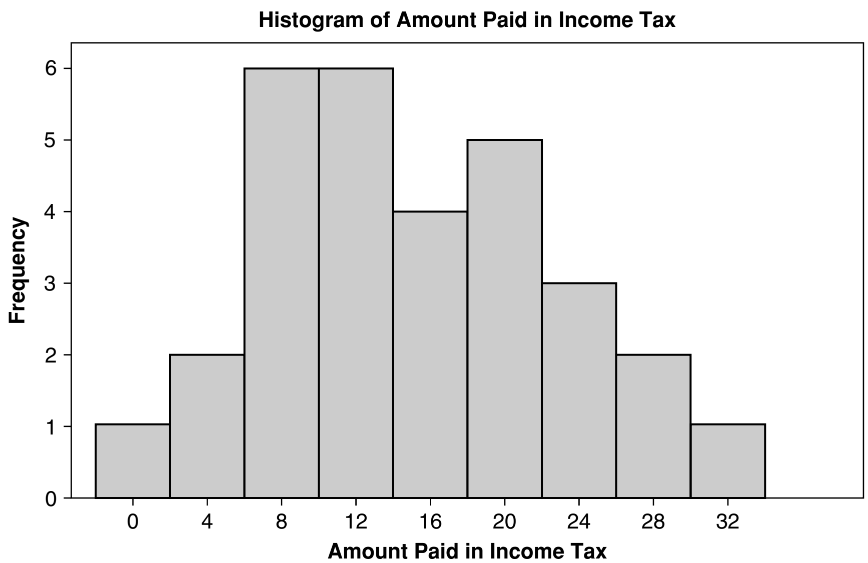
(b) The largest data value is 0.63.

(c) The smallest data value is 0.17.

(d) A "typical" data value would be 0.33. This value was chosen because almost half $(8+6=14)$ of the values are in the 0.30 to 0.39 range, and 0.33 is in the middle of this set. Also, 0.33 occurs three times in the data set.

8. By observing the stem-and-leaf plot, you would know the actual data values between the listed intervals on the histogram. By constructing a histogram, some information is lost since you would only know the frequency count within the intervals and not the actual values. You can also observe from the stem-and-leaf plot that the data values are ***almost*** evenly distributed within the classes. It would have been impossible to observe this on the histogram since only the frequency count for the classes is given. Also, many of the data values repeat, as evidenced by the stem-and-leaf plot but not the histogram.

10. (a) A histogram with 9 classes for the ***amounts paid in income tax.***



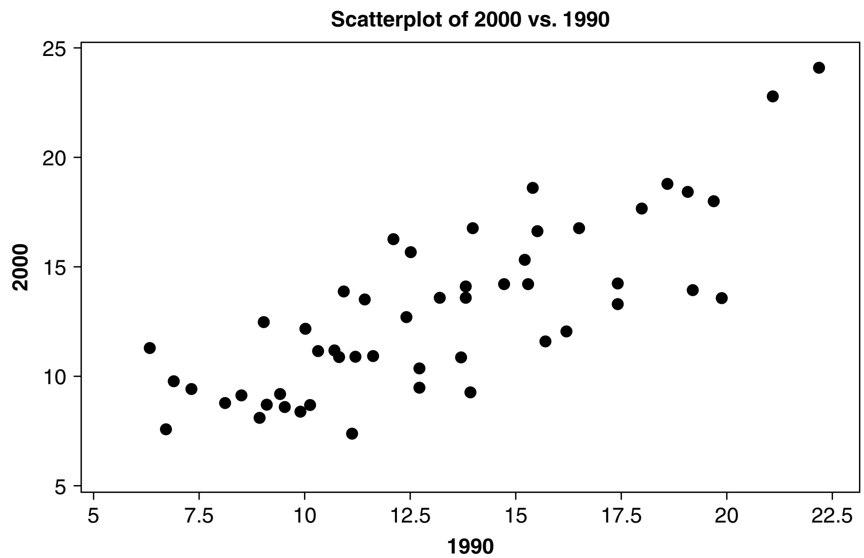
(b) A stem-and-leaf plot for the ***amounts paid in Social Security tax.***

0	1, 2, 2	(3)
0	6, 6, 6, 7, 7, 8	(9)
1	0, 1, 1, 1, 3, 3	(15)
1	5, 6, 7, 7, 9, 9	(15)
2	0, 1, 1, 2, 3, 4	(9)
2	6, 8	(3)
3	3	(1)

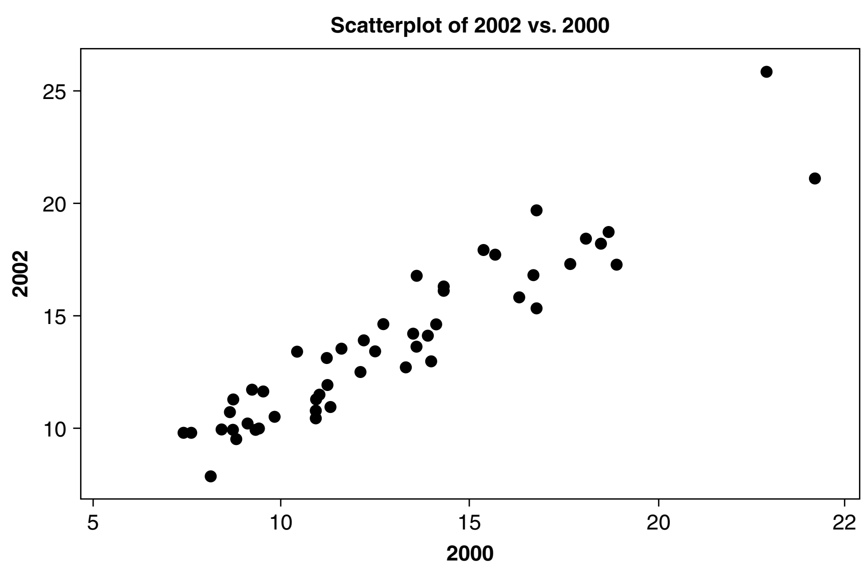
2.5 SETS OF PAIRED DATA

Problems

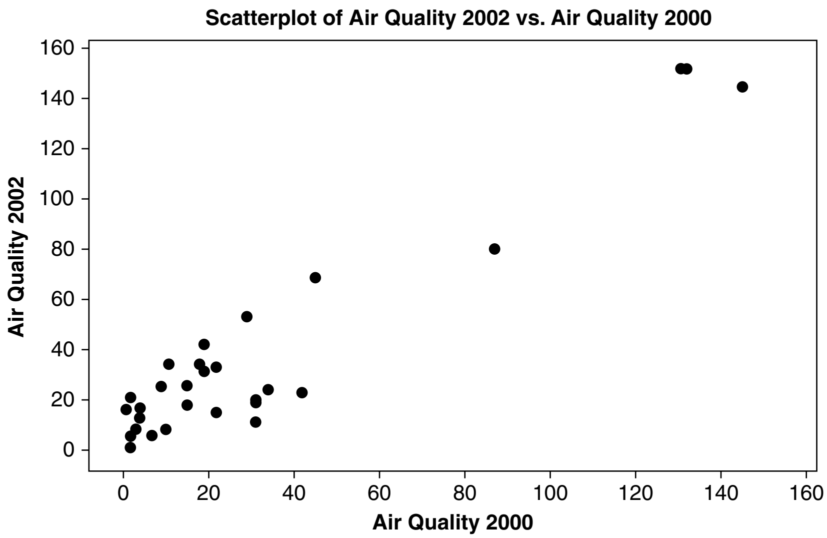
2. (a) The scatter diagram relating the 2000 and 1990 rates.



(b) The scatter diagram relating the 2000 and 2002 rates.

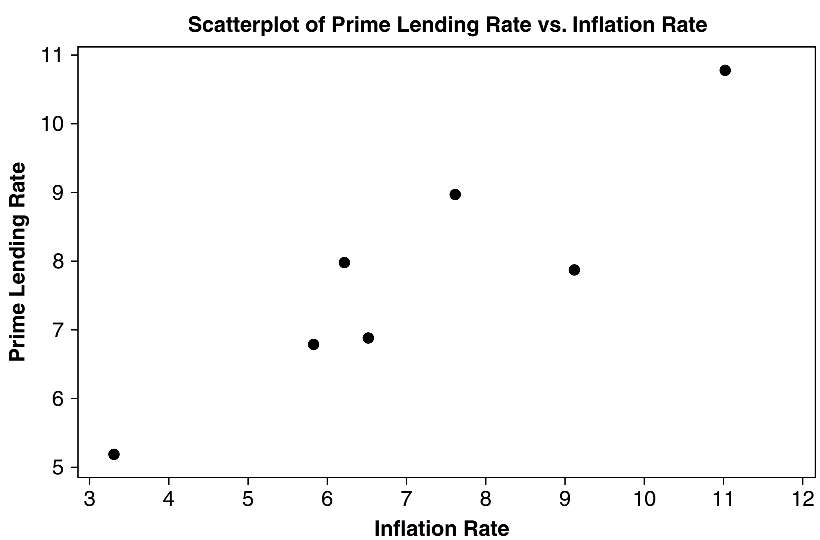


4. (a) The scatter diagram relating **the number of days in 2000 and 2002 that did not meet acceptable air quality standards** in a selection of U.S. metropolitan areas.



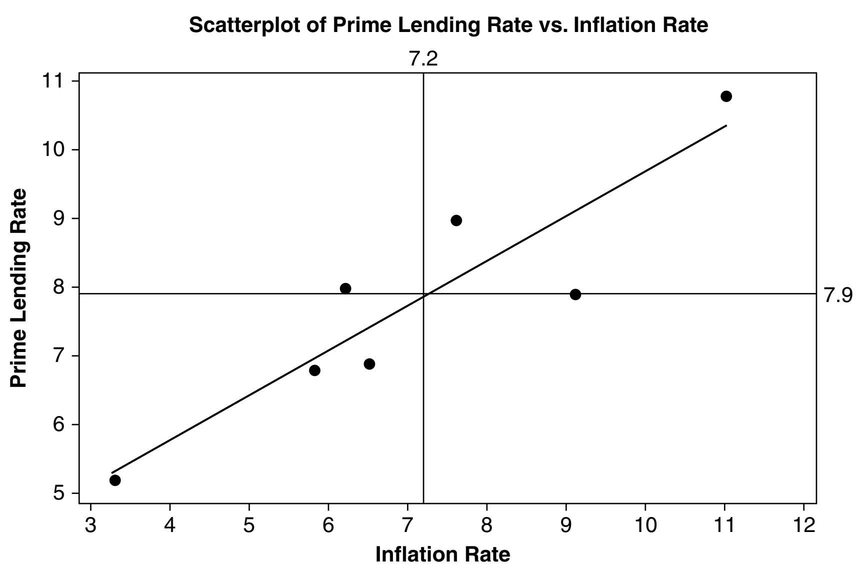
(b) Observe that higher values in 2000 appear "to go" with higher values in 2002.

6. (a) The scatter diagram of ***prime lending rate*** versus ***inflation rate*** is shown below.

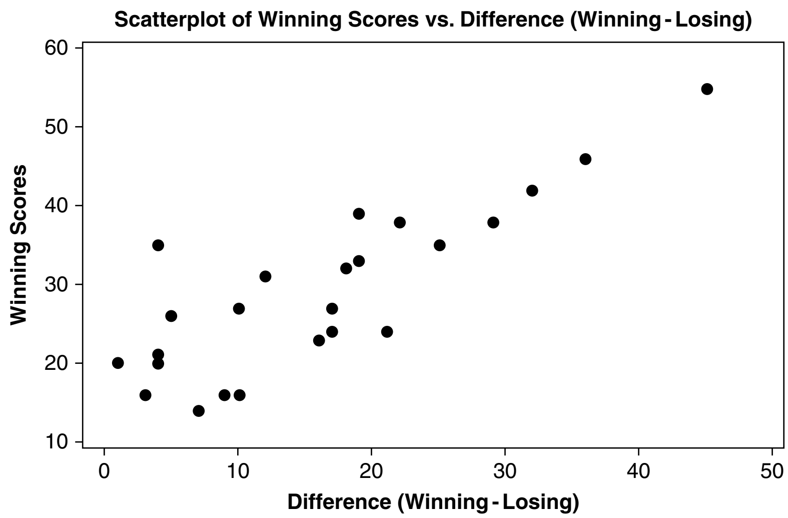


(b) The scatter diagram with a line drawn is shown below.

(c) The scatter diagram with lines converging at inflation rate = 7.2%. Observe from the graph that the corresponding prime lending rate is approximately 8%.

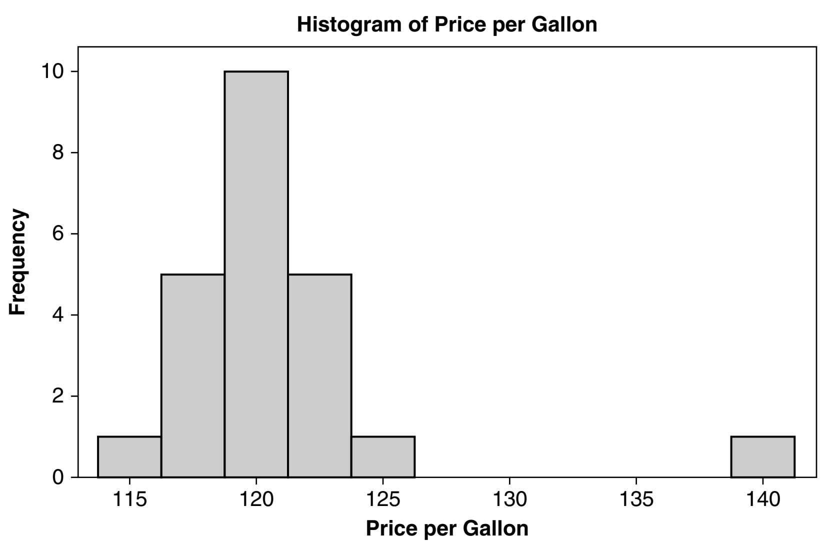


8. The scatter diagram for ***winning Super Bowl scores*** versus the ***amount by which the team won*** is shown below. From this diagram, it ***does*** appear that these two variables are dependent on each other and that larger values of the winning score tend to go with larger winning differences.

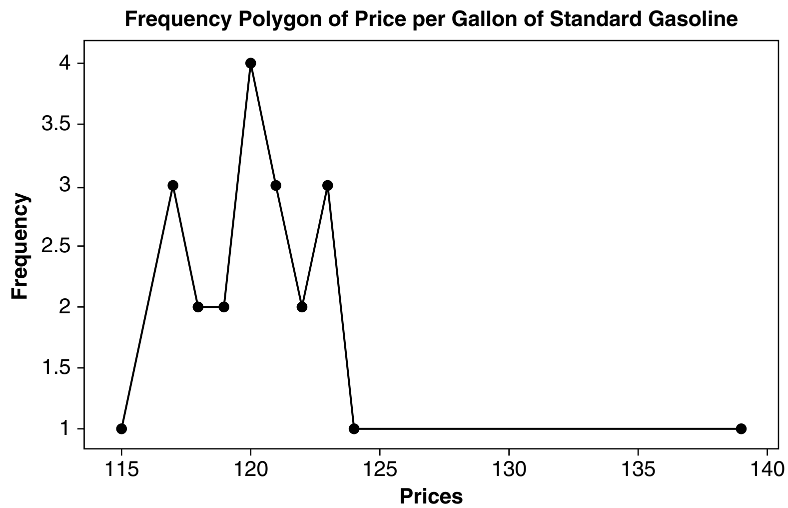


Review Problems

2. (a) A frequency histogram for the ***price per gallon of standard gasoline in the San Francisco Bay area in May of 1991*** is shown below.



(b) A frequency polygon for the data set is shown below.

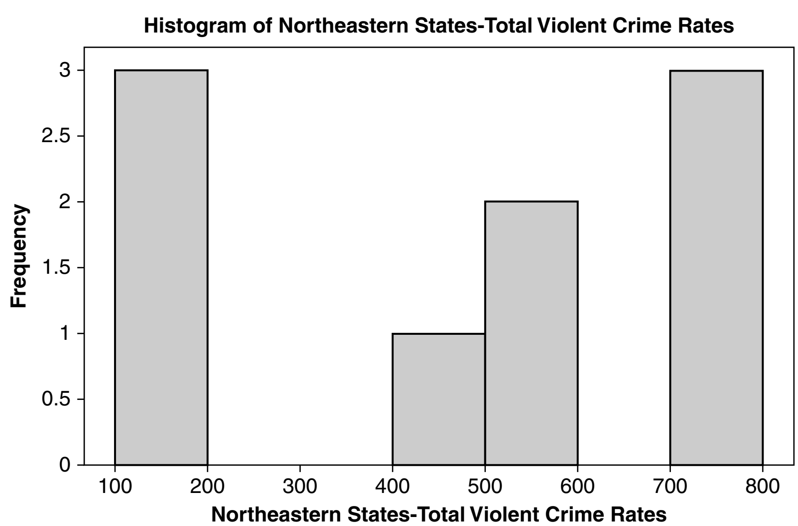


(c) A stem-and-leaf plot for the data is shown below.

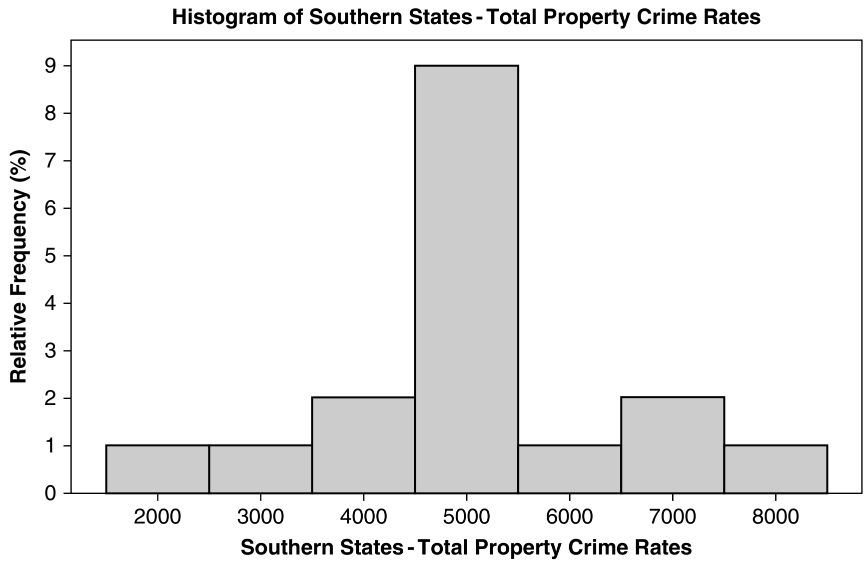
11	5, 7, 7, 7, 8, 8, 9, 9	(8)
12	0, 0, 0, 0, 1, 1, 1, 2, 2, 3, 3, 3, 4	(13)
13	9	(1)

(d) The value of 139 seems out of the "ordinary." This value seems somewhat larger than the other values.

4. (a) A frequency histogram for the ***total violent crime rates*** in the northeastern states in the United States is shown below. This histogram uses 7 classes.



(b) A relative frequency histogram for the ***total property crime rates*** in the southern states in the United States is shown below. This histogram uses 7 classes.



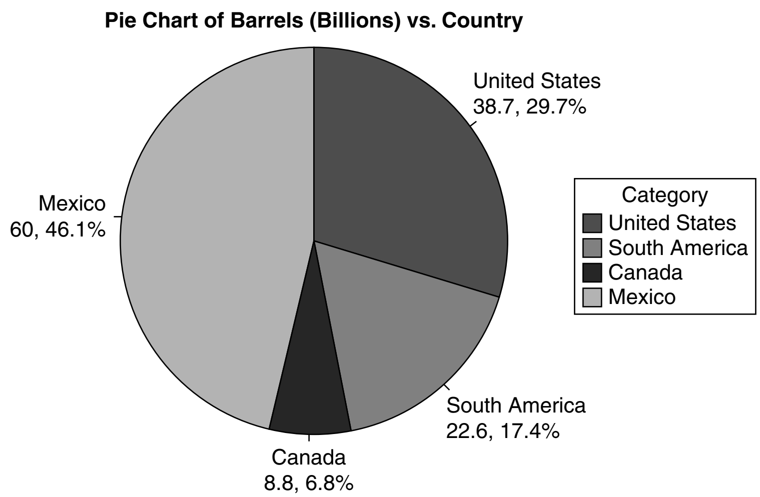
(c) A stem-and-leaf plot for the ***murder rates*** in the western states of the United States is shown below. This plot uses leaf units of 0.1.

1	.8	(1)
2	.6, .9	(2)
3	.3	(1)
4		(0)
5	.9	(1)
6		(0)
7	.8	(1)
8		(0)
9		(0)
10	.5	(1)
11	.8	(1)

(d) A stem-and-leaf plot for the ***burglary rates*** in the midwestern states of the United States is shown below.

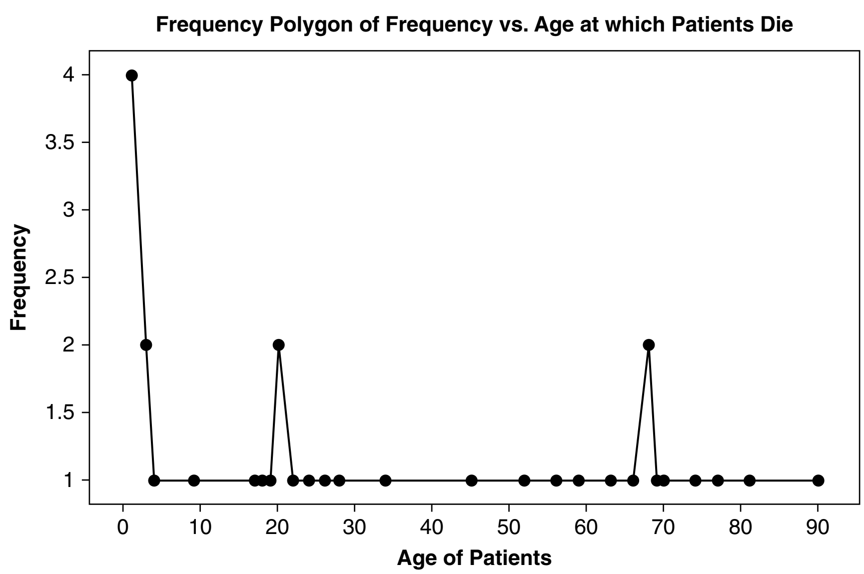
0	373, 590, 727, 752, 832, 854, 977	(7)
1	055, 120, 186, 253, 307	(5)

6. The pie chart for the ***estimated oil reserves*** (in billions of barrels) for the four regions in the western hemisphere is shown below. Note that there are two numbers associated with each country in the graph. For example, the two numbers associated with the United States are 38.7 and 29.796%. The first number of 38.7 represents the number of billions of barrels of oil reserve, and 29.796% represents the percentage of the total for the four regions.

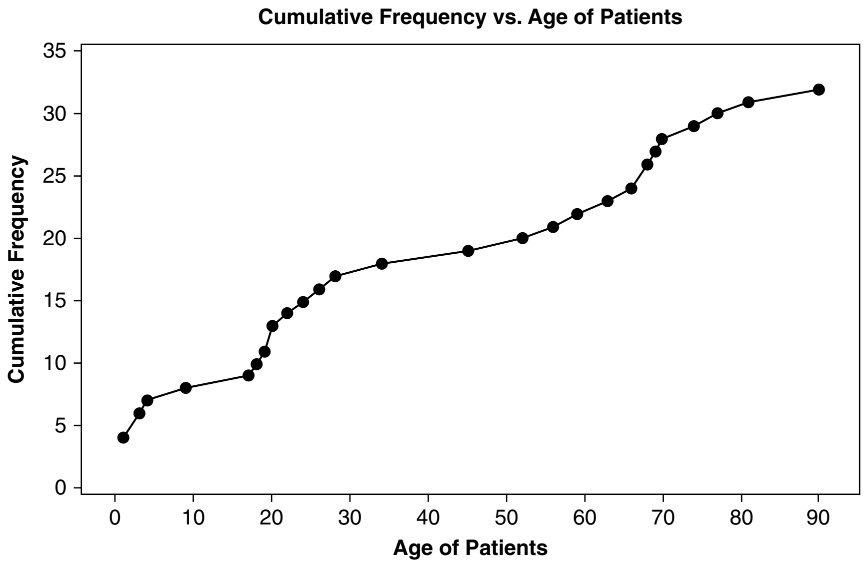


8. (a) A histogram (with 7 classes) for the ***ages at which patients died at a large inner-city (nonbirthing) hospital*** is shown below.

(b) A frequency polygon for the ***ages at which patients died at a large inner-city (nonbirthing) hospital*** is shown below.



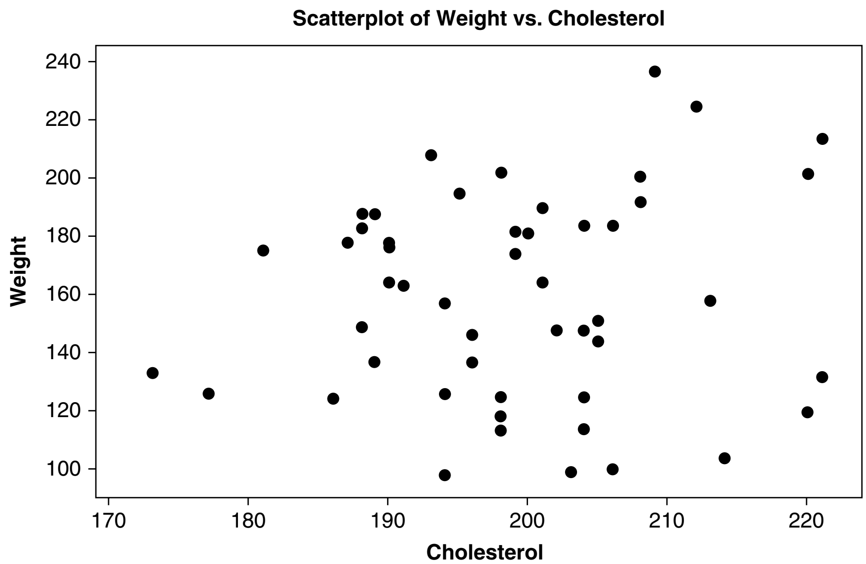
(c) A cumulative frequency polygon for the ***ages at which patients died at a large inner-city (nonbirthing) hospital*** is shown below.



(d) A stem-and-leaf diagram for the ***ages at which patients died at a large inner-city (nonbirthing) hospital*** is shown below.

0	1, 1, 1, 1, 3, 3, 4, 9	(8)
1	7, 8, 9	(3)
2	0, 0, 2, 4, 6, 8	(6)
3	4	(1)
4	5	(1)
5	2, 6, 9	(3)
6	3, 6, 8, 8, 9	(5)
7	0, 4, 7	(3)
8	1	(1)
9	0	(0)

10. (a) The scatter diagram for the ***weight versus the cholesterol level for the last 50 entries in Appendix A*** is shown below.



(b) From the scatter diagram in part (a), it appears that there is little or no association between these two variables.

12. (a) A side-by-side stem-and-leaf plot for the ***math SAT score and the verbal SAT score for the 14 seniors*** is shown below.

	Math Score Leaves	Stem	Verbal Score Leaves	
(3)	88, 75, 58	4	90	(1)
(5)	76, 75, 52 50, 05	5	04, 20, 24, 28, 30	(5)
(4)	72, 65, 22, 06	6	04, 05, 20, 30, 46, 90	(6)
(2)	80, 04	7	20, 20	(2)

(b) The students seem to have done better on the verbal portion of the SAT exam since 13 students scored above 500 as opposed to only 11 students who scored above 500 on the math portion.

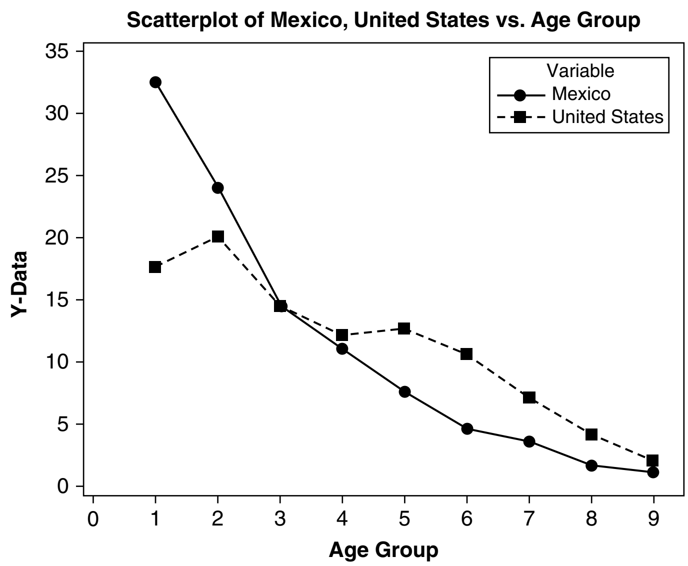
14. (a) Percentage of the Mexican population whose age is less than 30 years

$= 32.5 + 24 + 14.5 = 71.$

(b) Percentage of the U.S. population whose age is less than 30 years

$= 17.5 + 20 + 14.5 = 52.$

(c) The relative frequency (in %) polygons for Mexico and the U.S. populations are shown below.



(d) In general, the cumulative proportion of the population for the different age classifications for the United States is always below that of Mexico, except for the largest age groups, where the cumulative proportions both equal 100%.

16. One way of determining the outliers (as defined in this problem) is to arrange the data sets in order of magnitude and observe whether any value(s) fits (fit) this interpretation of an outlier.

(a) The ordered (ascending) data set is given below:

$$-17, 5, 10, 12, 14, 17, 18, 22, 22, 25, 28, 33$$

Thus, by the interpretation of "far away from the other values," -17 will be the only outlier. All other values are relatively close to one another except -17.

(b) The ordered (ascending) data set is given below:

$$2, 5, 7, 9, 10, 12, 12, 13, 15, 16, 18, 22, 54$$

Thus, by the interpretation of "far away from the other values," 54 will be the only outlier. All other values are relatively close to one another except 54.

(c) The ordered (ascending) data set is given below:

$$3, 6, 14, 17, 18, 20, 22, 24, 25, 27, 28, 43, 52$$

Thus, by the interpretation of "far away from the other values," 43 and 52 will be the only two outliers. All other values are relatively close to one another except 43 and 52.

Chapter 3 USING STATISTICS TO SUMMARIZE DATA SETS

3.2 SAMPLE MEAN

Problems

2. The sample mean $\bar{x} = (10 + 14.8 + 23.4 + 26.4 + 30.1)/5 = 20.94$ (pounds of cheese per capita consumption).

4. (a) Since the mean of the first four numbers is 14, then the total of these four numbers is $4 \times 14 = 56$. Thus, if the fifth number is 24, the mean of these five numbers will be $(56 + 24)/5 = 16$.

(b) Let x be the value of the fifth number. Since the mean of the five numbers is 24, then $(4 \times 14 + x)/5 = 24$ from which $x = 64$.
Alternatively, note that the fifth number will be $25 \times 5 - 14 \times 4 = 69$.

6. (a) The sum of the 10 values will be equal to $10 \times 20 = 200$. Since one of the values was 13, the sum of the remaining nine values will be $200 - 13 = 187$. Thus, with the actual value of 15, the new sum will be $187 + 15 = 202$. Thus, the revised $\bar{x}$ value will be $202/10 = 20.2$.

(b) With an additional value of 22, $\bar{x}$ will be equal to $(202 + 22)/11 = 20.36$. Thus, the sample mean will increase.
One should also note that since 22 is larger than $\bar{x}$, then adding 22 to the data set will increase the sample mean.

(c) Using the value of 13 as one of the observed values, the new value of $\bar{x}$ will be $(200 + 22)/11 = 20.18$.

8. The sample mean $\bar{x} = (182 + 184 + 187 + 170 + \cdots + 132 + 131 + 135 + 135)/15 = 157.33$ (bowling score).

10. Since $y_i = ax_i + b$ for $i = 1, 2, \ldots, n$, then the sample mean for the y_is will be given by

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = \frac{1}{n}\sum_{i=1}^{n}(ax_i) + \frac{1}{n}\sum_{i=1}^{n} b$$

That is,

$$\bar{y} = a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \frac{1}{n}(nb) = a\bar{x} + b$$

12. The sample mean $\bar{x} = (8010 + 11{,}653 + 9{,}783 + \cdots + 12,326)/6$

$$= 11{,}098.167 \text{ (in units of 1,000)}$$

That is, the average number of passenger cars produced during the 6 years is 11,098,167.

14. The sample mean $\bar{x} = (16 \times 9 + 17 \times 12 + \cdots + 20 \times 8)/54 = 17.9259$.

16. The sample size $n = 60$. Thus, the sample mean can be computed from $\bar{x} = (122 \times 30 + 160 \times 30)/60 = 141$. This is equivalent to using weights of 30/60. Alternatively, the sample mean can be computed as $(122 + 160)/2 = 141$.

18. The combined sample size is $n_1 + n_2$. So the weights will be $\frac{n_1}{n_1+n_2}$ and $\frac{n_2}{n_1+n_2}$. Thus, the sample mean of the combined sample will be $\bar{x} = \frac{n_1}{n_1+n_2}\bar{x}_1 + \frac{n_2}{n_1+n_2}\bar{x}_2$. Alternatively, we can write this as $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$.

20. Recall, the sample mean was $\bar{x} = 17.9259$. The observed values along with their deviations are shown in the table below.

x	16	17	18	19	20
$x - \bar{x}$	−1.9259	−0.9259	0.0741	1.0741	2.0741

Thus, the sum of the deviations is

$$\sum_{i=1}^{6} (x_i - \bar{x}) \times f = (-1.9259 \times 9 - 0.9259 \times 12 + 0.0741 \times 15 + 1.0741 \times 10 + 2.0741 \times 8) = 0.0022 \approx 0$$

Note: This sum did not equal zero because of rounding $\bar{x} = 968/54 \approx 17.9259$ to four decimal places. If a more precise value is used, the sum of the deviations will equal zero.

3.3 SAMPLE MEDIAN

Problems

2. (a) First arrange the data set in ascending order. Since the data set is odd ($n = 15$), the median will be the middle value (8th in the ordered set). That is, the median $m = 14$.

(b) The new sample median will be increased by 5. That is, $m = 19$.

(c) The new sample median will be multiplied by 3. That is, $m = 14 \times 3 = 42$.

4. First arrange the data set in ascending order. Since the data set is even ($n = 40$), the median will be the average of the two middle numbers (average of the 20th and the 21st in the ordered set). That is, the sample median $m = (29 + 30)/2 = 29.5$.

6. First arrange the data set in ascending order. Since the data set is odd ($n = 13$), the median will be the middle value (7th in the ordered set). That is, the median of the average annual number of days of precipitation for the cities $m = 100$.

8. First arrange the data set in ascending order. Since the data set is odd ($n = 51$), the median will be the middle value (26th in the ordered set). That is, we are finding the median of the State Sales Tax Rates for the 50 states and the District of Columbia, $m = 5.5$.

10. (a) Since $n = 10$, then the median $m = 5$ will be the average of the 5th and 6th values. Adding the value of 7 to the data set will make it the 6th ordered value, the 7th ordered value, or the 8th ordered value, etc. The median will increase (as it becomes the 6th value) if the 5th ordered value is strictly smaller than the (old) 6th value. Otherwise, the median remains unchanged.
(b) If the value of 3 is less than the 5th ordered value and the value of 42 is larger than the 6th ordered value, then the median remains the same. Otherwise, the new median will depend on the specific values of the 5th and 6th ordered values.

12. (a) Sample mean is higher for company A.
(b) No inference can be drawn about the sample medians.

14. (a) Sample mean, since total profit is of interest and since the total profit is a function of the mean profit.
(b) Sample median, since some scores may be outliers.
(c) Sample median, since some IQ scores may be outliers.
(d) Sample median, since the new more expensive homes may be outliers in terms of value.

16. Arrange the data sets in ascending order, and since the sample size $n = 30$, the median will be the average of the 15th and 16th values in the ordered sets.

(a) Weight: $\bar{x} = 148.07$ and the median $m = 145$.
(b) Cholesterol levels: $\bar{x} = 195.57$ and the median $m = 194.5$.
(c) Blood pressure: $\bar{x} = 115.20$ and the median $m = 112.5$.

3.3.1 Sample Percentiles

Problems

2. Q1 = 5.6, Q2 (median) = 6.7, Q3 = 8.05

4. First arrange the data set in ascending order.

(a) $n = 12$, $p = 0.4 \Rightarrow np = 12 \times 0.4 = 4.8$. Since np is not an integer round up to the next integer (5). Thus, the sample 40th percentile for the physician's rates per 100,000 will be located in the 5th position. Hence, the sample 40th percentile is 162.

(b) $n = 12,\ p = 0.6 \Rightarrow np = 12 \times 0.6 = 7.2$. Since np is not an integer round up to the next integer (8). Thus, the sample 60th percentile for the physician's rates per 100,000 will be located in the 8th position. Hence, the sample 60th percentile is 177.

(c) $n = 12,\ p = 0.8 \Rightarrow np = 12 \times 0.8 = 9.6$. Since np is not an integer round up to the next integer (10). Thus, the sample 80th percentile for the physician's rates per 100,000 will be located in the 10th position. Hence, the sample 80th percentile is 188.

6. It is given that the sample $100p$ percentile of a data set is 120. Adding a value of 30 to each data values will change the $100p$ percentile of 120 to $120 + 30 = 150$. Thus, the new sample $100p$ percentile will be 150.

8. First arrange the data values in ascending order. $n = 22,\ p = 0.9 \Rightarrow np = 22 \times 0.9 = 19.8$. Since np is not an integer round up to the next integer (20). Thus, the sample 90th percentile for the data set will be located in the 20th position. Hence, the sample 90th percentile is 121.

10. (a) An interval in which approximately 50% of the data will lie is 35–66.

(b) A value which is greater than approximately 50% of the data is 47.

(c) A value such that approximately 25% of the data are greater than it is 66.

3.4 SAMPLE MODE

Problems

2. From the frequency distribution for the winning Masters Golf Tournament scores shown in Example 2.2, the sample mode is 279.

4. If you had to guess the exact salary, then your greatest chance of being correct would be from using the sample mode. However, since rounding to the nearest \$1,000 allows you to guess to within \$500 in either direction, the sample median is the wiser choice. The sample mean will be a poor choice due to the possibility of outliers.

6. Since the sample mode for the data set $x_i,\ i = 1, 2, \ldots, n$ is 10, then the sample mode for the data set $y_i = 2x_i + 5,\ i = 1, 2, \ldots, n$ will be $2(10) + 5 = 25$.

8. Sample mean is 110. Sample median is at least 100 and at most 120. We can't say anything about the sample mode.

3.5 SAMPLE VARIANCE AND SAMPLE STANDARD DEVIATION

Problems

2. (a) To just guess as to which data set has the larger sample variance, one can simply compute the range.

Range for data set A = $75 - 66 = 9$.
Range for data set B = $16 - 2 = 14$.

Thus, one can say that since the range for data set B is larger, it appears that data set B will have the larger sample variance.

(b) The sample mean for data set A is $\bar{x} = 70.6667$ with $n = 6$.

Now $\sum_{i=1}^{6} x_i^2 = 30,014$, $n\bar{x}^2 = 6 \times 70.6667^2 = 29{,}962.6949$; thus, the sample variance for data set A is

$$s^2 = (30,014 - 29{,}962.6949)/(6 - 1) = 10.26$$

correct to two decimal places.

Note: If you had used two decimal places for the mean to start with, the variance would have been 9.70. So be careful – you should use four decimal places throughout your computations, and then round your answer to two decimal places.

(c) The sample mean for data set B is $\bar{x} = 8.6667$ with $n = 6$.

Now $\sum_{i=1}^{6} x_i^2 = 566$, $n\bar{x}^2 = 6 \times 8.6667^2 = 450.6667$, thus the sample variance for data set B is

$$s^2 = (566 - 450.6667)/(6 - 1) = 23.07$$

correct to two decimal places.

4. The yearly number of physicians is consistently a bit less than 3 (say roughly 2.7) times the number of dentists, so the sample variance of the doctors should be about $(2.7)^2 = 7.29$ times the variance of the doctors.

6. (a) The sample mean $\bar{x} = (720 + 880 + \cdots + 1260 + 800)/10 = 832.5$.

(b) First arrange the data set in ascending order. Since $n = 10$, then the sample median will be the average of the values in the 5th and 6th positions. Thus, the sample median $m = (720 + 800)/2 = 760$.

(c) $\bar{x} = 832.5$, $\sum_{i=1}^{10} x_i^2 = 7{,}366{,}193$, $n\bar{x}^2 = 6{,}930{,}563$, $n = 10$, so the sample variance $s^2 = (7{,}366{,}193 - 6{,}930{,}563)/(10 - 1) = 48{,}403.3333$. Thus, the sample standard deviation $s = \sqrt{48{,}403.333} = 220$.

10. Let $y_i = ax_i + b,\ i = 1, 2, \ldots, n$. Thus, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} (ax_i + b)}{n}$. That is, $\bar{y} = \frac{\sum_{i=1}^{n} ax_i}{n} + \frac{\sum_{i=1}^{n} b}{n} = a\frac{\sum_{i=1}^{n} x_i}{n} + \frac{nb}{n} = a\bar{x} + b$. Now, $s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$

$$\Rightarrow s_y^2 = \frac{\sum_{i=1}^{n} \left[(ax_i + b) - (a\bar{x} + b)\right]^2}{n-1} = \frac{\sum_{i=1}^{n} [(ax_i - a\bar{x})]^2}{n-1} = \frac{a^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$= a^2 s_x^2 = a^2 s^2$ if s_x^2 is replaced with s^2.

12. Let $\overline{F} = 40$ and $s_F^2 = 12$. From Prob. 10,

(a) $\overline{C} = \frac{5}{9}(\overline{F} - 32)$, so $\overline{C} = \frac{5}{9}(40 - 32) = 4.44\,°\text{C}$.

(b) $s_C^2 = \left(\frac{5}{9}\right)^2 s_F^2 = \left(\frac{5}{9}\right)^2 (12) = 3.70\,°\text{C}$.

14. Based on Prob. 10, $s_y^2 = a^2 s_x^2$; then, the standard deviation for y will be $s_y = |a|\, s_x$.

16. $\bar{x} = 4.75,\ \sum_{i=1}^{8} x_i^2 = 188,\ n\bar{x}^2 = 180.5,\ n = 8$, so the sample variance

$s^2 = (188 - 180.5)/(8 - 1) = 1.0714 \Rightarrow$ the sample standard deviation $s = 1.0351$.

18. (a) Range for data set A $= 10 - 0 = 10$.
Range for data set B $= 10 - 0 = 10$.

(b) Data set A : $\bar{x} = 4.97,\ \sum_{i=1}^{5} x_i^2 = 198.3,\ n\bar{x}^2 = 148.007,\ n = 6$, so the sample variance $s^2 = (198.3 - 148.007)/(6 - 1) = 10.0586 \Rightarrow$ the sample standard deviation $s = 3.1715$.

Data set B : $\bar{x} = 4.83,\ \sum_{i=1}^{5} x_i^2 = 271.42,\ n\bar{x}^2 = 140.167,\ n = 6$, so the sample variance $s^2 = (271.42 - 140.167)/(6 - 1) = 26.2506 \Rightarrow$ the sample standard deviation $s = 5.1235$.

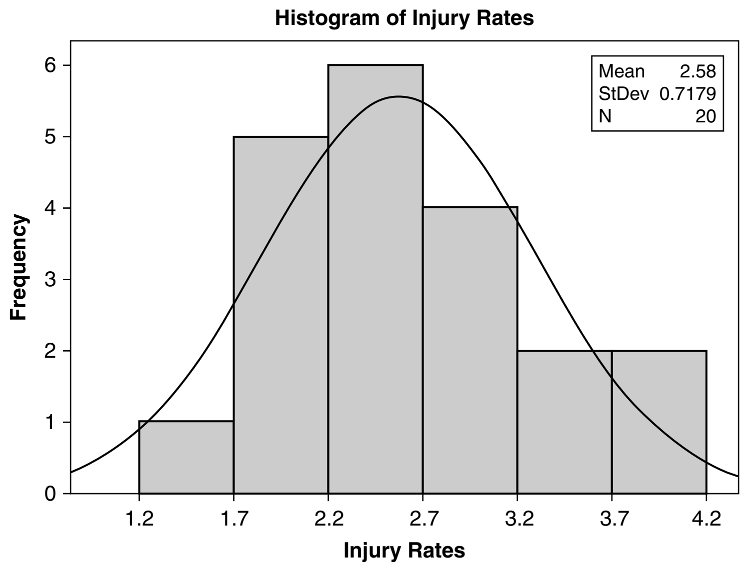
(c) Interquartile range for data set A = 75th percentile − 25th percentile
$= 5.2 - 4.5 = 0.7$.

Interquartile range for data set B = 75th percentile − 25th percentile
$= 9.5 - 0.1 = 9.4$.

3.6 NORMAL DATA SETS AND THE EMPIRICAL RULE

Problems

2. (a) The histogram using 6 classes is shown below.



(b) The histogram does not appear symmetric.

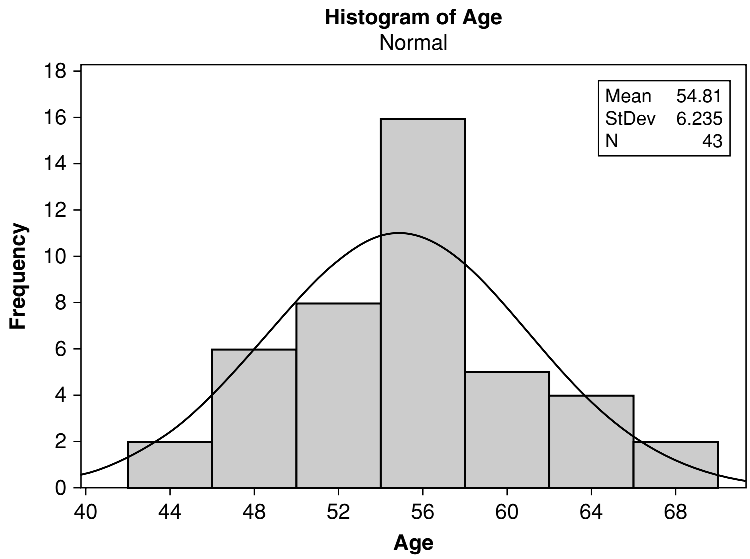
(c) The data set is skewed to the right.

(d) No response because of the response of *no* in part (b).

6. (a) $\bar{x} = 54.6364$ and the sample standard deviation $s = 6.273$.

(b) The histogram for the age at inauguration of all 44 presidents of the United States is shown below.

(c) From the histogram, the data appears to be approximately normal.



(d) Since the data is approximately normal, then approximately 95% of the data values will lie within two standard deviations of the mean. That is, approximately 95% of the data will lie between $54.6364 \pm 2(6.273)$ or from 42.0904 to 67.1824. (This computation used three decimal places throughout.)

Note: If you do not round during the intermediate steps but then round the final answer to three decimal places you will get 42.345 to 67.283.

(e) When the data set is arranged in ascending order, 41 out of the 44 values will lie in the interval of 42.0904 to 67.1824. Thus, the percentage of values that actually lie in the interval is 93.2 %.

8. (a) The sample mean $\bar{x} = 54.97$ and the sample standard deviation $s = 10.10$.

(b) Since the stem-and-leaf plot appears normally distributed, approximately 68% of the data values will lie between $\bar{x} - s$ and $\bar{x} + s$, and approximately 95% of the data values will lie between $\bar{x} - 2s$ and $\bar{x} + 2s$.

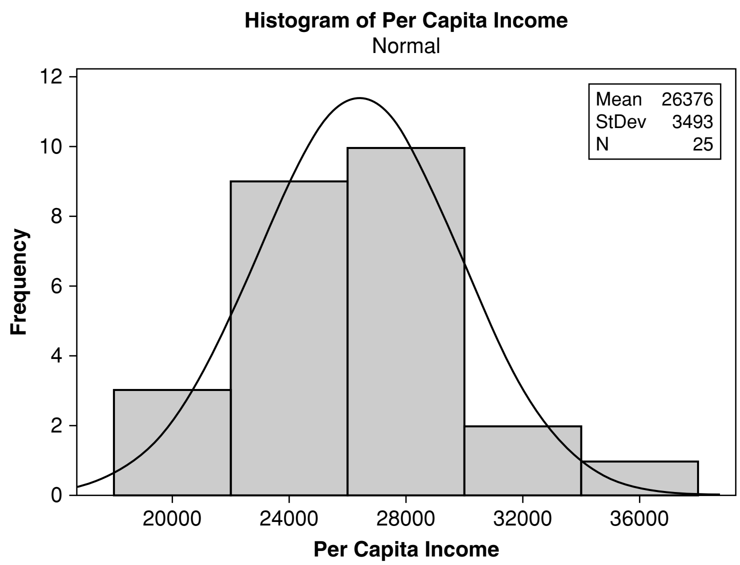
(c) Approximately 68% of the data values will lie between one standard deviation from the mean. That is, approximately 68% of the values will lie between 44.87 and 65.08. There are 25 values in the interval; thus, the actual percentage of values in this interval is $(24/36) \times 100\% = 66.7\%$.

Approximately 95% of the data values will lie between two standard deviations from the mean. That is, approximately 95% of the values will lie between 34.77 and 75.18. All 36 values are in this interval; thus, the actual percentage of values in this interval is 100%.

10. (a) The sample mean $\bar{x} = 27$.
(b) The sample median $m = 22$.
(c) The sample standard deviation $s = 15.44$.
(d) The histogram with 7 classes for the data set is shown below. From the histogram, the data appears to be right skewed.

(e) One standard deviation from the mean will constitute the interval from 11.56 to 42.44. The number of data values in this interval is 31. Thus, the percentage of actual values in this interval is $(31/36) \times 100\% = 86.11\%$.
(f) The Empirical Rule stipulates that approximately 68% of the data values will lie within one standard deviation from the mean. Because of the percentage computed in part (e), it is rather obvious that the approximation of a normal distribution is not a very good one.

12. (a) A histogram using 5 classes is shown below for the per capita income for the final 25 states in the table.



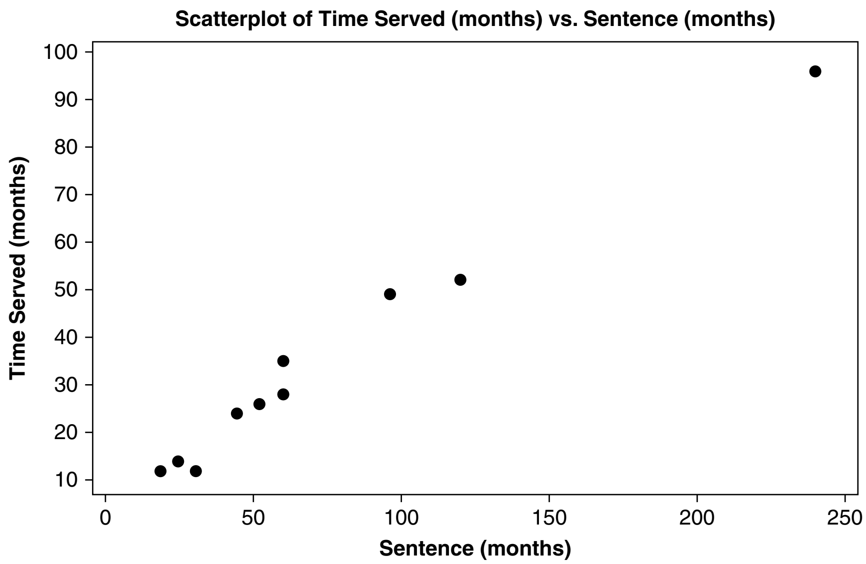
(b) The sample mean $\bar{x} = 26{,}376$.
(c) The sample median $m = 26{,}474$.
(d) The sample variance is $s^2 = 12{,}202{,}420$.
(e) From the histogram in part (a), the data appears to be approximately normal.
(f) Approximately 68% of the data values should lie between 22,882 and 29,868.
(g) Approximately 95% of the data values should lie between 19,389 and 33,361.
(h) The actual number of values in the interval given in part (f) is 18. Thus, the actual proportion of values that are in the interval is $(17/25) \times 100\% = 68\%$.
(i) The actual number of values in the interval given in part (g) is 24. Thus, the actual proportion of values that are in the interval is $(24/25) \times 100\% = 96\%$.

3.7 SAMPLE CORRELATION COEFFICIENT

Problems

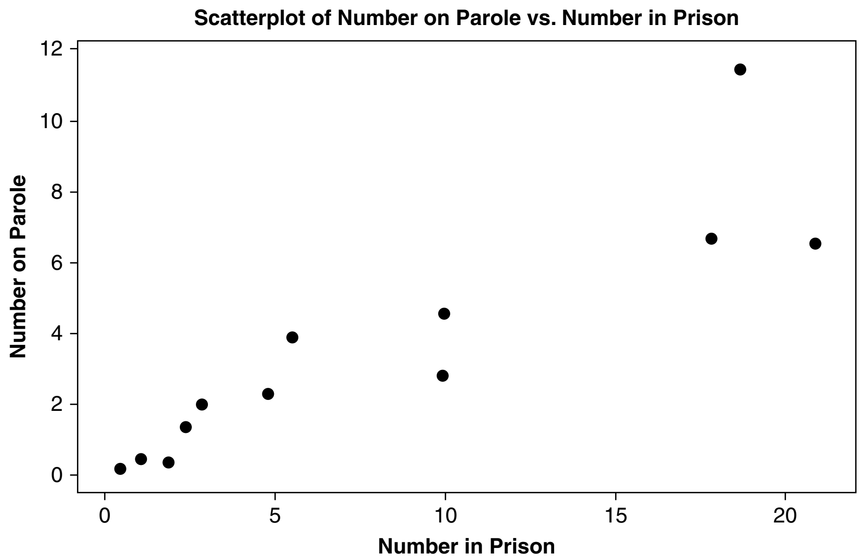
2. The sample correlation coefficient $r = 1$ for all three data sets.

4. The scatter diagram of sentence time versus the amount of time that was actually served is shown below.



The computed sample correlation $r = 0.9905$. Based on the scatter diagram and the value of r, we can conclude that there is a very strong positive correlation between these two variables.

6. (a) The scatter diagram for the number of adults in prison versus the number on parole for 12 midwestern states is shown below. The data are in thousands of adults.

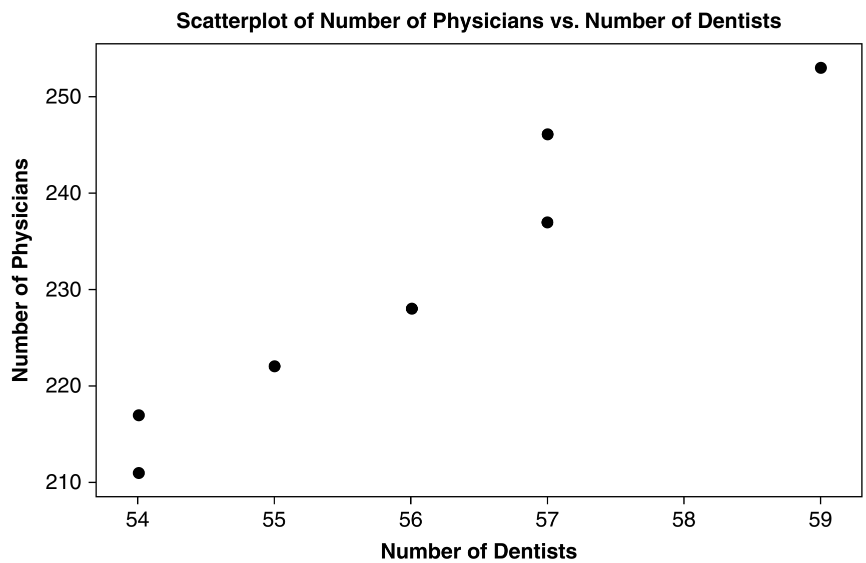


(b) The sample correlation coefficient between the number of adults in state prison and on parole in that state for the 12 midwestern states is $r = 0.8982$.

(c) Large.

8. The sample correlation coefficient $r = 0.32566$.

10. (a) From the scatter diagram for the number of physicians versus the number of dentists per 100,000 population for the given years, we can see that these variables are positively correlated.



(b) No. Certain time periods may be more prosperous (for individuals attending medical or dental school) or may be related to overall publicity regarding both medical and dental care. It may be just coincidence that these variables are correlated.

12. (a) Using all the data values, the sample correlation coefficient between the death rates of stomach cancer and female breast cancer is $r = -0.6805$.

(b) Using the first seven data values, the sample correlation coefficient between the death rates of stomach cancer and female breast cancer is $r = -0.6289$.

14. (a) No.

(b) Correlation is not causation. One possible explanation for the correlation is perhaps the more responsible or more educated parents encouraged their children to focus both on academics and oral hygiene.

16. Offensive players (especially the very good ones), who tend to have more playing time than a defensive player, has more opportunity to both score points and to commit fouls. Thus, this gives a possible explanation for the positive correlation between the number of points scored and the number of fouls committed.

18. Not necessarily; foreign-born immigrants could just be more likely to become residents of a high income state.

Review Problems

2. (a) The sample mean $\bar{x} = 74.13$.
(b) The sample median $m = 73$.
(c) The sample mode is 89.
(d) The sample standard deviation $s = 12.65$.
(e) Yes. The data appears to be approximately normal.
(f) Now $\bar{x} - 2s = 48.83$ and $\bar{x} + 2s = 99.44$; thus, 100% of the data values lie within two standard deviations of the mean.
(g) The Empirical Rule says that approximately 95% of the data values will lie within two standard deviations of the mean when the distribution is normal.

4. (a) 88
(b) 74
(c) 71
(d) 37 and 60

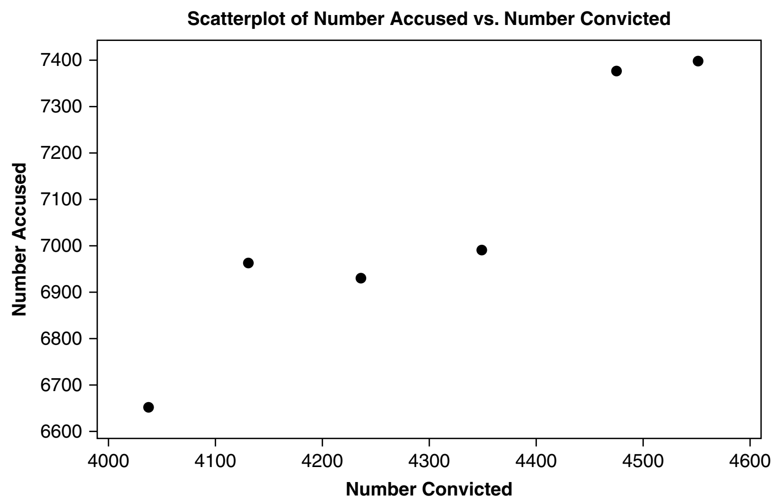
6. (a) A stem-and-leaf diagram for the birth weights is shown below.

2	.4	(1)
3	.3	(1)
4	.1	(1)
5	.0, .1, .2, .6, .8, .9, .9	(7)
6	.0, .1, .2, .3, .3, .4, .4, .5, .7, .8	(9)
7	.2, .4, .5, .5, .6, .6, .7, .8, .8, .9, .9	(11)
8	.3, .5, .8	(3)
9	.2, .7, .8, .9	(3)
10	.0, .3, .5	(3)

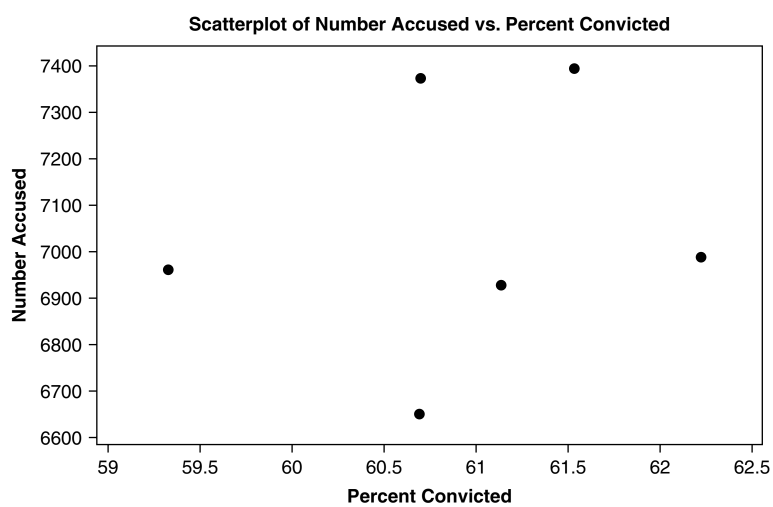
(b) The sample mean $\bar{x} = 7.095$.
(c) The sample median $m = 7.0$.
(d) The sample standard deviation $s = 1.848$.
(e) Now $\bar{x} - 2s = 3.40$ and $\bar{x} + 2s = 10.79$. Since 39 out of the 41 values fall within this interval, then the proportion of data that lies within these two numbers is $(39/41) \times 100\% = 95.1\%$.
(f) Yes. The stem-and-leaf diagram in part (a) indicates that the data set is approximately normal. However, others may view the graph as skewed to the left.
(g) 95% of the data values should lie between 3.39 and 10.666.

8. (a) Number accused: the sample mean $\bar{x} = 7,050$
the sample median $m = 6,975$.
(b) Number convicted: the sample mean $\bar{x} = 4,296.2$
the sample median $m = 4292$.
(c) The sample standard deviation of the number accused $s = 286$.

(d) The sample standard deviation of the number convicted $s = 199$.
(e) Positive correlation.
(f) Sample correlation coefficient of the numbers accused and convicted $r = 0.9413$.
(g) Since the "percentage" is the number convicted divided by the number accused, then the number convicted and the percentage convicted do not differ by a multiplicative constant. Hence, the "percentage" should be calculated prior to the computation of r. The sample correlation coefficient of the number accused and the percent of these that are convicted is $r = 0.1909$.
(h) (i) The scatter diagram for part (f) is shown below.

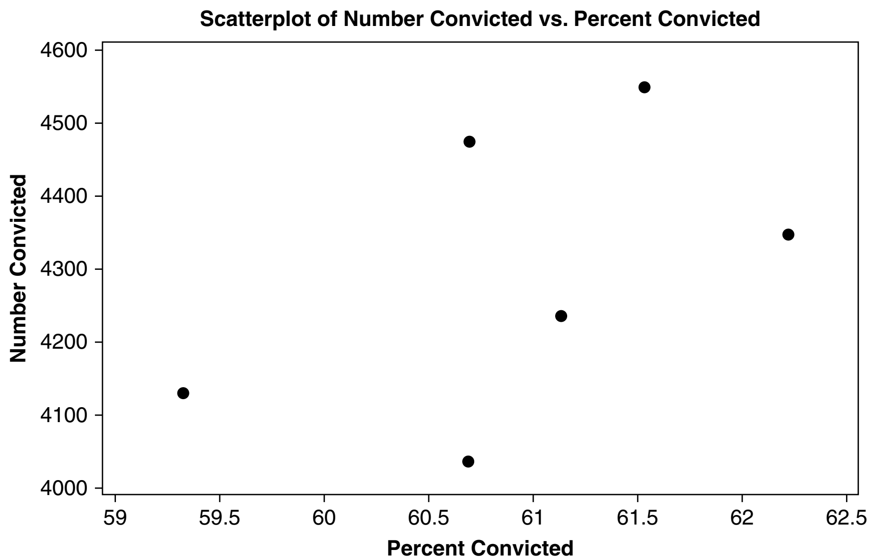


(ii) The scatter diagram for part (g) is shown below.



(i) Based on the scatter diagram in part (j), a guess for the correlation coefficient between the number convicted and the percentage convicted is $r = 0.5$. Perhaps a "reasonable" guess should be anywhere from $r = 0.3$ to $r = 0.7$.

(j) The scatter plot for the number convicted versus the percentage convicted is shown below.



(k) The sample correlation coefficient between the number convicted and the percentage convicted is $r = 0.510$.

10. No. It means that there is an association between marriage and one's lifespan not that marriage causes a longer lifespan. Other explanations would be the type of diet, exercise, or any other factors that may influence a longer lifespan and how interested a person is in marrying. In fact, causation might occur in the opposite direction. Perhaps healthy individuals are more likely to marry than unhealthy individuals.

12. (a) Mean = 15%

(b) No, because the states have differing populations.

14. Not necessarily; those people having back problems may be trying to improve their posture.

Chapter 4 PROBABILITY

4.2 SAMPLE SPACE AND EVENTS OF AN EXPERIMENT

Problems

2. This experiment is without replacement.

Let the event of a red ball be defined as R.

Let the event of a blue ball be defined by B.

Let the event of a yellow ball be defined by Y.

(a) The sample space S for this experiment is

$S = \{(R, B), (R, Y), (B, R), (B, Y), (Y, R), (Y, B)\}$

(b) Let A be the event that the first ball drawn is yellow. Then,

$A = \{(Y, B), (Y, R)\}$

(c) Let B be the event that the same ball is drawn twice. Then, $B = \emptyset$. That is, this event will be a null event since the experiment is without replacement.

4. Let H be the event of a head on any toss of the coin; T be the event of a tail on any toss of the coin.

(a) The sample space S of this experiment is

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

(b) Let A be the event that a tail occurs more often than a head, then

$A = \{HTT, THT, TTH, TTT\}$

6. Let N be the event that the New York Yankees win and C be the event that the Chicago White Sox win.

The sample space S for the three games is

$S = \{NNN, NNC, NCN, NCC, CNN, CNC, CCN, CCC\}$

Since A is the event that the Yankees win more games than the Sox, then

$A = \{NNN, NNC, NCN, CNN\}$

8. (a) Since the outcomes of this experiment will consist of one course from each category, then the sample space is

$S =$ {(chicken, pasta, ice cream), (chicken, pasta, gelatin), (chicken, pasta, apple pie), (chicken, rice, ice cream), (chicken, rice, gelatin), (chicken, rice, apple pie), (chicken, potatoes, ice cream), (chicken, potatoes, gelatin), (chicken, potatoes, apple pie), (roast beef, pasta, ice cream), (roast beef, pasta, gelatin), (roast beef, pasta, apple pie), (roast beef, rice, ice cream), (roast beef, rice, gelatin), (roast beef, rice, apple pie), (roast beef, potatoes, ice cream), (roast beef, potatoes, gelatin), (roast beef, potatoes, apple pie)}

(b) Let A be the event that ice cream is chosen, then

$A =$ {(chicken, pasta, ice cream), (chicken, rice, ice cream), (chicken, potatoes, ice cream), (roast beef, pasta, ice cream), (roast beef, rice, ice cream), (roast beef, potatoes, ice cream)}

(c) Let B be the event that chicken is chosen, then

$B =$ {(chicken, pasta, ice cream), (chicken, pasta, gelatin), (chicken, pasta, apple pie), (chicken, rice, ice cream), (chicken, rice, gelatin), (chicken, rice, apple pie), (chicken, potatoes, ice cream), (chicken, potatoes, gelatin), (chicken, potatoes, apple pie)}

(d) $A \cap B =$ {(chicken, pasta, ice cream), (chicken, rice, ice cream), (chicken, potatoes, ice cream)}

(e) Let C be the event that rice is chosen, then

$C =$ {(chicken, rice, ice cream), (chicken, rice, gelatin), (chicken, rice, apple pie), (roast beef, rice, ice cream), (roast beef, rice, gelatin), (roast beef, rice, apple pie)}

(f) $A \cap B \cap C =$ {(chicken, rice, ice cream)}

10. (a) E and F are disjoint events.

(b) E and F are disjoint events.

(c) E and F are not disjoint events.

(d) E and F are not disjoint events.

(e) E and F are disjoint events.

12. The sample space for the rolling of two dice (die 1 and die 2) is shown below with the outcomes given as ordered pairs.

Let A be the event that the sum of the faces is even.

Let B be the event that first die (die 1) lands on 1.

Let C be the event that the sum of the faces is 6.

Die 1\Die2	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

(a) $A \cap B = \{(1, 1), (1, 3), (1, 5)\}$

(b) $A \cup B = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 2), (2, 4), (2, 6), (3, 1),$
$(3,3), (3,5), (4,2), (4,4), (4,6), (5,1), (5,3), (5,5)(6,2), (6,4), (6,6)\}$

(c) $B \cap C = \{(1, 5)\}$

(d) $B^c = \{(2, 1), (2, 2), (2, 3), \ldots, (6, 4), (6, 5), (6, 6)\}$

(e) $A^c \cap C = \emptyset$

(f) $A \cap B \cap C = \{(1, 5)\}$

4.3 PROPERTIES OF PROBABILITY

Problems

2. $P(A) = 0.2$, $P(B) = 0.5$, and A and B are disjoint events.

(a) $P(A^c) = 1 - P(A) = 1 - 0.2 = 0.8$

(b) $P(A \cup B) = P(A) + P(B)$ (since the events are disjoint)
$= 0.2 + 0.5 = 0.7$

(c) $P(A \cap B) = P(\text{null set})$ (since the events are disjoint)
$= 1 - P(S) = 1 - 1 = 0$

(d) $P(A^c \cap B) = P(B) = 0.5$ (***Hint:*** Draw a Venn diagram.)

4. (a) Let A be the event of stopping for at least 1 red light. Then,

$P(A) = P(1 \text{ red light}) + P(2 \text{ red light}) + P(3 \text{ red light})$

$= 0.36 + 0.34 + 0.16 = 0.86$

Alternatively, $P(A) = 1 - P(\text{stopping for zero red lights})$

$= 1 - 0.14 = 0.86$

(b) Let B be the event of stopping for more than 2 red lights. Then,

$P(B) = P(3 \text{ red lights}) = 0.16$

6. Let A be the event of drawing a king and B be the event of drawing an ace.
$P(A) = 1/6$ and $P(B) = 1/6$

Since A and B are disjoint events, then

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) = 1/6 + 1/6 = 1/3 = 0.3333$

8. (a) $P(\text{2 or fewer defects}) = P(\text{0 defect or 1 defect or 2 defects})$
$= P(\text{0 defect}) + P(\text{1 defect}) + P(\text{2 defects})$
$= 0.12 + 0.18 + 0.25 = 0.55$ (disjoint events)

(b) $\text{P(4 or more defects)} = \text{P\{4 defects or (5 or more defects)\}}$
$= P(\text{4 defects}) + P(\text{5 or more defects})$
$= 0.15 + 0.10 = 0.25$

(c) $P(\text{between 1 and 3 defects inclusive})$
$= P(\text{1 defect or 2 defects or 3 defects})$
$= P(\text{1 defect}) + P(\text{2 defects}) + P(\text{3 defects})$
$= 0.18 + 0.25 + 0.20 = 0.63$

(d) Largest value of p is $0.12 + 0.25 + 0.15 + 0.1 = 0.62$

(e) Smallest value of p is $0.12 + 0.25 + 0.15 = 0.52$

10. (a) $P(\text{newborn will die between ages 30 and 60})$
$= 0.033 + 0.063 + 0.124 = 0.220$

(b) $P(\text{newborn will not survive to age 40}) = P(\text{dying before age 40})$
$= 0.062 + 0.012 + 0.024 + 0.033 = 0.131$

(c) $P(\text{newborn will survive to age 80}) = 1 - P(\text{living beyond age 80})$
$= 1 - \{P(\text{living in decade 9 or decade 10})\}$
$= 1 - \{P(\text{living in decade 9}) + P(\text{living in decade 10})\}$
$= 1 - (0.168 + 0.028) = 0.804$

12. Let A be the event that the customer has an American Express card.
Let B be the event that the customer has a Visa card.
$P(A) = 0.22, P(B) = 0.58, P(A \cap B) = 0.14$. Since,
$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.22 + 0.58 - 0.14 = 0.66$, then

$P(\text{customer has neither an American Express nor a Visa Card})$
$= 1 - P(\text{customer has at least one of the credit cards})$
$= 1 - 0.66 = 0.34$

14. $P(A) = 0.064, P(B) = 0.043, P(A \cap B) = 0.025$.

(a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.064 + 0.043 - 0.025 = 0.082$

(b) $P(\text{welds have neither defect}) = 1 - P(\text{welds have at least one defect})$
$= 1 - P(A \cup B) = 1 - 0.082 = 0.918$

16. Let A be the event of receiving an A in statistics.

Let B be the event of receiving an A in physics.

$P(A) = 0.4, P(B) = 0.6, P(A \cup B) = 0.86$

(a) P(does not receive an A in either statistics or physics)
$= 1 - P$(receive an A in at least one of the courses)
$= 1 - P(A \cup B) = 1 - 0.86 = 0.14$

(b) Since $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, then
$0.86 = 0.4 + 0.6 - P(A \cap B)$, so $P(A \cap B) = 1 - 0.86 = 0.14$

4.4 EXPERIMENTS HAVING EQUALLY LIKELY OUTCOMES

Problems

2. (a) Let A be the event that the person has gained weight. Then,

$$P(A) = \frac{5}{32} = 0.1563$$

(b) Let B be the event that the person has lost weight. Then,

$$P(B) = \frac{18}{32} = 0.5625$$

(c) Let C be the event that the person neither lost nor gained weight. Then,

$$P(C) = \frac{9}{32} = 0.2813$$

4. (a) Let A be the event that the chairperson is selected from a country whose meat production exceeds 10,000. Then,

$$P(A) = \frac{3}{10}$$

(b) Let B be the event that the chairperson is selected from a country whose meat production is under 3,500. Then,

$$P(B) = \frac{5}{10} = \frac{1}{2}$$

(c) Let C be the event that the chairperson is selected from a country whose meat production is between 4,000 and 6,000. Then,

$$P(C) = \frac{0}{10} = 0$$

(d) Let D be the event that the chairperson is selected from a country whose meat production is less than 2,000. Then,

$$P(D) = \frac{0}{10} = 0$$

6. Let x be the number of pennies in the bag. Since the bag contains 4 times as many dimes as pennies, the number of dimes will be $4x$. Thus, the total number of coins in the bag is $x + 4x = 5x$. Thus,

$P(\text{dime being drawn}) = (4x)/(5x) = 4/5 = 0.8$

8. Let A be the event of owning a cat.

Let B be the event of owning a dog.

Given: $P(A) = 0.20$; $P(B) = 0.32$; $P(A \cap B) = 0.12$.

(a) $P(\text{family has neither a cat nor a dog}) = [P(A \cup B)]^c = 1 - P(A \cup B)$
Now, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, thus
$P(A \cup B) = 0.2 + 0.32 - 0.12 = 0.4$, and so
$P(\text{family has neither a cat nor a dog}) = [P(A \cup B)]^c = 1 - 0.4 = 0.6$

(b) Number of families who have neither a cat nor a dog
$= 0.6 \times 1000 = 600$

10. Let A be the event that a club member plays tennis.

Let B be the event that a club member plays squash.

Given: $P(A) = 44/120$; $P(B) = 30/120$; $P(A \cap B) = 18/120$.

(a) $P(\text{member does not play tennis}) = 1 - P(A) = 1 - 44/120$
$= 76/120 = 19/30 = 0.6333$

(b) $P(\text{member does not play squash}) = 1 - P(B) = 1 - 30/120$
$= 90/120 = 3/4 = 0.75$

(c) $P(\text{member plays neither tennis nor squash}) = 1 - P(A \cup B)$.
Now, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, so
$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 44/120 + 30/120 - 18/120$
$= 56/120 = 0.4667$. Thus,
$P(\text{member plays neither tennis nor squash}) = 1 - P(A \cup B)$
$= 1 - 56/120 = 1 - 0.4667 = 0.5333$

12. Note the number of sample points is 36.

Die 1\Die 2	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

(a) $P(\text{sum is either 7 or 11}) = 6/36 + 2/36 = 8/36 = 0.2222$

Note: The event of a sum of 7 and the event of a sum of 11 are mutually exclusive.

(b) $P(\text{sum is 2, 3, or 12}) = 1/36 + 2/36 + 1/36 = 4/36 = 1/9$

(c) $P(\text{sum is an even number}) = 18/36 = 1/2$

14. (a) $P(\text{worker earns under \$ 15,000})$
$= (427 + 440 + 1{,}274 + 548 + 358 + 889)/(31{,}340 + 49{,}678)$
$= 0.0486$

(b) $P(\text{female who earns between \$20,000 and \$40,000})$
$= (6{,}291 + 6{,}555)/(31{,}340 + 49{,}678) = 0.1586$

(c) $P(\text{worker earns under \$50,000})$
$= 1 - (8{,}255 + 947 + 20{,}984 + 7{,}377)/(31{,}340 + 49{,}678) = 0.5364$

16. (a) $P(\text{person in the second position is a boy}) = 4/9$

(b) $P(\text{Charles is in the second position}) = 1/9$

4.5 CONDITIONAL PROBABILITY AND INDEPENDENCE

Problems

2. Let H represent the outcome of a head and T represent the outcome of a tail. The sample space $S = \{HH,\ HT,\ TH,\ TT\}$.

Let A represent the event of a head on both tosses. So, $A = \{HH\}$.

Let B be the event of head on the first toss. So, $B = \{HH,\ HT\}$.

Then, $P(A|B) = [P(A \cap B)]/P(B) = P(A)/P(B)$ since $A \cap B = A$. Thus,

$P(A|B) = (1/4)/(2/4) = 1/2$

4. Let A be the event of a female student.

Let B be the event of student majoring in computer science.

Given: $P(A) = 0.52;\ P(B) = 0.05;\ P(A \cap B) = 0.02$.

(a) $P(A|B) = [P(A \cap B)]/P(B) = 0.02/0.05 = 0.4$

(b) $P(B|A) = [P(A \cap B)]/P(A) = 0.02/0.52 = 0.0385$

6. (a) Let A be the event that the resident is between 10 and 20 years old.
Let B be the event that the resident is less than 30 years old.
$P(A|B) = [P(A \cap B)]/P(B) = 5,100/(4{,}200 + 5{,}100 + 6{,}200) = 0.3290$

(b) Let C be the event that the resident is between 30 and 40 years old.
Let D be the event that the resident is older than 30 years.
$P(C|D) = [P(C \cap D)]/P(D)$
$= 4{,}400/(4{,}400 + 3{,}600 + 2{,}500 + 1{,}800 + 1{,}100) = 0.3284$

8. Let A be the event of a male student.

Let B be the event of a female student.

Let C be the event that the student is less than 25 years old.

(a) $P(C|A) = [P(A \cap C)]/P(A) = (91 + 1{,}309 + 1{,}089 + 1{,}080)/5,882$
$= 0.6068$

(b) $P(A|C) = [P(A \cap C)]/P(C) = (91 + 1{,}309 + 1{,}089 + 1{,}080)/(91+$
$1{,}309 + 1{,}089 + 1{,}080 + 119 + 1{,}455 + 1{,}135 + 968)$
$= 0.4925$

(c) $P(C|B) = [P(B \cap C)]/P(B) = (119 + 1{,}455 + 1{,}135 + 968)/6,663$
$= 0.5519$

(d) $P(B|C) = [P(B \cap C)]/P(C) = (119 + 1{,}455 + 1{,}135 + 968)/(91 + 1{,}309$
$+ 1{,}089 + 1.080 + 119 + 1{,}455 + 1{,}135$
$+ 968) = 0.5075$

10. (a) P(firstborn) $= 167/400 = 0.4175$

(b) P(student rated as confident) $= 122/400 = 0.305$

(c) Let A be the event that the student is rated as being confident.
Let B be the event of a firstborn.
$P(A|B) = [P(A \cap B)]/P(B) = 62/167 = 0.3713$

(d) $P(A|\text{not } B) = 60/233 = 0.2575$

(e) $P(B|A) = [P(A \cap B)]/P(A) = 62/122 = 0.5082$

12. Let R be the event that an eligible voter registered.

Let V be the event that a registered voter actually voted.

Given: $P(R) = 0.683; P(V|R) = 0.599.$

(a) $P(V) = P(R \cap V) = P(R) \times P(V|R) = 0.683 \times 0.599 = 0.409$

(b) $P(R|V^c) = P(R \cap V^c)/P(V^c) = [P(R) - P(R \cap V)]/[1 - P(V)]$
$= P(R) - P(V)/[1 - P(V) = [0.683 - 0.409]/[1 - 0.409] = 0.4635$

14. Let R be the event of a red sock. Let B be the event of blue sock.

Let G be the event of a green sock.

(a) $P(RR) = (5/12)\ (4/11) = 0.1515$
(b) $P(BB) = (4/12)\ (3/11) = 0.0909$
(c) $P(GG) = (3/12)\ (2/11) = 0.0455$
(d) $P(RR \text{ or } BB \text{ or } GG) = P(RR) + P(BB) + P(GG)$
$= 0.1515 + 0.0909 + 0.0455 = 0.2879$

16. Let R be the event of a red sock. Then, $P(RR) = 1/2$. Thus,

$P(RR) = (3/n)\ [2/(n-1)] = 1/2$. Simplifying gives $n(n-1) = 12$ or $n^2 - n - 12 = 0$. Solving this quadratic equation gives $n = 4$ or $n = -3$. The admissible solution is $n = 4$.

18. (a) $P(\text{at least one } 6) = 11/36$

(b) $P(\text{at least one } 6 \mid \text{sum is } 9)$
$= P(\text{at least one 6 and sum is 9})/P(\text{sum is 9}) = 2/4 = 1/2$

(c) $P(\text{at least one } 6 \mid \text{sum is } 10)$
$= P(\text{at least one 6 and sum is 10})/P(\text{sum is 10}) = 2/3$

20. Let A be the event of a plot containing oil.

Let B be the event of hitting oil with the first well.

Given: $P(A) = 0.7$; $P(B|A) = 0.5$. So,

$P(\text{first well hits oil}) = P(A \cap B) = P(B|A) \times P(A) = (0.5)(0.7) = 0.35$

22. Let B be the event of a black ball. Then, $P(BB) = (6/10)\ (7/11) = 0.3818$.

24. Let A be the event that the male student is more than 30 years old.

Let B be the event that the female student is more than 30 years old.

(a) $P(\text{exactly one is more than 30 years old}) = P[(A \cap B^C) \cup (A^C \cap B)]$
$= P(A \cap B^C) + P(A^C \cap B)$. Now,
$P(A) = (613 + 684)/5882 = 1297/5882$ and $P(A^C) = 4585/5882$
$P(B) = (716 + 1339)/6663 = 2055/6663$ and $P(B^C) = 4608/6663$
Thus, $P(A \cap B^C) + P(A^C \cap B) = (1297/5882)(4608/6663)+$
$(4585/5882)\ (2055/6663) = 0.3929$

(b) $P(\text{male is older} \mid \text{exactly one is more than 30 years old})$
$= P(\text{male is older and exactly one is more than 30 years old})/P(\text{exactly one is more than 30 years old}) = P(A \cap B^C)/0.3929 = 0.3881$

26. Let A be the event of the oldest child being a girl.

Let B be the event of the younger child being a boy.

The sample space $S = \{AA,\ AB,\ BA,\ BB\}$.

Now, $P(A|B) = [P(A \cap B)]/P(B) = (1/4)/(2/4) = 1/2$. Also, $P(A) = 1/2$.

Since $P(A|B) = P(A) = 1/2$, this implies that A and B are independent.

28. (a) Independent
(b) Dependent
(c) Independent

30. Let S be the event of a 5. Let F be the event of a 1, 2, 3, 4, or 6.

P(more than 6 throws to observe a 5) $= P$(at least 7 throws are required)

$= P(FFFFFF) = (5/6)^6$

32. Let A be the event of a corner plot from the 9-plot field.

Let B be the event of a corner plot from the 12-plot field.

Given: $P(A) = 4/9; P(B) = 4/12$; A and B are independent.

(a) $P(A \cap B) = P(A) \times P(B) = (4/9)(4/12) = 4/27$
(b) $P(A^C \cap B^C) = (5/9)(8/12) = 10/27$
(c) P(at least one is a corner plot) $= P(A \cup B)$. Now,
$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 4/9 + 4/12 - 4/27 = 17/27$

34. Let A be the event that the sum is 7. Let B, C, D, E, F, and G be the events that the first die lands on 1, 2, 3, 4, 5, and 6, respectively.

$P(A) = 6/36 = 1/6$; $P(B) = P(C) = P(D) = P(E) = P(F) = P(G) = 6/36 = 1/6$

Now, $P(A \cap B) = 1/36 = P(A) \times P(B) = (1/6)(1/6) = 1/36$. Hence, A and B are independent. Similarly, you can show that A is independent of C, D, E, F, and G.

36. Let A be the event of an accidental death in 1988. Then, $P(A) = 0.0478$. Assuming the deaths were independent of each other, then P(3 randomly chosen deaths were all due to accidents) $= P(A)P(A)P(A) = 0.0478^3$.

38. Let A be the event that the first ball drawn is white.

Let B be the event that the second ball drawn is black.

The sample space is $S = \{AA,\ AB,\ BA,\ BB\}$.

(We should assume balls are drawn without replacement because of problem number 39.)

$P(A|B) = [P(A \cap B)]/P(B) = P(AB)/P(B)$. Now, $P(AB) = (1/2)(5/9)$ and

$P(B) = P(AB \text{ or } BB) = (1/2)(5/9) + (1/2)(4/9)$, so $P(A|B) = 5/9$. Also,

$P(A) = P(AA \text{ or } AB) = (1/2)(4/9) + (1/2)(5/9) = 1/2$. Since

$P(A|B) \neq P(A)$, then A and B are not independent if the first ball is not replaced before the second ball is drawn.

40. Let A be the event of a YES response.

Let B be the event of a NO response.

Given: $P(A) = 0.7; P(B) = 0.3.$

(a) $P(A \cap A \cap A \cap A \text{ or } B \cap B \cap B \cap B) = [P(A)]^4 + [P(B)]^4$
$= (0.7)^4 + (0.3)^4 = 0.2482$

(b) $P(B \cap B \cap A \cap A) = (0.3)(0.3)(0.7)(0.7) = 0.0441$

(c) $P(\text{at least 1 NO}) = 1 - P(\text{all 4 YES}) = 1 - (0.7)^4 = 0.7599$

(d) $P(\text{exactly 3 YES}) = P(A \cap A \cap A \cap B) + P(A \cap A \cap B \cap A) +$
$P(A \cap B \cap A \cap A) + P(B \cap A \cap A \cap A)$
$= 4(0.7)^3(0.3) = 0.4116$

(e) $P(\text{at least one YES}) = 1 - P(\text{all 4 NO}) = 1 - P(B \cap B \cap B \cap B)$
$= 1 - (0.3)^4 = 0.9919$

42. Let A be the event of a defective chip from machine A.

Let B be the event of a defective chip from machine B.

Given: $P(A) = 0.10; P(B) = 0.05.$

(a) $P(A \cap B) = P(A) \times P(B) = (0.1)(0.05) = 0.005$

(b) $P(A^C \cap B^C) = P(A^C) \times P(B^C) = (0.9)(0.95) = 0.855$

(c) $P(\text{exactly 1 chip is defective}) = P(A^C \cap B \text{ or } A \cap B^C) = P(A^C \cap B) +$
$P(A \cap B^C) = P(A^C)P(B) + P(A)P(B^C) = (0.9)(0.05) + (0.1)(0.95) = 0.14$

(d) The sample space here is $S = \{AB, A^C B, AB^C, A^C B^C\}$.
$P(\text{a defective chip came from machine } A \mid \text{exactly 1 chip is defective})$
$= P(AB \text{ or } AB^C | A^C B \text{ or } AB^C) = P(AB^C)/[P(A^C B) + P(AB^C)]$
$= [(0.1)(0.95)]/[(0.9)(0.05) + (0.1)(0.95)] = 0.6786$

(e) $P(\text{a defective chip came from machine B} \mid \text{exactly 1 chip is defective})$
$= 0.3214$ (Solve similarly as in part (d))

4.6 BAYES' THEOREM

Problems

2. Let G be the event of guessing the answer. Let K be the event of knowing the answer. Let C be the event of a correct response.

Given: $P(C|G) = 1/5 = 0.2;\ P(C|K) = 1;\ P(K) = 0.6;\ P(G) = 0.4.$

$P(K|C) = [P(C|K)P(K)]/[P(C|K)P(K) + P(C|G)P(G)] =$

$[(1)(0.6)]/[(1)(0.6) + (0.2)(0.4)] = 0.8824$

4. Let $R1$ be the event that the ball drawn from urn I is red. Let $R2$ be the event that the ball drawn from urn II is red. Similarly, define $B1$ and $B2$ for blue ball.

(a) $P(R2) = P(R2|R1)P(R1) + P(R2|B1)P(B1) = (3/5)(4/7) + (2/5)(3/7)$
$= 18/35 = 0.5143$

(b) $P(R1|B2) = [P(B2|R1)P(R1)]/[P(B2|R1)P(R1) + P(B2|B1)P(B1)]$
$= [(2/5)(4/7)]/[(2/5)(4/7) + (3/5)(3/7)] = 8/17 = 0.4706$

6. Let R represent red and B black. Thus, the three cards may be denoted by RR, BB, and RB. We need to determine $P(RB|R)$. Now,

$P(RB|R) = P(RB \cap R)/P(R) = P(RB \cap R)/[P(RB \cap R) + P(RR \cap R)] =$

$[P(R|R)P(RB)]/[P(R|RB)P(RB) + P(R|RR)P(RR)]$

$= [(1/2)(1/3)]/[(1/2)(1/3) + (1)(1/3)] = 1/3$

8. (a) Since Susan has blue eyes and both her parents have brown eyes, the eye gene pair for the mother and father has to be blue and brown.

(b) ***Note:*** *Since the husband has blue eyes, he must have two eye genes that are blue. Thus, the baby definitely will receive one blue gene from the father. So this problem is equivalent to finding the probability that Susan's sister has a blue-eyed gene given she has brown eyes.*

Since Susan's sister has brown eyes, then she must have received a blue-eyed gene from her mother and a brown-eyed gene from her father; or a brown-eyed gene from both parents; or a brown-eyed gene from her mother and a blue-eyed gene from her father.

So, P(Susan's sister having a brown eyes) $= P$(blue from mother and brown from father) $+ P$(brown from both) $+ P$(blue from father and brown from mother) $= (1/2)(1/2) + (1/2)(1/2) + (1/2)(1/2) = 3/4$.
Also, P(Susan's sister having a blue-eyed gene and having brown eyes) $= P$(blue-eyed gene from her mother and a brown-eyed gene from her father or a brown-eyed gene from her mother and a blue-eyed gene from her father) $= (1/2 \times 1/2) + (1/2 \times 1/2) = 2/4 = 1/2$.
Thus, P(Susan's sister has a blue-eyed gene | Susan's sister has brown eyes) $= (1/2)/(3/4) = 2/3$.

4.7 COUNTING PRINCIPLES

Problems

2. $9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 9! = 362{,}880$ batting orders

4. $36 \times 36 \times 36 = (36)^3 = 46,656$ possibilities

6. The traveler met $7^4 = 2,401$ kittens.

8. $\binom{n}{0} = n!/[0!(n-0)!] = n!/[0!n!] = 1$. This answer makes sense since there is only one way of selecting zero objects out of n.

10. Number of hand shakes $= 19 \times 18 \times 17 \times 16 \ldots 5 \times 4 \times 3 \times 2 \times 1 = 19!$

12. $P(\text{none of the trucks with faulty brakes}) = \dfrac{\binom{7}{2}\binom{3}{0}}{\binom{10}{2}} = 0.4667$

14. (a) There will be $\binom{40}{8}$ groups of numbers of which only one group is the winning set. Thus, $P(\text{player has all 8 numbers}) = 1/\binom{40}{8}$.

(b) Number of ways of getting 7 of the selected numbers is $\binom{8}{7}\binom{32}{1}$

Hence, $P(\text{getting 7 of the selected numbers}) = \dfrac{\binom{8}{7}\binom{32}{1}}{\binom{40}{8}}$

(c) $P(\text{getting at least 6 of the numbers}) = \dfrac{\binom{8}{6}\binom{32}{2}+\binom{8}{7}\binom{32}{1}+\binom{8}{8}\binom{32}{0}}{\binom{40}{8}} =$ 0.00018

16. $P(\text{randomly selected bridge hand is a Yarborough}) = \dfrac{\binom{36}{13}\binom{16}{0}}{\binom{52}{13}} = 0.0036$

18. Number of paths from A to $B = \binom{7}{3} = \binom{7}{4} = 35$

Review Problems

2. Let A be the event of making a foul shot. Let B be the event of missing.

Given: $P(A) = 0.8; P(B) = 0.2$.

(a) $P(\text{making both shots}) = P(A \cap A) = P(A)P(A) = (0.8)(0.8) = 0.64$

(b) $P(\text{missing both shots}) = P(B \cap B) = P(B)P(B) = (0.2)(0.2) = 0.04$

(c) P(making 2nd shot|missing the 1st) $=$ P(making 2nd shot and missing 1st)/P(missing 1st)

Now, the possible outcomes are $S = \{AA,\ AB,\ BA,\ BB\}$. So, P(making 2nd shot | missing the 1st) $= [P(AB)]/[P(BA) + P(BB)]$

$= [(0.2)(0.8)]/[(0.2)(0.8) + (0.2)(0.2)] = 0.8$

Another approach to the problem is to use the property of independence. Note that the probability of making the 2nd shot conditional that she missed the 1st shot is the same as the (unconditional) probability of making the 2nd shot, and this probability is $P(A) = 0.8$.

4. Let F be the event of a registered female voter. Let M be the event of a male registered voter. Let V be the event that a person voted in the last election.

Given: $P(F) = 0.54; P(M) = 0.46; P(V|F) = 0.68; P(V|M) = 0.62;$

$P(V^C|M) = 0.38;\ P(V^C|F) = 0.32.$

(a) P(female who voted) $= P(F \cap V) = P(V|F)P(F) = (0.68)(0.54)$
$= 0.3672$

(b) P(male who did not vote) $= P(M \cap V^C) = P(V^C|M)P(M)$
$= (0.38)(0.46) = 0.1748$

(c) $P(M|V) = [P(V|M)P(M)]/[P(V|M)P(M) + P(V|F)P(F)]$
$= [(0.62)(0.46)]/[(0.62)(0.46) + (0.68)(0.54)] = 0.4372$

6. Assume selection without replacement.

(a) P(both are aces) $= P$(ace on 1st draw and ace on 2nd draw)
P(ace on 1st) P(ace on 2nd|ace on 1st) $= (4/52)(3/51) = 0.0045$

(b) P(both spades) $= P$(spade on the 1st draw and spade on 2nd draw)
P(spade on first) P(spade on 2nd|spade on 1st) $= (13/52)(12/51) = 0.0588$

(c) ***Note:*** If the 1st card selected is a spade, then the next selection should be a club, or a diamond, or a heart for different suits. Similarly, you can analyze if you first selected a club, or a diamond or a heart. Thus,

P(different suits) $= 3(13/52)(13/51) + 3(13/52)(13/51) +$

$3(13/52)(13/51) + 3(13/52)(13/51) = 0.7647$

Alternatively, one can determine $[1 - P$(the two cards are from the same suit)$]$
$= 1 - 4(13/52)(12/51) = 0.7647$

(d) You can use a similar argument as in part (c) to verify
P(different denominations) $= (4/52)(4/51)(12)(13) = 0.9412$
Alternatively, one can determine $[1 - P$(the two cards are from the same denomination)$] = 1 - 13(4/52)(3/51) = 0.9412$

8. (a) $P(2 \text{ questions}) = (1/2)^2 = 1/4$

(b) $P(3 \text{ questions}) = (1/2)^3$

(c) $P(10 \text{ questions}) = (1/2)^{10}$

10. Let A be the event of a male. Let B be the event of a person being under 25.

(a) $P(A) = (117)/(240.3) = 0.4869; P(A^C) = 1 - 0.4869 = 0.5131$

(b) $P(B) = (99.2)/(240.3) = 0.4128; P(B^C) = 1 - 0.4128 = 0.5872$

(c) $P(A \cap B) = (50.4)/(240.3) = 0.2097$

(d) $P(A \cap B^C) = (66.6)/(240.3) = 0.2772$

(e) $P(A|B) = [PA \cap B]/P(B) = (0.2097)/(0.4128) = 0.5081$

(f) $P(B|A) = [PA \cap B]/P(A) = (0.2097)/(0.4869) = 0.4308$

12. Let A be the event of an ace. Let B be the event of a face card.

$P(\text{blackjack}) = P(A \cap B \text{ or } B \cap A) = P(A \cap B) + P(B \cap A)$

$= (4/52)(16/51) + (16/52(4/51) = 32/663 = 0.0483$

14. $P(A) = 0.8;\ P(A^C) = 0.2;\ P(B) = 0.6;\ P(B^C) = 0.4.$

(a) $P(A \cap B) = P(A)P(B) = (0.8)(0.6) = 0.48$

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.6 - 0.48 = 0.92$

(c) $P(B) = 0.6$. (d) $P(A^C \cap B) = P(A^C)P(B) = (0.2) + (0.6) = 0.12$

16. Let M be the event of the detection of a nondefective disk in the manual inspection. Let E be the event of the detection of a nondefective disk in the electronic inspection.

Given: $P(M) = 0.3;\ P(E|M) = 0.2$. So,

$P(M \cap E) = P(E|M)P(M) = (0.3)(0.2) = 0.06$

18. Let AF be the event that a child receives a blue gene from the father.
Let BF be the event that a child receives a brown gene from the father.
Let AM be the event that a child receives a blue gene from the mother.
Let BM be the event that a child receives a brown gene from the mother.

Given: $P(AF) = P(BF) = P(AM) = P(BM) = 1/2$.

(a) $P(\text{oldest child has blue eyes})\ P(AF \text{ and } AM) = P(AF)P(AM)$
$= (1/2)(1/2) = 1/4$

(b) $P(\text{oldest has blue and the youngest has brown}) =$
$P(\text{oldest has blue})\ P(\text{youngest has brown}) =$
$P(AF)P(AM)[P(AF)P(BM) + P(BF)P(AM) + P(BF)P(BM)]$
$= (1/4)(3/4) = 3/16$

(c) P(oldest has brown and the youngest has blue) =
P(oldest has brown) P(youngest has blue) =
$[P(AF)P(BM) + P(BF)P(AM) + P(BF)P(BM)]P(AF)P(AM)$
$= (3/4)(1/4) = 3/16$

(d) P(one child has blue eyes and the other has brown) =
$(b) + (c) = 3/16 + 3/16 = 6/16 = 3/8$

(e) P(both have blue eyes) = P(oldest has blue eyes and the youngest has blue eyes)
= P(oldest has blue eyes)P(youngest has blue eyes)
$= (1/4)(1/4) = 1/16$

(f) P(both have brown eyes) = P(oldest has brown eyes and the youngest has brown eyes)
= P(oldest has brown eyes)P(youngest has brown eyes)
$= (3/4)(3/4) = 9/16$

20. Let A be the event of an ace. Let B be the event of a spade.

(a) $P(A) = 4/52 = 1/13; P(B) = 13/52 = 1/4$.
Now, $P(A|B) = P(A \cap B)/P(B) = P(\text{ace of spades})/P(B) =$
$(1/52)/(1/4) = 4/52 = 1/13 = P(A)$. Thus, A and B are independent.

(b) $P(A) = 3/39 = 1/13; P(B) = 13/39 = 1/3$.
Now, $P(A|B) = P(A \cap B)/P(B) = P(\text{ace of spades})/P(B) =$
$(1/39)/(1/3) = 3/39 = 1/13 = P(A)$. Thus, A and B are independent.

(c) $P(A) = 4/44 = 1/11; P(B) = 13/44$.
Now, $P(A|B) = P(A \cap B)/P(B) = P(\text{ace of spades})/P(B) =$
$(1/44)/(13/44) = 1/13 \neq P(A)$. Thus, A and B are not independent.

22. Let A be the event that the battery lasts beyond 10,000 miles. Let B be the event that the battery lasts beyond 20,000 miles. Let C be the event that the battery lasts beyond 30,000 miles.

Given: $P(A) = 0.8; P(B) = 0.4; P(C) = 0.1$.

(a) $P(B|A) = P(A \cap B)/P(A) = P(B)/P(A) = 0.4/0.8 = 1/2$

(b) $P(C|A) = P(A \cap C)/P(A) = P(C)/P(A) = 0.1/0.8 = 1/8$

24. (a) P(both attended a jazz performance) $= (0.14)(0.1) = 0.014$

(b) P(exactly 1 attended a jazz performance)
$= (0.14)(0.9) + (0.86)(0.1) = 0.212$

(c) P(younger attended the jazz performance|exactly 1 attended)
= P (younger attended and the older did not attend)/P(exactly 1 attended)
$= 0.14 \times 0.9/0.212 = 0.5943$

26. No. We have to assume that the two events are independent. Under this assumption

P(attending a jazz and a classical musical performance)

$= (0.1)(0.13) = 0.013$

28. 1/3, 1/2

30. Let C be the event that the woman has breast cancer, and let M be the event she has a positive mammography. Then

$$P(C|M) = \frac{P(M|C)P(C)}{P(M|C)P(C) + P(M|C^c)P(C^c)} = \frac{.9(.02)}{.9(.02) + .1(.98)} = .155$$

Chapter 5 DISCRETE RANDOM VARIABLES

5.2 RANDOM VARIABLES

Problems

2. The sample space $S = \{(b,b,b),\ (b,b,g),\ (b,g,b),\ (g,b,b),\ (b,g,g),\ (g,b,g),$ $(g,g,b), (g,g,g)\}$. Since W = number of girls that came before the first boy, then the probability distribution for W is

i	$P\{W = i\}$
0	4/8
1	2/8
2	1/8
3	1/8
	Total $= 1$

4. Let X = sum on the 2 faces on the pair of dice. The possible values of X are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12. The probability distribution for X is shown below.

i	2	3	4	5	6	7	8	9	10	11	12
$P\{X = i\}$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

6. Let X = the amount of time the first one to arrive has to wait.

(a) The possible values of X are 0, 30, and 60 minutes.
(b) Note that $X = 0$ can occur 3 times, $X = 30$ can occur 2 times, and $X = 60$ can occur 2 times. Thus, $P\{X = 0\} = 3/9 = 1/3$, $P\{X = 30\} = 4/9$, and $P\{X = 60\} = 2/9$.

8. Let X = number of defective batteries chosen. The possible values of X are 0, 1, and 2. The probability distribution for X is shown below.

i	0	1	2
$P\{X = i\}$	7/15	7/15	1/15

Note: You can compute the probabilities by following this logic – for instance, to compute $P\{X = 0\} = P(2 \text{ nondefective}) = (7/10)(6/9)$, etc.

10. Let X = number of jobs the contractor wins. The possible values of X are 0, 1, and 2. The probability distribution for X is shown below.

i	0	1	2
$P\{X = i\}$	0.3	0.6	0.1

Note: You can compute the probabilities by following this logic – for instance, to compute $P\{X = 0\} = P(\text{losing the 1st and losing the 2nd}) = (0.5)(0.6) = 0.3$, etc.

12. No. The sum of the probabilities is 1.1. The sum of the probabilities must equal 1.

14. Yes. $P(i)$ all lie between 0 and 1 and the sum of the probabilities equals 1.

16. Let Y = number of children in the family of the selected child. The possible values for Y are 1, 2, 3, 4, and 5. The probability distribution for Y is shown below.

i	1	2	3	4	5
$P\{Y = i\}$	82/185	57/185	34/185	10/185	2/185

18. Let X = total amount paid. The possible values for X are 0, 1, and 2 in \$100,000. The probability distribution for X is shown below.

i	0	1	2
$P\{X = i\}$	0.855	0.14	0.005

5.3 EXPECTED VALUE

Problems

2. (a) $E[X] = (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.2) + (5)(0.1) = 2.9$
(b) $E[X] = (1)(0.3) + (2)(0.1) + (3)(0.2) + (4)(0.1) + (5)(0.3) = 3$
(c) $E[X] = (1)(0.2) + (2)(0) + (3)(0.6) + (4)(0) + (5)(0.2) = 3$
(d) $E[X] = (3)(1) = 3$

4. (a) Straight fee \$1,200. E[contingency fee] = \$5,000(1/2) = \$2,500. Thus, the contingency fee will result in a higher than expected fee.
(b) Straight fee \$1,200. E[contingency fee] = \$5,000(1/3) = \$1,666.67. Thus, the contingency fee will result in a higher than expected fee.
(c) Straight fee \$1,200. E[contingency fee] = \$5,000(1/4) = \$1,250. Thus, the contingency fee will result in a higher than expected fee.
(d) Straight fee \$1,200. E[contingency fee] = \$5,000(1/5) = \$1,000. Thus, the contingency fee will result in a lower than expected fee.

6. (a) $E[X] = (1)(1/2) + (2)(1/2) = 1.5$
(b) $E[X] = (1)(1/3) + (2)(1/3) + (3)(1/3) = 2$
(c) $E[X] = (1)(1/4) + (2)(1/4) + (3)(1/4) + (4)(1/4) = 5/2 = 2.5$

(d) $E[X] = \frac{\sum_{i=1}^{n} i}{n} = [n(n+1)/2]/n = (n+1)/2$. ***Note:*** $\sum_{i=1}^{n} i = n(n+1)/2$ as given in part (e) of the problem in the text.

(e) Replacing the numerator in part (d) with $n(n+1)/2$ gives $E[X] = [n(n+1)/2]/n = (n+1)/2$. For example, if $n = 2$, then $E[X] = (2+1)/2 = 1.5$ as in part (a).

8. Yes. The expected loss if the firm suffers a cutoff in power when the computers are being used is $(\$1{,}200)(0.25) = \$300$. Since this is less than \$400 and the firm wants to minimize the expected value of their loss, the firm should risk using their computers.

10. One test will be required if all 4 donors blood are acceptable. $P(\text{all 4 are acceptable}) = (0.9)^4 = 0.6561$, so the probability of having to do tests individually is $(1-0.6561) = 0.3439$. Thus, $E[\text{number of tests}] = (1)(0.6561) + (5)(0.3439) = 2.3756$.

12. Let X = number of games played. Then, $E[X] = (4)(1/8) + (5)(1/4) + (6)(5/16) + (7)(5/16) = 93/16 = 5.8125$.

14. $E[\text{earnings}] = (\$200)(1/4) + (\$300)(3/4) = \$275$

16. $E[\text{gain}] = (\$30{,}000)(0.4) - (\$15{,}000)(0.6) = \$3{,}000$

18. $E[\text{gain}] = (35)(1/38) - (1)(37/38) = -2/38 \neq 0$. This is not a fair bet.

20. $E[\text{gain}] = (1)(18/38) - (1)(20/38) = -2/38 = -0.0526$

22. Let F represent player I and S represent player II. The possible ways of winning is represented in $S = \{FF, FSF, SFF, SS, SFS, FSS\}$. Thus,

(a) $$E[\text{number of sets}] = 2(1/3)(1/3) + 3(2)(1/3)(1/3)(2/3) + 2(2/3)(2/3) + 3(2)(1/3)(2/3)(2/3) = 22/9 = 2.44$$

(b) $$P(\text{player I wins}) = P(FF \text{ or } FSF \text{ or } SFF) = P(FF) + P(FSF) + P(SFF) = (1/3)(1/3) + 2(1/3)(1/3)(2/3) = 1/9 + 4/27 = 7/27$$

24. (a) $$E[\text{sum of bonuses}] = E[X+Y] = 1500 + E[0.8X] = 1500 + 0.8(1500) = \$2{,}700$$

(b) $$E[\text{sum of bonuses}] = E[X+Y] = 1500 + E[X+1000] = 1500 + E[X] + 1000 = 1500 + 1500 + 1000 = \$4{,}000$$

26. (a) $$E[X] = (40)(40/148) + (33)(33/148) + (50)(50/148) + (25)(25/148) = 39.2838$$

$$E[Y] = (40)(1/4) + (33)(1/4) + (50)(1/4) + (25)(1/4) = 37$$

(b) The probabilities (weights) associated with the random variable X vary. Thus, the values of X have different weights unlike the weights associated with Y. The weights for X are largest for the most populated buses.

28. (a) $E[\text{profit}] = (1,200)(\$14) = \$16{,}800$

(b) $E[\text{profit}] = (\$14 \times 1200 - \$8 \times 300)(0.5) + (\$14 \times 1500)(0.2) + (\$14 \times 1500)(0.3) = \$17{,}700$

(c) $E[\text{profit}] = (\$14 \times 1200 - \$8 \times 600)(0.5) + (\$14 \times 1500 - \$8 \times 300)(0.2) + (\$14 \times 1800)(0.3) = \$17{,}280$

30. $E[X] = 5$ and $E[Y] = 12$, thus

(a) $E[3X + 4Y] = E[3X] + E[4Y] = 3E[X] + 4E[Y] = 3(5) + 4(12) = 63$
(b) $E[2 + 5Y + X] = E[2] + 5E[Y] + E[X] = 2 + 5(12) + 5 = 67$
(c) $E[4 + Y] = E[4] + E[Y] = 4 + 12 = 16$

32. Let X = the year end bonus for the husband, and Y = the year end bonus for the wife. Then,

$E[X] = (0)(0.3) + (\$1,000)(0.6) + (\$2{,}000)(0.1) = \$800$ and
$E[Y] = (\$1{,}000)(0.7) + (\$2{,}000)(0.3) = \$1{,}300$
Now, the sum of their bonuses $S = X + Y$. Thus, $E[S] = E[X + Y] =$

$E[X] + E[Y] = \$800 + \$1,300 = \$2{,}100$.

34. $E[\text{number of closed or assisted banks}] = (3)[(8)(1/8) + (6)(1/8) + \cdots + (4)(1/8) + (11)(1/8)] = 18$

36. Let W = number of defective batteries. Then, $P(W = 0) = 28/45$, $P(W = 1) = 16/45$, and $P(W = 2) = 1/45$. So,

(a) $E[W] = (0)(28/45) + (1)(16/45) + (2)(1/45) = 0.4$ batteries
Let D be the event of defective battery and N be the event of a nondefective battery. The sample space is $X = \{DD, DN, ND, NN\}$. So from the problem, $P(X = 0) = P\{ND \text{ or } NN\} = 72/90$ and $P(X = 1) = P\{DD \text{ or } DN\} = 18/90$. Also, $P(Y = 0) = 72/90$ and $P(Y = 1) = 18/90$.
(b) We can write $W = X + Y$.
(c) $E[W] = E[X + Y] = E[X] + E[Y] = 18/90 + 18/90 = 0.4$

5.4 VARIANCE OF RANDOM VARIABLES

Problems

2. Part (a) will have the largest variance and part (c) will have the smallest.

(a) $E[X] = 0.5$; $\text{Var}(X) = (1)^2(0.5) - (0.5)^2 = 0.25$
(b) $E[X] = 0.4$; $\text{Var}(X) = (1)^2(0.4) - (0.4)^2 = 0.24$
(c) $E[X] = 0.1$; $\text{Var}(X) = (1)^2(0.1) - (0.1)^2 = 0.09$

4. (a) $E[X] = 2; \mathrm{Var}(X) = (1)(1/3) + (4)(1/3) + (9)(1/3) - (2)^2 = 0.6667$
(b) $E[X] = 5/3;\ \mathrm{Var}(X) = (1)(1/2) + (4)(1/3) + (9)(1/6) - (5/3)^2 = 0.5556$
(c) $E[X] = 7/3;\ \mathrm{Var}(X) = (1)(1/6) + (4)(1/3) + (9)(1/2) - (7/3)^2 = 0.5556$

6. Let X = amount earned. Then,

$$E[X] = (\$300)(1/3) + (\$600)(2/3) = \$500 \text{ and}$$
$$\mathrm{Var}(X) = (300)^2(1/3) + (600)^2(2/3) - 500^2 = 20{,}000$$

8. (a) $E[X] = (1)(12/60) + (2)(25/60) + (3)(16/60) + (4)(7/60) = 2.3$
(b) $E[X] = (1)(12/60) + (4)(25/60) + (9)(16/60) + (16)(7/60) - 2.3^2 = 0.8433$

10. Var(Profit earned by the nursery; Problem 27(b), Sec. 5.4) $= [(15{,}000)]^2(0.5) + [(21{,}000)]^2(0.2) + [(21{,}000)]^2(0.3) - (18{,}000)^2 = 9{,}000{,}000$

12. Let X = number of tickets received per taxi per month, and let Y = total number of tickets received. Now, $E[X] = 0.9$ and $\mathrm{Var}(X) = 0.49$, and since $Y = \sum_{i=1}^{4} X_i$, and X_i are independent, then $\mathrm{Var}(Y) = \sum_{i=1}^{4} \mathrm{Var}(X_i) = 4(0.49) = 1.96$.

14. Let X = amount earned. $E[X] = 275$. $\mathrm{Var}(X) = (40{,}000)(1/4) + (90{,}000)(3/4) - 275^2 = 1875$. Thus, $SD(X) = \$43.30$.

16. Let X = amount of money Robert earns. Let Y = amount of money Sandra earns. $E[X] = \$30{,}000$, $E[Y] = \$32{,}000$, $SD(X) = \$3{,}000$ and $SD(Y) = \$5{,}000$.

(a) $E[X + Y] = E[X] + E[Y] = \$62{,}000$
(b) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = (3{,}000)^2 + (5{,}000)^2 = 34{,}000{,}000$. Thus, $SD(X + Y) = \sqrt{\mathrm{Var}(X + Y)} = \$5{,}830.95$.

18. $\mathrm{Var}(2X + 3) = 16$ implies $\mathrm{Var}(2X) + \mathrm{Var}(3) = 16$. That is, $4\mathrm{Var}(X) = 16$. Thus, $\mathrm{Var}(X) = 4$ and $SD(X) = 2$.

5.5 BINOMIAL RANDOM VARIABLES

Problems

2. (a) $8!/(3!5!) = 56$; (b) $7!/(3!4!) = 35$; (c) $9!/(4!5!) = 126$

4. $P\{X = i\} = n!/[i!(n - i)!]p^i(1 - p)^{n-i}$, so

$P\{X = n\} = n!/[n!(n - n)!]p^n(1 - p)^{n-n} = p^n$.

If we let p be the probability of success in a binomial experiment with n trials and X is the number of successes, then $P\{X = n\} = p^n$.

6. Let X = number of defective ball bearings. Then, X is a binomial random variable with parameters $n = 5$ and $p = 0.05$.

(a) $P\{X = 0\} = 0.7738$
(b) $P\{X \geq 2\} = 1 - P\{X < 2\} = 1 - P\{X \leq 1\} = 1 - 0.9774 = 0.0226$

8. Let X = number of components that work. Then, X is a binomial random variable with parameters $n = 4$ and $p = 0.8$.

$$P\{X \geq 2\} = 1 - P\{X < 2\} = 1 - P\{X \leq 1\} = 1 - 0.0272 = 0.9728$$

10. Let X = number of correct responses. Then, X is a binomial random variable with parameters $n = 5$ and $p = 1/3$.

$$P\{X \geq 4\} = P\{X = 4\} + P\{X = 5\} = 0.0412 + 0.0041 = 0.0453$$

12. Let X = number of defective diskettes. Then, X is a binomial random variable with parameters $n = 10$ and $p = 0.05$.

(a) $P(\text{of returning package}) = P\{X \geq 1\} = 1 - P\{X < 1\}$
$= 1 - P\{X = 0\}$
$= 1 - 0.5987 = 0.4013$

(b) Let Y = number of packages bought. Then, Y is binomial random variable with parameters $n = 3$ and $p = 0.4013$.

$$P\{Y = 1\} = 0.4315$$

14. Let X = number of alcohol-related fatal automobile accidents. Thus, X is a binomial random variable with parameters $n = 3$ and $p = 0.55$.

(a) $P\{X = 3\} = 0.1664$
(b) $P\{X = 2\} = 0.4084$
(c) $P\{X \geq 1\} = 1 - P\{X < 1\} = 1 - P\{X = 0\} = 1 - 0.0911 = 0.9089$

16. X is a binomial random variable with parameters $n = 20$ and $p = 0.6$.

(a) $P\{X \leq 14\} = 0.8744$
(b) $P\{X < 10\} = 0.1275$
(c) $P\{X \geq 13\} = 1 - P\{X \leq 12\} = 1 - 0.5841 = 0.4159$
(d) $P\{X > 10\} = 1 - P\{X \leq 10\} = 1 - 0.2447 = 0.7553$
(e) $P\{9 \leq X \leq 16\} = P\{X \leq 16\} - P\{X \leq 8\}$
$= 0.9840 - 0.0565 = 0.9275$
(f) $P\{7 < X < 15\} = P\{X \leq 14\} - P\{X \leq 7\}$
$= 0.8744 - 0.0210 = 0.8534$

18. (a) $\text{Var}(X) = (20)(1/6)(5/6) = 100/36 = 25/9$
(b) $\text{Var}(5 \text{ or } 6) = (20)(1/3)(2/3) = 40/9$
(c) $\text{Var}(\text{even number}) = (20)(1/2)(1/2) = 5$
(d) $\text{Var}(\text{all except } 6) = (20)(5/6)(1/6) = 25/9$

20. Let X = number of heads. Then, X is a binomial random variable with parameters $n = 500$ and $p = 0.5$.

$$\text{Var}(X) = (500)(0.5)(0.5) = 125, \text{ thus } SD(X) = \sqrt{125} = 11.1803$$

22. Let X = number of heads. Then, X is a binomial random variable with parameters n and p. Since the expected number of heads is 6, then

$np = 6$ or $10p = 6$, and thus, $p = 0.6$. Thus, $P\{X = 8\} = 0.121$.

24. X is a binomial random variable with expected value 4.5 and variance 0.45. That is, $np = 4.5$ and $np(1 - p) = 0.45$. Solving gives $n = 5$ and $p = 0.9$.

(a) $P\{X = 3\} = 0.0729$
(b) $P\{X \geq 4\} = P\{X = 4\} + P\{X = 5\} = 0.3280 + 0.5905 = 0.9185$

26. The probability that the first 4 games will be split 3 and 1 is $2\binom{4}{1}(1/2)^4 = 1/2$. Hence, the desired probability is $1/2(1/2)^3 = 1/16$.

5.6 HYPERGEOMETRIC RANDOM VARIABLES

Problems

2. X is a binomial random variable with $n = 20$ and $p = 0.85$.

4. X is a binomial random variable with parameters $n = 100$ and $p = 0.05$.

6. X is a hypergeometric random variable with parameters $N = 52$, $n = 4$, and $p = 4/52 = 0.0769$.

5.7 POISSON RANDOM VARIABLES

Problems

2. (a) Binomial: $P\{X = 2\} = 0.1937$
Poisson approximation: $P\{X = 2\} = 0.1839$ with $\lambda = np = 10(0.1) = 1$
(b) Binomial: $P\{X = 2\} = 0.0746$
Poisson approximation: $P\{X = 2\} = 0.0758$ with $\lambda = np = 10(0.05) = 0.5$
(c) Binomial: $P\{X = 2\} = 0.0042$
Poisson approximation: $P\{X = 2\} = 0.0045$ with $\lambda = np = 10(0.01) = 0.1$
(d) Binomial: $P\{X = 2\} = 0.2335$
Poisson approximation: $P\{X = 2\} = 0.2240$ with $\lambda = np = 10(0.3) = 0.3$

4. (a) $E(X) = 144$
(b) $\text{Var}(X) = 144$, so $SD(X) = 12$

Review Problems

2. (a) Since $P\{X \leq 6\} = 0.7$ and $P\{X < 6\} = 0.5$, then $P\{X = 6\}$
$= 0.7 - 0.5 = 0.2$
(b) $P\{X > 6\} = 1 - P\{X \leq 6\} = 1 - 0.7 = 0.3$

4. X is either 1 or 2 and $E[X] = 1.6$. Let $P\{X = 1\} = p$ and $P\{X = 2\} = 1 - p$. Then, $(1)(p) + (2)(1 - p) = 1.6$ from which $p = 0.4 = P\{X = 1\}$.

6. Let $X =$ the amount of time the first one to arrive has to wait.

The possible values of X are 0, 1, and 2 hours. Note that $X = 0$ can occur 3 times, $X = 1$ can occur 4 times, $X = 2$ can occur 2 times. Thus,

$$P\{X = 1\} = 3/9,\ P\{X = 1\} = 4/9, \text{ and } P\{X = 2\} = 2/9. \text{ Hence,}$$

$$E[X] = (0)(3/9) + (1)(4/9) + (2)(2/9) = 8/9 \text{ hours}$$

8. Let $X =$ number of defective batteries, then X is a hypergeometric random variable with parameters $N = 12,\ n = 2,\ p = 4/12 = 1/3$. Since $E[X] = np$, then $E[X] = (2)(1/3) = 2/3$.

10. $E[3X + 10] = E[3X] + E[10] = 3E[X] + 10 = 70$, from which $E[X] = 20$.

12. Let $X =$ number of children in a randomly selected family and $Y =$ number of children in the family of a randomly selected child.

(a) Y will have the larger expected value since the probabilities in Y are proportional to Y, but the probabilities in X are equal.
(b) $E[X] = (4)(1/5) + (3)(1/5) + (2)(2/5) + (1)(1/5) = 2.4$
$E[Y] = (4)(4/12) + (3)(3/12) + (2)(4/12) + (1)(1/12) = 2.83333$

14. (a) $\text{Var}(2X + 14) = \text{Var}(2X) + \text{Var}(14) = 4\text{Var}(X) + 0 = 4(4) = 16$
(b) Now, $\text{Var}(2X) = 4\text{Var}(X) = 16$, hence $SD(2X) = \sqrt{16} = 4$
(c) $SD(2X + 14) = \sqrt{16} = 4$

16. Let $X =$ gross annual earnings of first client and $Y =$ gross annual earnings of second client. $E[X] = \$200{,}000$, $SD(X) = \$60{,}000$, $E[Y] = \$140{,}000$, $SD(Y) = \$50{,}000$.

(a) E[manager's total fee] $= E[0.15X + 0.2Y] = 0.15E[X] + 0.2E[Y] =$ [\$(200,000)(0.15) + (140,000)(0.2)] = \$58,000

(b) Var (manager's total fee) $= \text{Var}[0.15X + 0.2Y] = 0.0225\text{Var}(X) + 0.04\text{Var}(Y) = 181{,}000{,}000$. So, SD(manager's total fee) = \$13,453.62.

18. (a) Let X = number of normal size sets that were bought. Then, X is a binomial random variable with parameters $n = 5$ and $p = 0.3$. Thus, $P\{X = 3\} = 0.1323$.

(b) Let Y = number of extra-large sets that were bought. Then, Y is a binomial random variable with parameters $n = 5$ and $p = 0.1$.

$$P\text{ \{do not purchase any extra-large size sets\}} = P\{Y = 0\} = 0.5905$$

(c) Let Z = number of zero purchases. Then, Z is a binomial random variable with $n = 5$ and $p = 0.6$.

$$P\{\text{purchase of a total of 2 sets}\} = P\{\text{3 zero purchases}\} = 0.3456$$

20. X is a binomial random variable with $E[X] = 6$ and $\text{Var}(X) = 2.4$. Thus, $p = 0.6$ and $n = 10$.

(a) $P\{X > 2\} = 1 - P\{X \leq 2\} = 1 - [P\{X = 0\} + P\{X = 1\} + P\{X = 2\}]$ $= 1 - [0.0001 + 0.0016 + 0.0106] = 0.9877$

(b) $P\{X \leq 9\} = 1 - P\{X = 10\} = 0.9940$

(c) $P\{X = 12\} = 0$

Chapter 6 NORMAL RANDOM VARIABLES

6.2 CONTINUOUS RANDOM VARIABLES

Problems

2. (a) The length of the rectangle is $(b - a)$. Since the total area under the curve must be one, in order for this area to be one, then the height must be $1/(b - a)$. Recall: area of rectangle = length × height = $(b - a)[1/(b - a)] = 1$.

(b) $P\{X \leq (a + b)/2\} = [1/(b - a)] \times [(a + b)/2 - a]$
$= [1/(b - a)] \times [(b - a)/2] = 1/2$

4. Let X = amount of time in minutes that you will have to wait.

(a) $P\{X \geq 30\} = 30/60 = 1/2$ (b) $P\{X < 15\} = 15/60 = 1/4$
(c) $P\{10 \leq X \leq 35\} = 25/60 = 5/12$ (d) $P\{X < 45\} = 45/60 = 3/4$

6. Let X = amount of playing time in minutes.

(a) $P\{X > 20\} = 0.5 + 0.25 = 0.75$ (b) $P\{X < 25\} = 0.25 + 0.5/2 = 0.5$
(c) $P\{15 \leq X \leq 35\} = 0.25/2 + 0.5 + 0.25/2 = 0.75$
(d) $P\{X > 35\} = 0.25/2 = 0.125$

8. (a) Recall the area of a triangle = (1/2) × base × height.
Since the area of the triangle must be equal to 1 (probability density curve), then $1 = (1/2)(4)(\text{height})$ from which the height is equal to 1/2 unit.
(b) P {she will study more than 3 hours} = $(1/2)(1)(1/4) = 1/8$
(c) P {she will study between 1 and 3 hours} = $(2) \times (1/2 - 1/8) = 3/4$

6.3 NORMAL RANDOM VARIABLES

Problems

2. Let X = height of a male in inches.

Note: A height of 82 inches is two standard deviations above the mean. Because of symmetry, the area to the right of 82 inches is approximately 0.025. Thus, $P\{X < 82\} = 1 - 0.025 = 0.975$.

4. (c) **6.** (c) **8.** (c) **10.** (c) **12.** (d) **14.** (b) **16.** (c)

18. (a) $P\{X > 105\} \approx 0.0062$ and $P\{Y > 105\} = 0.5$. Hence, Y is more likely to exceed 105.

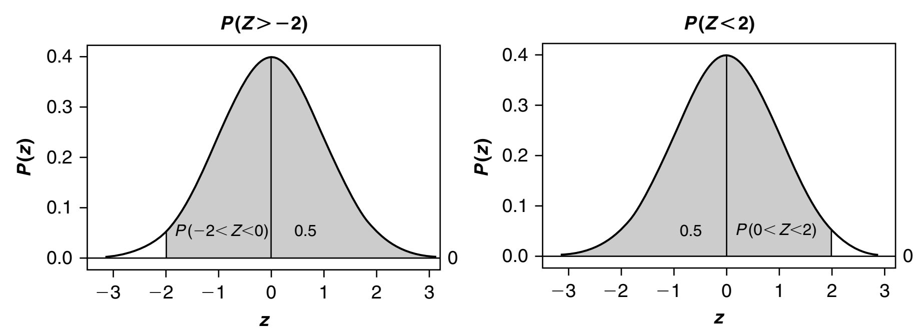
(b) $P\{X < 95\} \approx 0.0062$ and $P\{Y < 95\} \approx 0.16$. Hence, Y is more likely to be less than 95.

6.4 PROBABILITIES ASSOCIATED WITH A STANDARD NORMAL RANDOM VARIABLE

Problems

2. Now, $P\{-Z < x\} = P\{Z > -x\} = 0.5 + P\{-x < Z < 0\} = 0.5 + P\{0 < Z < x\} = P\{Z < x\}$, for all x. Hence, $-Z$ is also a standard random variable.

4. $P\{Z > -2\} = 0.5 + P\{-2 < Z < 0\} = 0.5 + P\{0 < Z < 2\} = P\{Z < 2\}$. This is depicted in the two displays.



6. (a) Using the results in problem 5, $P\{-1 < Z < 1\} = 2P\{Z < 1\} - 1 = 0.6826$.

(b) $P\{|Z| < 1.4\} = P\{-1.4 < Z < 1.4\} = 2P\{Z < 1.4\} - 1 = 0.8384$

6.6 ADDITIVE PROPERTY OF NORMAL RANDOM VARIABLES

Problems

2. (a) $P\{X > 12\} = P\{Z > (12 - 10)/3\} = P\{Z > 0.67\} = 1 - P\{Z < 0.67\} = 1 - 0.7486 = 0.2514$

(b) $P\{X < 13\} = P\{Z < (13 - 10)/3\} = P\{Z < 1\} = 0.8413$.

(c) $P\{8 < X < 11\} = P\{(8 - 10)/3 < Z < (11 - 10)/3\} =$
$P\{-0.67 < Z < 0.33\} = P\{Z < 0.33\} - P\{Z < -0.67\} =$
$P\{Z < 0.33\} - [1 - P\{Z < 0.67\}] = 0.6293 - [1 - 0.7486] = 0.3779$

(d) $P\{X > 7\} = P\{Z > (7 - 10)/3\} = P\{Z > -1\} = P\{Z < 1\} = 0.8413$

(e) $P\{|X - 10| > 5\} = P\{|Z| > 5/3\} = P\{|Z| > 1.67\} =$
$2(1 - P\{Z < 1.67\}) = 2(1 - 0.9525) = 2(0.0475) = 0.095$

(f) $P\{X > 10\} = P\{Z > (10 - 10)/3\} = P\{Z > 0\} = 0.5$

(g) $P\{X > 20\} = P\{Z > (20 - 10)/3\} = P\{Z > 3.33\} = 1 - P\{Z < 3.33\} =$
$1 - 0.9996 = 0.0004$

4. Let $X =$ score on the scholastic achievement test.

(a) $Z = (700 - 520)/94 = 1.9149$. Thus, the score of 700 is 1.9149 standard deviations above the mean of 520.

(b) $P\{X > 700\} = P\{Z > 1.91\} = 1 - P\{Z < 1.91\} = 1 - 0.9719 = 0.0281$

Thus, the percentage who scored greater than 700 is 2.81%.

6. Let $X =$ life of the tire.

(a) $P\{30 < X < 40) = P\{(30 - 35)/5 < Z < (40 - 35)/5\}$
$= P\{-1 < Z < 1\} = P\{Z < 1\} - P\{Z < -1\}$
$= P\{Z < 1\} - [1 - P\{Z < 1\}]$
$= 0.8413 - [1 - 0.8413] = 0.6826$

(b) $P\{X > 40\} = P\{Z > 1\} = 1 - P\{Z < 1\} = 1 - 0.8413 = 0.1587$

(c) $P\{X > 50\} = P\{Z > (50 - 35)/5\} = P\{Z > 3\} = 1 - P\{Z < 3\} = 1 -$
$0.9987 = 0.0013$

8. Let $X =$ pulse rate of young adults.
$P\{X > 95\} = P\{Z > (95 - 72)/9.5\} = P\{Z > 2.42\} = 1 - P\{Z < 2.42\} =$
$1 - 0.9922 = 0.0078$
Thus, 0.78% of the population of young adults do not meet the military standard.

10. Let $X =$ diameter(inches) of the bolts.
$P\{1.09 < X < 1.11\} = P\{(1.09 - 1.1)/0.005 < Z < (1.11 - 1.1)/0.005\} =$
$P\{-2 < Z < 2\} = P\{Z < 2\} - P\{Z < -2\} = P\{Z < 2\} - [1 - P\{Z < 2\}] =$
$0.9772 - [1 - 0.9772] = 0.9544$. Thus, the proportion of bolts that do not meet the specifications is $(1 - 0.9544) \times 100\% = 4.56\%$.

12. Let $X =$ lifetime of the first battery and $Y =$ lifetime of the second battery.

(a) $P\{X > 20\} = P\{Z > (20 - 24)/6\} = P\{Z > -0.67\} = P\{Z < 0.67\} =$
0.7486
$P\{Y > 20\} = P\{Z > (20 - 22)/2\} = P\{Z > -1\} = P\{Z < 1\} = 0.8413$
Since the second battery has the higher probability of lasting beyond 20,000 miles, purchase the second battery.

(b) $P\{X > 21\} = P\{Z > (21 - 24)/6\} = P\{Z > -0.5\} = P\{Z < 0.5\}$
$= 0.6915$
$P\{Y > 21\} = P\{Z > (21-22)/2\} = P\{Z > -0.5\} = P\{Z < 0.5\} = 0.6915$

Since the first battery has the same probability of lasting beyond 20,000 miles as the second, you may purchase either of the two batteries.

14. Let X = annual rainfall (inches) in Cincinnati.

(a) $P\{X > 42] = P\{Z > (42 - 40.14)/8.7\} = P\{Z > 0.21\} =$
$1 - P\{Z < 0.21\} = 1 - 0.5832 = 0.4168$

(b) Let Y = sum of the next 2 years' rainfall. Then, Y is also normal with mean $(40.14 + 40.14) = 80.28$ and standard deviation $\sqrt{(8.7^2 + 8.7^2)} = 12.3037$. So, $P\{Y > 84\} = P\{Z > (84 - 80.28)/12.3037\} = P\{Z > 0.3\} =$
$1 - P\{Z < 0.3\} = 1 - 0.6179 = 0.3821$

(c) Let W = sum of the next 3 years' rainfall. Then, W is also normal with mean of 120.42 and standard deviation 15.0688. So,
$P\{W > 126\} = P\{Z > (126 - 120.42)/15.0688\} = P\{Z > 0.37\} =$
$1 - P\{Z < 0.37\} = 1 - 0.6443 = 0.3557$

(d) Each year is independent.

16. Let X = weight of the chemistry text and Y = weight of the economics text.

(a) $X + Y$ will be normal with mean of 8.1 and standard deviation of 2.56. So,
$P\{(X+Y) > 9\} = P\{Z > (9 - 8.1)/2.56\} = P\{Z > 0.35\} = 1 - P\{Z < 0.35\} =$
$1 - 0.6368 = 0.3632$

(b) $P\{Y > X\} = P\{(Y - X) > 0\}$. Now, $Y - X$ will be normal with mean of 1.1 and a standard deviation of 2.56. Thus,
$P\{(Y - X) > 0\} = P\{Z > (0 - 1.1)/2.56\} = P\{Z > -0.43\} = P\{Z < 0.43\} =$
0.6664

(c) The random variables X and Y are independent. That is, the weights for the introductory chemistry and economic textbooks are independent.

6.7 PERCENTILES OF NORMAL RANDOM VARIABLES

Problems

2. (a) $P\{|Z| > x\} = 0.05 \Rightarrow P\{Z > x\} + P\{Z < -x\} = 0.05 \Rightarrow$
$2[1 - P\{Z < x\}] = 0.05 \Rightarrow P\{Z < x\} = 0.975 \Rightarrow x = 1.96$

(b) $P\{|Z| > x\} = 0.025 \Rightarrow P\{Z > x\} + P\{Z < -x\} = 0.025$
$\Rightarrow 2[1 - P\{Z < x\}] = 0.025 \Rightarrow P\{Z < x\} = 0.9875 \Rightarrow x = 2.24$

(c) $P\{|Z| > x\} = 0.005 \Rightarrow P\{Z > x\} + P\{Z < -x\} = 0.005$
$\Rightarrow 2[1 - P\{Z < x\}] = 0.005 \Rightarrow P\{Z < x\} = 0.9975 \Rightarrow x = 2.81$

4. Let X be the real estate brokers' exam scores, and let x be the required cutoff score. Thus, we have $P\{X > x\} = 0.1$ or $P\{Z > (x - 420)/66\} = 0.1$. That is, $(x - 420)/66 = 1.28$ or $x = 504.48$.

The "excellent" scores should begin at 505 assuming that only integers are used as values of the exam scores.

6. Let X be the time (seconds) it takes to run a mile by the high school boys, and let x be the critical time (seconds). Thus, we have $P\{X > x\} = 0.2$ or $P\{X < x\} = 0.8$. That is, $P\{Z < (x-460)/40\} = 0.8$ or $(x-460)/40 = 0.84$ from which $x = 493.6$ seconds. Thus, if a male student runs the mile in more than 493.6 seconds, then he should have additional training.

8. Let X be the time (seconds) it takes to run a mile by the high school boys, and let x be the cutoff time (seconds). Thus, we have $P\{X < x\} = 0.01$. That is, $P\{Z < (x-460)/40\} = 0.01$ or $(x-460)/40 = -2.33$ from which $x = 366.8$ seconds.

10. Let X be life (in miles) of a transmission on a new car. Let x be the length of the warranty period. Thus, we have $P\{X < x\} = 0.2$. That is, $P\{Z < (x-70{,}000)/10{,}000\} = 0.2$ from which $(x-70{,}000)/10{,}000 = -0.84$ or $x = 61{,}600$ (miles).

12. Let X be the scores on the quantitative part of the Graduate Record Exam.

(a) Let x be the cutoff score in order to be in the top 10%. Thus, we have $P\{X > x\} = 0.1$. That is, $P\{Z > (x-510)/92\} = 0.1$ from which $(x-510)/92 = 1.28$ or $x = 627.76$.

(b) Let x be the cutoff score in order to be in the top 5%. Thus, we have $P\{X > x\} = 0.05$. That is, $P\{Z > (x-510)/92\} = 0.05$ from which $(x-510)/92 = 1.645$ or $x = 661.34$.

(c) Let x be the cutoff score in order to be in the top 1%. Thus, we have $P(X > x\} = 0.01$. That is, $P\{Z > (x-510)/92\} = 0.01$ from which $(x-510)/92 = 2.33$ or $x = 724.36$.

Review Problems

2. (a) $z_{0.04} = 1.75$ (b) $z_{0.22} = 0.77$

(c) $P\{Z > 2.2\} = 1 - P\{Z < 2.2\} = 1 - 0.9861 = 0.0139$

(d) $P\{Z < 1.6\} = 0.9452$. (e) $z_{0.78} = -0.77$

4. Let X = jet pilot's blackout thresholds.

(a) $P\{X > 5\} = P\{Z > (5-4.5)/0.7\} = P\{Z > 0.71\} = 1 - P\{Z < 0.71\} = 1 - 0.7611 = 0.2389$

(b) $P\{X < 4\} = P\{Z < (4-4.5)/0.7\} = P\{Z < -0.71\} = 0.2389$

(c) $P\{3.7 < X < 5.2\} = P\{(3.7-4.5)/0.7 < Z < (5.2-4.5)/0.7\} = P\{-1.14 < Z < 1\} = P\{Z < 1\} - [1 - P\{Z < 1.14\}] = 0.7142$

6. Let $X =$ number of lost years due to smoking.

(a) $P\{X < 2\} = P\{Z < (2 - 5.5)/1.5\} = P\{Z < -2.33\} = 1 - P\{Z < 2.33\} = 1 - 0.9901 = 0.0099$

(b) $P\{X > 8\} = P\{Z > (8 - 5.5)/1.5\} = P\{Z > 1.67\} = 1 - P\{Z < 1.67\} = 1 - 0.9525 = 0.0475$

(c) $P\{4 < X < 7\} = P\{(4 - 5.5)/1.5 < Z < (7 - 5.5)/1.5\} = P\{-1 < Z < 1\} = 0.8413 - [1 - 0.8413] = 0.6826$

8. Let X be the speed (mph) of a car traveling on a New Jersey highway, and let x be the cutoff speed. Thus, we have $P\{X > x\} = 0.05$ or $P\{Z > (x - 60)/5\} = 0.05$. That is, $(x - 60)/5 = 1.645$ or $x = 68.225$. Tickets will be issued for speeds greater than 68.225 mph.

10. Let $X =$ yearly number of miles accumulated by the automobile. Now, $P\{X < 17{,}400\} = P\{Z < (17{,}400 - 18{,}000)/1{,}700\} = P\{Z < -0.35\} = 1 - P\{Z < 0.35\} = 1 - 0.6368 = 0.3632$. Since this probability is not less than 0.2 (20%), then 17,400 miles will not be in the 20% of lowest miles, and thus, this car is not likely to be kept.

An alternative solution is to let x be the cutoff mileage; thus, we have $P\{X < x\} = 0.2$ or $P\{Z < (x - 18{,}000)/1{,}700) = 0.2$. That is, $(x - 18{,}000)/1{,}700 = -0.84$ from which $x = 16{,}572$. Since $17{,}400 > 16{,}572$, then this car is not likely to be kept.

12. Let $X =$ amount (pounds) of tomatoes consumed per year by a randomly chosen female.
Let $Y =$ amount (pounds) of tomatoes consumed per year by a randomly chosen male.

(a) $P\{X > 14.6\} = P\{Z > (14.6 - 14)/2.7\} = P\{Z > 0.22\} = 1 - P\{Z < 0.22\} = 1 - 0.5871 = 0.4129$

(b) $P\{Y < 14\} = P\{Z < (14 - 14.6)/3\} = P\{Z < -0.2\} = 1 - P\{Z < 0.2\} = 1 - 0.5793 = 0.4207$

(c) $P\{(X > 15) \cap (Y < 151)\} = P\{X > 15\} \times P\{Y < 15\}$. Now, $P\{X > 15\} = P\{Z > (15 - 14) / 2.7\} = P\{Z > 0.37\} = 1 - P\{Z < 0.37\} = 1 - 0.6443 = 0.3557$. Also, $P\{Y < 15\} = P\{Z < (15 - 14.6)/3\} = P\{Z < 0.13\} = 0.5517$. Hence, $P\{(X > 15) \cap (Y < 151)\} = (0.3557)(0.5517) = 0.1962$.

(d) $P\{X > Y\} = P\{(X - Y) > 0\} = P\{Z > (0 - (-0.6))/4.0361\} = P\{Z > 0.15\} = 1 - P\{Z < 0.15\} = 1 - 0.5596 = 0.4404$

Note: $\mu_{X-Y} = 14 - 14.6 = -0.6$ and $\sigma_{X-Y} = \sqrt{(2.7^2 + 3^2)} = 4.0361$.

14. Let μ and σ be the mean and standard deviation (in units of a thousand dollars) of X, a physician's earnings. Then

$$0.25 = P(X < 180) = P\left(Z < \frac{180 - \mu}{\sigma}\right)$$

and

$$0.75 = P(X < 320) = P\left(Z < \frac{320 - \mu}{\sigma}\right)$$

Using that $0.25 = P(Z < -.675)$ and $0.75 = P(Z < .675)$ gives

$$\frac{180 - \mu}{\sigma} = -0.675$$

and

$$\frac{320 - \mu}{\sigma} = 0.675$$

giving that $\mu = 250$, $\sigma = 103.7$. Hence,

(a) 0

(b) $P(Z < 50/103.7) - P(Z < 10/103.7) = 0.147$

16. (a) Statistics exam, because $\frac{62-55}{10} > \frac{70-60}{20}$.

(b) $P(X < 70) = P(Z < \frac{70-60}{20}) = P(Z < 0.5) = 0.6915$

(c) $P(X < 62) = P(Z < \frac{62-55}{10}) = P(Z < 0.7) = 0.7580$

Chapter 7 DISTRIBUTIONS OF SAMPLING STATISTICS

7.3 SAMPLE MEAN

Problems

2. (a) $\mu = E[X] = (1)(0.7) + (2)(0.3) = 1.3$

(b) $\sigma^2 = \text{Var}(X) = E[(X-\mu)^2] = (1-1.3)^2(0.7) + (2-1.3)^2(0.3) = 0.21$

(c) The possible values of the pair (X_1, X_2) are $(1, 1)$, $(1, 2)$, $(2, 1)$, and $(2, 2)$. Thus, the possible averages $(\overline{X})$ for these pairs are 1, 1.5, and 2.

(d) $P\{\overline{X} = 1\} = P\{(1, 1)\} = (0.7)(0.7) = 0.49$

$P\{\overline{X} = 1.5\} = P\{(1, 2) \text{ or } (2, 1)\} = (0.7)(0.3) + (0.3)(0.7) = 0.42$

$P\{\overline{X} = 2\} = P\{(2, 2)\} = (0.3)(0.3) = 0.09$

(e) $E[\overline{X}] = (1)(0.49) + (1.5)(0.42) + (2)(0.09) = 1.3$

$$\begin{aligned}\text{Var}(\overline{X}) = E[(\overline{X} - 1.3)^2] &= (1-1.3)^2(0.49) \\ &+ (1.5-1.3)^2(0.42) + (2-1.3)^2(0.09) = 0.105\end{aligned}$$

(f) Yes. $E[\overline{X}] = 1.3 = \mu$ and $\text{Var}(\overline{X}) = 0.105 = \sigma^2/n$

4. $E[\overline{X}] = \$80$; $SD(\overline{X}) = \$40/\sqrt{(20)} = \$8.94$

6. (a) $E[\overline{X}] = 475$ hours; $SD(\overline{X}) = 60/\sqrt{(100)} = 6$

(b) $E[\overline{X}] = 475$ hours; $SD(\overline{X}) = 60/\sqrt{(200)} = 4.2426$

(c) $E[\overline{X}] = 475$ hours; $SD(\overline{X}) = 60/\sqrt{(400)} = 3$

7.4 CENTRAL LIMIT THEOREM

Problems

2. Let X = number of miles flown by frequent flyers on a particular airline each year.

(a) $P\{4500/200 \le \overline{X} \le 5000/200\} = P\{22.5 \le \overline{X} \le 25\}$

$= P\{(22.5-23)/(11/\sqrt{20}) \le \overline{X} \le (25-23)/(11/\sqrt{20})\}$

$= P\{-0.2 \le Z \le 0.81\} = P\{Z \le 0.81\} - [1 - P\{Z \le 0.2\}]$

$= 0.7910 - [1 - 0.5793] = 0.3703$

(b) $P\{\overline{X} > 5200/200\} = P\{\overline{X} > 26\} = P\{Z > (26 - 23)/(11/\sqrt{20})\}$
$= P\{Z > 1.22\} = 1 - P\{Z < 1.22\} = 1 - 0.8888 = 0.1112$

4. Let X = gain on the bet.

(a) $E[X] = (35)(1/38) - (1)(37/38) = -0.0526$

$\text{Var}(X) = (35)^2(1/38) + (-1)^2(37/38) - (-0.0526)^2 = 33.2078$, so
$SD(X) = \sqrt{33.2078} = 5.7626$

(b) $$P\left\{\sum_{i=1}^{1000} X_i > 0\right\} = P\left\{Z > \frac{0 - (-0.0526 \times 1000)}{5.7626 \times \sqrt{1000}}\right\} = P\{Z > 0.29\}$$

$$= 1 - 0.6141 = 0.3859 \approx 0.39$$

(c) $$P\left\{\sum_{i=1}^{1000} X_i > 0\right\} = P\left\{Z > \frac{0 - (-0.0526 \times 100000)}{5.7626 \times \sqrt{100000}}\right\} = P\{Z > 2.89\}$$

$$= 1 - 0.9981 = 0.0019 \approx 0.002$$

6. Let X = lifetime of a zircon semiconductor.
$P\{\sum X > 2000\} = P\{\overline{X} > 2000/22\} = P\{\overline{X} > 90.9091\}$

$= P\{Z > (90.9091 - 100)/(34/\sqrt{22})\} = P\{Z > -1.25\}$

$= P\{Z < 1.25\} = 0.8944$

8. Let X = total amount (inches) of snowfall.

(a) $P\{X < 80\} = P\{X < 80/50\} = P\{X < 1.6\}$
$= P\{Z < (1.6 - 1.5)/(0.3/\sqrt{50})\} = P\{Z < 2.36\} = 0.9909$

(b) Assumption: $\overline{X}$ has an approximate normal distribution with a mean of 1.5 and a standard deviation of $(0.3/\sqrt{50})$.

(c) The assumption is justified since the sample size $n = 50 > 30$.

10. Let X = total of the first 140 rolls.

$P\{X \le 400\} = P\{\overline{X} < 400 / 140\} = P\{\overline{X} < 2.86\}$
$= P\{Z < (2.86 - 3.5)/(1.7078/\sqrt{140})\} = P\{Z < -4.45\} \approx 0$

12. Let X = number of miles that an electric car battery can function.

(a) $P\{-20 < \overline{X} - \mu < 20\} = P\{-20/(100/\sqrt{10}) < Z < 20/(100/\sqrt{10})\}$
$= P\{-0.63 < Z < 0.63\} = 2P\{Z < 0.63\} - 1 = 2(0.7357) - 1 = 0.4714$

(b) $P\{-20 < \overline{X} - \mu < 20\} = P\{-207(100/\sqrt{20}) < Z < 207(100/\sqrt{20})\}$
$= P\{-0.89 < Z < 0.89\} = 2P\{Z < 0.89\} - 1 = 2(0.8133) - 1 = 0.6266$

(c) $P\{-20 < \overline{X} - \mu < 20\} = P\{-20/(100/\sqrt{40}) < Z < 20/(100/\sqrt{40})\}$
$= P\{-1.26 < Z < 1.26\} = 2P\{Z < 1.26\} - 1 = 2(0.8962) - 1 = 0.7924$

(d) $P\{-20 < \overline{X} - \mu < 20\} = P\{-20/(100/\sqrt{100}) < Z < 20/(100\sqrt{100})\}$
$= P\{-2 < Z < 2\} = 2P\{Z < 2\} - 1 = 2(0.9772) - 1 = 0.9544$

14. Let X = lifetime of the electric bulb.

$P\{\overline{X} < 480\} = P\{Z < (480 - 500)/(60/\sqrt{20})\} = P\{Z < -1.49\}$
$= 1 - P\{Z < 1.49\} = 1 - 0.9319 = 0.0681$

16. Let X = test scores for the group of 25 students and
Y = test scores for the group of 64 students.

(a) $P\{72 < \overline{X} < 82\} = P\{(72 - 77)/(15/\sqrt{25}) < Z < (82 - 77)/(15/\sqrt{25})\}$
$= P\{-1.67 < Z < 1.67\} = 2P\{Z < 1.67\} - 1 = 2(0.9525) - 1 = 0.905$

(b) $P\{72 < \overline{Y} < 82\} = P\{(72 - 77)/(15/\sqrt{64}) < Z < (82 - 77)/(15/\sqrt{64})\}$
$= P\{-2.67 < Z < 2.67\} = 2P\{Z < 2.67\} - 1 = 2(0.9962) - 1 = 0.9924$

(c) $P\{\overline{X} > \overline{Y}\} = P\{\overline{X} - \overline{Y} > 0\} = P\{Z > (0 - (77 - 77))/SD\}$
$= P\{Z > 0\} = 0.5$

(d) For $n = 25$, $SD = 3$ and for $n = 64$, $SD = 1.875$. The class with size 25 has greater standard deviation; hence, more variation from the mean is likely. Thus, the smaller class was more likely to have an average of at least 83.

7.5 SAMPLING PROPORTIONS FROM A FINITE POPULATION

Problems

2. Let X = number of defective batteries; thus, X is binomial with $n = 8$ and $p = 0.1$. We cannot use the normal approximation since $np = 0.8$ and $n(1 - p) = 7.2$.

(a) $P\{X = 0\} = 0.4305$
(b) 15% of 8 = 1.2, thus $P\{X > 1.2\} = P\{X \geq 2\} = 1 - [P\{X < 2\}]$
$= 1 - [P\{X = 0\} + P\{X = 1\}] = 1 - [0.4305 + 0.3826] = 0.1869$
(c) 8% of 8 = 0.64 and 12% of 8 = 0.96.
Hence, $P\{0.64 < X < 0.96\} = 0$.

4. (a) Since $n(1 - p) = 10 \times 0.4 = 4$, the normal approximation should not be used. Instead, use the binomial distribution. So, if X represents the binomial random variable with $n = 10$ and $p = 0.6$, then $P\{\overline{X} > 0.55\} = P\{X > 5\} = 1 - P\{X \leq 5\} = 1 - 0.3669 = 0.6331$.

(b) $P\{\overline{X} > 0.55\} \approx P\{Z > (0.55 - 0.6)/0.049\} = P\{Z > -1.02\}$
$= P\{Z < 1.02\} = 0.8461$

(c) $P\{\overline{X} > 0.55\} \approx P\{Z > (0.55 - 0.6)/0.0155\} = P\{Z > -3.23\}$
$= P\{Z < 3.23\} = 0.9994$

(d) $P\{\overline{X} > 0.55\} \approx P\{Z > (0.55 - 0.6)/0.0049\} = P\{Z > -10.2\}$
$= P\{Z < 10.2\} = 1$

6. Let X = number of unemployed Japanese workers.

(a) $P\{X \leq 30\} \approx P\{Z < (30.5 - 32.4)/\sqrt{((600)(0.054)(0.946))}\}$
$= P\{Z \leq -0.3432\} = 0.3657$ (using TI-83). ***Note:*** If you use statistical tables, the answer will be 0.3669.

(b) $P\{X > 40\} = P\{X \geq 41\} \approx$
$P\{Z > (40.5 - 32.4)/\sqrt{((600)(0.054)(0.946))}\}$
$= P\{Z > 1.4631\} = 1 - P\{Z < 1.4631\} = 0.0717$(using TI-83).
Note: If you use statistical tables, the answer will be 0.0721.

8. Let X = number of people who favor the proposed rise in school taxes.

(a) $P\{X \geq 45\} \approx P\{Z \geq (44.5 - 65)/\sqrt{((100)(0.65)(0.35))}\}$
$= P\{Z \geq -4.3\} \approx 1$

(b) $P\{X < 60\} = P\{X \leq 59\} \approx P\{Z < (59.5 - 65)/\sqrt{((100)(0.65)(0.35))}\}$
$= P\{Z < -1.15\} = 1 - P\{Z < 1.15\} = 1 - 0.8749 = 0.1251$

(c) $P\{55 \leq X \leq 75\} \approx P\{(54.5 - 65)/\sqrt{((100)(0.65)(0.35))}$

$$\leq Z \leq (75.5 - 65)/\sqrt{((100)(0.65)(0.35))}\}$$

$= P\{-22 \leq Z \leq 2.2\} = 2P\{Z \leq 2.2\} - 1 = 2(0.9861) - 1 = 0.9722$

10. Let X = number of passengers who show up for the flight.

$$P\{X \le 250\} \approx P\{Z \le (250.5 - 244.4)/\sqrt{((260)(0.94)(0.06))}\}$$
$$= P\{Z \le 1.59\} = 0.9441$$

12. Let X = number of students planning to major in business.

$$P\{X > 60\} = P\{X \ge 61\} \approx P\{Z \ge (60.5 - 54)/\sqrt{((200)(0.27)(0.73))}\}$$
$$= P\{Z \ge 1.04\} = 1 - P\{Z < 1.04\} = 1 - 0.8508 = 0.1492$$

14. Let X = number of students planning to major in engineering.

$$P\{X < 15\} = P\{X \le 14\} \approx P\{\, Z < (14.5\ - 20)/\sqrt{((200)(0.1)(0.9))}\}$$
$$= P\{Z \le -1.3\} = 1 - P\{Z < 1.3\} = 1 - 0.9032 = 0.0968$$

16. (a) $P\{X \le 100\} = P\{X \le 100.5\} \approx P\{Z \le (100.5 - 90)/6\} = P\{Z \le 1.75\}$
$= 0.9599$

(b) $P\{X > 75\} = P\{X \ge 76\} \approx P\{Z \ge (75.5 - 90)/6\} = P\{Z \ge -2.42\}$
$= P\{Z < 2.42\} = 0.9922$

(c) $P\{80 < X < 100\} = P\{81 \le X \le 99\}$
$\approx P\{(80.5 - 90)/6 \le Z \le (99.5 - 90)/6\}$
$= P\{-1.58 \le Z \le 1.58\} = P\{Z \le 1.58\} - [1 - P\{Z \le 1.58\}] = 0.8858$

18. (a) Let X = number of males who never eat breakfast.

$$P\{X \ge 75\} \approx P\{Z \ge (74.5 - 75.6)/7.52\} = P\{Z \ge -0.15\}$$
$$= 0.5596$$

(b) Let X = number of males who smoke.
$P\{X < 100\} = P\{X \le 99\} \approx P\{Z \le (99.5 - 97.8)/8.12\}$
$= P\{Z \le 0.21\} = 0.5832$

20. Let X = number of men who are smokers and Y = number of women who are smokers.
$E[X] = 300(0.326) = 97.8; E[Y] = 83.4;\ \mathrm{Var}(X) = 65.9172;$
$\mathrm{Var}(Y) = 60.2148;\ E[X - Y] = E[X] - E[Y] = 14.4;$
$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 126.132;\ SD(X - Y) = \sqrt{(126.132)}$
$P\{X - Y > 0\} = P\{X - Y \ge 1\} \approx P\{Z \ge (0.5 - 14.4)/11.23\}$
$= P\{Z \ge -1.24\} = P\{Z < 1.24\} = 0.8925$

7.6 DISTRIBUTION OF THE SAMPLE VARIANCE OF A NORMAL POPULATION

Problems

2. Recall, $(n-1)S^2/\sigma^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/\sigma^2$ has a chi-squared distribution with $(n-1)$ degrees of freedom.

Now, for large n, $(n-1) \to n$, so $(n-1)S^2/\sigma^2 \to nS^2/\sigma^2 \to S^2/(\sigma^2/n)$. Also, S^2 can now be considered an average of the squared deviations from the mean with variance (σ^2/n). Hence, for large n, we can consider $S^2/(\sigma^2/n)$ to be approximately normal by the central limit theorem.

Review Problems

2. (a) $\mu = \sum_{i=1}^{4} i \times P\{X = i\} = (1)(0.1) + (2)(0.2) + (3)(0.3) + (4)(0.4) = 3$

(b) $\sigma^2 = \mathrm{Var}(X) = \sum_{i=1}^{4} i^2 P\{X = i\} - \mu^2 = (1)^2(0.1) + (2)^2(0.2) + (3)^2(0.3) + (4)^2(0.4) - (3)^2 = 1$, so $\sigma = SD(X) = 1$

(c) $E[\overline{X}] = \mu = 3$

(d) $\mathrm{Var}(\overline{X}) = \mathrm{Var}(X)/n = 1/10 = 0.10$

(e) $SD(\overline{X}) = \sqrt{0.1} = 0.3162$

4. Let X = total life of the 10 batteries.

(a) $P\{X > 2200\} = P\{\overline{X} > 2200/10\} = P\{\overline{X} > 220\}$
$\approx P\{Z > (220 - 225)/(24/\sqrt{10})\} = P\{Z > -0.66\} = P\{Z \leq 0.66\}$
$= 0.7454$

(b) $P\{X > 2350\} = P\{\overline{X} > 2350/10\} = P\{\overline{X} > 235\}$
$\approx P\{Z > (235 - 225)/(24/\sqrt{10})\} = P\{Z > 1.32\} = 1 - P\{Z \leq 1.32\}$
$= 1 - 0.9066 = 0.0934$

(c) $P\{X > 2500\} = P\{\overline{X} > 2500/10\} = P\{\overline{X} > 250\}$
$\approx P\{Z > (250 - 225)/(24/\sqrt{10})\} = P\{Z > 3.29\}$
$= 1 - P\{Z \leq 3.29\} = 1 - 0.9995 = 0.0005$

(d) $P\{2200 < X < 2350\} = P\{220 < \overline{X} < 235\}$
$= P\{\overline{X} < 235\} - P\{\overline{X} < 220\} = 0.9066 - (1 - 0.7454) = 0.652$

6. Let X = weight of the total load.

$P\{X > 16{,}000\} = P\{\overline{X} > 16{,}000/100\} = P\{\overline{X} > 160\}$
$\approx P\{Z > (160 - 155)/(28/\sqrt{100})\} = P\{Z > 1.79\}$
$= 1 - P\{Z \leq 1.79\} = 1 - 0.9633 = 0.0367$

8. Let X = salary of newly chemical engineering majors.

$P\{\overline{X} \leq 45{,}000\} \approx P\{Z \leq (45{,}000 - 54{,}000)/(5{,}000/\sqrt{12})\}$
$= P\{Z \leq -6.2354\} = 0$

10. (a) Let X_A = number of club wins against the class A teams.
Let X_B = number of club wins against the class B teams.
Let X = total number of club wins against the class A and B teams.

Thus, X_A and X_B are independent binomial random variables, and since $X = X_A + X_B$, then X is not a binomial random variable (the sum of two binomials is not binomial).

(b) X_A is a binomial random variable with parameters $n_A = 32$ and $p_A = 0.5$;
X_B is a binomial random variable with parameters $n_B = 28$ and $p_B = 0.7$.
(c) $X = X_A + X_B$
(d) $P\{X \geq 40\} = P\{X \geq 39.5\}$
$\approx P\{Z \geq (39.5 - (16 + 19.6))/(\sqrt{((32)(0.5)(0.5) + (28)(0.7)(0.3))}\}$
$= P\{Z \geq 1.05\} = 1 - P\{Z < 1.05\} = 1 - 0.8531 = 0.1469$

12. (b) $E[X_i] = (1)(0.52) + (-1X0.48) = 0.04$
(c) $\text{Var}(X_i) = (1)^2(0.52) + (-1)^2(0.48) - (0.04)^2 = 0.9984$
(d) Let X = total gain after 100 days, then we need $P\{X > 210\}$. Now,

$$P\{X > 210\} = P\left\{200 + \sum_{i=1}^{100} X_i > 210\right\} = P\left\{\sum_{i=1}^{100} X_i > 10\right\}$$
$$= P\{\overline{X} > 10/100\} \approx P\{Z > (0.1 - 0.04)/\sqrt{[(0.9984)/(100)]}\}$$
$$= P\{Z > 0.6\}$$
$$= 1 - P\{Z < 0.6\} = 1 - 0.7257 = 0.2743$$

14. (a) Not necessarily. For instance, if there is a large section of a particular course, there is the possibility that many students from this sample may be enrolled in that section.

(b) A better way to achieve this estimate of the average class size is as follows: select at random 100 different courses and then select at random a student from each of these selected courses to report the number of students in their classes. ***Note:*** There are other methods for achieving a very good estimate for the average class size.

Chapter 8 ESTIMATION

8.2 POINT ESTIMATOR OF A POPULATION MEAN

Problems

2. No. The estimate would not be unbiased since this sample will not be a random sample.

4. Let $X =$ number of hours per day spent watching TV by preschoolers in the given neighborhood.

Then, $\overline{X} = 2.5$ hours.

6. $n = (4)^2(1000) = 16{,}000$

8. Let $X =$ number of minutes patient waited to see a physician at a medical clinic.

Then, $\overline{X} = 44.0833$ minutes.

10. (a) $\mathrm{SE}(\overline{X}) = \sigma/\sqrt{n}$

(b) $\mathrm{SE}(\overline{X}) = (\sqrt{2/3})[\sigma/\sqrt{n}] = 0.8165[\sigma/\sqrt{n}]$

Comparing the standard errors for X in (a) and (b), we see that the standard error in (a) is larger. Thus, the data in (b) will yield a more precise estimator for μ.

(c) Let n_1 be the sample size in part (a) and n_2 be associated with part (b). Then, we want $\sigma/\sqrt{n_1} = (\sqrt{2})[\sigma/\sqrt{n_2}] = \sigma/\sqrt{(n_2/2)}$. That is, we want $n_1 = n_2/2$, from which $n_2 = 2n_1$. Thus, to achieve the same precision in (b) of the estimator in (a), we would need twice the sample size in (a).

8.3 POINT ESTIMATOR OF A POPULATION PROPORTION

Problems

2. Let $X =$ number of Americans who felt that the communist party will win a free election in the Soviet Union (1985) in the sample.

$n = 1325, X = 510, \hat{p} = X/n = 510/1325 - 0.3849$. Thus, the estimate for

$\mathrm{SE}\,(\hat{p}) = \sqrt{((03849)(1 - 0.3849)/1325)} = .0134$.

4. Let $X =$ number of solitaire games won in the 20 games played,

$n = 20, X = 7$.

(a) $\hat{p} = X/n = 7/20 = 0.35$

(b) $\mathrm{SE}\,(\hat{p}) = \sqrt{(0.35)(1 - 0.35) \times 20} = 0.1067$

6. Let $X =$ number of parents who are in favor of raising the driving age to 18 in the sample $n = 100$, $X = 64$.

(a) $\hat{p} = X/n = 64/100 = 0.64$

(b) SE $(\hat{p}) = \sqrt{((0.64)(1-0.64)/100)} = 0.048$

8. Let $X =$ number of female architects in the sample, $n = 500$, $X = 104$.

(a) $\hat{p} = X/n = 104/500 = 0.208$

(b) SE $(\hat{p}) - \sqrt{((0.208)(1-0.208)/500)} = 0.0182$

10. (a) $\hat{p} = 0.0233$; SE$(\hat{p}) = \sqrt{((0.0233)(1-0.0233)/1200)} = 0.0044$

(b) $\hat{p} = 0.0375$; SE$(\hat{p}) = \sqrt{((0.0375)(1-0.0375)/1200)} = 0.0055$

(c) $\hat{p} = 0.0867$; SE$(\hat{p}) = \sqrt{((0.0867)(1-0.0867)/1200)} = 0.0081$

12. If n is the sample size, we need $1/(2\sqrt{n}) = 0.1$, from which $n = 25$.

14. Let $X =$ number of full-time African-American law enforcement officers who were employed in Chicago in 1990 in the sample $n = 600, X = 87$.

(a) $\hat{p} = X/n = 87/600 = 0.145$

(b) SE$(\hat{p}) = \sqrt{[12048(0.145)(1-0.145)]} = 38.6$

8.3.1 Estimating the Probability of a Sensitive Event

Problems

2. The estimate of p, $\hat{p} = [1 - 2(10/50)] = 0.6$.

8.4 ESTIMATING A POPULATION VARIANCE

Problems

2. Let $X =$ width of a slot.

Estimate of the population mean $= \overline{X} = 8.7509$ inches.

Estimate of the population standard deviation $= S = 0.0057$ inches.

4. Let $X =$ size (in inches) of a bounce.

$$\text{With } n = 30, \text{ then } S^2 = \frac{\sum_{i=1}^{30}(X_i - \overline{X})^2}{30-1} = \frac{\sum_{i=1}^{30}(X_i)^2 - 30\overline{X}^2}{30-1}$$

$$= [136.2 - 30(52.1/30)^2]/29 = 1.5765$$

Hence, the estimate for the population standard deviation of the size of a bounce is $S = \sqrt{1.5765} = 1.2556$.

6. Estimate of the population mean $= \overline{X} = 105.70$ pounds.

Estimate of the population variance $S^2 = 30.6778$.

10. Let $X =$ burn time for the chair.

(a) Estimate of the population mean $= X = 464.14$ (°F).

(b) Estimate of the population standard deviation $= S = 19.32$ (°F).

12. Let $X =$ systolic blood pressure for a worker in the mining industry.

(a) Estimate of the population mean $= X = 132.23$.

(b) Estimate of the population standard deviation $= S = 10.49$.

(c) $P\{X > 150\} \approx P\{Z > (150 - 132.23)/10.49\} = P\{Z > 1.69\}$
$= 1 - P\{Z < 1.69\} = 1 - 0.9545 = 0.0455$

14. Let $X =$ weight of a fish weighed in the particular scale.

$\overline{X} = 5.8750$ grams and $S = 0.2301$ grams. Now,

$S^2 = \text{Var(data)} = \text{Var}(X) = 0.0530$, and $\text{Var(error)} = 0.01$.

Since, $\text{Var(data)} = \text{Var(true weight)} + \text{Var(error)}$, then

$0.0529 = \text{Var(true weight)} + 0.01$, from which

$\text{Var(true weight)} = 0.0430$ or $\text{SD(true weight)} = 0.2073$.

8.5 INTERVAL ESTIMATORS OF THE MEAN OF A NORMAL POPULATION WITH KNOWN POPULATION VARIANCE

Problems

2. That is, we can assert with "90% confidence" that the ***average*** birth weights of all boys born at certain hospital will lie between 6.6 and 7.2 pounds.

4. Let $X =$ PCB level of a fish caught in Lake Michigan.

$n = 40$, $\overline{X} = 11.480$ ppm, $\sigma = 0.86$ ppm, $\alpha = 0.05$, $\alpha/2 = 0.025$, $Z_{0.025} = 1.96$.

Since the $(1 - \alpha)$ 100% confidence interval (CI) for μ (mean PCB levels) is given by $X \pm Z_{\alpha/2}\sigma\sqrt{n}$ we have $11.480 \pm (1.96)(0.86/\sqrt{40})$. Thus, the 95% CI for μ is (11.2123, 11.7477).

6. Let X = length of time for a component to function, $n = 9, \overline{X} = 10.80$ hours, $\sigma = 3.4$ hours.

(a) $\alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96$. Since the $(1 - \alpha)$ 100% CI μ (mean life) is given by $\overline{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$, we have $10.80 \pm (1.96)(3.4/\sqrt{9})$. Thus, the 95% CI for μ is (8.5787, 13.0213).

(b) $\alpha = 0.01, \alpha/2 = 0.005, Z_{0.005} = 2.575$ (using the standard normal probability table). Since the $(1-\alpha)$ 100% CI for μ (mean life) is given by $\overline{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$, we have $10.80 \pm (2.575)(3.4/\sqrt{9})$. Thus, the 99% CI for μ is (7.8817, 13.7183).

8. Let X = test score on a certain achievement test.

$n = 324, \overline{X} = 74.6, \sigma = 11.3, \alpha - 0.1, \alpha/2 = 0.05, Z_{0.05} = 1.645$.

Since the $(1 - \alpha)$ 100% CI for μ (mean test score) is given by $\overline{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$, we have $74.6 \pm (1.645)(11.3/\sqrt{324})$. Thus, the 90% CI for μ is (73.5673, 75.6327).

10. Let X = life of the tire.

$n = 10, \overline{X} = 28,400$ miles, $\sigma = 3{,}300$ miles, $\alpha = 0.05$, $\alpha/2 = 0.025$.

Since the $(1 - \alpha)$ 100% CI for μ (mean test score) is given by $\overline{X} \pm Z_{\alpha/2}\sigma/\sqrt{n}$, we have $28{,}400 \pm (1.96)(3{,}300/\sqrt{10})$. Thus, the 95% CI for μ is (26,354.6388, 30,445.3612).

12. $\alpha = 0.1, \alpha/2 = 0.05, Z_{0.05} = 1.645, \sigma = 180, b = 40$.

Since $n \geq \left(\dfrac{2Z_{\alpha/2}\sigma}{b}\right)^2 = [(2)(1.645)(180)/40]^2 = 219.188$. Thus, you would need a sample size of at least 220.

14. $\alpha = 0.05$, $\alpha/2 = 0.025$, $Z_{0.025} = 1.96$, $\sigma = 70$, $b = 4$. Since $n \geq \left(\dfrac{2Z_{\alpha\sigma/2}}{b}\right)^2 = [(2)(1.96)(70)/4]^2 = 4705.96 \approx 4706$. Thus, you would need a sample size of at least 4706.

16. (a) We need an upper confidence bound (UCB).
$n = 10, \overline{X} = 9.8, \alpha = 0.05, Z_{0.05} = 1.645, \sigma = 3$. Since the $(1-\alpha)100\%$ UCB for μ is given by $\overline{X} + Z_{\alpha}\dfrac{\sigma}{\sqrt{n}}$, the required UCB is $[9.8 + (1.645)(3)/\sqrt{10}] = 11.3606$.

(b) We need a lower confidence bound (LCB).
$n = 10, \overline{X} = 9.8, \alpha = 0.01, Z_{0.01} = 2.575, \sigma = 3$. Since the $(1 - \alpha)$ 100% UCB for μ is given by $\overline{X} + Z_{\alpha}\frac{\sigma}{\sqrt{n}}$, the required LCB is $[9.8 - (2.575)(3)/\sqrt{10}] = 8.2394$.

8.6 INTERVAL ESTIMATORS OF THE MEAN OF A NORMAL POPULATION WITH UNKNOWN POPULATION VARIANCE

Problems

2. Let X = number of days it takes for California customers to receive their orders.

(a) $n = 12, \overline{X} = 12.25, S = 3.67, \alpha = 0.1, \alpha/2 = 0.05, t_{11,0.05} = 1.796.$

Since the $(1 - \alpha)$ 100% CI for μ (mean number of days) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $12.25 \pm (1.796)(3.67/\sqrt{}12)$. Thus, the 90% CI for μ is $(10.3472, 14.1528)$. That is, we can be 90% confident that the average number of days it takes for the California customers to receive their orders is between 10.35 and 14.15 days (answer rounded to two decimal places).

(b) $n = 12, \overline{X} = 12.25, S = 3.67, \alpha = 0.05, \alpha/2 = 0.025, t_{11,0.025} = 2.201.$

Since the $(1 - \alpha)$ 100% CI for μ (mean number of days) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $12.25 \pm (2.201)(3.67/\sqrt{}12)$. Thus, the 95% CI for μ is (9.9182, 14.5818). That is, we can be 95% confident that the average number of days it takes for the California customers to receive their orders is between 9.92 and 14.58 days. (Answer rounded to two decimal places.)

4. Let $\overline{X}$ = number of daily intercity bus riders.

(a) $\overline{X} = 54.42$

(b) $S = 7.24$

(c) $n = 12, \alpha = 0.05, \alpha/2 = 0.025, t_{11,0.025} = 2.201.$

Since the $(1 - \alpha)$ 100% CI for μ (mean number of days) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $54.42 \pm (2.201)(7.24/\sqrt{}12)$. Thus, the 95% CI for μ is (49.82, 59.02). That is, we can be 95% confident that the average number of daily intercity bus riders is between 49.82 and 59.02.

6. Let $\overline{X}$ = lifetime of a General Electric transistor.

$n = 30, \overline{X} = 1210$ hours, $S = 92$ hours.

(a) $\alpha = 0.1, \alpha/2 - 0.05, t_{29,0.05} = 1 - 699.$

Since the $(1 - \alpha)$ 100% CI for μ (mean lifetime) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $1210 \pm (1.699)(92\sqrt{}30)$. Thus, the 90% CI for μ is (1181.4622, 1238.5378). That is, we can be 90% confident that the average lifetime for the transistors is between 1181.4622 and 1238.5378 hours.

(b) $n = 30, \alpha = 0.05, \alpha/2 = 0.025, t_{29,0.025} = 2.045.$

Since the $(1-\alpha)$ 100% CI for μ (mean lifetime) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $1210 \pm (2.045)(92/\sqrt{30})$. Thus, the 95% CI for μ is (1175.6505, 1244.3495). That is, we can be 95% confident that the average lifetime for the transistors is between 1175.6505 and 1244.3495 hours.

(c) $n = 30$, $\alpha = 0.01$, $\alpha/2 = 0.005$, $t_{29,0.005} = 2.756$.
Since the $(1-\alpha)$ 100% CI for μ (mean lifetime) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $1210 \pm (2.756)(92/\sqrt{30})$. Thus, the 99% CI for μ is (1163.708, 1256.292). That is, we can be 99% confident that the average lifetime for the transistors is between 1163.708 and 1256.292 hours.

8. Let X = losing score in a Super Bowl football game.

$n = 7, \overline{X} = 18.14, S = 6.57, \alpha = 0.05, \alpha/2 = 0.025, t_{6,0.025} = 2.447$.

Since the $(1-\alpha)$ 100% CI for μ (mean losing score) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $18.14 \pm (2.447)(6.57/\sqrt{7})$. Thus, the 95% CI for μ is (12.0635, 24.2165). That is, we can be 95% confident that the average losing score in Super Bowl games is between 12.0635 and 24.2165 points.

10. Let X = time it takes to perform the task.

$n = 20, \overline{X} = 12.4, S = 3.3, \alpha = 0.05, \alpha/2 = 0.025, t_{19,0.025} = 2.093$.

Since the $(1-\alpha)$ 100% CI for μ (mean time) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $12.4 \pm (2.093)(3.3/\sqrt{20})$. Thus, the 95% CI for μ is (10.8556, 13.9444). That is, we can be 95% confident that the average time to complete the task is between 10.8556 and 13.9444 minutes.

12. Let X = height of male from the certain tribe.

(a) $n = 64, \overline{X} = 72.4, S = 2.2, \alpha = 0.05, \alpha/2 = 0.025, t_{63,0.025} \approx 2$.
Since the $(1-\alpha)$ 100% CI for μ (mean height) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $72.4 \pm (2)(2.2/\sqrt{64})$. Thus, the 95% CI for μ is (71.85, 72.95). That is, we can be 95% confident that the average height of the males in this certain tribe is between 71.85 and 72.95 inches.

***Note*:** If you use $t_{63,0.025} \approx Z_{0.025} = 1.96$, then the CI will be (71.86, 72.94).

(b) $n = 64, \overline{X} = 72.4, S = 2.2, \alpha = 0.01, \alpha/2 = 0.005, t_{63,0.025} \approx 2.66$.
Since the $(1-\alpha)$ 100% CI for μ (mean height) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $72.4 \pm (2.66)(2.2/\sqrt{64})$. Thus, the 99% CI for μ is (71.6685, 73.1315). That is, we can be 99% confident that the average height of the males in this certain tribe is between 71.6685 and 73.1315 inches.

Note: If you use $t_{63,0.005} \approx Z_{0.005} = 2.576$, then the CI will be (71.6916, 73.1084).

14. Let X = melting point of lead.

(a) $n = 20, X = 330.2, S = 15.4, \alpha = 0.05, \alpha/2 = 0.025, t_{19,0.025} = 2.093.$

Since the $(1 - \alpha)$ 100% CI for μ (mean melting point) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $330.2 \pm (2.093)(15.4/\sqrt{20})$. Thus, the 95% CI for μ is (332.9927, 337.4073). That is, we can be 95% confident that the average melting point for lead is between 332.9927 and 337.4073 degrees centigrade.

(b) $n = 20, X = 330.2, S = 15.4, \alpha = 0.01, \alpha/2 = 0.005, t_{19,0.005} = 2.861.$

Since the $(1 - \alpha)$ 100% CI for μ (mean time) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $330.2 \pm (2.861)(15.4/\sqrt{20})$. Thus, the 99% CI for μ is (320.3480, 340.0520). That is, we can be 99% confident that the average melting point for lead is between 320.3480 and 340.0520 degrees centigrade.

16. Let X = length of time for an officer on the Chicago police force.

$n = 46, X = 14.8$ years, $S = 8.2$ years.

(a) $\alpha = 0.1, \alpha/2, = 0.05, t_{45,0.05} \approx 1.684.$

Since the $(1 - \alpha)$ 100% CI for μ (mean melting point) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $14.8 \pm (1.684)(8.2/\sqrt{46})$. Thus, the 90% CI for μ is (12.764, 16.7925). That is, we can be 90% confident that the average time for the officers on the Chicago police force is between 12.764 and 16.7925 years.

(b) $\alpha = 0.05, \alpha/2 = 0.025, t_{45,0.025} \approx 2.021.$

Since the $(1 - \alpha)$ 100% CI for μ (mean time) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $14.8 \pm (2.021)(8.2/\sqrt{46})$. Thus, the 95% CI for μ is (12.3566, 17.2434). That is, we can be 95% confident that the average time for the officers on the Chicago police force is between 12.3566 and 17.2434 years.

(c) $\alpha = 0.01, \alpha/2 = 0.005, t_{45,0.005} \approx 2.704.$

Since the $(1 - \alpha)$ 100% CI for μ (mean time) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $14.8 \pm (2.704)(8.2/\sqrt{46})$. Thus, the 99% CI for μ is (11.5308, 18.0692). That is, we can be 99% confident that the average times for the officers on the Chicago police force is between 11.5308 and 18.0692 years.

18. Let X = price of the crude oil stock.

$n = 20, \overline{X} = 17.465, S = 0.4252, \alpha = 0.05, \alpha/2 = 0.025, t_{19,0.025} = 2.093.$

Since the $(1 - \alpha)$ 100% CI for μ (mean price) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $17.465 \pm (2.093)(0.4252/\sqrt{20})$. Thus, the 95% CI for μ is (17.2660, 17.6640).

That is, we can be 95% confident that the average price of the crude oil stock is between 17.266 and 17.664.

20. Let X = number of days of school missed by a student.

(a) $n = 50, \overline{X} = 8.4, S = 5.1, \alpha = 0.05, \alpha/2 - 0.025, t_{49,0.025} \approx 2.01.$
Since the $(1-\alpha)$ 100% CI for μ (mean missed days) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $8.4 \pm (2.01)(5.1/\sqrt{50})$. Thus, the 95% CI for μ is (6.9503, 9.8497). That is, we can be 95% confident that the average number of school days missed by the students is between 6.9503 and 9.8497.

(b) We need to compute the UCB for X. Since the $(1 - \alpha)$ 100% UCB is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $8.4 + (1.675)(5.1/\sqrt{50})$. Thus, the 95% UCB for X is 9.6081. Thus, with 95% confidence, we can state that the average number of days missed is less than 9.6081.

22. Let X = number of days it takes for California customers to receive their orders.

$n = 12, \overline{X} = 12.25, S = 3.67, \alpha = 0.05, t_{11,0.05} = 1.796.$

Since the $(1 - \alpha)$ 100% UCB is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $12.25 + (1.796)(3.67/\sqrt{12})$. Thus, the 95% UCB for X is 14.1528. So, with 95% confidence, we can state that the average number of days it takes for the California customers to receive their orders is less than 14.1528 days.

24. Let X = amount of carbon monoxide measured.

$n = 7, \overline{X} = 100.7, S = 5.982, \alpha = 0.01, t_{6,0.01} = 3.143.$

Since the $(1-\alpha)$ 100% UCB is given by $\overline{X} + t_{n-1,\alpha} \times \frac{S}{\sqrt{n}}$, then the 99% UCB will be $100.7+3.143\times5.982/\sqrt{7}$. That is, the UCB is 107.8063. Thus, with 99% confidence, the inspector can state that the average amount of carbon monoxide measured less than 107.8065 parts per million. (Hopefully, this will calm the concerns of the group.)

8.7 INTERVAL ESTIMATORS OF A POPULATION PROPORTION

Problems

2. Let X = number of people who suffer an additional heart attack within 1 year of their first heart attack.

(a) $n = 300, X = 46, \hat{p} = 46/300 = 0.1533, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96.$

Since the $(1-\alpha)$ 100% CI for p (proportion of second heart attack within one year) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.1533 \pm (1.96)(0.0208)$. Thus, the 95% CI for p is (0.1125, 0.1941). That is, we can be 95% confident that the proportion of second heart attacks within 1 year is between 11.25% and 19.41%.

(b) Let X = number of people who suffer an additional heart attack within 1 year of their first heart attack.
$n = 300, X = 92, \hat{p} = 92/300 = 0.3067, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96$.
Since the $(1-\alpha)$ 100% CI for p (proportion of second heart attack within 1 year) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.3067 \pm (1.96)(0.0266)$. Thus, the 95% CI for p is (0.2546, 0.3588). That is, we can be 95% confident that the proportion of second heart attacks within 1 year is between 25.46% and 35.88%.

4. $n = 1200, \hat{p} = 0.57, \alpha = 0.01, \alpha/2 = 0.005, Z_{0.005} = 2.575$.
Since the $(1-\alpha)$ 100% CI for p (proportion of population that favored Reagan at the time of the poll) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.57 \pm (2.75)(0.0143)$. Thus, the 99% CI for p is (0.5332, 0.6068). That is, we can be 99% confident that the proportion of the population that favored Reagan at the time of the 1980 poll was between 53.32% and 60.68%.

6. Let X = number of recent science PhDs who are optimistic.

(a) $n = 100, X = 42, \hat{p} = 42/100 = 0.42, \alpha = 0.1, \alpha/2 = 0.05, Z_{0.05} = 1.645$.
Since the $(1-\alpha)$ 100% CI for p (proportion of recent science PhDs who are optimistic) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.42 \pm (1.645)(0.0494)$. Thus, the 90% CI for p is (0.3387, 0.5013). That is, we can be 90% confident that the proportion of recent science PhDs who are optimistic is between 33.87% and 50.13%.

(b) $n = 100$, $X = 42$, $\hat{p} = 42/100 = 0.42$, $\alpha = 0.01$, $\alpha/2 = 0.005$, $Z_{0.005} = 2.575$.
Since the $(1-\alpha)$ 100% CI for p (proportion of recent science PhDs who are optimistic) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.42 \pm (2.575)(0.0494)$. Thus, the 99% CI for p is (0.2928, 0.5472). That is, we can be 99% confident that the proportion of recent science PhDs who are optimistic is between 29.28% and 54.72%.

8. Let X = number of solitaire games won.
$n = 20, X = 7, \hat{p} = 7/20 = 0.35, \alpha = 0.1, \alpha/2 = 0.05, Z_{0.005} = 1.645$.
Since the $(1-\alpha)$ 100% CI for p (proportion of solitaire games won) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.35 \pm (1.645)(0.1067)$. Thus, the 90% CI for p is (0.1745, 0.5255). That is, we can be 90% confident that the probability of winning at solitaire will be between 17.45% and 52.55%.

10. Let X = number of cups of coffee that had less than the specified amount.

$n = 100, X = 9, \hat{p} = 9/100 = 0.09, \alpha = 0.1, \alpha/2 = 0.05, Z_{0.05} = 1.645.$

Since the $(1-\alpha)$ 100% CI for p(proportion of cups of coffee that were underfilled) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.09 \pm (1.645)(0.0286)$. Thus, the 90% CI for p is (0.0430, 0.1370). That is, we can be 90% confident that the proportion of cups of coffee that had less than the specified amount will be between 4.3% and 13.7%.

12. Let X = number of male authors.

$n = 300, X = 117, \hat{p} = 117/300 = 0.39, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96.$

Since the $(1-\alpha)$ 100% CI for p (proportion of male authors) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.39 \pm (1.96)(0.0282)$. Thus, the 95% CI for p is (0.3347, 0.4453). That is, we can be 95% confident that the proportion of male authors will be between 33.47% and 44.53%.

14. Let X = number of male psychologists.

$n = 1000, X = 457, \hat{p} = 457/1000 = 0.457, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96.$

Since the $(1-\alpha)$ 100% CI for p (proportion of male authors) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.457 \pm (1.96)(0.0158)$. Thus, the 95% CI for p is (0.4260, 0.4880). That is, we can be 95% confident that the proportion of male psychologists will be between 42.6% and 48.8%.

16. Let X = number of people who were in favor of the war against Iraq on January 22, 2004.

(a) $n = 600, X = 450, \hat{p} = 450/600 = 0.75, \alpha = 0.1, \alpha/2 = 0.05, Z_{0.05} = 1.645.$

Since the $(1-\alpha)$ 100% CI for p(proportion of people who favored the war) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.75 \pm (1.645)(0.0177)$. Thus, the 90% CI for p is (0.7210, 0.7791). That is, we can be 90% confident that the proportion of the population who favored the war against Iraq on January 22, 2004 was between 72.10% and 77.91%.

(b) $n = 600, X = 450, \hat{p} = 450/600 = 0.75, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96.$

Since the $(1-\alpha)$ 100% CI for p (proportion of people who favored the war) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.75 \pm (1.96)(0.0177)$. Thus, the 95% CI for p is (0.7153, 0.7847). That is, we can be 95% confident that the proportion of the population who favored the war against Iraq on January 22, 2004, was between 71.53% and 78.47%.

(c) $n = 600, X = 450, \hat{p} = 450/600 = 0.75, \alpha = 0.01, \alpha/2 = 0.005, Z_{0.005} = 2.575$.

Since the $(1 - \alpha)$ 100% CI for p (proportion of people who favored the war) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}$, we have $0.75 \pm (2.575)(0.0177)$. Thus, the 99% CI for p is (0.7044, 0.7956). That is, we can be 99% confident that the proportion of the population who favored the war against Iraq on January 22, 2004, was between 70.44% and 79.56%.

18. No. The error of ±4% if applied to both estimates will be 47% ± 4% and 53% ± 4%. That is, we have intervals of 43%, 51% and 49%, 57%. Since these intervals overlap, we cannot say with certainty that candidate A is the current choice.

20. $\alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96, b = 2(0.01) = 0.02$.

Since $n > \left(\frac{Z_{\alpha/2}}{b}\right)^2$, $n > (1.96/0.02)^2 = 9{,}604$. That is, a sample size of at least 9,604 is needed if we want to estimate the required proportion to within 0.01 with 95% confidence.

22. Let X = number of male psychologists.

$n = 1000, X = 457, \hat{p} = 457/1000 = 0.457, \alpha = 0.05, Z_{0.05} = 1.645$.

Since the $(1 - \alpha)$ 100% LCB for p (proportion of male authors) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}$, we have $0.457 - (1.645)(0.0158)$. Thus, the 95% LCB for p is 0.4310. That is, we can be 95% confident that the proportion of male psychologists will be more than 43.1%.

24. We need to compute a LCB.

Let X = number of consumers who are satisfied with the product.

$n = 500, \hat{p} = 0.92, \alpha = 0.05, Z_{0.05} = 1.645$.

Since the $(1 - \alpha)$ 100% LCB for p (proportion of satisfied customers) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}$, we have $0.92 - (1.645)(0.0121)$. Thus, the 95% LCB for p is 0.9001. That is, we can be 95% confident that the proportion of satisfied customers will be more than 90.01%. Thus, let x be equal to 90.01%.

When $\alpha = 0.1, Z_{0.1} = 1.28$, from which $x = 90.45\%$.

26. (a) Let X = number who favor the war at the time of the poll (see problem 16).

$n = 600, X = 450, \hat{p} = 450/600 = 0.75, \alpha = 0.05, Z_{0.05} = 1.645$.

Since the $(1-\alpha)$ 100% UCB for p (proportion of the population who favored the war) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}$, we have $0.75 + (1.645)(0.0177)$. Thus, the 95% UCB for p is 0.7791. That is, we can be 95% confident that the proportion of the population who favored the war will be less than 77.91%.

(b) $n = 600, X = 450, \hat{p} = 450/600 = 0.75, \alpha = 0.05, Z_{0.05} = 1.645$.
Since the $(1-\alpha)$ 100% LCB for p (proportion of the population who favored the war) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.75 - (1.645)(0.0177)$. Thus, the 95% LCB for p is 0.7209. That is, we can be 95% confident that the proportion of the population who favored the war will be more than 72.09%.

28. We need to compute the LCB and UCB.

(a) Let X = number of Los Angeles residents who favor strict gun control legislation.
$n = 100, X = 64, \hat{p} = 64/100 = 0.64, \alpha = 0.05, Z_{0.05} = 1.645$.
Since the $(1-\alpha)$ 100% LCB for p (proportion of Los Angeles population who favored the legislation) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.64 - (1.645)(0.048)$. Thus, the 95% LCB for p is 0.5610. That is, with 95% confidence, we can state that more than 56.10% of all Los Angeles residents favor gun control.

(b) Let X = number of Los Angeles residents who favor strict gun control legislation.
$n = 100, X = 64, \hat{p} = 64/100 = 0.64, \alpha = 0.05, Z_{0.05} = 1.645$.
Since the $(1-\alpha)$ 100% UCB for p (proportion of Los Angeles population who favored the legislation) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.64 + (1.645)(0.048)$. Thus, the 95% UCB for p is 0.7190. That is, with 95% confidence, we can state that less than 71.90% of all Los Angeles residents favor gun control.

Review Problems

2. Let X = weight of a ball bearing.

(a) $\sigma = 0.5, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96, b = 2(0.1) = 0.2$.
Since $n \geq \left(\frac{2Z_{\alpha/2}\sigma}{b}\right)^2$, $n \geq (2 \times 1.96 \times 0.5/0.2)^2 = 96.04 \approx 97$.

(b) $\sigma = 0.5, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96, b = 2(0.01) = 0.02$.
Since $n \geq \left(\frac{2Z_{\alpha/2}\sigma}{b}\right)^2$, $n \geq (2 \times 1.96 \times 0.5/0.02)^2 = 9604$.

(c) $n = 8, \overline{X} = 3.938, S = 0.403, \alpha = 0.05, \alpha/2 = 0.025, t_{7,0.025} = 2.365$.
Since the $(1-\alpha)$ 100% CI for μ (mean weight) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}}\frac{S}{\sqrt{n}}$, we have $3.938 \pm (2.365)(0.403/\sqrt{8})$. Thus, the 95% CI for μ is (3.601, 4.275). That is, we can be 95% confident that the average weight of the ball bearings is between 3.601 and 4.275 units.

4. The 90% CIs for the variables all, male, and female are given below in the table.

Variable	n	$\overline{X}$	S	**Std. Error**	**90% CI**
All	30	195.57	12.13	2.21	(191.92, 199.21)
Male	15	191.00	12.80	3.30	(185.18, 196.82)
Female	15	200.13	9.81	2.53	(195.67, 204.60)

Thus, we are 90% confident that the average blood cholesterol levels for both males and females lie between the intervals given in the table.

6. $\alpha = 0.1, \alpha/2 = 0.05, Z_{0.05} = 1.645, b = 2(0.02) = 0.04.$

Since $n > \left(\frac{Z_{\alpha/2}}{b}\right)^2$, $n > (1.645/0.04)^2 = 1691.2656$. That is, a sample size of at least 1692 is needed if we want to estimate the required proportion to within 0.02 (2%) with 90% confidence.

8. Let X = length (in minutes) of a rapid eye movement (REM) interval.

$n = 7, X = 41.43, S = 7.32, \alpha = 0.01, \alpha/2 = 0.005, t_{6,0.005} = 3.707.$

Since the $(1 - \alpha)$ 100% CI for μ (mean length of an interval) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $41.43 \pm (3.707)(7.32/\sqrt{7})$. Thus, the 99% CI for μ is (31.1738, 51.6862). That is, we can be 99% confident that the average length of the REM interval is between 31.1738 and 51.6862 minutes.

10. Let X = number of farm workers who were in favor of unionizing.

$n = 300, X = 144, \hat{p} = 144/300 = 0.48, \alpha = 0.1, \alpha/2 = 0.05, Z_{0.05} = 1.645.$

Since the $(1-\alpha)$ 100% CI for p (proportion of farm workers who favored unionizing) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.48 \pm (1.645)(0.0288)$. Thus, the 90% CI for p is (0.4326, 0.5274). That is, we can be 90% confident that the proportion of the population of farm workers who favored unionizing is between 43.26% and 52.74%.

12. (a) $n = 9, \overline{X} = 35, \sigma = 3, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96.$
Since the $(1 - \alpha)$ 100% CI for μ is given by $\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, we have $35 \pm (1.96)(3/\sqrt{9})$. Thus, the 95% CI for μ is (33.04, 36.96).

(b) $n = 9, \overline{X} = 35, \sigma = 6, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96.$
Since the $(1 - \alpha)$ 100% CI for μ is given by $\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, we have $35 \pm (1.96)(6/\sqrt{9})$. Thus, the 95% CI for μ is (31.08, 38.92).

(c) $n = 9, \overline{X} = 35, \sigma = 12, \alpha = 0.05, \alpha/2 = 0.025, Z_{0.025} = 1.96.$
Since the $(1 - \alpha)$ 100% CI for μ is given by $\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, we have $35 \pm (1.96 \times 12/\sqrt{9})$. Thus, the 95% CI for μ is (27.16, 42.84).

14. Let X = IQ scores of students at the large eastern university.

(a) $n = 18, \overline{X} = 133.22, S = 10.21, \alpha = 0.1, \alpha/2 = 0.05, t_{17,0.05} = 1.74$.
Since the $(1-\alpha)$ 100% CI for μ (mean IQ score for the students at this large eastern university) is given by $\overline{X} \pm t_{n-1,\alpha/\sqrt{n}} \frac{S}{\sqrt{n}}$, we have $133.22 \pm (1.74)(10.21/\sqrt{8})$. Thus, the 90% CI for μ is (129.0327, 137.4073). That is, we can be 90% confident that the average IQ score for these students is between 129.0327 and 137.4073.

(b) $n = 18, \overline{X} = 133.22, S = 10.21, \alpha = 0.05, \alpha/2 = 0.025, t_{17,0.025} = 2.11$.
Similarly, the 95% CI is (128.1422, 138.2978).

(c) $n = 18, \overline{X} = 133.22, S = 10.21, \alpha = 0.01, \alpha/2 = 0.005, t_{17,0.005} = 2.898$.

Similarly, the 99% CI is (126.2459, 140.1941).

16. $\alpha = 0.01, \alpha/2 = 0.005, Z_{0005} = 2.575, b = 0.03$.

Since $n > \left(\frac{Z_{\alpha/2}}{b}\right)^2$, $n > (2.575/0.03)^2 = 7367.3611$. That is, a sample size of at least 7368 is needed if we want the 99% CI to have a length of at most 0.03.

18. For a given $n = 1600$ and α, the largest possible margin of error is $\pm\frac{Z_{\alpha/2}}{2\sqrt{n}} = \pm\frac{Z_{\alpha/2}}{2\sqrt{1600}} = \pm\frac{Z_{\alpha/2}}{80}$.

20. Let X = number of secondary school teachers who are female.

(a) $n = 1000, X = 518, \hat{p} = 518/1000 = 0.518, \alpha = 0.1, Z_{0.1} = 1.28$.
Since the $(1-\alpha)$ 100% UCB for p (proportion of female secondary school teachers) is given by $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, we have $0.518 + (1.28 \times 0.0158)$. Thus, the 90% UCB for p is 0.5382. That is, with 90% confidence, less than 53.82% of the secondary school teachers are females.

(b) $n = 1000, X = 518, \hat{p} = 518/1000 = 0.518, \alpha = 0.05, Z_{0.05} = 1.645$.
Similarly, the 95% UCB is 0.5440.

(c) $n = 1000, X = 518, \hat{p} = 518/1000 = 0.518, \alpha = 0.01, Z_{0.01} = 2.33$.

Similarly, the 99% UCB is 0.5548.

22. Let X = price of a house in the given city.

$n = 9, \overline{X} = \$222{,}000, S = \$12{,}000, \alpha = 0.05, t_{8,0.05} = 1.86$.

Since the $(1-\alpha)$ 100% UCB is given by $\overline{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$, we have $222{,}000 + (1.86)(12{,}000/\sqrt{9})$. Thus, the 95% UCB for X is 229,440. Thus, with 95% confidence, we can state that the average price of all recently sold houses in the given city is less than \$229,440.

Chapter 9 TESTING STATISTICAL HYPOTHESES

9.2 HYPOTHESIS TESTS AND SIGNIFICANCE LEVELS

2. Let μ be the mean time it takes the drug (somatriptan) to enter the bloodstream. Then,

H_0: $\mu \geq 10$ minutes against H_1: $\mu < 10$ minutes

4. (a) No – this is not generally true. This can only be true if the test statistic in H_0: $\mu \leq 1.5$ against H_1: $\mu > 1.5$ was a relatively small value.

(b) Yes. This is true since the test statistic will now be in the "do not reject" region for H_0: $\mu > 1.5$ against H_1: $\mu \leq 1.5$. Since the test statistic must be relatively *large* in order to reject the first null hypothesis, the test statistic cannot be *small* enough to reject the second null hypothesis.

9.3 TESTS CONCERNING THE MEAN OF A NORMAL POPULATION: CASE OF KNOWN VARIANCE

Problems

2. $n = 10,\ \overline{X} = 11.72,\ \sigma = 2,\ SE(\overline{X}) = 0.6325,\ \alpha = 0.05,\ Z_{\alpha/2} = 1.96.$
H_0: $\mu = 11.2$ against H_1: $\mu \neq 11.2$
Test statistic: $Z = 0.8222$
p value $= 2 \times 0.2061 = 0.4122$
Conclusion: Since $|Z| = 0.8222 < Z_{\alpha/2} = 1.96$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average PCB concentration is equal to 11.2 ppm. ***Note*:** This decision is also confirmed by the p value.

4. (a) $n = 36,\ \overline{X} = 100,\ \sigma = 5,\ SE(\overline{X}) = 0.833,\ \alpha = 0.05,\ Z_{\alpha/2} = 1.96.$
H_0: $\mu = 105$ against H_1: $\mu \neq 105$
Test statistic: $Z = -6$
p value $= 2P\{Z > 6\} \approx 0$
Conclusion: Since $|Z| = 6 > Z_{\alpha/2} = 1.96$, reject H_0.
***Note*:** When $\alpha = 0.01$, $Z_{\alpha/2} = 2.575$, so we will still reject H_0. This conclusion is also supported by the p value.

(b) $n = 36,\ \overline{X} = 100,\ \sigma = 10,\ SE(\overline{X}) = 1.67,\ \alpha = 0.05,\ Z_{\alpha/2} = 1.96.$
H_0: $\mu = 105$ against H_1: $\mu \neq 105$
Test statistic: $Z = -3$
p value $= 2P\{Z > 3\} \approx 0.0026$
Conclusion: Since $|Z| = 3 > Z_{\alpha/2} = 1.96$, reject H_0.

***Note*:** When $\alpha = 0.01,\ Z_{\alpha/2} = 2.575$, so we will still reject H_0. This conclusion is also supported by the p value.

(c) $n = 36,\ \overline{X} = 100,\ \sigma = 15,\ SE(\overline{X}) = 2.5,\ \alpha = 0.05,\ Z_{\alpha/2} = 1.96.$
H_0: $\mu = 105$ against H_1: $\mu \neq 105$
Test statistic: $Z = -2$
p value $= 2P\{Z > 2\} = 2(1 - 0.9772) = 0.0456$
Conclusion: Since $|Z| = 2 > Z_{\alpha/2} = 1.96$, reject H_0.
***Note*:** When $\alpha = 0.01,\ Z_{\alpha/2} = 2.575$, so we will not reject H_0. This conclusion is also supported by the p value.

6. $n = 7,\ \overline{X} = 14.7429,\ \sigma = 2,\ SE(\overline{X}) = 0.7559.$
H_0: $\mu = 14$ against H_1: $\mu \neq 14$
Test statistic: $Z = 0.9827$
p value $= 2P\{Z > 0.9827\} = 2(1 - 0.8371) = 0.3258$

8. $n = 10,\ \overline{X} = 124.2,\ \sigma = 9,\ SE(\overline{X}) = 2.846.$
H_0: $\mu = 122$ against H_1: $\mu \neq 122$
Test statistic: $Z = 0.7730$
p value $= 2P\{Z > 0.773\} = 2(1 - 0.7794) = 0.4412$
Conclusion: Since the p value $= 0.4412$ is large, H_0 is not rejected. That is, the data are not inconsistent with the null hypothesis that the average is equal to 122.

10. (a) $n = 64,\ \overline{X} = 52.5,\ \sigma = 20,\ SE(\overline{X}) = 2.5.$
H_0: $\mu = 50$ against H_1: $\mu \neq 50$
Test statistic: $Z = 1$
p value $= 2P\{Z > 1\} = 2(1 - 0.8413) = 0.3174$

(b) $n = 64,\ \overline{X} = 55,\ \sigma = 20,\ SE(\overline{X}) = 2.5.$
H_0: $\mu = 50$ against H_1: $\mu \neq 50$
Test statistic: $Z = 2$
p value $= 2P\{Z > 2\} = 2(1 - 0.9772) = 0.0456$

(c) $n = 64,\ \overline{X} = 57.5,\ \sigma = 20,\ SE(\overline{X}) = 2.5.$
H_0: $\mu = 50$ against H_1: $\mu \neq 50$
Test statistic: $Z = 3$
p value $= 2P\{Z > 3\} = 2(1 - 0.9987) = 0.0026$

12. $n = 1,\ \overline{X} = 32,\ \sigma = 4.7,\ SE(\overline{X}) = 4.7.$
H_0: $\mu = 16.2$ against H_1: $\mu \neq 16.2$
Test statistic: $Z = 3.3617$
p value $= 2P\{Z > 3.3617\} = 2(1 - 0.9996) = 0.0008$
Conclusion: Since the p value $= 0.0008$ is small, H_0 is rejected. That is, the data are consistent with the alternative hypothesis that the average number of cases of childhood cancer in the communities near high-level electromagnetic field is not equal to 16.2.

14. $n=10,\ \sigma=0.8,\ SE(\overline{X})=0.253,\ \alpha=0.01,\ Z_{\alpha/2}=2.575$. Thus, the significance level of 0.01 test of $H_0: \mu = 14$ against $H_1: \mu \neq 14$ is to reject H_0 if $\dfrac{\sqrt{10}}{0.8}\left|\overline{X} - 14\right| \geq Z_{0.005}$ from which $\overline{X} \geq 14 + 0.6512$ or $\overline{X} \leq 14 - 0.6512$. That is, if either $\overline{X} \geq 14.6512$ or $\overline{X} \leq 13.3488$, H_0 will be rejected. Thus, $P\{\text{rejecting } H_0\} = P\{\overline{X} \geq 14.6512\} + P\{\overline{X} \leq 13.3488\} \Rightarrow P\{Z \geq (14.6512 - 14.8)/0.253\} + P\{Z \leq (13.3488 - 14.8)/0.253\} = P\{Z \geq -0.5881\} + P\{Z \leq -5.736\} = 0.7224$.

9.3.1 One-Sided Tests

Problems

2. (a) $n = 20,\ \overline{X} = 105,\ \sigma = 5,\ SE(\overline{X}) = 1.1180$.
$H_0: \mu \leq 100$ against $H_1: \mu > 100$
Test statistic: $Z = 4.4722$
p value $= P\{Z > 4.4722\} \approx 0$

(b) $n = 20,\ \overline{X} = 105,\ \sigma = 10,\ SE(\overline{X}) = 2.2361$.
$H_0: \mu \leq 100$ against $H_1: \mu > 100$
Test statistic: $Z = 2.2361$
p value $= P\{Z > 2.24\} = (1 - 0.9875) = 0.0125$

(c) $n = 20,\ \overline{X} = 105,\ \sigma = 15,\ SE(\overline{X}) = 3.3541$.
$H_0: \mu \leq 100$ against $H_1: \mu > 100$
Test statistic: $Z = 1.4907$
p value $= P\{Z > 1.49\} = (1 - 0.9319) = 0.0681$

4. $n = 10,\ \sigma = 0.05,\ \overline{X} = 8.399,\ SE(\overline{X}) = 0.0158,\ \alpha = 0.05,\ Z_\alpha = 1.645$.

(a) $H_0: \mu \leq 8.4$ (b) $H_1: \mu > 8.4$

(c) Test statistic: $Z = -0.0632$
Conclusion: Since $Z = -0.0632 < Z_\alpha = 1.645$, do not reject H_0. Should not use the solution.

(d) p value $= P\{Z > -0.0632\} = 0.5239$

6. $n = 36,\ \overline{X} = 9.1,\ \sigma = 2.4,\ SE(\overline{X}) = 0.4,\ \alpha = 0.05,\ Z_\alpha = 1.645$.
$H_0: \mu \leq 8.2$ against $H_1: \mu > 8.2$
Test statistic: $Z = 2.25$
Conclusion: Since $Z = 2.25 > Z_\alpha = 1.645$, H_0 is rejected. That is, the data are consistent with the alternative hypothesis that the average diameter of the tomatoes grown by the farmer is greater than 8.2 cm.

8. (a) $n = 12,\ \overline{X} = 109.5,\ \sigma = 4,\ SE(\overline{X}) = 1.1547,\ \alpha = 0.05,\ Z_\alpha = 1.645$.
$H_0: \mu \leq 100$ against $H_1: \mu > 100$
Test statistic: $Z = 8.2272$
Conclusion: Since $Z = 8.2272 > Z_\alpha = 1.645$, H_0 is rejected. That is, the data are consistent with the alternative hypothesis that the average is greater than 100.

(b) $n = 12,\ \overline{X} = 109.5,\ \sigma = 4,\ SE(\overline{X}) = 1.1547,\ \alpha = 0.01,\ Z_\alpha = 2.33.$
H_0: $\mu \leq 100$ against H_1: $\mu > 100$
Test statistic: $Z = 8.2272$
Conclusion: Since $Z = 8.2272 > Z_\alpha = 2.33,\ H_0$ is rejected. That is, the data are consistent with the alternative hypothesis that the average is greater than 100.

(c) p value $= P\{Z > 8.2272\} = 0$. The p value is the same for both (a) and (b).

10. Rejecting the null hypothesis in the first case and accepting the alternative implies basically that at the level of significance α, the true mean is not less than or equal to the mean specified in the null hypothesis. Thus, it seems reasonable that the level of significance for both tests is the same.

9.4 THE *t* TEST FOR THE MEAN OF A NORMAL DISTRIBUTION: CASE OF UNKNOWN VARIANCE

Problems

2. $n = 8,\ \overline{X} = 2081.7,\ S = 120.2,\ SE(\overline{X}) = 42.5,\ t_{7,0.025} = 2.365,\ t_{7,0.005} = 3.499.$

(a) H_0: $\mu = 2000$ against H_1: $\mu \neq 2000$

(b) Test statistic: $T = 1.92$
Conclusion: Since $T = 1.92 < t_{7,0.025} = 2.365,\ H_0$ is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the average is equal to \$2, 000.

(c) The same conclusion as in (b) is reached at the 1% level of significance.

(d) p value $= 0.096$

4. (a) $n = 12,\ \overline{X} = 283.83,\ S = 16.60,\ SE(\overline{X}) = 4.79,\ t_{11,0.05} = 1.796.$
H_0: $\mu = 300$ against H_1: $\mu \neq 300$
Test statistic: $T = -3.38$
Conclusion: Since $|T| = 3.38 > t_{11,0.05} = 1.796, H_0$ is rejected at the 10% level of significance. That is, the data are consistent with the alternative hypothesis that the average number of lunches served daily is not equal to 300.

(b) $t_{11,0.025} = 2.201$. Same decision as in (a).

(c) $t_{11,0.005} = 3.106$. Same decision as in (a).

6. $n = 36,\ \overline{X} = 22.5,\ S = 3.1,\ SE(\overline{X}) = 0.5167,\ t_{35,0.025} \approx 2.0315.$
H_0: $\mu = 24$ against H_1: $\mu \neq 24$
Test statistic: $T = -2.903$
Conclusion: Since $|T| = 2.903 > t_{35,0.025} = 2.0315,\ H_0$ is rejected at the 5% level of significance. That is, the data are consistent with the alternative hypothesis that the

average number of pushups in 60 seconds done by entering high-school freshmen is not equal to 24.

8. $n = 36,\ \overline{X} = 4.5,\ S = 0.9,\ SE(\overline{X}) = 0.15,\ t_{35,0.005} \approx 2.727.$
$H_0\colon \mu = 5$ against $H_1\colon \mu \neq 5$
Test statistic: $T = -3.3333$
Conclusion: Since $|T| = 3.3333 > t_{35,0.005} = 2.727,\ H_0$ is rejected at the 1% level of significance. That is, the data are not consistent with the null hypothesis that the average weight gain is equal to 5 grams.

10. No solution presented since the Sunday major league baseball scores will change from week to week.

12. $n = 10,\ \overline{X} = 91,050,\ S = 7,797,\ SE(\overline{X}) = 2,466,\ t_{9,0.1} = 1.383,\ t_{9,0.05} = 1.833.$

(a) $H_0\colon \mu \leq 87{,}800$ against $H_1\colon \mu > 87{,}800$
Test statistic: $T = 1.32$
Conclusion: Since $T = 1.32 < t_{9,0.1} = 1.383,\ H_0$ is not rejected at the 10% level of significance. That is, the data are not inconsistent with the null hypothesis that the average salary is less than or equal to $87,800.

(b) Same conclusion as in (a) at the 5% level of significance.

14. $n = 20,\ \overline{X} = 246.15,\ S = 8.66,\ SE(\overline{X}) = 1.94,\ t_{19,0.05} = 1.729,\ t_{19,0.05} = 2.539.$

(a) $H_0\colon \mu \geq 250$ against $H_1\colon \mu < 250$
Test statistic: $T = -1.99$
Conclusion: Since $T = -1.99 < -t_{19,0.05} = 1.729,\ H_0$ is rejected at the 5% level of significance. That is, the data are consistent with the alternative hypothesis that the average lifetime of the batteries is less than 250 hours.

(b) Since $T = -1.99$ is not less than $-t_{19,0.01} = -2.539,\ H_0$ is not rejected at the 1% level of significance. That is, the data are not inconsistent with the null hypothesis that the average lifetime of the batteries is greater than or equal to 250 hours.

16. $n = 12,\ \overline{X} = 29.792,\ S = 2.414,\ SE(\overline{X}) = 0.697,\ t_{11,0.05} = 1.796.$

(a) $H_0\colon \mu \leq 30$ against $H_1\colon \mu > 30$
Test statistic: $T = -0.30$
Conclusion: Since $T = -0.30 < t_{11,0.05} = 1.796,\ H_0$ is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the average stress resistance of the plastic sheets is less than or equal to 30 pounds per square inch.

(b) H_0: $\mu \geq 30$ against H_1: $\mu < 30$
Test statistic: $T = -0.30$
Conclusion: Since $T = -0.30$ is not less than $-t_{11,0.05} = -1.796$, H_0 is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the average stress resistance of the plastic sheets is greater than or equal to 30 pounds per square inch.

18. $n = 20$, $\overline{X} = 530.3$, $S = 31.87$, $SE(\overline{X}) = 7.13$, $t_{19,0.05} = 1.729$.
H_0: $\mu \geq 542$ against H_1: $\mu < 542$
Test statistic: $T = -1.64$
Conclusion: Since $T = -1.64$ is not less than $-t_{19,0.05} = -1.729$, H_0 is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the average verbal SAT score is greater than or equal to 542.

9.5 HYPOTHESIS TESTS CONCERNING POPULATION PROPORTIONS

Problems

2. $n = 450$, $x = 294$, $p_0 = 0.6$, $\mu = np_0 = 270$, $\sigma = \sqrt{np_0(1-p_0)} = 10.3923$, $\alpha = 0.05$.
H_0: $p \geq 0.6$ against H_1: $p < 0.6$
Test statistic: $X = 294$
p value $= P\{X \leq 294\} = P\{X \leq 294.5\}$
$\approx P\{Z \leq (294.5 - 270)/10.3923\} = P\{Z \leq 2.3575\} = 0.9909$
Conclusion: Since p value $= 0.9909 > \alpha = 0.05$, H_0 is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the proportion (of immigrants working in the health profession in the United States for more than 1 year who feel that they are underemployed with respect to their training) is greater than or equal to 60%.
***Note*:** Same conclusion at the 1% level of significance.

4. (a) $n = 100$, $x = 50$, $p_0 = 0.6$, $\mu = np_0 = 60$, $\sigma = \sqrt{np_0(1-p_0)} = 4.8990$, $\alpha = 0.1$.
H_0: $p \geq 0.6$ against H_1: $p < 0.6$
Test statistic: $X = 50$
p value $= P\{X \leq 50\} = P\{X \leq 50.5\}$
$\approx P\{Z \leq (50.5 - 60)/4.899\} = P\{Z \leq -1.9392\} = 0.0262$
Conclusion: Since p value $= 0.0262 < \alpha = 0.1$, H_0 is rejected at the 10% level of significance. That is, the data are consistent with the alternative hypothesis that the proportion of voters in favor of restructuring the city government is less than 60%.

(b) $n = 100$, $x = 50$, $p_0 = 0.6$, $\mu = np_0 = 60$, $\sigma = \sqrt{np_0(1 - p_0)} = 4.8990$, $\alpha = 0.05$.
H_0: $p \geq 0.6$ against H_1: $p < 0.6$
Test statistic: $X = 50$
p value $= P\{X \leq 50\} = P\{X < 50.5\}$
$\approx P\{Z \leq (50.5 - 60)/4.899\} = P\{Z \leq -1.9392\} = 0.0262$
Conclusion: Since p value $= 0.0262 < \alpha = 0.05$, H_0 is rejected at the 5% level of significance. That is, the data are consistent with the alternative hypothesis that the proportion of voters in favor of restructuring the city government is less than 60%.

(c) $n = 100$, $x = 50$, $p_0 = 0.6$, $\mu = np_0 = 60$, $\sigma = \sqrt{np_0(1 - p_0)} = 4.8990$, $\alpha = 0.01$.
H_0: $p \geq 0.6$ against H_1: $p < 0.6$
Test statistic: $X = 50$
p value $= P\{X \leq 50\} = P\{X \leq 50.5\}$
$\approx P\{Z \leq (50.5 - 60)/4.899\} = P\{Z \leq -1.9392\} = 0.0262$
Conclusion: Since p value $= 0.0262 > \alpha = 0.01$, H_0 is not rejected at the 1% level of significance. That is, the data are not inconsistent with the null hypothesis that the proportion of voters in favor of restructuring the city government is less than 60%.

(d) $n = 200$, $x = 100$, $p_0 = 0.6$, $\mu = np_0 = 120$, $\sigma = \sqrt{np_0(1 - p_0)} = 6.9282$, $\alpha = 0.01$.
H_0: $p \geq 0.6$ against H_1: $p < 0.6$
Test statistic: $X = 100$
p value $= P\{X \leq 100\} = P\{X \leq 100.5\}$
$\approx P\{Z \leq (100.5 - 120)/6.9282\} = P\{Z \leq -2.8146\} = 0.0025$
Conclusion: Since p value $= 0.0025 < \alpha = 0.01$, H_0 is rejected at the 1% level of significance. That is, the data are consistent with the alternative hypothesis that the proportion of voters in favor of restructuring the city government is less than 60%.

6. (a) $n=200, x=50, p_0=0.24, \mu=np_0=48, \sigma=\sqrt{np_0(1 - p_0)}=6.0399, \alpha=0.05$.
H_0: $p \geq 0.24$ against H_1: $p < 0.24$
Test statistic: $X = 50$
p value $= P\{X \leq 50\} = P\{X \leq 50.5\}$
$\approx P\{Z \leq (50.5 - 48)/6.0399\} = P\{Z \leq 0.4139\} = 0.6591$
Conclusion: Since p value $= 0.6591 > \alpha = 0.05$, H_0 is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the proportion of viewers tuned in to the show is greater than or equal to 24%.

(b) $n = 200$, $x = 50$, $p_0 = 0.24$, $\mu = np_0 = 48$, $\sigma = \sqrt{np_0(1-p_0)} = 6.0399$, $\alpha = 0.05$.
H_0: $p \leq 0.24$ against H_1: $p > 0.24$
Test statistic: $X = 50$
p value $= P\{X \geq 50\} = P\{X \geq 50.5\}$
$\approx P\{Z \geq (50.5 - 48)/6.0399\} = P\{Z \geq 0.4139\} = 0.3409$
Conclusion: Since p value $= 0.3409 > \alpha = 0.05, H_0$ is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the proportion of viewers tuned in to the show is less than or equal to 24%.

(c) Since in the above two tests the null hypothesis was not rejected, we cannot say that there is sample evidence for or against the claim.

(d) Should take a larger sample.

8. (a) $n = 200, x = 70, p_0 = 0.45, \mu = np_0 = 90, \sigma = \sqrt{np_0(1-p_0)} = 7.0356$, $\alpha = 0.05$.
H_0: $p \geq 0.45$ against H_1: $p < 0.45$
Test statistic: $X = 70$
p value $= P\{X \leq 70\} = P\{X \leq 70.5\}$
$\approx P\{Z \leq (70.5 - 90)/7.0356\} = P\{Z \leq -2.7716\} = 0.0028$
Conclusion: Since p value $= 0.0028 < \alpha = 0.05,\ H_0$ is rejected at the 5% level of significance. That is, the data are consistent with the alternative hypothesis that the proportion of life-threatening emergency calls is less than 45%.

(b) $n = 200$, $x = 70$, $p_0 = 0.45$, $\mu = np_0 = 90$, $\sigma = \sqrt{np_0(1-p_0)} = 7.0356$, $\alpha = 0.01$.
H_0: $p \geq 0.45$ against H_1: $p < 0.45$
Test statistic: $X = 70$
p value $= P\{X \leq 70\} = P\{X \leq 70.5\}$
$\approx P\{Z \leq (70.5 - 90)/7.0356\} = P\{Z \leq -2.7716\} = 0.0028$
Conclusion: Since p value $= 0.0028 < \alpha = 0.01, H_0$ is rejected at the 1% level of significance. That is, the data are consistent with the alternative hypothesis that the proportion of life-threatening emergency calls is less than 45%.

10. $n = 50, x = 32, p_0 = 0.75, \mu = np_0 = 37.5, \sigma = \sqrt{np_0(1-p_0)} = 3.0619, \alpha = 0.05$.
H_0: $p \geq 0.75$ against H_1: $p < 0.75$
Test statistic: $X = 32$
p value $= P\{X \leq 32\} = P\{X \leq 32.5\}$
$\approx P\{Z \leq (32.5 - 37.5)/3.0619\} = P\{Z \leq -1.633\} = 0.0516$
Conclusion: Since p value $= 0.0516 > \alpha = 0.05, H_0$ is not rejected at the 5% level of significance. That is, the data are not inconsistent with the null hypothesis that the proportion of students who favor traditional course grades rather than pass–fail is greater than or equal to 75%.

12. $n = 200,\ x = 54,\ p_0 = 0.22,\ \mu = np_0 = 44,\ \sigma = \sqrt{np_0(1-p_0)} = 5.8583,$
$\alpha = 0.05.$
H_0: $p = 0.22$ against H_1: $p \neq 0.22$
Test statistic: $X = 54$
p value $= 2[\mathrm{Min}(P\{X \le 54\}, P\{X \ge 54\})]$
Now, $P\{X \le 54\} = P\{X \le 54.5\} \approx P\{Z \le (54.5 - 44)/5.8583\}$
$= P\{Z \le 1.7923\} = 0.9633$, and $P\{X > 53.5\} \approx 0.0549$. Hence,
p value $= 2 \times 0.0549 = 0.1098$.
Conclusion: Since p value $= 0.1098 > \alpha = 0.05, H_0$ is not rejected at the 5% level of significance. That is, the data are consistent with the null hypothesis that the proportion of the population who have firearm at home is equal to 22%.

14. (a) $n = 200, x = 116, p_0 = 0.5, \mu = np_0 = 100, \sigma = \sqrt{np_0(1-p_0)} = 7.0711,$
$\alpha = 0.05$
H_0: $p = 0.5$ against H_1: $p \neq 0.5$
Test statistic: $X = 116$ (successes were taken to be heads)
p value $= 2[\mathrm{Min}(P\{X \le 116\}, P\{X \ge 116\})]$
Now, $P\{X \le 116\} = P\{X \le 116.5\} \approx P\{Z \le (116.5 - 100)/7.0711\}$
$= P\{Z \le 2.3334\} = 0.9901$, and $P\{X \ge 116\} \approx 0.0141$.
Hence, p value $= 2 \times 0.0141 = 0.0282$.
Conclusion: Since p value $= 0.0282 < \alpha = 0.05,\quad H_0$ is rejected at the 5% level of significance. That is, the data are consistent with the alternative hypothesis that the proportion of heads (tails) is not equal to 50%.

Review Problems

2. (a) p value $= 0.11 > \alpha = 0.05$, hence the null hypothesis H_0 will not be rejected.
(b) No, the test does not provide evidence of the ***truth*** of H_0. It implies that the sample data are not inconsistent with the null hypothesis.

4. Answers will vary.

6. (a) $n = 12,\ \overline{X} = 14.4,\ \sigma = 0.5,\ SE(\overline{X}) = 0.1443.$
H_0: $\mu = 15$ against H_1: $\mu \neq 15$
Test statistic: $Z = -4.157$
p value $= 2P\{Z > 4.157\} \approx 0$
(b) $n = 12,\ \overline{X} = 14.4,\ \sigma = 1,\ SE(\overline{X}) = 0.2887.$
H_0: $\mu = 15$ against H_1: $\mu \neq 15$
Test statistic: $Z = -2.0785$
p value $= 2P\{Z > 2.0785\} = 0.0377$
(c) $n = 12,\ \overline{X} = 14.4,\ \sigma = 2,\ SE(\overline{X}) = 0.5774.$
H_0: $\mu = 15$ against H_1: $\mu \neq 15$
Test statistic: $Z = -1.0392$
p value $= 2P\{Z > 1.0392\} = 0.2984$

8. (a) $n = 102,\ x = 55\%$ of $102 = 56.1,\ p_0 = 0.37,\ \mu = np_0 = 37.74,$
$\sigma = \sqrt{np_0(1 - p_0)} = 4.8761.$
H_0: $p \le 0.37$ against H_1: $p > 0.37$
Test statistic: $X = 56.1 \approx 56$ (first born)
p value $= P\{X \ge 56\} = P\{X \ge 55.5\}$
$\approx P\{Z \ge (55.5 - 37.74)/4.8761\} = P\{Z \ge 3.6423\} \approx 0.0000$
Conclusion: Since p value $\approx 0.000,\ H_0$ is rejected. That is, the data are consistent with the alternative hypothesis that the proportion of firstborn on the U.S. Supreme Court is indeed greater than 37%. (This, I guess according to Alfred Adler, would suggest that firstborn tend to be more self-confident and success oriented than later-born children.)

(b) No. All one can really say here is that the proportion of firstborn on the U.S. Supreme Court is significantly more than the proportion of firstborn in the population at large. Many factors measure self-confidence and success, other than merely the presence or absence of service as a Justice on the Supreme Court. Such factors may include (but not limited to) education, athletics, financial and family-related success, and achievements in religious and political activities.

(c) Answer will vary.

10. (a) P\{at least one of the tests will be rejected at the 5% level of significance\}
$= 1 - (0.95)^{25} = 0.7226$

(b) Answers will vary.

12. Answers will vary.

14. (a) No. Since p value $= 0.063 > \alpha = 0.05.$

(b) Yes. Since p value $= 0.063 < \alpha = 0.01.$

(c) p value $= 0.063$

16. Yes, the probability that a Poisson random variable with mean 6.7 would be as large as 27 (more than 4 standard deviations beyond its mean) is miniscule.

Chapter 10 HYPOTHESIS TESTS CONCERNING TWO POPULATIONS

10.2 TESTING EQUALITY OF MEANS OF TWO POPULATIONS: CASE OF KNOWN VARIANCES

Problems

2. (a) $n = m = 10,\ \overline{X} = 6.2670,\ \overline{Y} = 6.2850,\ \sigma_x = \sigma_y = 0.05,$

$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = 0.0224,\ \alpha = 0.05, Z_{\alpha/2} = 1.96.$

$H_0\colon \mu_x = \mu_y$ against $H_1\colon \mu_x \neq \mu_y$

Test statistic: $Z = (6.267 - 6.285)/(0.0224) = -0.8036$

Conclusion: Since $|Z| = 0.8036 < Z_{\alpha/2} = 1.96$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the averages are the same.

(b) p value $= 2P\{Z > 0.8036\} = 0.4238$

4. (a) $n = m = 9,\ \overline{X} = 137.67,\ \overline{Y} = 127.78,\ \sigma_x = 10,\ \sigma_y = 5,$

$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = 3.7268.$

$H_0\colon \mu_x \le \mu_y$ against $H_1\colon \mu_x > \mu_y$

Test statistic: $Z = (137.67 - 127.78)/(3.7268) = 2.6538$

p value $= P\{Z > 2.6538\} = (1 - 0.9960) = 0.004$

(b) $n = m = 9,\ \overline{X} = 137.67,\ \overline{Y} = 127.78,\ \sigma_x = 10,\ \sigma_y = 10,$

$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = 4.7140.$

$H_0\colon \mu_x \le \mu_y$ against $H_1\colon \mu_x > \mu_y$

Test statistic: $Z = (137.67 - 127.78)/(4.7140) = 2.0980$

p value $= P\{Z > 2.0980\} = (1 - 0.9821) = 0.0179$

(c) $n = m = 9,\ \overline{X} = 137.67,\ \overline{Y} = 127.78,\ \sigma_x = 10,\ \sigma_y = 20,$

$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = 7.4536$

H_0: $\mu_x \leq \mu_y$ against H_1: $\mu_x > \mu_y$

Test statistic: $Z = (137.67 - 127.78)/(7.4536) = 1.3269$

p value $= P\{Z > 1.3269\} = (1 - 0.9082) = 0.0918$

6. $n = 8,\ m = 12, \overline{X} = 22.4, \overline{Y} = 21.3, \sigma_x = 0.5, \sigma_y = 0.5, \alpha = 0.05, Z_\alpha = 1.645$

$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = 0.2282.$

H_0: $\mu_x \leq \mu_y$ against H_1: $\mu_x > \mu_y$

Test statistic: $Z = (22.4 - 21.3)/(0.2282) = 4.8203$

Conclusion: Since $Z = 4.8203 > Z_\alpha = 1.645$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average distance of asteroid A is greater than the average distance of asteroid B as measured from earth.

p value $= P\{Z > 4.8203\} \approx 0$

8. (a) $n = 12,\ m = 14,\ \overline{X} = 32,\ \overline{Y} = 22,\ \sigma_x = 8,\ \sigma_y = 8,\ \alpha = 0.05,\ Z_\alpha = 1.645$

$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = 3.1472.$

H_0: $\mu_x \leq \mu_y$ against H_1: $\mu_x > \mu_y$

Test statistic: $Z = (32 - 22)(3.1472) = 3.1774$

Conclusion: Since $Z = 3.1774 > Z_\alpha = 1.645$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average PCB level at the end of the river where the firm is located is greater than the average PCB level at the other end of the river.

(b) p value $= P\{Z > 3.1774\} = (1 - 0.9993) = 0.0007$

10.3 TESTING EQUALITY OF MEANS: UNKNOWN VARIANCES AND LARGE SAMPLE SIZES

Problems

2. $n = 53,\ m = 44,\ \overline{X} = 6.8,\ \overline{Y} = 7.2,\ S_x^2 = 5.2,\ S_y^2 = 4.9,$

$\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} = 0.4577.$

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$

Test statistic: $Z = (6.8 - 7.2)(0.4577) = -0.8739$

p value $= 2P\{Z > 0.8739\} = 2(1 - 0.8078) = 0.3844$

Conclusion: Since the p value is rather large, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average weights (in pounds) of the newborn babies in the two adjacent counties are equal.

4. (a) $n = 40,\ m = 40,\ \overline{X} = 38,\ \overline{Y} = 42,\ S_x^2 = 16,\ S_y^2 = 49,\ a = 0.05,$

$Z_{\alpha 2} = 1.96,\ \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} = 0.4577.$

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$

Test statistic: $Z = (38 - 42)(1.2748) = -3.1377$

Conclusion: Since $|Z| = 3.1377 > Z_{\alpha/2} = 1.96$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average time to work is not the same as the average time from work.

(b) p value $= 2P\{Z > 3.1377\} = 2(1 - 0.9992) = 0.0016$

6. $n = 100,\ m = 60,\ \overline{X} = 102.2,\ \overline{Y} = 105.3,\ S_x^2 = (11.8)^2,\ S_y^2 = (10.6)^2,$

$\alpha = 0.05, Z_{\alpha/2} = 1.96,\ \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} = 1.8069.$

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$

Test statistic: $Z = (102.2 - 105.3)(1.8069) = -1.7156$

Conclusion: Since $|Z| = 1.7156 < Z_{\alpha/2} = 1.96$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average IQ scores for the urban and rural students from upper Michigan are the same.

8. $n = 36,\ m = 36,\ \overline{X} = 12.4,\ \overline{Y} = 14.2,\ S_x^2 = (1.6)^2,\ S_y^2 = (1.8)^2,\ \alpha = 0.05,$

$Z_{\alpha 2} = 1.96,\ \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} = 0.4014.$

H_0: $\mu_x \leq \mu_y$ against H_1: $\mu_x > \mu_y$

Test statistic: $Z = (14.2 - 12.4)/(0.4014) = 4.4843$

Conclusion: Since $Z = 4.4843 > Z_\alpha = 1.645$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average yield for the second variety is greater than the average yield for the first variety.

10. $n = 50,\ m = 50,\ \overline{X} = 85.2,\ \overline{Y} = 74.8,\ S_x^2 = 26.4,\ S_y^2 = 24.5,$

$\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} = 1.0090.$

(a) H_0: $\mu_x - \mu_y \geq 5$ against H_1: $\mu_x - \mu_y < 5$

(b) Test statistic: $Z = [(85.2 - 74.8) - 5]/(1.009) = 5.3518$

p value $= P\{Z < 5.3518\} = 1$

(c) Conclusion: Since p value $= 1$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average salary for the graduates from the college is at least \$5,000 higher as compared to the graduates from the rival institution.

12. (a) $n = 30,\ m = 100,\ \overline{X} = 48.5,\ \overline{Y} = 26.6,\ S_x^2 = (14.5)^2,\ S_y^2 = (12.3)^2,$

$\alpha = 0.01,\ Z_{\alpha/2} = 2.33,\ \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} = 2.9191.$

H_0: $\mu_x \le \mu_y$ against H_1: $\mu_x > \mu_y$

Test statistic: $Z = (48.5 - 26.6)/(2.9191) = 7.5023$

Conclusion: Since $Z = 7.5023 > Z_{\alpha/2} = 2.33$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average lead content in the human hair during the period of 1880 to 1920 is greater than the average lead content in the human hair today.

(b) p value $= P\{Z > 7.5023\} = 0$

10.4 TESTING EQUALITY OF MEANS: SMALL-SAMPLE TESTS WHEN THE UNKNOWN POPULATION VARIANCES ARE EQUAL

Problems

2. (a) $n = 12,\ m = 12,\ \overline{X} = 30.83,\ \overline{Y} = 35.13,\ S_x^2 = (4.3)^2,\ S_y^2 = (4.51)^2,$

$\alpha = 0.05,\ t_{22,\,\alpha/2} = 2.074,\ S_p^2 = (4.41)^2.$

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$

Test statistic: $T = (30.83 - 35.13)/(1.8004) = -2.3884$

Conclusion: Since $|T| = 2.3884 > 2.074$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average fat intake of females is not the same during winter and summer at the 5% level of significance.

(b) $n = 12,\ m = 12,\ \overline{X} = 30.83,\ \overline{Y} = 35.13,\ S_x^2 = (4.3)^2,\ S_y^2 = (4.51)^2,$

$\alpha = 0.01,\ t_{22,\,\alpha/2} = 2.819,\ S_p^2 = (4.41)^2.$

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$

Test statistic: $T = (30.83 - 35.13)/(1.8004) = -2.3884$

Conclusion: Since $|T| = 2.3884 > 2.074$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average fat intake of females is not the same during winter and summer at the 1% level of significance.

4. $n = 12,\ m = 10,\ \overline{X} = 180,\ \overline{Y} = 136,\ S_x^2 = (92)^2,\ S_y^2 = (86)^2,$

$\alpha = 0.05,\ t_{20,\,\alpha/2} = 2.086,\ S_p^2 = 7983.4.$

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$

Test statistic: $T = (180 - 136)(38.2564) = 1.1501$

Conclusion: Since $|T| = 1.1501 < 2.086$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average distance flown between feedings is the same for both the male and female bats.

6. (a) $n = 53,\ m = 44,\ \overline{X} = 6.8,\ \overline{Y} = 7.2,\ S_x^2 = 5.2,\ S_y^2 = 4.9,$

$\alpha = 0.05,\ t_{95,\,\alpha/2} = 1.98,\ S_p^2 = (2.2504)^2.$

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \neq \mu_y$

Test statistic: $T = (6.8 - 7.2)(0.4590) = -0.8715$

Conclusion: Since $|T| = 0.8715 < 1.98$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average weights (in pounds) of the newborn babies in the two adjacent counties are equal.

(b) p value $= 2P\{T > 0.8715\}$ with 95 degrees of freedom. If we use 120 degrees of freedom (closest in the table), observe that $0.8715 < 1.289$. That is, $T = 0.8715$ is to the left of 1.289. Thus, the p value $> 2(0.1) = 0.2$. In the original problem, the p value was 0.3844. In both cases, these values are large enough in order not to reject the null hypothesis. ***Note:*** If technology is used, the p value $= 0.5977$.

8. (a) $n = 20,\ m = 20,\ \overline{X} = 57.4,\ \overline{Y} = 52.8,\ S_x^2 = (12.4)^2,\ S_y^2 = (13.8)^2,$

$\alpha = 0.1,\ t_{38,\,\alpha} = 1.303,\ S_p^2 = 172.1.$

H_0: $\mu_x \leq \mu_y$ against H_1: $\mu_x > \mu_y$

Test statistic: $T = (57.4 - 52.8)/(4.1485) = 1.1088$

Conclusion: Since $|T| = 1.1088 < 1.303$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average distance driven by Los Angeles commuters is less than or equal to the average distance driven by San Francisco commuters at the 10% level of significance.

(b) $n = 20,\ m = 20,\ \overline{X} = 57.4,\ \overline{Y} = 52.8,\ S_x^2 = (12.4)^2,\ S_y^2 = (13.8)^2,$

$\alpha = 0.05,\ t_{38,\,\alpha} = 1.684,\ S_p^2 = 172.1.$

H_0: $\mu_x \leq \mu_y$ against H_1: $\mu_x > \mu_y$

Test statistic: $T = (57.4 - 52.8)/(4.1485) = 1.1088$

Conclusion: Since $|T| = 1.1088 < 1.684$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average distance driven by

Los Angeles commuters is less than or equal to the average distance driven by San Francisco commuters at the 5% level of significance.

(c) $\alpha = 0.01,\ t_{38,\,\alpha} \approx 2.423$. Same conclusion as in (a) and (b).

10. $n = 10,\ m = 10,\ \overline{X} = 9.4$ (no break), $\overline{Y} = 7.10$ (coffee break),

$S_x^2 = (2.84)^2,\ S_y^2 = (1.91)^2,\ \alpha = 0.05,\ t_{18,\,\alpha} = 1.734,\ S_p^2 = 5.8569.$

$H_0\colon \mu_x \le \mu_y$ against $H_1\colon \mu_x > \mu_y$

Test statistic: $T = (9.4 - 7.1)/(1.0823) = 2.125$

Conclusion: Since $T = 2.125 > 1.734$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average number of mistakes made by the "no break group" is greater than the average number of mistakes made by the "coffee break group" at the 5% level of significance.

10.5 PAIRED-SAMPLE *t* TEST

Problems

2. Difference used: Time for old shoes – Time for new shoes.

$n = 10,\ \overline{D} = 0.22,\ S_d = 0.2394,\ \alpha = 0.1,\ t_{9,\,\alpha} = 1.383.$

$H_0\colon \mu_d \le 0$ against $H_1\colon \mu_d > 0$

Test statistic: $T = 2.91$

Conclusion: Since $T = 2.91 > 1.383$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average time is greater for the old shoes than for the new shoes at the 5% level of significance.

For $\alpha = 0.05, t_{9,\,\alpha} = 1.833$. Same conclusion.

p value $= P\{T > 2.91\} = 0.0087$

4. Difference used: Blood pressure before – Blood pressure after.

$n = 8,\ \overline{D} = 0.875,\ S_d = 4.29,\ \alpha = 0.05,\ t_{7,\,\alpha} = 1.895.$

(a) $H_0\colon \mu_d \le 0$ against $H_1\colon \mu_d > 0$

(b) No. The test statistic: $T = 0.58$. Since $T = 0.58 < 1.895$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average blood pressure before jogging is less than or equal to the average blood pressure after jogging at the 5% level of significance.

(c) No. There is insufficient sample evidence to conclude that the average blood pressure after jogging is less than the average blood pressure before jogging at the 5% level of significance.

(d) The data are insufficient for proving or disproving the idea that jogging for 1 month leads to a reduction in systolic blood pressure.

6. Difference used: Calories from fat in July – Calories from fat in January.

$n = 12,\ \overline{D} = -4.31,\ S_d = 6.39,\ \alpha = 0.05, t_{11,\,\alpha 2} = 2.201.$

(a) H_0: $\mu_d = 0$ against H_1: $\mu_d \neq 0$

Test statistic: $T = -2.34$

Conclusion: Since $|T| = 2.34 > 2.201$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average calories derived from fat intake for women during the months of July and January are not equal at the 5% level of significance.

(b) $\alpha = 0.01,\ t_{11,\,\alpha 2} = 3.106.$

H_0: $\mu_d = 0$ against H_1: $\mu_d \neq 0$

Test statistic: $T = -2.34$

Conclusion: Since $|T| = 2.34 < 3.106$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the average calories derived from fat intake for women during the months of July and January are equal at the 1% level of significance.

p value $= 2P\{|T| > 2.34\} = 0.039$

8. Difference used: Salary for men – Salary for women.

$n = 8,\ \overline{D} = 1.9125,\ S_d = 2.3,\ \alpha = 0.1,\ t_{7,\,\alpha 2} = 1.895.$

H_0: $\mu_d = 0$ against H_1: $\mu_d \neq 0$

Test statistic: $T = 2.35$

Conclusion: Since $T = 2.35 > 1.895$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average starting salary for the female and male law school graduates is not the same at the 5% level of significance.

p value $= 2P\{T > 2.35\} = 0.051$

10. (a) Difference used: Death rates in 1985 – Death rates in 1989.
$n = 8,\ \overline{D} = 0.40,\ S_d = 0.363,\ a = 0.05,\ t_{7,\,\alpha} = 1.895.$

H_0: $\mu_d \leq 0$ against H_1: $\mu_d > 0$

Test statistic: $T = 3.12$

Conclusion: Since $T = 3.12 > 1.895$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average death rate in 1985 was greater than the average death rate in 1989 at the 5% level of significance.

(b) Difference used: Death rates in 1985 – Death rates in 2001.

$n = 8,\ \overline{D} = 0.9875,\ S_d = 0.391,\ \alpha = 0.05,\ t_{7,\alpha} = 1.895.$

H_0: $\mu_d \leq 0$ against H_1: $\mu_d > 0$

Test statistic: $T = 7.15$

Conclusion: Since $T = 7.15 > 1.895$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average death rate in 1985 was greater than the average death rate in 2001 at the 5% level of significance.

(c) p value $= P\{T > 3.12\} = 0.0084$ for part (a)

p value $= P\{T > 7.15\} = 0.0$ for part (b)

10.6 TESTING EQUALITY OF POPULATION PROPORTIONS

Problems

2. $n_1 = 210,\ n_2 = 220,\ x_1 = 58,\ x_2 = 71,\ \hat{p}_1 = 0.2762,\ \hat{p}_2 = 0.3227,$

$\hat{p} = 0.3,\ \alpha = 0.05,\ Z_{\alpha/2} = 1.96.$

H_0: $p_1 = p_2$ against H_1: $p_1 \neq p_2$

Test statistic: $Z = -1.0526$

Conclusion: Since $|Z| = 1.0526 < 1.96$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the proportion of male and female decaffeinated coffee drinkers is equal at the 5% level of significance.

p value $= 2P\{Z > 1.0518\} = 2(1 - 0.8531) = 0.2938.$

4. (a) $n_1 = 50{,}000{,}000,\ n_2 = 170{,}000{,}000,\ x_1 = 202,\ x_2 = 181,$

$\hat{p}_1 = 4.04 \times 10^{-6},\ \hat{p}_2 = 1.065 \times 10^{-6}, \hat{p} = 1.7409 \times 10^{-6},\ \alpha = 0.05,$

$Z_\alpha = 1.645.$

H_0: $p_1 = p_2$ against H_1: $p_1 > p_2$

Test statistic: $Z = 14.0166$

Conclusion: Since $|Z| = 14.0166 > 1.645$, reject H_0. That is, the data are consistent with the alternative hypothesis that the proportion contracting swine flu is greater for the vaccinated portion of the population than for the unvaccinated portion at the 5% level of significance.

(b) No. Association does not imply causation. For example, at-risk people may be more likely to have been vaccinated than other people.

6. (a) $n_1 = 100,\ n_2 = 100,\ x_1 = 56,\ x_2 = 45,\ \hat{p}_1 = 0.56,\ \hat{p}_2 = 0.45,$

$\hat{p} = 0.505,\ \alpha = 0.1,\ Z_{\alpha/2} = 1.645.$

$H_0\colon p_1 = p_2$ against $H_1\colon p_1 \neq p_2$

Test statistic: $Z = 1.5557$

Conclusion: Since $|Z| = 1.5557 < 1.645$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the proportion of the population that favored the raising of the driving age is the same for San Francisco and Los Angeles at the 10% level of significance.

(b) For $\alpha = 0.05, Z_{\alpha/2} = 1.96$. Same conclusion as in part (a).

8. In this solution, we use the event of a boy as the third child as a successful outcome.

$n_1 = 830,\ n_2 = 1104,\ x_1 = 412,\ x_2 = 560,\ \hat{p}_1 = 0.4964,\ \hat{p}_2 = 0.5072,$

$\hat{p} = 0.5026,\ \alpha = 0.05,\ Z_{\alpha/2} = 1.96.$

$H_0\colon p_1 = p_2$ against $H_1\colon p_1 \neq p_2$

Test statistic: $Z = -0.4728$

Conclusion: Since $|Z| = 0.4728 < 1.96$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that the proportion of boys as the third child is equal at the 5% level of significance for the two types of families.

10. $n_1 = 480,\ n_2 = 360,\ x_1 = 72,\ x_2 = 30,\ \hat{p}_1 = 0.15,\ \hat{p}_2 = 0.0833,\ \hat{p} = 0.1214.$

$H_0\colon p_1 = p_2$ against $H_1\colon p_1 \neq p_2$

Test statistic: $Z = -2.927$

p value $= 2P\{Z > 2.927\} = 0.0034$

12. (a) The value of the test statistic is 0.2228, with resulting

p value $= 2P(Z \geq 0.2228) \approx 0.82$

Thus, the null hypothesis is not rejected.

(b) Yes, when $p = 0.5$, 72 is about one standard deviation below its mean. The exact p value is about 0.38.

14. (a) $n_1 = 2500,\ n_2 = 2000,\ x_1 = 738,\ x_2 = 640,\ \hat{p}_1 = 0.2952,\ \hat{p}_2 = 0.32,$

$\hat{p} = 0.3062,\ \alpha = 0.05,\ Z_\alpha = 1.645.$

H_0: $p_2 \le p_1$ against H_1: $p_2 > p_1$

Test statistic: $Z = 1.7935$

Conclusion: Since $Z = 1.7935 > 1.645$, reject H_0. That is, the data are consistent with the alternative hypothesis that the proportion of smokers before 1986 was more than the current proportion of smokers. Hence, one can imply that the proportion of smokers has decreased since 1986 at the 5% level of significance.

(b) For $\alpha = 0.01$, $Z_\alpha = 2.33$. Since $1.7935 < 2.33$, do not reject the null hypothesis at the 1% level of significance.

16. $n_1 = 286$, $n_2 = 310$, $x_1 = 252$, $x_2 = 270$, $\hat{p}_1 = 0.8811$, $\hat{p}_2 = 0.8710$

H_0: $p_1 = p_2$ against H_1: $p_1 \ne p_2$

Test statistic: $Z = 0.38$

p value $= 0.707$

Since 0.707 is large relative to any significance level, do not reject H_0.

Review Problems

2. $n = 14$, $m = 12$, $\overline{X} = 15.2$, $\overline{Y} = 14$, $S_x^2 = (9.2)^2, S_y^2 = (9)^2$,

$\alpha = 0.05$, $t_{24,\,\alpha/2} = 2.064$, $S_p^2 = 82.9717$.

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \ne \mu_y$

Test statistic: $T = (15.2 - 14)/(3.5834) = 0.3349$

Conclusion: Since $|T| = 0.3349 < 2.064$, do not reject H_0. That is, the data are not inconsistent with the null hypothesis that both treatments are equally effective in the population of rats at the 5% level of significance.

4. (a) $n = 26$ (female), $m = 34$ (male), $\overline{X} = 201.9$, $\overline{Y} = 195.4$, $S_x^2 = (11)^2$,

$S_y^2 = (12.8)^2$, $\alpha = 0.05$, $t_{58,\,\alpha/2} \approx 2$, $S_p^2 = (12.1)^2$.

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \ne \mu_y$

Test statistic: $T = (201.9 - 195.49)/(3.1524) = 2.062$

Conclusion: Since $|T| = 2.062 > 2$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average cholesterol level is different for males and females at the 5% level of significance.

p value $= 0.044$

(b) $n = 26$ (female), $m = 34$ (male), $\overline{X} = 112.27$, $\overline{Y} = 119.27$,

$S_x^2 = (9.32)^2$, $S_y^2 = (14.7)^2$, $\alpha = 0.05$, $t_{58,\,\alpha/2} \approx 2$, $S_p^2 = (12.7)^2$.

H_0: $\mu_x = \mu_y$ against H_1: $\mu_x \ne \mu_y$

Test statistic: $T = (112.27 - 119.27)(3.3087) = -2.1157$

Conclusion: Since $|T| = 2.1157 > 2$, reject H_0. That is, the data are consistent with the alternative hypothesis that the average blood pressure is different for males and females at the 5% level of significance.

p value ≈ 0.034

6. $n_1 = 1583,\ n_2 = 1242,\ x_1 = 113,\ x_2 = 123,\ \hat{p}_1 = 0.0714,\ \hat{p}_2 = 0.0990,$

$\hat{p} = 0.0835,\ \alpha = 0.01,\ Z_\alpha = 2.33.$

H_0: $p_1 \leq p_2$ against H_1: $p_1 > p_2$

Test statistic: $Z = -2.636$

Conclusion: Since $|Z| = 2.636 > 2.33$, reject H_0. That is, the data are inconsistent with the null hypothesis that the proportion of people dying from cancer who were exposed to dioxins is the same as those who were not exposed to dioxins at the 1% level of significance.

p value $= P\{Z > 2.632\} = 0.0043$

8. $n_1 = 56,\ n_2 = 64,\ x_1 = 38,\ x_2 = 32,\ \hat{p}_1 = 0.6786,\ \hat{p}_2 = 0.5,$

$\hat{p} = 0.5833,\ \alpha = 0.05,\ Z_{\alpha 2} = 1.96.$

H_0: $p_1 = p_2$ against H_1: $p_1 \neq p_2$

Test statistic: $Z = 1.9795$

Conclusion: Since $|Z| = 1.9795 > 1.96$, reject H_0. That is, the data are inconsistent with the null hypothesis that the proportion of men and women who favor gun control is the same at the 5% level of significance.

p value $= 2P\{Z > 1.9798\} = 0.0478$

10. $n_1 = 92,\ n_2 = 93,\ x_1 = 6,\ x_2 = 21, \hat{p}_1 = 0.0652,\ \hat{p}_2 = 0.2258,$

$\hat{p} = 0.1459.$

H_0: $p_1 = p_2$ against H_1: $p_1 \neq p_2$

Test statistic: $Z = -3.0933$

p value $= 2P\{Z > 3.0933\} = 2(1 - 0.999) = 0.002$

12. (b)

Chapter 11 ANALYSIS OF VARIANCE

11.2 ONE-FACTOR ANALYSIS OF VARIANCE

2. Below is a portion of a ***Minitab*** output for this problem.

One-way ANOVA: Sample 1, Sample 2, Sample 3

```
Source    DF     SS     MS     F      P
Factor     2   72.0   36.0  2.61  0.127
Error      9  124.0   13.8
Total     11  196.0

S = 3.712  R - Sq = 36.73%   R - Sq(adj) = 22.68%

                               Individual 95% CIs for
                               Mean Based on Pooled
                               StDev
Level      N    Mean  StDev  +-----+----+----+-----
Sample 1   4   8.000  3.162  (-----*-----)
Sample 2   4  14.000  4.082         (-----*-----)
Sample 3   4  11.000  3.830      (-----*-----)
                             +-----+----+-----+----
                           4.0   8.0 12.0   16.0
Pooled StDev = 3.712
```

From this output, $4\bar{S}^2 = 36$ and $\sum_{i=1}^{3} S_i^2/3 = 13.8$. Numerator degrees of freedom $= 2$, denominator degrees of freedom $= 9$, $\alpha = 0.05$,

$F_{2,9,0.05} = 4.26$.

$H_0: \mu_1 = \mu_2 = \mu_3$ against H_1: not all the means are equal

Test statistic: $F = 36/13.8 = 2.61$

Conclusion: Since $F = 2.61 < 4.26$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the means are different at the 5% level of significance.

***Note*:** This result is confirmed by the p value of 0.127 from the output. Also, observe that the confidence intervals for the three populations overlap, and hence, you cannot claim that they are different.

4. $20\bar{S}^2 = 128.6, \sum_{i=1}^{3} S_i^2/3 = 117.333$. Numerator degrees of freedom = 2, denominator degrees of freedom = 57, $\alpha = 0.05$, $F_{2,57,0.05} \approx 3.15$.
H_0: $\mu_1 = \mu_2 = \mu_3$ against H_1: not all the means are equal
Test statistic: $F = 128.6/117.3333 = 1.096$
Conclusion: Since $F = 1.096 < 3.15$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the mean calories due to fat for individuals living in the three regions are different at the 5% level of significance.

6. $25\bar{S}^2 = 2166.6667, \sum_{i=1}^{4} S_i^2/4 = 24.5$. Numerator degrees of freedom = 3, denominator degrees of freedom = 96, $\alpha = 0.05$, $F_{3,96,0.05} \approx 2.68$.
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ against H_1: not all the means are equal
Test statistic: $F = 2166.6667/24.5 = 88.4354$
Conclusion: Since $F = 88.4354 > 2.68$, reject H_0. That is, there is sufficient sample evidence to conclude that the mean distance traveled by the four types of golf balls is different at the 5% level of significance.

8. $12\ \bar{S}^2 = 336, \sum_{i=1}^{3} S_i^2/3 = 144.3333$. Numerator degrees of freedom = 2, denominator degrees of freedom = 33, $\alpha = 0.05, F_{2,33,0.05} \approx 3.32$.
H_0: $\mu_1 = \mu_2 = \mu_3$ against H_1: not all the means are equal
Test statistic: $F = 336/144.333 = 2.3279$
Conclusion: Since $F = 2.3279 < 3.32$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the mean times to clear a mild asthmatic attack for the three steroids are different at the 5% level of significance.

10. $6\bar{S}^2 = 294, \sum_{i=1}^{3} S_i^2/3 = 201$. Numerator degrees of freedom = 2, denominator degrees of freedom = 15, $\alpha = 0.05$, $F_{2,15,0.05} = 3.68$.
H_0: $\mu_1 = \mu_2 = \mu_3$ against H_1: not all the means are equal
Test statistic: $F = 294/201 = 1.4627$
Conclusion: Since $F = 1.4627 < 3.68$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the mean cholesterol levels for the runners in the three categories are different at the 5% level of significance.

12. $8\,\bar{S}^2 = 72.5067,\ \sum_{i=1}^{3} S_i^2/3 = 112.5$. Numerator degrees of freedom = 2, denominator degrees of freedom = 21, $\alpha = 0.05,\ F_{2,\,21,\,0.05} = 3.47$.

H_0: $\mu_1 = \mu_2 = \mu_3$ against H_1: not all the means are equal

Test statistic: $F = 72.5067/112.5 = 0.6445$

Conclusion: Since $F = 0.6445 < 3.47$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the mean test scores for the different types of training are different at the 5% level of significance.

11.3 TWO-FACTOR ANALYSIS OF VARIANCE: INTRODUCTION AND PARAMETER ESTIMATION

Problems

2. $X_{..} = (78 + 75 + 66 + \cdots + 61 + 52)/15 = 68.8$

$\sum_{i=1}^{5} X_i/5 = (249/3 + 226/3 + 196/3 + 196/3 + 165/3)/5 = 68$

$\sum_{i=1}^{3} X_i/3 = (348/5 + 332/5 + 352/5)/3 = 68.8$

These are all different ways of computing the average of all the data values (grand mean).

4. $E[X_{11}] = 75.3,\ E[X_{12}] = 68.8,\ E[X_{13}] = 61.05,\ E[X_{14}] = 71.8,$
$E[X_{15}] = 87.05,\ E[X_{21}] = 77.5,\ E[X_{22}] = 71,\ E[X_{23}] = 63.25,$
$E[X_{24}] = 74,\ E[X_{25}] = 89.25,\ E[X_{31}] = 75.7,\ E[X_{32}] = 69.2,$
$E[X_{33}] = 61.45,\ E[X_{34}] = 72.2,\ E[X_{35}] = 87.45,\ E[X_{41}] = 77.5,$
$E[X_{42}] = 71,\ E[X_{43}] = 63.25,\ E[X_{44}] = 74,\ E[X_{45}] = 89.25$

6. $X_1 = 9.725,\ X_2 = 5.95,\ X_3 = 2.125,\ X_1 = 5.13,\ X_2 = 7.33,\ X_3 = 5.57,$
$X_4 = 5.7,\ X = 5.933$

(a) $\hat{\mu} = 5.933$
(b) $\hat{\alpha}_1 = 9.725 - 5.933 = 3.792, \hat{\alpha}_2 = 5.95 - 5.933 = 0.017, \hat{\alpha}_3 = 2.125 - 5.933 = -3.808$
(c) $\hat{\beta}_1 = 5.13 - 5.9334 = -0.803, \hat{\beta}_2 = 7.333 - 5.933 = 1.4,\ \hat{\beta}_3 = 5.567 - 5.933\ = -0.367, \hat{\beta}_4 = 5.7 - 5.933 = -0.233$

8. $\begin{bmatrix} 5 & 9 & 13 & 17 \\ 6 & 10 & 14 & 18 \\ 7 & 11 & 15 & 19 \end{bmatrix}$

11.4 TWO-FACTOR ANALYSIS OF VARIANCE: TESTING HYPOTHESES

Problems

2. Below is a partial ***Minitab*** output for this problem.

Two-way ANOVA: Response versus System, User

```
Source       DF       SS       MS      F      P
System        2   90.167  45.0833  41.62  0.000
User          3   54.250  18.0833  16.69  0.003
Error         6    6.500   1.0833
Total        11  150.917

S = 1.041   R - Sq = 95.69%   R - Sq (adj) = 92.10%

                 Individual 95% CIs for Mean
                 Based on Pooled StDev
System     Mean  -+--------+-----+-----+-----
1         19.50     (-----*-----)
2         18.50  (-----*-----)
3         24.75                (-----*-----)
                 -+--------+-----+-----+-----
               17.5     20.0  22.5  25.0

                 Individual 95% CIs for Mean
                 Based on Pooled StDev
User       Mean  ----+-----+-----+-----+-----
1       22.6667              (-----*-----)
2       23.3333                (-----*-----)
3       19.3333     (-----*-----)
4       18.3333  (-----*-----)
                 ----+-----+-----+-----+-----
                  18.0  20.0  22.0  24.0
```

(a) $\alpha = 0.05,\ F_{2,6,0.05} = 5.14.$
H_0: $\alpha_i = 0$ for $i = 1,\ 2,\ 3$ against H_1: not all the α_i are equal to 0
Test statistic: $F = 45.0833/1.0833 = 41.62$
Conclusion: Since $F = 41.62 > 5.14$, reject H_0 and claim that the mean training time is not the same for all the systems.
***Note*:** This result can be observed from the 95% confidence intervals for the systems. Observe that one of them is way to the right and does not overlap with the other two.

(b) $\alpha = 0.05, F_{3,6,0.05} = 4.76.$
H_0: $\beta_j = 0$ for $j = 1,\ 2,\ 3,\ 4$ against H_1: not all the β_j are equal to 0

Test statistic: $F = 18.0833/1.0833 = 16.69$
Conclusion: Since $F = 16.69 > 4.76$, reject H_0 and claim that the mean training time is not the same for all the users.
***Note*:** This result can be observed from the 95% confidence intervals for the users. Observe that they do not all intersect.

4. Below is a partial ***Minitab*** output for the problem.

Two-way ANOVA: Exam Score versus Exam, Student

```
Source      DF      SS        MS      F       P
Exam         3    20.4     6.800   0.59   0.633
Student      4  1457.5   364.375  31.66   0.000
Error       12   138.1    11.508
Total       19  1616.0

S = 3.392  R - Sq = 91.45%   R - Sq (adj) = 86.47%
                Individual 95% CIs for Mean
                Based on Pooled StDev
Exam      Mean  --+-------+-----+----+-----
1         72.8  (-----*-----)
2         75.0         (-----*-----)
3         73.2    (-----*-----)
4         75.0         (-----*-----)
                --+-------+-----+----+-----
               70.0    72.5  75.0  77.5

                Individual 95% CIs for Mean
                Based on Pooled StDev
Student   Mean  -+-------+-----+-----+-----
1        76.50        (---*--)
2        70.00    (---*---)
3        62.25  (--*---)
4        73.00         (---*---)
5        88.25                (--*---)
                -+-------+-----+-----+-----
                60      70     80   90
```

$\alpha = 0.05,\ F_{3,12,0.05} = 3.49.$

H_0: $\alpha_i = 0$ for $i = 1,\ 2,\ 3,\ 4$ against H_1: not all the α_i are equal to 0

Test statistic: $F = 6.8/11.508 = 0.59$

Conclusion: Since $F = 0.59 < 3.49$, do not reject H_0. That is, there is insufficient sample evidence to claim that the mean exam scores for the different exams are not the same at the 5% level of significance.

***Note*:** This result can be observed from the 95% confidence intervals for the exam scores. Observe that they all overlap.

6. Below is a partial ***Minitab*** output for this problem.

Two-way ANOVA: Boxes versus Shift, Man

```
Source       DF       SS        MS     F       P
Shift         2   18.000    9.0000  2.08   0.241
Man           2   68.667   34.3333  7.92   0.041
Error         4   17.333    4.3333
Total         8  104.000

S = 2.082   R - Sq = 83.33%    R - Sq (adj) = 66.67%

                 Individual 95% CIs for Mean
                 Based on Pooled StDev
Shift     Mean   -----+-----+-----+------+-
1      29.3333                (-----*-----)
2      26.3333   (-----*-----)
3      29.3333                (-----*-----)
                 -----+-----+-----+------+-
                   25.0  27.5   30.0   32.5

                 Individual 95% CIs for Mean
                 Based on Pooled StDev
Man       Mean   -----+-----+------+-----+----
1      32.0000                (-----*-----)
2      27.6667       (-----*-----)
3      25.3333   (-----*-----)
                 -----+-----+------+------+--
                   24.5  28.0   31.5   35.0
```

(a) $\alpha = 0.05,\ F_{2,4,0.05} = 9.94$.
H_0: $\beta_j = 0$ for $j = 1,\ 2,\ 3$ against H_1: not all the β_j are equal to 0
Test statistic: $F = 34.3333/4.3333 = 7.92$
Conclusion: Since $F = 7.92 < 9.94$, do not reject H_0 and claim that there is insufficient sample evidence to conclude that the mean number of boxes packed is not the same for all workers.

(b) $\alpha = 0.05, F_{2,4,0.05} = 6.94$.
H_0: $\alpha_i = 0$ for $i = 1,\ 2,\ 3$ against H_1: not all the α_i are equal to 0
Test statistic: $F = 9/4.3333 = 2.08$
Conclusion: Since $F = 2.08 < 9.94$, do not reject H_0. That is, there is insufficient sample evidence to claim that the mean number of boxes packed is not the same for all shifts.
***Note*:** This result can be observed from the 95% confidence intervals for the shifts. Observe that they all overlap.

8. Below is a partial ***Minitab*** output for this problem.

Two-way ANOVA: Birth Rates versus Place, Year

```
Source       DF       SS       MS      F      P
Place         3  28.7245  9.57482  31.98  0.000
Year          3  23.1013  7.70042  25.72  0.000
Error         9   2.6945  0.29939
Total        15  54.5202

S = 0.5472  R - Sq = 95.06%   R - Sq (adj) = 91.76%

                  Individual 95% CIs for Mean
                  Based on Pooled StDev
Place      Mean   --+------+------+-----+-----
1       13.3925                  (-----*-----)
2       10.0800   (-----*-----)
3       11.0800             (-----*-----)
4       10.1475   (-----*-----)
                  --+------+------+-----+-----
                 9.6   10.8   12.0  13.2

                  Individual 95% CIs for Mean
                  Based on Pooled StDev
Year       Mean   -------+-----+-----+-------
1990    13.2500             (-----*-----)
2001    10.6125     (-----*-----)
2002    10.4875    (-----*-----)
2003    10.3500   (-----*-----)
                  ------+-------+-----+-----+
                     10.8    12.0  13.2  14.4
```

(a) $\alpha = 0.05,\ F_{3,9,0.05} = 3.86.$

H_0: $\alpha_i = 0$ for $i = 1,\ 2,\ 3,\ 4$ against H_1: not all the α_i are equal to 0

Test statistic: $F = 9.57482/0.29939 = 31.98$

Conclusion: Since $F = 31.98 > 3.86$, reject H_0. That is, there is sufficient sample evidence to claim that the mean birth rate is not the same for the four countries at the 5% level.

***Note*:** This result can be observed from the 95% confidence intervals for the shifts. Observe that they all overlap.

(b) $\alpha = 0.05,\ F_{3,\ 9,\ 0.05} = 3.86.$

H_0: $\beta_j = 0$ for $j = 1,\ 2,\ 3,\ 4$ against H_1: not all the β_j are equal to 0

Test statistic: $F = 7.70042/0.29939 = 25.72$

Conclusion: Since $F = 25.72 > 3.86$, reject H_0. That is, there is sufficient sample evidence to claim that the average birth rates for the different years are significantly different at the 5% level.

***Note*:** Observe that all the confidence intervals for year overlap.

Review Problems

2. $15\bar{S}^2 = 150$, $\sum_{i=1}^{4} S_i^2/4 = 235.75$. Numerator degrees of freedom $= 3$, denominator degrees of freedom $= 56$, $\alpha = 0.05$, $F_{3,56,0.05} \approx 2.76$.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ against H_1: at least one of the means is different

Test statistic: $F = 150/235.75 = 0.6363$

Conclusion: Since $F = 0.6363 < 2.76$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the reading courses are not equally effective at the 5% level of significance.

4. Below is a portion of a ***Minitab*** output for this problem.

Two-way ANOVA: Score versus Detergent, Machine

```
Source          DF       SS        MS       F       P
Detergent        3  102.333   34.1111    9.23   0.011
Machine          2  135.167   67.5833   18.29   0.003
Error            6   22.167    3.6944
Total           11  259.667
S = 1.922  R - Sq = 91.46%   R - Sq (adj) = 84.35%
                    Individual 95% CIs for Mean
                    Based on Pooled StDev
Detergent     Mean  ---+-----+------+----+-----
1          54.0000          (-----*-----)
2          56.0000              (-----*-----)
3          58.6667                 (-----*----)
4          50.6667  (-----*-----)
                    ---+-----+------+----+-----
                    49.0  52.5   56.0 59.5
                    Individual 95% CIs for Mean
                    Based on Pooled StDev
Machine       Mean  ------+-------+-----+-----+
1            53.25     (-----*-----)
2            51.75  (-----*-----)
3            59.50               (-----*-----)
                    ------+-------+-----+-----+
                       52.5   56.0   59.5   63.0
```

Note: For parts (a) and (b), it was assumed that low mean scores were better.

(a) The mean for detergent 1 = 54, the mean for detergent 2 = 56, the mean for detergent 3 = 58.7, and the mean for detergent 4 = 50.7.

  (i) Mean improvement of detergent 1 over 2 = 56 − 54 = 2
  (ii) Mean improvement of detergent 1 over 3 = 58.7 − 54 = 4.7
  (iii) Mean improvement of detergent 1 over 4 = 50.7 − 54 = −3.3 (interpret the negative value as a non-improvement)

(b) The mean for machine 1 = 53.25, the mean for machine 2 = 51.75, and the mean for machine 3 = 59.5.

  (i) Mean improvement of machine 3 over 1 = 53.25 − 59.5 = −6.25
  (ii) Mean improvement of machine 3 over 2 = 51.75 − 59.5 = −7.75

(c) $\alpha = 0.05,\ F_{3,6,0.05} = 4.76$.
H_0: $\alpha_i = 0$ for $i = 1,\ 2,\ 3,\ 4$ against H_1: not all the α_i are equal to 0
Test statistic: $F = 34.1111/3.6944 = 9.23$
Conclusion: Since $F = 9.23 > 4.76$, reject H_0. That is, there is sufficient sample evidence to claim that the mean score is not the same for the four detergents at the 5% level.

(d) $\alpha = 0.05,\ F_{2,6,0.05} = 5.14$.
H_0: $\beta_j = 0$ for $j = 1,\ 2,\ 3$ against H_1: not all the β_j are equal to 0
Test statistic: $F = 67.5833/3.6944 = 18.29$
Conclusion: Since $F = 18.29 > 5.14$, reject H_0. That is, there is sufficient evidence to claim that the average score for the different machines are significantly different at the 5% level.

6. See partial ***Minitab*** output for solution to Prob. 4 in Section 11.4.

$\alpha = 0.05,\ F_{4,12,0.05} = 3.26$.

H_0: $\beta_j = 0$ for $j = 1,\ 2,\ 3,\ 4,\ 5$ against H_1: not all the β_j are equal to 0

Test statistic: $F = 364.375/11.508 = 31.66$

Conclusion: Since $F = 31.66 > 3.26$, reject H_0. That is, there is sufficient sample evidence to claim that the mean exam scores for the different tests are not the same at the 5% level of significance.

***Note*:** This result can be observed from the 95% confidence intervals for the student means. Observe that they do not overlap. (Refer to the output in Prob. 4 in Section 11.4.)

8. Below is a partial ***Minitab*** output for this problem.

```
Two-way ANOVA: Values versus Row, Column
Source     DF      SS        MS     F       P
Row         3   168.8    56.267  0.99   0.430
Column      4  1946.3   486.575  8.57   0.002
Error      12   681.7    56.808
Total      19  2796.8
S = 7.537  R - Sq = 75.63%   R - Sq (adj) = 61.41%
               Individual 95% CIs for Mean
               Based on Pooled StDev
Row      Mean  ---+-----+-----+-----+-----
1        23.8  (-----*-----)
2        28.6         (-----*-----)
3        29.4          (-----*-----)
4        31.8             (-----*-----)
               ---+-----+-----+-----+-----
                 18.0  24.0   30.0    36.0
               Individual 95% CIs for Mean
               Based on Pooled StDev
Column   Mean  ------+------+-----+----+
1       30.50            (-----*-----)
2       28.50          (-----*-----)
3       27.50        (-----*-----)
4       43.25               (-----*-----)
5       12.25  (-----*-----)
               ------+------+-----+----+
                    12     24    36   48
```

(a) $\alpha = 0.05,\ F_{3,12,0.05} = 3.49$.
H_0: $\alpha_i = 0$ for $i = 1,\ 2,\ 3,\ 4$ against H_1: not all the α_i are equal to 0
Test statistic: $F = 56.267/56.808 = 0.99$
Conclusion: Since $F = 0.99 < 3.49$, do not reject H_0. That is, there is insufficient sample evidence to claim any row effect at the 5% level of significance.

(b) $\alpha = 0.05,\ F_{4,12,0.05} = 3.26$.
H_0: $\beta_j = 0$ for $j = 1,\ 2,\ 3,\ 4,\ 5$ against H_1: not all the β_j are equal to 0
Test statistic: $F = 486.575/56.808 = 8.57$
Conclusion: Since $F = 8.57 > 3.26$, reject H_0. That is, there is sufficient evidence to claim a column effect at the 5% level.

10. $8\bar{S}^2 = 155.5467, \sum_{i=1}^{3} S_i^2/3 = 21.4333$. Numerator degrees of freedom $= 2$, denominator degrees of freedom $= 21$, $\alpha = 0.05$, $F_{2,21,0.05} = 3.47$.

H_0: $\mu_1 = \mu_2 = \mu_3$ against H_1: at least one of the means is different

Test statistic: $F = 155.5467/21.4333 = 7.2572$

Conclusion: Since $F = 7.2572 > 3.47$, reject H_0. That is, there is sufficient sample evidence to conclude that the average lifetime (months) of a rat is affected by its diet at the 5% level of significance.

For $\alpha = 0.01$, $F_{2,21,0.01} = 5.78$. Same conclusion as above at the 1% level of significance.

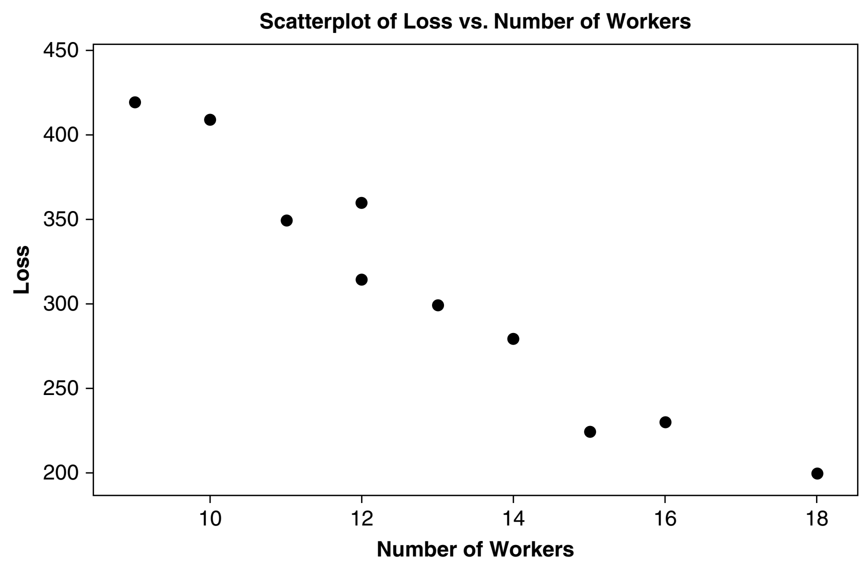
Chapter 12 LINEAR REGRESSION

12.2 SIMPLE LINEAR REGRESSION MODEL

Problems

2. (a) Let Y = value of the merchandise lost to shoplifters.
Let x = number of workers on duty.

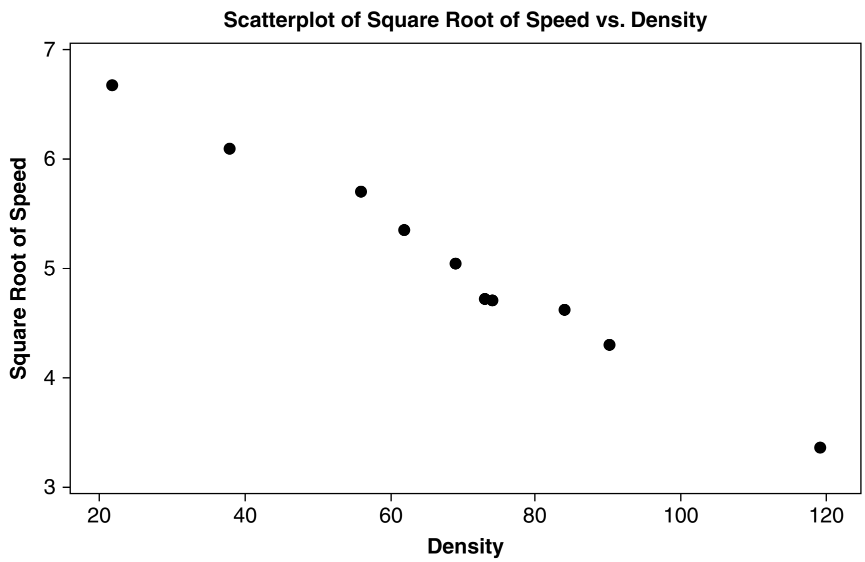
(b) The scatter diagram is shown below.



(c) A linear regression model will be a reasonable model to use since the data form a linear pattern. Observe that the amount of loss decreases as the number of workers increases.

4. (a) Let Y = square root of the speed.
Let x = traffic density.

(b) The scatter diagram for the square root of the speed versus traffic density is shown below.



(c) The simple linear regression model will be a reasonable model to use since the data form a linear pattern.

12.3 ESTIMATING THE REGRESSION PARAMETERS

Problems

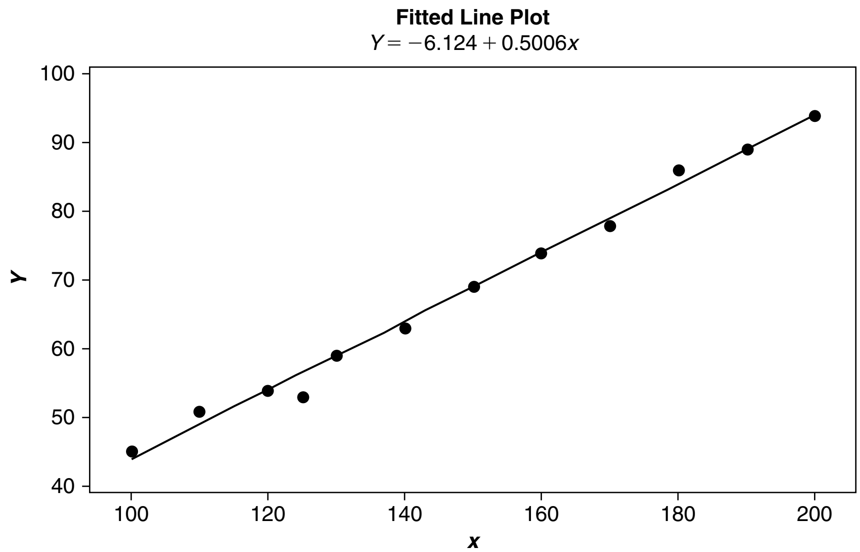
2. $n = 8, \sum_{i=1}^{8} x = 155, \sum_{i=1}^{8} Y = 150.8, \sum_{i=1}^{8} (xY) = 2796.4, \sum_{i=1}^{8} x^2 = 3285,$

$\sum_{i=1}^{8} Y^2 = 2903.66,\ S_{xY} = -125.35, S_{xx} = 281.875, S_{YY} = 61.08.$

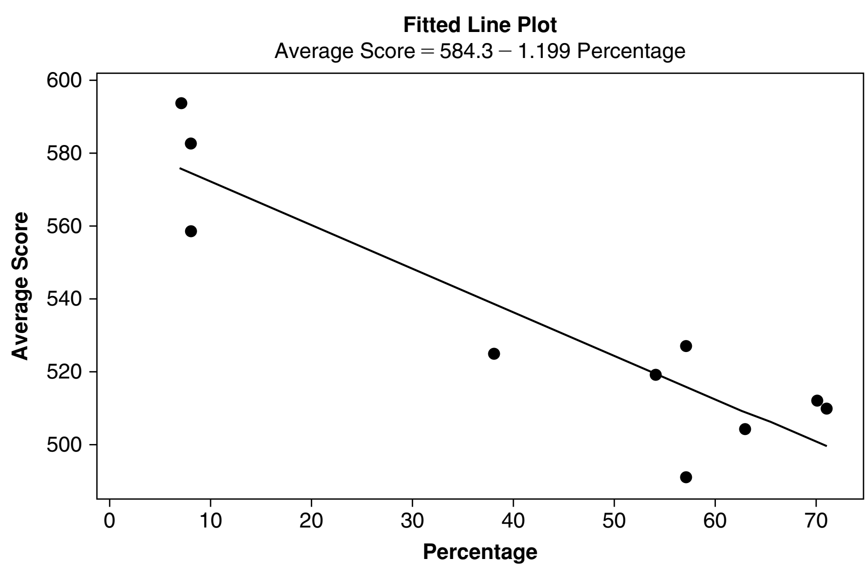
So, $\hat{\beta} = S_{xY}/S_{xx} = -125.35/281.875 = -0.4447007,$ and

$\hat{\alpha} = (150.8/8) - (-0.4447)(155/8) = 27.46606.$

4. (a) and (b)

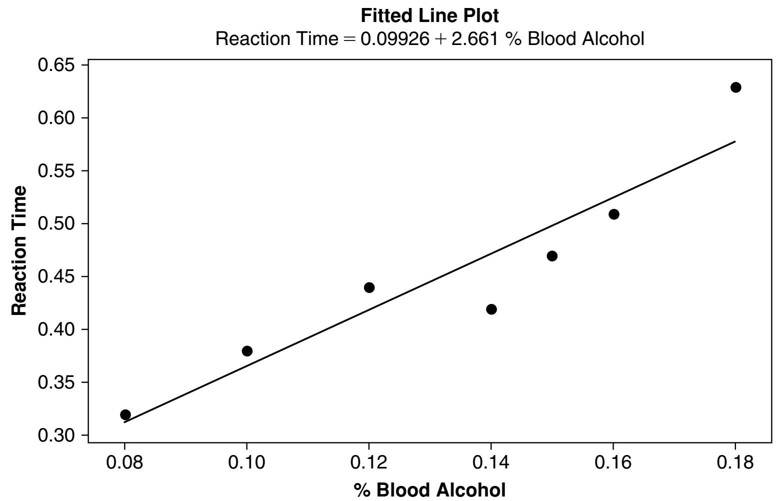


6. (a) Scatter plot of average SAT math scores versus percentage of graduating seniors who took the test.



(b) The estimated regression line is $Y = 584.3 - 1.199x$ (Average score $= 584.3 -$ 1.199 percentage).

8. Solutions to parts (a) and (c) are shown in the diagram below. Parts (b) and (d) will vary.

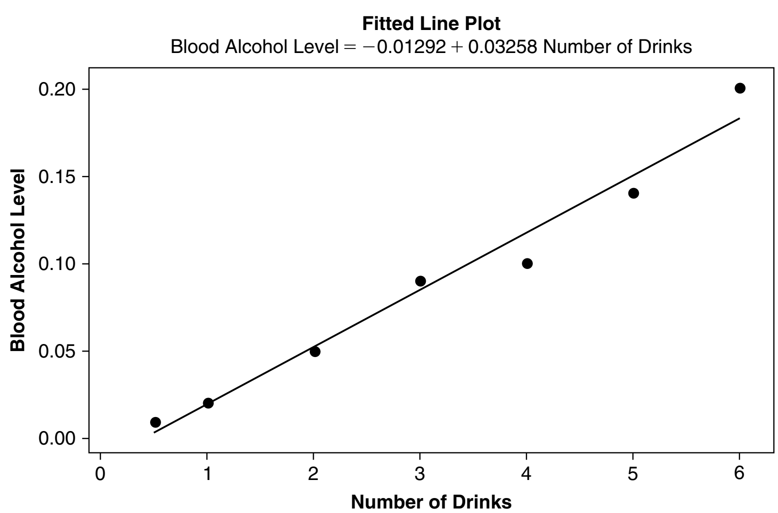


(c) $Y = 0.0993 + 2.661x$

(e) $x = 0.15$, the predicted reaction time (seconds) is
$Y = 0.0993 + 2.66(0.15) = 0.4985$

(f) $x = 0.17$, the predicted reaction time (seconds) is
$Y = 0.0993 + 2.661(0.17) = 0.5517$

10. (a) and (b) are shown in the diagram below.



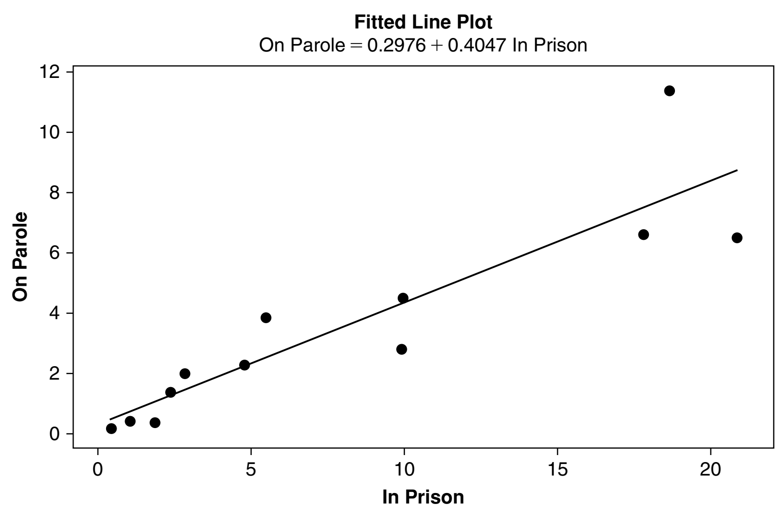
(c) Since $Y = -0.013 + 0.0326x$, then for $x = 3$, the predicted blood alcohol concentration is $Y = -0.013 + 0.0326(3) = 0.0848$.

(d) Since $Y = -0.0129 + 0.0326x$, then for $x = 7$, the predicted blood alcohol concentration is $Y = -0.0129 + 0.0326(7) = 0.2153$.

12. If we code the years x such that we start at 0 for 1994, 2 for 1996, 4 for 1998, etc., then the estimated regression line is $y = 507.6 + 0.7799x$. Thus the predictions for years 1997, 2006, and 2008 are 509.9397, 516.9588, and 519.2985 respectively.

14. The estimated regression line is $Y = 0.298 + 0.405x$, where Y is the number on parole and x is the number in prison in thousands.

For $x = 14.5$, the predicted number on parole is
$Y = 0.2976 + 0.4047(14.5) = 6.16575$ or $6,165.75$.



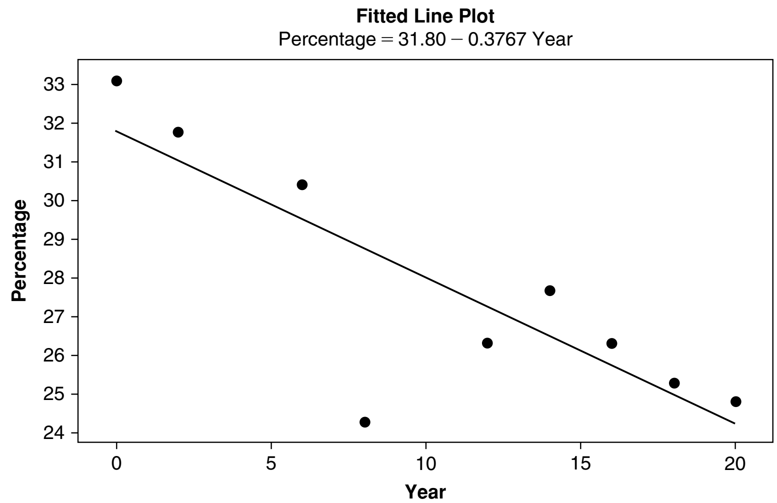
12.4 ERROR RANDOM VARIABLE

Problems

2. $n = 10, S_{xx} = 70, S_{YY} = 52,940, S_{xY} = -1865, SS_R = 3251.072$. Hence, the estimate of $\sigma^2 = 3251.072/(10 - 2) = 406.384$.

4. $n = 10, S_{xx} = 6594.102, S_{YY} = 842.6548, S_{xY} = -2320.261$,
$SS_R = 26.2282$. Hence, the estimate of $\sigma^2 = 26.2282/(10 - 2) = 3.2785$.

6. (a) and (b)



From the regression analysis, the estimate for $\sigma^2 = (1.9311)^2 = 3.7291$.

Fitted Line: Percentage versus Year
Regression Analysis: Percentage versus Year

```
The Regression Equation Is
Percentage = 31.8 - 0.377 Year

Predictor         Coef  SE Coef       T      P
Constant        31.796    1.215   26.18  0.000
Year          -0.37667  0.09656   -3.90  0.006


S = 1.93112    R-Sq = 68.5%    R-Sq(adj) = 64.0%

Analysis of Variance

Source              DF       SS      MS      F      P
Regression           1   56.751  56.751  15.22  0.006
Residual Error       7   26.104   3.729
Total                8   82.856
```

Also, from the regression equation for the year 1997, $x = 15$. Thus,

$Y = 31.8 - 0.377(15) = 26.145$

8. (a) Now, $\hat{\beta} = S_{xY}/S_{xx} = 12.2/46.8 = 0.2607,$ and
$\hat{\alpha} = (14.6) - (0.2607)(13.9) = 10.9763.$
Thus, the estimated regression equation is
$Y = 10.9763 + 0.2607x$, where Y is the age at which the son began to shave and x is the age at which the father began to shave.

(b) If $x = 15.1$, then the predicted age at which the son will begin to shave is $Y = 10.9763 + 0.2607(15.1) = 14.9129$ (years).

(c) $SS_R = [(S_{XX})(S_{YY}) - (S_{XY})^2]/(S_{XX}) = [(46.8)(53.3) - (12.2)^2]/46.8 = 50.1197$. Hence, the estimate of $\sigma^2 = 50.1197/(25 - 2) = 2.1791$.

12.5 TESTING THE HYPOTHESIS THAT $\beta = 0$

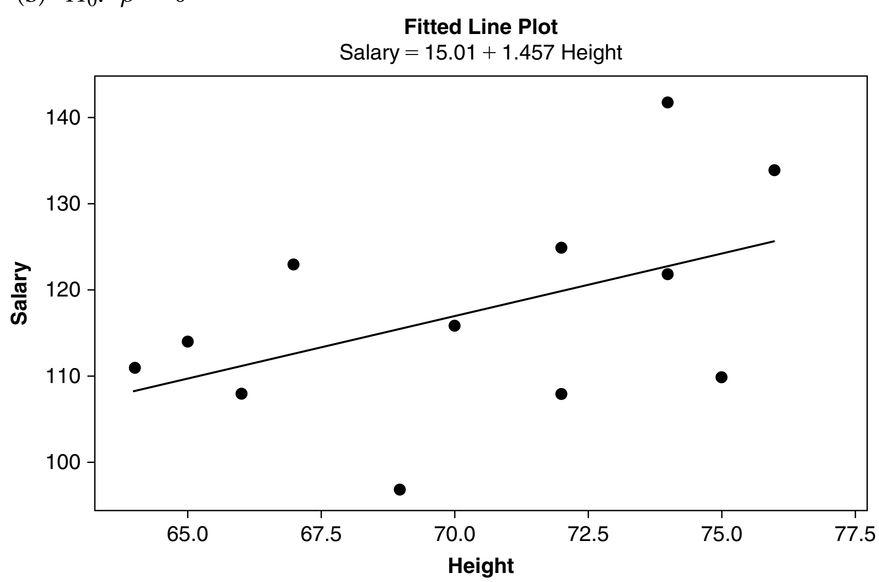
Problems

2. $n = 12, \gamma = 0.05, t_{10,0.025} = 2.228,\ S_{xx} = 186.667,\ S_{YY} = 1693,$

$S_{xY} = 272, SS_R = 1296.66, \sqrt{(n-2)S_{xx}/SS_R} = 1.1998, \quad \hat{\alpha} = 15.0143,$
$\hat{\beta} = 1.4571.$

(a) H_0: $\beta = 0$ against H_1: $\beta \neq 0$
TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\ \hat{\beta} = (1.1998)(1.4571) = 1.7482$
Conclusion: Since $|T| = 1.7482 < 2.228$, do not reject H_0. That is, there is insufficient sample evidence to claim that a male lawyer's salary is related to his height at the 5% level of significance.
The p value for this test is 0.111 as shown in the ***Minitab*** output.
Hence again, you will not reject H_0: $\beta = 0$.

(b) H_0: $\beta = 0$



Regression Analysis: Salary versus Height

```
The Regression Equation Is
Salary = 15.0 + 1.46 Height


Predictor       Coef  SE Coef      T      P
Constant       15.01    58.71   0.26  0.803
Height        1.4571   0.8334   1.75  0.111


S = 11.3871   R-Sq = 23.4%      R-Sq(adj) = 15.8%


Analysis of Variance

Source            DF       SS     MS      F      P
Regression         1    396.3  396.3   3.06  0.111
Residual Error    10   1296.7  129.7
Total             11   1693.0
```

4. $n = 14, \gamma = 0.05, t_{12,0.025} = 2.179, S_{xx} = 501.4297, S_{YY} = 28.0894,$
$S_{xY} = 104.7441,\ SS_R = 6.2093,\ \sqrt{(n-2)S_{xx}/SS_R} = 31.1298,\ \hat{\alpha} = 4.9476,$
$\hat{\beta} = 0.2089.$

(a) $Y = 4.9476 + 0.2089x$

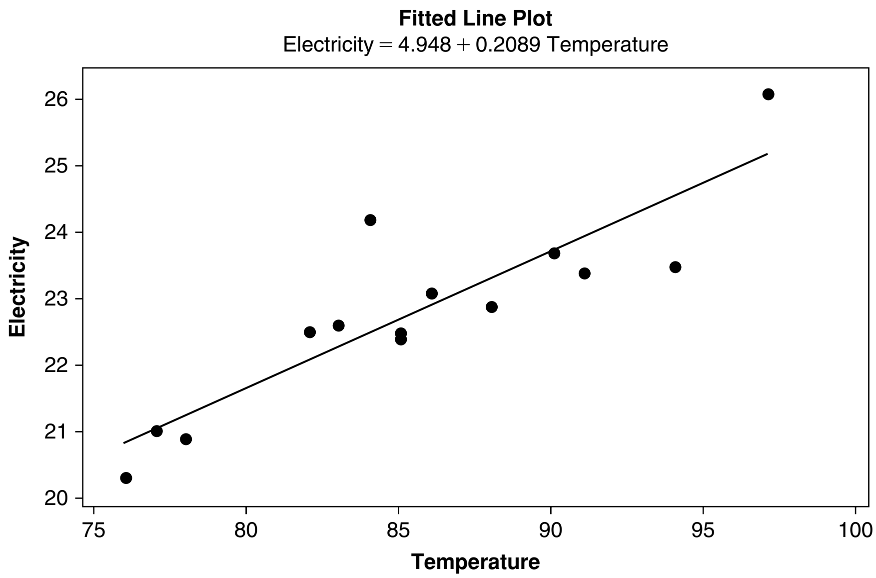
(b) For $x = 93$ (temperature), the predicted electricity consumption will be $Y = 4.9476 + (0.2089)(93) = 24.3753$ (millions of kilowatts).

(c) $H_0{:}\, \beta = 0$ against $H_1{:}\, \beta \neq 0$

TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\ \hat{\beta} = (31.1298)(0.2089) = 6.5030$

Conclusion: Since $|T| = 6.5030 > 2.179$, reject H_0. That is, there is sufficient sample evidence to claim that the daily temperature has an effect on the amount of electricity consumed at the 5% level of significance.

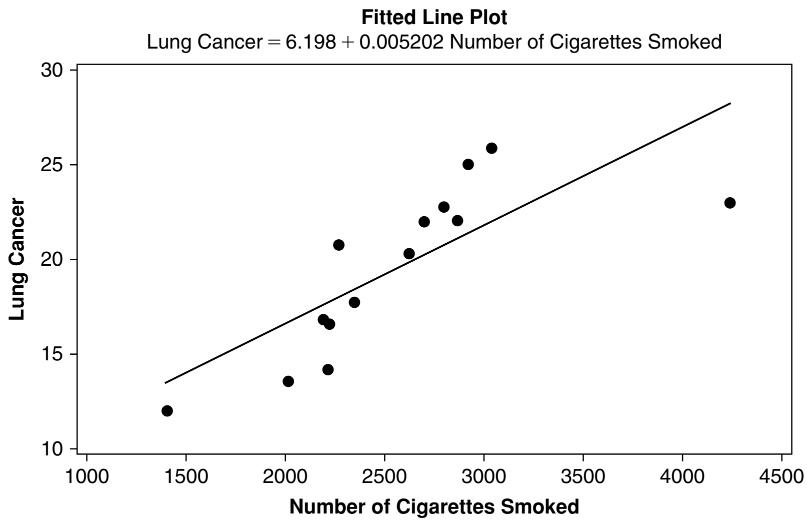
Following are ***Minitab*** outputs that support the computations using the formulas.



Regression Analysis: Electricity versus Temperature

```
The Regression Equation Is
Electricity = 4.95 + 0.209 Temperature


Predictor        Coef  SE Coef      T      P
Constant        4.948    2.751   1.80  0.097
Temperature   0.20889  0.03212   6.50  0.000


S = 0.719355 R-Sq = 77.9% R-Sq(adj) = 76.1%


Analysis of Variance

Source           DF      SS      MS      F      P
Regression        1  21.880  21.880  42.28  0.000
Residual Error   12   6.210   0.517
Total            13  28.089
```

6. (a) The scatter diagram of cigarettes smoked versus death rate from lung cancer is shown below.



(c) and (d)

From the ***Minitab*** output for the regression analysis, observe that the p value for H_0: $\beta = 0$ is 0.001. Hence, at the 5% or the 1%, H_0 will be rejected. Thus, one can conclude that cigarette consumption does affect the death rate from lung cancer.

Regression Analysis: Lung Cancer versus Number of Cigarettes Smoked

```
The Regression Equation Is
Lung Cancer = 6.20 + 0.00520 Number of
              Cigarettes Smoked

Predictor          Coef   SE Coef       T       P
Constant          6.198     3.212    1.93   0.078
Number of Cigarettes Smoked
             0.005202  0.001221    4.26   0.001

S = 2.86660 R-Sq = 60.2% R-Sq(adj) = 56.9%

Analysis of Variance

Source             DF        SS       MS       F      P
Regression          1    149.13   149.13   18.15  0.001
Residual Error     12     98.61     8.22
Total              13    247.74
```

(b) The estimated regression line is $Y = 6.198 + 0.0052x$.

(c) Using formulas,

$n = 14, \gamma = 0.05, t_{12,0.025} = 2.179, S_{xx} = 5510456, S_{YY} = 247.7412,$
$S_{xY} = 28666.88, SS_R = 98{,}6084, \sqrt{(n-2)S_{xx}/SS_R} = 818.8934,$
$\hat{\alpha} = 6.1976,\ \hat{\beta} = 0.0052.$

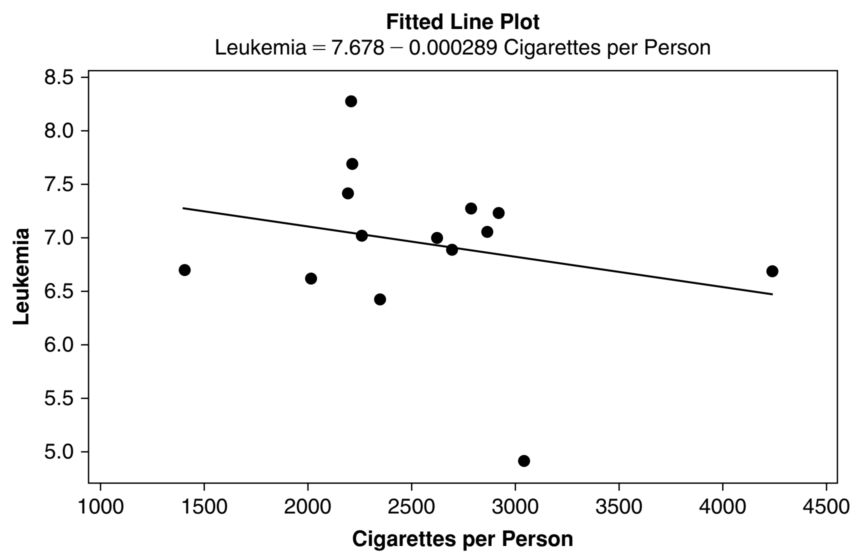
H_0: $\beta = 0$ against H_1: $\beta \neq 0$

TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\ \hat{\beta} = (818.8934)(0.0052) = 4.26$

Conclusion: Since $T = 4.26 > 2.179$, reject H_0. That is, there is sufficient sample evidence to claim that cigarette consumption does affect the death rate from lung cancer at the 5% level of significance.

(d) For $\gamma = 0.01,\ t_{12,0.005} = 3.055$. Hence same conclusion as in part (c).

8. (a) The scatter diagram of cigarettes smoked versus death rate from leukemia is shown below.



(b) The estimated regression line is $Y = 7.6785 - 2.89 \times 10^{-4}x$.

(c) $n = 14, \gamma = 0.05,\ t_{12,\,0025} = 2.179,\ S_{xx} = 5510456,\ S_{YY} = 7.4791,$
$S_{xY} = -1591.547,\ SS_R = 7.0195,\ \sqrt{(n-2)S_{xx}/SS_R} = 3069.254,$
$\hat{\alpha} = 7.6785,\ \hat{\beta} = -2.8882x10.$

H_0: $\beta = 0$ against H_1: $\beta \neq 0$

TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\ \hat{\beta} = (3069.254)(-0.00028882) = -0.8865$

Conclusion: Since $|T| = 0.8865 < 2.179$, do not reject H_0. That is, there is insufficient sample evidence to claim that cigarette consumption does affect the death rate from leukemia at the 5% level of significance.

(d) For $\gamma = 0.01$, $t_{12,0.005} = 3.055$. Hence same conclusion as in part (c).

Following is a ***Minitab*** output with some regression analysis that confirms the previous computations.

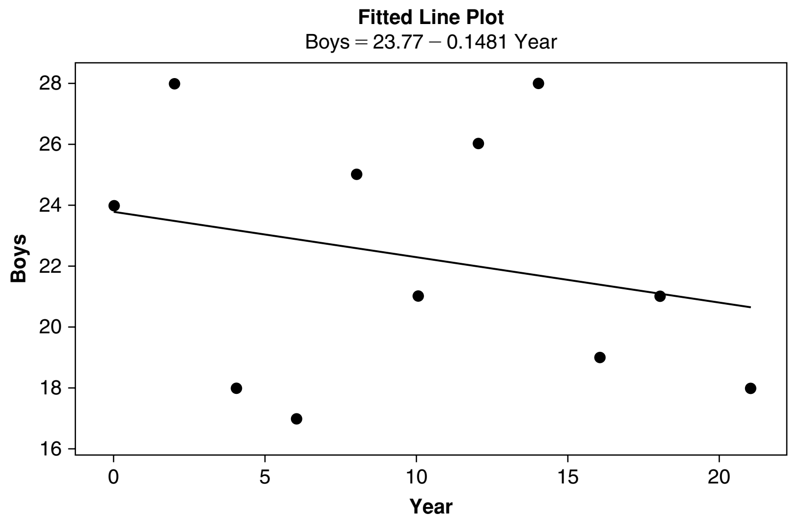
Regression Analysis: Leukemia versus Cigarettes per Person

```
The Regression Equation is
Leukemia = 7.68 - 0.000289 Cigarettes per Person

Predictor            Coef     SE Coef        T       P
Constant           7.6785      0.8570     8.96   0.000
Cigarettes     -0.0002888   0.0003258    -0.89   0.393
per Person

S = 0.764822 R-Sq = 6.1% R-Sq(adj) = 0.0%
Analysis of Variance

Source           DF       SS       MS      F       P
Regression        1   0.4597   0.4597   0.79   0.393
Residual Error   12   7.0194   0.5850
Total            13   7.4791
```

10. (a) Following is a ***Minitab*** output to help with the analysis.

Regression Analysis: Boys versus Year

```
The Regression Equation Is
Boys = 23.8 - 0.148 Year

Predictor        Coef  SE Coef       T       P
Constant       23.767    2.340   10.16   0.000
Year          -0.1481   0.1952   -0.76   0.467

S = 4.19085 R-Sq = 6.0%  R-Sq(adj) = 0.0%

Analysis of Variance

Source            DF       SS     MS        F       P
Regression         1    10.11  10.11     0.58   0.467
Residual Error     9   158.07  17.56
Total             10   168.18
```



Note: Year is used here as the input variable with $1982 \equiv 0$, etc.
From the ***Minitab*** output, we have

H_0: $\beta = 0$ against H_1: $\beta \neq 0$

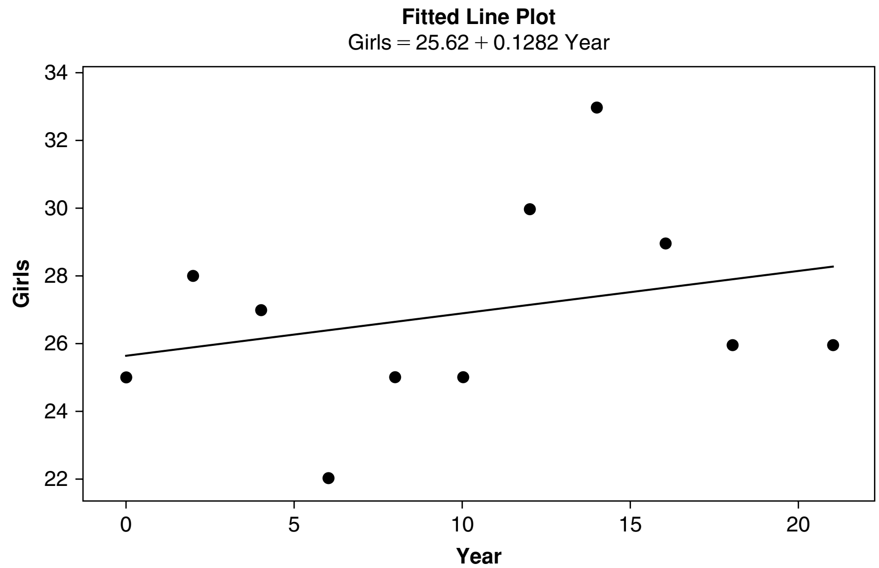
TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\hat{\beta} = -0.76$ with a p value $= 0.467$

Conclusion: Since p value $= 0.467 > 0.05$, do not reject H_0. That is, there is insufficient sample evidence to refute the claim that the percentage of boys who smoke is unchanging over time at the 5% level of significance.

(b) Following is a ***Minitab*** output to help with the analysis.

Regression Analysis: Girls versus Year

```
The Regression Equation Is
Girls = 25.6 + 0.128 Year

Predictor       Coef   SE Coef       T       P
Constant      25.615     1.679   15.26   0.000
Year          0.1282    0.1400    0.92   0.384

S = 3.00617   R-Sq = 8.5%  R-Sq(adj) = 0.0%

Analysis of Variance

Source            DF        SS      MS       F       P
Regression         1     7.576   7.576    0.84   0.384
Residual Error    9    81.333   9.037
Total             10    88.909
```

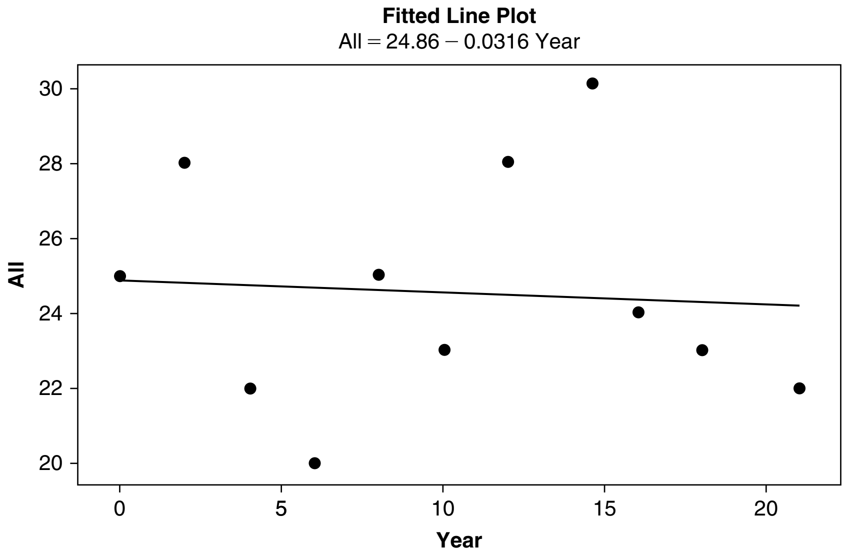


From the ***Minitab*** output, we have
H_0: $\beta = 0$ against H_1: $\beta \neq 0$
TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\hat{\beta} = 0.92$ with a p value $= 0.384$
Conclusion: Since p value $= 0.384 > 0.05$, do not reject H_0. That is, there is insufficient sample evidence to refute the claim that the percentage of girls who smoke is unchanging over time at the 5% level of significance.

(c) Following is a ***Minitab*** output to help with the analysis.

Regression Analysis: All versus Year

```
The Regression Equation Is
All = 24.9 - 0.032 Year

Predictor       Coef   SE Coef      T       P
Constant      24.864     1.788  13.91   0.000
Year         -0.0316    0.1491  -0.21   0.837

S = 3.20188   R-Sq = 0.5%   R-Sq(adj) = 0.0%

Analysis of Variance

Source            DF        SS     MS       F       P
Regression         1      0.46   0.46    0.04   0.837
Residual Error     9     92.27  10.25
Total             10     92.73
```



From the ***Minitab*** output, we have

H_0: $\beta = 0$ against H_1: $\beta \neq 0$

TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\hat{\beta} = -0.21$ with a p value $= 0.837$

Conclusion: Since p value $= 0.837 > 0.05$, do not reject H_0. That is, there is insufficient sample evidence to refute the claim that the percentage of 15-year-olds who smoke is unchanging over time at the 5% level of significance.

12.6 REGRESSION TO THE MEAN

Problems

2. $n = 12,\ \gamma = 0.05,\ t_{10,0.05} = 1.812,\ S_{xx} = 1706,\ S_{YY} = 872.25,$

$S_{xY} = 471,\ SS_R = 742.2142,\ \sqrt{(n-2)S_{xx}/SS_R} = 4.7943,$

$\hat{\alpha} = 74.5894,\ \hat{\beta} = 0.2761.$

H_0: $\beta \geq 1$ against H_1: $\beta < 1$

TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\ (\hat{\beta} - \beta) = (4.7943)(0.2761 - 1) = -3.4706$

Conclusion: Since $T = -3.4706 < -1.812$, reject H_0. That is, there is sufficient sample evidence to claim a regression toward the mean number of deaths at the 5% level of significance.

4. $n = 7,\ \gamma = 0.05,\ t_{5,\,0.05} = -2.015,\ S_{xx} = 97.71429,\ S_{YY} = 61.71429,$

$S_{xY} = -3.7143, SS_R = 61.5731, \sqrt{(n-2)S_{xx}/SS_R} = 2.8169,$

$\hat{\alpha} = 22.3860, \hat{\beta} = -3.8012 \times 10^{-2}.$

H_0: $\beta = 1$ against H_1: $\beta < 1$

TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\ (\hat{\beta} - \beta) = (2.8169)(-0.0380 - 1) = -2.9239$

Conclusion: Since $T = -2.9239 < -2.015$, reject H_0. That is, there is sufficient sample evidence to claim a regression toward the mean at the 5% level of significance.

6. The distribution of the heights appears to be approximately normally distributed.

12.7 PREDICTION INTERVALS FOR FUTURE RESPONSES

Problems

2. $n = 10, \gamma = 0.05, t_{8,0.025} = 2.306, S_{xx} = 967.5996, S_{YY} = 148.9,$

$S_{xY} = -324.8001, SS_R = 39.8724, \sqrt{(n-2)S_{xx}/SS_R} = 13.9334,$

$\hat{\alpha} = 17.1102, \hat{\beta} = -0.3357,\ W = 2.4328.$

(a) When $x = 42$ (age), the predicted Y (number of days missed) is 3.0100.

(b) The 95% prediction interval when $x = 42$ is $3.0100 \pm (2.306)(2.4328)$ or 3.0100 ± 5.603.
That is, we can be 95% confident that the number of days missed by a 42-year-old worker will lie between -2.593 and 8.613.

4. $n = 10, \gamma = 0.1, t_{8,0.05} = 1.86, S_{xx} = 598.3984, S_{YY} = 1.1610,$

$S_{xY} = 22.08, SS_R = 0.3463, \sqrt{(n-2)S_{xx}/SS_R} = 117.5784$

$\hat{\alpha} = -0.1147, \hat{\beta} = 3.6899 \times 10^{-2}, W = 0.2214.$

(a) When $x = 88$ (entrance exam score), the predicted Y (grade point average) is 3.1324.

(b) The 90% prediction interval when $x = 88$ is $3.1324 \pm (1.86)(0.2214)$ or 3.1324 ± 0.4118.
That is, we can be 90% confident that the grade point average for a student with an entrance exam score of 88 will lie between 2.7206 and 3.5442.

(c) $\gamma = 0.05, t_{8,0.025} = 2.306.$
H_0: $\beta = 0$ against H_1: $\beta \neq 0$
TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\ \hat{\beta} = (117.5784)(0.0369) = 4.3386$
Conclusion: Since $|T| = 4.3386 > 2.306$, reject H_0. That is, there is sufficient sample evidence to claim that the student's grade point average is not independent of his or her score on the entrance exam at the 5% level of significance.

6. $\overline{X} = 12.8, \overline{Y} = 12.9, S_{xx} = 36.5, S_{YY} = 42.4, S_{xY} = 24.4.$

(a) Now, $\hat{\beta} = S_{xY}/S_{xx} = 24.4/36.5 = 0.6685$, and $\hat{\alpha} = (12.9) - (0.6685)(12.8) = 4.3432$.
Thus, the estimated regression equation is $Y = 4.3432 + 0.6685x$, where Y is the daughter's age at puberty and x is the mother's age at puberty.

(b) $SS_R = [(S_{XX})(S_{YY}) - (S_{XY})^2]/(S_{XX}) = [(36.5)(42.4) - (24.4)^2]/36.5 = 26.0888$

(c) $\gamma = 0.05, t_{18;0.025} = 2.101$
H_0: $\beta = 0$ against H_1: $\beta \neq 0$
TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\hat{\beta} = (5.0183)(0.6685) = 3.3547$
Conclusion: Since $|T| = 3.3547 > 2.101$, reject H_0. That is, there is sufficient sample evidence to claim that the mother's age at puberty and the daughter's age at puberty are significantly related at the 5% level.

(d) Now $W = 1.7237$ and for $x = 13.8$, the predicted Y value is 13.5685. Thus, the 95% prediction interval for this predicted value is $13.5685 \pm (2.101)(1.7237)$ or 13.5685 ± 3.6215. That is, we can be 95% confident that the age at which the daughter will experience puberty will lie between 9.947 and 17.19 years if the age at which the mother experience puberty is at age 13.8 years.

12.8 COEFFICIENT OF DETERMINATION

Problems

2. $n = 3, S_{xx} = 4.6667, S_{YY} = 72, S_{xY} = 18, SS_R = 2.5714,$

$\sqrt{(n-2)\ S_{xx}/SS_R} = 1.3471, \hat{\alpha} = 3.1429, \hat{\beta} = 3.8571.$

The coefficient of determination:

$R^2 = \frac{S_{YY} - SS_R}{S_{YY}} = (72 - 2.5714)/72 = 0.9643$

4. $n = 12, S_{xx} = 11572.91, S_{YY} = 2922.918, S_{xY} = 5792.914,$

$SS_R = 23.2270, \sqrt{(n-2)S_{xx}/SS_R} = 70.587, \hat{\alpha} = -6.1242, \hat{\beta} = 0.5006.$

The coefficient of determination:

$R^2 = \frac{S_{YY} - SS_R}{S_{YY}} = (2922.918 - 23.2270)/2922.918 = 0.9919$

6. $n = 6, \gamma = 0.05, t_{4,0.025} = 2.776, S_{xx} = 24, S_{YY} = 160, S_{xY} = 56,$

$SS_R = 29.3333, \sqrt{(n-2)S_{xx}/SS_R} = 1.8091, \hat{\alpha} = 6.3333, \hat{\beta} = 2.3333.$

(a) The estimated regression line is $Y = 6.3333 + 2.3333x$, where Y is the number of cars sold and x is the number of salespeople.

(b) The coefficient of determination:
$R^2 = \frac{S_{YY} - SS_R}{S_{YY}} = (160 - 29.3333)/160 = 0.8167$

(c) From part (b), 81.67% of the variation in the number of cars sold is explained by the number of salespeople working.

(d) H_0: $\beta = 0$ against H_1: $\beta \neq 0$

TS: $T = [\sqrt{(n-2)S_{xx}/SS_R}]\hat{\beta} = (1.8091)(2.3333) = 4.2212$

Conclusion: Since $|T| = 4.2212 > 2.776$, reject H_0. That is, there is sufficient sample evidence to claim that the number of cars sold and the number of salespeople working are related at the 5% significance level.

12.9 SAMPLE CORRELATION COEFFICIENT

Problems

2. Because multiplication is commutative, $S_{uv} = S_{vu}$ and $S_{UU}S_{VV} = S_{VV}S_{UU}$.

Thus, $r = \frac{S_{uv}}{\sqrt{S_{uu}S_{vv}}} = \frac{S_{vu}}{\sqrt{S_{vv}S_{uu}}}$. That is, it does not matter whether u is the input variable or v is the input variable, the correlation between u and v will be the same.

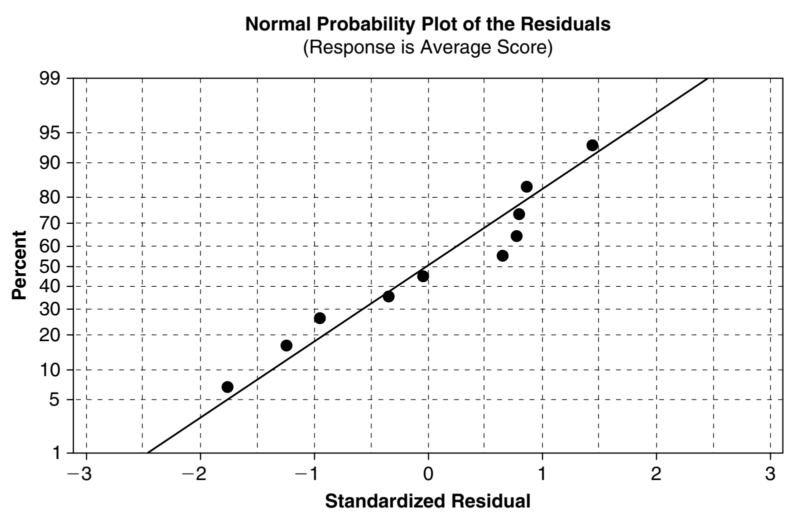
4. $r = 0.95$, then $R^2 = (0.95)^2 = 0.9025$. That is, 90.25% of the variability in the responses is explained by the different input variables.

6. $S_{xx} = 36.5, S_{YY} = 42.4, S_{xY} = 24.4$. Since the sample correlation coefficient $r = \frac{S_{xY}}{\sqrt{S_{xx}S_{YY}}}$, $r = 0.6202$.

12.10 ANALYSIS OF RESIDUALS: ASSESSING THE MODEL

Problems

2. Example 6, Section 12.3



The plot of the standardized residuals seems to follow a linear pattern. This would indicate that a linear model might be appropriate for the regression fit.

12.11 MULTIPLE LINEAR REGRESSION MODEL

Problems

2. The estimated multiple regression equation is $Y = -169.5333 + 0.3933X_1 + 0.8667X_2$, where Y is the yield of the experiment, X_1 is the temperature (degrees Fahrenheit), and X_2 is the pressure (pounds per square inch). For $X_1 = 150$ and $X_2 = 215$, the predicted yield is

$$Y = -169.5333 + (0.3933)(150) + (0.8667)(215) = 75.8022$$

4. (a) The estimated multiple regression equation is $Y = 1063 - 234.8X_1 - 6.22X_2$, where Y is the survival time in days, X_1 is the mismatch score, and X_2 is the age (years) of the patient.

(b) For $X_1 = 1.46$ and $X_2 = 50$, the predicted survival time is $Y = 1063 - (234.8)(1.46) - (6.22)(50) = 409.192$ days.

Review Problems

2. $n = 6, \gamma = 0.01, t_{4,0.005} = 4.604, S_{xx} = 28683.33, S_{yy} = 2.8601,$

$S_{xY} = -285, SS_R = 0.0282, \sqrt{(n-2)S_{xx}/SS_R} = 2014.843,$
$\hat{\alpha} = 9.4777, \hat{\beta} = -0.0099, \overline{X} = 88.3333.$

(a) The estimated regression line is $Y = 9.4777 - 0.0099x$, where Y is the cost per unit and x is the number of units. Thus, the predicted cost per unit when $x = 125$ is $Y = 8.2402$.

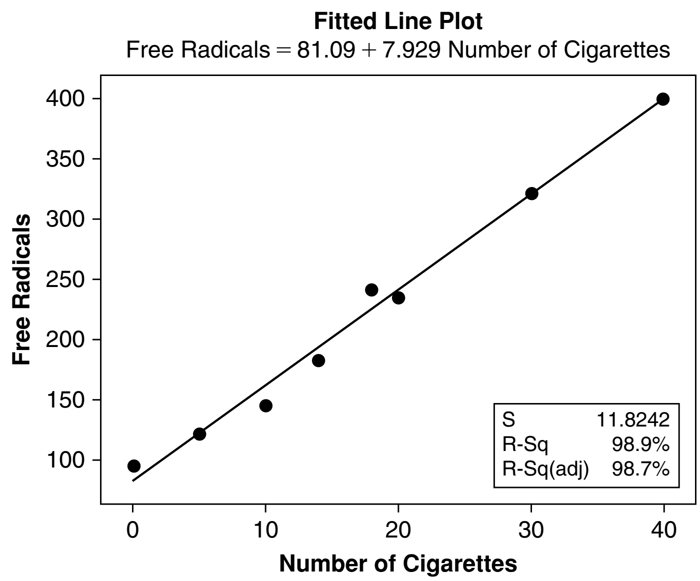
(b) The estimated variance for the cost when $x = 125$ is $W^2 = 0.0086$.

(c) When $x = 110$ (number of units), the predicted Y (cost per unit) is 8.3847 and W is 0.0914. The 99% prediction interval for the cost per unit when $x = 110$ is $8.3847 \pm (4.604 \times 0.0914)$ or 8.3847 ± 0.4208. That is, we can be 99% confident that the cost per unit will lie between 7.9639 and 8.8055 dollars when the production level is 125.

4. Yes, if people with low IQ scores marry people with low IQ scores and people with high IQ scores marry people with high IQ scores. No, otherwise.

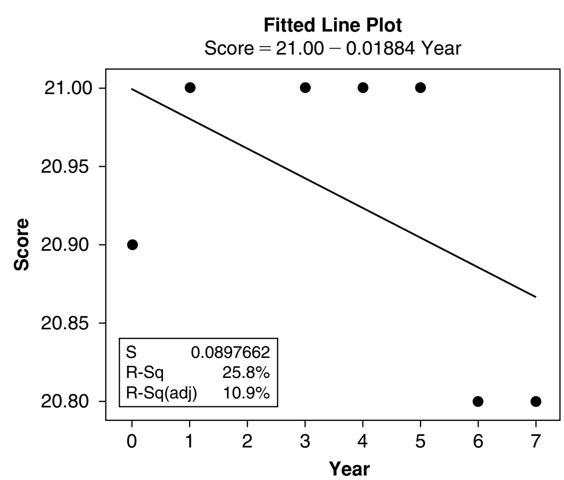
6. (a) and (c) are shown in the scatter plot below.

(d) The estimated regression line is $Y = 81.0874 + 7.9292x$, where Y is the free radicals and x is the number of cigarettes. Thus, the predicted amount of free radicals when $x = 26$ is $Y = 287.2467$.



(e) When $x = 26$, the predicted Y is 287.2467 and W is 12.9024. The 95% prediction interval for the free radicals when $x = 26$ is $287.2467 \pm (2.447)(12.9024)$ or 287.2467 ± 31.5722. That is, we can be 95% confident that the number of free radicals will lie between 255.6745 and 318.8189 when the average number of cigarettes smoked is 26.

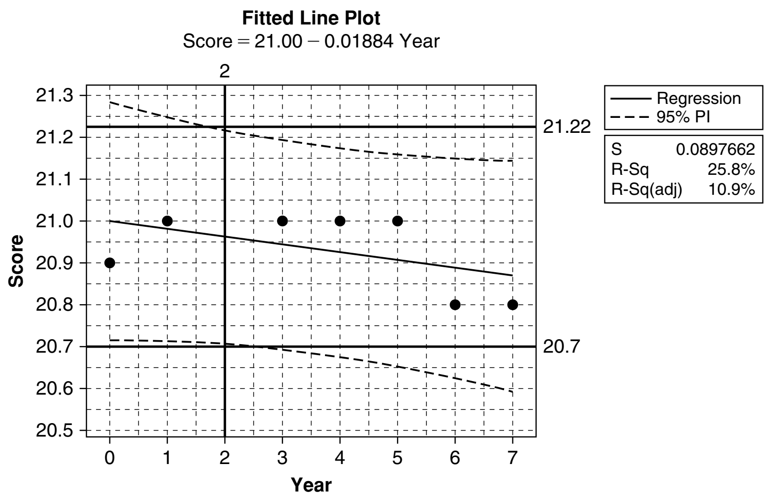
8. Following is a ***Minitab*** fitted line plot.



(a) The estimated regression line is $Y = 21.00 - 0.0188x$, where Y is the average score on the ACT exam and x is the year. Thus, the predicted score when $x = 1998$ is when $x = 2$. Thus, $Y = 20.9624$.

(b) The following ***Minitab*** graph shows the prediction interval around the regression line.

For the year 1998 (i.e., when $x = 2$), the 95% prediction interval for the ACT score will lie between 20.7 and 21.22 (approximately).



10. (111.565, 146.459)

12. (a) The estimated regression equation is $Y = -14.7 + 1.35x_1 + 0.419x_2$, where Y represents the wheat yield, x_1 represents the amount of rainfall, and x_2 represents the amount of fertilizer.

(b) The estimate of the additional yield in wheat for each additional inch of rain is 1.35 (assuming that the amount of fertilizer does not change).

(c) The estimate of the additional yield in wheat for each additional pound of fertilizer is 0.419 (assuming that the amount of rainfall does not change).

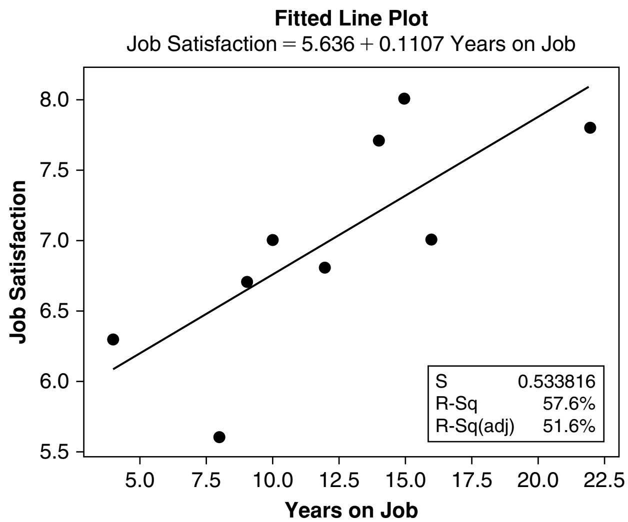
(d) The predicted yield will be $Y = -14.7 + (1.35)(26) + (0.419)(130) = 74.87$.

14. $n = 9, S_{xx} = 221.5556, S_{yy} = 4.7089, S_{xY} = 24.5223, SS_R = 1.9947,$

$\sqrt{(n-2)S_{xx}/SS_R} = 27.8839$.

(a) $\hat{\alpha} = 5.6361$, and $\hat{\beta} = 0.1107$.

(b) Since $\hat{\beta} = 0.1107 > 0$, it appears that job satisfaction increases as the years of service increase. This is depicted in the scatter plot below.



(c) The estimated regression equation for the job satisfaction ratings in problem 13 is $Y = -0.3953 + 0.162x_1 - 0.113x_2$, where x_1 represents the yearly income and x_2 represents the years on the job. Here, when the yearly income remains fixed, the job satisfaction rating decreases by 0.113 when the number of years on the job increases by 1 year.

(d) Job satisfaction is not only a function of years of service as was given in this problem. Other factors need to be considered such as income. Specifically, longevity at a job tends to increase one's income, which may lead to greater job satisfaction.

Chapter 13 CHI-SQUARED GOODNESS-OF-FIT TESTS

13.2 CHI-SQUARED GOODNESS-OF-FIT TESTS

Problems

2. (a) H_0: $P_i = 1/4$, for $i = 1, 2, 3, 4$ against H_1: $P_i \neq 1/4$ for at least one i

(b) $n = 200, N_1 = 44, N_2 = 38, N_3 = 57, N_4 = 61, P_1 = 1/4,$
$e_i = (200)(1/4) = 50.$
TS: $\chi^2 = 7$

(c) $\alpha = 0.1, \chi^2_{3,0.1} = 6.25.$
Conclusion: Since $7 > 6.25$, reject H_0. That is, there is sufficient evidence to claim that the observed values (1, 2, 3, 4) are not equally likely to occur at the 10% significance level.

(d) $\alpha = 0.05, \chi^2_{3,0.05} = 7.81.$
Conclusion: Since $7 < 7.81$, do not reject H_0. That is, there is insufficient evidence to claim that the observed values (1, 2, 3, 4) are not equally likely to occur at the 5% significance level.

(e) $\alpha = 0.01, \chi^2_{3,0.01} = 11.34.$
Conclusion: Since $7 < 11.34$, do not reject H_0. That is, there is insufficient evidence to claim that the observed values (1, 2, 3, 4) are not equally likely to occur at the 1% level of significance.

4. $n = 100, N_1 = 27, N_2 = 19, N_3 = 13, N_4 = 15, N_5 = 26, P_i = 1/5,$
$e_i = (100)(1/5) = 20, \alpha = 0.05, \chi^2_{4,0.05} = 9.49.$
H_0: $P_i = 1/5$, for $i = 1, 2, 3, 4, 5$ against H_1: $P_i \neq 1/5$ for at least one i
TS: $\chi^2 = 8$
Conclusion: Since $8 < 9.49$, do not reject H_0. That is, there is insufficient evidence to claim that the observed absences are not equally likely to occur at the 5% level of significance.

6. $n = 400, N_1 = 308, N_2 = 66, N_3 = 26, P_1 = 0.84, P_2 = 0.14, P_3 = 0.02,$
$e_1 = (400)(0.84) = 336, e_2 = (400)(0.14) = 56, e_3 = (400 \times 0.02) = 8.$
H_0: $P_1 = 0.84, P_2 = 0.14, P_3 = 0.02$ versus H_2: At least one P_i is different
TS: $\chi^2 = 44.6190$
Conclusion: Since $44.169 > 5.99$, reject H_0. That is, there is sufficient evidence to claim that the lawyers do not exhibit the same accident profile as the rest of the drivers in the region at the 5% level of significance.

8. Observed values for the die will vary.

10. $n = 1100, N_1 = 156, N_2 = 144, N_3 = 170, N_4 = 158, N_5 = 172,$
$N_6 = 148, N_7 = 152, P_1 = 1/7, e_i = (1100)(1/7) = 157.1429, \alpha = 0.05,$
$\chi^2_{6,0.05} = 12.59.$
H_0: $P_i = 1/7$, for $i = 1, 2, \ldots, 7$ against H_1: $P_i :\neq 1/7$ for at least one i
TS: $\chi^2 = 4.2691$
Conclusion: Since $4.2691 < 12.59$, do not reject H_0. That is, there is insufficient evidence to claim that the earthquakes are not equally likely to occur on any of the 7 days at the 5% level of significance.

12. $n = 240, N_1 = 24, N_2 = 94, N_3 = 48, N_4 = 35, N_5 = 39, P_1 = 0.137,$
$P_2 = 0.325, P_3 = 0.263, P_4 = 0.171, P_5 = 0.104,$
$e_1 = (240)(0.137) = 32.88, e_2 = (240)(0.325) = 78,$
$e_3 = (240)(0.263) = 63.12, e_4 = (240)(0.171) = 41.04,$
$e_5 = (240)(0.104) = 24.96, \alpha = 0.05, \chi^2_{4,0.05} = 9.49.$
H_0: $P_1 = 0.137, P_2 = 0.325, P_3 = 0.263, P_4 = 0.171, P_5 = 0.104$ against
H_1: At least one P_i is different
TS: $\chi^2 = 18.0886$
Conclusion: Since $18.0886 > 9.49$, reject H_0. That is, there is sufficient evidence to claim that the age breakdown of Sacramento workers on flexible schedules is different from the national breakdown at the 5% level of significance.

14. $n = 24{,}000, N_1 = 12{,}012, N_2 = 11{,}988, P_1 = 0.5, P_2 = 0.5,$
$e_1 = (24{,}000)(0.5) = 12{,}000, e_2 = (24{,}000)(0.5) = 12{,}000, \alpha = 0.05,$
$\chi^2_{1,0.05} = 3.84.$
H_0: $P_1 = 0.5, P_2 = 0.5$ against H_1: At least one P_i is different
TS: $\chi^2 = 0.024$
Conclusion: Since $0.024 < 3.84$, do not reject H_0. That is, there is insufficient evidence to claim that the coin is not fair at the 5% level. Also, since 0.024 is greater than $\chi^2_{1,0.95} = 0.004$, it follows that a value at least as small as 0.024 would occur over 5% of the time. Thus, the data is not indicative of any manipulation.

13.3 TESTING FOR INDEPENDENCE IN POPULATIONS CLASSIFIED ACCORDING TO TWO CHARACTERISTICS

Problems

2. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Had Pet, No Pet

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

           Had Pet    No Pet    Total
     1          28        44       72
             27.28     44.72
             0.019     0.011

     2           8        15       23
              8.72     14.28
             0.059     0.036

  Total         36        59       95

  Chi-square = 0.125, DF = 1, p value = 0.724
```

H_0: Owning a pet and survival after 1 year of a heart attack are independent.

H_1: Owning a pet and survival after 1 year of a heart attack are not independent.

TS: $\chi^2 = 0.125$ with a p value $= 0.724$

Conclusion: Since p value $= 0.724 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that owning a pet and survival after 1 year of a heart attack are not independent.

4. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Smokers, Nonsmokers

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

        Smokers    Nonsmokers    Total
  1          35           170      205
          58.72        146.28
          9.581         3.846

  2          79           190      269
          77.05        191.95
          0.049         0.020
```

(*Continued*)

(*Continued*)

```
          3          57         66      123
                  35.23      87.77
                 13.451      5.399

      Total         171        426      597

      Chi-square = 32.346, DF = 2, p value = 0.000
```

H_0: Accident frequency and smoking habits of a policyholder are independent.

H_1: Accident frequency and smoking habits of a policyholder are not independent.

TS: $\chi^2 = 32.346$ with a p value $= 0.000$

Conclusion: Since p value $= 0.000 < \alpha = 0.05$, reject H_0. That is, there is sufficient sample evidence to conclude that accident frequency and smoking habits of the policyholders are not independent.

6. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: 1, 2, 3 or more

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

              1       2   3 or more   Total
       1     12      10           4      26
           7.37    8.96        9.68
          2.914   0.122       3.331

       2     32      40          38     110
          31.17   37.89       40.94
          0.022   0.118       0.212

       3      7      12          25      44
          12.47   15.16       16.38
          2.397   0.657       4.539

   Total     51      62          67     180

   Chi-square = 14.312, DF = 4, p value = 0.006
```

H_0: Teaching performance and the number of courses taught are independent.

H_1: Teaching performance and the number of courses taught are not independent.

TS: $\chi^2 = 14.312$ with a p value $= 0.006$

Conclusion: Since p value $= 0.006 < \alpha = 0.05$, reject H_0. That is, there is sufficient sample evidence to conclude that teaching performance and the number of courses taught are not independent.

8. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Lower, Middle

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.
          Lower   Middle   Total
      1      87       63     150
          79.48    70.52
          0.711    0.801

      2      46       55     101
          53.52    47.48
          1.056    1.190

   Total    133      118     251

   Chi-square = 3.759, DF = 1, p value = 0.053
```

H_0: Attitude toward new clinic and socioeconomic status is independent.

H_1: Attitude toward new clinic and socioeconomic status is not independent.

TS: $\chi^2 = 3.759$ with a p value $= 0.053$

Conclusion: Since p value $= 0.053 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that attitude toward the new clinic and socioeconomic status is not independent.

10. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Low, Moderate, High

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

          Low   Moderate    High   Total
     1      6         14      24      44
        11.28      18.98   13.75
        2.468      1.304   7.641

     2     12         23      15      50
        12.81      21.56   15.63
        0.052      0.096   0.025

     3     23         32      11      66
        16.91      28.46   20.63
        2.191      0.440   4.492

  Total    41         69      50     160
  Chi-square = 18.708, DF = 4, p value = 0.001
```

(a) H_0: Blood cholesterol count and type of smoker are independent.
H_1: Blood cholesterol count and type of smoker are not independent.
TS: $\chi^2 = 18.708$ with a p value $= 0.001$
Conclusion: Since p value $= 0.001 < \alpha = 0.05$, reject H_0. That is, there is sufficient sample evidence to conclude that blood cholesterol count and type of smoker are not independent.

(b) For $\alpha = 0.01$, same conclusion as in part (a) since the p value $= 0.001$.

(c) No. Such a study is not a randomized, controlled experiment, and association does not imply causation. Confounding variables, such as lifestyle habits, may exist. For example, people who exercise frequently may be more likely to have low cholesterol and be nonsmokers than people who exercise rarely. Hence, a confounding variable, such as exercise, may affect cholesterol level and be related to smoking habits as well.

12. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: A, B, C

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

              A        B        C   Total
      1      26       44       30     100
          25.27    42.96    31.77
          0.021    0.025    0.098

      2      14       30       25      69
          17.44    29.64    21.92
          0.677    0.004    0.433

      3      30       45       33     108
          27.29    46.40    34.31
          0.269    0.042    0.050

   Total     70      119       88     277

   Chi-square = 1.620, DF = 4, p value = 0.805
```

H_0: Row characteristic and column characteristic are independent.

H_1: Row characteristic and column characteristic are not independent.

TS: $\chi^2 = 1.62$ with a p value $= 0.805$

Conclusion: Since p value $= 0.805$ is rather large, do not reject H_0. That is, there is insufficient sample evidence to conclude that the row characteristic and the column characteristic are not independent.

14. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: 2A, 2B, 2C

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

             A       B       C   Total
     1      52      88      60     200
         50.54   85.92   63.54
         0.042   0.050   0.197

     2      28      60      50     138
         34.87   59.29   43.84
         1.355   0.009   0.865

     3      60      90      66     216
         54.58   92.79   68.62
         0.537   0.084   0.100

  Total    140     238     176     554

  Chi-square = 3.239, DF = 4, p value = 0.519
```

H_0: Row characteristic and column characteristic are independent.

H_1: Row characteristic and column characteristic are not independent.

TS: $\chi^2 = 3.239$ with a p value $= 0.519$

Conclusion: Since p value $= 0.519$ is rather large, do not reject H_0. That is, there is insufficient sample evidence to conclude that the row characteristic and the column characteristic are not independent.

13.4 TESTING FOR INDEPENDENCE IN CONTINGENCY TABLES WITH FIXED MARGINAL TOTALS

Problems

2. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Upper Income, Lower Income

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

          Upper    Lower   Total
         Income   Income
      1      22       19      41
          20.50    20.50
          0.110    0.110

      2      31       39      70
          35.00    35.00
          0.457    0.457

      3      47       42      89
          44.50    44.50
          0.140    0.140

  Total     100      100     200

  Chi-square = 1.415, DF = 2, p value = 0.493
```

H_0: School preferences and family income are independent.

H_1: School preferences and family income are not independent.

TS: $\chi^2 = 1.415$ with a p value $= 0.493$

Conclusion: Since p value $= 0.493$ is rather large, do not reject H_0. That is, there is insufficient sample evidence to conclude that school preferences and family income are not independent.

4. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Newspaper 1, Newspaper 2, Newspaper 3

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

     Newspaper 1  Newspaper 2  Newspaper 3  Total
  1           22           25           28     75
           25.00        25.00        25.00
           0.360        0.000        0.360
```

(*Continued*)

(*Continued*)

```
         2        41        37        44      122
               40.67     40.67     40.67
               0.003     0.331     0.273

         3        37        38        28      103
               34.33     34.33     34.33
               0.207     0.392     1.168

     Total       100       100       100      300

     Chi-square = 3.094, DF = 4, p value = 0.542
```

H_0: The newspaper an individual reads and the economic class to which that person belongs are independent.

H_1: The newspaper an individual reads and the economic class to which that person belongs are not independent.

TS: $\chi^2 = 3.094$ with a p value $= 0.542$

Conclusion: Since p value $= 0.542 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the newspaper an individual reads and the economic class to which that person belongs are not independent.

6. Below is a ***Minitab*** output with the relevant computations for this problem.

H_0: Final grade a student receives and whether the student watches the lectures on TV or is present in the lecture hall are independent

H_1: Final grade a student receives and whether the student watches the lectures on TV or is present in the lecture hall are not independent

TS: $\chi^2 = 2.914$ with a p value $= 0.405$

Conclusion: Since p value $= 0.405 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the final grades a student receives and whether the student watches on TV or is present in the lecture hall are not independent.

Same conclusion at the 1% level of significance.

Chi-Square Test: A, B, C. Less than C

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

            A       B       C  Less than C  Total
     1     22      38      35            5    100
        20.00   35.00   37.50         7.50
        0.200   0.257   0.167        0.833

     2     18      32      40           10    100
        20.00   35.00   37.50         7.50
        0.200   0.257   0.167        0.833

  Total     40      70      75           15    200

  Chi-square = 2.914, DF = 3, p value = 0.405
```

8. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Red, White, Blue, Green

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

           Red   White    Blue   Green  Total
     1     108     106     105     127    446
        112.63  112.63  108.12  112.63
         0.190   0.390   0.090   1.834

     2     142     144     135     123    544
        137.37  137.37  131.88  137.37
         0.156   0.320   0.074   1.504

  Total     250     250     240     250    990

  Chi-square = 4.558, DF = 3, p value = 0.207
```

H_0: Proportion of response for the different color postcards is the same.

H_1: Proportion of response for the different color postcards is not the same.

TS: $\chi^2 = 4.558$ with a p value $= 0.207$

Conclusion: Since p value $= 0.207 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the proportion of response for the different color postcards is not the same.

10. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Number Sampled, Number Leading to a Lawsuit

```
Expected counts arc printed below observed counts.
Chi-square contributions are printed below
expected counts.

          Number  Number Leading
         Sampled    to a Lawsuit  Total
      1      400              16    416
          399.23           16.77
           0.001           0.035

      2      300              19    319
          306.14           12.86
           0.123           2.934

      3      300               7    307
          294.63           12.37
           0.098           2.334

  Total     1000              42   1042

  Chi-square = 5.526, DF = 2, p value = 0.063
```

H_0: Proportion of surgical operations that lead to lawsuits is the same.

H_1: Proportion of surgical operations that lead to lawsuits is not the same.

TS: $\chi^2 = 5.526$ with a p value $= 0.063$

Conclusion: Since p value $= 0.063 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the proportion of surgical operations that lead to lawsuits is not the same.

Same conclusion at the 1% level of significance.

Review Problems

Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: New York, Chicago, Phoenix, Seattle

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below
expected counts.

         New York  Chicago  Phoenix  Seattle  Total
     1        234      141      108      142    625
           222.49   168.87   109.33   124.30
            0.595    4.600    0.016    2.519

     2        303      256      165      170    894
           318.25   241.55   156.39   177.80
            0.731    0.864    0.474    0.343

     3        102       88       41       45    276
            98.25    74.57    48.28    54.89
            0.143    2.417    1.098    1.783

  Total       639      485      314      357   1795

  Chi-square = 15.584, DF = 6, p value = 0.016
```

H_0: The audience's reaction and location are independent.

H_1: The audience's reaction and location are not independent.

TS: $\chi^2 = 15.584$ with a p value $= 0.016$

Conclusion: Since p value $= 0.016 < \alpha = 0.05$, reject H_0. That is, there is sufficient sample evidence to conclude that the audience's reaction and location are not independent.

For $\alpha = 0.01$, p value $= 0.016 > 0.01$, so do not reject the null hypothesis. That is, there is insufficient sample evidence to conclude that the audience's reaction and location are not independent.

4. $n = 50, N_1 = 18, N_2 = 22, N_3 = 10, P_1 = 0.40, P_2 = 0.42, P_3 = 0.18,$

$e_1 = (50)(0.4) = 20, e_2 = (50)(0.42) = 21, e_3 = (50)(0.18) = 9,$

$\alpha = 0.05, \chi^2_{2,0.05} = 5.99.$

H_0: $P_1 = 0.4$, $P_2 = 0.42$, $P_3 = 0.18$ versus H_1: At least one P_i is different

TS: $\chi^2 = 0.3587$

Conclusion: Since $0.3587 < 5.99$, do not reject H_0. That is, there is insufficient evidence to claim that the proportions are different from those specified in the null hypothesis at the 5% level of significance.

6. $n = 1000, N_1 = 167, N_2 = 165, N_3 = 167, N_4 = 166, N_5 = 167, N_6 = 168,$

$P_1 = 1/7, e_i = (1000)\left(1/6\right) = 166.6667, \alpha = 0.05,$

$\chi^2_{5,0.05} = 11.07.$

H_0: $P_i = 1/6$, for $i = 1, 2, \ldots, 7$ against H_1: $P_i \neq 1/6$ for at least one i

TS: $\chi^2 = 0.032$

Conclusion: Since $0.032 < 11.07$, do not reject H_0. That is, there is insufficient evidence to claim that the die is not fair at the 5% level. Also, since 0.032 is less than $\chi^2_{5,0.995} = 0.412$, it follows that a value at least as small as 0.032 would occur less than 0.5% of the time. Thus, the data is indicative of manipulation.

8. $n = 664, N_1 = 109, N_2 = 74, N_3 = 97, N_4 = 94, N_5 = 83,$

$N_6 = 107, N_7 = 100, P_i = 1/7, e_i = (664)(1/7) = 94.8571, \alpha = 0.05,$
$\chi^2_{6,\,0.05} = 12.59.$

H_0: $P_i = 1/7$, for $i = 1, 2, \ldots, 7$ against H_1: $P_i \neq 1/7$ for at least one i

TS: $\chi^2 = 10.0663$

Conclusion: Since $10.0663 < 12.59$, do not reject H_0. That is, there is insufficient evidence to claim that a murder was not equally likely to occur on any of the 7 days of the week at the 5% level of significance.

10. $n = 556, N_1 = 315, N_2 = 101, N_3 = 108, N_4 = 32, e_1 = 313,$

$e_2 = 104$, $e_3 = 104$, $e_4 = 35$.

Computed chi-square statistic is $\chi^2 = 0.5103$.

Since 0.5103 is greater than $\chi^2_{3,0.95} = 0.35$, it follows that a value as small as or smaller than 0.5103 would occur over 5% of the time. Thus, the data is not indicative of any manipulation.

12. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: 1, 2, 3

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

           1      2      3  Total
    1     10     28     22     60
        9.85  30.23  19.92
       0.002  0.164  0.216

    2     19     63     38    120
       19.69  60.46  39.85
       0.025  0.107  0.086

    3     14     41     27     82
       13.46  41.31  27.23
       0.022  0.002  0.002

 Total    43    132     87    262

 Chi-square = 0.626, DF = 4, p value = 0.960
```

H_0: The size of the family and the educational level of the father are independent.

H_1: The size of the family and the educational level of the father are not independent.

TS: $\chi^2 = 0.626$ with a p value $= 0.96$

Conclusion: Since p value $= 0.96 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that the size of the family and the educational level of the father are not independent.

14. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Less than 2500, More than 2500

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below expected
counts.

         Less than  More than  Total
              2500       2500
      1         24        100    124
             18.15     105.85
             1.888      0.324

      2         36        250    286
             41.85     244.15
             0.819      0.140

  Total         60        350    410

  Chi-square = 3.171, DF = 1, p value = 0.075
```

H_0: The birth weight of the baby and the age of the mother are independent.

H_1: The birth weight of the baby and the age of the mother are not independent.

TS: $\chi^2 = 3.171$ with a p value $= 0.075$

Conclusion: Since p value $= 0.075 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that birth weight of the baby and the age of the mother are not independent.

16. Below is a ***Minitab*** output with the relevant computations for this problem.

Chi-Square Test: Front, Middle, Back

```
Expected counts are printed below observed counts.
Chi-square contributions are printed below
expected counts.

         Front  Middle   Back  Total
      1     22      40     18     80
         16.62   40.52  22.86
         1.739   0.007  1.032

      2     10      38     26     74
         15.38   37.48  21.14
         1.880   0.007  1.116

  Total     32      78     44    154

  Chi-square = 5.781, DF = 2, p value = 0.056
```

H_0: The gender of the student and the seat location of the student are independent.

H_1: The gender of the student and the seat location of the student are not independent.

TS: $\chi^2 = 5.781$ with a p value $= 0.056$

Conclusion: Since p value $= 0.056 > \alpha = 0.05$, do not reject H_0. That is, there is insufficient sample evidence to conclude that gender of the student and the seat location of the student are not independent at the 5% level of significance.

18. Observed data values will vary.

Chapter 14 NONPARAMETRIC HYPOTHESES TESTS

14.2 SIGN TEST

Problems

2. Below is a partial ***Minitab*** output for this problem.

```
Sign Test for Median
Sign test of median = 52758 versus not = 52758
 N        Below       Equal       Above          P
250        105          0          145       0.0136
```

H_0: $\eta = \$52,758$ against H_1: $\eta \neq \$52,758$
TS: $N = 105$ (42% of 250)
p value $= 2$ Minimum $[P\{N \geq 105\},\ P\{N \leq 105\}] = 0.0136$
Conclusion: For a significance level of 5%, since p value $= 0.0136 < 0.05$, reject the null hypothesis and conclude that the median household income for the state of Connecticut is no longer \$52,758.

4. (a) Below is a partial ***Minitab*** output for this problem.

```
Sign Test for Median
Sign test of median = 0.00000 versus not = 0.00000
 N        Below        Equal        Above          P
10          6            0            4         0.7539
```

H_0: $\eta = 0$ against H_1: $\eta \neq 0$
TS: $N = 6$ (60% of 10)
p value $= 2$ Minimum $[P\{N \geq 6\}, P\{N \leq 6\}] = 0.7539$
Conclusion: For a significance level of 5%, since p value $= 0.7539 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median is not equal to zero or that the effects of the two creams are not significantly different.

(b) Below is a partial ***Minitab*** output for this problem.

H_0:$\eta = 0$ against H_1:$\eta \neq 0$

Sign Test for Median

Sign test of median = 0.00000 versus not = 0.00000

N	Below	Equal	Above	P
20	12	0	8	0.5034

TS: $N = 12$(60% of 20)

p value $= 2$ Minimum $[P\{N \geq 6\}, P\{N \leq 6\}] = 0.5034$

Conclusion: For a significance level of 5%, since p value $= 0.5034 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median is not equal to zero or that the effects of the two creams are not significantly different.

(c) Below is a partial ***Minitab*** output for this problem.

Sign Test for Median

Sign test of median = 0.00000 versus not = 0.00000

N	Below	Equal	Above	P
50	30	0	20	0.2026

$n = 50$

H_0: $\eta = 0$ against H_1: $\eta \neq 0$

TS: $N = 30$ (60% of 50)

p value $= 2$ Minimum $[P\{N \geq 6\},\ P\{N \leq 6\}] = 0.2026$

Conclusion: For a significance level of 5%, since p value $= 0.2026 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median is not equal to zero or that the effects of the two creams are not significantly different.

(d) Below is a partial ***Minitab*** output for this problem.

Sign Test for Median

Sign test of median = 0.00000 versus not = 0.00000

N	Below	Equal	Above	P
100	60	0	40	0.0574

$n = 100$

H_0: $\eta = 0$ against H_1: $\eta \neq 0$

TS: $N = 60$ (60% of 100)

p value $= 2$ Minimum $[P\{N \geq 6\},\ P\{N \leq 6\}] = 0.0574$

Conclusion: For a significance level of 5%, since p value $= 0.0574 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median is not equal to zero or that the effects of the two creams are not significantly different.

(e) Below is a partial ***Minitab*** output for this problem.

```
Sign Test for Median
Sign test of median = 0.00000 versus not = 0.00000
 N        Below       Equal       Above          P
500        300          0          200       0.0000
```

$n = 500$

H_0: $\eta = 0$ against H_1: $\eta \neq 0$

TS: $N = 300$ (60% of 500)

p value $= 2$ Minimum $[P\{N \geq 6\}, P\{N \leq 6\}] = 0.0000$

Conclusion: For a significance level of 5%, since p value $= 0.0000 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to claim that the median is not equal to zero or that the effects of the two creams are significantly different.

6. Below is a partial ***Minitab*** output for this problem.

```
Sign Test for Median: Selling Price
Sign test of median = 122000 versus > 122000
               N  Below Equal Above    P    Median
Selling Price 20   7      0     13   0.1316 129000
```

H_0: $\eta \leq \$122,000$ against H_1: $\eta > \$122,000$

TS: $N = 7$

p value $= P\{N \leq 7\} = 0.1316$

Conclusion: For a significance level of 5%, since p value $= 0.1316 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median price has increased.

8. (a), (b), and (c)

Below is a partial ***Minitab*** output with some of the computations for this problem.

```
Sign Test for Median
Sign test of median = 0.00000 versus < 0.00000
N      Below     Equal     Above       P
24      16         0         8       0.0758
```

H_0: $p \geq 1/2$ against H_1: $p > 1/2$
TS: $N = 16$
p value $= P\{N \geq 16\} = 0.0758$
Conclusion: For a significance level of 5%, since p value $= 0.0758 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median blood cholesterol level has been reduced by the consumption of fish oil.

14.3 SIGNED-RANK TEST

Problems

2. (a) $E[\text{TS}] = (15)(16)/4 = 60$; $\text{Var}(\text{TS}) = (15)(16)(31)/24 = 310$
(b) $E[\text{TS}] = (13)(14)/4 = 45.5$; $\text{Var}(\text{TS}) = (13)(14)(27)/24 = 204.75$
(c) $E[\text{TS}] = (12)(13)/4 = 39$; $\text{Var}(\text{TS}) = (12)(13)(25)/24 = 162.5$

4. (a) The ordered data in increasing absolute values:
$-5, 7, -8, 14, -17, -18, 22, 33, 39, 40, -41, 55, -88, 99, 102.$
$\text{TS} = 1 + 3 + 5 + 6 + 11 + 13 = 39$
p value $= 2P\{\text{TS} \leq 39\} = 2P\{\text{TS} \leq 39.5\} = 2P\{Z \leq -1.16\} = 0.246$
p value using ***Minitab*** is 0.293
(b) The ordered data in increasing absolute values:
$0.01, -0.4, 1, 1.1, 2, -2.2, -3, -4, -6.6, -13, 44, 44, 50.$
$\text{TS} = 2 + 6 + 7 + 8 + 9 + 10 = 42$
p value $= 2P\{\text{TS} \leq 42\} = 2P\{\text{TS} \leq 42.5\} = 2P\{Z \leq -0.21\} = 0.8336$
p value using ***Minitab*** is 0.834
(c) The ordered data in increasing absolute values:
$-3, -7, 8, 12, 15, 19, -22, 31, 48, -55, 89, 92.$
$\text{TS} = 1 + 2 + 7 + 10 = 20$
p value $= 2P\{\text{TS} \leq 20\} = 2P\{\text{TS} \leq 20.5\} = 2P\{Z \leq -1.45\} = 0.147$
p value using ***Minitab*** is 0.147

6. $n = 100, \mu = 2525, \sigma = 290.8393.$

H_0: The sealant is not effective at reducing cavities.
H_1: The sealant is effective at reducing cavities.

TS: $D = 1830$
p value $= P\{\text{TS} \leq 1830\} = P\{\text{TS} \leq 1830.5\} = P\{Z \leq -2.39\} = 0.0084$
Conclusion: For a significance level of 5%, since p value $= 0.0084 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to claim that the sealants make a difference.
At the 1% level of significance, you will not reject the null hypothesis.

8. The ordered data in increasing absolute values of (before – after):

$$-0.04, -0.12, 0.14, 0.19, 0.21, 0.22, 0.25, 0.45, -0.58, 0.58, 0.83.$$

(a) $\text{TS} = 1 + 9.5 + 2 = 12.5$
(b) $n = 11, \mu = 33, \sigma = 11.2472$

H_0: The differences are symmetric about 0.
H_1: The differences are not symmetric about 0.
p value $= 2P\{\text{TS} \leq 12.5\} = 2P\{Z \leq -1.82\} = 0.0688$

10. If you let $D_i = X_i - v$, then you are transforming the data to be symmetric about 0. All you have to do is to use the usual signed-rank test on these transformed differences.

14.4 RANK-SUM TEST FOR COMPARING TWO POPULATIONS

Problems

2. Since $\sum_{i=1}^{n+m} i = k(k+1)/2$, then if $k = (n+m)$, $\sum_{i=1}^{n+m} i = (n+m)(n+m+1)/2$, where n is the size of sample 1 and m is the size of sample 2. Below is a listing of the sample values and their corresponding ranks.

Sample 1	142	155	237	244	202	111	326	334	350	247
Rank	3	4	9	10	7	1	15	16	18	11
Sample 2	212	277	175	138	341	255	303	188		
Rank	8	13	5	2	17	12	14	6		

The sum of the ranks for sample 1: $3 + 4 + \cdots + 18 + 11 = 94$.
The sum of the ranks for sample 2: $8 + 13 + \cdots + 14 + 6 = 77$.
(Sum of ranks for sample 1 + Sum of ranks for sample 2)
$= 94 + 77 = 171$.

Now, if we use the combined sample, then $(n + m) = (10 + 8) = 18$.
Substitute in $(n + m)(n + m + 1)/2 = (18)(19)/2 = 171$.
Thus, the sum of the ranks of the individual samples is the same as the sum of the first $(n + m)$ integers.

4. The samples and their ranks are given in the table below.

Vitamin C	6	12	14	2	7	7	1	8
Rank	10	15	16	7	11.5	11.5	6	13
Placebo	9	−3	0	−1	5	3	−4	−1
Rank	14	2	5	3.5	9	8	1	3.5

Let n be the sample size for the vitamin C group and m for the placebo group. We will use the vitamin C group to compute the TS.
$n = m = 8$. Sum of ranks for vitamin C group $= 90$.

$E[\text{TS}] = 68;\ \text{Var(TS)} = 90.6667$

H_0: Vitamin C and the placebo are equally effective in reducing cholesterol.
H_1: Vitamin C and the placebo are not equally effective in reducing cholesterol.
TS: $W = 90$
p value $= 2P\{W \geq 90\} = 2P\{W \geq 89.5\} \approx 2P\{Z \geq 2.26\} = 0.0238$
Conclusion: Since the p value $= 0.0238 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to conclude that vitamin C and the placebo are not equally effective in reducing cholesterol.

6. The samples and their ranks are given in the table below.

Untreated	18	12.4	13.5	14.6	24	21	23	17.5	
Rank	5	1	2	3	12	9	11	4	
Vitamin B1	34	27	21.2	29	20.5	19.6	28	33	19
Rank	17	13	10	15	8	7	14	16	6

Let n be the sample size for the untreated group and m for the vitamin B1 group. We will use the untreated group to compute the TS.
$n = 8,\ m = 9$. Sum of ranks for the untreated group $= 47$.

$E[\text{TS}] = 72\ \text{Var(TS)} = 108$

H_0: Vitamin B1 is not effective in stimulating growth in mushrooms.
H_1: Vitamin B1 is effective in stimulating growth in mushrooms.

TS: $W = 47$
p value $= 2P\{W \leq 47\} = 2P\{W \leq 47.5\} \approx 2P\{Z \leq -2.36\} = 0.0182$
Conclusion: Since the p value $= 0.0182 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to conclude that vitamin B1 is effective in stimulating growth in mushrooms.

8. p value $= 0.0237$ to four decimal places for a two-sided test.

10. Let n be the sample size for the smokers and m for the nonsmokers group. We will use the smokers group to compute the TS.
$n = 11,\ m = 14$. Sum of ranks for the smokers $= 184.5$.

$E[\text{TS}] = 143$ Var(TS) $= 333.6667$

H_0: Smoking does not have an effect on the systolic blood pressure of a male.
H_1: Smoking has an effect on the systolic blood pressure of a male.
TS: $W = 184.5$
p value $= 2P\{W \geq 184.5\} \approx 2P\{Z \geq 2.27\} = 0.0232$
Conclusion: Since the p value $= 0.0232 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to conclude that smoking affects the systolic blood pressure of a male. At the 1% significance level, you will fail to reject the null hypothesis.

14.5 RUNS TEST FOR RANDOMNESS

Problems

2. (a) Number of runs of 0s = 4; number of runs of 1s = 5
(b) Number of runs of 0s = 4; number of runs of 1s = 4
(c) Number of runs of 0s = 5; number of runs of 1s = 4

4. (a) Total number of runs $R = 14$.
(b) Let n be the number of defective items and m the number of nondefective items.
$n = 12,\ m = 48,\ R = 14,\ E[\text{TS}] = 20.2,\ \text{Var(TS)} = 5.9227.$
H_0: Number of defectives in the production run is random.
H_1: Number of defectives in the production run is not random.
TS: $R = 14$
p value $= 2P\{R \leq 14\} = 2P\{R \leq 14.5\} \approx 2P\{Z \leq -2.34\} = 0.0192$
Conclusion: Since the p value $= 0.0192 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to conclude that the number of defectives in the production run is not random.

6. Let n be the number of days of increase and m the number of days of decrease.
$n = 32,\ m = 18,\ R = 22,\ E[\text{TS}] = 24.04,\ \text{Var(TS)} = 10.3633.$
H_0: The increase and decrease of the Dow Jones industrial average constitutes a random sample.
H_1: The increase and decrease of the Dow Jones industrial average does not constitute a random sample.
TS: $R = 22$
p value $= 2P\{R \le 22\} = 2P\{R \le 22.5\} \approx 2P\{Z \le -0.48\} = 0.6312$

8. Let n be the number of values below the sample median, and let m be the number above the sample median.

The sample median $S_m = 1177.5.$ $n = m = 5,\ R = 2, E[\text{TS}] = 6,$ $\text{Var(TS)} = 2.2222.$
H_0: The observed Dow Jones industrial average constitutes a random sample.
H_1: The observed Dow Jones industrial average does not constitute a random sample.
TS: $R = 2$
p value $= 2P\{R \le 2\} = 2P\{R \le 2.5\} \approx 2P\{Z < -2.35\} = 0.0188$
Conclusion: Since the p value $= 0.0188 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to conclude that the observed Dow Jones industrial average does not constitute a random sample.

Note: Small sample size may make the approximation invalid.

14.7 PERMUTATION TESTS

Problems

2. The value of the test statistic is $T = \Sigma_j jX_j = 840$. Under the null hypothesis that all orderings are equally likely $E_{H_0}[T] = 840$, showing that the p value is approximately 0.5, which is not a validation of the player's reputation.

Review Problems

2. The samples and their ranks are given in the table below.

Generic	14.2	14.7	13.9	15.3	14.8	13.6	14.6	14.9	14.2
Rank	4.5	10.5	3	18	12	1	9	14	4.5
Name	14.3	14.9	14.4	13.8	15.0	15.1	14.4	14.7	14.9
Rank	6	14	7.5	2	16	17	7.5	10.5	14

(a) Let n be the sample size for the generic group and m for the name brand group. We will use the generic group to compute the TS.
$n = 9,\ m = 9$. Sum of ranks for the generic group $= 76.5$.

$E[\text{TS}] = 90$ $\text{Var}(\text{TS}) = 128.25$

H_0: Time to dissolve for the generic brand is less than or equal to the time to dissolve for the name brand tablets.
H_1: Time to dissolve for the generic brand is greater than the time to dissolve for the name brand tablets.
TS: $W = 76.5$
p value $= P\{W \leq 76.5\} \approx P\{Z \leq -1.1921\} = 0.4238$
Conclusion: Since the p value $= 0.4238 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to conclude that the time to dissolve for the generic brand is greater than the time to dissolve for the name brand tablets at the 5% significance level.

(b) Same conclusion at the 10% level of significance.

4. (a) To use the sign test, the difference between mpg without additive and mpg with additive was used. A partial ***Minitab*** output for this test is shown below.

```
Sign Test for Median
Sign test of median = 0.00000 versus not = 0.00000
N       Below       Equal       Above          P
8         2           0           6       0.2891
```

$H_0{:}\eta = 0$ against $H_1{:}\eta \neq 0$

TS: $N = 2$

p value $= 2P\{N \leq 2\} = 0.2891$ (N is binomial with parameters $n = 8$ and $p = 0.5$)
Conclusion: For a significance level of 5%, since p value $= 0.2891 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median is not equal to zero or that the effect of the additive is significant.

(b) To use the signed-rank test, the difference that was used in part (a) was also used here.

$n = 8,\ E[\text{TS}] = 18,\ \text{Var}(\text{TS}) = 51.$
H_0: The distribution of the differences is symmetric about 0.
H_1: The distribution of the differences is not symmetric about 0.
TS: $W = 4$
p value $= 2P\{W \leq 4\} = 2P\{W \leq 4.5\} \approx 2P\{Z \leq -1.89\} = 0.0588$

Conclusion: For a significance level of 5%, since p value $= 0.0588 > 0.05$, do not reject the null hypothesis. That is, there is insufficient sample evidence to claim that the median is not equal to zero or that the effect of the additive is significant.
The conclusions are the same for these two nonparametric tests at the 5% level of significance although the small sample sizes might make the approximate p value inaccurate.

6. Samples and test statistics will vary. Even the conclusions may vary.

8. Let n be the number of blood samples that does not show the virus and m the number of samples that shows the virus.
$n = 14,\ m = 11,\ R = 8,\ E[\text{TS}] = 13.32,\ \text{Var(TS)} = 5.8109.$

H_0: Samples were tested in a random order.
H_1: Samples were tested in a nonrandom order.
TS: $R = 8$
p value $= 2P\{R \leq 8\} = 2P\{R \leq 8.5\} \approx 2P\{Z \leq -2\} = 0.0456$
Conclusion: Since the p value $= 0.0456 < 0.05$, reject the null hypothesis. That is, there is sufficient sample evidence to conclude that the samples were not tested in a random order.

10. Here, we can assume that the two populations (Student 1 and Student 2) are classified according to 7 characteristics ($i = 4, 5, 6, \ldots, 10$). Thus, we have a 7 by 2 contingency table.

H_0: The number of letters used in a word is independent of the student.
H_1: The number of letters used in a word is not independent of the student.
TS: $\chi^2 = 6.796$
p value $= 0.3401$
Conclusion: For a level of significance of 0.05, $0.3401 > 0.05$. Thus do not reject the null hypothesis. That is, there is insufficient evidence to conclude that the number of letters used in a word is not independent of the student.

Chapter 15 QUALITY CONTROL

15.2 THE $\overline{X}$ CONTROL CHART FOR DETECTING A SHIFT IN THE MEAN

Problems

2. $n = 4, \mu = 35, \sigma = 4$.

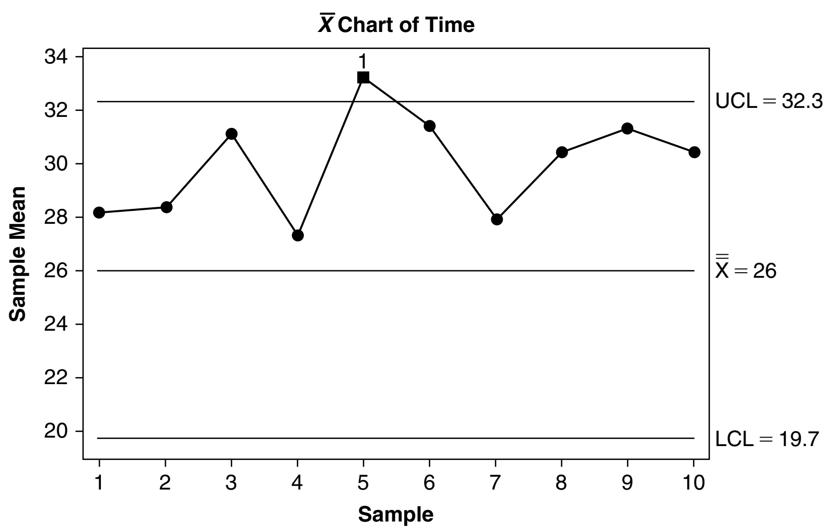
LCL $= \mu - 3\sigma/\sqrt{n} = 29$ and UCL $= \mu + 3\sigma/\sqrt{n} = 41$. The process is out of control since the average for the 17th subgroup of $41.2 > 41$ (UCL).

4. $n = 4, \mu = 26, \sigma = 4.2$.

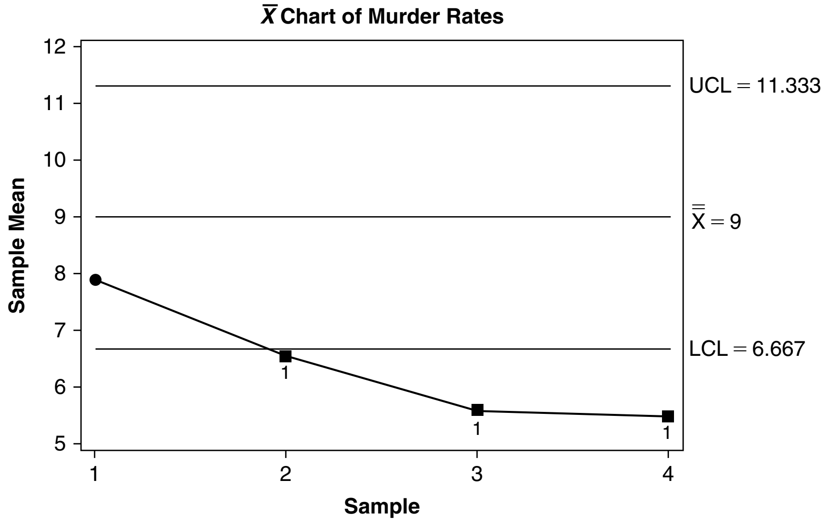
(a) LCL $= \mu - 3\sigma/\sqrt{n} = 19.7$ and UCL $= \mu + 3\sigma/\sqrt{n} = 32.3$

(b) The process is out of control since the average for the 5th subgroup of $33.2 > 32.3$ (UCL).

See the following $\overline{X}$ chart.



6. From the $\overline{X}$ chart for the murder rates, observe that the LCL $= 6.667$ and the UCL $= 11.333$. Also, observe that the murder rates for subgroups of size 2 display a downward trend below the LCL. This would indicate that the murder rate has changed from its historical average value of 9.0.



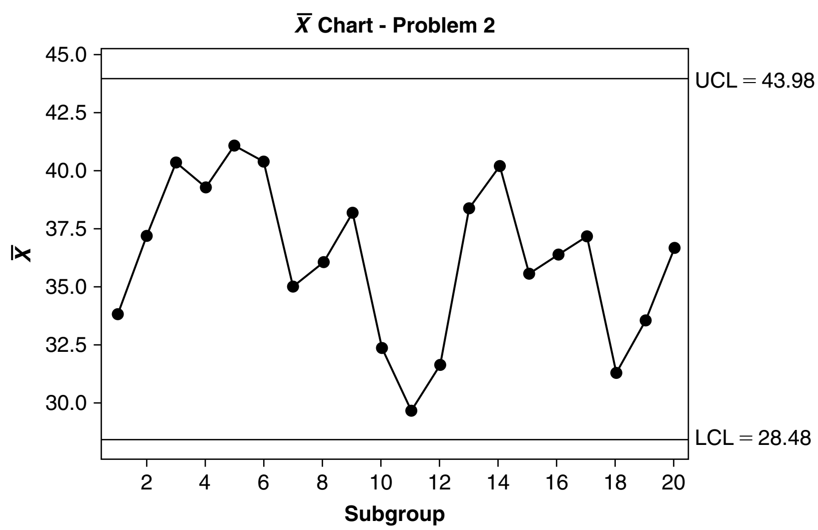
15.2.2 *S* Control Charts

Problems

2. (a) Estimate for $\mu = 36.23$; estimate for $\sigma = 5.43/0.9399851 = 5.7767$.

Also $3(5.7767)/\sqrt{5} = 7.7502$. Thus, LCL $= 36.23 - 7.7502 = 28.4798$ and UCL $= 36.23 + 7.7502 = 43.9802$. The process is in control.

(b) The following $\overline{X}$ control chart shows the process is in control.



(c) $\overline{\overline{X}} = 36.23, \frac{\overline{S}}{c(5)} = 5.43/0.9399851 = 5.7767$. If we let X be the number of items produced, then $P\{25 \leq X \leq 45\} = P\{(25 - 36.23)/5.7767 \leq X \leq (45 - 36.23)/5.7767)\} = P\{-1.9440 \leq Z \leq 1.5182\} = 0.9096$. Thus, we can say that 90.96% of the items produced will have values between 25 and 45.

4. (a) $\mu = 26$ and $\sigma = 8.3$. Thus, $\text{LCL} = 26 - (3)(8.3)/\sqrt{5} = 14.8644$ and $\text{UCL} = 26 + (3)(8.3)/\sqrt{5} = 37.1356$.

(b) $c(5) = 0.9399851$ and $\sigma = 8.3$. Thus, the

$\text{LCL} = c(5) \times 8.3 - 3(8.3) \times \sqrt{[1 - c^2(5)]} = -0.6944$ and

$\text{UCL} = c(5) \times 8.3 + 3(8.3) \times \sqrt{[1 - c^2(5)]} = 16.2981$

6. $\overline{\overline{X}} = 29.16, \overline{S} = 3.710, c(5) = 0.9399851$.

(a) $\overline{X}$: $\text{LCL} = 29.16 - (3)(3.71)/\sqrt{(5 \times 0.9399851)} = 24.0261$

$\overline{X}$: $\text{UCL} = 29.16 + (3)(3.71)/\sqrt{(5 \times 0.9399851)} = 34.2939$

S-chart: $\text{LCL} = 3.71\left[1 - (3)\sqrt{\frac{1}{c^2(5)} - 1}\right] = -0.3302$

S-chart: $\text{UCL} = 3.71\left[1 + (3)\sqrt{\frac{1}{c^2(5)} - 1}\right] = 7.7502$

(b) The process appears to be in control based on the limits in part (a).

(c) We can estimate μ by 29.16 and σ by $3.71/c(5) = 3.71/0.9399851 = 3.9469$. We can then use the estimated control limits

$\text{LCL} = \overline{S}\left[1 - 3\sqrt{\frac{1}{c^2(n)} - 1}\right]$ and $\text{UCL} = \overline{S}\left[1 + 3\sqrt{\frac{1}{c^2(n)} - 1}\right]$

8. $\overline{S} = 5.44/25 = 0.2176, c(6) = 0.9515332$

S-chart: $\text{LCL} = 0.2176\left[1 - (3)\sqrt{\frac{1}{c^2(6)} - 1}\right] = 0.0066$

S-chart: $\text{UCL} = 0.2176\left[1 + (3)\sqrt{\frac{1}{c^2(6)} - 1}\right] = 0.4286$

15.3 CONTROL CHARTS FOR FRACTION DEFECTIVE

Problems

2. $n = 500, p = 0.04$.

$\text{LCL} = np - 3\sqrt{(np(1 - p))} = 6.8547$

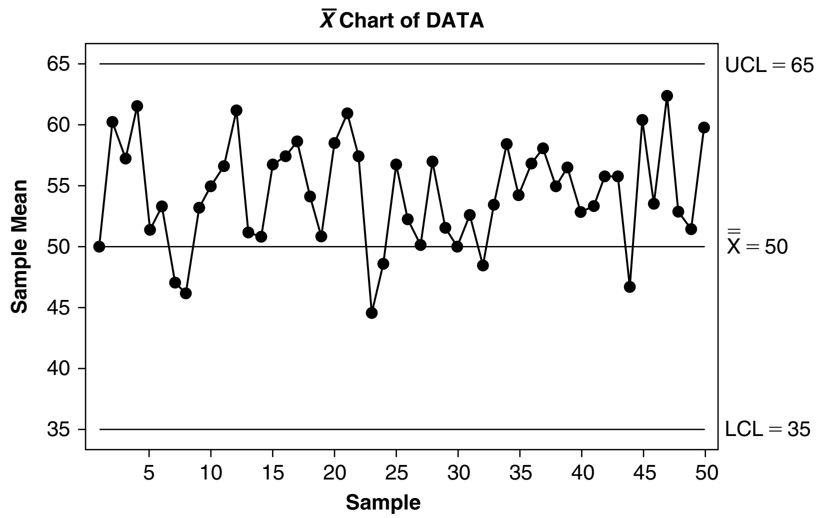
$\text{UCL} = np + 3\sqrt{(np(1 - p))} = 33.1453$

15.4 EXPONENTIALLY WEIGHTED MOVING-AVERAGE CONTROL CHARTS

Problems

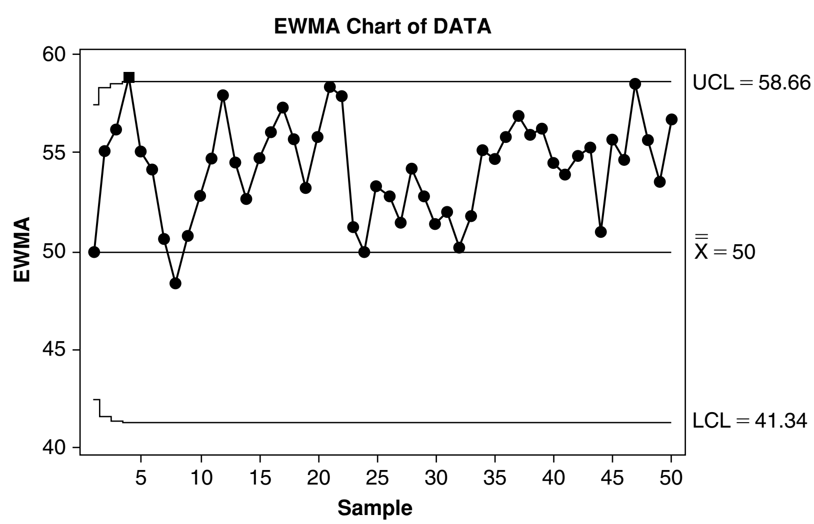
2. (a) Using the original parameters of $\mu = 50$, $\sigma = 5$, and $n = 1$ $\left(\text{since } \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = 5\right)$, following is the $\overline{X}$ chart.

Process is not out of control.



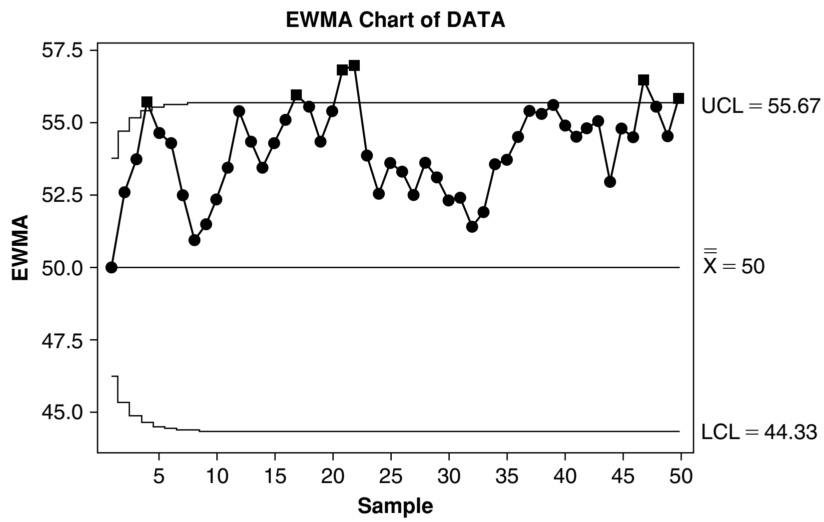
(b) $\mu = 50, \sigma = 5, \beta = 0.5, n = 1.$

Process is out of control.



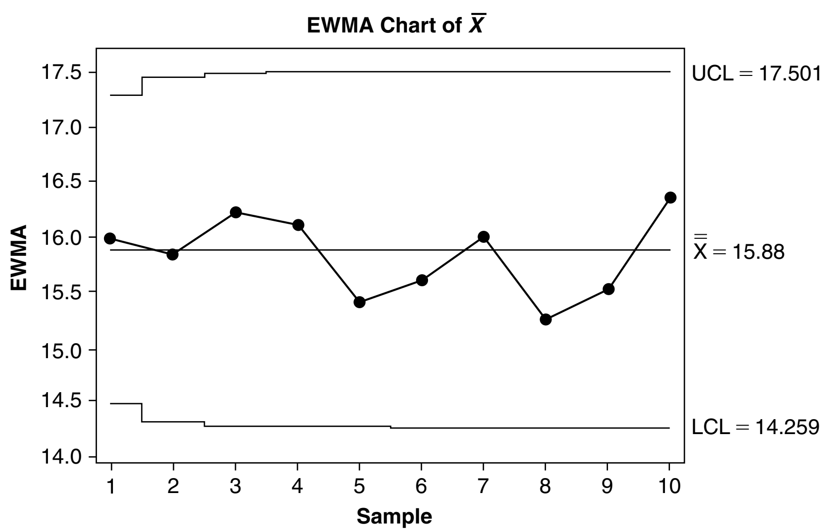
(c) $\mu = 50, \sigma = 5, \beta = 0.25, n = 1.$

Process is out of control

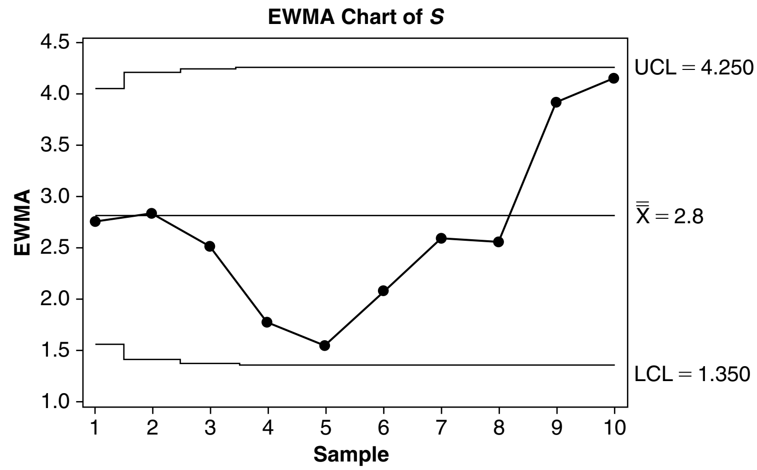


4. EWMA control chart with $\beta = 0.5$. Process is in control

Process is not out of control.



Process is not out of control but getting close to being out of control. Observe the last point in the EWMA chart for S.



15.5 CUMULATIVE-SUM CONTROL CHARTS

Problems

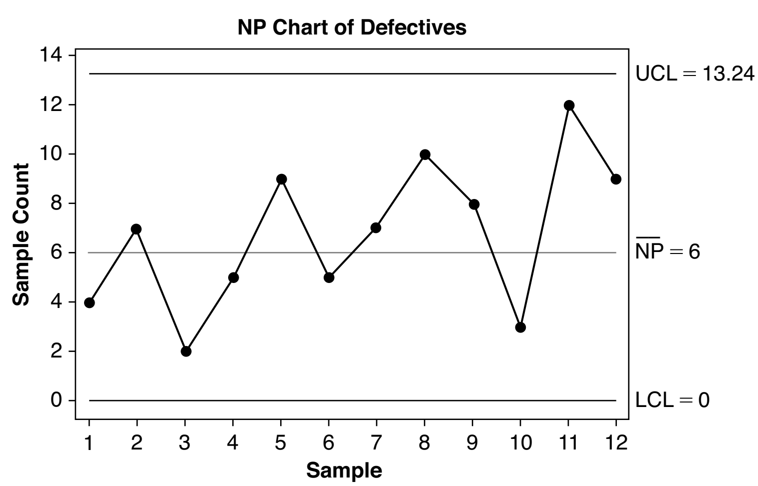
2. $\overline{\overline{X}} = 36.22, \overline{S} = 5.43; d = 1, B = 2.49$, control limit $= \frac{B\sigma}{\sqrt{n}} = \frac{2.49\times 5.43}{\sqrt{5}} = 6.0466;$

$Y_j = \overline{X}_j - \mu - \frac{d\sigma}{\sqrt{n}} = \overline{X}_j - 38.6484.$

$Y_1 = -4.8484, Y_2 = -1.4484, Y_3 = 1.7516, Y_4 = 0.6516, Y_5 = 2.4516, Y_6 = 1.7516, \ldots$

$S_1 = 0, S_2 = 0, S_3 = 1.7516, S_4 = 2.4002, S_5 = 4.8548, S_6 = 6.6064, \ldots$

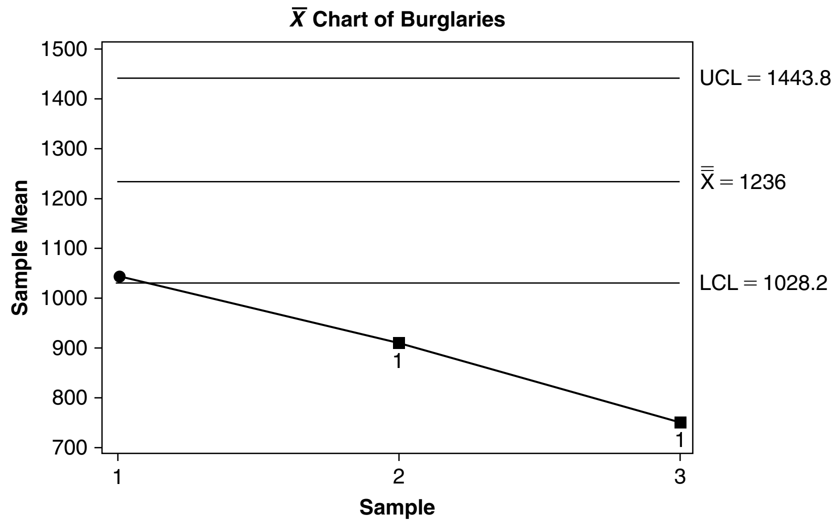
Since the control limit is $\frac{B\sigma}{\sqrt{n}} = 6.0466$, the cumulative-sum chart would declare that the mean has increased after observing the first six subgroup averages.



Review Problems

2. Subgroup size $n = 3, \mu = 1236, \sigma = 120$.

LCL $= \mu - 3\sigma/\sqrt{n} = 1028.1539$ and UCL $= \mu + 3\sigma/\sqrt{n} = 1443.8461$. From the $\overline{X}$ chart, one can conclude that the burglary rate has changed from its historical value. As a matter of fact, it is decreasing.



4. $n = 200, p = 0.03$.

LCL $= np - 3\sqrt{(np(1-p))} = -1.2374$ (0 for practical purposes)

UCL $= np + 3\sqrt{(np(1-p))} = 13.2374$

Since none of the numbers: 4, 7, 2, 5, 9, 5, 7, 10, 8, 3, 12, 9 fall outside the limits, then the process is in control. The following chart shows that the process is in control.