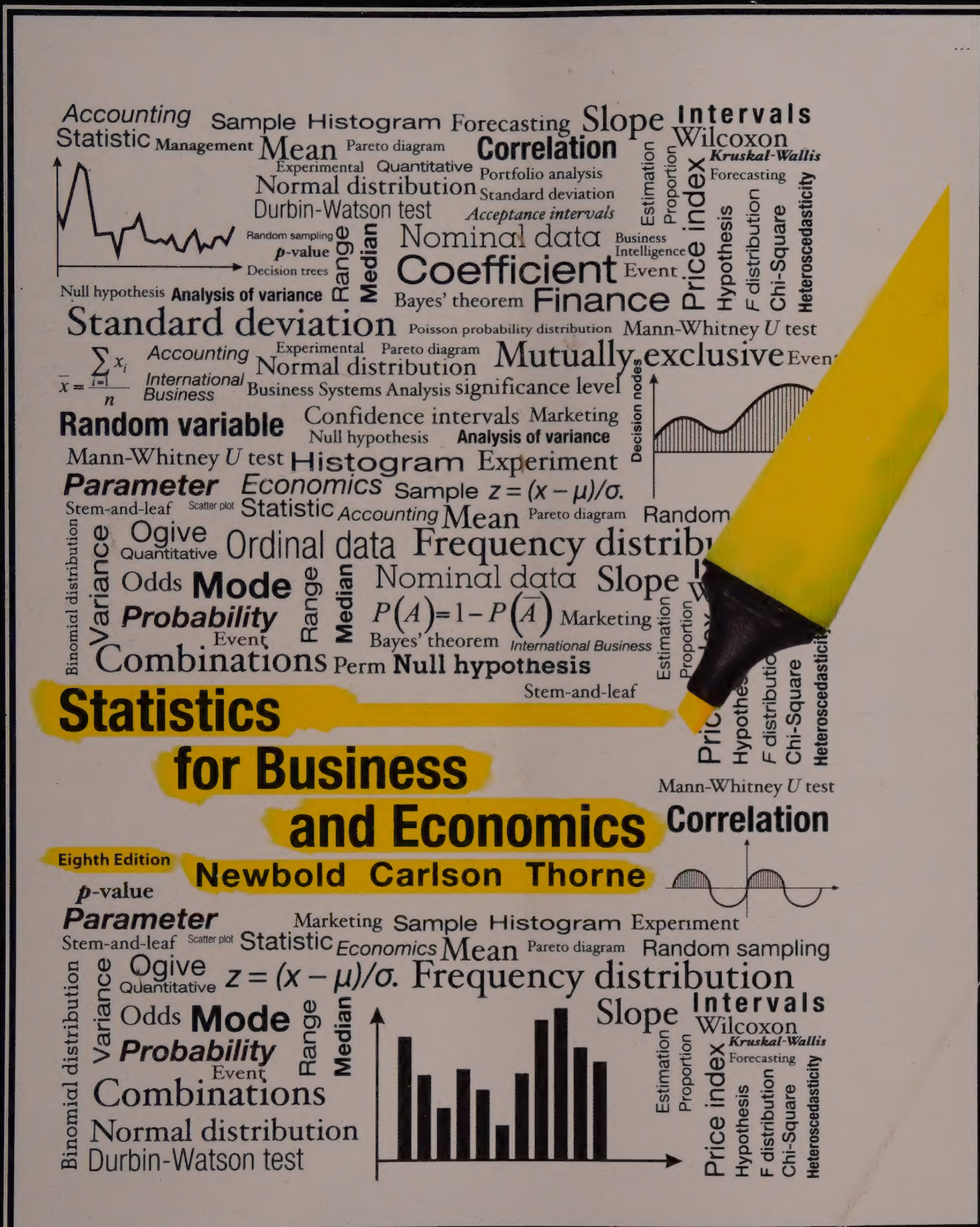
Student Solutions Manual



Paul Newbold • William L. Carlson
Betty M. Thorne

PEARSON

Student Solutions Manual

Statistics for Business and Economics

Eighth Edition

Paul Newbold
University of Nottingham
William L. Carlson
St. Olaf College
Betty M. Thorne
Stetson University

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editor in Chief: Donna Battista
Senior Executive Editor: Chuck Synovec
Editorial Project Manager: Mary Kate Murray
Production Project Manager: Jacqueline A. Martin
Operations Specialist: Cathleen Petersen

PEARSON

ISBN-13: 978-0-13-274568-0
ISBN-10: 0-13-274568-2

STUDENT SOLUTIONS MANUAL

Table of Contents

Table of Contents

Chapter 1:

Describing Data: Graphical

1.2

a. Categorical, nominal (The response is categorical because the responses can be grouped into classes or categories, in this case yes/no. The measurement levels are nominal because the responses are words that describe the categories.)
b. Categorical, ordinal (The response is categorical because the responses can be grouped into classes or categories. The measurement levels are ordinal because these are rankings of the data.)
c. Numerical, discrete (The response is numerical because the responses cannot be grouped into classes or categories. Since the response is an actual cost, it is discrete because the value comes from a counting process.)

1.4

a. Categorical – Qualitative – ordinal
b. Numerical – Quantitative – discrete
c. Categorical – Qualitative – nominal
d. Categorical – Qualitative – nominal

1.6

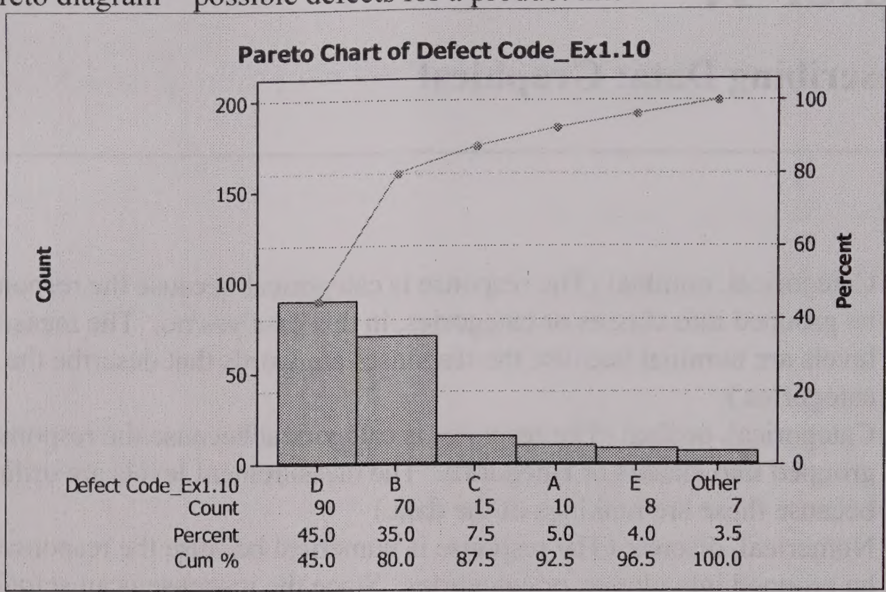
a. Categorical – Qualitative – nominal
b. Numerical – Quantitative - discrete
c. Categorical – Qualitative – nominal: yes/no response
d. Categorical – Qualitative – ordinal

1.8

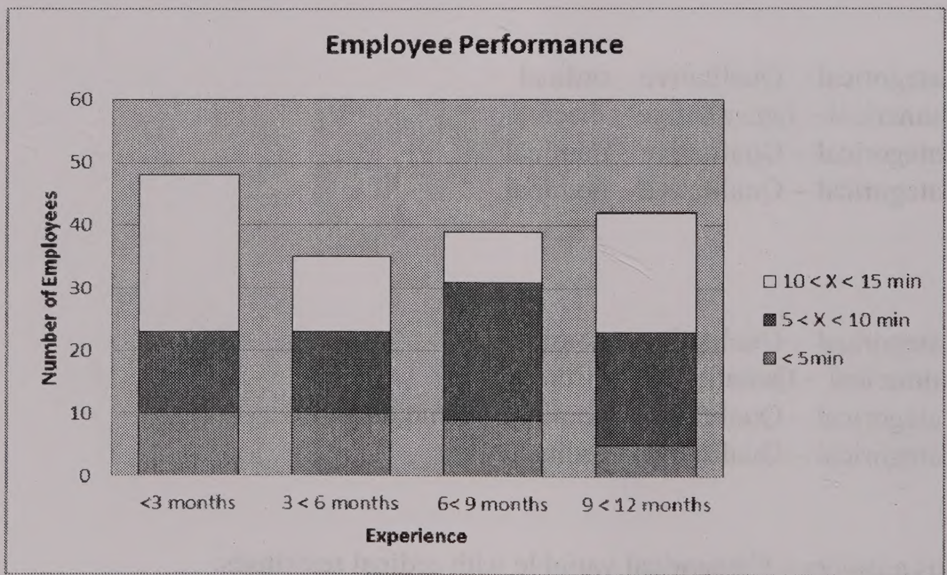
a. Various answers – Categorical variable with ordinal responses: PIR_grp (1-5)
b. Various answers – Categorical variable with nominal responses: female (male/female – 0/1)
c. Various answers – Numerical variable with continuous responses: daily_cost
d. Various answers – Numerical variable with discrete responses: PIR_p

1.10

Pareto diagram – possible defects for a product line

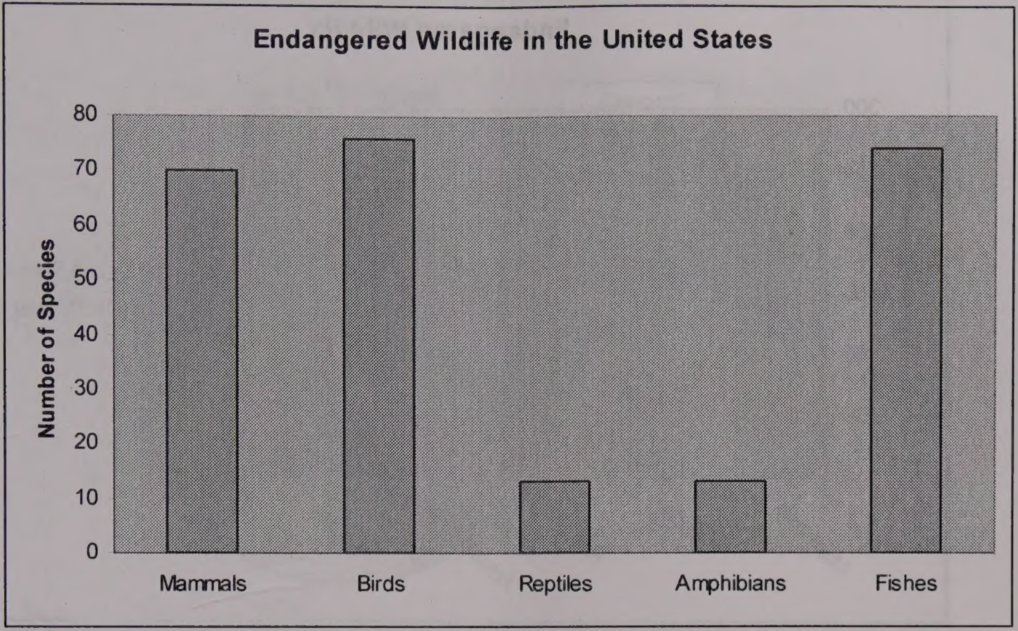


1.12

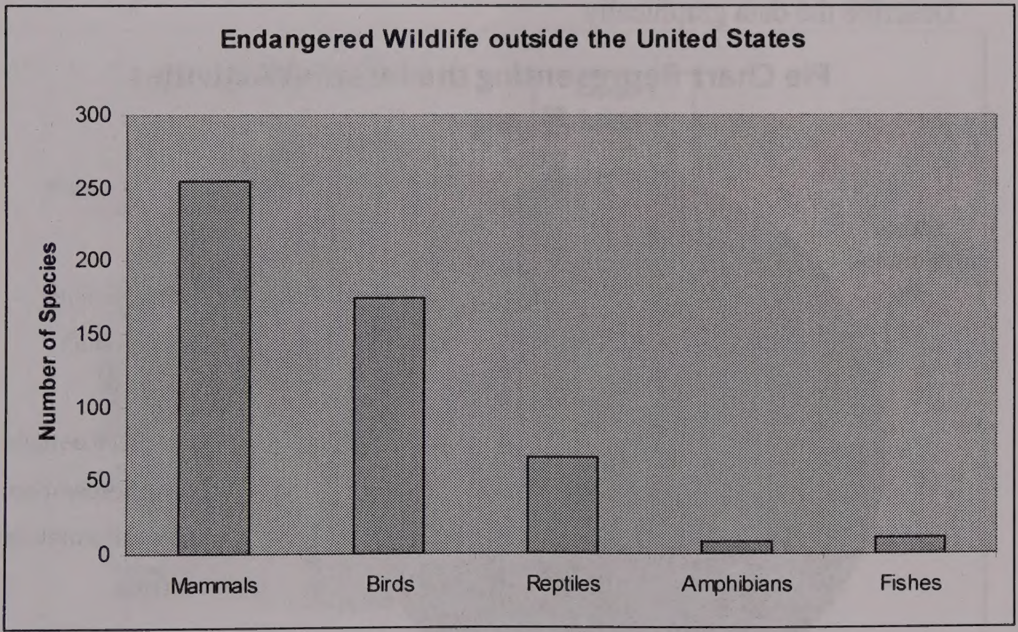


1.14

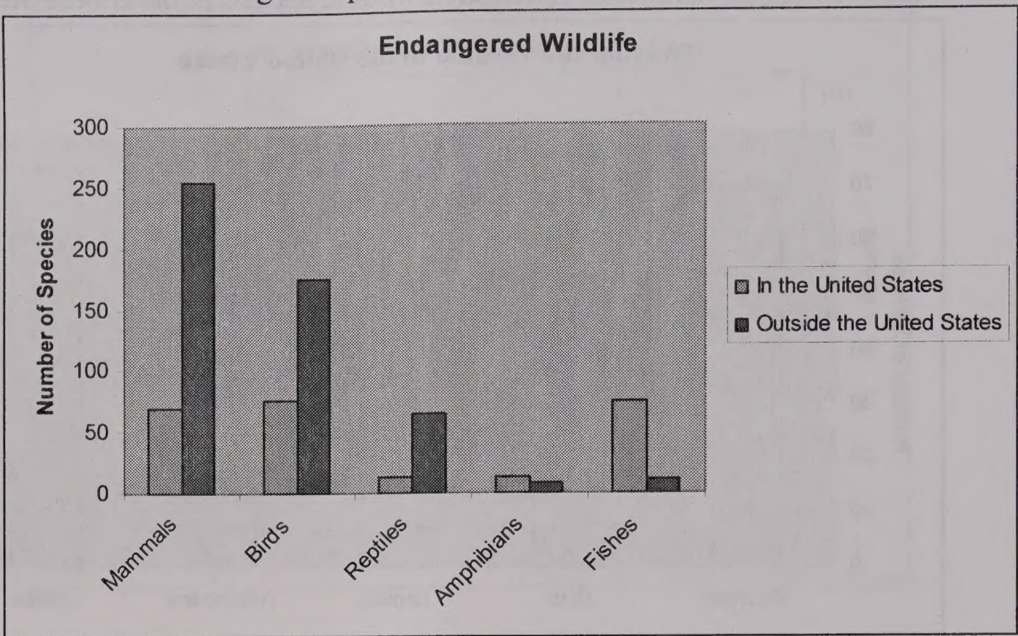
a. Bar chart of the number of endangered wildlife species in the United States



b. Bar chart of the number of endangered wildlife species outside the United States

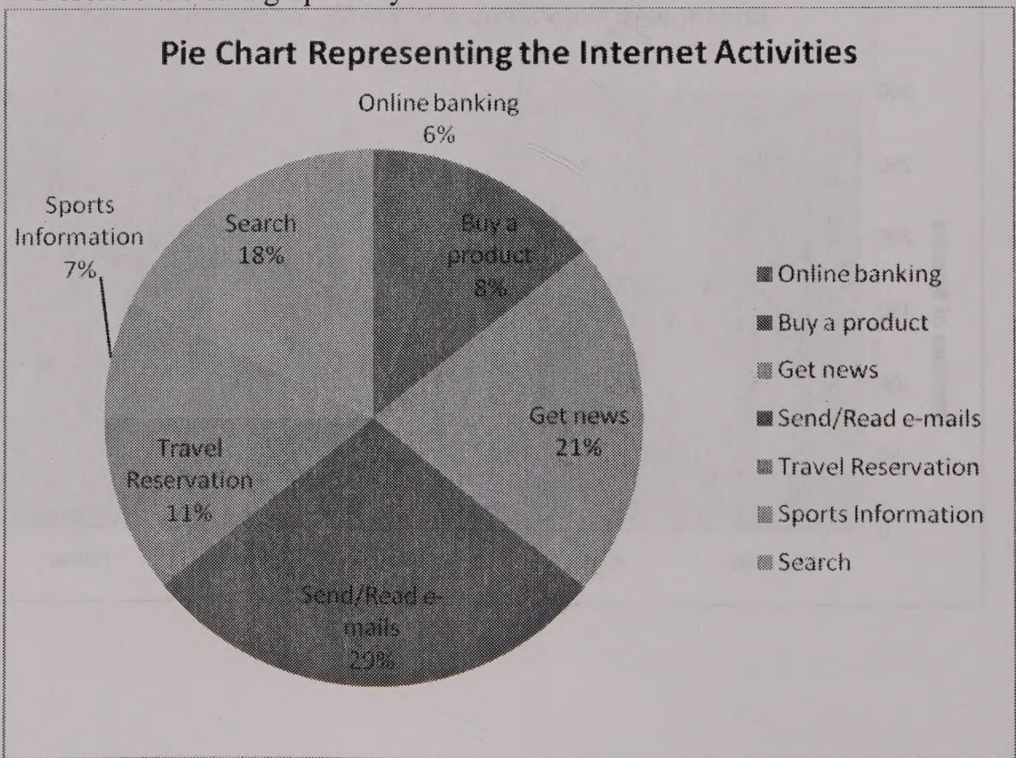


c. Bar chart to compare the number of endangered species in the United States to the number of endangered species outside the United States



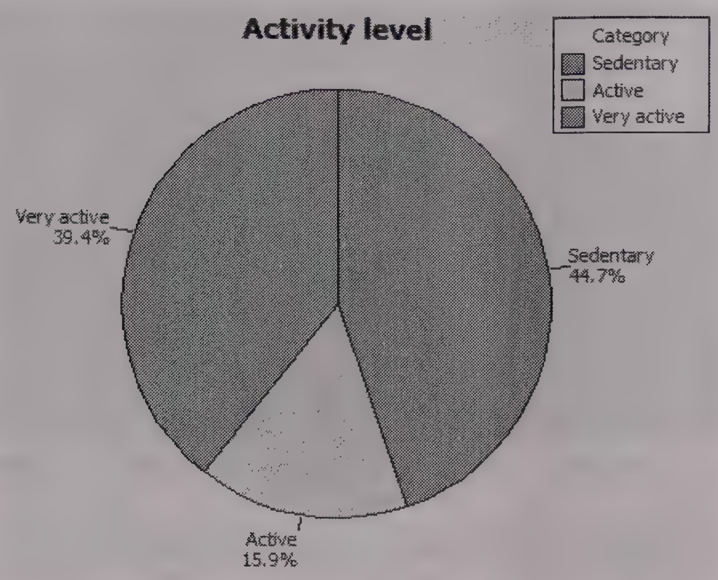
1.16

Describe the data graphically

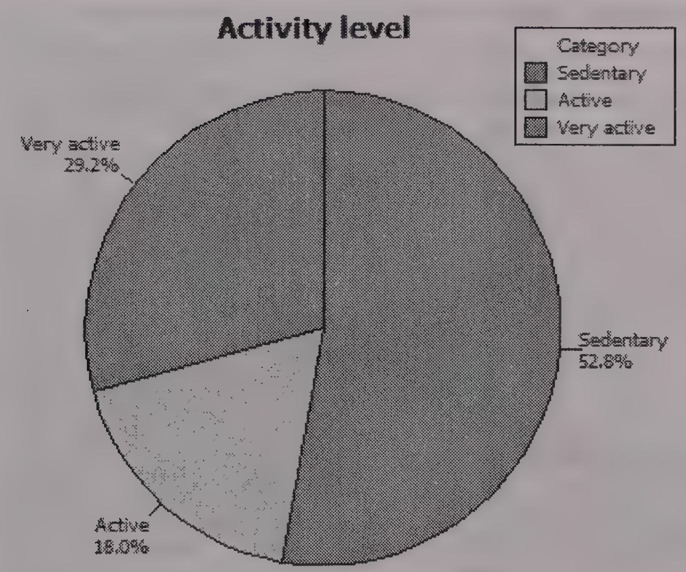


1.18

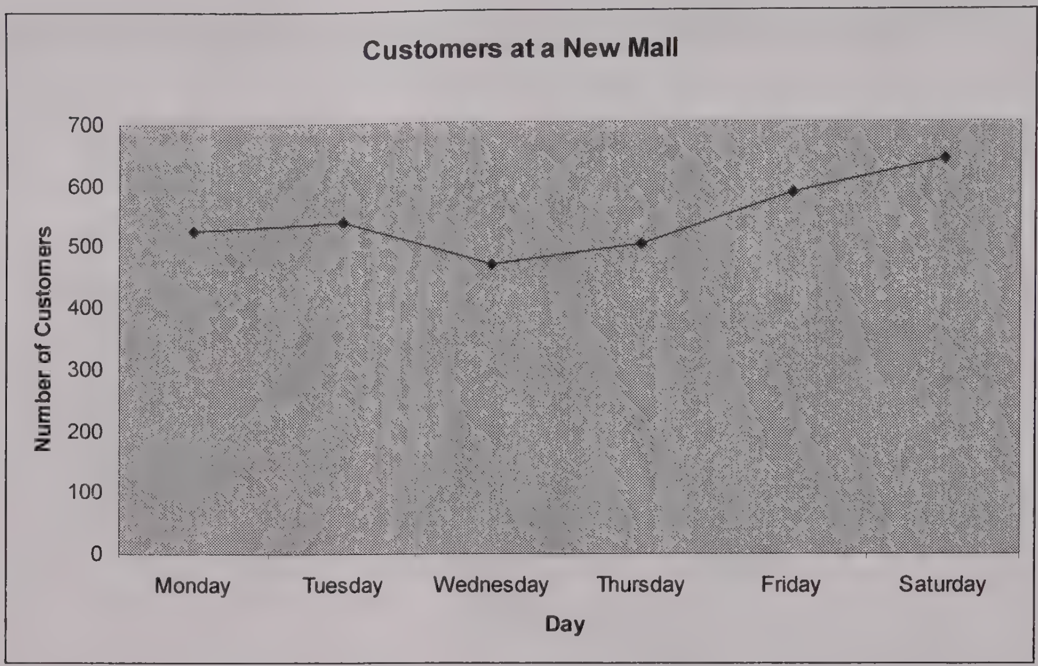
a. Pie chart of the percent of males in each of the activity level categories.



b. Pie chart of the percent of females in each of the activity level categories.

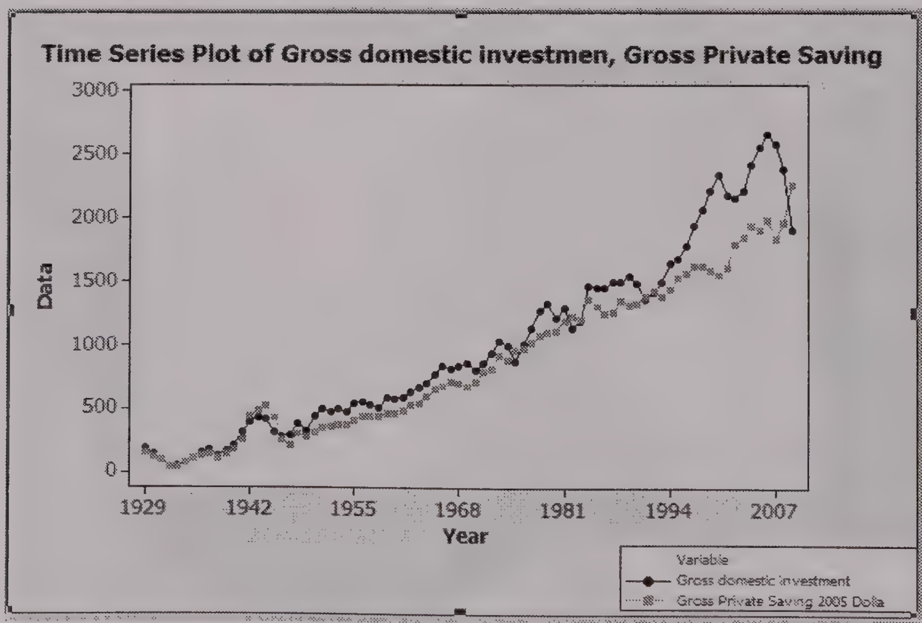


1.20



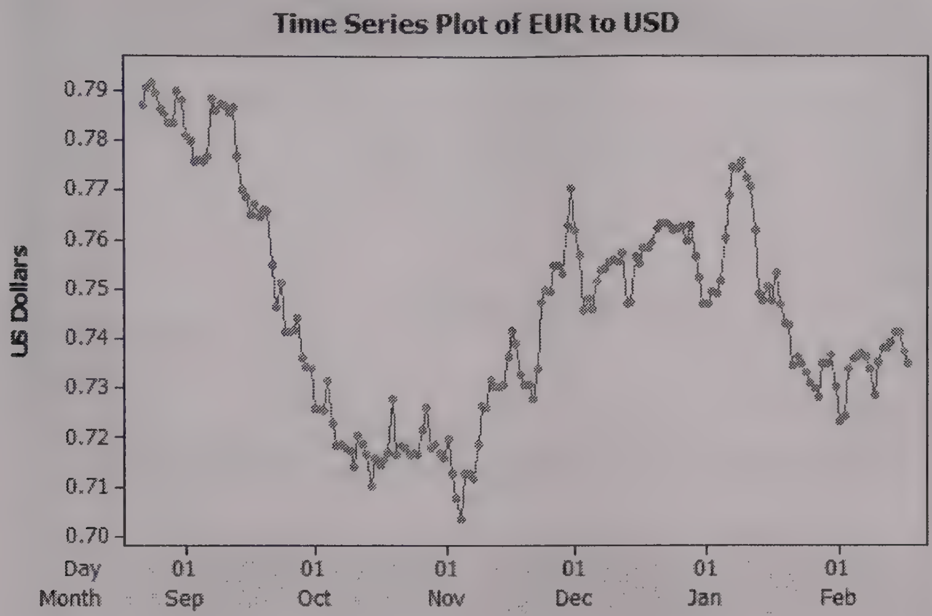
1.22

a. Time-series plot of Gross domestic investment and Gross private domestic



1.24

a. The Euro (EUR) compared to 1 U.S. Dollar (USD)

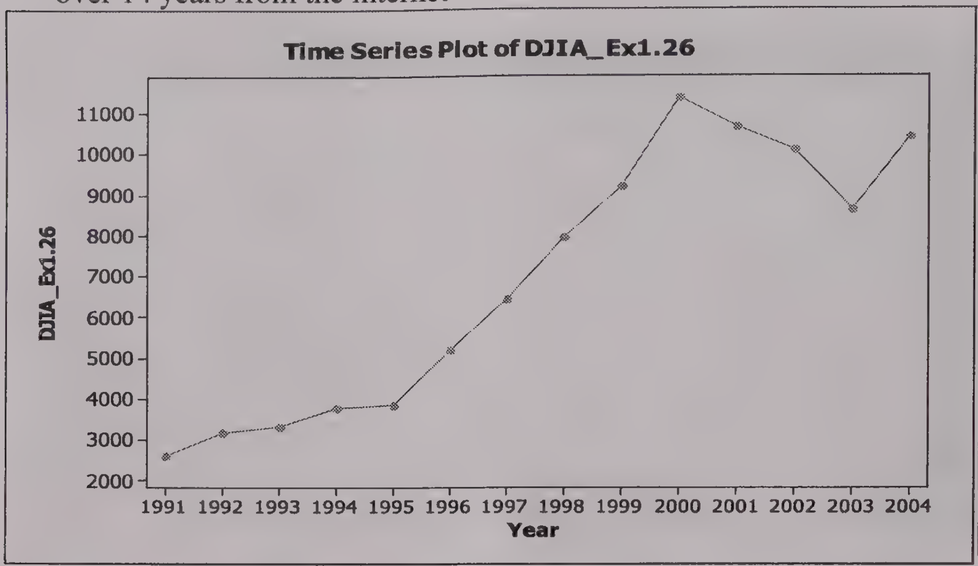


b. The GBP compared to 1 U.S. Dollar (USD)

c. Answers may vary.

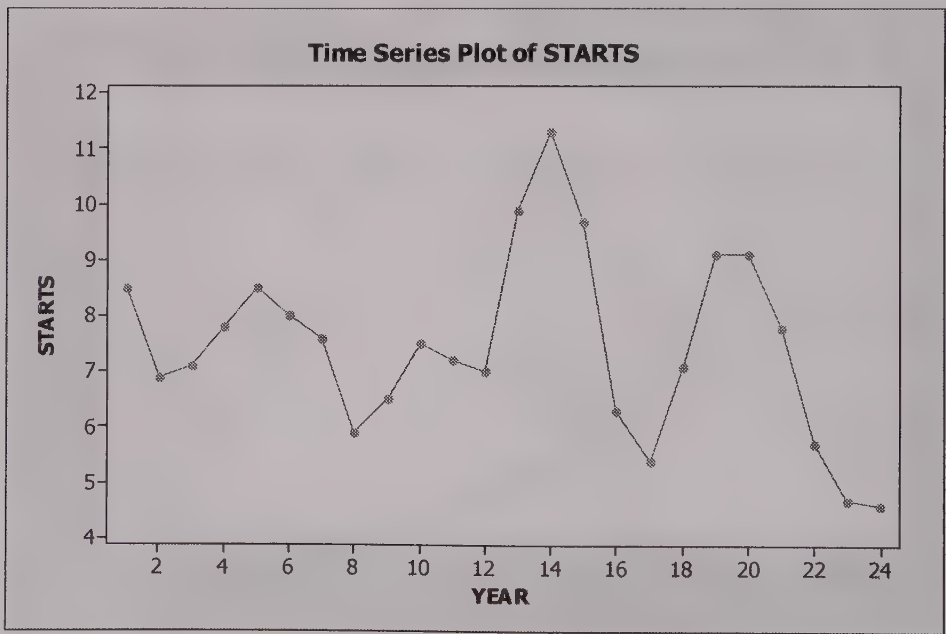
1.26

Time-series plot of a stock market index (Dow Jones Industrial Average) over 14 years from the internet



1.28

Time-series plot of Housing Starts data



1.30

a. 5 – 7 classes
b. 7 – 8 classes
c. 8 – 10 classes
d. 8 – 10 classes
e. 10 – 11 classes

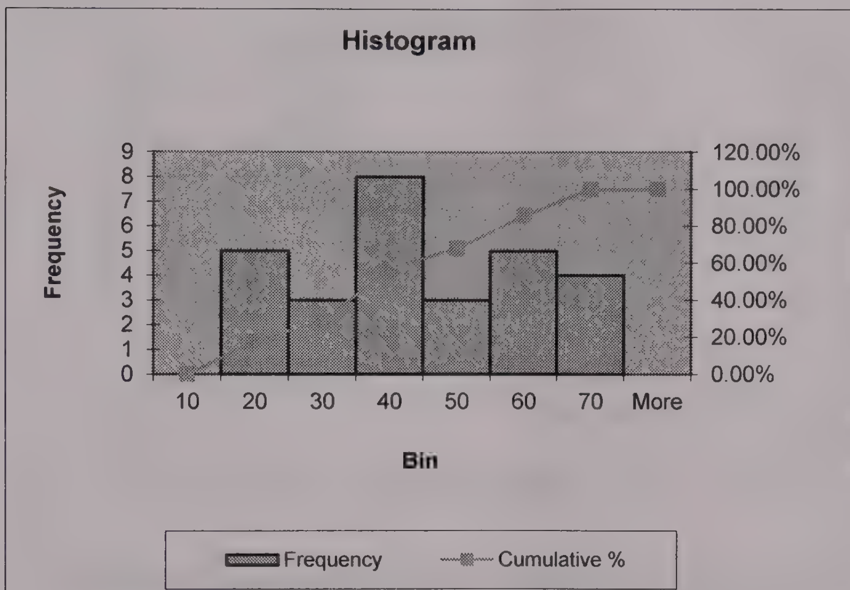
1.32

a. frequency distribution

Bin	Frequency
10	0
20	5
30	3
40	8
50	3
60	5
70	4
More	0

b. histogram and

c. ogive



d. stem-and-leaf display

Stem-and-Leaf Display: Data_Ex1.32

```
Stem-and-leaf of Data_Ex1.32  N  = 28
Leaf Unit = 1.0

  2   1  23
  5   1  557
  7   2  14
  8   2  8
  9   3  2
 (6)  3  567799
 13   4  0144
  9   4
  9   5  14
  7   5  699
  4   6  24
  2   6  55
```

1.34

Classes	Frequency	a. Relative Frequency	b. Cumulative Frequency	c. Cumulative Relative Frequency
0<10	8	16.33%	8	16.33%
10<20	10	20.41%	18	36.74%
20<30	13	26.53%	31	63.27%
30<40	12	24.49%	43	87.76%
40<50	6	12.24%	49	100.00%
Total	49	100.00%		

1.36

a. Construct a cumulative relative frequency distribution

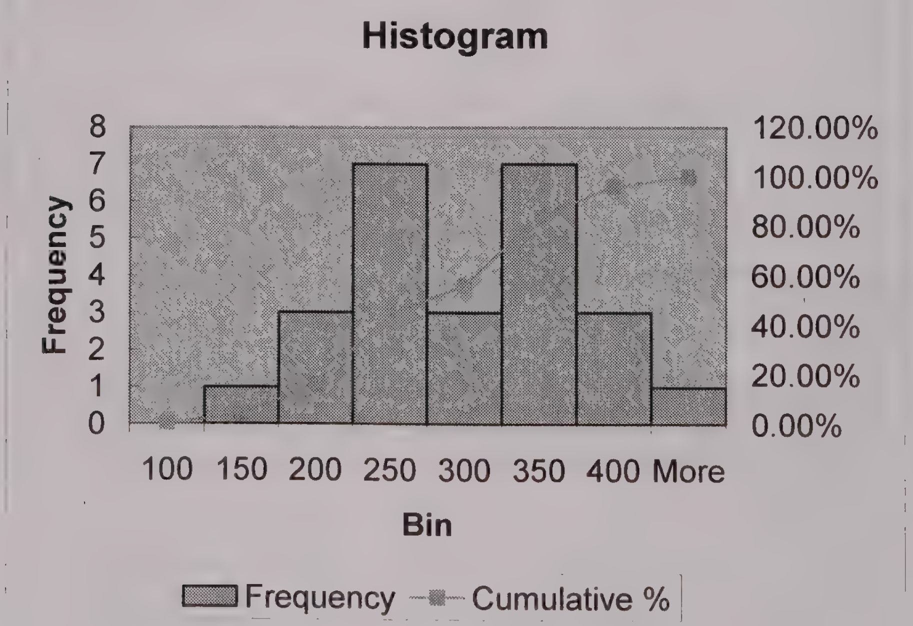
Age	Percent	Relative Cumulative Frequency
18-24	11.30%	11.30%
25-34	19.11%	30.41%
35-44	23.64%	54.05%
45-54	23.48%	77.53%
55+	22.48%	100.01%
Total		

b. 54.05%

c. 69.59%

1.38

a. Histogram of the Returns data



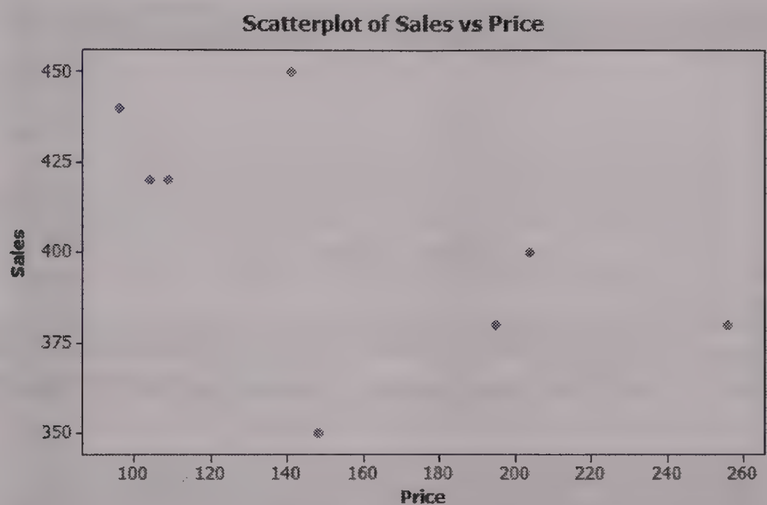
b. stem-and-leaf display

Stem-and-Leaf Display: Returns

```
Stem-and-leaf of Returns  N  = 25
Leaf Unit = 10

1    1  3
4    1  899
11   2  0014444
(3)   2  589
11   3  0000122
4    3  689
1    4
1    4
1    5  0
```

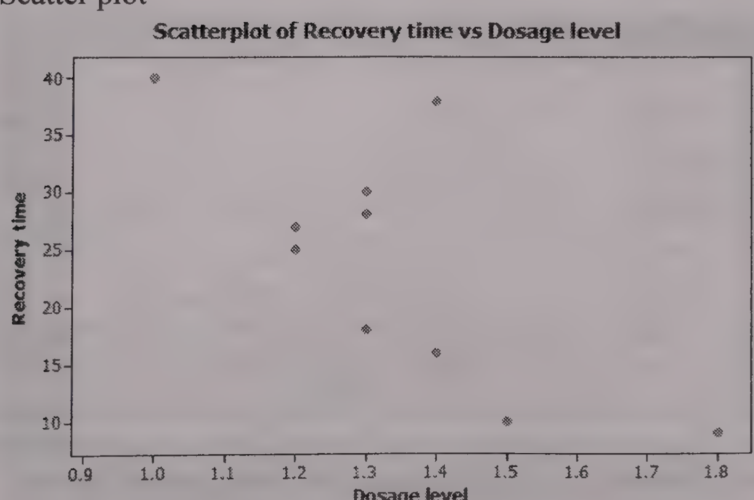
1.40

Scatter plot



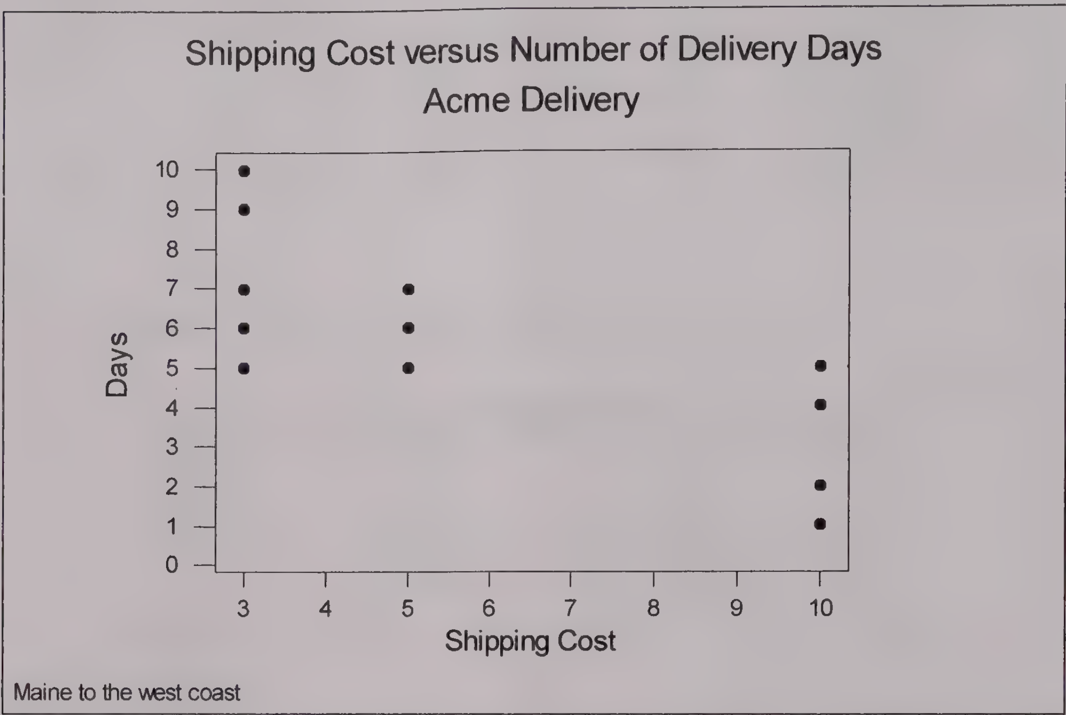
1.42

Scatter plot



1.44

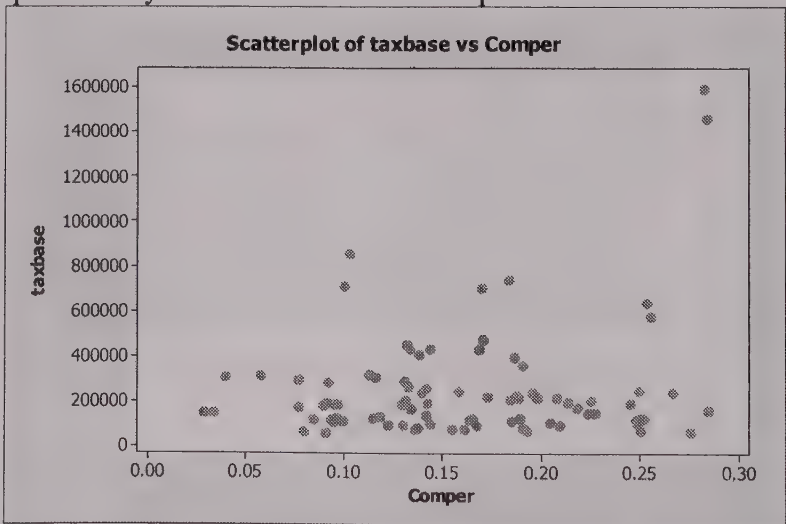
Acme Delivery – relation between shipping cost and number of delivery days



The relationship appears to be negative; however there is significant variability in delivery time at each of the three shipping costs – regular, $3; fast, $5; and lightning, $10.

1.46

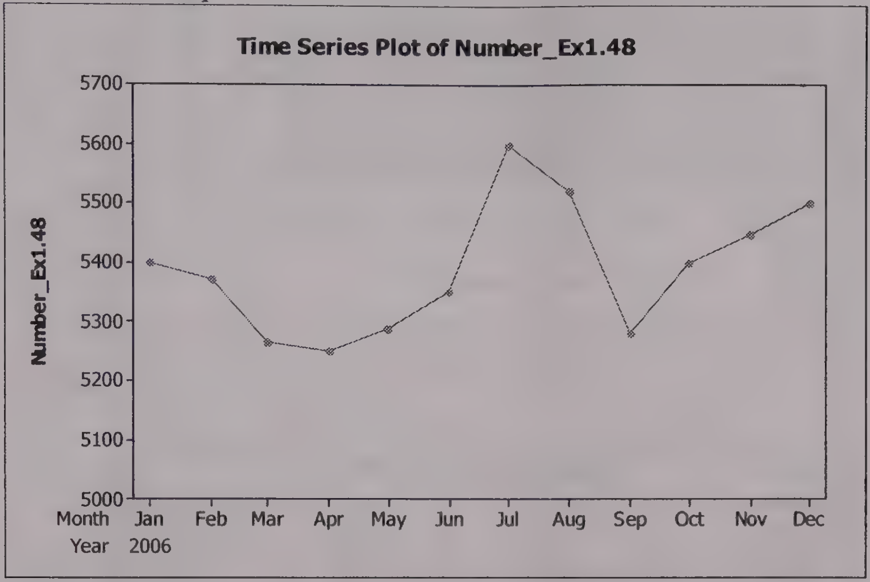
Scatter plot of Citydat – taxbase versus comper



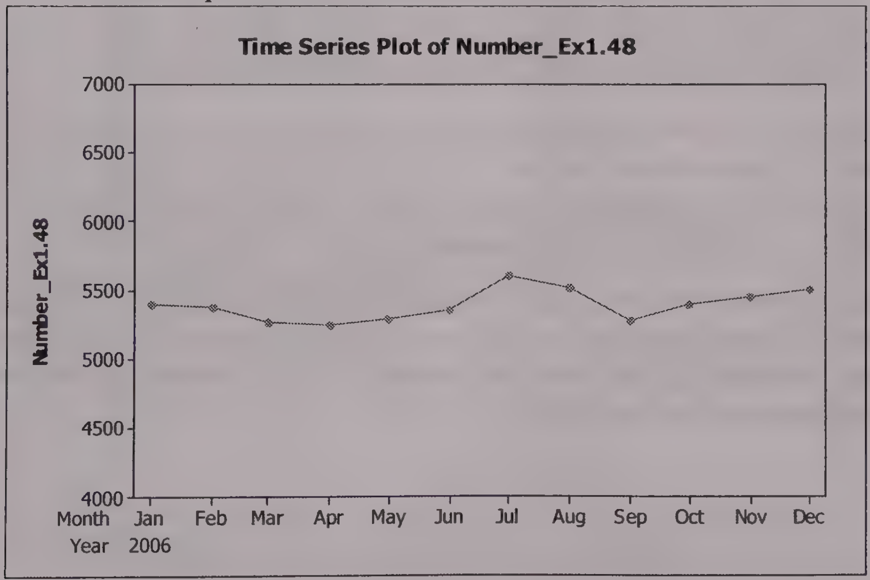
No relationship between the two variables and hence no evidence that emphasis on attracting a larger percentage of commercial property increases the tax base. The two outlier points on the right side of the plot might be used to argue that a very high percentage of commercial property will provide a larger tax base. That argument, however, is contrary to the overall pattern of the data.

1.48

a. Time-series plot with vertical scale from 5000 to 5700.



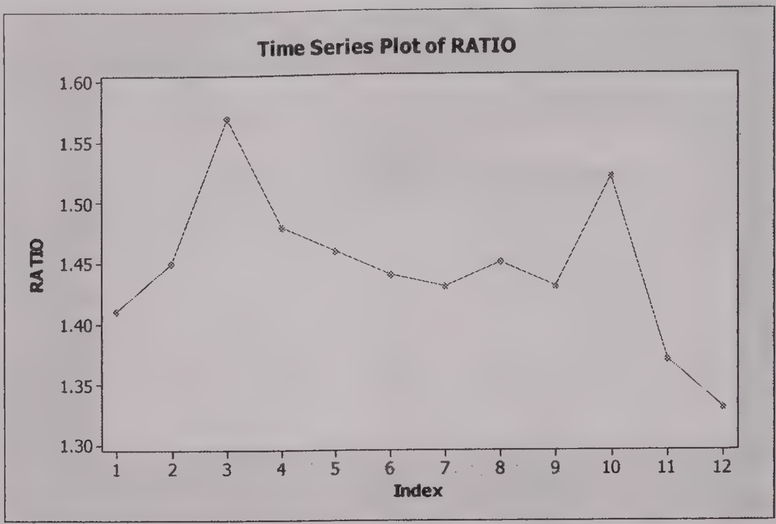
b. Time-series plot with vertical scale from 4000 to 7000.



b. Differences between the two graphs include the variability of the data series. One graph suggests greater variability in the data series while the other one suggests a relatively flat line with less variability. Keep in mind the scale on which the measurements are made.

1.50

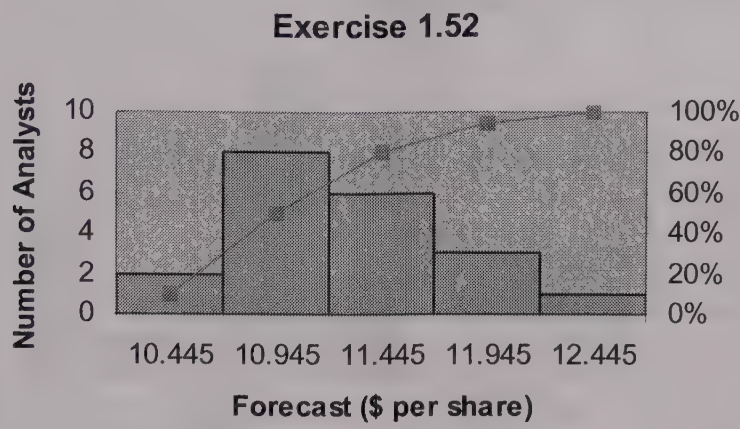
Draw two time-series plots for Inventory Sales with different vertical ranges.



Differences between the two graphs include the variability of the data series. One graph suggests greater variability in the data series while the other one suggests a relatively flat line with less variability. Keep in mind the scale on which the measurements are made.

1.52

a. Draw a histogram of 20 forecasted earnings per share.



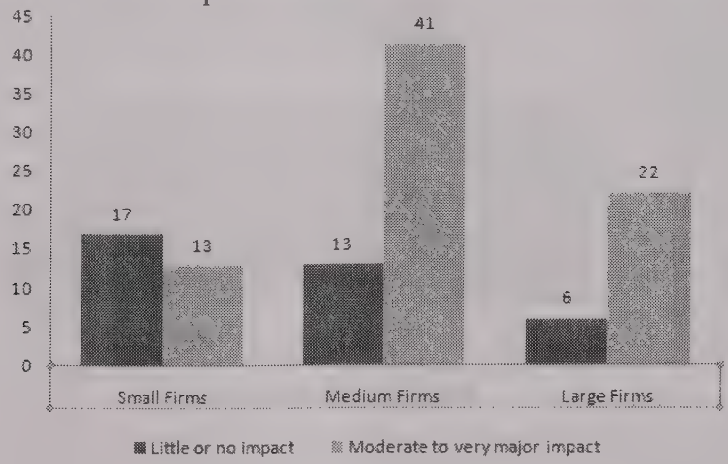
Answer to b., c. and d. are:

	(b)	(c)	(d)
Frequency	*Relative Freq.*	*Cumulative Freq.*	*Cumulative %*
2	0.1	2	10.00%
8	0.4	10	50.00%
6	0.3	16	80.00%
3	0.15	19	95.00%
1	0.05	20	100.00%

d. Cumulative relative frequencies are in the last column of the table above. These numbers indicate the percent of analysts who forecast that level of earnings per share and all previous classes, up to and including the current class. The third bin of 80% indicates that 80% of the analysts have forecasted up to and including that level of earnings per share.

1.54

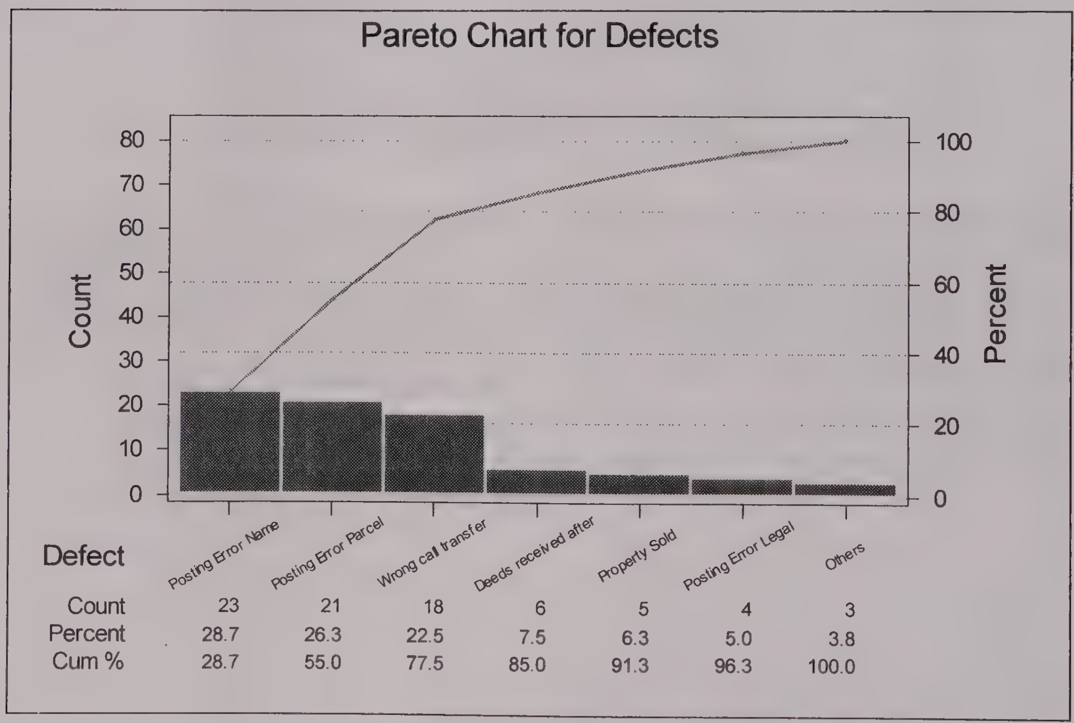
Cluster bar chart for impact of SOX



1.56

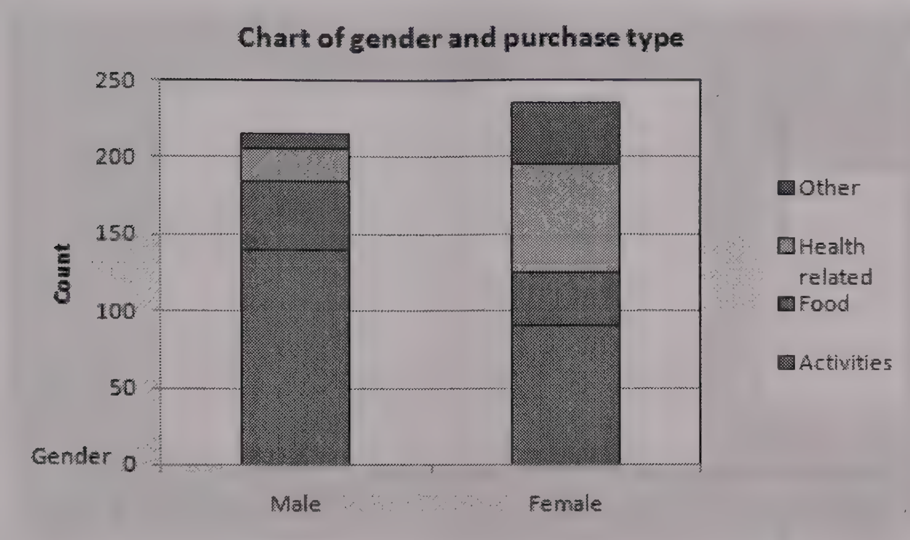
County Appraiser's Office – Data Entry Process

a. Pareto diagram



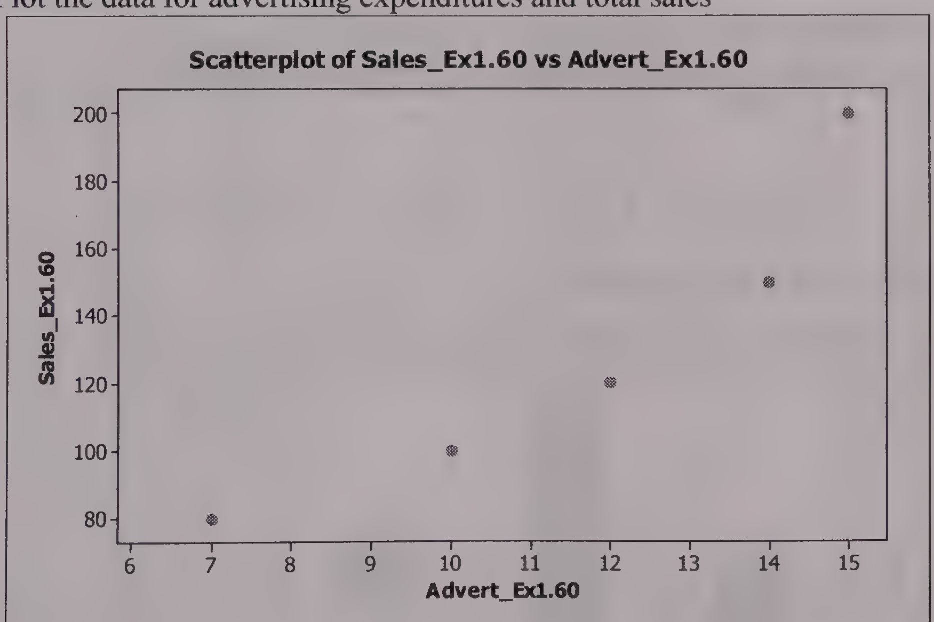
b. Recommendations should include a discussion of the data entry process. The data entry was being made by individuals with no knowledge of the data. Training of the data entry personnel should be a major recommendation. Increasing the size of the monitors used by the data entry staff would also reduce the number of errors.

1.58



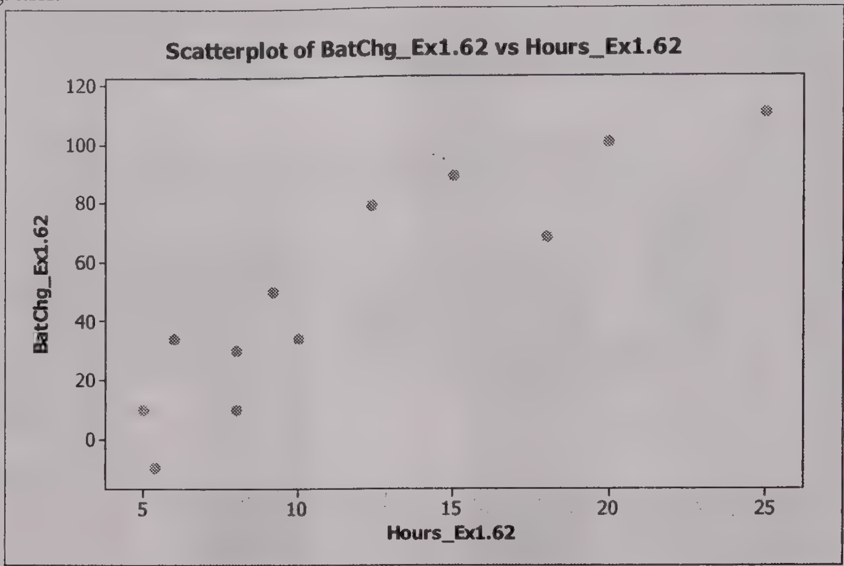
1.60

Plot the data for advertising expenditures and total sales



1.62

Plot the batting averages vs. hours spent per week in a weight-training program.



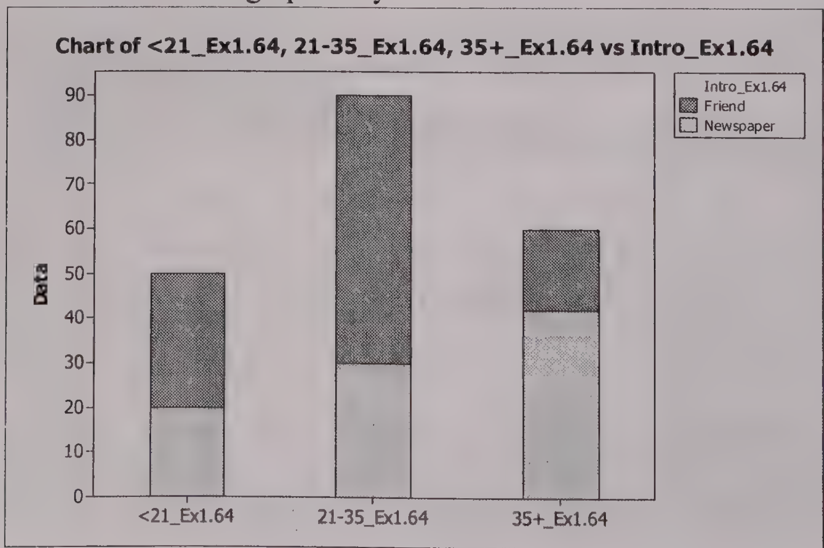
It appears that the number of hours spent per week in a special weight-training program is positively related to the change in their batting averages from the previous season.

1.64

a. Describe the new product data with a cross table

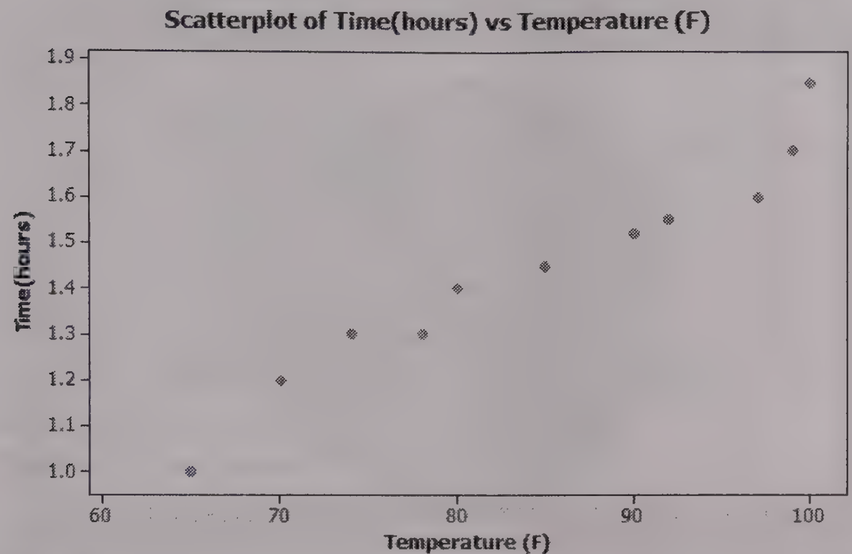
Age	Friend	Newspaper	Subtotal
<21 years	30	20	50
21-35	60	30	90
35+	18	42	60
Subtotal	108	92	200

b. Describe the data graphically



1.66

a. Scatterplot



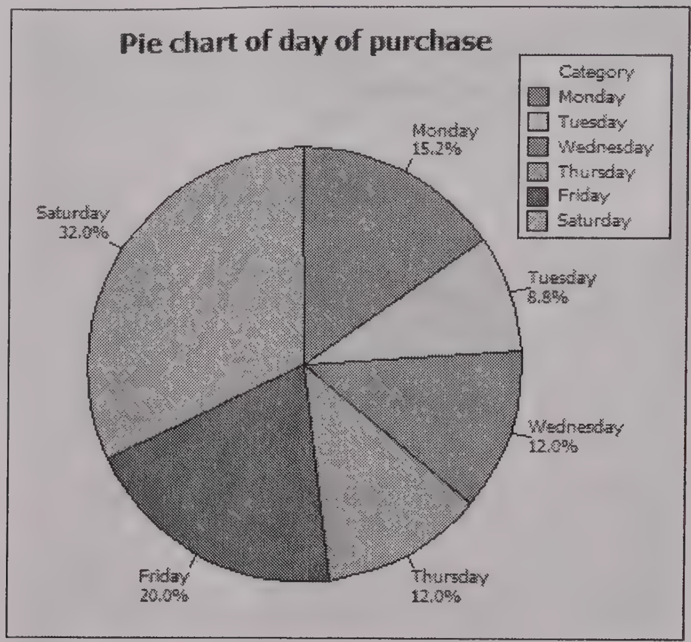
From the graph, it is evident that the temperature is positively related to the time it takes to mow.

1.68

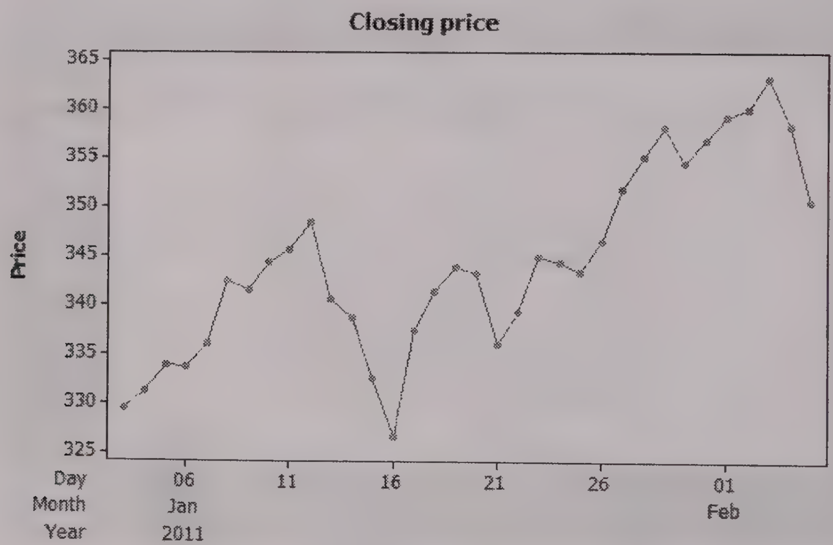
a. Cross table of method of payment and day of purchase for Florin data file.

Payment	M	T	W	Th	F	S	Tot
Am Ex	7	0	3	4	3	6	23
MC	1	4	4	2	4	9	24
Visa	6	6	4	5	8	10	39
Cash	3	1	0	0	3	9	16
Other	2	0	4	4	7	6	23
Subtotal	19	11	15	15	25	40	125

b. Pie chart of day of purchase



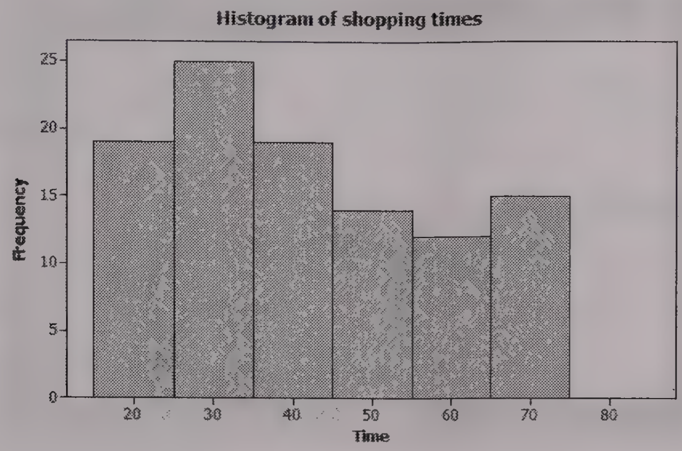
1.70 a. Time-series plot of the closing costs



b. Answers may vary.

1.72

a. Histogram of shopping times



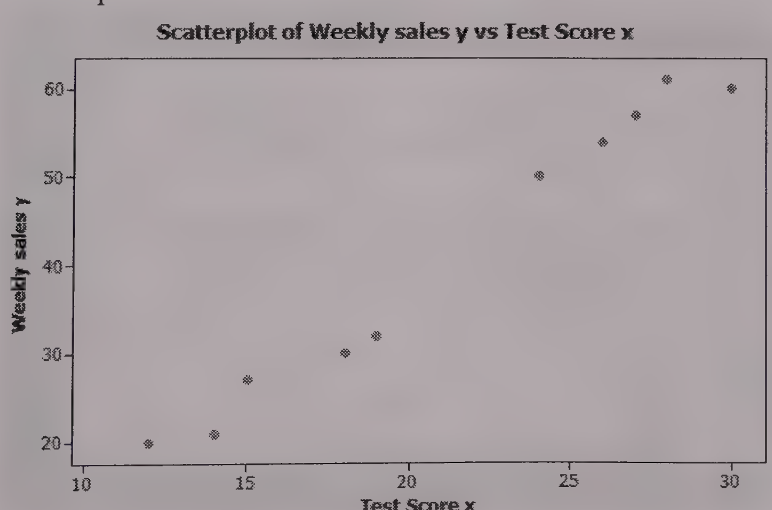
b. Stem-and-leaf display

Stem-and-Leaf Display: Time

```
Stem-and-leaf of Time  N  = 104
Leaf Unit = 1.0


9    1  888888899
19   2  0011111333
26   2  5555558
44   3  000011113344444444
52   3  77777788
52   4  00001222233
41   4  555678
35   5  00112222
27   5  5777799
20   6  00334
15   6  7889999
8    7  00001133
```

1.74

Scatter plot



Chapter 2:

Describing Data: Numerical

2.2

Number of complaints: 8, 8, 13, 15, 16

a. Compute the mean number of weekly complaints:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{60}{5} = 12$$

b. Calculate the median = middlemost observation = 13

c. Find the mode = most frequently occurring value = 8

2.4

Department store % increase in dollar sales: 2.9, 3.1, 3.7, 4.3, 5.9, 6.8, 7.0, 7.3, 8.2, 10.2

a. Calculate the mean number of weekly complaints: $\bar{x} = \frac{\sum x_i}{n} = \frac{59.4}{10} = 5.94$

b. Calculate the median = middlemost observation: $\frac{5.9+6.8}{2} = 6.35$

2.6

Daily sales (in hundreds of dollars): 6, 7, 8, 9, 10, 11, 11, 12, 13, 14

a. Find the mean, median, and mode for this store

Mean = $\bar{x} = \frac{\sum x_i}{n} = \frac{101}{10} = 10.1$

Median = middlemost observation = $\frac{10 + 11}{2} = 10.5$

Mode = most frequently occurring observation = 11

b. Find the five-number summary

Q1 = the value located in the $0.25(n + 1)^{th}$ ordered position
= the value located in the 2.75^{th} ordered position
= 7 + 0.25(8 – 7) = 7.25

Q3 = the value located in the $0.75(n + 1)^{th}$ ordered position
= the value located in the 8.25^{th} ordered position
= 12 + 0.75(13 – 12) = 12.75

Minimum = 6

Maximum = 14

Five - number summary:

minimum < Q1 < median < Q3 < maximum

6 < 7.25 < 10.5 < 12.75 < 14

2.8

Ages of 12 students: 18, 19, 21, 22, 22, 22, 23, 27, 28, 33, 36, 36

a. $\bar{x} = \sum \frac{x_i}{n} = \frac{307}{12} = 25.58$

b. Median = 22.50

c. Mode = 22

2.10

a. $\bar{x} = \sum \frac{x_i}{n} = \frac{282}{33} = 8.545$

b. Median = 9.0

c. The distribution is slightly skewed to the left since the mean is less than the median.

d. The five-number summary

Q1 = the value located in the $0.25(n + 1)^{th}$ ordered position
= the value located in the 8.5^{th} ordered position
$= 6 + 0.5(6 - 6) = 6$

Q3 = the value located in the $0.75(n + 1)^{th}$ ordered position
= the value located in the 25.5^{th} ordered position
$= 10 + 0.5(11 - 10) = 10.5$

Minimum = 2

Maximum = 21

Five - number summary:

minimum < Q1 < median < Q3 < maximum

$2 < 6 < 9 < 10.5 < 21$

2.12

The variance and standard deviation are

x_i	DEVIATION ABOUT THE MEAN, $(x_i - \bar{x})$	SQUARED DEVIATION ABOUT THE MEAN, $(x_i - \bar{x})^2$
6	–1	1
8	1	1
7	0	0
10	3	9
3	–4	16
5	–2	4
9	2	4
8	1	1
$\sum_{i=1}^{8} x_i = 56$	$\sum_{i=1}^{8} (x_i - \bar{x}) = 0$	$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 36$

$$\text{Sample mean} = \bar{x} = \frac{\sum_{i=1}^{8} x_i}{n} = \frac{56}{8} = 7$$

$$\text{Sample variance} = s^2 = \frac{\sum_{i=1}^{8}(x_i - \bar{x})^2}{n-1} = \frac{36}{8-1} = 5.143$$

$$\text{Sample standard deviation} = s = \sqrt{s^2} = \sqrt{5.143} = 2.268$$

2.14

x_i	DEVIATION ABOUT THE MEAN, $(x_i - \bar{x})$	SQUARED DEVIATION ABOUT THE MEAN, $(x_i - \bar{x})^2$
10	1	1
8	–1	1
11	2	4
7	–2	4
9	0	0
$\sum_{i=1}^{5} x_i = 45$	$\sum_{i=1}^{5}(x_i - \bar{x}) = 0$	$\sum_{i=1}^{5}(x_i - \bar{x})^2 = 10$

$$\text{Sample mean} = \bar{x} = \frac{\sum_{i=1}^{5} x_i}{n} = \frac{45}{5} = 9$$

$$\text{Sample variance} = s^2 = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^2}{n-1} = \frac{10}{4} = 2.5$$

$$\text{Sample standard deviation} = s = \sqrt{s^2} = \sqrt{2.5} = 1.581$$

$$\text{Coefficient of variation} = CV = \frac{s}{\bar{x}} \times 100\% = \frac{1.581}{9} \times 100\% = 17.57\%$$

2.16

Minitab Output

Stem-and-Leaf Display: Ex2.16

```
Stem-and-leaf of Ex2.16  N  = 35
Leaf Unit = 1.0

3    1  234
10   1  5577889
17   2  0012333
(4)  2  7799
14   3  1
13   3  557788
7    4  002
4    4  59
2    5  3
1    5
1    6
1    6  5
```

IQR = $Q_3 - Q_1$
Q_1 = the value located in the $0.25(35 + 1)^{th}$ ordered position
= the value located in the 9^{th} ordered position
= 18
Q_3 = the value located in the $0.75(35 + 1)^{th}$ ordered position
= the value located in the 27^{th} ordered position
= 38
IQR = $Q_3 - Q_1$ = 38 – 18 = 20 years

2.18

Mean = 250, $\sigma = 20$
a. To determine k, use the lower or upper limit of the interval:
Range of observation is 190 to 310.
$\mu + \sigma k = 310$ or $\mu - \sigma k = 190$
$250 + 20k = 310$ or $250 - 20k = 190$
Solving both the equations we arrive at $k = 3$.
According to the Chebyshev's theorem, proportion must be at least
$100[1-(1/k^2)]\% = 100[1-(1/3^2)]\% =$ 75%. Therefore, approximately 88.89% of the observations are between 190 and 310.

b. To determine k, use the lower or upper limit of the intervals:
Range of observation is 190 to 310.
$\mu + \sigma k = 290$ or $\mu - \sigma k = 210$
$250 + 20k = 290$ or $250 - 20k = 210$
Solving both the equations we arrive at $k = 2$.
According to the Chebyshev's theorem, proportion must be at least
$100[1-(1/k^2)]\% = 100[1-(1/2^2)]\% =$ 75%. Therefore, approximately 75% of the observations are between 210 and 290.

2.20

Compare the annual % returns on common stocks vs. U.S. Treasury bills

Minitab Output:

Descriptive Statistics: Stocks_Ex2.20, TBills_Ex2.20

Variable	N	N*	Mean	SE Mean	TrMean	StDev	Variance	CoefVar	Minimum
Stocks_Ex2.20	7	0	8.16	8.43	*	22.30	497.39	273.41	-26.50
TBills_Ex2.20	7	0	5.786	0.556	*	1.471	2.165	25.43	3.800

Variable	Q1	Median	Q3	Maximum	Range	IQR
Stocks_Ex2.20	-14.70	14.30	23.80	37.20	63.70	38.50
TBills_Ex2.20	4.400	5.800	6.900	8.000	4.200	2.500

a. Compare the means of the populations
Using the Minitab output

$\mu_{stocks} = 8.16$, $\mu_{Tbills} = 5.786$
Therefore, the mean annual % return on stocks is higher than the return for U.S. Treasury bills

b. Compare the standard deviations of the populations
Using the Minitab output,
$\sigma_{stocks} = 22.302$, $\sigma_{Tbills} = 1.471$
Standard deviations are not sufficient for comparision.
We need to compare the coefficient of variation rather than the standard deviations.

$$CV_{Stocks} = \frac{s}{x} \times 100 = \frac{8.16}{22.302} \times 100 = 70.93\%$$

$$CV_{Tbills} = \frac{s}{x} \times 100 = \frac{5.79}{1.471} \times 100 = 6.60\%$$

Therefore, the variability of the U.S. Treasury bills is much smaller than the return on stocks.

2.22
Minitab Output:
Descriptive Statistics: Weights

```
Variable    N  N*    Mean  SE Mean   StDev  Variance  CoefVar  Minimum      Q1
Weights    75   0  3.8079   0.0118  0.1024    0.0105     2.69   3.5700  3.7400

Variable   Median      Q3  Maximum   Range
Weights    3.7900  3.8700   4.1100  0.5400
```

a. Using the Minitab output, range = $4.11 - 3.57 = 0.54$, standard deviation = 0.1024, variance = 0.010486

b. $IQR = Q3 - Q1 = 3.87 - 3.74 = .13$. This tells that the range of the middle 50% of the distribution is 0.13

c. Coefficient of variation = $CV = \frac{s}{x} \times 100 = \frac{0.1024}{3.8079} \times 100 = 2.689\%$

2.24

a. Standard deviation (s) of the assessment rates:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{583.75}{39}} = \sqrt{14.974} = 3.8696$$

b. The distribution is approximately mounded. Therefore, the empirical rule applies. Approximately 95% of the distribution is expected to be within +/- 2 standard deviations of the mean.

2.26

a. mean without the weights $\bar{x} = \sum \frac{x_i}{n} = \frac{21}{5} = 4.2$

b. weighted mean

w_i	x_i	$w_i x_i$
8	4.6	36.8
3	3.2	9.6
6	5.4	32.4
2	2.6	5.2
5	5.2	26.0
24		110.0

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{110}{24} = 4.583$$

2.28

Class	m_i	f_i	$m_i f_i$	$(m_i - \bar{x})$	$(m_i - \bar{x})^2$	$f_i (m_i - \bar{x})^2$
4 < 10	7	8	56	–8.4	70.56	564.48
10 < 16	13	15	195	–2.4	5.76	86.4
16 < 22	19	10	190	3.6	12.96	129.6
22 < 28	25	7	175	9.6	92.16	645.12
		$\sum f_i = 40$	$\sum m_i f_i = 616$			$\sum f_i (m_i - \bar{x})^2 = 1425.6$

a. Sample mean $= \bar{x} = \frac{\sum m_i f_i}{n} = \frac{616}{40} = 15.4$

b. Sample variance $= s^2 = \frac{\sum_{i=1}^{K} f_i (m_i - x_i)^2}{n-1} = \frac{1425.6}{39} = 36.554$

Sample standard deviation $= s = \sqrt{s^2} = \sqrt{36.554} = 6.046$

2.30

Based on a sample of n=50:

m_i	f_i	$f_i m_i$	$(m_i - \bar{x})$	$(m_i - \bar{x})^2$	$f_i(m_i - \bar{x})^2$
0	21	0	-1.4	1.96	41.16
1	13	13	-0.4	0.16	2.08
2	5	10	0.6	0.36	1.8
3	4	12	1.6	2.56	10.24
4	2	8	2.6	6.76	13.52
5	3	15	3.6	12.96	38.88
6	2	12	4.6	21.16	42.32
Sum	**50**	**70**			**150**

a. Sample mean number of claims per day = $\bar{X} = \dfrac{\sum f_i m_i}{n} = \dfrac{70}{50} = 1.40$

b. Sample variance = $s^2 = \dfrac{\sum f_i(m_i - \bar{x})^2}{n-1} = \dfrac{150}{49} = 3.0612$

Sample standard deviation = $s = \sqrt{s^2} = 1.7496$

2.32

Estimate the sample mean and sample standard deviation

Class	m_i	f_i	$f_i m_i$	$(m_i - \bar{x})$	$(m_i - \bar{x})^2$	$f_i(m_i - \bar{x})^2$
9.95-10.45	10.2	2	20.4	-0.825	0.681	1.361
10.45-10.95	10.7	8	85.6	-0.325	0.106	0.845
10.95-11.45	11.2	6	67.2	0.175	0.031	0.184
11.45-11.95	11.7	3	35.1	0.675	0.456	1.367
11.95-12.45	12.2	1	12.2	1.175	1.381	1.381
Sum		**20**	**220.5**			**5.138**

a. sample mean = $\bar{X} = \dfrac{\sum f_i m_i}{n} = \dfrac{220.5}{20} = 11.025$

b. sample variance = $s^2 = \dfrac{\sum f_i(m_i - \bar{x})^2}{n-1} = \dfrac{5.138}{19} = 0.2704$

sample standard deviation = $s = \sqrt{s^2} = 0.520$

2.34

Using Table 1.7Minutes	m_i	f_i	$f_i m_i$	$(m_i - \bar{x})$	$(m_i - \bar{x})^2$	$f_i(m_i - \bar{x})^2$
220<230	225	5	1125	-36.545	1335.57	6677.851
230<240	235	8	1880	-26.545	704.6612	5637.289
240<250	245	13	3185	-16.545	273.7521	3558.777
250<260	255	22	5610	-6.5455	42.84298	942.5455
260<270	265	32	8480	3.45455	11.93388	381.8843
270<280	275	13	3575	13.4545	181.0248	2353.322
280<290	285	10	2850	23.4545	550.1157	5501.157
290<300	295	7	2065	33.4545	1119.207	7834.446
		110	28770			32887.27

a. Using Equation 2.21, Sample mean, $\bar{x} = \frac{\sum f_i m_i}{n} = \frac{28770}{110} = 261.54545$

b. Using Equation 2.22, sample variance

$$s^2 = \frac{\sum f_i(m_i - \bar{x})^2}{n-1} = \frac{32887.27}{109} = 301.718\,;\ \ s = \sqrt{s^2} = 17.370$$

c. From Exercise 2.23, $\bar{x} = 261.05$ and $s^2 = 306.44$. Therefore, the mean value obtained in both the Exercises are almost same, however variance is slightly lower by 4.7219 compared to Exercise 2.23.

2.36

a. Compute the sample covariance

x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})\,(y_i - \bar{y})$
12	200	-7	49	-156	24336	1092
30	600	11	121	244	59536	2684
15	270	-4	16	-86	7396	344
24	500	5	25	144	20736	720
14	210	-5	25	-146	21316	730
95	1780	0	236	0	133320	5570
$\bar{x} = 19.00$	$\bar{y} = 356.00$		$s_x^2 = 59$		$s_y^2 = 33330$	Cov(x,y) = 1392.5
			s_x =7.681146		s_y =182.5650569	

$$Cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{5570}{4} = 1392.5$$

b. Compute the sample correlation coefficient

$$r = \frac{Cov(x,y)}{s_x s_y} = \frac{1392.5}{(7.6811)(182.565)} = 0.9930$$

2.38 Minitab output

Covariances: x_Ex2.38, y_Ex2.38

```
          x_Ex2.38  y_Ex2.38
x_Ex2.38   31.8987
y_Ex2.38    4.2680   34.6536
```

Correlations: x_Ex2.38, y_Ex2.38

```
Pearson correlation of x_Ex2.38 and y_Ex2.38 = 0.128
```

Using Minitab outputa. $Cov(x,y) = 4.268$

b. $r = 0.128$

c. Weak positive association between the number of drug units and the number of days to complete recovery. Recommend low or no dosage units.

2.40

a. Compute the covariance

x_i	y_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\,(y_i-\bar{y})$
5	55	-2	4	12.4	153.76	-24.8
6	53	-1	1	10.4	108.16	-10.4
7	45	0	0	2.4	5.76	0
8	40	1	1	-2.6	6.76	-2.6
9	20	2	4	-22.6	510.76	-45.2
35	213	0	10	0	785.2	-83
$\mu_x = 7.00$	$\mu_y = 42.60$		$\sigma_x^2 = 2.0$		$\sigma_y^2 = 157.04$	$Cov(x,y) = -16.6$
			$\sigma_x = 1.4142$		$\sigma_y = 12.532$	

=

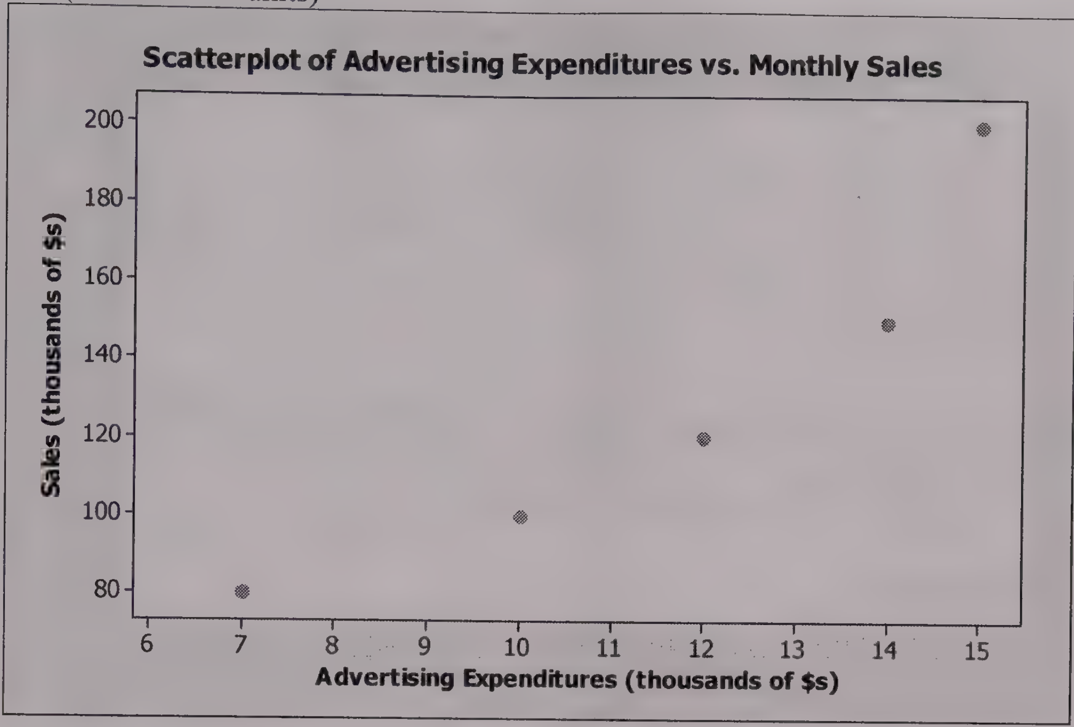
$$Cov(x,y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{N} = \frac{-83}{5} = -16.6$$

b. Compute the correlation coefficient

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x\sigma_y} = \frac{-16.6}{(1.4142)(12.5316)} = -.937$$

2.42

Scatter plot – Advertising expenditures (thousands of $s) vs. Monthly Sales (thousands of units)



x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})\ (y_i - \bar{y})$
10	100	-1.6	2.56	-30	900	48
15	200	3.4	11.56	70	4900	238
7	80	-4.6	21.16	-50	2500	230
12	120	0.4	0.16	-10	100	-4
14	150	2.4	5.76	20	400	48
58	650		41.2		8800	560
$\bar{x}$ = 11.60	$\bar{y}$ = 130.00		s_x^2 = 10.3		s_y^2 =2200	Cov(x,y) = 140
			s_x = 3.2094		s_y = 46.9042	

$$\text{Covariance} = Cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = 560 / 4 = 140$$

$$\text{Correlation} = \frac{Cov(x,y)}{s_x s_y} = \frac{140}{(3.2094)(46.9042)} = .93002$$

2.44

Air Traffic Delays (Number of Minutes Late)

m_i	f_i	$f_i m_i$	$(m_i - \bar{x})$	$(m_i - \bar{x})^2$	$f_i(m_i - \bar{x})^2$
5	30	150	-13.133	172.46	5173.90
15	25	375	-3.133	9.81	245.32
25	13	325	6.867	47.16	613.11
35	6	210	16.867	284.51	1707.07
45	5	225	26.867	721.86	3609.30
55	4	220	36.867	1359.21	5436.84
	83	1505			16785.54
$\bar{x}$	18.13			variance =	204.7017

a. Sample mean number of minutes late = 1505 / 83 = 18.1325

b. Sample variance = 16785.54/82 = 204.7017

Sample standard deviation = s = 14.307

2.46

For Location 2:

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	-9.2	84.64
19	8.8	77.44
2	-8.2	67.24
18	7.8	60.84
11	0.8	0.64
10	-0.2	0.04
3	-7.2	51.84
17	6.8	46.24
4	-6.2	38.44
17	6.8	46.24
102		473.6

$$\text{Mean} = \bar{x} = \frac{\sum x_i}{n} = \frac{102}{10} = 10.2$$

$$\text{Variance} = s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{473.6}{9} = 52.622$$

$$\text{Standard deviation} = s = \sqrt{s^2} = 7.254$$

For Location 3:

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
2	-16.4	268.96
3	-15.4	237.16
25	6.6	43.56
20	1.6	2.56
22	3.6	12.96
19	0.6	0.36
25	6.6	43.56
20	1.6	2.56
22	3.6	12.96
26	7.6	57.76
184		682.4
18.4		

$$\text{Mean} = \bar{x} = \frac{\sum x_i}{n} = \frac{184}{10} = 18.4$$

$$\text{Variance} = s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{682.4}{9} = 75.822$$

$$\text{Standard deviation} = s = \sqrt{s^2} = 8.708$$

For Location 4:

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
22	9.5	90.25
20	7.5	56.25
10	-2.5	6.25
13	0.5	0.25
12	-0.5	0.25
10	-2.5	6.25
11	-1.5	2.25
9	-3.5	12.25
10	-2.5	6.25
8	-4.5	20.25
125		200.5

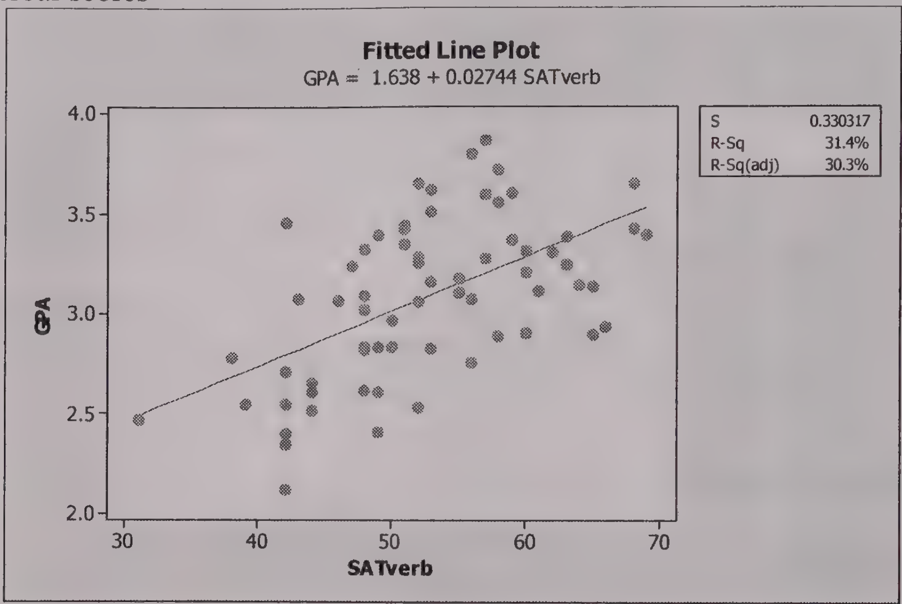
$$\text{Mean} = \bar{x} = \frac{\sum x_i}{n} = \frac{125}{10} = 12.5$$

$$\text{Variance} = s^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1} = \frac{200.5}{9} = 22.278$$

$$\text{Standard deviation} = s = \sqrt{s^2} = 4.720$$

2.48

a. Describe the data graphically between graduating GPA vs. entering SAT Verbal scores



b.

Correlations: GPA, SATverb

```
Pearson correlation of GPA and SATverb = 0.560
P-Value = 0.000
```

2.50

Mean of \$295 and standard deviation of \$63.

a. Find a range in which it can be guaranteed that 60% of the values lie.
   Use Chebyshev's theorem: at least 60% = $[1\text{-}(1/k^2)]$. Solving for k, $k = 1.58$. The interval will range from 295 +/- (1.58)(63) = 295 +/- 99.54. 195.46 up to 394.54 will contain at least 60% of the observations.

b. Find the range in which it can be guaranteed that 84% of the growth figures lie
   Use Chebyshev's theorem: at least 84% = $[1\text{-}(1/k^2)]$. Solving for k, $k = 2.5$. The interval will range from 295 +/- (2.50)(63) = 295 +/- 157.5. 137.50 up to 452.50 will contain at least 84% of the observations.

2.52

Tires have a lifetime mean of 29,000 miles and a standard deviation of 3,000 miles.

a. Find a range in which it can be guaranteed that 75% of the lifetimes of tires lies Use Chebyshev's theorem: at least 75% = $[1-(1/k^2)]$. Solving for $k = 2.0$. The interval will range from 29,000 ± (2.0)(3,000) = 29,000 ± 6,000 23,000 to 35,000 will contain at least 75% of the observations .

b. 95%: solve for $k = 4.47$. The interval will range from 29,000 ± (4.47)(3000) = 29,000 ± 13,416.41. 15,583.59 to 42,416.41 will contain at least 95% of the observations.

2.54

Minitab Output:

Descriptive Statistics: Time

```
           Total
Variable   Count    Mean   StDev  Variance  CoefVar  Minimum     Q1  Median
Time         104   41.68   16.86    284.35    40.46    18.00  28.50   39.00


Variable      Q3       Maximum
Time       56.50         73.00
```

Using the Minitab output

a. Mean shopping time = 41.68

b. Variance = 284.35

Standard deviation = 16.86

c. 95th percentile = the value located in the $0.95(n + 1)^{th}$ ordered position
= the value located in the 99.75th ordered position
= 70 + 0.75(70 –70) = 70.

d. Five - number summary:

minimum < Q1 < median < Q3 < maximum

18 < 28.50 < 39 < 56.50 < 73

e. Coefficient of variation = 40.46

f. Find the range in which ninety percent of the shoppers complete their shopping. Use Chebyshev's theorem: at least 90% = $[1-(1/k^2)]$. Solving for k, $k = 3.16$. The interval will range from 41.68 +/- (3.16)(16.86) = 41.88 +/- 53.28. -11.60 up to 94.96 will contain at least 90% of the observations.

2.56

x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
12	20	-9.3	86.49	-21.20	449.44	197.16
30	60	8.7	75.69	18.80	353.44	163.56
15	27	-6.3	39.69	-14.20	201.64	89.46
24	50	2.7	7.29	8.80	77.44	23.76
14	21	-7.3	53.29	-20.20	408.04	147.46
18	30	-3.3	10.89	-11.20	125.44	36.96
28	61	6.7	44.89	19.80	392.04	132.66
26	54	4.7	22.09	12.80	163.84	60.16
19	32	-2.3	5.29	-9.20	84.64	21.16
27	57	5.7	32.49	15.80	249.64	90.06
213	412		378.1		2505.6	962.4
$\bar{x} = 21.3$	$\bar{y} = 41.2$		$s_x^2 = 42.01$		$s_y^2 = 278.4$	
			$s_x = 6.4816$		$s_y = 16.6853$	

$$\text{Covariance} = Cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{962.4}{9} = 106.9333$$

$$\text{Correlation coefficient} = \frac{Cov(x,y)}{s_x s_y} = \frac{106.9333}{(6.4816)(16.6853)} = 0.9888$$

Chapter 3:

Probability

3.2 a. A intersection B contains the sample points that are in both A and B. The intersection $= (E_3, E_9)$

b. A union B contains the sample points in A or B or both. The union $= (E_1, E_2, E_3, E_7, E_8, E_9)$

c. A union B is not collectively exhaustive – it does not contain all of the possible sample points.

3.4 a. A intersection $B = (E_3, E_6)$

b. A union $B = (E_3, E_4, E_5, E_6, E_9, E_{10})$

c. A union B is not collectively exhaustive – it does not contain all of the possible sample points.

3.6 a. $(A \cap B)$ is the event that the Dow-Jones average rises on both days which is O_1. $(\bar{A} \cap B)$ is the event the Dow-Jones average does not rise on the first day but it rises on the second day which is O_3. The union between these two will be O_1 or O_3 either of which by definition is event B: the Dow-Jones average rises on the second day.

b. Since $(\bar{A} \cap B)$ is the event the Dow-Jones average does not rise on the first day but rises on the second day which is O_3 and because A is the event that the Dow-Jones average rises on the first day, then the union will be O_1, either the Dow-Jones average does not rise on the first day but rises on the second day or the Dow-Jones average rises on the first day or both. This is the definition of $A \cup B$.

3.8 The total number of outcomes in the sample space,

$$N = C_2^{12} = \frac{12!}{2!(12-2)!} = \frac{12!}{2!10!} = 66.$$

The number of ways to select 1 A from the 5 available, $C_1^5 = \dfrac{5!}{1!(5-1)!} = 5$.

The number of ways to select 1 B from the 7 available, $C_1^7 = \dfrac{7!}{1!(7-1)!} = 7$.

The number of outcomes that satisfy the condition of 1 A and 1 B is,
$N_A = 5 \times 7 = 35$.
Therefore, the probability that a randomly selected set of 2 will include 1A and 1 B is

$$P_A = \frac{N_A}{N} = \frac{C_1^5 C_1^7}{C_2^{12}} = \frac{5 \times 7}{66} = .53.$$

3.10 The total number of outcomes in the sample space,

$$N = C_4^{16} = \frac{16!}{4!(16-4)!} = 1,820.$$

The number of ways to select 2 As from the 10 available, $C_2^{10} = \dfrac{10!}{2!(10-2)!} = 45$.

The number of ways to select 2 Bs from the 6 available, $C_2^{6} = \dfrac{6!}{2!(6-2)!} = 15$.

The number of outcomes that satisfy the condition of 2 As and 2 Bs is,
$N_A = 45 \times 15 = 675$.
Therefore, the probability that a randomly selected set of 4 will include 2 As and 2 Bs is

$$P_A = \frac{N_A}{N} = \frac{C_2^{10} C_2^{6}}{C_4^{16}} = \frac{45 \times 15}{1,820} = .3709.$$

3.12 $P(A) = \dfrac{n_A}{n} = \dfrac{20,000}{180,000} = .1111$.

The probability of a random sample of 2 people from the city will contain 2 legal immigrants from Latin America is (.1111)(.1111) = .0123.

An alternative method to obtain the probability is,

$$P(A) = \frac{n_A}{n} = \frac{C_2^{20,000}}{C_2^{180,000}} = .0123.$$

3.14
a. $P(A) = P(10\% \text{ to} < 20\% \cup \text{more than } 20\%) = .33 + .21 = .54$
b. $P(B) = P(\text{less than } -10\% \cup -10\% \text{ to} < 0\%) = .04 + .14 = .18$
c. A complement is the event that the rate; of return is not more than 10%
d. $P(\bar{A}) = .04 + .14 + .28 = .46$
e. The intersection between 10% or more and return will be negative is the null or empty set.
f. $P(A \cap B) = 0$
g. The union of A and B is the event that are the rates of return of; less than –10%, -10% to < 0%, 10% to < 20% and more than 20%.
h. $P(A \cup B) = .04 + .14 + .33 + .21 = .72$
i. A and B are mutually exclusive because their intersection is the null set.
j. A and B are not collectively exhaustive because their union does not equal 1.

3.16 Events A and B of Exercise 3.2 are not mutually exclusive.
$P(A) = .4$, $P(B) = .4$, and $P(A \cup B) = .6$. $P(A \cup B) = P(A) + P(B) = .4 + .4 = .8 > .6$.
Therefore, if two events are not mutually exclusive, the probability of their union cannot equal the sum of their individual probabilities.

3.18 a. $P(X < 3) = .29 + .36 + .22 = .87$
b. $P(X > 1) = .22 + .10 + .03 = .35$
c. By the third probability postulate, the probabilities of all outcomes in the sample space must sum to one.

3.20 $P(A) = .40$, $P(B) = .45$, $P(A \cup B) = .85$
By the Addition Rule, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Therefore, $.85 = .40 + .45 - P(A \cap B)$
$$P(A \cap B) = .40 + .45 - .85 = 0$$

3.22 $P(A) = .60$, $P(B) = .45$, $P(A \cap B) = .30$
By the Addition Rule,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Therefore, $P(A \cup B) = .60 + .45 - .30 = .75$

3.24 $P(A) = .80$, $P(B) = .10$, $P(A \cap B) = .08$
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{.08}{.10} = .80$$
A and B are independent since the $P(A|B)$ of .80 equals the $P(A)$ of .80

3.26 $P(A) = .70$, $P(B) = .80$, $P(A \cap B) = .50$
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{.50}{.80} = .625$$
A and B are not independent since the $P(A|B)$ of .625 does not equal the $P(A)$ of .70

3.28 a. $P_7^7 = 7! = 5{,}040$
b. The probability that the guess will turn out to be correct is 1 / 5,040 = 0.0001984

3.30 $P_3^6 = 6!/3! = 120$. Therefore, the probability of selecting in the correct order, the three best performing stocks by chance is 1/120 = .00833

3.32 Ignoring the possibility of ties, the number of different predictions which could be done is $P_3^5 = 5!/2! = 60$. Therefore, the probability of making the correct prediction by chance is 1/60 = .0167

3.34 a. $P_2^7 = 7!/5! = 42$
b. $P_1^6 = 6!/5! = 6$
c. $P_1^6 = 6!/5! = 6$
d. Probability of being chosen as the heroine = 6 chances out of 42 = 6/42 = 1/7 = .1429. More direct way: Since there are seven candidates for 1 part – a randomly chosen candidate would have a 1 in 7 chance of getting any specific part.
e. Since being chosen as the heroine or as the best friend are mutually exclusive, the $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/7 + 1/7 - 0 = 2/7 = .2857$.
More direct way: Since there are seven candidates for 2 parts – a randomly chosen candidate would have a 2 in 7 chance of getting apart.

3.36 a. $C_2^6 = 6!/2!4! = 15$, $C_2^4 = 4!/2!2! = 6$. Because the selections are independent, there are (15)(6) = 90 different sets of funds from which to choose.
b. *P*(no U.S. fund under performs) = $C_2^5/15 = 5!/2!3!/15 = 10/15$. *P*(no foreign fund under performs) = $C_2^3/6 = 3!/1!2!/6 = 3/6$, *P*(at least one fund under performs) = 1 – *P*(no fund under performs) = 1-(10/15)(3/6) = 2/3 = 0.6667.

3.38 Let *A* – customer asks for assistance, *B* – customer makes a purchase, $A \cap B$ – both. Then, P(a customer does at least one of these two) = $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ = .30 + .20 - .15 = .35

3.40 $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C)$ = .02 + .01 + .04 - .0002 - .0008 - .000 = .069

3.42 Let *A* – watch a TV program oriented to business and financial issues, *B* – read a publication, then the $P(A) = .18$, $P(B) = .12$ and $P(A \cap B) = .10$
a. Find $P(B|A) = P(A \cap B) / P(A) = .10/.18 = .5556$
b. Find $P(A|B) = P(A \cap B) / P(B) = .10/.12 = .8333$

3.44 The number of ways of randomly choosing 2 stocks in order out of 4 is: $P_2^4 = 4!/2! =$ 12, the number of ways of randomly choosing 2 bonds in order from 5: $P_2^5 = 5!/3! =$ 20. Then the probability of choosing either the stocks in order or the bonds in order is the union between the two events which is equal to the sum of the individual probabilities minus the probability of the intersection. = 1/12 + 1/20 – 1/240 = .1292

3.46 Let event A—portfolio management was attended, B – Chartism attended, C – random walk attended, then $P(A) = .4$, $P(B) = .5$, $P(C) = .80$.

a. Find $P(A \cup B) = .4 + .5 - 0 = .9$

b. Find $P(A \cup C)$ if A and C are independent events $= .4 + .8 - .32 = .88$

c. If the $P(C|B) = .75$, then $P(B \cap C) = P(C|B)P(B) = (.75)(.5) = .375$. $P(C \cup B) = P(C) + P(B) - P(C \cap B) = .8 + .5 - .375 = .925$

3.48 Let A – work related problem occurs on Monday and B – work related problem occurs in the last hour of the day's shift, then $P(A) = .3$, $P(B) = .2$ and $P(A \cap B) = .04$, $P(A \cap \bar{B}) = P(A) - P(A \cap B) = .3 - .04 = .26$

a. $P(\bar{B}|A) = P(A \cap \bar{B})/P(A) = .26/.3 = .867$

b. Check if $P(A \cap B) = P(A)P(B)$. Since $.04 \neq .06$, the two events are not independent events.

3.50 Let A – new customer, B – call to a rival service customer, then $P(A) = .15$, $P(B) = .6$ and $P(B|A) = .8$. $P(A|B) = P(A \cap B)/ P(B)$ where $P(A \cap B) = P(B|A)P(A)$. $[(.8)(.15)]/.6 = .2$

3.52 $P(High\ Income \cap Never) = .05$

3.54 $P(Middle\ Income \cap Never) = .05$

3.56 $$P(High\ Income|Never) = \frac{P(High\ Income \cap Never)}{P(Never)} = \frac{.05}{.30} = .1667$$

3.58 $$P(Regular|High\ Income) = \frac{P(Regular \cap High)}{P(High)} = \frac{.10}{.25} = .40$$

3.60 $$\text{Odds} = \frac{.5}{1-.5} = 1 \text{ to } 1 \text{ odds}$$

3.62 P(High scores|>25 hours) = .40, P(Low scores|>25) = .20, $\frac{.4}{.2} = 2.00$. Studying increases the probability of achieving high scores.

3.64 Let F – frequent, I – Infrequent, O – Often, S – Sometimes and N- Never.

a. $P(F \cap O) = .12$

b. $P(F|N) = P(F \cap N)/P(N) = .19 / .27 = .7037$

c. Check if $P(F \cap N) = P(F)P(N)$. Since $.19 \neq .2133$, the two events are not independent.

d. $P(O|I) = P(I \cap O) / P(I) = .07/.21 = .3333$

e. Check if $P(I \cap O) = P(I)P(O)$. Since $.07 \neq .0399$, the two events are not independent.

f. $P(F) = .79$

g. $P(N) = .27$

h. $P(F \cup N) = P(F) + P(N) - P(F \cap N) = .79 + .27 - .19 = .87$

3.66 Let A – Regularly read business section, B – Occasionally, C – Never, TS – Traded stock

a. $P(C) = .25$
b. $P(\text{TS}) = .32$
c. $P(\text{TS}|C) = P(\text{TS}\cap C)/P(C) = .04/.25 = .16$
d. $P(C|\text{TS}) = P(\text{TS}\cap C)/\ P(\text{TS}) = .04/.32 = .125$
e. $P(\text{TS}|\bar{A}) = P(\text{TS}\cap(B\cup C))/P(B\cup C) = (.10 + .04)/(.41+.25) = .2121$

3.68 Let Y – Problems were worked, N – Problems not worked

a. $P(\text{Y}) = .32$
b. $P(A) = .25$
c. $P(A|\text{Y}) = P(A\cap\text{Y})/P(\text{Y}) = .12/.32 = .375$
d. $P(\text{Y}|A) = P(A\cap\text{Y})/P(A) = .12/.25 = .48$
e. $P(C \cup \text{below } C|\text{Y}) = P(C \cup \text{below } C \cap \text{Y})/P(\text{Y}) = (.12 + .02)/.32 = .4375$
f. No, since $P(A\cap\text{Y})$ which is $.12 \neq P(A)P(\text{Y})$ which is .08.

3.70 Let R – Readers, V – Voted in the last election

a. $P(\text{V}) = .76$
b. $P(\text{R}) = .77$
c. $P(\bar{V} \cap \bar{R}) = .1$

3.72 Let G – Growth, H – High, L – Low, S – Within. $P(\text{G}|\text{H}) = .1$, $P(\text{G}|\text{L}) = .8$, $P(\text{G}|\text{S}) = .5$, $P(\text{H}) = .25$, $P(\text{L}) = .15$, $P(\text{S}) = .6$

a. $P(\text{G} \cap \text{H}) = P(\text{G}|\text{H})P(\text{H}) = (.1)(.25) = .025$
b. $P(\text{G}) = P(\text{G} \cap \text{H}) + P(\text{G} \cap \text{L}) + P(\text{G} \cap \text{S}) = (.1)(.25) + (.8)(.15) + (.5)(.6) = .445$
c. $P(\text{L}|\text{G}) = P(\text{G} \cap \text{L}) / P(\text{G}) = (.8)(.15)/.445 = .2697$

3.74 Let T– Top quarter, M – Middle half, B – Bottom quarter. $P(10\%|\text{T}) = .7$, $P(10\%|\text{M}) = .5$, $P(10\%|B) = .2$

a. $P(10\%) = P(10\% \cap \text{T}) + P(10\% \cap \text{M}) + P(10\% \cap B) = (.7)(.25) + (.5)(.5) + (.2)(.25) = .475$
b. $P(\text{T}|10\%) = P(10\% \cap \text{T})/P(10\%) = (.7)(.25)/.475 = .3684$
c. $P(\bar{T}\,|\overline{10\%}) = P(\overline{10\%} \cap \bar{T})/P(\overline{10\%}) = P(\overline{10\% \cup T})\ /\ P(\overline{10\%}) =$

$[1 - P(10\% \cup \text{T})]/P(\overline{10\%}) = [1 - (.475 + .25 - (.7)(.25))]\ /\ .525 = .8571$

3.76 Let M – Faulty machine, I – Impurity

$P(\text{M}) = .4$, $P(\text{I}|\text{M}) = .1$, $P(\text{I}) = P(\text{I} \cap \text{M}) + P(\text{I} \cap \bar{M}) = (.4)(.1) + 0 = .04$ $P(\text{M}|\bar{I}) = P(\bar{I} \cap \text{M})\ /\ P(\bar{I}) = [P(\text{M}) - P(\text{I}\cap\text{M})]\ /\ P(\bar{I}) = (.4 - .04)/.96 = .375$

3.78 $P(A_1) = .4$, $P(B_1|A_1)=.6$, $P(B_1|A_2) = .7$

Complements: $P(A_2) = .6$, $P(B_2|A_1)=.4$, $P(B_2|A_2) = .3$

$$P(A_1 \mid B_1) = \frac{P(B_1 \mid A_1)P(A_1)}{P(B_1 \mid A_1)P(A_1) + P(B_1 \mid A_2)P(A_2)} = \frac{.6(.4)}{.6(.4) + .7(.6)} = .3636$$

3.80 $P(A_1) = .5$, $P(B_1|A_1)=.4$, $P(B_1|A_2) = .7$
Complements: $P(A_2) = .5$, $P(B_2|A_1)=.6$, $P(B_2|A_2) = .3$

$$P(A_1 \mid B_2) = \frac{P(B_2 \mid A_1)P(A_1)}{P(B_2 \mid A_1)P(A_1) + P(B_2 \mid A_2)P(A_2)} = \frac{.6(.5)}{.6(.5)+.3(.5)} = .6667$$

3.82 $P(A_1) = .6$, $P(B_1|A_1)=.6$, $P(B_1|A_2) = .4$
Complements: $P(A_2) = .4$, $P(B_2|A_1)=.4$, $P(B_2|A_2) = .6$

$$P(A_1 \mid B_1) = \frac{P(B_1 \mid A_1)P(A_1)}{P(B_1 \mid A_1)P(A_1) + P(B_1 \mid A_2)P(A_2)} = \frac{.6(.6)}{.6(.6)+.4(.4)} = .6923$$

3.84 E_1: Stock performs much better than the market average
E_2: Stock performs same as the market average
E_3: Stock performs worse than the market average
A: Stock is rated a 'Good Buy'
Given that $P(E_1) = .25$, $P(E_2) = .5$, $P(E_3) = .25$, $P(A|E_1) = .4$, $P(A|E_2) = .2$, $P(A|E_3) = .1$
Then, $P(E_1 \cap A) = P(A|E_1)P(E_1) = (.4)(.25) = .10$
$P(E_2 \cap A) = P(A|E_2)P(E_2) = (.2)(.5) = .10$
$P(E_3 \cap A) = P(A|E_3)P(E_3) = (.1)(.25) = .025$

$$P(E_1 \mid A) = \frac{P(A \mid E_1)P(E_1)}{P(A \mid E_1)P(E_1) + P(A \mid E_2)P(E_2) + P(A \mid E_3)P(E_3)} =$$

$$= \frac{(.40)(.25)}{(.4)(.25)+(.2)(.5)+(.1)(.25)} = .444$$

3.86 A_1: Chickens are purchased from Free Range Farms
A_2: Chickens are purchased from Big Foods Ltd
B_1: Chickens weigh less than 3 pounds
B_2: Chickens weigh more than 3 pounds

The following probabilites are given:
$P(B_1 \mid A_1) = 0.1$, $P(B_1 \mid A_2) = 0.2$, $P(A_1) = 0.4$,

Next, compute the complements of the given probabilities.
$P(B_2 \mid A_1) = 0.9$, $P(B_2 \mid A_2) = 0.8$, $P(A_2) = 0.6$

The probability the chicken came from Free Range Farms is $P(A_1 \mid B_2)$

$$P(A_1 \mid B_2) = \frac{P(B_2 \mid A_1)P(A_1)}{P(B_2 \mid A_1)P(A_1) + P(B_2 \mid A_2)P(A_2)}$$

$$= \frac{(0.9)(0.4)}{(0.9)(0.4)+(0.8)(0.6)} = 0.4286$$

b. As mentioned earlier,
A_1: Chickens are purchased from Free Range Farms
A_2: Chickens are purchased from Big Foods Ltd
$P(A_1) = 0.4$ and $P(A_2) = 0.6$.
P(Out of 5, probability that at least 3 came from Free Range Farms) = $P(A_1 \geq 3)$
$= P(A_1 = 3) + P(A_1 = 4) + P(A_1 = 5)$

P(selecting 3 Chickens are from Free Range Farms)
$= 0.4*0.4*0.4*0.6*0.6 = 0.02304$.
However, there are ${}^5C_3 = 10$ ways to select 3 chickens from 5.
Therefore $P(A_1 = 3) = 10* 0.02304 = 0.2304$.

Similarly, P(selecting 4 Chickens are from Free Range Farms)
$= 0.4*0.4*0.4*0.4*0.6 = 0.01536$.
However, there are ${}^5C_4 = 5$ ways to select 4 chickens from 5
$P(A_1 = 4) = {}^5C_4 * 0.01536 = 0.0768$.

Similarly, $P(A_1 = 5) = {}^5C_5 * 0.01024 = 0.01024$.
Therefore, $P(A_1 \geq 3) = 0.2304 + 0.0768 + 0.01024 = 0.31744$

If you purchase 5 chickens, then the probability that at least 3 came from Free Range Farms is 0.317

3.88 Mutually exclusive events are events such that if one event occurs, the other event cannot occur. For example, a U.S. Senator voting in favor of a tax cut cannot also vote against it. Independent events are events such that the occurrence of one event has no effect on the probability of the other event. For example, whether or not you ate breakfast this morning is unlikely to have any effect on the probability of a U.S. Senator voting in favor of a tax cut.

3.90 Conditional probability is determining the probability of one event given that another event has occurred. This is utilizing prior information that a specific event has occurred and then analyzing how that impacts the probability of another event. The importance is incorporating information about a known event. This has the impact of reducing the sample space in an experiment.

3.92

a. True: By definition, probabilities cannot be negative, hence the union between two events is $P(A) + P(B) - P(A \cap B)$ must be ≥ 0. Adding $P(A \cap B)$ to both sides, it follows that $P(A) + P(B)$ must be $\geq P(A \cap B)$
b. True: this follows from the probability of the union between two events. $P(A) + P(B) - P(A \cap B)$ cannot be larger than $P(A) + P(B)$. Only if the term $P(A \cap B)$ is negative can the statement be false. Since probabilities of events must be between 0 and 1 inclusive, the union between two events cannot be larger than the sum of the individual probabilities.
c. True: events cannot intersect with another event by more than their individual size

d. True: By definition of a complement, if an event occurs, then the event of 'not the event' (its complement) cannot occur.
e. False: Probabilities of any two events can sum to more than 1. Among other conditions, this statement will be true if the two events are mutually exclusive
f. False: there may be more events contained in the sample space. Hence, if two events are mutually exclusive, they may or may not be collectively exhaustive
g. False: the two events could contain common elements and hence their intersection would not be zero

3.94

a. False: Given that event B occurs has changed the sample space, hence, the revised probability may be less
b. False: If the probability of its complement is zero, then the event and its complement will be dependent events
c. True: This follows because the probabilities of events must be non-zero. By the Multiplicative Law of Probability: $P(A|B)P(B) = P(A \cap B)$. Since $0 \leq P(B) \leq 1$, then the $P(A|B) \geq P(A \cap B)$.
d. False: this statement is true only when the two events are independent
e. False: the posterior probability of any event could be smaller, larger, or equal to the prior probability

3.96 Let W – weather condition caused the accident, BI – bodily injuries were incurred. $P(\text{W}) = .3$, $P(\text{BI}) = .2$, $P(\text{W}|\text{BI}) = .4$
a. $P(\text{W} \cap \text{BI}) = P(\text{W}|\text{BI})P(\text{BI}) = (.4)(.2) = .08$
b. No, since $P(\text{W}\cap\text{BI}) = .08 \neq .06 = P(\text{W})P(B)$
c. $P(\text{BI}|\text{W}) = P(\text{W}\cap\text{BI})/P(\text{W}) = .08/.3 = .267$
d. $P(\overline{W} \cap \overline{BI}) = P(\overline{W \cup B}) = 1\text{-}P(\text{W}\cup\text{BI}) = 1 - P(\text{W}) - P(\text{BI}) + P(\text{W}\cap\text{BI}) = 1\text{-}.3\text{-}.2\text{+}.08 = .58$

3.98 Let M – analysts have an MBA, A – analysts are over age 35. $P(M) = .35$, $P(A) = .40$, $P(A|M) = .3$
a. $P(M\cap A) = P(A|M)P(M) = (.3)(.35) = .105$
b. $P(M|A) = P(M\cap A)P(A) = (.105)(.4) = .2625$
c. $P(M\cup A) = P(M) + P(A) - P(M\cap A) = .35 + .4 - .105 = .645$
d. $P(\overline{M} \,|\, \overline{A}) = P(\overline{M} \cap \overline{A})/P(\overline{A}) = [1 - P(M\cup A)]/P(\overline{A}) = .355/.4 = .5917$
e. No, because $P(M \cap A)$ which is $.105 \neq .14$ which is $P(M)P(A)$
f. No, their intersection is not zero, hence the two events cannot be mutually exclusive
g. No, the sum of their individual probabilities is .645 which is less than 1

3.100 Let 160 – farm size exceeds 160 acres, 50 – farm owner is over 50 years old. $P(160) = .2$, $P(50) = .6$, $P(50|160) = .55$

a. $P(160 \cap 50) = P(50|160)P(160) = (.55)(.2) = .11$

b. $P(160 \cup 50) = P(160) + P(50) – P(160 \cap 50) = .2 + .6 - .11 = .69$

c. $P(160|50) = P(160 \cap 50)/P(50) = .11/.6 = .1833$

d. No, since $P(160 \cap 50)$ which is $.11 \neq .12$ which is $P(160)P(50)$

3.102 Let NS – night shift worker, F – women, M – men, FP – favored plan. $P(\text{NS}) = .5$, $P(\text{F}) = .3$, $P(\text{M}) = .7$, $P(\text{F}|\text{NS}) = .2$, $P(\text{M}|\text{NS}) = .8$, $P(\text{FP}|\text{NS}) = .65$, $P(\text{FP}|\text{F}) = .4$. Therefore, $P(\text{NS} \cap \text{F}) = (.5)(.2) = .1$, $P(\text{FP}) = P(\text{FP}|\text{M})P(\text{M}) + P(\text{FP}|\text{F})P(\text{F}) = (.5)(.7) + (.4)(.3) = .47$, $P(\text{NS} \cap \text{FP}) = P(\text{FP}|\text{NS})P(\text{NS}) = (.65)(.5) = .325$

a. $P(\text{FP} \cap \text{F}) = P(\text{FP}|\text{F})P(\text{F}) = (.4)(.3) = .12$

b. $P(\text{NS} \cup \text{F}) = P(\text{NS}) + P(\text{F}) – P(\text{NS} \cap \text{F}) = .5 + .3 - .1 = .7$

c. No, since $P(\text{NS} \cap \text{F})$ which is $.1 \neq .15$ which is $P(\text{NS})P(\text{F})$

d. $P(\text{NS}|\text{F}) = P(\text{NS} \cap \text{F})/P(\text{F}) = .1/.3 = .3333$

e. $P(\overline{NS} \cap \overline{FP}) = 1 – P(\text{NS} \cup \text{FP}) = 1 – P(\text{NS}) – P(\text{FP}) + P(\text{NS} \cap \text{FP}) =$
$1 - .5 - .47 + .325 = .355$

3.104 a. $C_2^{12} = 12!/2!10! = 66$

b. $P(\text{faulty}) = C_1^{11}/66 = 11/66 = .1667$

3.106 Let T – treatment , C – patient was cured. $\text{P(T)} = 10/100 = .1$, $P(\text{C}) = .5$, $P(\text{C}|\text{T}) = .75$

a. $P(\text{C} \cap \text{T}) = P(\text{C}|\text{T})P(\text{T}) = (.75)(.1) = .075$

b. $P(\text{T}|\text{C}) = P(\text{C} \cap \text{T})/P(\text{C}) = .075/.5 = 0.15$

c. The probability will be $1 / [C_{10}^{100} = 1/ 100!/10!90! = 10!90!/100!$

3.108 Let T – passenger is identified by TPS, I – the passenger is carrying an illegal amount of liquor. $P(\text{I}|\text{T}) = P(\text{T} \cap \text{I})/P(\text{T}) = P(\text{T}|\text{I})P(\text{I})/[P(\text{T}|\text{I})P(\text{I}) + P(\text{T}|\bar{I})P(\bar{I})] = (.8)(.2)/[(.8)(.2) + (.2)(.8)] = .5$. There is a 50% chance that the passenger carrying an illegal amount of liquor is identified by TPS.

3.110 Let P – people who already own policies, S – sales made by salesman. $P(\text{S}|\text{P}) = P(\text{S} \cap \text{P})/P(\text{P}) = P(\text{P}|\text{S})P(\text{S})/[P(\text{P}|\text{S})P(\text{S}) + P(\text{P}|\bar{S})P(\bar{S})] = (.7)(.4)/[(.7)(.4) + (.5)(.6)] = .4828$

3.112 a. $P(\text{W}|\text{FW}) = P(\text{W} \cap \text{FW}) / P(\text{FW}) = .149/.29 = .5138$

b. $P(\bar{I}|\text{FI}) = P(\bar{I} \cap \text{FI})/P(\text{FI}) = .181/.391 = .4629$

3.114 a. $P(\text{G}) = P(\text{G}|\text{S})P(\text{S}) + P(\text{G}|\bar{S})P(\bar{S}) = (.7)(.6) + (.2)(.4) = .5$

b. $P(\text{S}|\text{G}) = P(\text{G} \cap \text{S})/P(\text{G}) = .42/.5 = .84$

c. No, since $P(\text{G} \cap \text{S})$ which is $.42 \neq .3$ which is $P(\text{G})P(\text{S})$

d. $P(\text{S} \geq 1) = 1 – P(\text{S} = 0) = 1 - .4^5 = .9898$

3.116 a. $P(\text{P}) = P(\text{P}|\text{HS})P(\text{HS}) + P(\text{P}|\text{C})P(\text{C}) + P(\text{P}|\text{O})P(\text{O}) = (.2)(.3)+(.6)(.5) + (.8)(.2) = .52$
b. $P(\text{HS}|\text{P}) = P(\text{HS}\cap\text{P})/P(\text{P}) = .06/.52 = .1154$

3.118 Let R_1 – Sally is guilty, R_2 – Sally is not guilty. Let G_1 – wearing gloves, G_2 – not wearing gloves. $P(R_1) = 0.50$, $P(R_2) = 0.50$ and $P(G_1|R_1) = .60$, $P(G_1|R_2) = .80$, $P(G_2|R_1) = .40$, $P(G_2|R_2) = .20$

a. $$P(R_1|G_1) = \frac{P(G_1 \mid R_1)P(R_1)}{P(G_1 \mid R_1)P(R_1) + P(G_1 \mid R_2)P(R_2)} = \frac{(.60)(.50)}{(.60)(.50) + (.80)(.50)} = 0.43$$

b. Not likely since wearing gloves reduces her probability

3.120 Let G_1 – Sales will grow, G_2 – Sales will not grow. Let N_1 – new operating system, N_2 – no new operating system. $P(G_1) = 0.70$, $P(G_2) = 0.30$ and $P(N_1|G_1) = .30$, $P(N_1|G_2) = .10$, $P(N_2|G_1) = .70$, $P(N_2|G_2) = .90$

$$P(G_1 \mid N_1) = \frac{P(N_1 \mid G_1)P(G_1)}{P(N_1 \mid G_1)P(G_1) + P(N_1 \mid G_2)P(G_2)} = \frac{(.30)(.70)}{(.30)(.70) + (.10)(.30)} = 0.875$$

3.122 Let $\boldsymbol{P}_1$ – regular plowing, $\boldsymbol{P}_2$ – minimal plowing, $\boldsymbol{H}$ – high yield, $\boldsymbol{L}$ – low yield. $P(\boldsymbol{P}_1) = .4$, $P(\boldsymbol{P}_2) = .6$, $P(\boldsymbol{H}|\boldsymbol{P}_2) = .5$, $P(\boldsymbol{L}|\boldsymbol{P}_1) = .4$

a. $P(H|P_1) = 1 - P(L|P_1) = 1 - .4 = .6$

b. $$P(P_1 \mid H) = \frac{P(H \mid P_1)P(P_1)}{P(H \mid P_1)P(P_1) + P(H \mid P_2)P(P_2)}$$
$$= \frac{(.6)(.4)}{(.6)(.4) + (.5)(.6)} = .4444$$

Chapter 4:

Discrete Random Variables and Probability Distributions

4.2

The number of defective parts produced in daily production is a discrete random variable that can take on no more than a countable number of values.

4.4

Discrete random variable – number of plays is countable.

4.6

Total sales, advertising expenditures, sales of competitors.

4.8

Discrete – the number of purchases is a countable number of values.

4.10

Probability distribution of the face values when a fair die is rolled

x-face values	P(x)
1	.166667
2	.166667
3	.166667
4	.166667
5	.166667
6	.166667

4.12

Various answers

x –# of times missing class	P(x)	F(x)
0	.65	.65
1	.15	.80
2	.10	.90
3	.09	.99
4	.01	1.00

4.14

a. Cumulative probability distribution:

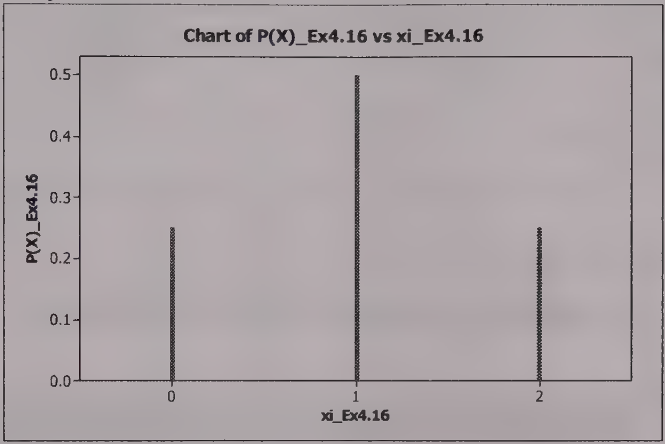
x	0	1	2	3	4	5	6	7	8	9
P(x)	.10	.08	.07	.15	.12	.08	.10	.12	.08	.10
F(x)	.10	.18	.25	.40	.52	.60	.70	.82	.90	1.00

b. $P(x \geq 5) = .08 + .10 + .12 + .08 + .10 = .48$

c. $P(3 \leq x \leq 7) = .15 + .12 + .08 + .10 + .12 = .57$

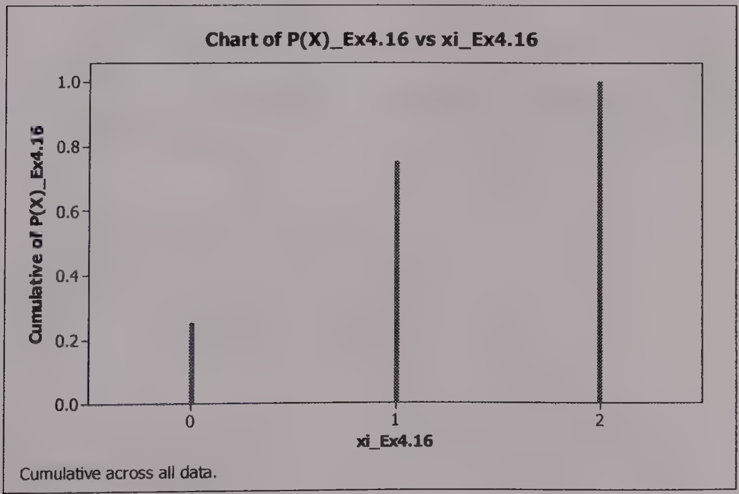
4.16

a. Probability distribution function



b. Cumulative probability distribution function:

x	P(x)	F(x)
0	.25	.25
1	.50	.75
2	.25	1.00



c. The mean of the random variable X

x	P(x)	xP(x)
0	.25	0
1	.50	.50
2	.25	.50
		1.00

$$\mu_x = E(X) = \sum xP(x) = 1.00$$

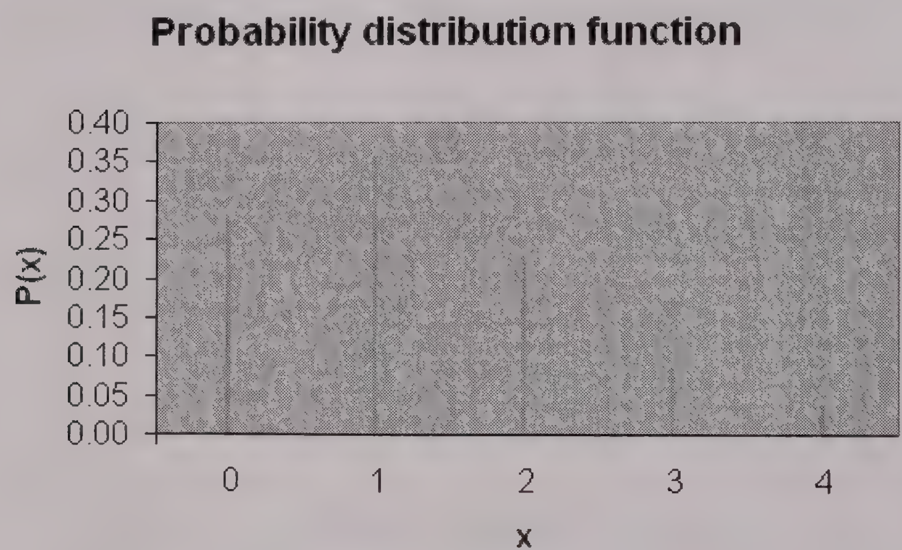
d. The variance of X

x	P(x)	xP(x)	(x-mu)^2	(x-mu)^2P(x)
0	.25	0	1	.25
1	.50	.50	0	0
2	.25	.50	1	.25
		1.00		.50

$$\sigma^2_x = E[(X-\mu_x)^2] = \sum (x-\mu_x)^2 P(x) = .50$$

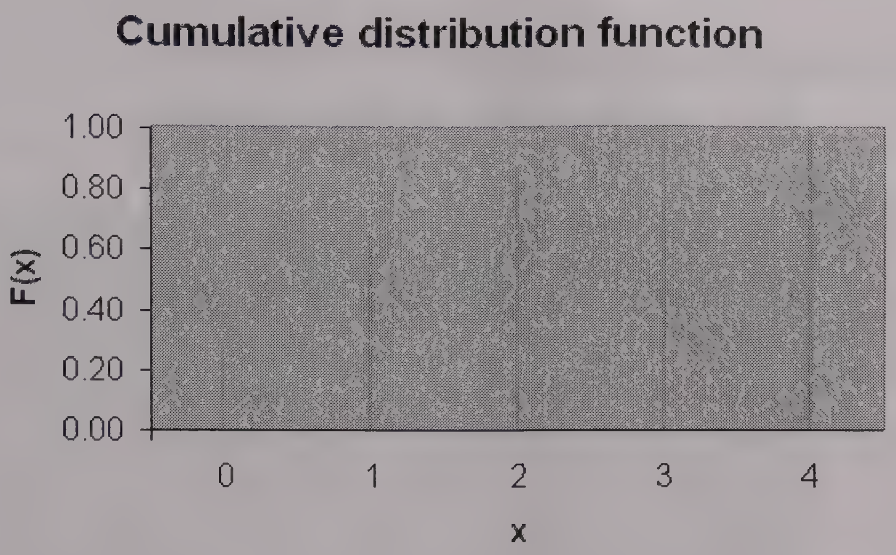
4.18

a. Probability distribution function



b. Cumulative probability distribution function

x	F(x)
0	0.28
1	0.28 + 0.36 = 0.64
2	0.64 + 0.23 = 0.87
3	0.87 + 0.09 = 0.96
4	0.96 + 0.04 = 1.00



c. The mean of the number of returns of an automobile

x	P(x)	xP(x)
0	0.28	0
1	0.36	0.36
2	0.23	0.46
3	0.09	0.27
4	0.04	0.16
		1.25

$$\mu_x = E(X) = \sum xP(x) = 1.25$$

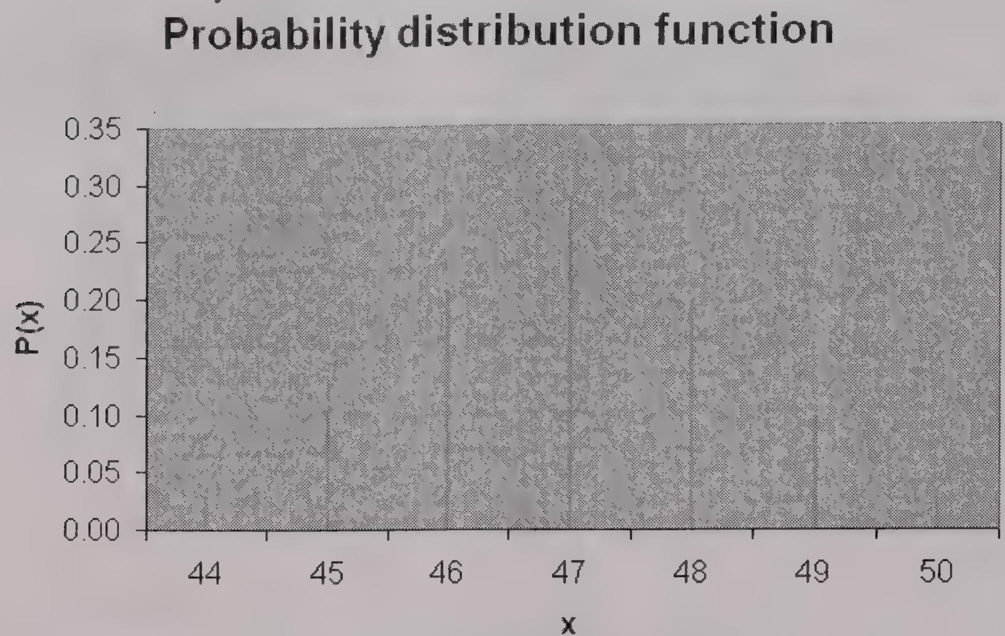
d. The variance of the number of returns of an automobile

x	P(x)	xP(x)	(x-mu)^2	(x-mu)^2P(x)
0	0.28	0	1.5625	0.4375
1	0.36	0.36	0.0625	0.0225
2	0.23	0.46	0.5625	0.129375
3	0.09	0.27	3.0625	0.275625
4	0.04	0.16	7.5625	0.3025
		1.25		1.1675

$$\sigma^2_x = E[(X-\mu_x)^2] = \sum (x-\mu_x)^2 P(x) = 1.1675$$

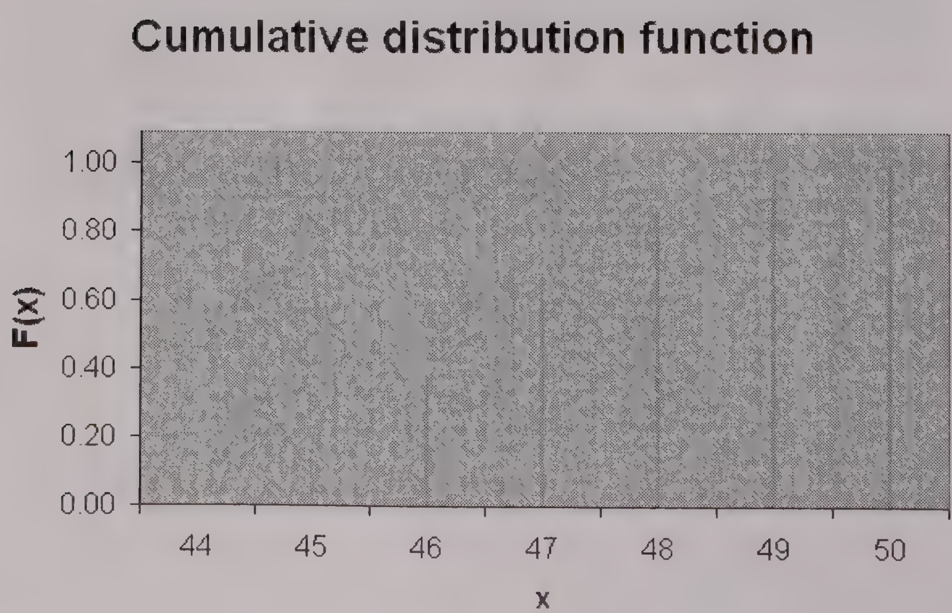
4.20

a. Probability distribution function



b. Cumulative probability distribution function

x	F(x)
44	0.04
45	0.04 + 0.13 = 0.17
46	0.17 + 0.21 = 0.38
47	0.38 + 0.29 = 0.67
48	0.67 + 0.20 = 0.87
49	0.87 + 0.10 = 0.97
50	0.97 + 0.03 = 1.00



c. $P(46 \le X \le 48) = P(X = 46) + P(X = 47) + P(X = 48) = 0.21 + 0.29 + 0.20 = 0.70$

d.

P(at least one package contains at least 47 lbs) $= 1 - P$(both packages contain fewer than 47 lbs) which is $1 - [P(X < 47)]^2 = 1 - [0.38]^2 = 1 - 0.1444 = 0.8556$

e. Excel output:

Microsoft Excel - CH 4.xls

	A	B	C	D	E
1	Weight	P(x)	F(x)	Mean	Variance
2	44	0.04	0.04	1.76	0.3364
3	45	0.13	0.17	5.85	0.4693
4	46	0.21	0.38	9.66	0.1701
5	47	0.29	0.67	13.63	0.0029
6	48	0.20	0.87	9.6	0.242
7	49	0.10	0.97	4.9	0.441
8	50	0.03	1.00	1.5	0.2883
9		1.00		46.9	1.95

$\mu_x = 46.9$ pounds per bag

$\sigma^2_x = 1.95$ and $\sigma_x = 1.3964$ pounds per bag

f. Mean and standard deviation of profit per bag:

Microsoft Excel - CH 4.xls

	J	K	L	M	N	O
1	Weight	Profit Per Bag	P(x)	F(x)	Mean	Variance
2	44	0.87	0.04	0.04	0.0348	0.00013456
3	45	0.85	0.13	0.17	0.1105	0.00018772
4	46	0.83	0.21	0.38	0.1743	0.00006804
5	47	0.81	0.29	0.67	0.2349	0.00000116
6	48	0.79	0.20	0.87	0.1580	0.00009680
7	49	0.77	0.10	0.97	0.0770	0.00017640
8	50	0.75	0.03	1.00	0.0225	0.00011532
9			1.00		0.8120	0.00078000

$\pi = 2.5 - (.75 + .02X)$

$\mu_\pi = E(\pi) = 2.5 - (.75 + (.02)(46.9)) = \$.812$

$\sigma_\pi = |.02|(1.3964) = \$.0279$

4.22

a. Probability distribution function

X	0	1	2
P(x)	0.81	0.18	.01

$P_x(0) = (.90)(.90) = .81$
$P_x(1) = (.90)(.10) + (.10)(.90) = .18$
$P_x(2) = (.10)(.10) = .01$

b. $P(Y = 0) = 18/20 \times 17/19 = 153/190$
$P(Y{=}1) = (2/20 \times 18/19) + (18/20 \times 2/19) = 36/190$
$P(Y{=}2) = 2/20 \times 1/19 = 1/190$

The answer in part b. is different from part a. because in part b. the probability of picking a defective part on the second draw depends upon the result of the first draw.

c. $\mu_x = 0(.81) + .18 + 2(.01) = 0.2$ defect
$\sigma^2{}_x = .22 - (.20)^2 = .18$

d. $\mu_y = 0(153/190) + (36/190) + 2(1/190) = 38/190 = 0.2$ defect
$\sigma^2{}_y = 40/190 - (.20)^2 = .1705$

4.24

"One and one": $E(X) = 1(.75)(.25) + 2(.75)^2 = 1.3125$
"Two-shot foul": $E(X) = 1((.75)(.25) + (.25)(.75)) + 2(.75)^2 = 1.50$
The "two-shot foul" has a higher expected value of number of points.

4.26 $\mu = 3.29$ $\sigma^2 = 1.3259$ $\sigma = 1.1515$

Rating	P(x)	Mean	Variance
1	0.07	0.07	0.367087
2	0.19	0.38	0.316179
3	0.28	0.84	0.023548
4	0.30	1.20	0.15123
5	0.16	0.80	0.467856
	1.00	3.29	1.3259
		S.D.	1.151477

4.28

a. $\mu = 1.82$ breakdowns $\sigma^2 = 1.0276$ $\sigma = 1.0137$ breakdowns

Breakdowns	P(x)	Mean	Variance	Cost	P(x)	Mean	Variance
0	0.1	0	0.33124	0	0.1	0	745290
1	0.26	0.26	0.174824	1500	0.26	390	393354
2	0.42	0.84	0.013608	3000	0.42	1260	30618
3	0.16	0.48	0.222784	4500	0.16	720	501264
4	0.06	0.24	0.285144	6000	0.06	360	641574
	1.00	1.82	1.0276		1.00	2730	2312100
		S.D.	1.013706			S.D.	1520.559

b. Cost: C = 1500X
E(C) = 1500(1.82) = μ = $2,730
σ = |1500|(1.0137) = $1,520.559

4.30 Mean and variance of a Bernoulli random variable with $P = .5$:

$$\mu_x = E(X) = \sum xP(x) = (0)(1-P) + (1)P = P = .5$$
$$\sigma^2_x = E[(X-\mu_x)^2] = \sum (x-\mu_x)^2 P(x) = (0-P)^2(1-P) + (1-P)^2 P = P(1-P)$$
$$\sigma^2_x = P(1-P) = .5(1-.5) = .25$$

4.32 Probability of a binomial random variable with $P = .3$ and $n = 14$, x = 7 and x less than 6

Cumulative Distribution Function
```
Binomial with n = 14 and p = 0.3

x   P( X <= x )
0       0.006782
1       0.047476
2       0.160836
3       0.355167
4       0.584201
5       0.780516
6       0.906718
7       0.968531
8       0.991711
```

$P(X = 7) = .968531 - .906718 = .06181$
$P(X < 6) = P(X \leq 5) = .7805$

4.34 Probability of a binomial random variable with $P = .7$ and $n = 18$, x = 12 and x less than 6

Cumulative Distribution Function
```
Binomial with n = 18 and p = 0.7

 x   P( X <= x )
 0       0.000000
 1       0.000000
 2       0.000000
 3       0.000004
 4       0.000039
 5       0.000269
 6       0.001430
 7       0.006073
 8       0.020968
 9       0.059586
10       0.140683
11       0.278304
12       0.465620
13       0.667345
```

$P(X = 12) = .465620 - .278304 = .1873$
$P(X < 6) = P(X \leq 5) = .000269$

4.36

Cumulative Distribution Function
Binomial with n = 5 and p = 0.250000

x	P(X <= x)
0.00	0.2373
1.00	0.6328
2.00	0.8965
3.00	0.9844
4.00	0.9990
5.00	1.0000

a. $P(X \geq 1) = 1 - P(X = 0) = 1 - .2373 = .7627$
b. $P(X \geq 3) = 1 - P(X \leq 2) = 1 - .8965 = .1035$

4.38

Cumulative Distribution Function
Binomial with n = 7 and p = 0.200000

x	P(X <= x)
0.00	0.2097
1.00	0.5767
2.00	0.8520
3.00	0.9667
4.00	0.9953
5.00	0.9996
6.00	1.0000
7.00	1.0000

$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.5767 = 0.4233$

4.40

Cumulative Distribution Function
Binomial with n = 5 and p = 0.500000

x	P(X <= x)
0.00	0.0312
1.00	0.1875
2.00	0.5000
3.00	0.8125
4.00	0.9688
5.00	1.0000

a. $P(X = 5) = P(X \leq 5) - P(X \leq 4) = 1.0000 - .9688 = .0312$
b. $P(X \geq 3) = P(X \leq 5) - P(X \leq 2) = 1.0000 - .5000 = .5$

Cumulative Distribution Function

```
Binomial with n = 4 and p = 0.500000

       x       P( X <= x)
    0.00         0.0625
    1.00         0.3125
    2.00         0.6875
    3.00         0.9375
    4.00         1.0000
```

c. $P(X \geq 2) = 1 - P(X \leq 1) = 1 - .3125 = .6875$

d. $E(X) = np = 5(.5) = 2.5$ wins

e. $E(X) = \mu = 1 + np = 1 + 4(.5) = 3$ wins

4.42

a.

Cumulative Distribution Function

```
Binomial with n = 4 and p = 0.350000

       x       P( X <= x)
    0.00         0.1785
    1.00         0.5630
    2.00         0.8735
    3.00         0.9850
    4.00         1.0000
```

$P(X \geq 2) = 1 - P(X \leq 1) = 1 - .5630 = .4370$

b. Let $Z = 2X$.

$E(Z) = 2 \times nP = 2 \times 4(0.35) = 2.8$, $\sigma_Z = |2|\sqrt{nP(1-P)} = |2|\sqrt{4(0.35)(0.65)} = 1.908$

c. Set $E(X) = np$ equal to 2.8, where $n = 4$, and solve for P.

$4P = 2.8$, $P = 0.7$

4.44

a. $E(X) = 2000(.032) = 64$

$\sigma_x = \sqrt{2000(.032)(.968)} = 7.871$

b. Let $Z = 10X$

$E(Z) = 10(64) = \$640$

$\sigma_z = |10|(7.871) = \$78.71$

4.46 a. $E(X) = \mu_x = np = 620(.78) = 483.6$, $\sigma_x = \sqrt{620(.78)(.22)} = 10.3146$

b. Let $Z = 2X$

$E(Z) = 2(483.6) = \$967.20$

$\sigma_z = |2|(10.3146) = \$20.6292$

4.48 The acceptance rules have the following probabilities:

(i) Rule 1: $P(X = 0) = (.8)^{10} = .1074$

(ii) Rule 2: $P(X \leq 1) = (.8)^{20} + 20(.2)(.8)^{19} = .0692$

Therefore, the second acceptance rule has the smaller probability of accepting a shipment containing 20% defectives.

4.50

Probability Density Function

Poisson with mean = 2.4

x	P(X = x)
0	0.090718
1	0.217723
2	0.261268
3	0.209014
4	0.125408
5	0.060196
6	0.024078
7	0.008255

$P(X = 4) = .125408$

4.52

Cumulative Distribution Function

Poisson with mean = 3.4

x	P(X <= x)
0	0.033373
1	0.146842
2	0.339740
3	0.558357
4	0.744182
5	0.870542
6	0.942147

$P(X < 6) = .870542$

4.54

Cumulative Distribution Function

Poisson with mu = 3.00000

x	P(X <= x)
0.00	0.0498
1.00	0.1991
2.00	0.4232
3.00	0.6472
4.00	0.8153
5.00	0.9161
6.00	0.9665
7.00	0.9881
8.00	0.9962
9.00	0.9989
10.00	0.9997

$P(X \leq 2) = .4232$

4.56

Cumulative Distribution Function

Poisson with mu = 4.20000

x	P(X <= x)
0.00	0.0150
1.00	0.0780
2.00	0.2102
3.00	0.3954
4.00	0.5898
5.00	0.7531
6.00	0.8675
7.00	0.9361
8.00	0.9721
9.00	0.9889
10.00	0.9959

$P(X \geq 3) = 1 - P(X \leq 2) = 1 - .2102 = .7898$

4.58

$n = 100, p = .045.$

Using the Poisson approximation to the binomial distribution,
$\lambda = 100 \times .045 = 4.5$

Cumulative Distribution Function

Poisson with mu = 4.50000

x	P(X <= x)
0.00	0.0111
1.00	0.0611
2.00	0.1736
3.00	0.3423
4.00	0.5321
5.00	0.7029
6.00	0.8311
7.00	0.9134
8.00	0.9597
9.00	0.9829
10.00	0.9933

$P(X < 3) = P(X \leq 2) = .1736$

4.60

$n = 6000, p = .001.$

Using the Poisson approximation to the binomial distribution,
$\lambda = 6000 \times .001 = 6$

Cumulative Distribution Function
Poisson with mu = 6.00000

x	P(X <= x)
0.00	0.0025
1.00	0.0174
2.00	0.0620
3.00	0.1512
4.00	0.2851
5.00	0.4457
6.00	0.6063
7.00	0.7440
8.00	0.8472
9.00	0.9161
10.00	0.9574

$P(X \geq 3) = 1 - P(X \leq 2) = 1 - .0620 = .9380$

4.62

Two models are possible – the poisson distribution is appropriate when the warehouse is serviced by many thousands of independent truckers where the mean number of 'successes' is relatively small. However, under the assumption of a small fleet of 10 trucks with a probability of any truck arriving during a given hour is .1, then the binomial distribution is the more appropriate model. Both models yield similar, although not identical, probabilities.

Cumulative Distribution Function
Poisson with mean = 1

x	P(X <= x)
0	0.36788
1	0.73576
2	0.91970
3	0.98101
4	0.99634
5	0.99941
6	0.99992
7	0.99999
8	1.00000
9	1.00000
10	1.00000

Cumulative Distribution Function

Binomial with n =`10 and p = 0.1

x	P(X <= x)
0	0.34868
1	0.73610
2	0.92981
3	0.98720
4	0.99837
5	0.99985
6	0.99999
7	1.00000
8	1.00000
9	1.00000
10	1.00000

4.64

Probability Density Function

Hypergeometric with N = 80, M = 42, and n = 20

x	P(X = x)
0	0.000000
1	0.000000
2	0.000008
3	0.000093
4	0.000704
5	0.003723
6	0.014348
7	0.041322
8	0.090392
9	0.151769

$P(X = 9) = .151769$

4.66

Probability Density Function

Hypergeometric with N = 100, M = 50, and n = 15

x	P(X = x)
0	0.000009
1	0.000185
2	0.001716
3	0.009392
4	0.033957
5	0.085911
6	0.157154
7	0.211677
8	0.211677

$P(X = 8) = .211677$

4.68

Cumulative Distribution Function

```
Hypergeometric with N = 16, X = 8, and n = 8

       x     P( X <= x )
    1.00          0.0051
    2.00          0.0660
    3.00          0.3096
    4.00          0.6904
    5.00          0.9340
    6.00          0.9949
    7.00          0.9999
    8.00          1.0000
```

$$P(X = 4) = P(X \le 4) - P(X \le 3) = .6904 - .3096 = .3808$$

4.70

Cumulative Distribution Function

```
Hypergeometric with N = 10, X = 5, and n = 6

       x     P( X <= x )
    0.00          0.0000
    1.00          0.0238
    2.00          0.2619
    3.00          0.7381
    4.00          0.9762
    5.00          1.0000
```

$$P(X < 3) = P(X \le 2) = .2619$$

4.72

a. Marginal probability distributions for X and Y:

Exercise_4.72			X_4.72				
Y_4.72		1	2	P(y)	Mean of Y	Var of Y	StDev of Y
	0	0.25	0.25	0.5	0	0.125	
	1	0.25	0.25	0.5	0.5	0.125	
P(x)		0.5	0.5	1	0.5	0.25	0.5
Mean of X		0.5	1	1.5			
Var of X		0.125	0.125	0.25			
StDev of X				0.5			
xyP(x,y)		0.25	0.5	0.75			
Cov(x,y) =							
$\sum\sum xyP(x,y)-\mu_x\mu_y$		0					

b. The covariance and correlation for X and Y:

$$Cov(X,Y)=\sum_x\sum_y xyP(x,y)-\mu_x\mu_y = .75-(1.5)(.5)=0.0$$

$$\rho = Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_x\sigma_y} = 0.0/(.5)(.5)=0.0$$

Note that when covariance between X and Y is equal to zero, it follows that the correlation between X and Y is also zero.

c. The mean and variance for the linear function W = X + Y:

$$\mu_W = a\mu_x + b\mu_y = 1(1.5)+1(.5)=2.0$$

$$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCov(X,Y) = 1^2(.25)+1^2(.25)+2(1)(1)(0.0)=.50$$

4.74

a. Marginal probability distributions for X and Y:

Exercise_4.74			X_4.74				
Y_4.74		1	2	P(y)	Mean of Y	Var of Y	StDev of Y
	0	0.7	0	0.7	0	0.063	
	1	0	0.3	0.3	0.3	0.147	
P(x)		0.7	0.3	1	0.3	0.21	0.458258
Mean of X		0.7	0.6	1.3			
Var of X		0.063	0.147	0.21			
StDev of X				0.458258			
xyP(x,y)		0	0.6	0.6			
Cov(x,y) =							
$\sum\sum xyP(x,y)-\mu_x\mu_y$		0.21					

b. The covariance and correlation for X and Y:

$$Cov(X,Y)=\sum_x\sum_y xyP(x,y)-\mu_x\mu_y = .60-(1.3)(.3)=.21$$

$$\rho = Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_x\sigma_y} = .21/(.458258)(.458258)=1.00$$

c. The mean and variance for the linear function W = 3X + 4Y:

$$\mu_W = a\mu_x + b\mu_y = 3(1.3)+4(.3)=5.1$$

$$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCov(X,Y) = 3^2(.21)+4^2(.21)+2(3)(4)(.21)=10.29$$

4.76

a. Marginal probability distributions for X and Y:

Exercise_4.76			X_4.76				
Y_4.76		1	2	P(y)	Mean of Y	Var of Y	StDev of Y
	0	0.7	0	0.7	0	0.063	
	1	0	0.3	0.3	0.3	0.147	
P(x)		0.7	0.3	1	0.3	0.21	0.458258
Mean of X		0.7	0.6	1.3			
Var of X		0.063	0.147	0.21			
StDev of X				0.458258			
xyP(x,y)		0	0.6	0.6			
Cov(x,y) =							
$\sum\sum xyP(x,y)-\mu_x\mu_y$		0.21					

b. The covariance and correlation for X and Y:

$$Cov(X,Y)=\sum_x\sum_y xyP(x,y)-\mu_x\mu_y = .60-(1.3)(.3)=.21$$

$$\rho=Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_x\sigma_y} = .21/(.458258)(.458258)=1.00$$

c. The mean and variance for the linear function $W = 10X - 8Y$

$$\mu_W = a\mu_x + b\mu_y = (10)(1.3)+(-8)(.3)=10.6$$

$$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCov(X,Y) = 10^2(.21)+(-8)^2(.21)+2(10)(-8)(.21)=.84$$

4.78 a. Joint cumulative probability function at $X = 1$, $Y = 4$:

$F_{X,Y}(1,4) = .09 + .07 + .14 + .23 = .53$

b. $P_{Y|X}(3|0) = .09/.19 = .4737$

$P_{Y|X}(4|0) = .07/.19 = .3684$

$P_{Y|X}(5|0) = .03/.19 = .1579$

c. $P_{X|Y}(0|4) = .07/.46 = .1522$

$P_{X|Y}(1|4) = .23/.46 = .5$

$P_{X|Y}(2|4) = .16/.46 = .3478$

d. $E(XY) = 0 + 1(3)(.14) + 1(4)(.23) + 1(5)(.10) + 2(3)(.07) + 2(4)(.16) + 2(5)(.11) = 4.64$

$\mu_x = 0 + .47 + 2(.34) = 1.15$

$\mu_y = 3(.3) + 4(.46) + 5(.24) = 3.94$

$Cov(X,Y) = 4.64 - (1.15)(3.94) = .109$

The covariance indicates that there is a positive association between the number of lines in the advertisement and the volume of inquiries.

e. No, because $Cov(X,Y) \neq 0$.

4.80 a. $P(0,0)$=.54, $P(0,1)$=.30, $P(1,0)$=.01, $P(1,1)$=.15

b. $P_{Y|X}(0|1) = 1/16 = .0625$; $P_{Y|X}(1|1) = 15/16 = .9375$

c. E(XY) = .15

$$\mu_x = 0 + 1(.16) = .16$$

$$\mu_y = 0 + 1(.45) = .45$$

$$Cov(X,Y) = .15 - (.16)(.45) = .078$$

The covariance indicates that there is a positive association between regular watchers of a late-night talk show and brand-name recognition.

4.82 Because of independence, the joint probabilities are the products of the marginal probabilities, so $P(0,0)$=.0216, and so on.

				Y Service					
X Food		0	1	2	3	P(x)	Mean of X	Var of X	StDev of X
	0	0.0216	0.0456	0.0408	0.012	0.12	0	0.322752	
	1	0.0522	0.1102	0.0986	0.029	0.29	0.29	0.118784	
	2	0.0756	0.1596	0.1428	0.042	0.42	0.84	0.054432	
	3	0.0306	0.0646	0.0578	0.017	0.17	0.51	0.314432	
P(y)		0.18	0.38	0.34	0.1	1	1.64	0.8104	0.900222
Mean of Y		0	0.38	0.68	0.3	1.36			
Var of Y		0.332928	0.049248	0.139264	0.26896	0.7904			
StDev of Y						0.889044			

4.84

				Y Small						
X Large		0	1	2	3	4	P(x)	Mean of X	Var of X	StDev of X
	0	0.0144	0.0208	0.0288	0.0104	0.0056	0.08	0	0.453152	
	1	0.0288	0.0416	0.0576	0.0208	0.0112	0.16	0.16	0.304704	
	2	0.0504	0.0728	0.1008	0.0364	0.0196	0.28	0.56	0.040432	
	3	0.0576	0.0832	0.1152	0.0416	0.0224	0.32	0.96	0.123008	
	4	0.018	0.026	0.036	0.013	0.007	0.1	0.4	0.26244	
	5	0.0108	0.0156	0.0216	0.0078	0.0042	0.06	0.3	0.411864	
P(y)		0.18	0.26	0.36	0.13	0.07	1	2.38	1.5956	1.263171
Mean of Y		0	0.26	0.72	0.39	0.28	1.65			
Var of Y		0.49005	0.10985	0.0441	0.23693	0.38658	1.2675			
StDev of Y							1.12583302			

$$\mu = 5\mu_x + 10\mu_y = 5(2.38) + 10(1.65) = 28.4$$

$$\sigma = \sqrt{(5^2)\sigma_x^2 + (10^2)\sigma_y^2} = \sqrt{25(1.5965) + 100(1.2675)} = 12.91$$

4.86

Days	P(x)	F(x)	Mean	Variance
1	0.05	0.05	0.05	0.242
2	0.2	0.25	0.4	0.288
3	0.35	0.60	1.05	0.014
4	0.3	0.90	1.2	0.192
5	0.1	1.00	0.5	0.324
Ex 4.86	1.00		3.2	1.06
			S.D.	1.029563

a. $P(X < 3) = .05 + .20 = .25$
b. E(X) = 3.2 days
c. $\sigma = 1.029563$ days
d. Cost = \$20,000 + \$2,000X
E(Cost) = \$20,000 + \$2,000E(X) = \$26,400,
standard deviation = (\$2,000)(1.029563) = \$2,059.13
e. The probability of a project taking at least 4 days to complete is .30 + .10 = .4. Given independence of the individual projects, the probability that at least two of three projects will take at least 4 days to complete is a binomial random variable with $n = 3, p = .4$. $P(2) + P(3) = 3(.4)^2(.6) + (1)(.4)^3(1) = .352$

4.88 a. $\mu = np = 9(.25) = 2.25$
b. $\sigma = \sqrt{np(1-p)} = \sqrt{9(.25)(.75)} = 1.299$
c. (i) E(X) = 1 + 2.25 = 3.25, (ii) $\sigma = 1.299$

4.90 a. $P(4) = (.95)(.90)(.90)(.80) = .6156$
$P(3) = (.05)(.90)(.90)(.8) + 2(.95)(.10)(.90)(.80) + (.95)(.90)(90)(.20) = .3231$
$P(2) = 2(.95)(.90)(.10)(.2) + 2(.05)(.90)(.10)(.80) + (.05)(.90)(.90)(.2) + (.95)(.10)(.10)(.8) = .0571$
$P(1) = (.95)(.10)(.10)(.2) + 2(.05)(.90)(.10)(.20) + (.05)(.10)(.10)(.8) = .0041$
$P(0) = (.05)(.10)(.10)(.20) = .0001$

b. E(X) = .0041 + 2(.0571) + 3(.3231) + 4(.6156) = 3.55 vehicles
c. $\sum x^2 P(x) = 12.99$, $\sigma_x = \sqrt{12.99 - (3.55)^2} = .6225$ vehicles

4.92 Assume that the shots are independent of each other

a. $P(X \ge 2) = 1 - P(X \le 1) = 1 - [\binom{6}{0}(.4)^0(.6)^6 + \binom{6}{1}(.4)(.6)^5] = 0.767$

b. $P(X = 3) = \binom{6}{3}(.4)^3(.6)^3 = 0.2765$

c. $\mu = np = (6)(.4) = 2.4$, $\sigma = \sqrt{6(.4)(.6)} = 1.2$

d. Mean of total points scored = $3(\mu) = 3(2.4) = 7.2$,
Std dev = $3\sigma = 3(1.2) = 3.6$

4.94 a. This is a binomial probability (assuming independence) with a $p = .6$ and $n = 7$. Then the $P(\text{A wins}) = P(X \geq 4) =$

$$\binom{7}{4}(.6)^4(.4)^3 + \binom{7}{5}(.6)^5(.4)^2 + 7(.6^6)(.4) + (.6)^7 = 0.71021$$

b. $\binom{6}{3}(.6)^3(.4)^3 = 0.27648$

c. (i) The outcome of the first four games are known with certainty. Therefore, the series is a best out of three games. To compute the probability that team A wins, find $P(X \geq 2) = 3(.6)^2(.4) + (.6)^3 = 0.648$,

(ii) $\binom{2}{1}(.6)(.4) = 0.48$

4.96 $P(X \geq 2) = 1 - \dfrac{C_0^4 C_4^{16} + C_1^4 C_3^{16}}{C_4^{20}} = 1 - \dfrac{1820 + 2240}{4845} = .16202$

4.98 $P(X > 2) = 1 - e^{-6.5} - e^{-6.5}(6.5) - e^{-6.5}(6.5)^2/2! = 0.95696$

4.100

The mean and variance for the total value of the stock portfolio:

Exercise_4.100			X_4.100						
Y_4.100		40	50	60	70	P(y)	Mean of Y	Var of Y	StDev of Y
	45	0	0	0.05	0.2	0.25	11.25	17.01563	
	50	0.05	0	0.05	0.1	0.2	10	2.1125	
	55	0.1	0.05	0	0.05	0.2	11	0.6125	
	60	0.2	0.1	0.05	0	0.35	21	15.94688	
P(x)		0.35	0.15	0.15	0.35	1	53.25	35.6875	5.973902
Mean of X		14	7.5	9	24.5	55			
Var of X		78.75	3.75	3.75	78.75	165			
StDev of X						12.84523			
xyP(x,y)		800	437.5	465	1172.5	2875			
Cov(x,y) =									
$\sum\sum xyP(x,y)-\mu_x\mu_y$		-53.75							

The mean and variance for the linear function W = aX + bY.

$\mu_W = a\mu_x + b\mu_y = (10)55 + (5)53.25 = \$816.25$

$$\sigma_W^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2abCov(x,y)$$
$$= 10^2(165) + 5^2(35.6875) + 2(10)(5)(-53.75) = \$12017.1875$$

4.102

a. Use the Poisson probability distribution because the random variable is the number of occurrences of a certain event (delivery failures) in a given continuous interval (one day).

b. First, use the given data to determine an estimate for λ, the expected number of failures per day.

$$\lambda = \frac{15+10+...+8+11}{20} = 10.75$$

Cumulative Distribution Function

```
Poisson with mu = 10.7500

      x          P( X <= x)
   1.00            0.0003
   2.00            0.0015
   3.00            0.0059
   4.00            0.0179
   5.00            0.0435
   6.00            0.0895
   7.00            0.1601
   8.00            0.2549
   9.00            0.3682
  10.00            0.4900
```

$$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.3682 = 0.6318$$

c. **Cumulative Distribution Function**

```
Poisson with mu = 10.7500

      x          P( X <= x)
   1.00            0.0003
   2.00            0.0015
   3.00            0.0059
   4.00            0.0179
   5.00            0.0435
   6.00            0.0895
   7.00            0.1601
   8.00            0.2549
   9.00            0.3682
  10.00            0.4900
```

$$P(X < 6) = P(X \leq 5) = 0.0435$$

d. Cumulative Distribution Function

```
Poisson with mu = 10.7500

        x        P( X <= x)
     1.00            0.0003
     2.00            0.0015
     3.00            0.0059
     4.00            0.0179
     5.00            0.0435
     6.00            0.0895
     7.00            0.1601
     8.00            0.2549
     9.00            0.3682
    10.00            0.4900
    11.00            0.6091
    12.00            0.7157
    13.00            0.8039
    14.00            0.8716
    15.00            0.9201
    16.00            0.9527
    17.00            0.9733
    18.00            0.9857
    19.00            0.9926
    20.00            0.9964
```

Use the cumulative distribution function table to construct a table of probabilities for $P(X > x)$.

x	$P(X > x)$
0	1.0000
1	0.9997
2	0.9985
3	0.9941
4	0.9821
5	0.9565
6	0.9105
7	0.8399
8	0.7451
9	0.6318
10	0.5100
11	0.3909
12	0.2843
13	0.1961
14	0.1284
15	0.0799
16	0.0473
17	0.0267
18	0.0143
19	0.0074
20	0.0036

Note that $P(X > 14)$ is greater than 10%, but $P(X > 15)$ is less than 10%. Thus, the number of failures such that the probability of exceeding this number is 10% or less is 15.

4.104

a. Probability distribution with $n = 20$ and $p = 1 - .2 = .8$.

Cumulative Distribution Function

```
Binomial with n = 20 and p = 0.8

 x  P( X <= x )
 0       0.0000
 1       0.0000
 2       0.0000
 3       0.0000
 4       0.0000
 5       0.0000
 6       0.0000
 7       0.0000
 8       0.0001
 9       0.0006
10       0.0026
11       0.0100
12       0.0321
13       0.0867
14       0.1958
15       0.3704
16       0.5886
17       0.7939
18       0.9308
19       0.9885
20       1.0000
```

$P(X \geq 17) = 1 - P(X \leq 16) = 1 - .5886 = .4114$

b. $P(X \geq 19) = 1 - P(X \leq 18) = 1 - .9308 = .0692$

c. Probability distribution with $n = 20$ and $p = 1 - .15 = .85$.

Cumulative Distribution Function

```
Binomial with n = 20 and p = 0.85

 x  P( X <= x )
 0       0.0000
 1       0.0000
 2       0.0000
 3       0.0000
 4       0.0000
 5       0.0000
 6       0.0000
 7       0.0000
 8       0.0000
 9       0.0000
10       0.0002
11       0.0013
12       0.0059
13       0.0219
14       0.0673
15       0.1702
16       0.3523
17       0.5951
18       0.8244
19       0.9612
20       1.0000
```

$P(X \geq 17) = 1 - P(X \leq 16) = 1 - .3523 = .6477$

d. $P(X \geq 19) = 1 - P(X \leq 18) = 1 - .8244 = .1756$

4.106

$n = 1{,}000{,}000$;
99.999% of the computers produced will perform exactly as promised in the descriptive literature. Therefore, $p = 1 - 0.99999 = .00001$.

Using the Poisson approximation to the binomial distribution,
$\lambda = 1{,}000{,}000 \times .00001 = 10$

Cumulative Distribution Function

```
Poisson with mu = 10.0000

       x      P( X <= x )
    0.00         0.0000
    1.00         0.0005
    2.00         0.0028
    3.00         0.0103
    4.00         0.0293
    5.00         0.0671
    6.00         0.1301
    7.00         0.2202
    8.00         0.3328
    9.00         0.4579
   10.00         0.5830
   11.00         0.6968
   12.00         0.7916
   13.00         0.8645
   14.00         0.9165
   15.00         0.9513
   16.00         0.9730
   17.00         0.9857
```

a. $P(X < 5) = P(X \le 4) = .0293$.
The probability of 100% computers produced will perform exactly as promised in the descriptive literature is 0.0293, which is very less.

b. $P(X > 15) = 1 - P(X \le 15) = 1 - .9513 = .0487$.
The probability of 99.999% computers produced will perform exactly as promised in the descriptive literature is $1 - 0.0487 = 0.9513$.

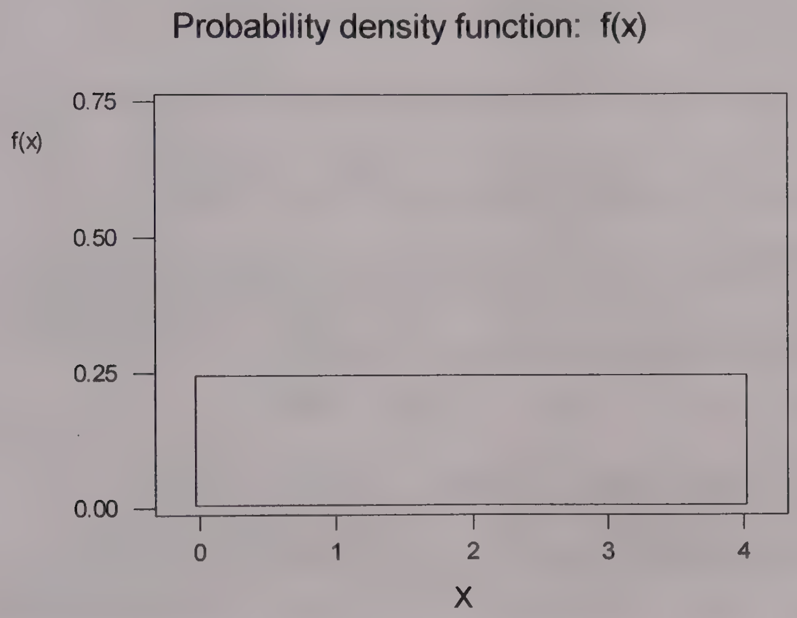
Chapter 5:

Continuous Random Variables and Probability Distributions

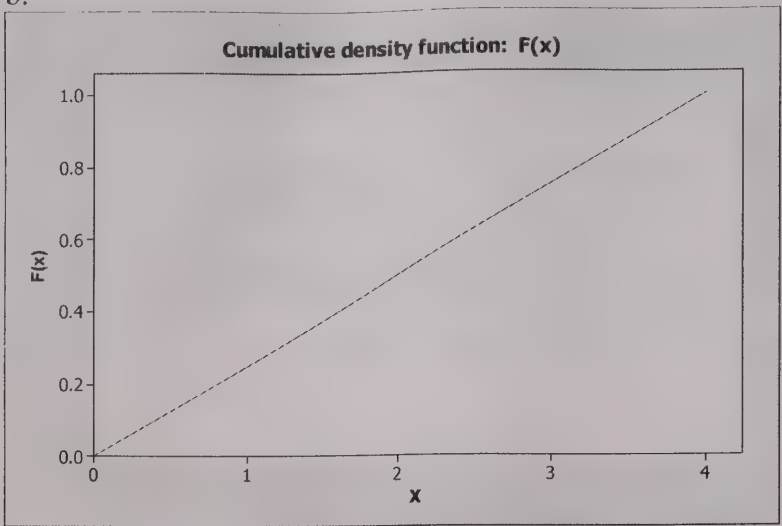
5.2 $P(1.0 < X < 1.9) = F(1.9) - F(1.0) = (.5)(1.9) - (.5)(1.0) = 0.45$

5.4 $P(X > 1.3) = F(1.3) = (.5)(2.0) - (.5)(1.3) = 0.35$

5.6 a.



b.



c. $P(X < 1) = .25$

d. $P(X < .5) + P(X > 3.5) = P(X < .5) + 1 - P(X < 3.5) = .25$

5.8 a. $P(380 < X < 460) = P(X < 460) - P(X < 380) = .6 - .4 = .2$

b. $P(X < 380) < P(X < 400) < P(X < 460)$; $.4 < P(X < 400) < .6$

5.10 $W = a + bX$. If Profit = 1000 - 2X, where X = number of units produced, find the mean and variance of the profit if the mean and variance for the number of units produced are 50 and 90 respectively. $\mu_W = a + b\mu_x =$ 1000 - 2(50) = 900. $\sigma^2_W = b^2\sigma^2_X = (-2)^2(90) = 360$.

5.12 $W = a + bX$. If Profit = 6000 - 3X, where X = number of units produced, find the mean and variance of the profit if the mean and variance for the number of units produced are 1000 and 900 respectively. $\mu_W = a + b\mu_x = 6000 - 3(1000) = 3000$. $\sigma^2_W = b^2\sigma^2_X = (-3)^2(900) = 8100$

5.14 $\mu_Y = 20 + \mu_X = 20 + 4 =$ \$24 million

$Bid = 1.1\ \mu_Y = 1.1(24) =$ \$26.4 million, $\sigma_\pi =$ \$1 million

5.16 $\mu_Y = 6{,}000 + .08\ \mu_X = 6{,}000 + 48{,}000 =$ \$54,000

$\sigma_Y = |.08|\ \sigma_X = .08(180{,}000) =$ \$14,400

5.18 a. Find Z_0 such that $P(Z < Z_0) = .7$, closest value of $Z_0 = .52$

b. Find Z_0 such that $P(Z < Z_0) = .25$, closest value of $Z_0 = -.67$

c. Find Z_0 such that $P(Z > Z_0) = .2$, closest value of $Z_0 = .84$

d. Find Z_0 such that $P(Z > Z_0) = .6$, closest value of $Z_0 = -.25$

5.20 X follows a normal distribution with $\mu = 80$ and $\sigma^2 = 100$

a. Find $P(X > 60)$. $P(Z > \frac{60-80}{10}) = P(Z > -2.00) = .5 + .4772 = .9772$

b. Find $P(72 < X < 82)$. $P(\frac{72-80}{10} < Z < \frac{82-80}{10}) = P(-.80 < Z < .20) =$ $.2881 + .0793 = .3674$

c. Find $P(X < 55)$. $P(Z < \frac{55-80}{10}) = P(Z < -2.50) = .5 - .4938 = .0062$

d. Probability is .1 that X is greater than what number? $Z = 1.28$.
$1.28 = \frac{X-80}{10}$; $X = 92.8$

e. Probability is .6826 that X is in the symmetric interval about the mean between which two numbers? $Z = +/- 1$. $\pm 1 = \frac{X-80}{10}$. $X = 70$ and 90.

5.22 a. $P(Z < \frac{400-380}{50}) = P(Z < .4) = .6554$

b. $P(Z > \frac{360-380}{50}) = P(Z > -.4) = F_Z(.4) = .6554$

c. The graph should show the property of symmetry – the area in the tails equidistant from the mean will be equal.

d. $P(\frac{300-380}{50} < Z < \frac{400-380}{50}) = P(-1.6 < Z < .4) = F_Z(.4) - [1-F_Z(1.6)] =$ $.6554 - .0548 = .6006$

e. The area under the normal curve is equal to .8 for an infinite number of ranges – merely start at a point that is marginally higher. The shortest range will be the one that is centered on the Z of zero. The Z that corresponds to an area of .8 centered on the mean is a Z of ± 1.28. This yields an interval of the mean plus and minus \$64: [\$316, \$444]

5.24 a. $P(Z > \frac{38-35}{4}) = P(Z > .75) = 1 - F_Z(.75) = .2266$

b. $P(Z < \frac{32-35}{4}) = P(Z < -.75) = 1 - F_Z(.75) = .2266$

c. $P(\frac{32-35}{4} < Z < \frac{38-35}{4}) = P(-.75 < Z < .75) = 2F_Z(.75) - 1 = 2(.7734) - 1 =$ $.5468$

d. (i) The graph should show the property of symmetry – the area in the tails equidistant from the mean will be equal.
(ii) The answers to a, b, c sum to one because the events cover the entire area under the normal curve which by definition, must sum to 1.

5.26 a. $P(Z < \frac{10-12.2}{2.8}) = P(Z < -.79) = 1 - F_z(.79) = .2148$

b. $P(Z > \frac{15-12.2}{2.8}) = P(Z > 1) = 1 - F_z(1) = .1587$

c. $P(\frac{12-12.2}{2.8} < Z < \frac{15-12.2}{2.8}) = P(-.07 < Z < 1) = F_z(1) - [1 - F_z(.07)]$
$= .8413 - .4721 = .3692$

d. The answer to a. will be larger because 10 grams is closer to the mean than is 15 grams. Thus, there would be a greater area remaining less than 10 grams than will be the area above 15 grams.

5.28 $P(Z > 1.5) = 1 - F_z(1.5) = .0668$

5.30 $P(Z > .67) = .25, \; .67\sigma = 17.8 - \mu$
$P(Z > 1.04) = .15, \; 1.04\sigma = 19.2 - \mu$
Solving for μ, σ: $\mu = 15.265, \sigma^2 = (3.7838)^2 = 14.317$

5.32 For Investment A, the probability of a return higher than 10%:
$P(Z > \frac{10-10.4}{1.2}) = P(Z > -.33) = F_Z(.33) = .6293$
For Investment B, the probability of a return higher than 10%
$P(Z > \frac{10-11.0}{4}) = P(Z > -.25) = F_Z(.25) = .5987$
Therefore, Investment A is a better choice

5.34 a. $P(Z > -1.28) = .9, \; -1.28 = \frac{Xi-150}{40}$, Xi = 98.8

b. $P(Z < .84) = .8, .84 = \frac{Xi-150}{40}$, Xi = 183.6

c. $P(X \geq 1) = 1 - P(X = 0) = 1 - [P(Z < \frac{120-150}{40})]^2 = 1 - [P(Z < -.75)]^2 =$
$1 - (.2266)^2 = .9487$

5.36 a. $P(\frac{400-420}{80} < Z < \frac{480-420}{80}) = P(-.25 < Z < .75) = F_z(.75) - [1 - F_Z$
$(.25)] = .7734 - .4013 = .3721$

b. $P(Z > 1.28) = .1, \; 1.28 = \frac{Xi-420}{80}, Xi = 522.4$

c. 400 – 439
d. 520 – 559

e. $P(X \geq 1) = 1 - P(X = 0) = 1 - [P(Z < \frac{500-420}{80})]^2 = 1 - (.8413)^2 = .2922$

5.38 $P(Z < 1.5) = .9332,\ 1.5 = \dfrac{85-70}{\sigma},\ \sigma = 10$

$P(Z > \dfrac{80-70}{10}) = P(Z > 1) = .1587$

$P(X \geq 1) = 1 - P(X{=}0) = 1 - [F_Z(1)]^4 = 1 - (.8413)^4 = .4990$

5.40 $n = 1600$ from a binomial probability distribution with $P = .40$

a. Find $P(X > 1650)$. $E[X] = \mu = 1600(.4) = 640$, $\sigma = \sqrt{(1600)(.4)(.6)} = 19.5959$ $P(Z > \dfrac{1650-640}{19.5959}) = P(Z > 51.5414) = 1 - F_Z(51.5414) = 0$

b. Find $P(X < 1530)$. $P(Z < \dfrac{1530-640}{19.5959}) = P(Z{<}45.4177) = 1$

c. $P(\dfrac{1550-640}{19.5959} < Z < \dfrac{1650-640}{19.5959}) = P(46.4383 < Z < 51.5414) = =1\text{-}1\text{=}0$

d. Probability is .09 that the number of successes is less than how many? Z = -1.34. $-1.34 = \dfrac{X-640}{19.5959}$ X = 613.74≈614 successes

e. Probability is .20 the number of successes is greater than? Z = .84. $.84 = \dfrac{X-640}{19.5959}$. X = 656.46≈656successes

5.42 $n = 1600$ from a binomial probability distribution with $P = .40$

a. Find $P(P > .45)$. $E[P] = \mu = P = .40$, $\sigma = \sqrt{\dfrac{P(1-P)}{n}} = \sqrt{\dfrac{.4(1-.4)}{1600}} = .01225$ $P(Z > \dfrac{.45-.40}{.01225}) = P(Z > 4.082) = 1 - F_Z(4.082) = .0000$

b. Find $P(P < .35)$. $P(Z < \dfrac{.35-.40}{.01225}) = P(Z < -4.08) = 1 - F_Z(4.08) = .0000$

c. $P(\dfrac{.44-.40}{.01225} < Z < \dfrac{.37-.40}{.01225}) = P(3.27 < Z < -2.45) = 1 - [(2)[1\text{-}F_z(3.27)]] = 1 - (2)[1 - .9995] = .9995 - .0071 = .9924$

d. Probability is .20 that the percentage of successes is less than what percent? Z = -.84.

$$-.84 = \frac{X - .40}{.01225} \quad P = 38.971\%$$

e. Probability is .09 the percentage of successes is greater than? Z = 1.34. $1.34 = \dfrac{X - .40}{.01225}$. P = 41.642%

5.44 a. $E[X] = \mu = 900(.2) = 180,\ \sigma = \sqrt{(900)(.2)(.8)} = 12$

$$P(Z > \frac{200-180}{12}) = P(Z > 1.67) = 1 - F_Z(1.67) = .0475$$

b. $P(Z < \dfrac{175-180}{12}) = P(Z < -.42) = 1 - F_Z(.42) = .3372$

5.46 $E[X] = (100)(.6) = 60,\ \sigma = \sqrt{(100)(.6)(.4)} = 4.899$

$$P(Z < \frac{50-60}{4.899}) = P(Z < -2.04) = 1 - F_Z(2.04) = 1 - .9793 = .0207$$

5.48 $P(Z > \dfrac{38-35}{4}) = P(Z > .75) = 1 - F_Z(.75) = 1 - .7734 = .2266$

$E[X] = 100(.2266) = 22.66,\ \sigma = \sqrt{(100)(.2266)(.7734)} = 4.1863$

$$P(Z > \frac{25-22.66}{4.1863}) = P(Z > .56) = 1 - F_Z(.56) = 1 - .7123 = .2877$$

$$P(Z > 1.71) = 1 - F_Z(1.71) = 1 - .9564 = .0436$$

5.50 $\lambda = 1.0$, what is the probability that an arrival occurs in the first t=2 time units?

Cumulative Distribution Function

```
Exponential with mean = 1
x   P( X <= x )
0      0.000000
1      0.632121
2      0.864665
3      0.950213
4      0.981684
5      0.993262
```

$P(T < 2) = .864665$

5.52 $\lambda = 5.0$, what is the probability that an arrival occurs after t=7 time units?

Cumulative Distribution Function

```
Exponential with mean = 5
x   P( X <= x )
0       0.000000
1       0.181269
2       0.329680
3       0.451188
4       0.550671
5       0.632121
6       0.698806
7       0.753403
8       0.798103
```

P(T>7) = 1-[P(T ≤7)] = 1 - .7534 =.2466

5.54 $\lambda = 3.0$, what is the probability that an arrival occurs after t=2 time units?

Cumulative Distribution Function

```
Exponential with mean = 3
x   P( X <= x )
0       0.000000
1       0.283469
2       0.486583
3       0.632121
```

$P(T<2) = .4866$

5.56 $P(X > 18) = e^{-(18/15)} = .3012$

5.58 a. $P(X > 3) = 1 - [1 - e^{-(3/\mu)}] = e^{-3\lambda}$ since $\lambda = 1/\mu$

b. $P(X > 6) = 1 - [1 - e^{-(6/\mu)}] = e^{-(6/\mu)} = e^{-6\lambda}$

c. $P(X>6|X>3) = P(X > 6)/P(X > 3) = e^{-6\lambda} / e^{-3\lambda}] = e^{-3\lambda}$

The probability of an occurrence within a specified time in the future is not related to how much time has passed since the most recent occurrence.

5.60 Let $\lambda = 20 \text{ trucks} / 60 \text{ minutes} = 1 \text{ truck} / 3 \text{ minutes}$.

a. $P(t \geq 5) = 1 - P(t < 5) = 1 - [1 - e^{-(1/3)(5)}] = 0.1889$

b. $P(t \leq 1) = 1 - e^{-(1/3)(2)} = 0.4866$

c. $P(4 \leq t \leq 10) = [1 - e^{-(1/3)(10)}] - [1 - e^{-(1/3)(4)}] = 0.2279$

5.62 Find the mean and variance of the random variable: W = 5X + 4Y with correlation = -.5

$\mu_W = a\mu_x + b\mu_y = 5(100) + 4(200) = 1300$

$\sigma^2{}_W = a^2\sigma^2{}_X + b^2\sigma^2{}_Y + 2abCorr(X,Y)\sigma_X\sigma_Y$

$= 5^2(100) + 4^2(400) + 2(5)(4)(-.5)(10)(20) = 4{,}900$

5.64 Find the mean and variance of the random variable: W = 5X – 4Y with correlation = .5.

$$\mu_W = a\mu_x - b\mu_y = 5(500) - 4(200) = 1700$$

$$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y - 2abCorr(X,Y)\sigma_X\sigma_Y$$
$$= 5^2(100) + 4^2(400) - 2(5)(4)(.5)(10)(20) = 4900$$

5.66 $\mu_Z = 100{,}000(.1) + 100{,}000(.18)$. $\mu_Z = 10{,}000 + 18{,}000 = 28{,}000$

$\sigma_Z = 0$. Note that the first investment yields a certain profit of 10% which is a zero standard deviation. $\sigma_X = 100{,}000(.06) = 6{,}000$

5.68 $\mu_Z = \mu_1 + \mu_2 + \mu_3 = 50{,}000 + 72{,}000 + 40{,}000 = 162{,}000$

$$\sigma_Z = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \sqrt{(10{,}000)^2 + (12{,}000)^2 + (9{,}000)^2} = 18{,}027.76$$

5.70 The calculation of the mean is correct, but the standard deviations of two random variables cannot be summed. To get the correct standard deviation, add the variances together and then take the square root. The standard deviation: $\sigma_z = \sqrt{5(15)^2} = 33.5410$.

5.72 a. Compute the mean and variance of the portfolio with correlation of +.5

$$\mu_W = a\mu_x + b\mu_y = 50(25) + 40(40) = 2850$$

$$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCorr(X,Y)\sigma_X\sigma_Y$$
$$= 50^2(121) + 40^2(225) + 2(50)(40)(.5)(11)(15) = 992{,}500$$

b. Recompute with correlation of -.5

$$\mu_W = a\mu_x + b\mu_y = 50(25) + 40(40) = 2850$$

$$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCorr(X,Y)\sigma_X\sigma_Y$$
$$= 50^2(121) + 40^2(225) + 2(50)(40)(-.5)(11)(15) = 332{,}500$$

5.74 a. W = aX – bY = 10X – 10Y

$$\mu_W = a\mu_x - b\mu_y = 10(100) - 10(90) = 100$$

$$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y - 2abCorr(X,Y)\sigma_X\sigma_Y$$

$=10^2(100) + 10^2(400) - 2(10)(10)(-.4)(10)(20) = 66{,}000$ $\sigma_W = \sqrt{66{,}000}$ $= 256.90465$

b. $P(Z < \dfrac{0-100}{256.90465}) = P(Z < -.39) = 1 - F_Z(.39) = 1 - .6517 = .3483$

5.76 a. W = aX – bY = 1X – 1Y

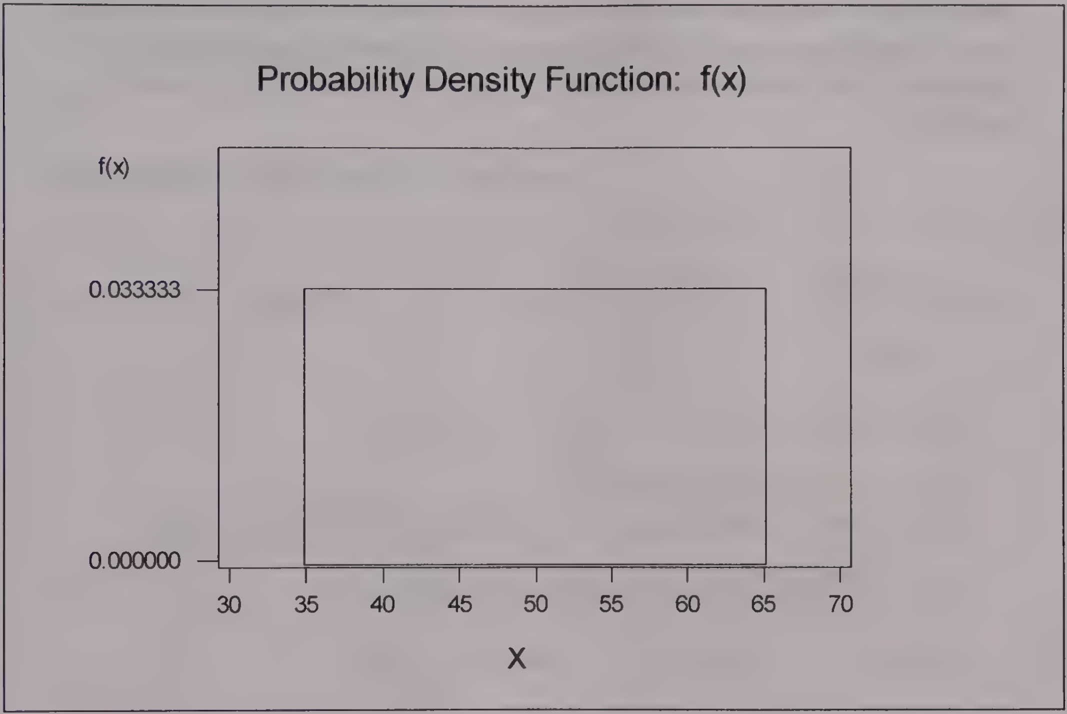
$\mu_W = a\mu_x - b\mu_y$ = 1(100) – 1(105) = -5

$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y - 2abCorr(X,Y)\sigma_X\sigma_Y$

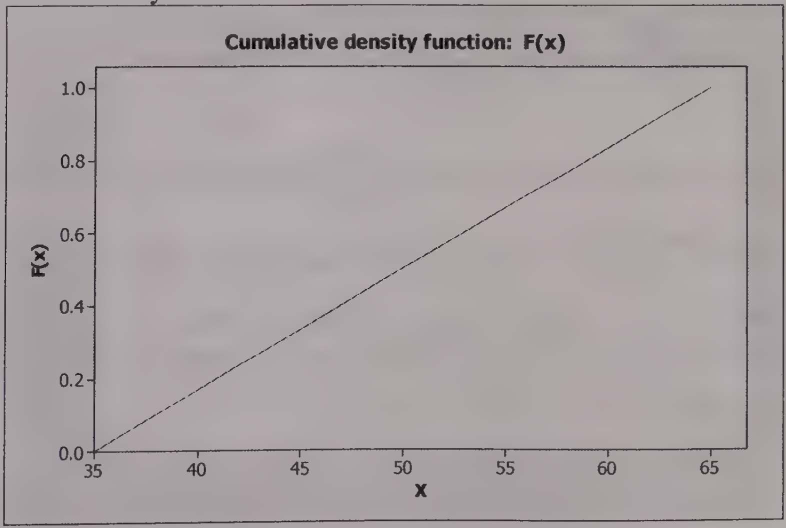
$=1^2(900) + 1^2(625) - 2(1)(1)(.7)(30)(25) = 475$ $\sigma_W = \sqrt{475}$ =21.79449

b. $P(Z > \frac{0-(-5)}{21.79449}) = P(Z > .23) = 1 - F_Z(.23) = 1 - .5910 = .4090$

5.78 a.



b. Cumulative density function



c. $P(40 < X < 50) = (50/30) - (40/30) = 10/30$

d. $E[X] = \dfrac{65+35}{2} = 50$

5.80 a. $\mu_Y = 2000(0.1) + 1000(1 + \mu_x) = 200 + 1{,}160 = 1{,}360$

b. $\sigma_Y = |1000|\, \sigma_Y = 1000(.08) = 80$

5.82 Given that the variance of both predicted earnings and forecast error are both positive and given that the variance of actual earnings is equal to the sum of the variances of predicted earnings and forecast error, then the variance of predicted earnings must be less than the variance of actual earnings

5.84 a. $P(Z > \dfrac{3-2.6}{.5}) = P(Z > .8) = 1 - F_Z(.8) = .2119$

b. $P(\dfrac{2.25-2.6}{.5} < Z < \dfrac{2.75-2.6}{.5}) = P(-.7 < Z < .3) = F_z(.3) - [1 - F_Z(.3)]$
$= .3759$

c. $P(Z > 1.28) = .1,\ 1.28 = \dfrac{Xi-2.6}{.5},\ Xi = 3.24$

d. $P(Xi > 3) = .2119$ (from part a)
$E[X] = 400(.2119) = 84.76,\ \sigma_X = \sqrt{(400)(.2119)(.7881)} = 8.173$
$P(Z > \dfrac{80-84.76}{8.173}) = P(Z > -.58) = F_Z(.58) = .7190$

e. $P(X \geq 1) = 1 - P(X = 0) = 1 - (.7881)^2 = .3789$

5.86 a. $P(Z < \dfrac{85-100}{30}) = P(Z < -.5) = .3085$

b. $P(\dfrac{70-100}{30} < Z < \dfrac{130-100}{30}) = P(-1 < Z < 1) = 2\,F_z(1) - 1 = .6826$

c. $P(Z > 1.645) = .05,\ 1.645 = \dfrac{Xi-100}{30}$, Xi = 149.35

d. $P(Z > \dfrac{60-100}{30}) = P(Z > -1.33) = F_Z(1.33) = .9032$
$P(X \geq 1) = 1 - P(X = 0) = 1 - (.0918)^2 = .9916$

e. Use the binomial formula: $P(X = 2) = C_2^4(.9082)^2(.0918)^2 = 0.0417$

f. 90 – 109

g. 130 - 149

5.88 $P(Z > 1.28) = .1$, $1.28 = \dfrac{130-100}{\sigma}$, $\sigma = 23.4375$

$P(Z > \dfrac{140-100}{23.4375}) = P(Z > 1.71) = 1 - F_Z(1.71) = .0436$

4.36% of members spend more than \$140 in a year.

5.90 $E[X] = 1000(.4) = 400$, $\sigma_X = \sqrt{(1000)(.4)(.6)} = 15.4919$

$P(Z < \dfrac{500-400}{15.4919}) = P(Z < 6.45) \approx 1.0000$

5.92 The number of calls per 12-hour time period follows a Poisson distribution with $\lambda = 15 \text{ calls/12-hour}$ time period.

Cumulative Distribution Function

```
Poisson with mu = 15.0000

     x        P( X <= x)
   0.00          0.0000
   1.00          0.0000
   2.00          0.0000
   3.00          0.0002
   4.00          0.0009
   5.00          0.0028
   6.00          0.0076
   7.00          0.0180
   8.00          0.0374
   9.00          0.0699
  10.00          0.1185
  11.00          0.1848
  12.00          0.2676
  13.00          0.3632
  14.00          0.4657
  15.00          0.5681
  16.00          0.6641
  17.00          0.7489
```

$P(x < 10) = P(x \le 9) = 0.0699$

$P(x > 17) = 1 - P(x \le 17) = 1 - 0.7489 = 0.2511$

5.94 a. $E[X] = 600(.4) = 240$, $\sigma_X = \sqrt{(600)(.4)(.6)} = 12$

$P(Z > \dfrac{260-240}{12}) = P(Z > 1.67) = 1 - F_Z(1.67) = .0475$

b. $P(Z > -.254) = .6$, $-.254 = \dfrac{Xi - 240}{12}$, $Xi = 236.95$ (237 listeners)

5.96 $P(Z>1.28)=.1$, $1.28=\frac{3.5-2.4}{\sigma}$, $\sigma=.8594$. Probability that 1 executive spends 3+ hours on task: $P(Z>\frac{3-2.4}{.8594}) = P(Z>.7) = 1 - F_Z(.7) = 0.3$

$E[X] = 400(0.3)=120$, $\sigma_X = \sqrt{(400)(.3)(.7)} = 9.165$; $P(Z>\frac{80-120}{9.165}) =$ $P(Z>-4.36) = F_Z(4.36) = 1$

5.98 Portfolio consists of 10 shares of stock A and 8 shares of stock B

a. Find the mean and variance of the portfolio value: W = 10X + 8Y with correlation of .3.

$\mu_W = a\mu_x + b\mu_y = 10(12) + 8(10) = 200$

$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCorr(X,Y)\sigma_X\sigma_Y$

$= 10^2(14) + 8^2(12) + 2(10)(8)(.5)(3.74166)(3.4641) = 3{,}204.919$

b. Option 1: Stock 1 with mean of 12, variance of 25, correlation of -.2.

$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCorr(X,Y)\sigma_X\sigma_Y$

$= 10^2(25) + 8^2(12) + 2(10)(8)(-.2)(5)(3.4641) = 2713.744$

Option 2: Stock 2 with mean of 10, variance of 9, correlation of .6.

$= 10^2(9) + 8^2(12) + 2(10)(8)(0.5\ )(3)(3.4641) = 2499.384$

To reduce the variance of the portfolio, select Option 2

5.100 a. $\mu_W = a\mu_x + b\mu_y = 1(40) + 1(35) = 75$

$\sigma^2_W = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCorr(X,Y)\sigma_X\sigma_Y$

$= 1^2(100) + 1^2(144) + 2(1)(1)(.5)(10)(12) = 364$

$\sigma_W = \sqrt{364} = 19.07878$

Probability that all seats are filled:

$\frac{100-75}{19.07878} = 1.31$ Fz =0.9049. $1 - 0.9049 = 0.0951$

b. Probability that between 75 and 90 seats will be filled:

$\frac{90-75}{19.07878} = .79$ $.5 - Fz(.79) = 0.2852$

5.102: Results are obtained using Minitab.
Mean and variance for stock prices:

	AB Volvo (ADR)	Alcoa Inc.	Pentair Inc.	TCF Financial Corporation
Mean	8.6143	31.9829	28.9543	25.1643
Variance	25.4171	27.4188	95.4157	20.4361

Covariances:

	AB Volvo (ADR)	Alcoa Inc.	Pentair Inc.
Alcoa Inc.	6.5180		
Pentair Inc.	31.2128	5.4712	
TCF Financial Corporation	-4.3594	-2.7947	20.6897

Let the total value of the portfolio be represented by variable W.

μ_W = (0.3333)(8.6143) + (0.1667)(31.9829) + (0.3333)(28.9543) + (0.1667)(25.1643) = 22.05

σ_W^2 = $(0.3333^2)(25.4171) + (0.1667^2)(27.4188) + (0.3333^2)(95.4157) + (0.1667^2)(20.4361) + 2[(0.3333)(0.1667)(6.5180) + (0.3333)(0.3333)(31.2128) + (0.3333)(0.1667)(-4.3594) + (0.1667)(0.3333)(5.4712) + (0.1667)(0.1667)(-2.7947) + (0.3333)(0.1667)(20.6897)] = 24.68$

We can confirm these results by finding the portfolio price for each year, shown next, and then by finding the mean and variance of these prices.

Portfolio Price
26.1833
24.6217
24.0983
27.7533
20.2700
14.3017
17.1033

Descriptive Statistics: Portfolio Price

```
Variable          N  N*    Mean  SE Mean  StDev  Variance  Minimum     Q1  Median
Portfolio Price   7   0   22.05     1.88   4.97     24.68    14.30  17.10   24.10

Variable             Q3  Maximum
Portfolio Price   26.18    27.75
```

The previous output confirms that $\mu_W = 22.05$ and $\sigma_W^2 = 24.68$.

Assuming that the portfolio price is normally distributed, the narrowest interval that contains 95% of the distribution of portfolio value is centered at the mean. Therefore, it is $\mu_W \pm z_{\alpha/2}\sigma_W$. Using $z_{\alpha/2} = 1.96$ and $\sigma_W = 4.97$, the interval is $22.05 \pm (1.96)(4.97)$ or $(12.31, 31.79)$.

5.104: Results are obtained using Minitab.
Mean and variance for stock price growth:

	3M Company	Alcoa Inc.	Intel Corporation	Potlatch Corporation	General Motors Corporation	Sea Containers Ltd.
Mean	0.001992	0.004389	-0.000082	0.007449	-0.014355	-0.146323
Variance	0.002704	0.005060	0.006727	0.006674	0.014518	0.176663

Covariances:

	3M Company	Alcoa Inc.	Intel Corporation	Potlatch Corporation	General Motors Corporation
Alcoa Inc.	0.00153782				
Intel Corporation	0.00163165	0.00184360			
Potlatch Corporation	0.00012217	0.00197600	0.00144736		
General Motors Corporation	-0.00005101	0.00103371	-0.00006588	0.00246545	
Sea Containers Ltd.	0.00075015	0.00706908	-0.00131221	-0.00151704	0.0107742

Let the portfolio growth be represented by variable W. The mean and variance for this portfolio, μ_W and σ_W^2, can be found using the following equations or by using technology.

$$\mu_W = \sum_{i=1}^{k} a_i\mu_i, \quad \sigma_W^2 = \sum_{i=1}^{k} a_i^2\sigma_i^2 + 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} a_i a_j \, Cov(X_i, X_j)$$

Descriptive Statistics: Portfolio Growth

```
Variable           N  N*     Mean  SE Mean   StDev  Variance  Minimum       Q1
Portfolio Growth  60   0  -0.0245   0.0111  0.0862    0.0074  -0.4182  -0.0688

Variable          Median      Q3  Maximum
Portfolio Growth  -0.0062  0.0303   0.1212
```

As previously shown, μ_W = -0.0245 and $\sigma_W^2 = 0.0074$.

5.106: Results are obtained using Minitab.
Mean and variance for stock price growth:

	AB Volvo	Pentair Inc.	Reliant Energy Inc.	TCF Financial Corporation	3M Company	Restoration Hardware Inc.
Mean	0.019592	0.007641	0.019031	-0.004087	0.001992	-0.013406
Variance	0.004805	0.006227	0.012686	0.004001	0.002704	0.027618

Covariances:

	AB Volvo	Pentair Inc.	Reliant Energy Inc.	TCF Financial Corporation	3M Company
Pentair Inc.	0.00074848				
Reliant Energy Inc.	0.00228027	0.00105381			
TCF Financial Corporation	-0.00001514	-0.00021080	-0.00041228		
3M Company	0.00099279	0.00087718	0.00031032	0.00072435	
Restoration Hardware Inc.	0.00117969	0.00169410	0.00055922	-0.00041072	0.00204408

Let the portfolio growth be represented by variable W. The mean and variance for this portfolio, μ_W and σ_W^2, can be found using the following equations or by using technology.

$$\mu_W = \sum_{i=1}^{k} a_i\mu_i \, , \; \sigma_W^2 = \sum_{i=1}^{k} a_i^2\sigma_i^2 + 2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} a_i a_j \, Cov(X_i, X_j)$$

Descriptive Statistics: Portfolio Growth

```
Variable           N  N*     Mean  SE Mean    StDev  Variance   Minimum
Portfolio Growth  60   0  0.00513  0.00612  0.04740   0.00225  -0.16714

Variable                Q1   Median       Q3  Maximum
Portfolio Growth  -0.02762  0.00631  0.04184  0.10438
```

As previously shown, $\mu_W = 0.00513$ and $\sigma_W^2 = 0.00225$.

For the second portfolio (20% AB Volvo, 30% Pentair, 30% Reliant Energy, and 20% 3M Company), we get the following output:

Descriptive Statistics: Portfolio Growth

```
Variable           N  N*     Mean  SE Mean    StDev  Variance   Minimum
Portfolio Growth  60   0  0.01232  0.00680  0.05270   0.00278  -0.15522

Variable                Q1   Median       Q3  Maximum
Portfolio Growth  -0.02121  0.01386  0.05357  0.10539
```

For the second portfolio, as previously shown, $\mu_W = 0.01232$ and $\sigma_W^2 = 0.00278$.

The second portfolio has a higher mean and a higher variance. Recall that risk is directly related to variance. Since the second portfolio has a significantly larger mean and only a slightly larger variance, it would be the better choice.

Chapter 6:

Sampling and Sampling Distributions

6.2 The sampling distribution of the sample means for flipping of the coins can be generated by use of the binomial formula.

a. Binomial random variable with $n = 2, p = .5$

Probability Density Function

```
Binomial with n = 2 and p = 0.5

x  P( X = x )
0        0.25
1        0.50
2        0.25
```

b. Binomial random variable with $n = 4, p = .5$

Probability Density Function

```
Binomial with n = 4 and p = 0.5

x  P( X = x )
0      0.0625
1      0.2500
2      0.3750
3      0.2500
4      0.0625
```

c. Binomial random variable with $n = 10, p = .5$

Probability Density Function

```
Binomial with n = 10 and p = 0.5

 x  P( X = x )
 0    0.000977
 1    0.009766
 2    0.043945
 3    0.117188
 4    0.205078
 5    0.246094
 6    0.205078
 7    0.117188
 8    0.043945
 9    0.009766
10    0.000977
```

6.4 The response should note that there will be errors in taking a census of the entire population as well as errors in taking a sample. Improved accuracy can be achieved by using sampling methods versus taking a complete census (see reference to Hogan, 90). By using sample information, we can make valid inferences about the entire population without the time and expense involved in taking a census.

6.6 a. Mean and variance of the sampling distribution for the sample mean

$$\mu_{\bar{x}} = \mu = 100$$

$$\sigma^2_{\bar{x}} = \sigma^2/n = 900/30 = 30;\ \sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{30}$$

b. $z_{\bar{x}} = \dfrac{109-100}{\sqrt{30}} = 1.64$

Probability that $\bar{x} > 109 = 1 - Fz(1.64) = .0505$

c. $z_{\bar{x}} = \dfrac{96-100}{\sqrt{30}} = -.73;\ 1 - Fz(.73) = .2327.$

$z_{\bar{x}} = \dfrac{110-100}{\sqrt{30}} = 1.83;\ Fz = .9664.$

Probability that $96 \le \bar{x} \le 110 = .9664 - .2327 = .7337$

d. $z_{\bar{x}} = \dfrac{107-100}{\sqrt{30}} = 1.28$

Probability that $\bar{x} \le 107 = Fz = .8997$

6.8 a. Mean and variance of the sampling distribution for the sample mean

$$\mu_{\bar{x}} = \mu = 400$$

$$\sigma^2_{\bar{x}} = \sigma^2/n = 1600/35 = 45.7143;\ \sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{45.7143}$$

b. $z_{\bar{x}} = \dfrac{412-400}{\sqrt{45.7143}} = 1.77$

Probability that $\bar{x} > 412 = 1 - Fz(1.77) = .0384$

c. $z_{\bar{x}} = \dfrac{407-400}{\sqrt{45.7143}} = 1.04;\ Fz(1.04) = .8508.$

$z_{\bar{x}} = \dfrac{393-400}{\sqrt{45.7143}} = -1.04;\ 1 - Fz(1.04) = .1492.$

Probability that $393 \le \bar{x} \le 407 = .8508 - .1492 = .7016$

d. $z_{\bar{x}} = \dfrac{389-400}{\sqrt{45.7143}} = -1.63$

Probability that $\bar{x} \le 389 = 1 - Fz(1.63) = 1 - .9484 = .0516$

6.10 a. $E(\overline{X}) = \mu_{\bar{x}} = 1{,}200$ hours

b. $\sigma_{\bar{x}}^{2} = \frac{\sigma^2}{n} = \frac{(400)^2}{9} = 17{,}778$ hours

c. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{400}{3} = 133.33$ hours

d. $P(Z < \frac{1{,}050 - 1{,}200}{133.33}) = P(Z < -1.13) = .1292$

6.12

a. $P(\bar{x} > 210{,}000) = P\left(Z > \frac{210{,}000 - 215{,}000}{25{,}000/\sqrt{100}}\right) = P(Z > -2) = 0.9772$

b. $P(213{,}000 < \bar{x} < 217{,}000)$

$= P\left(\frac{213{,}000 - 215{,}000}{25{,}000/\sqrt{100}} < Z < \frac{217{,}000 - 215{,}000}{25{,}000/\sqrt{100}}\right) = P(-0.8 < Z < 0.8)$

$= 0.5763$

c. $P(214{,}000 < \bar{x} < 216{,}000)$

$= P\left(\frac{214{,}000 - 215{,}000}{25{,}000/\sqrt{100}} < Z < \frac{216{,}000 - 215{,}000}{25{,}000/\sqrt{100}}\right) = P(-0.4 < Z < 0.4)$

$= 0.3108$

d. The sample mean selling price is most likely to lie in the range \$214,000 to \$216,000 since it is centered about the given population mean.

e. The results were still valid because of the central limit theorem with the sample size being larger than 30.

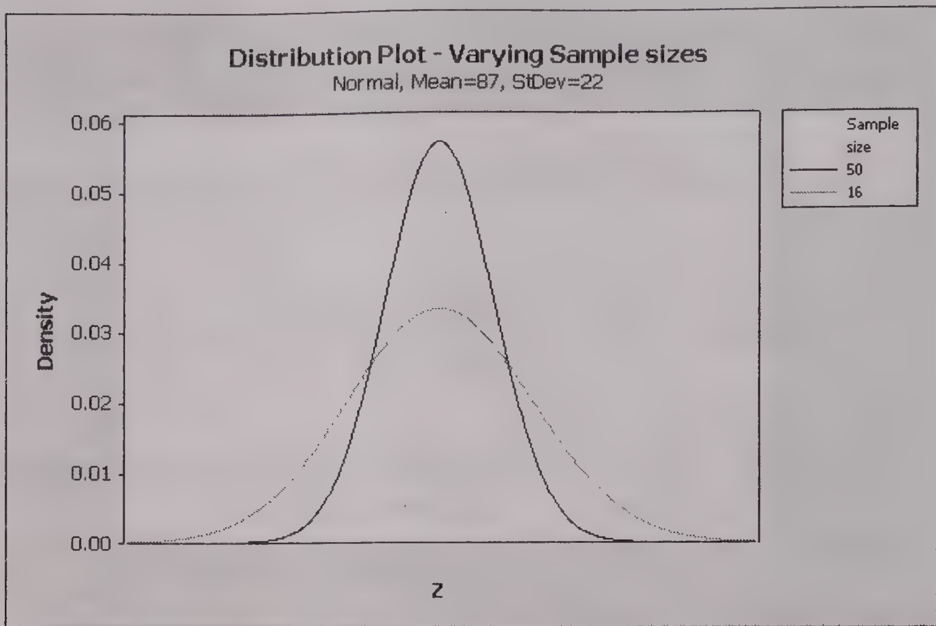
6.14 a. $\sigma_{\bar{x}} = \frac{22}{\sqrt{16}} = 5.5$ minutes

b. $P(Z < \frac{100 - 87}{5.5}) = P(Z < 2.36) = .9909$

c. $P(Z > \frac{80 - 87}{5.5}) = P(Z > -1.27) = .8980$

d. $P(\frac{85 - 87}{5.5} > Z > \frac{95 - 87}{5.5}) = P(-.36 > Z > 1.45) = .4329$

e. Higher, higher, lower. The graph will show that the standard error of the sample means will decrease with an increased sample size.



6.16 a. $\sigma_{\bar{x}} = \dfrac{40}{\sqrt{100}} = \$4$

b. $P(Z > 5/4) = P(Z > 1.25) = .1056$

c. $P(Z < -4/4) = P(Z < -1) = .1587$

d. $P(-3/4 > Z > 3/4) = P(-.75 > Z > .75) = .4532$

6.18 a. $\sigma_{\bar{x}} = \dfrac{1.6}{\sqrt{100}} = .16$; $P(Z > 1.645) = .05$, $1.645 = \dfrac{Difference}{.16}$, Difference = .2632

b. $P(Z < -1.28) = .1$, $-1.28 = \dfrac{Difference}{.16}$, Difference = $-.2048$

c. $P(-1.44 > Z > 1.44) = .15$, $P(Z > 1.44) = .075$, $1.44 = \dfrac{Difference}{.16}$, Difference = $\pm.2304$

6.20 a. $P(Z > 1.96) = .025$, $1.96 = \dfrac{2}{8.4/\sqrt{n}}$, $n = 67.766$, take $n = 68$

b. smaller

c. larger

6.22 a. $N = 20$, correction factor $= \frac{0}{19}$

$N = 40$, correction factor $= \frac{20}{39}$

$N = 100$, correction factor $= \frac{80}{99}$

$N = 1{,}000$, correction factor $= \frac{980}{999}$

$N = 10{,}000$, correction factor $= \frac{9{,}980}{9{,}999}$

b. When the population size (N) equals the sample size (n), then there is no variation away from the population mean and the standard error will be zero. As the sample size becomes relatively small compared to the population size, the correction factor tends towards 1 and the correction factor becomes less significant in the calculation of the standard error.

c. The correction factor tends toward a value of 1 and becomes progressively less important as a modifying factor when the sample size decreases relative to the population size.

6.24 $\sigma_{\bar{x}} = \frac{30}{\sqrt{50}}\sqrt{\frac{200}{249}} = 3.8023$

a. $P(Z > \frac{2.5}{3.8023}) = P(Z > .66) = .2546$

b. $P(Z < \frac{-5}{3.8023}) = P(Z < -1.31) = .0951$

c. $P(\frac{-10}{3.8023} < Z < \frac{10}{3.8023}) = P(-2.63 < Z < 2.63) = 1 - .9914 = .0086$

6.26 $E(\hat{p}) = .4$; $\sigma_{\hat{p}} = \sqrt{\frac{(.4)(.6)}{100}} = .04899$

a. Probability that the sample proportion is greater than .45

$z = \frac{.45 - .4}{.04899} = P(Z > 1.02) = .1539$

b. Probability that the sample proportion is less than .29

$z = \frac{.29 - .4}{.04899} = P(Z < -2.25) = .0122$

c. Probability that the sample proportion is between .35 and .51

$P(\frac{.35 - .4}{.04899} < Z < \frac{.51 - .4}{.04899}) = P(-1.02 < Z < 2.25) = .8339$

6.28 $E(\hat{p}) = .60$; $\sigma_{\hat{p}} = \sqrt{\dfrac{(.6)(.4)}{100}} = .04899$

a. Probability that the sample proportion is greater than .66

$$z = \frac{.66-.6}{.04899} = P(Z > 1.22) = .1112$$

b. Probability that the sample proportion is less than .48

$$z = \frac{.48-.6}{.04899} = P(Z < -2.45) = .0071$$

c. Probability that the sample proportion is between .52 and .66

$$P(z = \frac{.52-.6}{.04899} < Z < z = \frac{.66-.6}{.04899}) = P(-1.63 < Z < 1.22) = .8372$$

6.30 a. $E(\hat{p}) = .424$

b. $\sigma_{\hat{p}}^2 = \dfrac{(.424)(.576)}{100} = .00244$

c. $\sigma_{\hat{p}} = .0494$

6.32 a. $E(\hat{p}) = .20$

b. $\sigma_{\hat{p}}^2 = \dfrac{(.2)(.8)}{180} = .000889$

c. $\sigma_{\hat{p}} = .0298$

d. $P(Z < \dfrac{.15-.2}{.0298}) = P(Z < -1.68) = .0465$

6.34 $\sigma_{\hat{p}} = \sqrt{\dfrac{(.4)(.6)}{120}} = .0447$

$$P(\frac{.35-.4}{.0447} < Z < \frac{.45-.4}{.0447}) = P(-1.12 < Z < 1.12) = .7372$$

6.36 a. $\sigma_{\hat{p}} = \sqrt{\dfrac{(.2)(.8)}{130}} = .0351$

b. $P(Z > \dfrac{.15-.2}{.0351}) = P(Z > -1.42) = .9222$

c. $P(\dfrac{.18-.2}{.0351} < Z < \dfrac{.22-.2}{.0351}) = P(-.57 < Z < .57) = .4314$

d. Higher, higher

6.38 The largest value for $\sigma_{\hat{p}}$ is when $P = .5$. In this case, $\sigma_{\hat{p}} = \sqrt{\dfrac{(.5)(.5)}{100}} =$ .05

6.40 $\sigma_{\hat{p}} = \sqrt{\dfrac{(.25)(.75)}{120}} = .0395$

a. $P(Z > 1.28) = .1,\ 1.28 = \dfrac{Difference}{.0395}$, Difference = .0506

b. $P(Z < -1.645) = .05,\ -1.645 = \dfrac{Difference}{.0395}$, Difference = .065

c. $P(Z > 1.036) = .15,\ 1.036 = \dfrac{Difference}{.0395}$, Difference = .0409

6.42 $\sigma_{\hat{p}} = \sqrt{\dfrac{(.5)(.5)}{250}} = .03162,\ P(Z > \dfrac{.58-.5}{.03162}) = P(Z > 2.53) = .0057$

6.44 a. $\sigma_{\hat{p}} = \sqrt{\dfrac{0.4(1-0.4)}{120}} = 0.0447$

b. $P(\hat{p} < 0.33) = P\left(Z < \dfrac{0.33-0.40}{0.0447}\right) = P(Z < -1.57) = 0.0582$

c. $P(0.38 < \hat{p} < 0.46) = P\left(\dfrac{0.38-0.40}{0.0447} < Z < \dfrac{0.46-0.40}{0.0447}\right)$

$= P(-0.45 < Z < 1.34) = 0.5835$

6.46 $P(Z < \dfrac{.1-.122}{.036} = P(Z < -.61) = .2709,\ \sigma_{\hat{p}} = \sqrt{\dfrac{(.2709)(.7291)}{81}} = .04938$

$P(Z > \dfrac{.5-.2709}{.04938}) = P(Z > 4.64) \approx .0000$

6.48 a. Probability that the sample mean is > 200.

$z_{\bar{x}} = \dfrac{200-198}{10/\sqrt{25}} = 1.00$

Probability that $\bar{x} > 200 = 1 - Fz(1.00) = .1587$

b. 5% of the sample variances would be less than this value of the sample variance

$$P(s^2 > k) = P\left[\frac{(n-1)s^2}{\sigma^2}\right] = 0.95$$

$$\chi^2_{24,.95} = 13.85,\ \frac{24s^2}{100} < 13.85,\ s^2 < 57.702$$

c. 5% of the samples variances would be greater than this value of the sample variance

$$P(s^2 > k) = P\left[\frac{(n-1)s^2}{\sigma^2}\right] = 0.05$$

$$\chi^2_{24,.05} = 36.42,\ \frac{24s^2}{100} > 36.42,\ s^2 > 151.729$$

6.50 $P(\frac{(n-1)s^2}{\sigma^2} > \frac{19(3.1)}{1.75}) = P(\chi^2_{(19)} > 33.66)$ = between .01 and .025 (.0201 exactly)

6.52 a. $P(\frac{(n-1)s^2}{\sigma^2} > \frac{15(3,000)^2}{(2,500)^2}) = P(\chi^2_{(15)} > 21.6)$ = greater than .1 (.1187 exactly)

b. $P(\frac{(n-1)s^2}{\sigma^2} < \frac{15(1,500)^2}{(2,500)^2}) = P(\chi^2_{(15)} < 5.4)$ = between .01 and .025 (.0118 exactly)

6.54 a. $P(\frac{(n-1)s^2}{\sigma^2} < \frac{24(75)^2}{(100)^2}) = P(\chi^2_{(24)} < 13.5)$ = between .025 and .05 (.0429 exactly)

b. $P(\frac{(n-1)s^2}{\sigma^2} > \frac{24(150)^2}{(100)^2}) = P(\chi^2_{(24)} > 54)$ = less than .005 (.0004 exactly)

6.56

Descriptive Statistics: C20, C21, C22, C23, C24, C25, C26, C27, ...

Variable	Mean	Variance
C20	3.00	2.00
C21	4.00	8.00
C22	4.00	8.00
C23	4.50	12.50
C24	5.00	18.00
C25	5.00	2.00
C26	5.00	2.00
C27	6.00	8.00
C28	6.0000	0.000000000
C29	6.500	0.500
C30	7.00	2.00
C31	6.500	0.500
C32	7.00	2.00
C33	7.500	0.500
C34	5.50	4.50

Descriptive Statistics: Variance

```
Variable  Mean  StDev  Variance    Sum
Variance  4.72   5.26     27.62  70.80
```

$\overline{x} = \dfrac{70.8}{15} = 4.72$; Therefore, $E(s^2) \neq \sigma^2$.

$E(s^2) = {6(3.91667)}\big/{(5)} = 4.7$; Therefore, $E(s^2) = {N\sigma^2}\big/{(N-1)}$

6.58 a. $P(\chi^2_{(9)} > 14.68) = .10$, $14.68 = 9(\text{Percentage})$, Percentage = 1.6311 (163.11%)
b. $P(\chi^2_{(9)} < 2.7) = .025$, $P(\chi^2_{(9)} > 19.02) = .025$,
$2.7 = 9a$, $a = .3$, $19.02 = 9b$, $b = 2.1133$
The probability is .95 that the sample variance is between 30% and 211.33% of the population variance.
c. The interval in part b. will be smaller.

6.60 a. $P(\chi^2_{(11)} > 4.57) = .95$, $4.57 = 11\text{Percentage}$, Percentage = .4155 (41.55%)
b. $P(\chi^2_{(11)} > 5.58) = .90$, $5.58 = 11\text{Percentage}$, Percentage = .5073 (50.73%)
c. $P(\chi^2_{(11)} < 3.82) = .025$, $P(\chi^2_{(11)} > 21.92) = .025$,
$3.82 = 11a$, $a = .34727$, $21.92 = 11b$, $b = 1.9927$
The probability is .95 that the sample variance is between 34.727% and 199.27% of the population variance.

6.62 $P(\dfrac{(n-1)s^2}{\sigma^2} < \dfrac{24(12.2)}{15.4}) = P(\chi^2_{(24)} < 19.01)$ = less than .90 (.2485 exactly)

6.64 a. $C_2^6 = \dfrac{6!}{2!4!} = 15$ possible samples
b. (41, 39), (41, 35), (41, 35), (41, 33), (41, 38), (39, 35), (39, 35), (39, 33), (39, 38), (35, 35), (35, 33), (35, 38), (35, 33), (35, 38), (33, 38)
c. $34P_{\overline{X}}(34) = 34\dfrac{2}{15} = 4.5333$

$35P_{\overline{X}}(35) = \dfrac{35}{15} = 2.3333$

$35.5P_{\overline{X}}(35.5) = \dfrac{35.5}{15} = 2.3667$

$36P_{\overline{X}}(36) = \dfrac{36}{15} = 2.4$

$$36.5P_{\bar{X}}(36.5) = 36.5\frac{2}{15} = 4.8667$$

$$37P_{\bar{X}}(37) = 37\frac{3}{15} = 7.4$$

$$38P_{\bar{X}}(38) = 38\frac{2}{15} = 5.0667$$

$$38.5P_{\bar{X}}(38.5) = \frac{38.5}{15} = 2.5667$$

$$39.5P_{\bar{X}}(39.5) = \frac{39.5}{15} = 2.6333$$

$$40P_{\bar{X}}(40) = \frac{40}{15} = 2.6667$$

d. The mean of the sampling distribution of the sample mean is $\sum \bar{x}P_{\bar{x}}(\bar{x}) = 36.8333$ which is exactly equal to the population mean: $\frac{1}{N}\sum x_i = 36.8333$.

6.66 a. $P(Z > \frac{450-420}{100/\sqrt{25}}) = P(Z > 1.5) = .0668$

b. $P(\frac{400-420}{100/\sqrt{25}} < Z < \frac{450-420}{100/\sqrt{25}}) = P(-1 < Z < 1.5) = .7745$

c. $P(Z > 1.28) = .1,\ 1.28 = \frac{\bar{x}-420}{100/\sqrt{25}},\ \bar{x} = 445.6$

d. $P(Z < -1.28) = .1,\ -1.28 = \frac{\bar{x}-420}{100/\sqrt{25}},\ \bar{x} = 394.4$

e. $P(\chi^2_{(24)} > 36.42) = .05,\ 36.42 = \frac{24s^2}{(100)^2},\ s = 123.1868$

f. $P(\chi^2_{(24)} < 13.85) = .05,\ 13.85 = \frac{24s^2}{(100)^2},\ s = 75.966$

g. Smaller. A larger sample size would lead to a smaller standard error and the graph of the normal distribution would be tighter with less area in the tails.

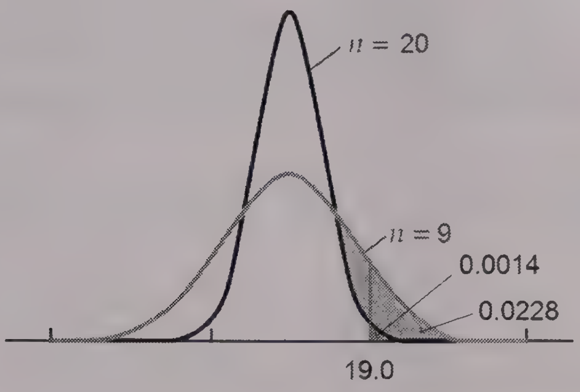
6.68 a. $P(Z > \frac{19-14.8}{6.3/\sqrt{9}}) = P(Z > 2) = .0228$

b. $P(\frac{10.6-14.8}{6.3/\sqrt{9}} < Z < \frac{19-14.8}{6.3/\sqrt{9}}) = P(-2 < Z < 2) = .9544$

c. $P(Z < -.67) = .25,\ -.67 = \frac{\bar{x}-14.8}{6.3/\sqrt{9}},\ \bar{x} = 13.3930$

d. $P(\chi^2_{(8)} > 13.36) = .1,\ 13.36 = \dfrac{8s^2}{(6.3)^2},\ s = 8.1414$

e. Smaller



6.70 Let $n = N$, then $\bar{X} = \mu_x$:

$$E[\sum_{i=1}^{N}(X_i - \bar{X})^2] = n\sigma_x^2 - n\frac{\sigma_x^2}{n}\frac{N-n}{N-1} = n\sigma_x^2 - \frac{N-n}{N-1}\sigma_x^2 =$$

$$\frac{\sigma_x^2}{N-1}(nN - n - N + n) = \frac{N\sigma_x^2}{N-1}(n-1)$$

Therefore, $E[s_x^2] = E[\frac{1}{n-1}\sum(X_i - \bar{X})^2] = \frac{1}{n-1}E[\sum(X_i - \bar{X})^2] = \frac{N\sigma_x^2}{N-1}$

6.72a. $P(\hat{p} < 0.7) = P\left(Z < \dfrac{0.7 - 0.8}{\sqrt{(0.8)(0.2)/60}}\right) = P(Z < -1.94) = 0.0262$

b. $P(\hat{p} < 0.7) = P\left(Z < \dfrac{0.7 - 0.8}{\sqrt{(0.8)(0.2)/6}}\right) = P(Z < -0.61) = 0.2709$

c. $P(\bar{x} > 38{,}000) = P\left(Z > \dfrac{38{,}000 - 37{,}000}{4{,}000/\sqrt{6}}\right) = P(Z > 0.61) = 0.2709$

d. $P(x > 38{,}000) = P\left(Z > \dfrac{38{,}000 - 37{,}000}{4{,}000}\right) = P(Z > 0.25) = 0.4013$

6.74 $P(\frac{(n-1)s^2}{\sigma^2} > 20(2)) = P(\chi^2_{(20)} > 40) = .005$

6.76 $\bar{X} - 10 < \mu_{\bar{x}} < \bar{X} + 10, \ -10 < \bar{X} - \mu_{\bar{x}} < 10$

$$P(\frac{-10}{40/\sqrt{16}} < Z < \frac{10}{40/\sqrt{16}}) = P(-1 < Z < 1) = .6826$$

6.78 a. $\sigma_{\hat{p}} = \sqrt{\frac{(.4)(.6)}{250}} = .03098$

$P(Z > -.84) = .8, \ -.84 = \frac{p-.4}{.03098}, p = .374$

b. $P(Z < 1.28) = .9, \ 1.28 = \frac{p-.4}{.03098}, p = .4397$

c. $P(Z > .39) = .35, \ .39 = \frac{Difference}{.03098}$, Difference $= \pm.0121$

6.80 a. $P(\frac{(n-1)s^2}{\sigma^2} > \frac{24(4,000)^2}{(6,600)^2}) = P(\chi^2_{(24)} > 8.82)$ = more than .995 (.9979 exactly)

b. $P(\frac{(n-1)s^2}{\sigma^2} < \frac{24(8,000)^2}{(6,600)^2}) = P(\chi^2_{(24)} < 35.62)$ = between .9 and .95 (.9354 exactly)

6.82 a. The sample mean is $\bar{x} = \frac{\sum x}{n} = \frac{70803}{100} = 708.03$.
The sample standard deviation is $s = 8.106$.

To find the standard deviation of the sample mean for each sample, use the equation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Use $s = 8.106$ as an estimate for σ and note that $n = 5$ for each sample. Thus, $\sigma_{\bar{x}} = \frac{8.106}{\sqrt{5}} = 3.625$.

b. $P(\bar{x} < 685) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} < \frac{685 - 710}{3.625}\right) = P(Z < -6.90) = 0.0000$

c. $P(\bar{x} > 720) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} > \frac{720 - 710}{3.625}\right) = P(Z > 2.76) = 0.0029$

6.84 a. The sample mean is $\bar{x} = \frac{\sum x}{n} = \frac{156654.6}{138} = 1135.178.$

The sample variance is $s^2 = 11.3382.$

To find the variance of the sample mean for each sample, use the equation

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, which gives us $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$. Use $s^2 = 11.3382$ as an estimate for σ^2 and

note that $n = 6$ for each sample. Thus, $\sigma_{\bar{x}}^2 = \frac{11.3382}{6} = 1.8898.$

b. Use $\sigma_{\bar{x}} = \sqrt{1.8898} = 1.3747.$

$$P(1120 < \bar{x} < 1150) = P\left(\frac{1120-1134}{1.3747} < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{1150-1134}{1.3747}\right)$$

$$= P(-10.18 < z < 11.64) = 1.0000$$

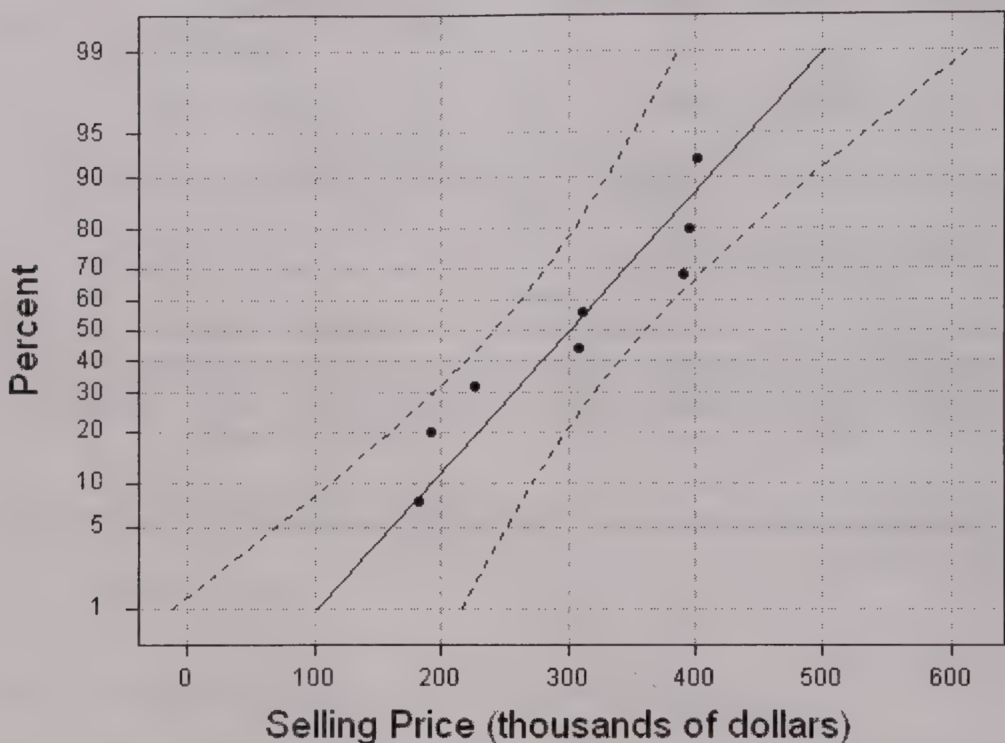
b. Answers will vary. The sample means for our 23 random samples :
1138.32, 1137.23, 1136.15, 1134.98, 1135.90, 1137.58, 1133.73, 1134.05,
1134.98, 1133.22, 1136.45, 1135.37, 1134.82, 1134.02, 1136.17, 1133.90,
1133.15, 1135.30, 1133.90, 1133.63, 1134.18, 1135.22, 1136.85

There are no sample means that are outside the acceptance limits, 1,120 grams and 1,150 grams.

Chapter 7:

Estimation: Single Population

7.2 a. There appears to be no evidence of nonnormality, as shown by the normal probability plot.



b. Assuming normality, the unbiased and most efficient point estimator for the population mean is the sample mean, $\overline{X}$.

$$\overline{X} = \frac{\sum X_i}{n} = \frac{2411}{8} = 301.375 \text{ thousand dollars}$$

c. Assuming normality, the sample mean is an unbiased estimator of the population mean with variance $\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$. Also, assuming normality, the unbiased and most efficient point estimator for the population variance is the sample variance, s^2.

$$\text{Var}(\overline{X}) = \frac{s^2}{n} = \frac{8373.125}{8} = 1046.64$$

d. The unbiased and most efficient point estimator for a proportion is the sample proportion, $\hat{p}$.

$$\hat{p} = \frac{x}{n} = \frac{3}{8} = 0.375$$

7.4 $n = 12$ employees. Number of hours of overtime worked in the last month:

a. Unbiased point estimator of the population mean is the sample mean:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{293}{12} = 24.42$$

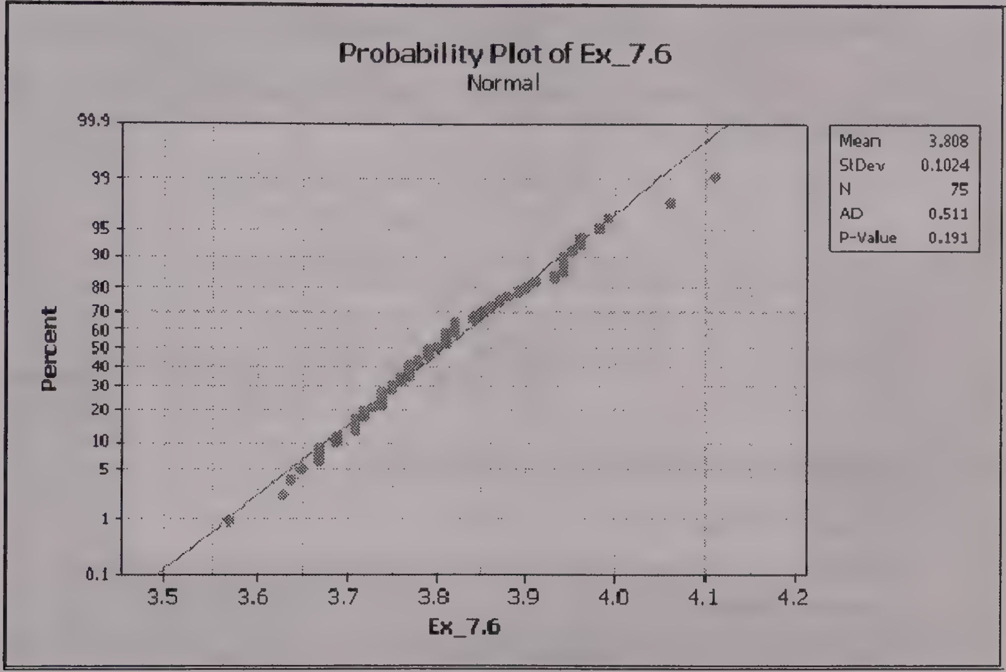
b. The unbiased point estimate of the population variance: $s^2 = 85.72$

c. Unbiased point estimate of the variance of the sample mean

$$Var(\bar{X}) = \frac{s^2}{n} = \frac{85.72}{12} = 7.1433$$

d. Unbiased estimate of the population proportion: $\hat{p} = \frac{x}{n} = \frac{3}{12} = .25$

7.6 a.



No evidence of the data distribution coming from a nonnormal population.

b. The minimum variance unbiased point estimate of the population mean is the sample mean: $\bar{X} = \frac{\sum X_i}{n} = \frac{285.59}{75} = 3.8079$

Descriptive Statistics: Volumes

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Volumes	75	3.8079	3.7900	3.8054	0.1024	0.0118

Variable	Minimum	Maximum	Q1	Q3
Volumes	3.5700	4.1100	3.7400	3.8700

c. The minimum variance unbiased point estimate of the population variance is the sample variance $s^2 = 0.1024^2 = .0105$.

7.8 Reliability factor for each of the following:

a. 93% confidence level: $z_{\alpha/2} = +/- 1.81$

b. 96% confidence level: $z_{\alpha/2} = +/- 2.05$

c. 80% confidence level: $z_{\alpha/2} = +/- 1.28$

7.10 Calculate the margin of error to estimate the population mean

a. 98% confidence level, $n = 64$; variance = 144

$$ME = z_{\alpha/2}\,{}^{\sigma}\!/\!\sqrt{n} = 2.33\left({}^{12}\!/\!\sqrt{64}\right) = 3.495$$

b. 99% confidence interval, $n = 120$; standard deviation = 100

$$ME = z_{\alpha/2}\,{}^{\sigma}\!/\!\sqrt{n} = 2.58\left({}^{100}\!/\!\sqrt{120}\right) = 23.552$$

7.12 Calculate the LCL and UCL:

a. $\bar{x} \pm z_{\alpha/2}\,{}^{\sigma}\!/\!\sqrt{n} = 50 \pm 1.96\left({}^{40}\!/\!\sqrt{64}\right) = 40.2 \text{ to } 59.8$

b. $\bar{x} \pm z_{\alpha/2}\,{}^{\sigma}\!/\!\sqrt{n} = 85 \pm 2.58\left({}^{20}\!/\!\sqrt{225}\right) = 81.56 \text{ to } 88.44$

c. $\bar{x} \pm z_{\alpha/2}\,{}^{\sigma}\!/\!\sqrt{n} = 510 \pm 1.645\left({}^{50}\!/\!\sqrt{485}\right) = 506.27 \text{ to } 513.73$

7.14 a. Calculate the standard error of the mean

$${}^{\sigma}\!/\!\sqrt{n} = {}^{5}\!/\!\sqrt{100} = .5$$

b. Calculate the margin of error of the 90% confidence interval for the population mean

$$ME = z_{\alpha/2}\,{}^{\sigma}\!/\!\sqrt{n} = 1.645(.5) = .8225$$

c. Calculate the width of the 98% confidence interval for the population mean

$$\text{width} = 2ME = 2\left[z_{\alpha/2}\,{}^{\sigma}\!/\!\sqrt{n}\right] = 2\left[2.33(.5)\right] = 2.33$$

7.16 $n = 25, \quad \bar{x} = 19.8, \quad \sigma = 1.2, \quad z_{.005} = 2.58$

99% confidence interval:

$$\bar{x} \pm z\left({}^{\sigma}\!/\!\sqrt{n}\right) = 19.8 \pm 2.58(1.2/5) = 19.1808 \text{ up to } 20.4192$$

7.18 Find the *ME*

a. $n = 4$, 99% confidence level, $x_1 = 25, x_2 = 30, x_3 = 33, x_4 = 21$

$s = 5.3151$

$$ME = t_{v,\alpha/2} \, s/\sqrt{n} = 5.841\left(5.3151/\sqrt{4}\right) = 15.5227$$

b. $n = 5$, 90% confidence level, $x_1 = 15, x_2 = 17, x_3 = 13, x_4 = 11, x_5 = 14$

$s = 2.2361$

$$ME = t_{v,\alpha/2} \, s/\sqrt{n} = 2.132\left(2.2361/\sqrt{5}\right) = 2.1320$$

7.20 Find the LCL and UCL for each of the following:

a. $\alpha = .05$, $n = 25$, sample mean = 560, $s = 45$

$$\bar{x} \pm t_{v,\alpha/2}\left(s/\sqrt{n}\right) = 560 \pm 2.064\left(45/\sqrt{25}\right) = 541.424 \text{ to } 578.576$$

b. $\alpha/2 = .05$, $n = 9$, sample mean = 160, sample variance = 36

$$\bar{x} \pm t_{v,\alpha/2}\left(s/\sqrt{n}\right) = 160 \pm 1.860\left(6/\sqrt{9}\right) = 156.28 \text{ to } 163.72$$

c. $1 - \alpha = .98$, $n = 22$, sample mean = 58, $s = 15$

$$\bar{x} \pm t_{v,\alpha/2}\left(s/\sqrt{n}\right) = 58 \pm 2.518\left(15/\sqrt{22}\right) = 49.9474 \text{ to } 66.0526$$

7.22 Calculate the width for each of the following:

a. $\alpha = 0.05$, $n = 6$, $s = 40$

$$w = 2ME = 2\left[t_{v,\alpha/2} \, s/\sqrt{n}\right] = 2\left[2.571\left(40/\sqrt{6}\right)\right] = 2(41.98425) = 83.9685$$

b. $\alpha = 0.01$, $n = 22$, sample variance = 400

$$w = 2ME = 2\left[t_{v,\alpha/2} \, s/\sqrt{n}\right] = 2\left[2.831\left(20/\sqrt{22}\right)\right] = 2(12.07142) = 24.1428$$

c. $\alpha = 0.10$, $n = 25$, $s = 50$

$$w = 2ME = 2\left[t_{v,\alpha/2} \, s/\sqrt{n}\right] = 2\left[1.711\left(50/\sqrt{25}\right)\right] = 2(17.11) = 34.22$$

7.24 **Results for: Sugar.xls**
Descriptive Statistics: Weights

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Weights	100	520.95	518.75	520.52	9.45	0.95

Variable	Minimum	Maximum	Q1	Q3
Weights	504.70	544.80	513.80	527.28

90% confidence interval:
Results for: Sugar.xls
One-Sample T: Weights

Variable	N	Mean	StDev	SE Mean	90.0% CI
Weights	100	520.948	9.451	0.945	(519.379, 522.517)

b. Narrower since a smaller value of t will be used in generating the 80% confidence interval.

7.26 $n = 28, \quad \bar{x} = 60.893, \quad s = 5.2305, \quad t_{27,.025} = 2.052$

$$ME = t_{v,\alpha/2}\, {}^{s}\!\big/\!_{\sqrt{n}} = 2.052\left(5.2305\big/\sqrt{28}\right) = 2.0283$$

7.28 $n = 25, \quad \bar{x} = 42,740, \quad s = 4,780, \quad t_{24,.05} = 1.711$

90% confidence interval:
$42,740 \pm 1.711(4780/5) = \$41,104.28$ up to $\$44,375.72$

7.30 Find the margin of error to estimate the population proportion for each of the following:

a. $n = 350, \ \hat{p} = .3, \ \alpha = .01$

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2.576\sqrt{\frac{.3(.7)}{350}} = .0631$$

b. $n = 275, \ \hat{p} = .45, \ \alpha = .05$

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{.45(.55)}{275}} = .0588$$

c. $n = 500, \ \hat{p} = .05, \ \alpha = .10$

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645\sqrt{\frac{.05(.95)}{500}} = .01603$$

7.32 $n = 250 + 75 + 25 = 350;$

a. Estimate the percent of alumni in favor of the program:

$\hat{p} = 250/350 = .7143$

$\alpha = .05$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .7143 \pm 1.96\sqrt{\frac{.7143(.2857)}{350}} = .667 \text{ up to } .7616$$

b. Estimate the percent of alumni in opposed to the program:
$\hat{p} = 75/350 = .2143$
90% confidence interval

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .2143 \pm 1.645\sqrt{\frac{.2143(.7857)}{350}} = .1782 \text{ up to } .2504$$

7.34 $n = 95$, $\hat{p} = 67/95 = .7053$, $z_{.005} = 2.58$
99% confidence interval:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .7053 \pm (2.58)\sqrt{\frac{.7053(.2947)}{95}} = .5846 \text{ up to } .8260$$

7.36 $n = 350$, $\hat{p} = 73/350 = .2086$, $z_{.025} = 1.96$
95% confidence interval:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .2086 \pm (1.96)\sqrt{.2086(.7914)/350} = .166 \text{ to } .2512$$

7.38 $width = .76 - .68 = .08$; $ME = .04$, $\hat{p} = 180/250 = 0.72$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{.72(.28)}{250}} = .0284$$

$ME = .04 = z_{\alpha/2}\,(.0284)$, $z_{\alpha/2} = 1.41$
$\alpha = 2[1 - F_z(1.41)] = 2[.0793] = .1586$
Confidence level: $100(1 - .1586) = 84.14\%$

7.40 a. $n = 460$, $\hat{p} = (50 + 120 + 80)/460 = .5435$, $z_{.025} = 1.96$
95% confidence interval:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .5435 \pm (1.96)\sqrt{\frac{.5435(.4565)}{460}} = .49798 \text{ up to } .58898$$

b. $n = 460$, $\hat{p} = (60 + 50 + 40)/460 = .3261$, $z_{.05} = 1.645$
90% confidence interval:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .3261 \pm (1.645)\sqrt{\frac{.3261(.6739)}{460}} = .29009 \text{ up to } .36209$$

7.42 a. $n = 21$, $s^2 = 16$, *taking* $\alpha = .05$, $\chi^2_{20,.025} = 34.17$

$$LCL = \frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} = \frac{20(16)}{34.17} = 9.3649$$

b. $n=16, \quad s=8, \quad taking \quad \alpha=.05, \quad \chi^2_{15,.025}=27.49$

$$LCL=\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}=\frac{15(8)^2}{27.49}=34.9218$$

c. $n=28, \quad s=15, \quad taking \quad \alpha=.01, \quad \chi^2_{27,.005}=49.64$

$$LCL=\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}=\frac{27(15)^2}{49.64}=122.3811$$

7.44 Random sample from a normal population

a. Find the 90% confidence interval for the population variance

Descriptive Statistics: Ex7.44

Variable	Mean	SE Mean	StDev	Variance	Minimum	Maximum
Ex7.44	11.00	1.41	3.16	10.00	8.00	16.00

$n=5, \quad s^2=10, \quad \chi^2_{4,.95}=.711, \quad \chi^2_{4,.05}=9.49$

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}<\sigma^2<\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}=\frac{4(10)}{9.49}<\sigma^2<\frac{4(10)}{.711}=4.21496<\sigma^2<$$
56.25879

b. Find the 95% confidence interval for the population variance

$n=5, \quad s^2=10, \quad \chi^2_{4,.975}=.484, \quad \chi^2_{4,.025}=11.14$

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}<\sigma^2<\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}=\frac{4(10)}{11.14}<\sigma^2<\frac{4(10)}{.484}=3.59066<\sigma^2<$$
82.64463

7.46 $n=10, \quad s^2=28.3023, \quad \chi^2_{9,.05}=16.92, \quad \chi^2_{9,.95}=3.33$

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}<\sigma^2<\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}=$$

$$\frac{9(28.3023)}{16.92}<\sigma^2<\frac{9(28.3023)}{3.33}=15.0544 \text{ up to } 76.4927$$

7.48 $n=18, \quad s^2=(10.4)^2=108.16, \quad \chi^2_{17,.05}=27.59, \quad \chi^2_{17,.95}=8.67$

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}<\sigma^2<\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}=\frac{17(108.16)}{27.59}<\sigma^2<\frac{17(108.16)}{8.67}$$

$= 66.6444$ up to 212.0784. Assume that the population is normally distributed.

7.50 7.50 $n = 9$, $s^2 = .7875$, $\chi^2_{8,.05} = 15.51$, $\chi^2_{8,.95} = 2.73$

$$\frac{8(.7875)}{15.51} < \sigma^2 < \frac{8(.7875)}{2.73} = .4062 \text{ up to } 2.3077$$

7.52 a. The confidence interval is $\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$. Use

$t_{79,0.025} = 1.990$, and $\hat{\sigma}^2_{\bar{x}} = 1.1676$ from exercise 7.51 part a.

$\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$

$142 - (1.990)(1.0806) < \mu < 142 + (1.990)(1.0806)$

or (139.85, 144.15)

b. The confidence interval is $\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$. Use

$t_{89,0.025} = 1.987$, and $\hat{\sigma}^2_{\bar{x}} = 0.6667$ from exercise 7.51 part b.

$\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$

$232.4 - (1.987)(0.8165) < \mu < 232.4 + (1.987)(0.8165)$

or (230.78, 234.02)

c. The confidence interval is $\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$. Use

$t_{199,0.025} = 1.972$, and $\hat{\sigma}^2_{\bar{x}} = 0.6049$ from exercise 7.51 part c.

$\bar{x} - t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + t_{n-1,\alpha/2}\hat{\sigma}_{\bar{x}}$

$59.3 - (1.972)(0.7777) < \mu < 59.3 + (1.972)(0.7777)$

or (57.77, 60.83)

7.54 a. A $100(1-\alpha)\%$ confidence interval for the population proportion is obtained from the following formula:

$\hat{p} - z_{\alpha/2}\hat{\sigma}_{\hat{p}} < P < \hat{p} + z_{\alpha/2}\hat{\sigma}_{\hat{p}}$

As stated, $N = 1058$, $n = 160$, and $x = 40$.

For a 95% confidence level, note that $z_{\alpha/2} = z_{0.025} = 1.96$. Next calculate $\hat{p}$ and $\hat{\sigma}_{\hat{p}}$.

$$\hat{p} = \frac{x}{n} = \frac{40}{160} = 0.25$$

$$\hat{\sigma}^2_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n-1} \times \frac{(N-n)}{(N-1)} = \frac{0.25(1-0.25)}{160-1} \times \frac{1058-160}{1058-1} = 0.0010019$$

$$\hat{\sigma}_{\hat{p}} = \sqrt{0.0010019} = 0.03165$$

So the 95% confidence interval is

$\hat{p} - z_{\alpha/2}\hat{\sigma}_{\hat{p}} < P < \hat{p} + z_{\alpha/2}\hat{\sigma}_{\hat{p}}$

or

$0.25 - (1.96)(0.03165) < P < 0.25 + (1.96)(0.03165)$

or

(0.188, 0.312)

b. As stated, $N = 854,\ n = 81,$ and $x = 50$.

For a 99% confidence level, note that $z_{\alpha/2} = z_{0.005} = 2.576$. Next calculate $\hat{p}$ and $\hat{\sigma}_{\hat{p}}$.

$$\hat{p} = \frac{x}{n} = \frac{50}{81} = 0.6173$$

$$\hat{\sigma}_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n-1} \times \frac{(N-n)}{(N-1)} = \frac{0.6173(1-0.6173)}{81-1} \times \frac{854-81}{854-1} = 0.002676$$

$$\hat{\sigma}_{\hat{p}} = \sqrt{0.002676} = 0.05173$$

So the 99% confidence interval is

$$\hat{p} - z_{\alpha/2}\hat{\sigma}_{\hat{p}} < P < \hat{p} + z_{\alpha/2}\hat{\sigma}_{\hat{p}}$$

or

$$0.6173 - (2.576)(0.05173) < P < 0.6173 + (2.576)(0.05173)$$

or

$$(0.484,\ 0.751)$$

7.56 a. $\bar{x} = 9.7, s = 6.2$, $\hat{\sigma}_{\bar{x}} = \sqrt{\frac{(s)^2}{n} \times \frac{N-n}{N-1}} = \sqrt{\frac{(6.2)^2}{50} \times \frac{139}{188}} = .7539$

95% confidence interval:
$9.7 \pm 2.01\ (.7539)$
or (8.1847, 11.2153)

b. 99% confidence interval:
$N\bar{x} - t_{\alpha/2}N\hat{\sigma}_{\bar{x}} < N\mu < N\bar{x} + t_{\alpha/2}N\hat{\sigma}_{\bar{x}}$
where, $N\bar{x} = (189)(9.7) = 1833.30$
$N\hat{\sigma}_{\bar{x}} = 189 \times .7539 = 142.4871$
$1833.30 \pm 2.68(142.4871)$
$1451.4346 < N\mu < 2215.1654$

7.58 a. 99% confidence interval:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{(5.32)^2}{40} \times \frac{85}{124}} = .6964$$

$7.28 \pm 2.708\ (.6964) = (5.3941,\ 9.1659)$

b. 90% confidence interval:
$N\bar{x} - t_{\alpha/2}N\hat{\sigma}_{\bar{x}} < N\mu < N\bar{x} + t_{\alpha/2}N\hat{\sigma}_{\bar{x}}$, where $N\bar{x} = (125)(7.28) = 910$
$N\hat{\sigma}_{\bar{x}} = 125 \times .6964 = 87.05$
$910 \pm 1.685(87.05) = 763.3208 < N\mu < 1{,}052.6793$

7.60 $\bar{x} = 143/35 = 4.0857$

90% confidence interval:

$N\bar{x} - t_{\alpha/2}N\hat{\sigma}_{\bar{x}} < N\mu < N\bar{x} + t_{\alpha/2}N\hat{\sigma}_{\bar{x}}$, where

$N\bar{x} = (120)(4.0857) = 490.2857$

$$N\hat{\sigma}_{\bar{x}} = 120\sqrt{\frac{(3.1)^2}{35} \times \frac{(120-35)}{(120-1)}} = 53.1429$$

490.2857 ±1.691(53.1429) = 400.4211< $N\mu$ < 580.1503

7.62 7.62 $\hat{p} = x/n = 56/100 = .56$

$\hat{\sigma}_{\hat{p}} = \sqrt{[\hat{p}(1-\hat{p})/(n-1)][(N-n)/(N-1)]}$

$= \sqrt{[(.56)(.44)/99][(420-100)/(420-1)]} = .0436$

90% confidence interval: $.56 \pm 1.645(.0436)$: .4883 up to .6317

7.64 $\hat{p} = x/n = 31/80 = .3875$

$\hat{\sigma}_{\hat{p}} = \sqrt{[\hat{p}(1-\hat{p})/(n-1)][(N-n)/(N-1)]}$

$= \sqrt{[(.3875)(.6125)/(79)][(420-80)/(420-1)]} = .0494$

90% confidence interval: $.3875 \pm 1.645(.0494)$: .3063 up to .4687

Therefore, $128.646 < NP < 196.854$ or between 129 and 197 students intend to take the final.

7.66 To estimate the population proportion:

a. $n = \dfrac{.25\left(z_{\alpha/2}\right)^2}{ME^2} = \dfrac{.25(1.96)^2}{.03^2} = 1067.111$. Take a sample of size $n =$ 1068.

b. $n = \dfrac{.25\left(z_{\alpha/2}\right)^2}{ME^2} = \dfrac{.25(1.96)^2}{.05^2} = 384.16$. Take a sample of size $n = 385$.

c. In order to reduce the *ME* in half, the sample size must be increased by a larger proportion.

7.68 a. $z_{.05} = 1.645, \quad ME = .04$

$$n = \frac{.25(z_{\alpha/2})^2}{ME^2} = \frac{(.25)(1.645)^2}{(.04)^2} = 422.8, \text{ take } n = 423$$

b. $\dfrac{(.25)(1.96)^2}{(.04)^2} = 600.25$, take $n = 601$

c. $\dfrac{(.25)(2.33)^2}{(.05)^2} = 542.89$, take $n = 543$

7.70 $z_{.05} = 1.645, \quad ME = .03$

$$n = \frac{.25(z_{\alpha/2})^2}{ME^2} = \frac{(.25)(1.645)^2}{(.03)^2} = 751.7\text{, take } n = 752$$

7.72 Use the equation $n = \dfrac{N\sigma^2}{(N-1)\sigma_{\bar{x}}^2 + \sigma^2}$ to find the sample size needed.

a. Since $1.96\sigma_{\bar{x}} = 50,\ \sigma_{\bar{x}} = \dfrac{50}{1.96} = 25.51$.

$$n = \frac{N\sigma^2}{(N-1)\sigma_{\bar{x}}^2 + \sigma^2} = \frac{(3300)(500)^2}{(3300-1)(25.51)^2 + (500)^2} = 344.2\text{, take } n = 345$$

b. From part (a), $\sigma_{\bar{x}} = 25.51$.

$$n = \frac{N\sigma^2}{(N-1)\sigma_{\bar{x}}^2 + \sigma^2} = \frac{(4950)(500)^2}{(4950-1)(25.51)^2 + (500)^2} = 356.6,\ \text{take } n = 357$$

c. From part (a), $\sigma_{\bar{x}} = 25.51$.

$$n = \frac{N\sigma^2}{(N-1)\sigma_{\bar{x}}^2 + \sigma^2} = \frac{(5000000)(500)^2}{(5000000-1)(25.51)^2 + (500)^2} = 384.1,\ \text{take } n = 385$$

d. The required sample size for part (b) is larger than that for part (a), and the required sample size for part (c) is larger than that for parts (a) and (b). This shows that the required sample size increases as N increases.

7.74 $\sigma_{\bar{x}} = \dfrac{2000}{1.96} = 1020.4,\ n = \dfrac{812(20000)^2}{811(1020.4)^2 + (20000)^2} = 261.0038$. Take 262 observations.

7.76 $\sigma_{\hat{p}} = \dfrac{.05}{2.575} = .0194$

$$n = \frac{.25N}{(N-1)\sigma^2_{\hat{p}} + .25} = \frac{(.25)320}{319(.0194)^2 + .25} = 216.18 = 217 \text{ observations.}$$

7.78 a. $n = 25,\quad \bar{x} = 227.60,\quad s = 41.86,\quad t_{24,.05} = 1.711$

90% confidence interval: $\bar{x} \pm t_{\alpha/2}\left(s/\sqrt{n}\right) = 227.60 \pm 1.711(41.86/\sqrt{25}) =$

213.2741 up to 241.9259. Assume that the population is normally distributed.

b. Width for 95% and 98% confidence intervals:

[95%]: Width $= 2ME = 2\left[t_{\alpha/2}\, {}^{s}\!/\!\sqrt{n}\right] = 2\left[2.064(8.3728)\right] = 34.563$

[98%]: Width $= 2ME = 2\left[t_{\alpha/2}\, {}^{s}\!/\!\sqrt{n}\right] = 2\left[2.492(8.3728)\right] = 41.73$

7.80 $n = 25, \quad \bar{x} = 28.5, \quad s = 6.8, \quad t_{24,.05} = 1.711$

90% confidence interval $= 28.5 \pm 1.711(6.8/\sqrt{25}\,) = 26.173$ up to 30.827

7.82 $n = 25, \quad \bar{x} = 12.5, \quad s = 3.8, \quad t_{24,.025} = 2.064$

b. For a 95% confidence interval, $12.5 \pm 2.064(3.8/5)$
$= 10.9314$ up to 14.0686

7.84 Width $= 3.69 - 3.49 = .2$, $ME = .1 = z_{\alpha/2}(1.045/\sqrt{457})$, $\quad z_{\alpha/2} = 2.05$

$\alpha = 2[1 - F_z(2.05)] = .0404$

$100(1-.0404) = 95.96\%$

7.86 $n = 33$ accounting students who recorded study time

a. An unbiased, consistent, and efficient estimator of the population mean is the sample mean, $\bar{x} = 8.545$

b. The sampling error for a 95% confidence interval using degrees of freedom = 32,

$$ME = 2.037\left({}^{3.817}\!/\!\sqrt{33}\right) = 1.3536$$

7.88 $n = 250, x = 100, \ \hat{p} = \dfrac{x}{n} = \dfrac{100}{250} = .4$

a. The standard error to estimate population proportion of first timers

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{.4(1-.4)}{250}} = .03098$$

b. Since no confidence level is specified, we find the sampling error (Margin of Error) for a 95% confidence interval.

$ME = 1.96\ (.03098) = .0607$

c. For a 92% confidence interval,

$ME = 1.75\ (.03098) = .05422$

92% confidence interval for estimating the proportion of repeat fans, $.6 \pm .05422$ giving $.5458$ up to $.6542$.

7.90 $n = 500$ motor vehicle registrations, 200 were mailed, 160 paid in person, remainder paid online.

a. 90% confidence interval to estimate the population proportion to pay for vehicle registration renewals in person.

Test and CI for One Proportion

```
Test of p = 0.5 vs p not = 0.5
Sample      X     N   Sample p              90% CI
1         160   500   0.320000    (0.285686, 0.354314)
```

The 90% confidence interval is from 28.5686% up to 35.4314%

b. 95% confidence interval to estimate the population proportion of online renewals.

Test and CI for One Proportion

```
Test of p = 0.5 vs p not = 0.5
Sample      X     N   Sample p              95% CI
1         140   500   0.280000    (0.240644, 0.319356)
```

The 95% confidence interval is from 24.0644% up to 31.9356%

7.92 From the data in 7.90, find the confidence level if the interval extends from 23.7% up to 32.3%.
ME = ½ the width of the confidence interval = (.323 – .237)/2 = .086 / 2 = .043 and $\hat{p} = .28$

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.043 = z_{\alpha/2}\sqrt{\frac{.28(1-.28)}{500}}$$

solving for z: $z_{\alpha/2} = 2.14$

Area from the z-table = (.5 – .0162) × 2 = .4838 × 2 = .9676. The confidence level is 96.76%

7.94 The 98% confidence interval of the mean age of online renewal users. $n =$ 460, sample mean = 42.6, $s = 5.4$.

$$\bar{x} \pm t_{\alpha/2}\, {s}/{\sqrt{n}} = 42.6 \pm 2.33\left({5.4}/{\sqrt{460}}\right) = 42.6 \pm .58664 = 42.0134 \text{ up to } 43.18664$$

7.96 a. $\hat{\sigma}_{\bar{x}}^2 = \dfrac{(149.92)^2}{50}\left(\dfrac{272-50}{272-1}\right) = 368.242$

99% confidence interval for the population mean:

$492.36 \pm 2.68\sqrt{368.242}$

440.9319 up to 543.7881

b. $N\hat{\sigma}_{\bar{x}} = (272)(\sqrt{368.242}) = 5219.577$

95% confidence interval for the population total:

$(272 \times 492.36) \pm 2.01(5219.577)$

123430.57 up to 144413.27

c. The 90% interval is narrower; the t-score would decline to 1.677

7.98. $N = 20, \quad n = 10, \quad \bar{x} = 257, \quad s = 37.16, \quad t_{9,.05} = 1.833$

$$\hat{\sigma}^2_{\bar{x}} = \frac{(37.16)^2}{10}\left(\frac{20-10}{20-1}\right) = 72.677$$

a. 90% confidence interval for the average number of new prescriptions:
$257 \pm 1.833\sqrt{72.677} = 241.3735$ up to 272.6265, assuming that the population of the number of new prescriptions written for the new drug is normal.

b. Reducing ME in part (a) by half = 15.6265/2 = 7.8133.

We know that $n_0 = \dfrac{z^2_{\alpha/2}\sigma^2}{ME^2} = \dfrac{(1.64^2)(72.677^2)}{7.8133^2} = 3.202$

$$n = \frac{n_0 N}{n_0 + (N-1)} = \frac{3.202 \times 20}{3.202 + (20-1)} = 2.88 \approx 3$$

7.100 $\sigma_{\bar{x}} = \dfrac{2000}{1.645} = 1215.8$, $n = \dfrac{328(12000)^2}{327(1215.8)^2 + (12000)^2} = 75.28$. Take 76 observations

7.102 We know that $n_0 = \dfrac{z^2_{\alpha/2}\sigma^2}{ME^2}$, where $ME = z_{\alpha/2}\sigma_{\bar{X}}$. This gives

$$n_0 = \frac{z^2_{\alpha/2}\sigma^2}{(z_{\alpha/2}\sigma_{\bar{X}})^2} = \frac{\sigma^2}{\sigma^2_{\bar{X}}} \text{ or } \sigma^2 = n_0\sigma^2_{\bar{X}}.$$

Substituting this into $\dfrac{N\sigma^2}{(N-1)\sigma^2_{\bar{X}} + \sigma^2}$ gives

$$\frac{N(n_0\sigma^2_{\bar{X}})}{(N-1)\sigma^2_{\bar{X}} + n_0\sigma^2_{\bar{X}}} = \frac{\sigma^2_{\bar{X}} \cdot Nn_0}{\sigma^2_{\bar{X}}[(N-1) + n_0]} = \frac{n_0 N}{n_0 + (N-1)}.$$

Chapter 8:

Estimation: Additional Topics

8.2 Two normally distributed populations based on dependent samples of n=5 observations

a. Margin of error for 90% confidence level

Paired T-Test and CI: Before_Ex8.2, After_Ex8.2

```
Paired T for Before_Ex8.2 - After_Ex8.2
                 N       Mean     StDev   SE Mean
Before_Ex8.2     5     8.4000    2.6077    1.1662
After_Ex8.2      5    10.2000    3.1145    1.3928
Difference       5   -1.80000   0.83666   0.37417

90% CI for mean difference: (-2.59766, -1.00234)
```

$$ME = t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}} = ME = 2.132\left(\frac{.83666}{\sqrt{5}}\right) = .79772$$

b. UCL and LCL for a 90% confidence interval

UCL = –1.00234, LCL = –2.59766

c. Width of a 95% confidence interval

$$width = 2\left[ME = t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}\right] = 2\left[2.776\left(\frac{.83666}{\sqrt{5}}\right)\right] = 2.07737$$

8.4 Let X = Without Passive Solar; Y = With Passive Solar; $d_i = x_i - y_i$

$n = 10,\quad \sum d_i = 373,\quad \bar{d} = 37.3,\quad t_{9,.05} = 1.833$

$$s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}} = \sqrt{2806.1/9} = 17.6575$$

90% confidence interval: $37.3 \pm 1.833(17.6575)/\sqrt{10}$

$27.0649 < \mu_x - \mu_y < 47.5351$

8.6 Using independent random samples, find a 90% confidence interval estimate of the difference in the means of the two populations

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

$$= (400 - 360) \pm 1.645\sqrt{\frac{20^2}{64} + \frac{25^2}{36}}$$

$= 40 \pm 7.9933$

$= 32.0067$ up to 47.9933

8.8 Assuming equal population variances, the number of degrees of freedom for each of the following:

a. degrees of freedom = $n_x + n_y - 2 = 16 + 9 - 2 = 23$

b. degrees of freedom = $n_x + n_y - 2 = 12 + 14 - 2 = 24$

c. degrees of freedom = $n_x + n_y - 2 = 20 + 8 - 2 = 26$

8.10 Assuming unequal population variances, the number of degrees of freedom for each of the following:

a. $$v = \frac{\left[\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)\right]^2}{\left(\frac{s_1^2}{n_1}\right)^2 / (n_1 - 1) + \left(\frac{s_2^2}{n_2}\right)^2 / (n_2 - 1)}$$

$$v = \frac{\left[\left(\frac{5}{16}\right) + \left(\frac{36}{4}\right)\right]^2}{\left(\frac{5}{16}\right)^2 / (16 - 1) + \left(\frac{36}{4}\right)^2 / (4 - 1)} \approx 3$$

b. $$v = \frac{\left[\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)\right]^2}{\left(\frac{s_1^2}{n_1}\right)^2 / (n_1 - 1) + \left(\frac{s_2^2}{n_2}\right)^2 / (n_2 - 1)}$$

$$v = \frac{\left[\left(\frac{30}{9}\right) + \left(\frac{4}{16}\right)\right]^2}{\left(\frac{30}{9}\right)^2 / (9 - 1) + \left(\frac{4}{16}\right)^2 / (16 - 1)} \approx 9$$

8.12 Let X = machine A and Y = machine B.

$\bar{x} = 130,\ \sigma_x = 8.4, n_x = 40;\ \bar{y} = 120, \sigma_y = 11.3, n_y = 36$

95% confidence interval for the difference in means parts produced per hour by the two machines:

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} = (130 - 120) \pm 1.96\sqrt{\frac{8.4^2}{40} + \frac{11.3^2}{36}} = 10 \pm 4.5169,$$

5.4831 up to 14.5169

8.14

Descriptive Statistics: Machine 1, Machine 2

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Machine 1	100	520.95	518.75	520.52	9.45	0.95
Machine 2	100	513.75	514.05	513.91	5.49	0.55

Variable	Minimum	Maximum	Q1	Q3
Machine 1	504.70	544.80	513.80	527.28
Machine 2	496.50	527.00	510.33	517.68

95% confidence level: assuming normal populations and unequal population variances $(520.95-513.75)\pm(1.96)\sqrt{\frac{(9.45)^2}{100}+\frac{(5.49)^2}{100}}=$ 5.0579 up to 9.3421

8.16 $n_1=200,\quad \bar{x}=.517,\quad s_1=.148,\quad z_{.005}=2.58$

$n_2=400,\quad \bar{y}=.489,\quad s_2=.158$

99% confidence interval assuming equal variances

$$(.517-.489)\pm(2.58)(.1547)\sqrt{\frac{1}{200}+\frac{1}{400}}=-.0066 \text{ up to } .0626$$

8.18 Margin of error, assuming 95% confidence level:

a. $ME=z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}=1.96\sqrt{\frac{.75(1-.75)}{280}+\frac{.68(1-.68)}{320}}=$ .072

b. $ME=z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}=1.96\sqrt{\frac{.51(1-.51)}{210}+\frac{.48(1-.48)}{200}}=$ .0968

8.20

$n_x=120,\quad \hat{p}_x=\frac{x}{n}=\frac{85}{120}=.7083,\quad n_y=163,\quad \hat{p}_y=\frac{y}{n}=\frac{78}{163}=.4785,\quad z_{.01}=2.33$

98% confidence interval: $(\hat{p}_x-\hat{p}_y)\pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x}+\frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$

$=(.7083-.4785)\pm(2.33)\sqrt{\frac{(.7083)(.2917)}{120}+\frac{(.4785)(.5215)}{163}}$

$=.2298\pm.13288=.0969$ up to $.3627$

8.22 $\hat{p}_{freshmen} = 80/138 = .5797, \quad \hat{p}_{sophs} = 73/96 = .7604$

$$(.5797 - .7604) \pm (1.96)\sqrt{\frac{(.5797)(.4203)}{138} + \frac{(.7604)(.2396)}{96}}$$

$= -.1807 \pm .1186 = -.2993$ up to $-.0621$

8.24 $n_x = 510, \quad \hat{p}_x = .6275, \quad n_y = 332, \quad \hat{p}_y = .6024, \quad z_{.05} = 1.645$

$$(.6275 - .6024) \pm (1.645)\sqrt{\frac{(.6275)(.3725)}{510} + \frac{(.6024)(.3976)}{332}}$$

$.0251 \pm .0565 = -.0314$ up to $.0816$

8.26 Independent random samples from two normally distributed populations. Assuming unequal variances, the 90% confidence interval:

$$(\bar{X} - \bar{Y}) \pm t_{(v,\alpha/2)}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} =$$

$$\text{where } v = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\frac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$

$$v = \frac{\left[\left(\frac{10^2}{15}\right) + \left(\frac{40^2}{13}\right)\right]^2}{\left(\frac{10^2}{15}\right)^2 /(15-1) + \left(\frac{40^2}{13}\right)^2 /(13-1)} = 13.30176 \approx 13$$

$$(400 - 360) \pm 1.771\sqrt{\frac{10^2}{15} + \frac{40^2}{13}} = 40 \pm 20.1726 = 19.8274 \text{ up to } 60.1726$$

8.28 $n_x = 16, \quad \bar{x} = 625{,}000, \quad s_x = 80{,}000$

$n_y = 10, \quad \bar{y} = 608{,}000, \quad s_y = 73{,}000, \quad t_{24,.05} = 1.711$

90% confidence interval:

$$(\bar{x} - \bar{y}) \pm t_{n_1+n_2-2,\alpha/2}s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \quad \text{where } s_p = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}$$

$$s_p = \sqrt{\frac{(16-1)80{,}000^2 + (10-1)73{,}000^2}{16+10-2}} = 77{,}449.17688$$

$$(625{,}000 - 608{,}000) \pm 1.711(77{,}449.17688)\sqrt{\frac{1}{16} + \frac{1}{10}}$$

$= 17{,}000 \pm 53{,}418.72262 = -36{,}418.72262$ up to $70{,}418.72262$

8.30 90% confidence interval for the difference in mean amount spent on textbooks for accounting majors vs. management majors.
Assuming independent random samples taken from each type of major from normal populations with a common variance.

$n_x = 40, \quad \bar{x} = 340, \quad s_x = 20$

$n_y = 50, \quad \bar{y} = 285, \quad s_y = 30, \quad t_{88,.05} = 1.662$

$$(\bar{x}-\bar{y}) \pm t_{n_1+n_2-2,\alpha/2} s_p \sqrt{\frac{1}{n_x}+\frac{1}{n_y}} \text{ where } s_p = \sqrt{\frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}}$$

$$s_p = \sqrt{\frac{(40-1)20^2+(50-1)30^2}{40+50-2}} = 26.0463$$

$$(340-285) \pm 1.662(26.0463)\sqrt{\frac{1}{40}+\frac{1}{50}} = 55 \pm 9.1830 = 45.8170 \text{ to } 64.1830$$

8.32 98% confidence interval for student pair:

Paired T-Test and CI: COURSE, NO COURSE

```
Paired T for COURSE - NO COURSE
                N       Mean      StDev   SE Mean
COURSE          6   70.6667    16.0333    6.5456
NO COURSE       6   66.1667    14.1904    5.7932
Difference      6   4.50000    4.13521   1.68819

98% CI for mean difference: (-1.18066, 10.18066)
```

8.34 Assuming that the population variances are equal, the 95% confidence interval estimate of the difference in the mean HEI–2005 scores between male and female participants at the time of their first interview.

Two-Sample T-Test and CI: HEI2005_male, HEI2005_female

```
Two-sample T for HEI2005_male vs HEI2005_female

                   N  Mean  StDev  SE Mean
HEI2005_male     2139  50.6   13.8     0.30
HEI2005_female   2321  53.3   14.4     0.30

Difference = mu (HEI2005_male) - mu (HEI2005_female)
Estimate for difference:  -2.769
95% CI for difference:  (-3.600, -1.939)
```

Chapter 9:

Hypothesis Testing: Single Population

9.2 H_0: No change in interest rates is warranted

H_1: Reduce interest rates to stimulate the economy

9.4 a. European perspective:

H_0: Genetically modified food stuffs are safe

H_1: They are not safe

b. U.S. farmer perspective:

H_0: Genetically modified food stuffs are not safe

H_1: They are safe

9.6 $H_0: T_B \geq T_G$ No difference in the total number of votes between Bush and Gore

$H_1: T_B < T_G$ Gore with more votes

9.8 A random sample of n = 25 is obtained from a population with a variance σ^2 and the sample mean is computed. Test the null hypothesis $H_0: \mu = 100$ versus the alternative $H_1: \mu > 100$ with alpha = .05. Compute the critical value $\bar{x}_c$ and state the decision rule

a. $\sigma^2 = 225$. Reject H_0 if $\bar{x} > \bar{x}_c = \mu_0 + z_\alpha \sigma/\sqrt{n} = 100 + 1.645(15)/\sqrt{25} = 104.935$

b. $\sigma^2 = 900$. Reject H_0 if $\bar{x} > \bar{x}_c = \mu_0 + z_\alpha \sigma/\sqrt{n} = 100 + 1.645(30)/\sqrt{25} = 109.87$

c. $\sigma^2 = 400$. Reject H_0 if $\bar{x} > \bar{x}_c = \mu_0 + z_\alpha \sigma/\sqrt{n} = 100 + 1.645(20)/\sqrt{25} = 106.58$

d. $\sigma^2 = 600$. Reject H_0 if $\bar{x} > \bar{x}_c = \mu_0 + z_\alpha \sigma/\sqrt{n} = 100 + 1.645(24.4949)/\sqrt{25} = 108.0588$

9.10 A random sample of n = 25, variance = σ^2 and the sample mean is = 70. Consider the null hypothesis $H_0: \mu = 80$ versus the alternative $H_1: \mu < 80$. Compute the p-value

a. $\sigma^2 = 225$. $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{70-80}{15/\sqrt{25}} = -3.33$. $p-value = P(z_p < -3.33) = .0004$

b. $\sigma^2 = 900$. $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{70-80}{30/\sqrt{25}} = -1.67$. $p-value = P(z_p < -1.67) = .0475$

c. $\sigma^2 = 400$. $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{70-80}{20/\sqrt{25}}$ = -2.50. $p-value = P(z_p < -2.50) = .0062$

d. $\sigma^2 = 600$. $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{70-80}{24.4949/\sqrt{25}}$ = -2.04. $p-value = P(z_p < -2.04) =$

.0207

9.12 $H_0: \mu \geq 50$; $H_1: \mu < 50$; reject H_0 if $Z_{.10} < -1.28$

$Z = \dfrac{48.2-50}{3/\sqrt{9}}$ = -1.8, therefore, Reject H_0 at the 10% level.

9.14 Test $H_0: \mu \leq 100$; $H_1: \mu > 100$, using n = 25 and alpha = .05

a. $\bar{x} = 106, s = 15$. Reject if $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$ $\dfrac{106-100}{15/\sqrt{25}}$ = 2.00. Since 2.00 is greater than the critical value of 1.711, there is sufficient evidence to reject the null hypothesis.

b. $\bar{x} = 104, s = 10$. Reject if $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$, $\dfrac{104-100}{10/\sqrt{25}}$ = 2.00. Since 2.00 is greater than the critical value of 1.711, there is sufficient evidence to reject the null hypothesis.

c. Assuming a one-tailed test, $\bar{x} = 95, s = 10$. Reject if $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$, $\dfrac{95-100}{10/\sqrt{25}}$ = -2.50. Since -2.50 is less than the critical value of 1.711, there is insufficient evidence to reject the null hypothesis.

d. Assuming a one-tailed upper test, $\bar{x} = 92, s = 18$. Reject if $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$,

$\dfrac{92-100}{18/\sqrt{25}}$ = -2.22. Since -2.22 is less than the critical value of 1.711, there is insufficient evidence to reject the null hypothesis.

9.16 $H_0: \mu \geq 3; H_1: \mu < 3;$

$t = \dfrac{2.4-3}{1.8/\sqrt{100}}$ = -3.33, p-value is < .005. Reject H_0 at any common level of alpha

9.18 $H_0: \mu = 0; H_1: \mu \neq 0;$

$t = \dfrac{.078-0}{.201/\sqrt{76}}$ = 3.38, p-value is < .010. Reject H_0 at any common level of alpha

9.20 $H_0: \mu = 0; H_1: \mu < 0;$

$t = \dfrac{-2.91 - 0}{11.33/\sqrt{170}}$ = -3.35, p-value is < .005. Reject H_0 at any common level of alpha.

9.22 a. No, the 95% confidence level provides for 2.5% of the area in either tail. This does not correspond to a one-tailed hypothesis test with an alpha of 5% which has 5% of the area in one of the tails.
b. Yes.

9.24 $H_0: \mu = 20; H_1: \mu \neq 20;$ reject H_0 if $|t_{8,.05/2}| > 2.306$

$t = \dfrac{20.3556 - 20}{.6126/\sqrt{9}}$ = 1.741, therefore, do not reject H_0 at the 5% level

9.26 The population values must be assumed to be normally distributed.

$H_0: \mu \geq 50; H_1: \mu < 50;$ reject H_0 if $t_{19,.05} < -1.729$

$t = \dfrac{41.3 - 50}{12.2/\sqrt{20}}$ = -3.189, therefore, reject H_0 at the 5% level

9.28 A random sample is obtained to test the null hypothesis of the proportion of women who said yes to a new shoe model. $H_0: p \leq .25; H_1: p > .25;$. What value of the sample proportion is required to reject the null hypothesis with alpha = .03?

a. n = 400. Reject H_0 if $\hat{p} > \hat{p}_c = p_0 + z_\alpha \sqrt{p_0(1-p_0)/n}$ = .25 +1.88 $\sqrt{(.25)(1-.25)/400}$ = .2907

b. n = 225. Reject H_0 if $\hat{p} > \hat{p}_c = p_0 + z_\alpha \sqrt{p_0(1-p_0)/n}$ = .25 +1.88 $\sqrt{(.25)(1-.25)/225}$ = .30427

c. n = 625. Reject H_0 if $\hat{p} > \hat{p}_c = p_0 + z_\alpha \sqrt{p_0(1-p_0)/n}$ = .25 +1.88 $\sqrt{(.25)(1-.25)/625}$ = .28256

d. n = 900. Reject H_0 if $\hat{p} > \hat{p}_c = p_0 + z_\alpha \sqrt{p_0(1-p_0)/n}$ = .25 +1.88 $\sqrt{(.25)(1-.25)/900}$ = .2771

9.30 $H_0: p \leq .25; H_1: p > .25;$

$z = \dfrac{.2908 - .25}{\sqrt{(.25)(.75)/361}}$ = 1.79, *p*-value = 1 – $F_Z(1.79)$ = 1 - .9633 = .0367

Therefore, reject H_0 at alpha greater than 3.67%

9.32 $H_0: p = .5; H_1: p \neq .5;$

$z = \frac{.45-.5}{\sqrt{(.5)(.5)/160}}$ = -1.26, p-value = 2[1 – F_Z(1.26)] = 2[1 – .8962] = .2076

The probability of finding a random sample with a sample proportion this far or further from .5 if the null hypothesis is really true is .2076

9.34 $H_0: p = .5; H_1: p > .5;$

$z = \frac{.56-.5}{\sqrt{(.5)(.5)/50}}$ = .85, p-value = 1 – F_Z(.85) = 1 – .8023 = .1977

Therefore, reject H_0 at alpha levels in excess of 19.77%

9.36 $H_0: p \geq .75; H_1: p < .75;$

$z = \frac{.6931-.75}{\sqrt{(.75)(.25)/202}}$ = -1.87, p-value = 1 – F_Z(1.87) = 1 – .9693 = .0307

Therefore, reject H_0 at alpha levels in excess of 3.07%

9.38 What is the probability of Type II error if the actual proportion is

a. $P = .52$. $\beta = P(.46 \leq \hat{p} \leq .54 \mid p = p^*) = P\left[\frac{.46-p^*}{\sqrt{\frac{p^*(1-p^*)}{n}}} \leq z \leq \frac{.54-p^*}{\sqrt{\frac{p^*(1-p^*)}{n}}}\right]$

$= P\left[\frac{.46-.52}{\sqrt{\frac{.52(1-.52)}{600}}} \leq z \leq \frac{.54-.52}{\sqrt{\frac{.52(1-.52)}{600}}}\right]$ = P(-2.94 $\leq$ z $\leq$ .98) = .4984 + .3365 = .8349

b. $P = .58$. $\beta = P(.46 \leq \hat{p} \leq .54 \mid p = p^*) = P\left[\frac{.46-.58}{\sqrt{\frac{.58(1-.58)}{600}}} \leq z \leq \frac{.54-.58}{\sqrt{\frac{.58(1-.58)}{600}}}\right]$

= P(-5.96 $\leq$ z $\leq$ -1.99) = .5000 – .4767 = .0233

c. $P = .53$. $\beta = P(.46 \leq \hat{p} \leq .54 \mid p = p^*) = P\left[\frac{.46-.53}{\sqrt{\frac{.53(1-.53)}{600}}} \leq z \leq \frac{.54-.53}{\sqrt{\frac{.53(1-.53)}{600}}}\right]$

= P(-3.44 $\leq$ z $\leq$ .49) = .4997 + .1879 = .6876

d. $P = .48$. $\beta = P(.46 \le \hat{p} \le .54 \mid p = p^*) = P\left[\dfrac{.46-.48}{\sqrt{\dfrac{.48(1-.48)}{600}}} \le z \le \dfrac{.54-.48}{\sqrt{\dfrac{.48(1-.48)}{600}}}\right]$

$= P(-.98 \le z \le 2.94) = .3365 + .4984 = .8349$

e. $P = .43$. $\beta = P(.46 \le \hat{p} \le .54 \mid p = p^*) = P\left[\dfrac{.46-.43}{\sqrt{\dfrac{.43(1-.43)}{600}}} \le z \le \dfrac{.54-.43}{\sqrt{\dfrac{.43(1-.43)}{600}}}\right]$

$= P(1.48 \le z \le 5.44) = .5000 - .4306 = .0694$

9.40 a. H_0 is rejected when $\dfrac{\bar{X}-3}{.4/\sqrt{64}} > 1.645$ or when $\bar{X} > 3.082$. Since the sample mean is 3.07% which is less than the critical value, the decision is do not reject the null hypothesis.

b. The $\beta = P(Z < \dfrac{3.082-3.1}{.4/\sqrt{64}}) = 1 - F_Z(.36) = .3594$. Power of the test = $1 - \beta =$ .6406

9.42 H_0 is rejected when $\dfrac{p-.5}{\sqrt{.25/802}} < -1.28$ or when $p < .477$

The power of the test = $1 - \beta = 1 - P(Z > \dfrac{.477-.45}{\sqrt{(.45)(.55)/802}}) = 1 - P(Z > 1.54) =$ .9382

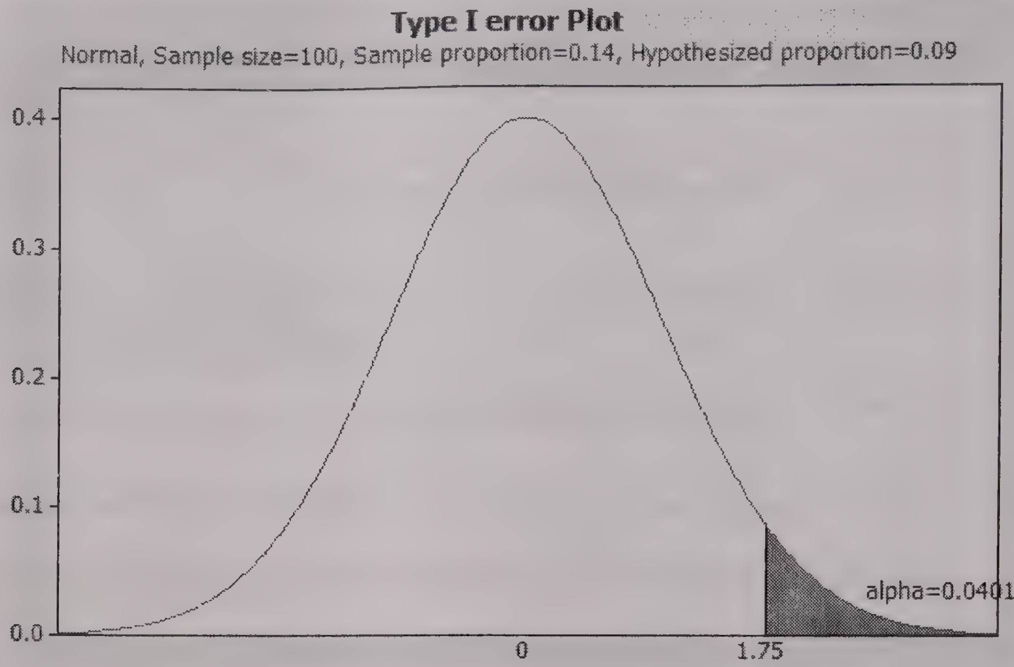
9.44 a. H_0 is rejected when $-1.645 > \dfrac{p-.5}{\sqrt{.25/199}} > 1.645$ or when $.442 > p > .558$.

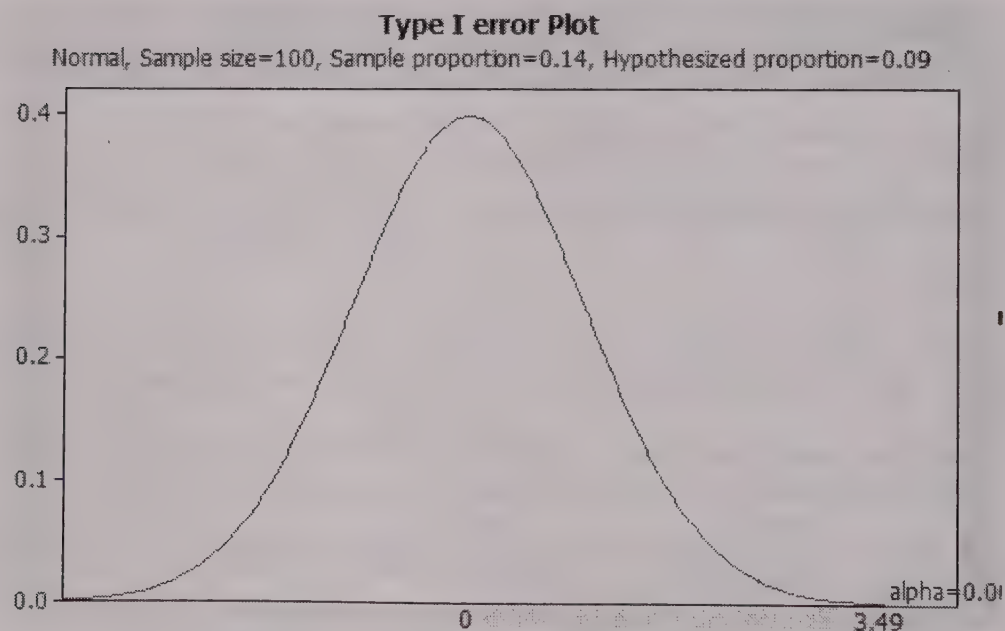
Since the sample proportion is .5226 which is within the critical values. The decision is that there is insufficient evidence to reject the null hypothesis.

b. $\beta = P(\dfrac{.442-.6}{\sqrt{(.6)(.4)/199}} < Z < \dfrac{.558-.6}{\sqrt{(.6)(.4)/199}}) = 1 - P(-4.55 < Z < -1.21) = .1131$

9.46 a. $\alpha = P(Z > \dfrac{.14-.09}{\sqrt{(.09)(.91)/100}}) = P(Z > 1.75) = 0.0401$

b. $\alpha = P(Z > \dfrac{.14-.09}{\sqrt{(.09)(.91)/400}}) = P(Z > 3.49) = 0.0002$. The smaller probability of a Type I error is due to the larger sample size which lowers the standard error of the mean.





c. $\beta = P(Z < \frac{.14-.20}{\sqrt{(.2)(.8)/100}}) = P(Z < -1.5) = .0668$

d. i) lower, ii) higher

9.48 $H_0: \sigma^2 \le 500; H_1: \sigma^2 > 500$; reject H_0 if $\chi^2_{(7,.10)} > 12.02$

$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{7(933.982)}{500} = 13.0757$, Therefore, reject H_0 at the 10% level

9.50 $H_0: \sigma^2 = 300; H_1: \sigma^2 \neq 300;$

$\chi^2 = \frac{29(480)}{300} = 46.4$, p-value = .0214. Reject H_0 at the 5% level

9.52 $H_0: \sigma \geq 18.2; H_1: \sigma < 18.2;$

$\chi^2 = \frac{24(15.3)^2}{(18.2)^2} = 16.961.$

Do not reject H_0 at the 10% level since $\chi^2 > 15.66 = \chi^2_{(24,.10)}$

9.54 The p-value indicates the likelihood of getting the sample result at least as far away from the hypothesized value as the one that was found, assuming that the distribution is really centered on the null hypothesis. The smaller the p-value, the stronger the evidence against the null hypothesis.

9.56

a. False. The significance level is the probability of making a Type I error – falsely rejecting the null hypothesis when in fact the null is true.
b. True
c. True
d. False. The power of the test is the ability of the test to correctly reject a false null hypothesis.
e. False. The rejection region is farther away from the hypothesized value at the 1% level than it is at the 5% level. Therefore, it is still possible to reject at the 5% level but not at the 1% level.
f. True
g. False. The p-value tells the strength of the evidence against the null hypothesis.

9.58 a. $\alpha = P(Z < \frac{776-800}{120/\sqrt{100}}) = P(Z < -2) = .0228$

b. $\beta = P(Z > \frac{776-740}{120/\sqrt{100}}) = P(Z > 3) = .0014$

c. i) smaller ii) smaller

d. i) smaller ii) larger

9.60 $H_0: p = .5; H_1: p \neq .5;$

$z = \frac{.4808-.5}{\sqrt{(.5)(.5)/104}} = -.39$, p-value = $2[1-F_Z(.39)] = 2[1-.6517] = .6966$

Therefore, reject H_0 at levels in excess of 69.66%

9.62 $H_0: p \le .25; H_1: p > .25$; reject H_0 if $z_{.05} > 1.645$

$z = \frac{.3333 - .25}{\sqrt{(.25)(.75)/150}} = 2.356$, therefore, reject H_0 at the 5% level

9.64 Cost Model where W = Total Cost: W = 1,000 + 5X

$\mu_W = 1,000 + 5(400) = 3,000$

$\sigma^2_W = (5)^2(625) = 15,625, \quad \sigma_W = 125, \quad \sigma_{\overline{W}} = \frac{125}{\sqrt{25}} = 25$

$H_0: W \le 3000; \quad H1: W > 3000;$

Using the test statistic criteria: (3050 – 3000)/25 = 2.00 which yields a p-value of .0228, therefore, reject H_0 at the .05 level.

Using the sample statistic criteria: $\overline{X}_{crit} = 3,000 + (25)(1.645) = 3041.1$, $\overline{X}_{calc} = 3,050$, since $\overline{X}_{calc} = 3,050 > \overline{X}_{crit} = 3041.1$, therefore, reject H_0 at the .05 level.

9.66 Per capita consumption of fruits and vegetables

$\overline{x} = 172.79$; s = 19.254

$H_0: \mu \le 170; H_1: \mu > 170;$. Reject if $\frac{\overline{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$

$\frac{172.79 - 170}{19.254/\sqrt{3108}} = 8.075$ Since 8.075 is greater than the critical value of 1.645, there is sufficient evidence to reject the null hypothesis.

Per capita consumption of snack foods

$\overline{x} = 114.11$; s = 9.541

$H_0: \mu \ge 114; H_1: \mu < 114;$. Reject if $\frac{\overline{x} - \mu_0}{s/\sqrt{n}} < t_{n-1,\alpha}$

$\frac{114.11 - 114}{9.541/\sqrt{3108}} = 0.66$. Since 0.66 is greater than the critical value of -1.645, there is no sufficient evidence to reject the null hypothesis.

Per capita consumption of soft drinks

$\overline{x} = 66.81$; s = 7.5

$H_0: \mu \ge 65; H_1: \mu < 65;$. Reject if $\frac{\overline{x} - \mu_0}{s/\sqrt{n}} < t_{n-1,\alpha}$

$\frac{66.81 - 65}{7.5/\sqrt{3108}} = 13.487$. Since 13.487 is greater than the critical value of -1.645, there is no sufficient evidence to reject the null hypothesis.

Per capita consumption of meat

$\bar{x} = 70.38$; s = 12.694

$H_0: \mu \le 70; H_1: \mu > 70;$. Reject if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$

$\frac{70.38-70}{12.694/\sqrt{3108}} = 1.69$. Since 1.69 is greater than the critical value of 1.645, there is sufficient evidence to reject the null hypothesis.

9.68 $H_0: \mu \le 40, H_1: \mu > 40; \bar{X} = 49.73 > 42.86$ reject H_0

One-Sample T: Salmon Weight

```
Test of mu = 40 vs mu > 40

Variable           N       Mean      StDev   SE Mean
Salmon Weigh      39      49.73      10.60      1.70

Variable       95.0% Lower Bound          T         P
Salmon Weigh               46.86       5.73     0.000
```

At the .05 level of significance we have strong enough evidence to reject Ho that the true mean weight of salmon is no different than 40 in favor of Ha that the true mean weight is significantly greater than 40.

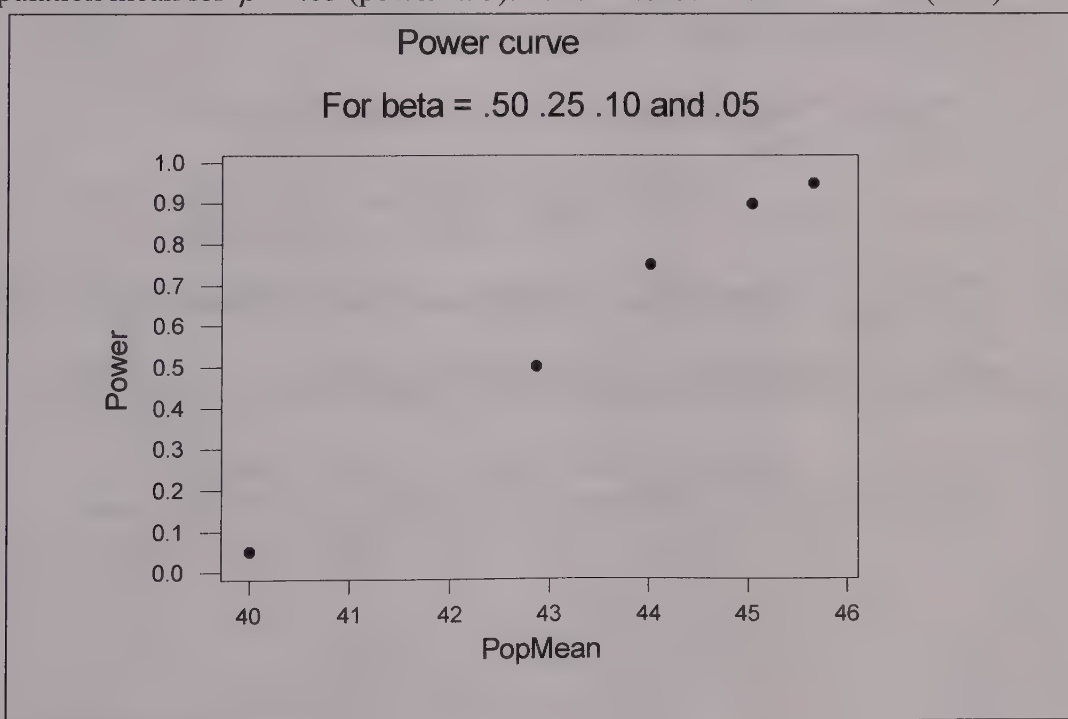
$\bar{X}_{crit} = Ho + t_{crit}(S_{\bar{x}})$: 40 + 1.686(1.70) = 42.8662

Population mean for β = .50 (power=.50): tcrit = 0.0: 42.8662 + 0.0(1.70) = 42.8662

Population mean for β = .25 (power=.75): tcrit = .681: 42.8662 + .681(1.70) = 44.0239

Population mean for β = .10 (power=.90): tcrit = 1.28: 42.8662 + 1.28(1.70) = 45.0422

Population mean for β = .05 (power=.95): tcrit = 1.645: 42.8662 + 1.645(1.70) = 45.6627



9.70 a. Assume that the population is normally distributed

One-Sample T: Grams:

```
Test of mu = 5 vs mu not = 5
Variable          N       Mean      StDev   SE Mean
Grams:11-34      12     4.9725     0.0936    0.0270

Variable              95.0% CI             T        P
Grams:11-34     ( 4.9130,   5.0320)    -1.02    0.331
```

$\bar{x} = 4.9725; s = .0936$, $H_0: \mu = 5; H_1: \mu \neq 5;$ reject H_0 if $|t_{(11,.025)}| > 2.201$

$t = \dfrac{4.9725 - 5}{.0936/\sqrt{12}} = -1.018$. Do not reject H_0 at the 5% level

b. Assume that the population is normally distributed

$H_0: \sigma = .025; H_1: \sigma > .025;$ reject H_0 if $\chi^2_{(11,.05)} > 19.68$

$\chi^2 = \dfrac{11(.0936)^2}{(.025)^2} = 154.19$. Therefore, reject H_0 at the 5% level

9.72 Obesity rates of adults in the U.S. population.

$\bar{x} = 28.29$; s = 3.625

$H_0: \mu \leq 28; H_1: \mu > 28;$. Reject if $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$

$\dfrac{28.29 - 28}{3.625/\sqrt{3140}} = 4.461$. Since 4.461 is greater than the critical value of 1.645, there is sufficient evidence to reject the null hypothesis.

Low-income preschool obesity rate in the U.S. population

$\bar{x} = 14.19$; s = 3.716

$H_0: \mu \leq 13; H_1: \mu > 13;$. Reject if $\dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$

$\dfrac{14.19 - 13}{3.716/\sqrt{2691}} = 16.589$. Since 16.589 is greater than the critical value of 1.645, there is sufficient evidence to reject the null hypothesis.

9.74 Mean weights of Men in the first interview

$\bar{x} = 27.98$; s = 5.468

$H_0: \mu \geq 30; H_1: \mu < 30;$. Reject if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$

$\frac{27.98-30}{5.468/\sqrt{2108}}$ = -16.985. Since -16.985 is smaller than the critical value of -1.645, there is sufficient evidence to reject the null hypothesis.

Mean weights of Men in the second interview

$\bar{x} = 28.07$; s = 5.447

$H_0: \mu \geq 30; H_1: \mu < 30;$. Reject if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$

$\frac{28.07-30}{5.447/\sqrt{1925}}$ = -15.562. Since -15.562 is smaller than the critical value of -1.645, there is sufficient evidence to reject the null hypothesis.

There is no difference in the results obtained from the first and second Interviews for men.

Mean weights of Women in the first interview

$\bar{x} = 28.93$; s = 7.022

$H_0: \mu \geq 30; H_1: \mu < 30;$. Reject if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$

$\frac{28.93-30}{7.022/\sqrt{2274}}$ = -7.238. Since -7.238 is smaller than the critical value of -1.645, there is sufficient evidence to reject the null hypothesis.

Mean weights of Women in the second interview

$\bar{x} = 29.02$; s = 7.071

$H_0: \mu \geq 30; H_1: \mu < 30;$. Reject if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$

$\frac{29.02-30}{7.071/\sqrt{2132}}$ = -6.385. Since -6.385 is smaller than the critical value of -1.645, there is sufficient evidence to reject the null hypothesis.

There is no difference in the results obtained from the first and second Interviews for women.

9.76 Mean weights of White people in the first interview

$\bar{x} = 27.85$; s = 6.065

$$H_0: \mu \geq 30; H_1: \mu < 30;. \text{ Reject if } \frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$$

$\frac{27.85-30}{6.065/\sqrt{2357}}$ = -17.18. Since -17.18 is smaller than the critical value of -1.645, there is sufficient evidence to reject the null hypothesis.

Mean weights of White people in the second interview

$\bar{x} = 27.92$; s = 6.117

$$H_0: \mu \geq 30; H_1: \mu < 30;. \text{ Reject if } \frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$$

$\frac{27.92-30}{6.117/\sqrt{2213}}$ = -16.019. Since -16.019 is smaller than the critical value of -1.645, there is sufficient evidence to reject the null hypothesis.

There is no difference in the results obtained from the first and second Interviews for White people.

9.78 Mean weights of people who have been diagnosed with high blood pressure in the first interview

$\bar{x} = 30.15$; s = 6.613

$$H_0: \mu \geq 30; H_1: \mu < 30;. \text{ Reject if } \frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$$

$\frac{30.15-30}{6.613/\sqrt{1522}}$ = 0.913. Since 0.913 is greater than the critical value of -1.645, there is no sufficient evidence to reject the null hypothesis.

Mean weights of people who have been diagnosed with high blood pressure in the second interview

$\bar{x} = 30.29$; s = 6.651

$$H_0: \mu \geq 30; H_1: \mu < 30;. \text{ Reject if } \frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$$

$\frac{30.29-30}{6.651/\sqrt{1420}}$ = 1.656. Since 1.656 is greater than the critical value of -1.645, there is no sufficient evidence to reject the null hypothesis.

There is no difference in the results obtained from the first and second Interviews for people diagnosed with high blood pressure.

Chapter 10:

Hypothesis Testing: Additional Topics

10.2 n = 25 paired observations with standard deviation of the difference between sample means = 25. Can you reject the null hypothesis at an alpha of .05 if

a. The sample means are 56 and 50, $H_0: \mu_1 - \mu_2 \geq 0; H_1: \mu_1 - \mu_2 < 0;$

$$t = \frac{6-0}{25/\sqrt{25}} = 1.2, \text{ p-value} = 0.879$$

Do not reject H_0 at alpha of .05

Paired T-Test and CI

```
             N   Mean   StDev   SE Mean
Difference  25   6.00   25.00      5.00
95% upper bound for mean difference: 14.55
T-Test of mean difference = 0 (vs < 0): T-Value = 1.20  P-Value = 0.879
```

b. The sample means are 59 and 50, $H_0: \mu_1 - \mu_2 \geq 0; H_1: \mu_1 - \mu_2 < 0;$

$$t = \frac{9-0}{25/\sqrt{25}} = 1.8, \text{ p-value} = 0.958.$$ Do not reject H_0 at alpha of .05

Paired T-Test and CI

```
             N   Mean   StDev   SE Mean
Difference  25   9.00   25.00      5.00
95% upper bound for mean difference: 17.55
T-Test of mean difference = 0 (vs < 0): T-Value = 1.80  P-Value = 0.958
```

c. The sample means are 56 and 48, $H_0: \mu_1 - \mu_2 \geq 0; H_1: \mu_1 - \mu_2 < 0;$

$$t = \frac{8-0}{25/\sqrt{25}} = 1.60, \text{ p-value} = .939.$$ Do not reject H_0 at alpha of .05

Paired T-Test and CI

```
             N   Mean   StDev   SE Mean
Difference  25   8.00   25.00      5.00
95% upper bound for mean difference: 16.55
T-Test of mean difference = 0 (vs < 0): T-Value = 1.60  P-Value = 0.939
```

d. The sample means are 54 and 50, $H_0: \mu_1 - \mu_2 \geq 0; H_1: \mu_1 - \mu_2 < 0;$

$$t = \frac{4-0}{25/\sqrt{25}} = 0.8, \text{ p-value} = .784.$$ Do not reject H_0 at alpha of .05

Paired T-Test and CI

```
             N   Mean   StDev   SE Mean
Difference  25   4.00   25.00      5.00
95% upper bound for mean difference: 12.55
T-Test of mean difference = 0 (vs < 0): T-Value = 0.80  P-Value = 0.784
```

10.4 Let x – Initial urban home selling prices; y – Urban home selling prices over time
Urban home selling prices in Atlanta, Chicago, Dallas, and Oakland,
$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0;$

Paired T-Test and CI: Sale 1 Price, Sale 2 Price

```
Paired T for Sale 1 Price - Sale 2 Price

                 N     Mean   StDev  SE Mean
Sale 1 Price  4000    61323  119893     1896
Sale 2 Price  4000    83585  118721     1877
Difference    4000   -22262   26231      415

95% lower bound for mean difference: -22944
T-Test of mean difference = 0 (vs > 0): T-Value = -53.68  P-Value = 1.000
```

Since, p-value is equal to 1.000, we do not reject the null hypothesis.

b)
Urban home selling prices in Atlanta, $H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0;$

$t = \dfrac{-16564 - 0}{16660.979/\sqrt{1000}}$ = -31.44, p-value = 1.00. Do not reject H_0 at any levels of alpha.

```
Paired T for Sale 1 Price - Sale 2 Price

                 N    Mean  StDev  SE Mean
Sale 1 Price  1000   45451  23581      746
Sale 2 Price  1000   62015  29515      933
Difference    1000  -16564  16661      527
95% lower bound for mean difference: -17431
T-Test of mean difference = 0 (vs > 0): T-Value = -31.44  P-Value = 1.000
```

Since, p-value is equal to 1.000, we do not reject the null hypothesis.

10.6 Let x – Process 1; y – Process 2

a. Reject H_0 if $\dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}} > z_\alpha$. For $\alpha = 0.05$, $z_\alpha = z_{0.05} = 1.645$.

$$z = \frac{\bar{x} - \bar{y} - D_0}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}} = \frac{50 - 60}{\sqrt{\dfrac{900}{25} + \dfrac{1600}{28}}} = -1.04$$

Do not reject H_0 at alpha of .05.

b. Reject H_0 if $\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} > z_\alpha$. For $\alpha = 0.05, z_\alpha = z_{0.05} = 1.645$.

$$z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} = \frac{20}{\sqrt{\frac{900}{25}+\frac{1600}{28}}} = 2.07$$

Reject H_0 at alpha of .05.

c. Reject H_0 if $\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} > z_\alpha$. For $\alpha = 0.05, z_\alpha = z_{0.05} = 1.645$.

$$z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} = \frac{45-50}{\sqrt{\frac{900}{25}+\frac{1600}{28}}} = -0.52$$

Do not reject H_0 at alpha of .05.

d. Reject H_0 if $\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} > z_\alpha$. For $\alpha = 0.05, z_\alpha = z_{0.05} = 1.645$.

$$z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} = \frac{15}{\sqrt{\frac{900}{25}+\frac{1600}{28}}} = 1.55$$

Do not reject H_0 at alpha of .05.

10.8 Let x – male financial analysts; y – female financial analysts

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y > 0;$

$$z = \frac{85.8-71.5}{\sqrt{(19.13)^2/151+(12.2)^2/108}} = 7.334.$$

Reject H_0 at all common levels of alpha

10.10 Let x – students who vote; y – students who do not vote

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$

$$z = \frac{2.71-2.79}{\sqrt{(.64)^2/114+(.56)^2/123}} = -1.0207,$$

p-value = $2[1-F_Z(1.02)] = 2[1-.8461] = .3078$

Therefore, reject H_0 at levels of alpha in excess of 30.78%

10.12 Let x – prospectuses in which sales forecasts were disclosed; y – prospectuses in which sales earnings forecasts were not disclosed

Assuming both populations are normal with equal variances:

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$

$$s^2_p = \frac{69(6.14)^2 + 50(4.29)^2}{70+51-2} = 29.592247$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{3.97 - 2.86}{\sqrt{\frac{29.592247}{70} + \frac{29.592247}{51}}} = 1.108$$

Therefore, do not reject H_0 at the 10% alpha level since $1.108 < 1.645 = t_{(119,.05)}$

10.14 a. $H_0: P_x - P_y = 0; H_1: P_x - P_y < 0;$

$$\hat{p}_o = \frac{500(.42) + 600(.50)}{500+600} = .4636,\ z = \frac{.42 - .50}{\sqrt{\frac{(.4636)(1-.4636)}{500} + \frac{(.4636)(1-.4636)}{600}}}$$

$= -2.65$ p-value = .004. Therefore, reject H_0 at all common levels of alpha

b. $H_0: P_x - P_y = 0; H_1: P_x - P_y < 0;$

$$\hat{p}_o = \frac{500(.60) + 600(.64)}{500+600} = .6218,$$

$$z = \frac{.60 - .64}{\sqrt{\frac{(.6218)(1-.6218)}{500} + \frac{(.6218)(1-.6218)}{600}}} = -1.36$$

p-value = .0869. Therefore, reject H_0 at .10, but do not reject at the .05 level

c. $H_0: P_x - P_y = 0; H_1: P_x - P_y < 0;$

$$\hat{p}_o = \frac{500(.42) + 600(.49)}{500+600} = .4582,$$

$$z = \frac{.42 - .49}{\sqrt{\frac{(.4582)(1-.4582)}{500} + \frac{(.4582)(1-.4582)}{600}}} = -2.32$$

p-value = .0102.

Therefore, reject H_0 at the .05 level, but do not reject at the .01 level

d. $H_0: P_x - P_y = 0; H_1: P_x - P_y < 0;$

$$\hat{p}_o = \frac{500(.25) + 600(.34)}{500+600} = .299,\ z = \frac{.25 - .34}{\sqrt{\frac{(.299)(1-.299)}{500} + \frac{(.299)(1-.299)}{600}}} = -3.25$$

p-value = .0006. Therefore, reject H_0 at all common levels of alpha

e. $H_0: P_x - P_y = 0; H_1: P_x - P_y < 0;$

$$\hat{p}_o = \frac{500(.39) + 600(.42)}{500 + 600} = .4064,$$

$$z = \frac{.39 - .42}{\sqrt{\frac{(.4064)(1-.4064)}{500} + \frac{(.4064)(1-.4064)}{600}}} = -1.01$$

p-value = .1562. Therefore, do not reject H_0 at any common level of alpha

10.16 Let x – agreed with the statement in country A; y – agreed with the statement in country B

$H_0: P_x - P_y = 0; H_1: P_x - P_y < 0;$

$$\hat{p}_o = \frac{1556(.384) + 1108(.52)}{1556 + 1108} = .44, \quad z = \frac{.384 - .52}{\sqrt{\frac{(.44)(.56)}{1556} + \frac{(.44)(.56)}{1108}}} = -6.97$$

Reject H_0 at all common levels of alpha

10.18 Let x – people who had pledged had already been laid off; y – people who had not pledged had already been laid off

$H_0: P_x - P_y = 0; H_1: P_x - P_y \neq 0;$ reject H_0 if $|z_{.025}| > 1.96$

$$\hat{p}_o = \frac{78 + 208}{175 + 604} = .36714$$

$$z = \frac{.446 - .344}{\sqrt{\frac{(.36714)(.63286)}{175} + \frac{(.36714)(.63286)}{604}}} = 2.465.$$ Reject H_0 at the 5% level

10.20 Let x – When asked how satisfied they were; y – When asked how dissatisfied they were

$H_0: P_x - P_y = 0; H_1: P_x - P_y > 0;$ reject H_0 if $|z_{.05}| > 1.645$

$$\hat{p}_o = \frac{138 + 128}{240 + 240} = .554$$

$$z = \frac{.575 - .533}{\sqrt{\frac{(.554)(.446)}{240} + \frac{(.554)(.446)}{240}}} = .926.$$

Do not reject H_0 at the 5% level

10.22 a. $H_0: \sigma^2_x = \sigma^2_y; H_1: \sigma^2_x > \sigma^2_y$

F = 125/51 = 2.451. Reject H_0 at the 1% level since 2.451 > 2.11 ≈ $F_{(44,40,.01)}$

b. $H_0: \sigma^2_x = \sigma^2_y; H_1: \sigma^2_x > \sigma^2_y$

F = 235/125 = 1.88. Reject H_0 at the 5% level since 1.88 > 1.69 ≈ $F_{(43,44,.05)}$

c. $H_0: \sigma^2_x = \sigma^2_y; H_1: \sigma^2_x > \sigma^2_y$

F = 134/51 = 2.627. Reject H_0 at the 1% level since 2.627 > 2.11 ≈ $F_{(47,40,.01)}$

d. $H_0: \sigma^2_x = \sigma^2_y; H_1: \sigma^2_x > \sigma^2_y$

F = 167/88 = 1.90. Reject H_0 at the 5% level since 1.90 > 1.79 ≈ $F_{(24,38,.05)}$

10.24 Let x – active price competition; y – duopoly and tacit collusion

$H_0: \sigma^2_x = \sigma^2_y; H_1: \sigma^2_x > \sigma^2_y$; reject H_0 if $F_{(3,6,.05)} > 4.76$

F = 114.09/16.08 = 7.095. Reject H_0 at the 5% level

10.26 Let x – Books having more than 100 data files; y – books with at most 100 data files

$H_0: \sigma^2_x = \sigma^2_y; H_1: \sigma^2_x \neq \sigma^2_y$;

F = $(2107)^2/(1681)^2$ = 1.57

Therefore, do not reject H_0 at the 10% level since 1.57 < 3.18 ≈ $F_{(9,9,.05)}$

10.28 No. The probability of rejecting the null hypothesis given that it is true is 5%.

10.30 Assuming population variances are equal,

a. $H_0: \mu \le 4; H_1: \mu > 4$; reject H_0 if $t_{.05} > 1.671$

$$t = \frac{4.4-4}{1.3/\sqrt{70}} = 2.574.$$ Reject H_0 at the 5% level

b. Let x – response for business managers ; y – response for college economics

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y < 0$; reject H_0 if $t_{.05} < -1.645$

$$s_p^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2} = \frac{69(1.3)^2 + 105(1.4)^2}{70+106-2} = 1.853$$

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{4.4-5.3}{\sqrt{\frac{1.853}{70} + \frac{1.853}{106}}} = -4.293.$$

Reject H_0 at levels in excess of 5%

10.32 Let x – four-member groups; y – eight-member groups

Presuming the populations are normally distributed with equal variances, the samples must be independent random samples:

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y < 0$; reject H_0 if $t_{(10,.01)} < -2.764$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{3(24.4)^2 + 7(14.6)^2}{4 + 8 - 2} = 327.82$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{78 - 114.7}{\sqrt{\frac{327.82}{4} + \frac{327.82}{8}}} = -3.31 .$$

Reject H_0 at levels in excess of 1%

10.34

Let x – Obesity rate in the metro; y – Obesity rate in the non-metro

Assuming the populations are normally distributed with equal variances , in all the cases and independent random samples:

Percent of obese adults in metro and non-metro counties

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0$; reject H_0 if $t_{.05} <-1.645$

$n_x = 1089,\ n_y = 2051,\ s_x = 3.63, s_y = 3.58$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{1088(3.63)^2 + 2050(3.58)^2}{1089 + 2051 - 2} = 12.95$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{27.68 - 28.61}{\sqrt{\frac{12.95}{1089} + \frac{12.95}{2051}}} = -6.86$$

Reject H_0 at 5% level.

Percent of low-income preschool obesity in metro and non-metro counties

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0$; reject H_0 if $t_{.05} >1.645$

$n_x = 1015,\ n_y = 1676,\ s_x = 3.49, s_y = 3.85$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{974(3.49)^2 + 1715(3.85)^2}{1015 + 1676 - 2} = 13.81$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{14.27 - 14.14}{\sqrt{\frac{13.81}{1015} + \frac{13.81}{1676}}} = 0.87$$

Do not reject H_0 at 5% level.

10.36

Let x – knee patients; y – hip patients

Assuming the populations are normally distributed with equal variances,

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$. Sample sizes less than 100, use the t-test

$$s_p^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2} = \frac{82(.649)^2 + 53(.425)^2}{83+54-2} = .32675$$

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}} = \frac{6.543-6.733}{\sqrt{\frac{.32675}{83}+\frac{.32675}{54}}} = -1.901.$$ p-value is between (.05 and .025)

x 2 = .10 and .05. Reject H_0 at any alpha of .10 or higher.

10.38 Let x – firms with substantial earnings; y – firms without substantial earnings

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y < 0;$ reject Ho if $t_{(44,.05)} < -1.684$

$$s_p^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2} = \frac{22(.055)^2 + 22(.058)^2}{23+23-2} = .00319$$

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}} = \frac{.058-.146}{\sqrt{\frac{.00319}{23}+\frac{.00319}{23}}} = -5.284.$$

Reject H_0 at any common level of alpha

10.40 Let x – health insurance firms; y – casualty insurance firms

$H_0: P_x - P_y = 0; H_1: P_x - P_y \neq 0;$

$$\hat{p}_o = \frac{47+40}{69+69} = .630435,\ z = \frac{.6812-.5797}{\sqrt{(.630435)(.369565)(\frac{1}{69}+\frac{1}{69})}} = 1.235, \text{ p-value} =$$

$2[1-F_Z(1.24)] = 2[1-.8925] = .1075.$

Reject H_0 at levels of alpha in excess of 10.75%

10.42 Let x – Obesity rate in the metro; y – Obesity rate in the non-metro

Assuming the populations are normally distributed with equal variances in all the cases and independent random samples:

Percent of obese adults in metro and non-metro counties of California

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 37$, $n_y = 21$, $s_x = 3.83$, $s_y = 1.69$

$$s_p^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2} = \frac{36(3.83)^2 + 20(1.69)^2}{37+21-2} = 10.47$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{23.09 - 23.78}{\sqrt{\frac{10.47}{37} + \frac{10.47}{21}}} = -0.78$$

Do not reject H_0 at 5% level.

Percent of low-income preschool obesity in metro and non-metro counties of California

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} > 1.645$

$n_x = 37,\ n_y = 20,\ s_x = 1.98, s_y = 3.03$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{36(1.98)^2 + 19(3.03)^2}{37 + 20 - 2} = 5.72$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{16.83 - 14.97}{\sqrt{\frac{5.72}{37} + \frac{5.72}{20}}} = 2.81$$

Reject H_0 at 5% level.

Percent of obese adults in metro and non-metro counties of Michigan

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 26,\ n_y = 57,\ s_x = 2.11, s_y = 0.94$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{25(2.11)^2 + 56(0.94)^2}{26 + 57 - 2} = 1.98$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{29.48 - 29.50}{\sqrt{\frac{1.98}{26} + \frac{1.98}{57}}} = -0.06$$

Do not reject H_0 at 5% level.

Percent of low-income preschool obesity in metro and non-metro counties of Michigan

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 26,\ n_y = 56,\ s_x = 1.61, s_y = 2.85$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{25(1.61)^2 + 55(2.85)^2}{26 + 56 - 2} = 6.38$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{13.54 - 14.25}{\sqrt{\frac{6.38}{26} + \frac{6.38}{56}}} = -1.18$$

Do not reject H_0 at 5% level.

Percent of obese adults in metro and non-metro counties of Minnesota

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 21,\ n_y = 66,\ s_x = 1.23, s_y = 0.71$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{20(1.23)^2 + 65(0.71)^2}{21 + 66 - 2} = 0.74$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{26.71 - 27.41}{\sqrt{\frac{0.74}{21} + \frac{0.74}{66}}} = -3.27$$

Reject H_0 at 5% level.

Percent of low-income preschool obesity in metro and non-metro counties of Minnesota

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 21,\ n_y = 66,\ s_x = 1.94, s_y = 3.11$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{20(1.94)^2 + 65(3.11)^2}{21 + 66 - 2} = 8.27$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{11.46 - 12.84}{\sqrt{\frac{8.27}{21} + \frac{8.27}{66}}} = -1.91$$

Reject H_0 at 5% level.

Percent of obese adults in metro and non-metro counties of Florida

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 38,\ n_y = 29,\ s_x = 3.38, s_y = 3.39$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{37(3.38)^2 + 28(3.39)^2}{38 + 29 - 2} = 11.44$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{24.84 - 29.40}{\sqrt{\frac{11.44}{38} + \frac{11.44}{29}}} = -5.46$$

Reject H_0 at 5% level.

Percent of low-income preschool obesity in metro and non-metro counties of Florida

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$ reject H_0 if $t_{.05} > 1.645$

$n_x = 38,\ n_y = 28,\ s_x = 2.73, s_y = 2.66$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{37(2.73)^2 + 27(2.66)^2}{38 + 28 - 2} = 7.27$$

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{13.42 - 13.08}{\sqrt{\frac{7.27}{38} + \frac{7.27}{28}}} = 0.5$$

Do not reject H_0 at 5% level.

10.44 Let x – used the old procedure; y – used the new procedure

a. $H_0: \mu_y - \mu_x = 0; H_1: \mu_y - \mu_x > 0;$

df = $n_1 + n_2 - 2 = 27 + 27 - 2 = 52$; $t_{52,.05} = 1.675$

$$s^2_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(27-1)100 + (27-1)150}{52} = 125$$

$$t_{calc} = \frac{\overline{x}_2 - \overline{x}_1}{\sqrt{\frac{s^2_p}{n_1} + \frac{s^2_p}{n_2}}} = \frac{64 - 60}{\sqrt{\frac{125}{60} + \frac{125}{64}}} = 1.99$$

At the .05 level of significance, reject Ho and accept the alternative that the mean output per hectare is significantly greater with the new procedure.

b. 95% acceptance interval:

$F_{26,26,.025} = 2.20$, $P(\frac{1}{2.20} \leq \frac{s_2^2}{s_1^2} \leq 2.20) = .95$, $F_{calc} = \frac{150}{100} = 1.50$, because F calc is within the acceptance interval, there is not sufficient evidence against the null hypothesis that the sample variances are not significantly different from each other.

10.46 Let x – students eligible for free lunches in rural area; y – students eligible for free lunches in urban area
Presuming the populations are normally distributed with equal variances, the samples must be independent random samples:

Eligibility for free lunches between rural and urban residents of California

$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0;$ reject H_0 if $t_{.05} > 1.645$
$n_x = 21,\ n_y = 37, s_x = 12.1, s_y = 12.43$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{20(12.1)^2 + 36(12.43)^2}{21 + 37 - 2} = 151.64$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{34.86 - 37.01}{\sqrt{\frac{151.64}{21} + \frac{151.64}{37}}} = -0.64$$

Do not reject H_0 at 5% level.

Eligibility for free lunches between rural and urban residents of Texas

$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0;$ reject H_0 if $t_{.05} > 1.645$
$n_x = 176,\ n_y = 77,\ s_x = 12.85, s_y = 10.71$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{175(12.85)^2 + 76(10.71)^2}{176 + 77 - 2} = 149.89$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{44.03 - 37.4}{\sqrt{\frac{149.89}{176} + \frac{149.89}{77}}} = 3.99$$

Reject H_0 at 5% level.

Eligibility for free lunches between rural and urban residents of Florida

$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0;$ reject H_0 if $t_{.05} > 1.645$
$n_x = 29,\ n_y = 38,\ s_x = 9.34, s_y = 11.29$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{28(9.34)^2 + 37(11.29)^2}{29 + 38 - 2} = 110.09$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{44.08 - 33.8}{\sqrt{\frac{110.09}{29} + \frac{110.09}{38}}} = 3.98$$

Reject H_0 at 5% level.

10.48 Let x – Sales for Ole ice cream; y – sales for Carl's ice cream

a. $H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y > 0;$

Results for: Ole.MTW

Two-Sample T-Test and CI: Olesales, Carlsale

```
Two-sample T for Olesales vs Carlsale
            N      Mean     StDev   SE Mean
Olesales  156      3791      5364       429
Carlsale  156      2412      4249       340

Difference = mu Olesales - mu Carlsale
Estimate for difference:  1379
95% lower bound for difference: 475
T-Test of difference = 0 (vs >): T-Value = 2.52  P-Value = 0.006  DF =
310
Both use Pooled StDev =  4839
```

Reject H_0 at the .01 level of significance.

b. $H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y \neq 0;$

Two-Sample T-Test and CI: Oleprice, Carlpric

```
Two-sample T for Oleprice vs Carlpric
N       Mean      StDev    SE Mean
Oleprice  156      0.819      0.139      0.011
Carlpric  156      0.819      0.120     0.0096

Difference = mu Oleprice - mu Carlpric
Estimate for difference:   -0.0007
95% CI for difference: (-0.0297, 0.0283)
T-Test of difference = 0 (vs not =): T-Value = -0.05  P-Value = 0.962
DF = 310
Both use Pooled StDev = 0.130
```

Do not reject H_0 at any common level of significance. Note that the 95% confidence interval contains 0, therefore, no evidence of a difference.

10.50 Let x – American trade magazine advertisements; y – British trade magazine advertisements

$H_0: P_x - P_y = 0; H_1: P_x - P_y \neq 0;$ reject H_0 if $z < -z_{\alpha/2} = -1.96$ or $z > z_{\alpha/2} = 1.96$

Let $n_x = 270$ and $n_y = 203$. Then, $\hat{p}_x = 56/270 = 0.2074$ and $py = 52/203 = 0.2562$.

$$\hat{p}_0 = \frac{(270)(0.2074) + (203)(0.2562)}{270 + 203} = 0.2283$$

$$z = \frac{0.2074 - 0.2562}{\sqrt{\frac{(0.2283)(1-0.2283)}{270} + \frac{(0.2283)(1-0.2283)}{203}}} = -1.25$$

Do not reject H_0 at the 5% level. Conclude that there is a not a difference in the proportion of humorous ads in British versus American trade magazines.

10.52 Let x – diet of immigrants; y – diet of non-immigrants
Presuming the populations are normally distributed with equal variances, the samples must be independent random samples:
Difference in the diet of immigrants and natives in the first interview
$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0$; reject H_0 if $t_{.05} > 1.645$
$n_x = 885$, $n_y = 3575$, $s_x = 13.98, s_y = 13.95$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{884(13.98)^2 + 3574(13.95)^2}{885 + 3575 - 2} = 194.69$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{57.30 - 50.70}{\sqrt{\frac{194.69}{885} + \frac{194.69}{3575}}} = 12.61$$

Reject H_0 at 5% level.

Difference in the diet of immigrants and natives in the second interview
$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0$; reject H_0 if $t_{.05} > 1.645$
$n_x = 801$, $n_y = 3329$, $s_x = 14.11, s_y = 14.33$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{800(14.11)^2 + 3328(14.33)^2}{801 + 3329 - 2} = 204.2$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{60.07 - 53}{\sqrt{\frac{204.2}{801} + \frac{204.2}{3329}}} = 12.58$$

Reject H_0 at 5% level.

Hence the immigrants have strong interest for good diet in both the first and second interview

10.54 Let x – diet of single people; y – diet of married people
Presuming the populations are normally distributed with equal variances, the samples must be independent random samples:
Difference in the diet of individuals who are single and those who are married in the first interview
$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0$; reject H_0 if $t_{.05} > 1.645$
$n_x = 1785$, $n_y = 2673$, $s_x = 14.04, s_y = 14.26$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{1784(14.04)^2 + 2672(14.26)^2}{1785 + 2675 - 2} = 200.87$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{50.90 - 52.75}{\sqrt{\frac{200.87}{1785} + \frac{200.87}{2673}}} = -4.26$$

Do not reject H_0 at 5% level.

Difference in the diet of individuals who are single and those who are married in the second interview

$H_0: \mu_x - \mu_y \le 0; H_1: \mu_x - \mu_y > 0;$ reject H_0 if $t_{.05} > 1.645$

$n_x = 1597$, $n_y = 2531$, $s_x = 14.82, s_y = 14.33$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{1596(14.82)^2 + 2530(14.33)^2}{1597 + 2531 - 2} = 210.84$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{53.11 - 55.17}{\sqrt{\frac{210.84}{1597} + \frac{210.84}{2531}}} = -4.43$$

Do not reject H_0 at 5% level.

We cannot conclude about the quality of the diet of individuals who are single and those who are married in either of the interviews.

10.56 Let x – daily food cost for women; y – daily food cost for men

Presuming the populations are normally distributed with equal variances, the samples must be independent random samples:

Difference in the daily food cost between men and women in the first interview

$H_0: \mu_x - \mu_y \ge 0; H_1: \mu_x - \mu_y < 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 2321$, $n_y = 2139$, $s_x = 2.57, s_y = 3.42$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{2320(2.57)^2 + 2138(3.42)^2}{2321 + 2139 - 2} = 9.05$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{4.65 - 5.84}{\sqrt{\frac{9.05}{2321} + \frac{9.05}{2139}}} = -13.16$$

Reject H_0 at 5% level.

Difference in the daily food cost between men and women in the second interview

$H_0: \mu_x - \mu_y \ge 0; H_1: \mu_x - \mu_y < 0;$ reject H_0 if $t_{.05} < -1.645$

$n_x = 2176$, $n_y = 1954$, $s_x = 2.49, s_y = 3.22$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{2175(2.49)^2 + 1953(3.22)^2}{2176 + 1954 - 2} = 8.17$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{4.49 - 5.58}{\sqrt{\frac{8.17}{2176} + \frac{8.17}{1954}}} = -12.26$$

Reject H_0 at 5% level.

Hence there is difference in the daily food cost quality of diet between men and women in both the interviews.

10.58 We wanted to test whether the immigrant population have a lower percentage of people that are overweight compared to the remainder of the population in the first interview.

Let x – immigrant population

y- non-immigrant population

$H_0: P_x - P_y \geq 0; H_1: P_x - P_y < 0;$

Two Proportions - 1 = Immigrant 2 = non-immigrant Day 1

```
Sample      X      N  Sample p
1         404    881  0.458570
2        1993   3571  0.558107


       Difference = p (1) - p (2)
Estimate for difference:  -0.0995372
95% upper bound for difference:  -0.0687260
Test for difference = 0 (vs < 0):  Z = -5.31  P-Value = 0.000
```

Reject H_0 at all levels of alpha.

We wanted to test whether the immigrant population have a lower percentage of people that are overweight compared to the remainder of the population in the second interview.

Let x – immigrant population

y- non-immigrant population

$H_0: P_x - P_y \geq 0; H_1: P_x - P_y < 0;$

Two Proportions - 1 = Immigrant 2 = non-immigrant Day 2

```
Sample      X      N  Sample p
1         375    797  0.470514
2        1891   3325  0.568722


Difference = p (1) - p (2)
Estimate for difference:  -0.0982074
95% upper bound for difference:  -0.0658764
Test for difference = 0 (vs < 0):  Z = -5.00  P-Value = 0.000
```

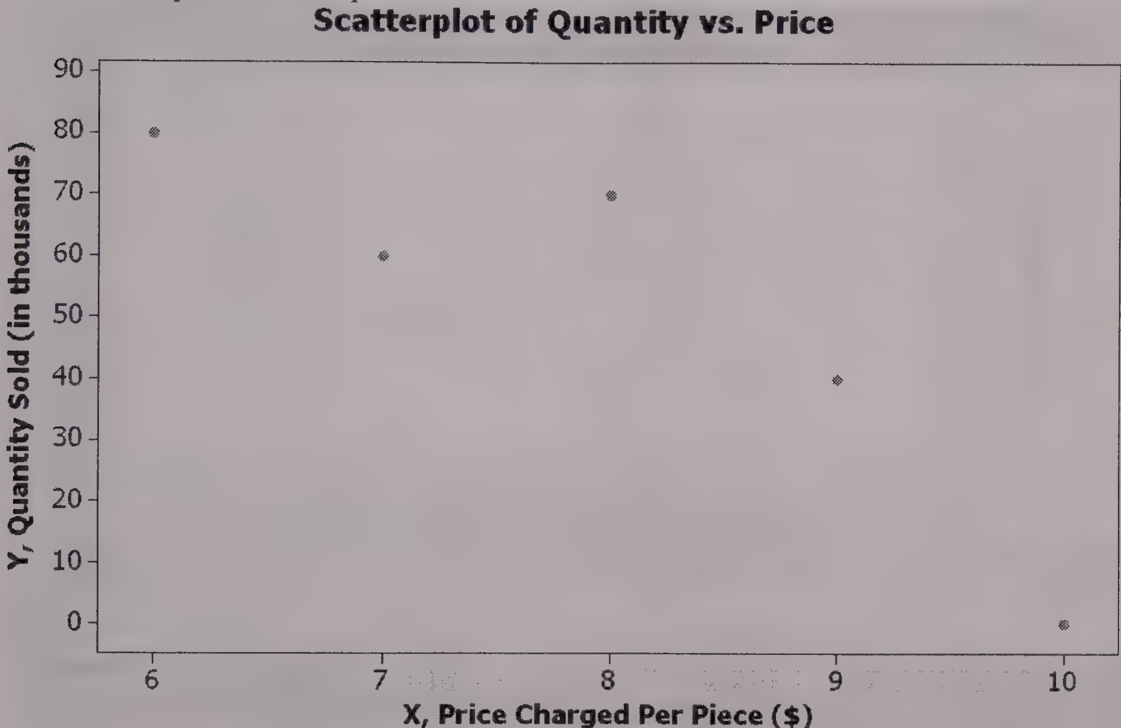
Reject H_0 at all levels of alpha.

Hence the percentage of overweight people is higher for immigrant than for non-immigrant populations in both the first and second interviews.

Chapter 11:

Simple Regression

11.2 a. Prepare a scatter plot.



b. Compute the covariance = -45

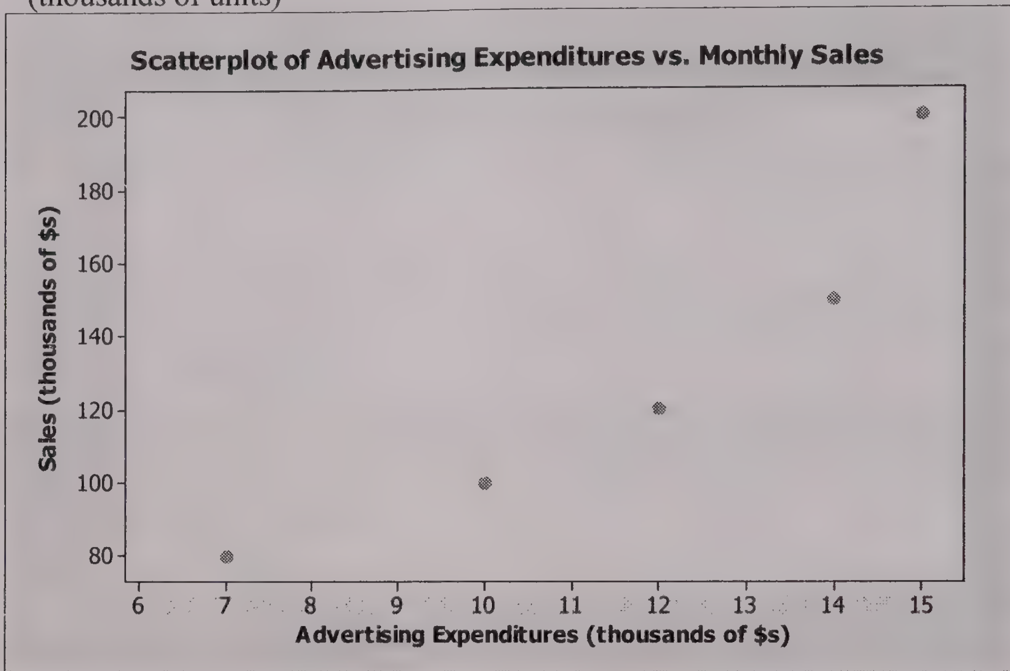
x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
6	80	-2	4	30	900	-60
7	60	-1	1	10	100	-10
8	70	0	0	20	400	0
9	40	1	1	-10	100	-10
10	0	2	4	-50	2500	-100
40	250	0	10	0	4000	-180
$\bar{x}$ = 8.00	$\bar{y}$ = 50.00		s_x^2 = 2.5		s_y^2 =1000	Cov(x,y) = -45
			s_x = 1.5811		s_y =31.623	

c. Compute and interpret $b_1 = \dfrac{Cov(x,y)}{s_x^2} = \dfrac{-45}{2.5} = -18.0$. For a one dollar increase in the price per piece of plywood, the quantity sold of plywood is estimated to decrease by 18 thousand pieces.

d. Compute $b_0 = \bar{y} - b_1\bar{x} = 50.0 - (-18)(8.0) = 194.00$

e. What quantity of plywood is expected to be sold if the price were \$7 per piece?
$\hat{y} = b_0 + b_1 x = 194.00 + -18.0(7) = 68$ thousand pieces sold

11.4 a. Scatter plot – Advertising expenditures (thousands of \$s) vs. Monthly Sales (thousands of units)



x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
10	100	-1.6	2.56	-30	900	48
15	200	3.4	11.56	70	4900	238
7	80	-4.6	21.16	-50	2500	230
12	120	0.4	0.16	-10	100	-4
14	150	2.4	5.76	20	400	48
58	650		41.2		8800	560
$\bar{x}$ = 11.60	$\bar{y}$ = 130.00		s_x^2 = 10.3		s_y^2 =2200	Cov(x,y) = 140
			s_x = 3.2094		s_y = 46.9042	
				b1 =	13.5922	
				b0 =	-27.67	

$$\text{Covariance} = Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = 560 / 4 = 140$$

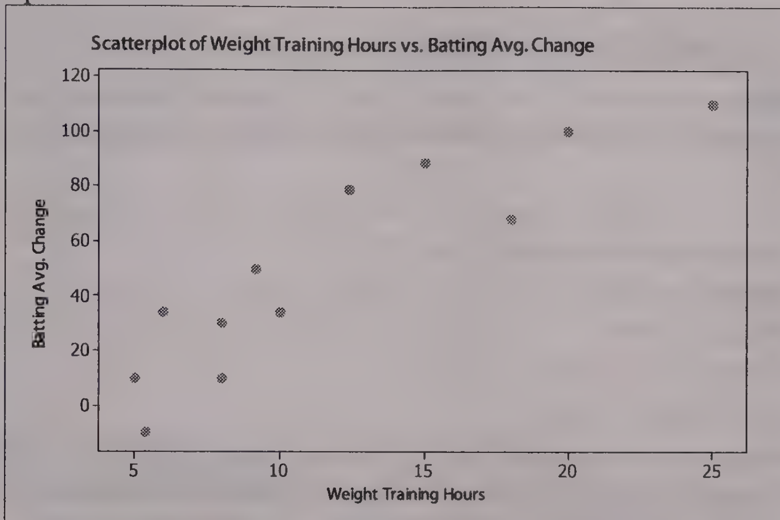
$$\text{Correlation} = \frac{Cov(x, y)}{s_x s_y} = \frac{140}{(3.2094)(46.9042)} = .93002$$

b. It appears that advertising has a positive effect on sales. Note that correlation does not imply causation. Other factors could have been changing at the same time that advertising changed. For example, prices of competitive goods could have been changing or tastes and preferences of consumers, or the number of buyers in the market.

c. Compute the regression $b_1 = \dfrac{Cov(x,y)}{s_x^2} = 140 / 10.3 = 13.5922$ and coefficients

$b_0 = \bar{y} - b_1\bar{x} = 130.0 - (13.5922)(11.6) = \text{-}27.669$

11.6 a. Scatter plot



The data show a positive relationship between the average number of hours spent in the weight training program versus the increment to each player's batting average. It appears that the weight training program has been effective, although correlation does not necessarily imply causation.

b. Estimate the regression equation.

x_i	y_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
8	10	-3.83333	14.69444	-40.3333	1626.778	154.61111
20	100	8.166667	66.69444	49.66667	2466.778	405.61111
5.4	-10	-6.43333	41.38778	-60.3333	3640.111	388.14444
12.4	79	0.566667	0.321111	28.66667	821.7778	16.244444
9.2	50	-2.63333	6.934444	-0.33333	0.111111	0.8777778
15	89	3.166667	10.02778	38.66667	1495.111	122.44444
6	34	-5.83333	34.02778	-16.3333	266.7778	95.277778
8	30	-3.83333	14.69444	-20.3333	413.4444	77.944444
18	68	6.166667	38.02778	17.66667	312.1111	108.94444
25	110	13.16667	173.3611	59.66667	3560.111	785.61111
10	34	-1.83333	3.361111	-16.3333	266.7778	29.944444
5	10	-6.83333	46.69444	-40.3333	1626.778	275.61111
142	604		450.2267		16496.67	2461.2667

$\bar{x}$ = 11.83 $\bar{y}$ = 50.33 s_x^2 = 40.9297 s_y^2 = 1499.697 Cov(x,y) = 223.75152

6.397632 38.72592 s_x = 6.397632 s_y = 38.72592

$$\text{Covariance} = Cov(x,y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = 2461.2667 / 11 = 223.75152$$

$$\text{Correlation} = \frac{Cov(x,y)}{s_x s_y} = \frac{223.75152}{(6.397632)(38.72592)} = .90312$$

$$\text{Regression coefficients } b_1 = \frac{Cov(x,y)}{s_x^2} = 223.75152 / 40.9297 = 5.4667 \text{ and}$$

$$b_0 = \bar{y} - b_1\bar{x} = 50.33 - (5.4667)(11.83) = -14.3411$$

Regression equation = $\hat{Y}$ = -14.4 + 5.47X. Each additional hour of the weight training program yields an expected improvement in batting average of 5.47 points.

11.8 Given the regression equation $\hat{y} = -50 + 12X$

a. Y changes by +36
b. Y changes by -48
c. $\hat{y} = -50 + 12(12) = 94$
d. $\hat{y} = -50 + 12(23) = 226$
e. Regression results do not "prove" that increased values of X "causes" increased values of Y. Theory will help establish conclusions of causation.

11.10 Given the regression equation $\hat{y} = 100 + 21X$

a. Y changes by +105
b. Y changes by -147
c. $\hat{y} = 100 + 21(14) = 394$
d. $\hat{y} = 100 + 21(27) = 667$
e. Regression results do not "prove" that increased values of X "causes" increased values of Y. Theory will help establish conclusions of causation.

11.12 There are insufficient data points to create an accurate regression equation. The estimate is significantly outside previous data values so any estimate would be faulty and not reliable.

11.14 A population regression equation consists of the true regression coefficients β_i and the true model error ε_i. By contrast, the estimated regression model consists of the estimated regression coefficients b_i and the residual term e_i. The population regression equation is a model that purports to measure the actual value of Y, while the sample regression equation is an estimate of the predicted value of the dependent variable Y.

11.16 The constant represents an adjustment for the estimated model and not the number sold when the price is zero.

11.18 Compute the coefficients for a least squares regression equation

$$b_1 = r_{xy}\frac{s_y}{s_x} \quad b_0 = \bar{y} - b_1\bar{x} \quad \hat{y}_i = b_0 + b_1 x_i$$

a. $b_1 = .6\frac{75}{25} = 1.80 \quad b_0 = 100 - 1.80(50) = 10 \quad \hat{y}_i = 10 + 1.80x_i$

b. $b_1 = .7\frac{65}{35} = 1.30 \quad b_0 = 210 - 1.30(60) = 132 \quad \hat{y}_i = 132 + 1.30x_i$

c. $b_1 = .75\frac{78}{60} = .975 \quad b_0 = 100 - .975(20) = 80.5 \quad \hat{y}_i = 80.5 + .975x_i$

d. $b_1 = .4\frac{75}{100} = .30 \quad b_0 = 50 - .30(10) = 47 \quad \hat{y}_i = 47 + .30x_i$

e. $b_1 = .6\frac{70}{80} = .525 \quad b_0 = 200 - .525(90) = 152.75 \quad \hat{y}_i = 152.75 + .525x_i$

11.20 a. $n = 20, \bar{X} = 25.4/20 = 1.27, \bar{Y} = 22.6/20 = 1.13$

$$b_1 = \frac{150.5 - (20)(1.13)(1.27)}{145.7 - (20)(1.27)^2} = 1.0737, \; b_0 = 1.13 - 1.0737(1.27) = -.2336$$

b. For a one unit increase in the rate of return of the S&P 500 index, we estimate that the rate of return of the corporation's stock will increase by 1.07%

c. When the percentage rate of return of the S&P 500 index is zero, we estimate that the corporation's rate of return will be -.2336%

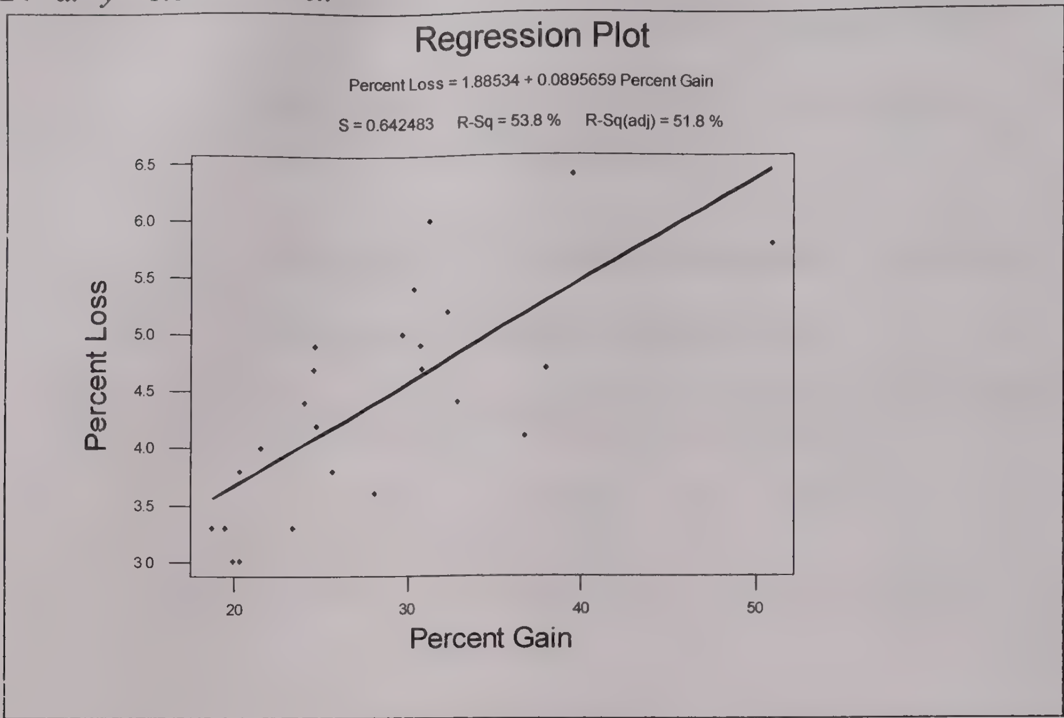
11.22 a. $b_1 = 180/350 = .5143$, $b_0 = 16 - .5143(25.5) = 2.8854$

$\hat{y} = 2.8854 + .5143x$

b. For a one unit increase in the average cost of a meal, we would estimate that the number of bottles sold would increase by .5143%

c. Yes. 2.8854 bottles are estimated to be sold, regardless of the price paid for a meal.

11.24 a. $\hat{y} = 1.89 + 0.0896x$



b. 0.0896%. For a one percent pre-November 13 gain, we would estimate that there would be a loss of .0896% on November 13.

11.26 a.

$$y_i = a + bx_i + e_i$$
$$e_i = y_i - a - bx_i$$
$$e_i = y_i - \bar{y} + b\bar{x} - bx_i$$
$$e_i = y_i - \bar{y} - b(x_i - \bar{x})$$

b.

$$\sum e_i = \sum (y_i - \bar{y} - b(x_i - \bar{x}))$$
$$\sum e_i = \sum y_i - n\bar{y} - b\sum x_i + bn\bar{x}$$
$$\sum e_i = n\bar{y} - n\bar{y} - bn\bar{x} + bn\bar{x}$$
$$\sum e_i = 0$$

c.

$$\sum e_i^2 = \sum (y_i - \bar{y} - b(x_i - \bar{x}))^2$$
$$\sum e_i^2 = \sum (y_i - \bar{y})^2 + b^2 \sum (x_i - \bar{x})^2 - 2b\sum (x_i - \bar{x})(y_i - \bar{y})$$
$$\sum e_i^2 = \sum (y_i - \bar{y})^2 + b^2 \sum (x_i - \bar{x})^2 - 2b^2 \sum (x_i - \bar{x})^2$$
$$\sum e_i^2 = \sum (y_i - \bar{y})^2 - b^2 \sum (x_i - \bar{x})^2$$

d.

$$\hat{y}_i - \overline{y} = a + bx_i - \overline{y}$$
$$\hat{y}_i - \overline{y} = \overline{y} - b\overline{x} + bx_i - \overline{y}$$
$$\hat{y}_i - \overline{y} = b(x_i - \overline{x})$$

e.

$$SSE = \sum e_i^2 = \sum (y_i - \overline{y})^2 - b^2 \sum (x_i - \overline{x})^2$$
$$\sum e_i^2 = \sum (y_i - \overline{y})^2 - \sum (\hat{y}_i - \overline{y})^2$$
$$\sum e_i^2 = SST - SSR$$
$$SST = SSR + SSE$$

11.28 $n = 8$, $\overline{y} = 405$, $SST = \sum (y_i - \overline{y})^2 = 8000$, $SSR = \sum (\hat{y}_i - \overline{y})^2 = 4802.845$

$$r^2 = \frac{4802.845}{8000} = 0.60$$

The coefficient of determination is $r^2 = 0.60$. This means that 60% of the variation is explained by the regression equation.

11.30 $n = 25, \overline{x} = 28.344, \sum x^2 = 21464.7, \overline{y} = 4.424, \sum y^2 = 509.86, b = .0896$

Based on the result from Exercise 11.28 a:

$$R^2 = (.0896)^2 \frac{21464.7 - 25(28.344)^2}{509.86 - 25(4.424)^2} = .5384$$

Or, from the Minitab output:

Regression Analysis: Percent Loss Y versus Percent Gain X

```
The regression equation is
Percent Loss Y = 1.89 + 0.0896 Percent Gain X
Predictor        Coef     SE Coef        T        P
Constant       1.8853      0.5067     3.72    0.001
Percent       0.08957     0.01729     5.18    0.000

S = 0.6425      R-Sq = 53.8%      R-Sq(adj) = 51.8%
Analysis of Variance
Source            DF        SS        MS        F        P
Regression         1    11.072    11.072    26.82    0.000
Residual Error    23     9.494     0.413
Total             24    20.566
```

11.32 Given the following simple regression model $Y = \beta_0 + \beta_1 X$

a. $t = \dfrac{b_1}{s_{b1}} = \dfrac{5}{2.1} = 2.381$. $t_{36,.05} \approx 1.684$. Reject H_0 at the 5% level.

$b_1 \pm t_{(n-2,\alpha/2)} s_{b_1}$. 95% confidence interval is: $5 \pm 2.021(2.1) = .7559$ up to 9.2441.

99% confidence interval is: $5 \pm 2.704(2.1) =$ -.6784 up to 10.6784.

b. $t = \dfrac{b_1}{s_{b1}} = \dfrac{5.2}{2.1} = 2.476.$ $t_{44,.05} \approx 1.684.$

Reject H_0 at the 5% level. $b_1 \pm t_{(n-2,\alpha/2)} s_{b_1}$.

95% confidence interval is: $5.2 \pm 2.021(2.1) = .9559$ up to 9.4441.
99% confidence interval is: $5.2 \pm 2.704(2.1) = -.4784$ up to 10.8784.

c. $t = \dfrac{b_1}{s_{b1}} = \dfrac{2.7}{1.87} = 1.444.$ $t_{36,.05} \approx 1.684.$

Do not reject H_0 at the 5% level. $b_1 \pm t_{(n-2,\alpha/2)} s_{b_1}$.

95% confidence interval is: $2.7 \pm 2.021(1.87) = -1.079$ up to 6.4793.
99% confidence interval is: $2.7 \pm 2.704(1.87) = -2.3565$ up to 7.7565.

d. $t = \dfrac{b_1}{s_{b1}} = \dfrac{6.7}{1.8} = 3.722.$ $t_{27,.05} \approx 1.703.$

Reject H_0 at the 5% level. $b_1 \pm t_{(n-2,\alpha/2)} s_{b_1}$.

95% confidence interval is: $6.7 \pm 2.052(1.8) = 3.0064$ up to 10.3936.
99% confidence interval is: $6.7 \pm 2.771(1.8) = 1.7122$ up to 11.6878.

11.34 a. The standard error is 26.41.

b. $s_{b_1} = 0.1661$

c. $(-0.6877, -0.0421)$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6676789
R Square	0.44579511
Adjusted R Square	0.35342763
Standard Error	26.4128764
Observations	8

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3367.03475	3367.03475	4.82632096	0.07040329
Residual	6	4185.84025	697.640041		
Total	7	7552.875			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper*
Intercept	469.810751	30.8896257	15.2093378	5.0985E-06	394.226504	545.
Price	-0.36494077	0.16611708	-2.19688893	0.07040329	-0.77141491	0.04

Lower 90.0%	*Upper 9*
409.78662	529.834
-0.6877363	-0.0421

11.36 The regression equation is $\hat{y}$=43154+1.33x

For one \$ increase in the per capita disposable income, we estimate or predict that there will be an increase of \$1.33 in retail sales.

95% confidence interval (-111,907, 200,866) (Using the below Minitab output)

Regression Analysis: retail sales versus per capita disposable income

```
The regression equation is
retail sales = 43154 + 1.33 per capita disposable income
Predictor                          Coef  SE Coef     T      P
Constant                          43154    80581  0.54  0.595
per capita disposable income      1.326    2.797  0.47  0.638


S = 90949.9   R-Sq = 0.5%   R-Sq(adj) = 0.0%
Analysis of Variance

Source          DF           SS           MS      F      P
Regression       1   1859655400   1859655400   0.22  0.638
Residual Error  49  4.05322E+11   8271877993
Total           50  4.07182E+11
Predicted Values for New Observations
New
Obs    Fit   SE Fit          95% CI                  95% PI
  1  44480    77821  (-111907, 200866)  (-196065, 285025)XX

XX denotes a point that is an extreme outlier in the predictors
```

11.38

Regression Analysis: Percent Loss Y versus Percent Gain X

```
The regression equation is
Percent Loss Y = 1.89 + 0.0896 Percent Gain X

Predictor          Coef  SE Coef     T      P
Constant         1.8853   0.5067  3.72  0.001
Percent Gain X  0.08957  0.01729  5.18  0.000

S = 0.642483   R-Sq = 53.8%   R-Sq(adj) = 51.8%
Analysis of Variance
Source          DF      SS      MS      F      P
Regression       1  11.072  11.072  26.82  0.000
Residual Error  23   9.494   0.413
Total           24  20.566
```

a. $\sum e_i^2 = SSE = 9.494$, $s_e^2 = SSE/(n-2) = 9.494/23 = .413$

b. $s_b^2 = (.01729)^2 = .000299$

c. $t_{23,.05} = 1.714$, $t_{23,.025} = 2.069$, $t_{23,.005} = 2.807$

Therefore, the 90% confidence interval is: $.0896 \pm 1.714(.01729)$, .05996 up to .1192

Therefore, the 95% confidence interval is: $.0896 \pm 2.069(.01729)$, .05382 up to .1254

Therefore, the 99% confidence interval is: $.0896 \pm 2.807(.01729)$, .0411 up to .1381

11.40 Given a simple regression: $\hat{y}_{n+1} = 14 + 7(11) = 91$

95% Prediction Interval: $\hat{y}_{n+1} \pm t_{n-2,\alpha/2}\sqrt{\left[1+\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}(s_e)$

$$91 \pm 2.069\sqrt{\left[1+\frac{1}{25}+\frac{(11-8)^2}{300}\right]}(7.45) = 91 \pm 16.051,\ (74.949,\ 107.051)$$

95% Confidence Interval: $\hat{y}_{n+1} \pm t_{n-2,\alpha/2}\sqrt{\left[\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}(s_e)$

$$91 \pm 2.069\sqrt{\left[\frac{1}{25}+\frac{(11-8)^2}{300}\right]}(7.45) = 91 \pm 4.078,\ (86.922,\ 95.078)$$

11.42 Given a simple regression: $\hat{y}_{n+1} = 8 + 10(17) = 178$

95% Prediction Interval: $\hat{y}_{n+1} \pm t_{n-2,\alpha/2}\sqrt{\left[1+\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}(s_e)$

$$178 \pm 2.021\sqrt{\left[1+\frac{1}{44}+\frac{(17-8)^2}{800}\right]}(11.23) = 178 \pm 21.0613,\ (153.939,\ 202.061)$$

95% Confidence Interval: $\hat{y}_{n+1} \pm t_{n-2,\alpha/2}\sqrt{\left[\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}(s_e)$

$$178 \pm 2.021\sqrt{\left[\frac{1}{44}+\frac{(17-8)^2}{800}\right]}(11.23) = 178 \pm 7.992,\ (170.008,\ 185.992)$$

11.44 a. $Y = 73.65 - 36.98X$

b. The probability is 0.90 that the interval -59.70 to -14.26 includes the true slope coefficient.

c. No. The prediction is outside the range of the data used for estimation. There is no guarantee that the same relationship holds outside the range of the data and predictions cannot be made for a dosage of 2.5 grams.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.730652
R Square	0.533853
Adjusted R Square	0.475584
Standard Error	7.765435
Observations	10

ANOVA

	df	SS	MS	F	Significance F
Regression	1	552.4841584	552.4841584	9.161957	0.016387198
Residual	8	482.4158416	60.3019802		
Total	9	1034.9			

	Coefficients	Standard Error	t Stat	P-value	Lower 90.0%	Upper 90.0%
Intercept	73.65347	16.55432369	4.449198091	0.002142	42.86990055	104.4370301
X Variable 1	-36.9802	12.21729623	-3.0268725	0.016387	-59.6988507	-14.26154534

11.46 $s^2{}_b = .000299, b = .0896$

$$H_o : \beta = 0, H_1 : \beta \neq 0,\ t = \frac{.0896}{\sqrt{.000299}} = 5.1817,$$

Therefore, reject H_0 at the 1% level since $t = 5.1817 > 2.807 = t_{23,.005}$

11.48 $b_1 = 180/350 = 0.514,\ b_0 = 16-(0.514)25.5 = 2.893$

$\hat{y} = 2.893+0.514x$

$\hat{y}_{n+1} = 2.893 + .514(27) = 16.771$

95% Prediction Interval: $\hat{y}_{n+1} \pm t_{n-2,\alpha/2} \sqrt{\left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]}(s_e)$

SST = 250, SSR = $(0.514)^2 350 = 92.571$, SSE = 157.429, $s_e^2 = 10.495$

$$16.771 \pm 2.131 \sqrt{\left[1 + \frac{1}{17} + \frac{(27 - 25.5)^2}{350(17-1)}\right]}(3.240) = 16.771 \pm 7.1261,\ (14.578, 18.964)$$

95% Confidence Interval: $\hat{y}_{n+1} \pm t_{n-2,\alpha/2}\sqrt{\left[\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}(s_e)$

$$16.771 \pm 2.131\sqrt{\left[\frac{1}{17}+\frac{(27-25.5)^2}{350}\right]}(3.240) = 16.771 \pm 1.7637,\ (8.731,\ 12.259)$$

11.50 $H_o:\beta=0, H_1:\beta<0,\ t=\frac{-6.025}{\sqrt{.2827}}=-11.332$

Therefore, reject H_0 at the 5% level since $t = -11.332 < -1.943 = -t_{6,.05}$

11.52

Regression Analysis: Health versus Percap Disp

```
The regression equation is
Health = 699 + 0.152 Percap Disp
Predictor        Coef  SE Coef     T      P
Constant          699     5270  0.13  0.895
Percap Disp    0.1515   0.1829  0.83  0.411


S = 5947.60   R-Sq = 1.4%   R-Sq(adj) = 0.0%


Analysis of Variance
Source          DF          SS        MS     F      P
Regression       1    24277272  24277272  0.69  0.411
Residual Error  49  1733322810  35373935
Total           50  1757600082
```

$\hat{Y}_{n+1} = 699 + 0.152(32000) = 5563$

The 95% prediction interval for prediction of the actual value:

$$5563 \pm 2.021\sqrt{[1+\frac{1}{51}+\frac{(32000-28447)^2}{1{,}057{,}275{,}916.296}]}(5947.60)$$

5563 ± 12208.2311, -6645.2311 up to 17771.2311

The 95% confidence interval for prediction of the expected value:

$$5563 \pm 2.021\sqrt{[\frac{1}{51}+\frac{(32000-28447)^2}{1{,}05{,}72{,}75{,}916.296}]}(5947.60)$$

5563 ± 2134.9734, 3428.0266 up to 7697.9734

11.54 a. Compute the sample correlation

x	y	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
2	5	-1.8	3.24	-2.4	5.76	4.32
5	8	1.2	1.44	0.6	0.36	0.72
3	7	-0.8	0.64	-0.4	0.16	0.32
1	2	-2.8	7.84	-5.4	29.16	15.12
8	15	4.2	17.64	7.6	57.76	31.92
19	**37**		**30.8**		**93.2**	**52.4**

$\bar{x} = 19/5 = 3.8$, $\bar{y} = 37/5 = 7.4$, $s_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\dfrac{30.8}{4}} = 2.7749$,

$s_y = \sqrt{\dfrac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{\dfrac{93.2}{4}} = 4.827$, $s_{xy} = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = 52.4/4 =$
13.1

$r = \dfrac{s_{xy}}{s_x s_y} = 13.1/(2.7749)(4.827) = .97802$

b. Compute the sample correlation

x	y	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
7	5	-1.8	3.24	-2.4	5.76	4.32
10	8	1.2	1.44	0.6	0.36	0.72
8	7	-0.8	0.64	-0.4	0.16	0.32
6	2	-2.8	7.84	-5.4	29.16	15.12
13	15	4.2	17.64	7.6	57.76	31.92
44	**37**	**0**	**30.8**	**0**	**93.2**	**52.4**

$\bar{x} = 44/5 = 8.8$, $\bar{y} = 37/5 = 7.4$, $s_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\dfrac{30.8}{4}} = 2.7749$,

$s_y = \sqrt{\dfrac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{\dfrac{93.2}{4}} = 4.827$, $s_{xy} = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$
$= 52.4/4 = 13.1$

$r = \dfrac{s_{xy}}{s_x s_y} = 13.1/(2.7749)(4.827) = .97802$

c. Compute the sample correlation

x	y	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
12	4	-3.6	12.96	-1.8	3.24	6.48
15	6	-0.6	0.36	0.2	0.04	-0.12
16	5	0.4	0.16	-0.8	0.64	-0.32
21	8	5.4	29.16	2.2	4.84	11.88
14	6	-1.6	2.56	0.2	0.04	-0.32
78	**29**		**45.2**		**8.8**	**17.6**

$\bar{x} = 78/5 = 15.6$, $\bar{y} = 29/5 = 5.8$, $s_x = \sqrt{\dfrac{\sum(x_i-\bar{x})^2}{n-1}} = \sqrt{\dfrac{45.2}{4}} = 3.36155$,

$s_y = \sqrt{\dfrac{\sum(y_i-\bar{y})^2}{n-1}} = \sqrt{\dfrac{8.8}{4}} = 1.48324$, $s_{xy} = \dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1} = 17.6/4$
$= 4.4$

$$r = \frac{s_{xy}}{s_x s_y} = 4.4/(3.36155)(1.48324) = .88247$$

d. Compute the correlation coefficient

x	y	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
2	8	-1.8	3.24	-5	25	9
5	12	1.2	1.44	-1	1	-1.2
3	14	-0.8	0.64	1	1	-0.8
1	9	-2.8	7.84	-4	16	11.2
8	22	4.2	17.64	9	81	37.8
19	**65**	**0**	**30.8**	**0**	**124**	**56**

$\bar{x} = 19/5 = 3.8$, $\bar{y} = 65/5 = 13$, $s_x = \sqrt{\dfrac{\sum(x_i-\bar{x})^2}{n-1}} = \sqrt{\dfrac{30.8}{4}} = 2.77488$,

$s_y = \sqrt{\dfrac{\sum(y_i-\bar{y})^2}{n-1}} = \sqrt{\dfrac{124}{4}} = 5.56776$, $s_{xy} = \dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1} = 56/4 =$
14

$$r = \frac{s_{xy}}{s_x s_y} = 14/(2.77488)(5.56776) = .90615$$

11.56 Let x = Examination and y = Project

x	y	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
81	76	2.4	5.76	-0.7	0.49	-1.68
62	71	-16.6	275.56	-5.7	32.49	94.62
74	69	-4.6	21.16	-7.7	59.29	35.42
78	76	-0.6	0.36	-0.7	0.49	0.42
93	87	14.4	207.36	10.3	106.09	148.32
69	62	-9.6	92.16	-14.7	216.09	141.12
72	80	-6.6	43.56	3.3	10.89	-21.78
83	75	4.4	19.36	-1.7	2.89	-7.48
90	92	11.4	129.96	15.3	234.09	174.42
84	79	5.4	29.16	2.3	5.29	12.42
786	**767**		**824.4**		**668.1**	**575.8**

$\bar{x} = 78.6$, $\bar{y} = 76.7$, $s_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\dfrac{824.4}{9}} = 9.5708$,

$s_y = \sqrt{\dfrac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{\dfrac{668.1}{9}} = 8.6159$, $s_{xy} = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} =$

$575.8/9 = 63.97778$, $r = \dfrac{s_{xy}}{s_x s_y} = 63.97778/(9.5708)(8.6159) = .7759$

11.58 $H_o: \rho = 0, H_1: \rho > 0, t = \dfrac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} = t = \dfrac{.11\sqrt{351}}{\sqrt{1-(.11^2)}} = 2.073$

Therefore, reject H_0 at the 2.5% level since $2.073 > 1.96 = z_{.025} \approx t_{351,.025}$

11.60 a. Using the computer, calculate the sample correlation

Correlations: Dow5day, Dow1Yr

```
Pearson correlation of Dow5day and Dow1Yr = -0.407
P-Value = 0.168
```

b. $H_o: \rho = 0, H_1: \rho \neq 0$, $t = \dfrac{-.4066\sqrt{11}}{\sqrt{1-(-.4066)^2}} = -1.4761$

$t_{11,.05} = \pm 1.796$, p-value of .168 > alpha of .10. Do not reject H_0 at the 10% level

11.62 Using the computer, calculate the correlation and test against a two-sided alternative

Results for: Advertising Revenue.MTW

Correlations: Cost of advertisement, Revenue from Inquiries

```
Pearson correlation of X and Y = 0.0575
P-Value = 0.827
```

$H_o: \rho = 0, H_1: \rho \neq 0$, $t = \dfrac{.0575\sqrt{15}}{\sqrt{1-(.0575)^2}} = .2231$,

Do not reject H_0 at the 20% level since $.2231 < 1.341 = t_{15,.10}$

11.64 The Senior Housing return has a beta of 1.369 with a coefficient Student's $t = 3.87$ and an overall R-squared of 20.5%. This means the nondiversifiable risk response for Senior Housing is significantly above the overall market. This firm's return is more responsive to the market.

Senior Housing	mean	0.008144
SUMMARY OUTPUT	return	0.81%

Regression Statistics	
Multiple R	0.45332
R Square	0.205499
Adjusted R Square	0.191801
Standard Error	0.068181
Observations	60

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.069738	0.069738	15.0018	0.000275
Residual	58	0.269623	0.004649		
Total	59	0.339362			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.00082	0.009101	-0.08991	0.928668	-0.01904	0.0174
SP 500	1.368854	0.353415	3.873216	0.000275	0.661417	2.076292

11.66 The Seagate return has a beta of 1.810 with a coefficient Student's $t = 2.98$ and an overall R-squared of 13.2%. The nondiversifiable risk response for Seagate is substantially above the overall market. For the 60-month period the average monthly return for Seagate was –0.03%, with the average variance equal to 0.0021.

The Microsoft return has a beta of 0.967 with a coefficient Student's $t = 3.31$ and an overall R-squared of 15.9%. The nondiversifiable risk response for Microsoft is slightly below the overall market. For the 60-month period the average monthly return for Microsoft was 0.00%, with the average variance equal to 0.0006.

The Tata return has a beta of 2.796 with a coefficient Student's $t = 4.14$ and an overall R-squared of 22.8%. The nondiversifiable risk response for Tata is extremely above the overall market. For the 60-month period the average monthly return for Tata was 2.23%, with the average variance equal to 0.0049.

The Tata stock has the most risk and the most return. Microsoft is the least risky but has a low return. If the market is trending upward Tata is the recommended stock. If the market is volatile or trending downward Microsoft is the recommended stock.

Seagate	mean	-0.00026676	-0.03%
	Variance	0.01560793	
Microsoft	mean	4.59393E-07	0.00%
	Variance	0.003712483	
Tata	mean	0.022260121	2.23%
	Variance	0.021645412	

SUMMARY OUTPUT Seagate!

Regression Statistics	
Multiple R	0.363972
R Square	0.132476
Adjusted R Square	0.117518
Standard Error	0.117361
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.121992712	0.121993	8.856925	0.004253
Residual	58	0.798875163	0.013774		
Total	59	0.920867875			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.01212	0.015666099	-0.77368	0.44226	-0.04348	0.019238
SP 500	1.810456	0.608340183	2.976059	0.004253	0.592732	3.02818

SUMMARY OUTPUT Microsoft

Regression Statistics	
Multiple R	0.398804
R Square	0.159045
Adjusted R Square	0.144546
Standard Error	0.056355
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.034836621	0.034837	10.96919	0.001598
Residual	58	0.184199855	0.003176		
Total	59	0.219036476			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.00633	0.007522567	-0.842	0.403246	-0.02139	0.008724
SP 500	0.967473	0.292113539	3.311977	0.001598	0.382745	1.552202

SUMMARY OUTPUT TATA

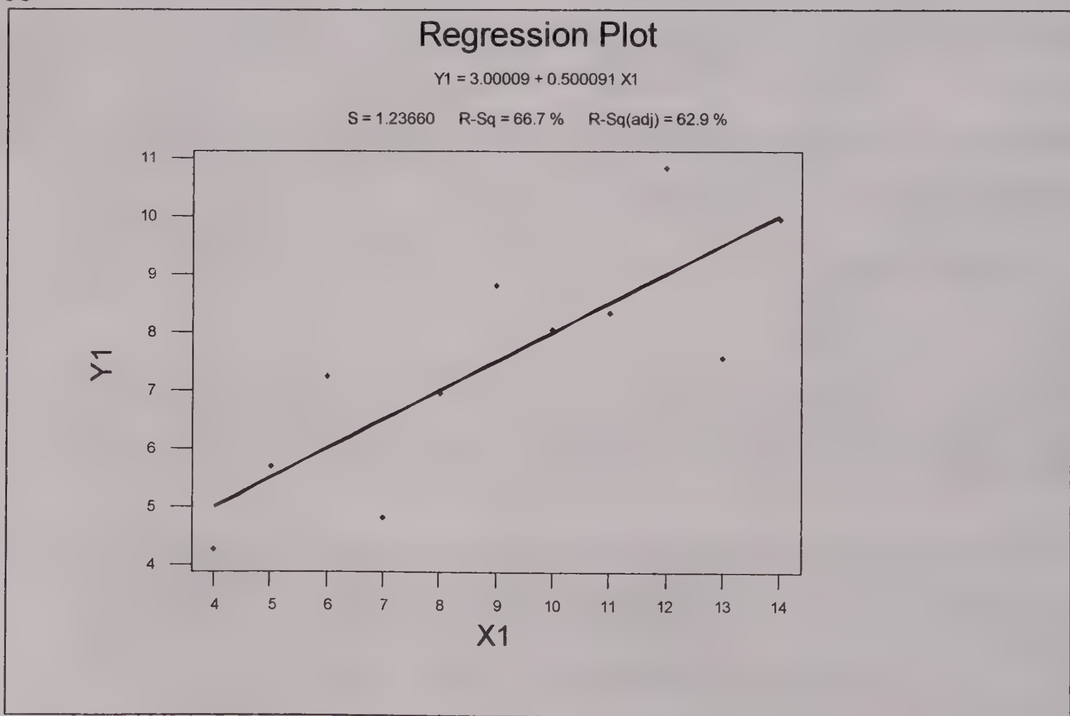
Regression Statistics	
Multiple R	0.477387
R Square	0.227898
Adjusted R Square	0.214586
Standard Error	0.130386
Observations	60

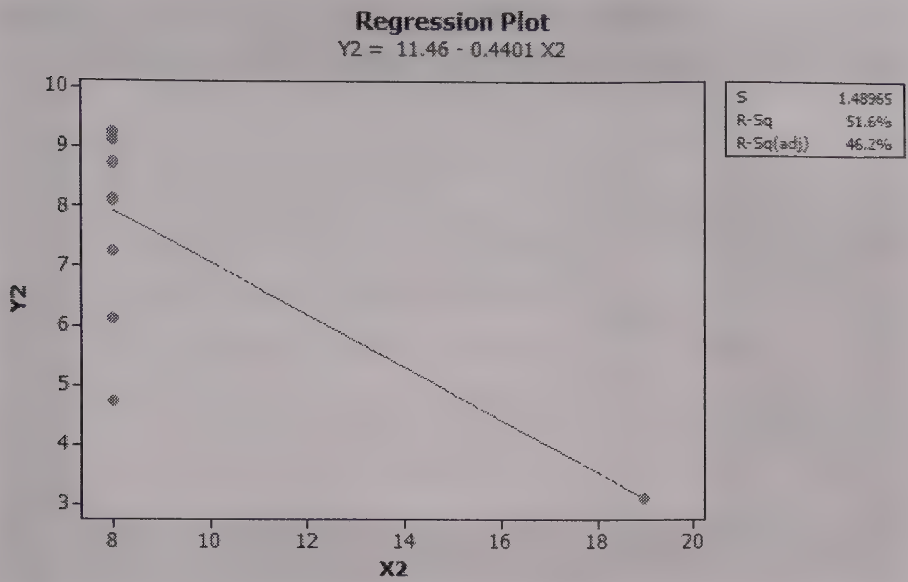
ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.291044447	0.291044	17.11966	0.000115
Residual	58	0.986034863	0.017001		
Total	59	1.27707931			

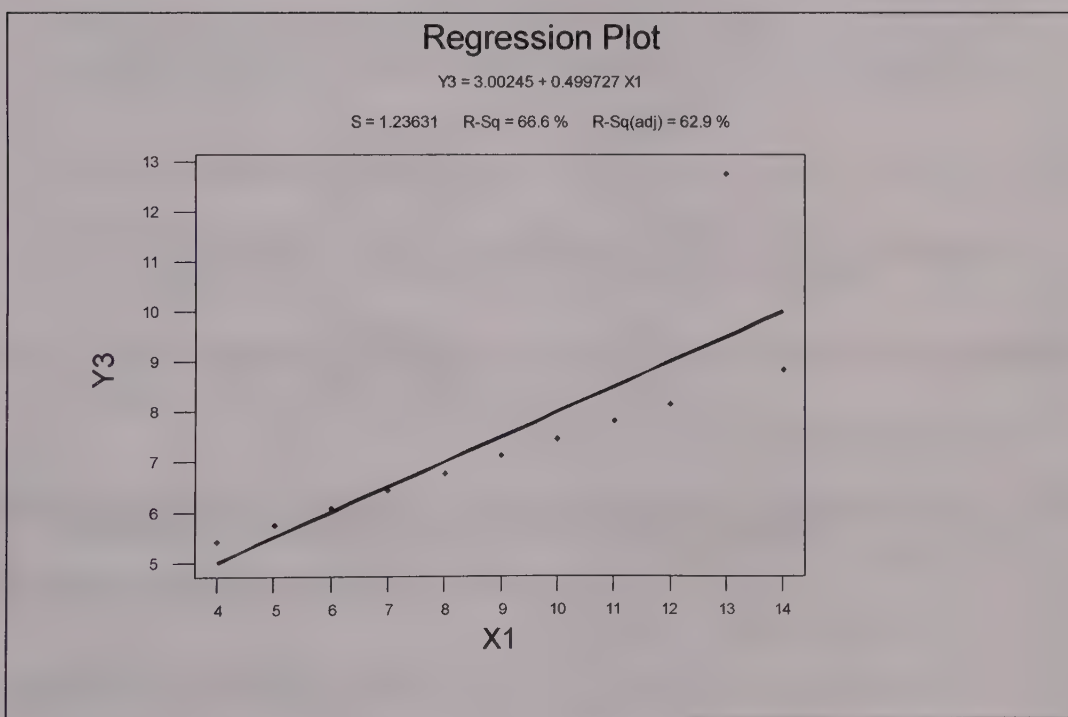
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.003951	0.01740474	0.226996	0.821225	-0.03089	0.03879
SP 500	2.796409	0.675854453	4.137591	0.000115	1.44354	4.149278

11.68

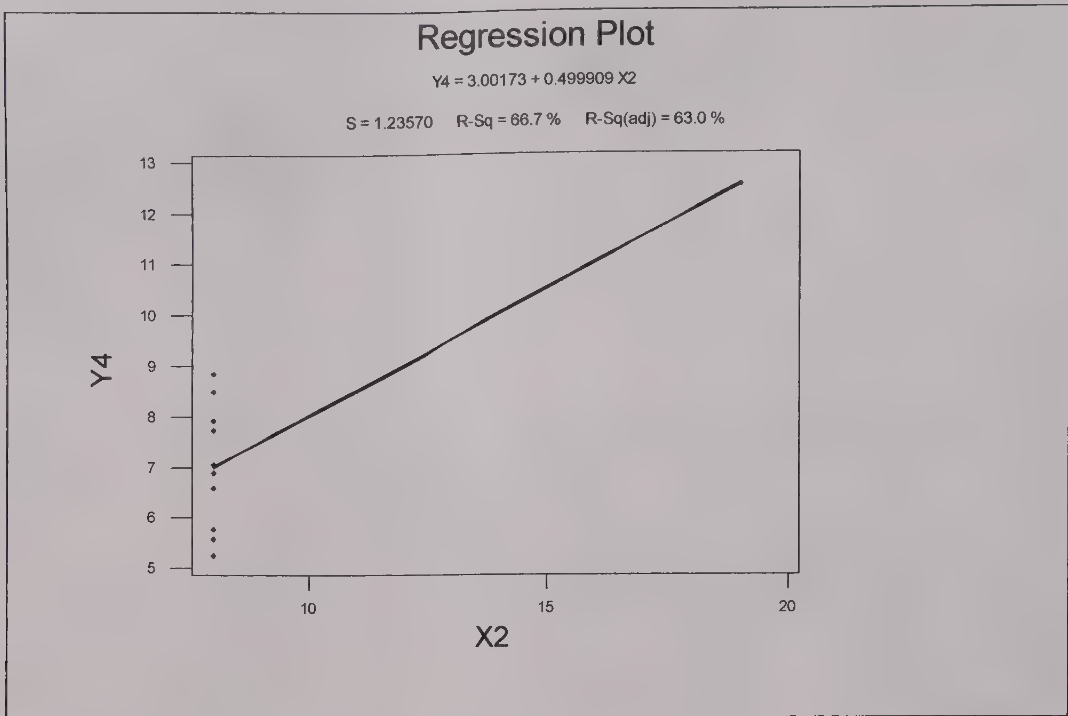




Regression Plot
Y2 = 11.46 - 0.4401 X2
S 1.48965
R-Sq 51.6%
R-Sq(adj) 46.2%
Y2
X2



Regression Plot
Y3 = 3.00245 + 0.499727 X1
S = 1.23631 R-Sq = 66.6 % R-Sq(adj) = 62.9 %
Y3
X1



The model of Y1 = f(X_1) is a good fit for a linear model
The model of Y2 = f(X_2) is a non-linear model
The model of Y3 = f(X_3) has a significant outlier at the largest value of X1
The model of Y4 = f(X_4) has only two values of the independent variable

11.70 $H_o: \rho = 0, H_1: \rho > 0,\ t = \dfrac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} = t = \dfrac{.37\sqrt{51}}{\sqrt{(1-.37^2)}} = 2.844$

Therefore, reject H_0 at the .5% level since t = 2.844 > 2.666 ≈ $t_{51,.005}$

11.72 $H_o: \rho = 0, H_1: \rho > 0,\ t = \dfrac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} = t = \dfrac{.293\sqrt{64}}{\sqrt{(1-.293^2)}} = 2.452$

Therefore, reject H_0 at the 1% level since t = 2.452 > 2.39 ≈ $t_{60,.01}$

11.74 a. For a one unit increase in the inflation rate, we estimate that the actual spot rate will increase by .7916 units.

b. $R^2 = 9.7\%$. 9.7% of the variation in the actual spot rate can be explained by the variations in the inflation rate.

c. $H_o: \beta = 0, H_1: \beta > 0,\ t = \dfrac{.7916}{.2759} = 2.8692$, Reject H_0 at the .5% level since t = 2.8692 > 2.66 = $t_{77,.005}$

d. $H_o: \beta = 0, H_1: \beta \neq 0,\ t = \dfrac{.7916 - 1}{.2759} = -.7553$,

Do not reject H_0 at any common level

11.76

a. For each unit increase in the diagnostic statistics test, we estimate that the final student score at the end of the course will increase by .2875 point.

b. 11.58% of the variation in the final student score can be explained by the variation in the diagnostic statistics test

c. The two methods are 1) the test of the significance of the population regression slope coefficient (β) and 2) the test of the significance of the population correlation coefficient (ρ)

1) $H_o: \beta = 0, H_1: \beta > 0$, $t = \dfrac{.2875}{.04566} = 6.2965$

Therefore, reject H_0 at any common level of alpha

2) $H_o: \rho = 0, H_1: \rho > 0$, $r = \sqrt{R^2} = \sqrt{.1158} = .3403$

$t = \dfrac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} = t = \dfrac{.3403\sqrt{304}}{\sqrt{(1-.3403^2)}} = 6.3098$, Reject H_0 at any common level

11.78 a. $R^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{204}{268} = .2388$. 23.88% of the variation in the dependent variable can be explained by the variation in the independent variable.

b. $$\sum (x_i - \bar{x})^2 = \frac{SST(R^2)}{b^2} = \frac{268(.2388)}{(1.3)^2} = 37.8689$$

$s^2_e = 204/23 = 8.8696$ $\quad s^2_b = 8.8696/37.8689 = .2342$

$H_o: \beta = 0, H_1: \beta \neq 0$, $t = \dfrac{1.3}{\sqrt{.2342}} = 2.6863$

Therefore, reject H_0 at the 5% level since $t = 2.6863 > 2.069 = t_{23,.025}$

c. $1.3 \pm 2.069\sqrt{.2342}$, the interval runs from .2987 up to 2.3013

11.80 If a linear regression was estimated, the slope would be negative and indicate that increased fertilizer reduces yield, which is not realistic.

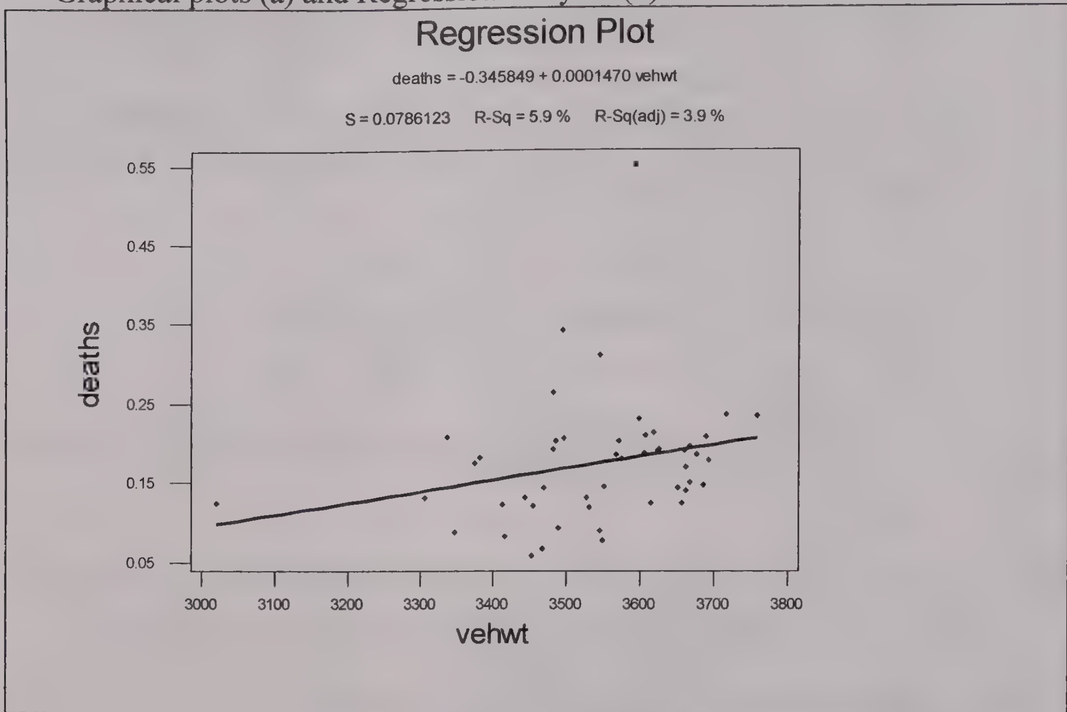
11.82

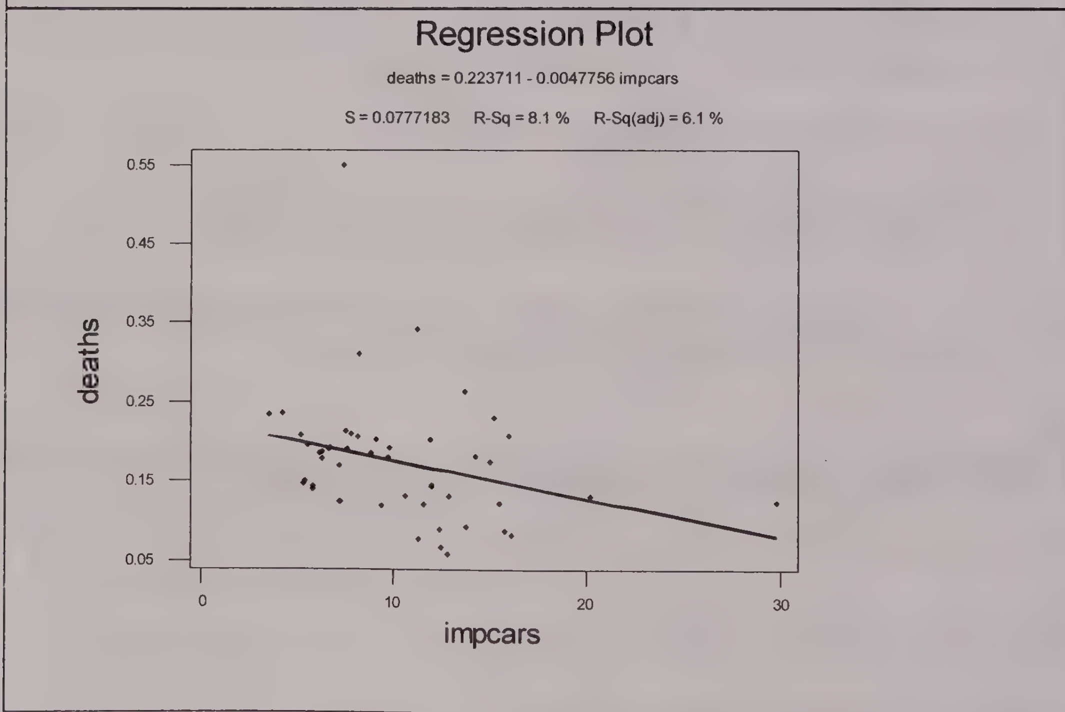
a. Relationships are shown below in the correlation matrix with graphical plots to follow

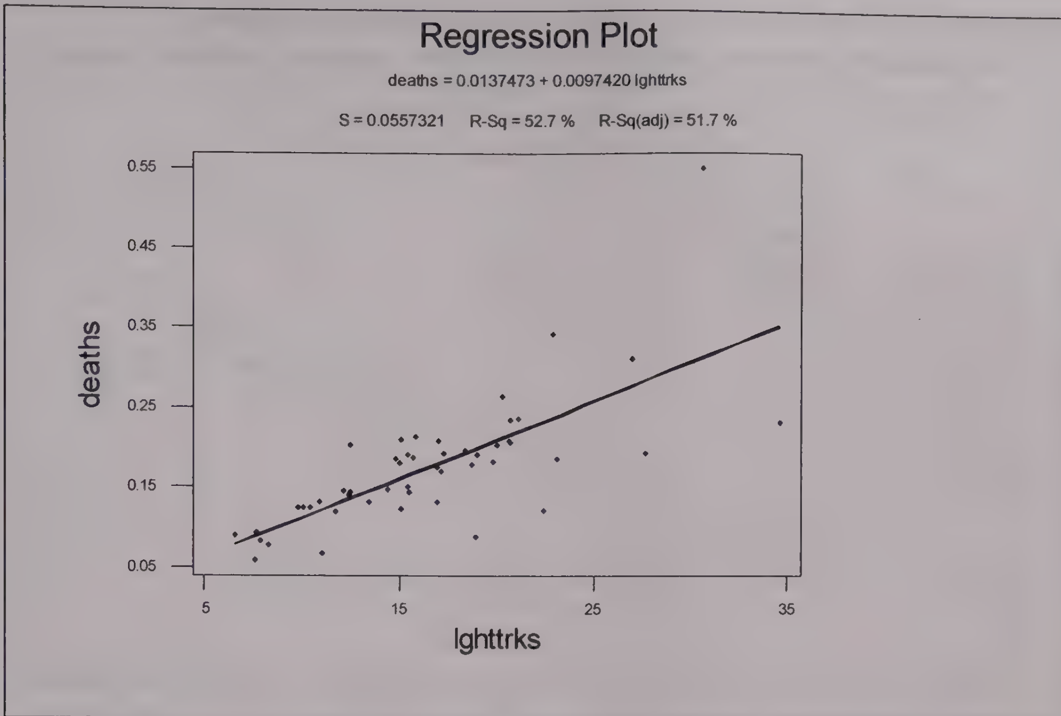
Correlations: deaths, vehwt, impcars, lghttrks, carage

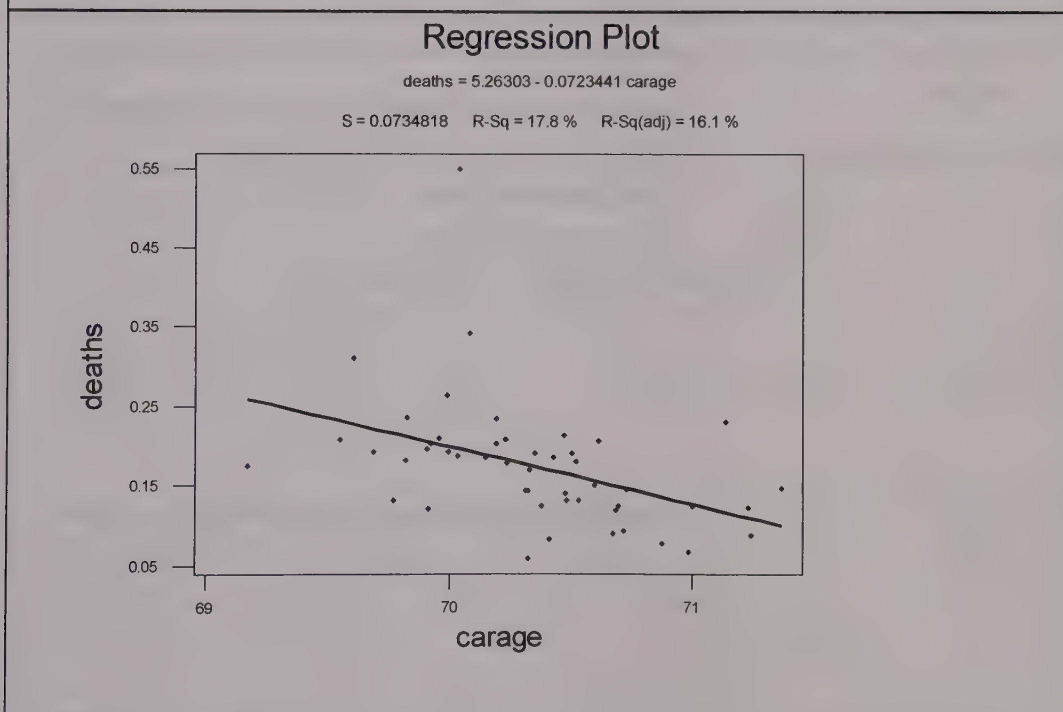
```
          deaths    vehwt  impcars lghttrks
vehwt      0.244
           0.091

impcars   -0.284   -0.943
           0.048    0.000

lghttrks   0.726    0.157   -0.175
           0.000    0.282    0.228

carage    -0.422    0.123    0.011   -0.329
           0.003    0.400    0.943    0.021

Cell Contents: Pearson correlation
               P-Value
```

Unusual data points include outliers of .55 crash deaths. This data point is much higher than expected given the levels of the independent variables.
Graphical plots (a) and Regression analyses (b):









Light trucks has the strongest linear association (52.7%) followed by age (17.8%), then imported cars (8.1%) and then vehicle weight (5.9%).

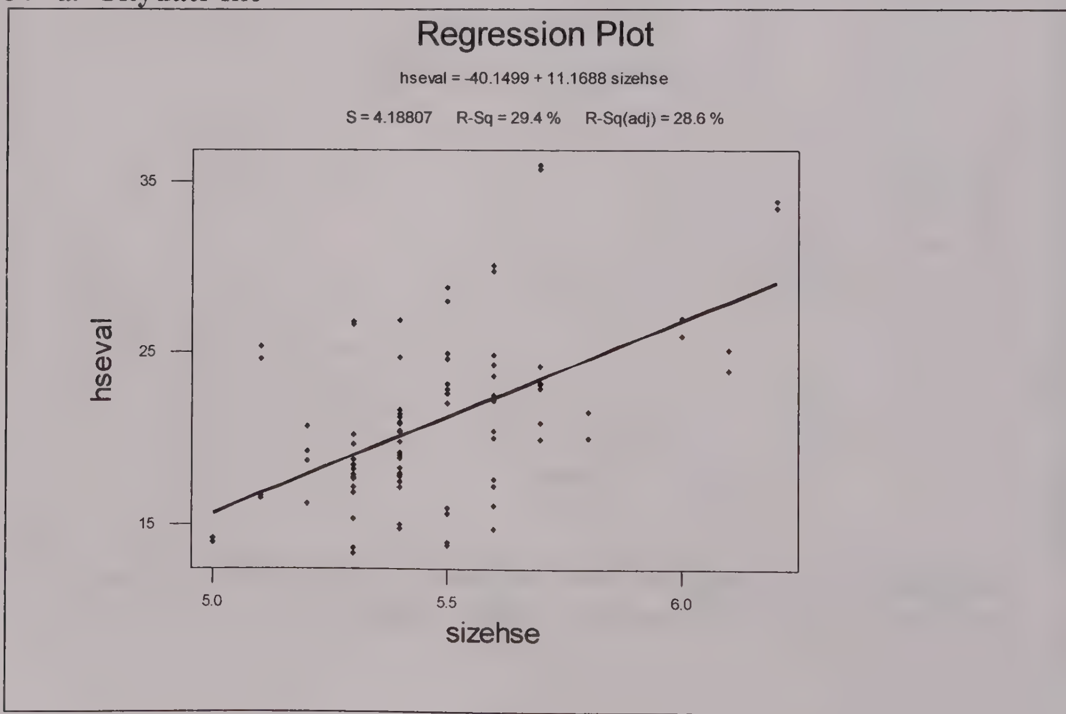
c. Rank predictor variables in terms of their relationship to crash deaths

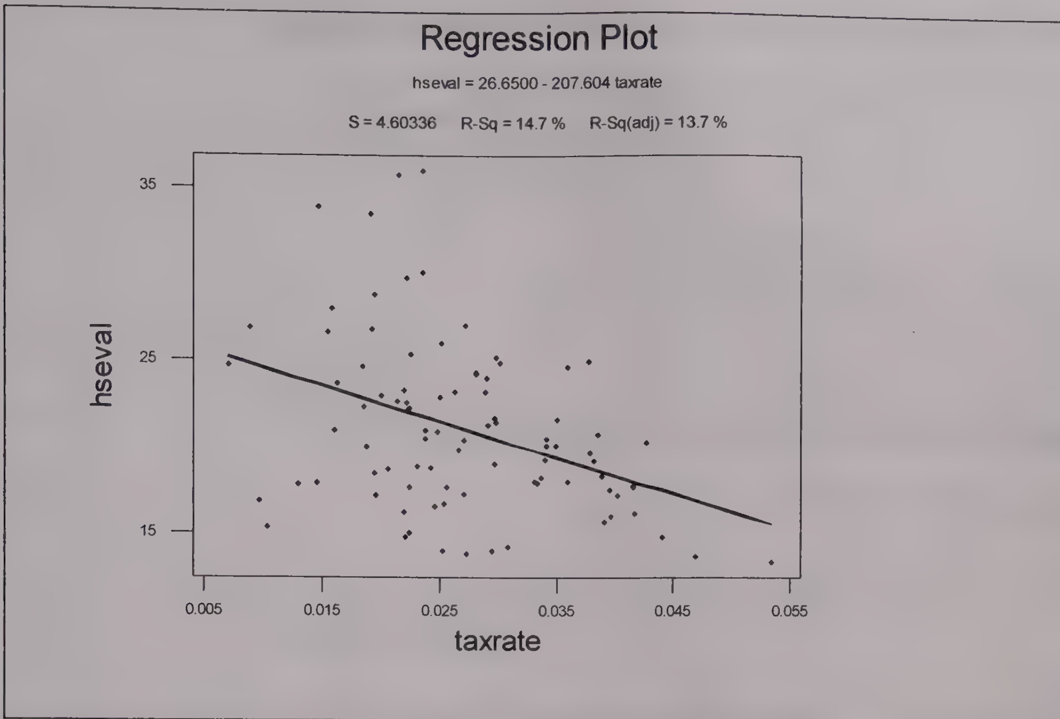
Stepwise Regression: deaths versus vehwt, impcars, lghttrks, carage

```
  Alpha-to-Enter: 0.15   Alpha-to-Remove: 0.15
 Response is  deaths  on  4 predictors, with N =   49

    Step           1         2         3
Constant     0.01375   2.50631   2.55472

lghttrks      0.0097    0.0088    0.0083
T-Value         7.24      6.39      6.02
P-Value        0.000     0.000     0.000

carage                  -0.035    -0.041
T-Value                  -2.00     -2.34
P-Value                  0.052     0.024

vehwt                            0.00011
T-Value                             1.80
P-Value                            0.079

S             0.0557    0.0540    0.0528
R-Sq           52.73     56.49     59.42
R-Sq(adj)      51.72     54.60     56.71
C-p              6.3       4.2       3.1
```

Crash deaths are positively related to both weight and percent of light trucks. Deaths are negatively related to percent import cars and the age of the vehicle. Light trucks has the strongest association followed by age and then vehicle weight.

11.84 a. **Citydatr** file





b.

Regression Analysis: hseval versus sizehse

The regression equation is
hseval = - 40.1 + 11.2 sizehse

Predictor	Coef	SE Coef	T	P
Constant	-40.15	10.11	-3.97	0.000
sizehse	11.169	1.844	6.06	0.000

S = 4.188 R-Sq = 29.4% R-Sq(adj) = 28.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	643.12	643.12	36.67	0.000
Residual Error	88	1543.51	17.54		
Total	89	2186.63			

Regression Analysis: hseval versus taxrate

The regression equation is
hseval = 26.6 - 208 taxrate

Predictor	Coef	SE Coef	T	P
Constant	26.650	1.521	17.52	0.000
taxrate	-207.60	53.27	-3.90	0.000

S = 4.603 R-Sq = 14.7% R-Sq(adj) = 13.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	321.83	321.83	15.19	0.000
Residual Error	88	1864.80	21.19		
Total	89	2186.63			

Size of house is a stronger predictor than is the taxrate.

c. Whether tax rates are lowered or not does not have as large an impact as does the size of the house on the evaluation.

11.86 a. **Regression Analysis: Residential versus Prime Rate**

```
The regression equation is
Residential = 285 + 10.6 Prime Rate


248 cases used, 8 cases contain missing values


Predictor      Coef  SE Coef      T      P
Constant     285.43    22.24  12.83  0.000
Prime Rate   10.598    2.879   3.68  0.000


S = 153.781   R-Sq = 5.2%   R-Sq(adj) = 4.8%
```

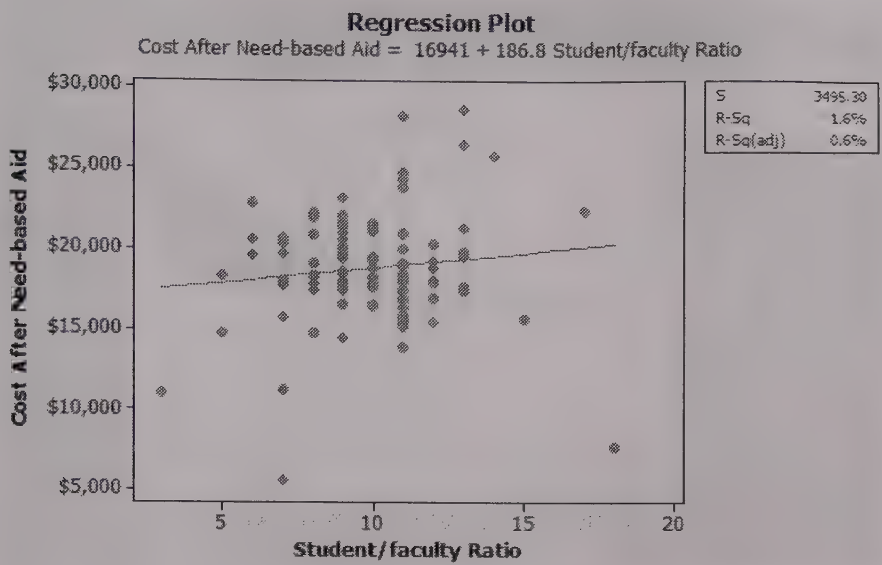
Regression Analysis: Residential versus Fed Funds Rate

```
The regression equation is
Residential = 399 - 3.77 Fed Funds Rate


226 cases used, 30 cases contain missing values


Predictor          Coef  SE Coef      T      P
Constant         399.02    18.96  21.04  0.000
Fed Funds Rate   -3.775    2.961  -1.27  0.204


S = 151.230   R-Sq = 0.7%   R-Sq(adj) = 0.3%
```

The $r^2 = 0.052$ for the regression with the prime rate is higher than the $r^2 = 0.007$ for the regression with the federal rate, so the regression with the prime rate explains more of the variation with the model, but neither is particularly strong. Also, the coefficient for the federal rate could statistically be equal to zero, i.e. the p-value is greater than 0.05.

b. Prime rate: (4.96, 16.24), Federal rate: (-9.58, 2.03)
c. Prime rate 306.63, Federal rate 391.5
d. Prime rate: (272.64, 340.62), Federal rate: (363.3, 419.6)

11.88



Regression Analysis: Cost After Need- versus Student/faculty

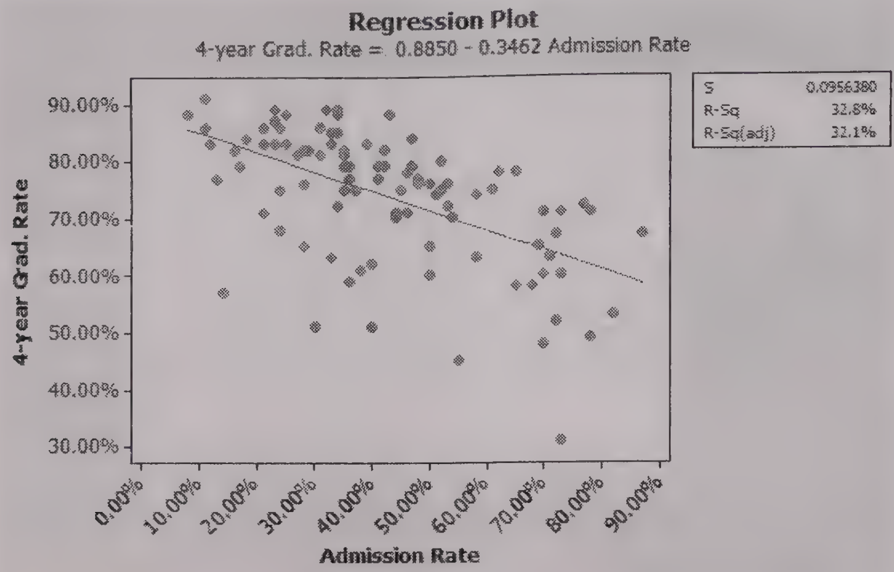
```
The regression equation is
Cost After Need-based Aid = 16941 + 187 Student/faculty Ratio
Predictor                 Coef  SE Coef      T      P
Constant                 16941     1518  11.16  0.000
Student/faculty Ratio   186.8    148.7   1.26  0.212


S = 3495.30   R-Sq = 1.6%   R-Sq(adj) = 0.6%


Analysis of Variance
Source          DF          SS        MS     F      P
Regression       1    19283471  19283471  1.58  0.212
Residual Error  96  1172846720  12217153
Total           97  1192130191
```

For a one unit increase in the student/faculty ratio, we estimate that the total annual cost after need-based financial aid will increase by $186.8.

The coefficient b_0=16941 states that when student/faculty ratio is 0, the expected value of the total annual cost after need-based financial aid will be $16941.

R^2 = 1.6%. 1.6 % of the variation in the total annual cost after need-based financial aid can be explained by variation in the student/faculty ratio.

11.90



Regression Analysis: 4-year Grad. Rate versus Admission Rate

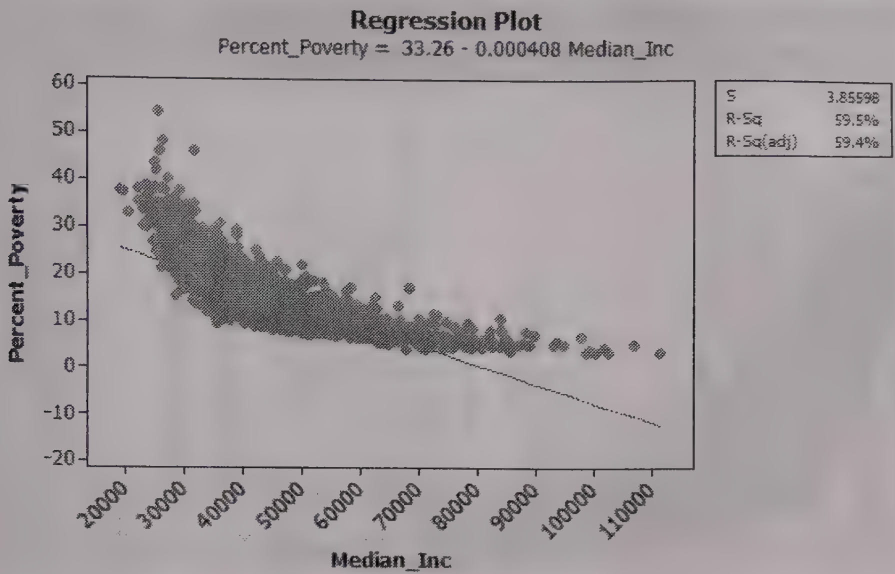
```
The regression equation is
4-year Grad. Rate = 0.885 - 0.346 Admission Rate
Predictor             Coef   SE Coef      T      P
Constant           0.88502   0.02368  37.37  0.000
Admission Rate    -0.34619   0.05063  -6.84  0.000


S = 0.0956380   R-Sq = 32.8%   R-Sq(adj) = 32.1%


Analysis of Variance
Source          DF        SS        MS      F      P
Regression       1   0.42772   0.42772  46.76  0.000
Residual Error  96   0.87808   0.00915
Total           97   1.30580
```

For a one % increase in the students admission rate, we estimate that the 4-year graduation rate will decrease by 0.346% .

The coefficient $b_0 = 0.885$ states that when the students admission rate is 0%, the expected value of the 4-year graduation rate will be 0.885%.

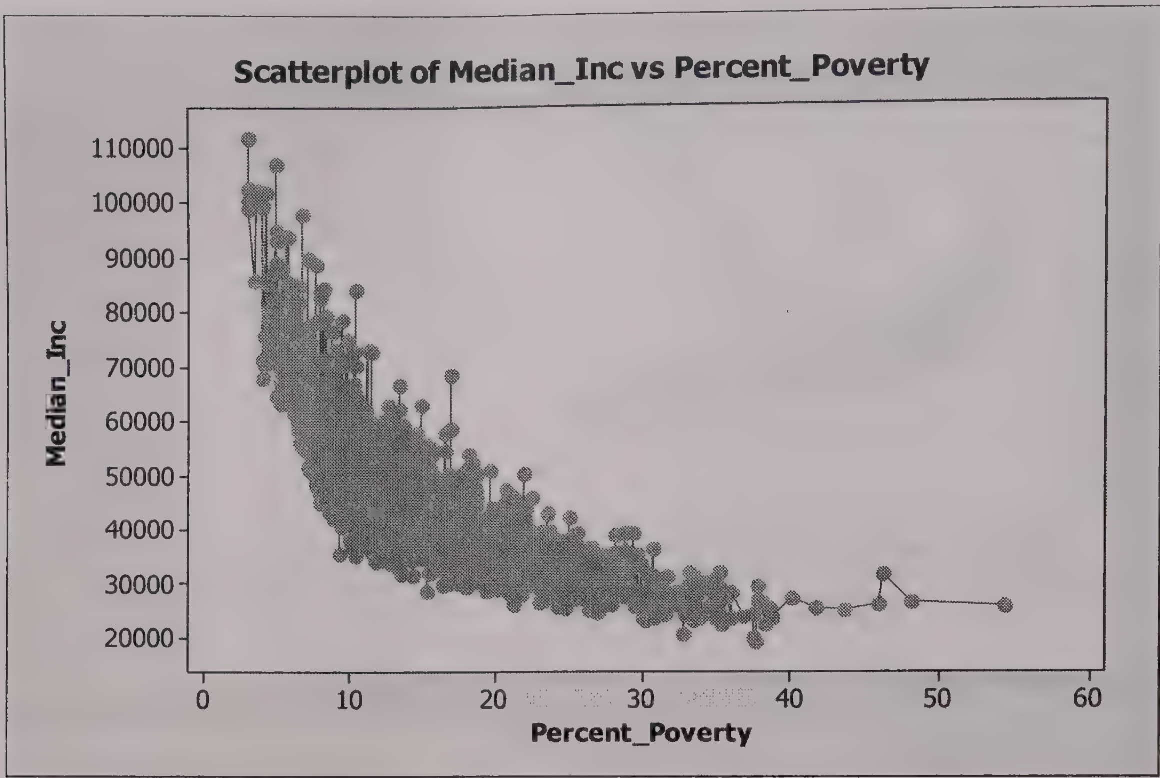
R^2 = 32.8%. 32.8%.of the variation in the 4-year graduation rate can be explained by variation in the students admission rate.

11.92



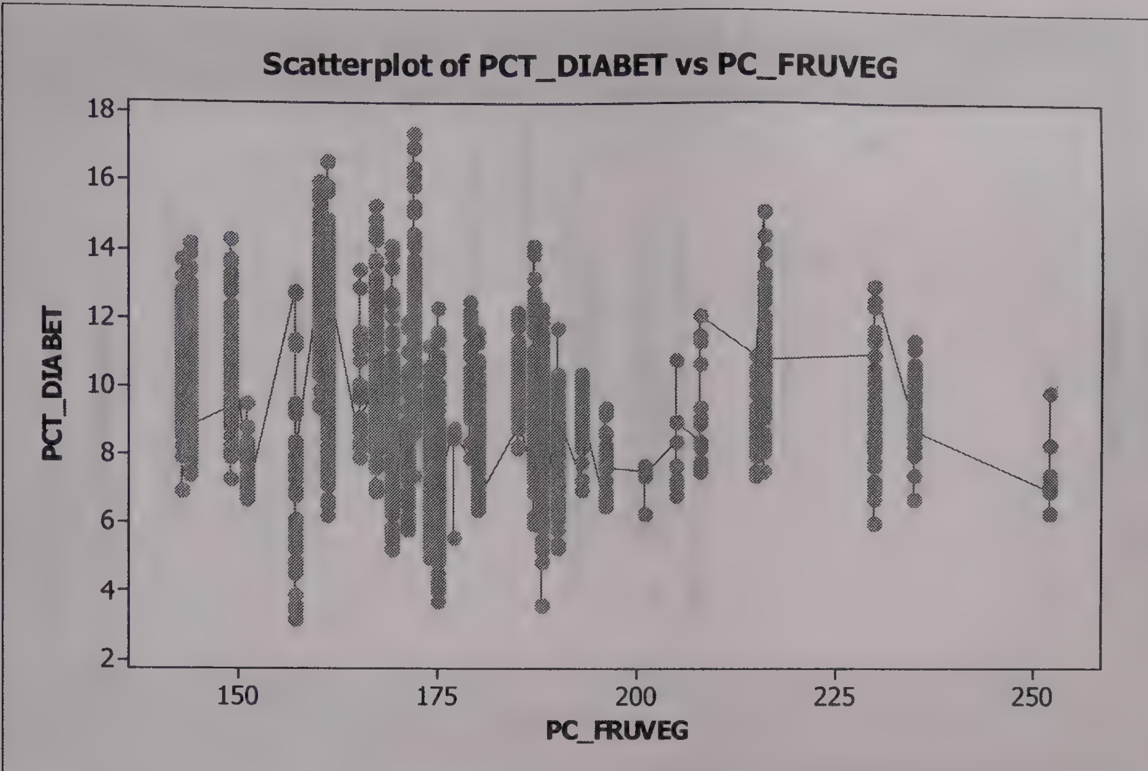
Regression Analysis: Median_Inc versus Percent_Poverty

```
The regression equation is
Median_Inc = 66379 - 1458 Percent_Poverty


3138 cases used, 2 cases contain missing values


Predictor            Coef  SE Coef       T      P
Constant          66378.9    352.5  188.29  0.000
Percent_Poverty  -1457.57    21.49  -67.81  0.000


S = 7289.15   R-Sq = 59.5%   R-Sq(adj) = 59.4%


Analysis of Variance

Source            DF           SS           MS        F      P
Regression         1  2.44313E+11  2.44313E+11  4598.25  0.000
Residual Error  3136  1.66621E+11     53131763
Total           3137  4.10934E+11
```



For a one % increase in the poverty level, we estimate that the median income will decrease by $1457.57.
The coefficient $b_0 = 66378.9$ states that when the poverty level is 0%, the expected value of the median income will be $66378.9.
$R^2 = 59.5\%$. 59.5% of the variation in the poverty level can be explained by variation in the median income.

11.94



Correlations: PC_FRUVEG, PCT_DIABET

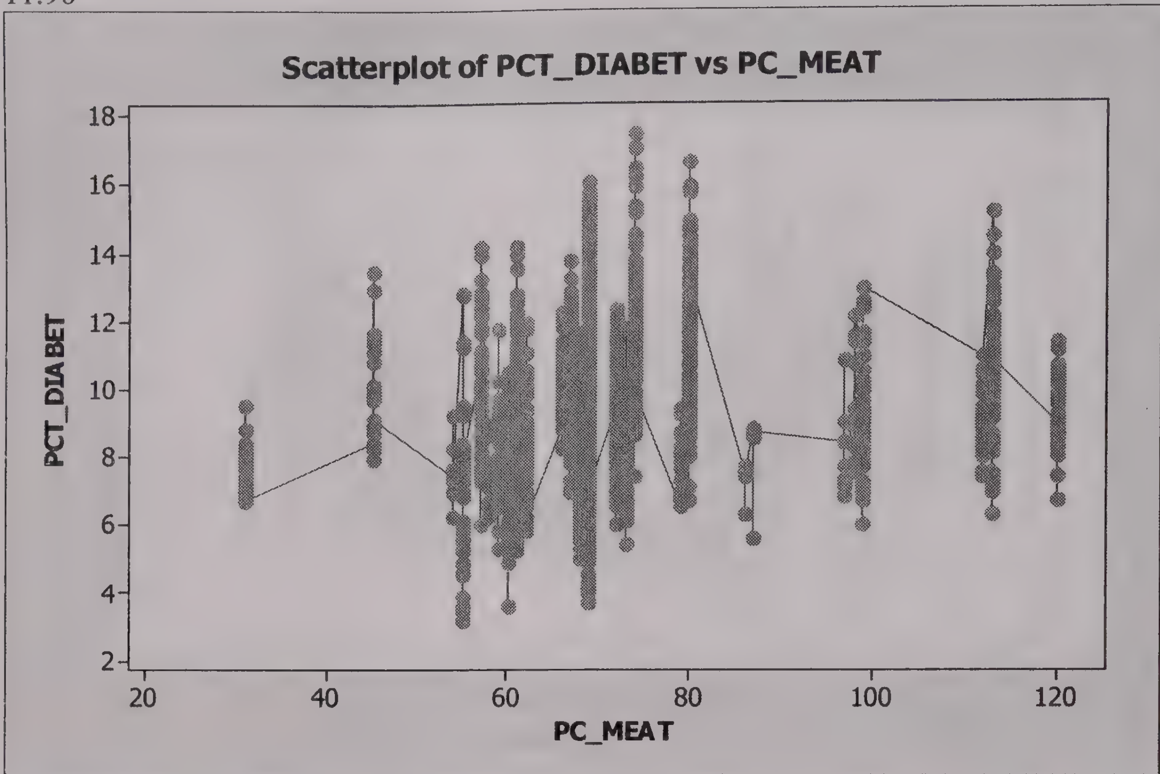
```
Pearson correlation of PC_FRUVEG and PCT_DIABET = -0.236
P-Value = 0.000
```

Regression Analysis: PCT_DIABET versus PC_FRUVEG

```
The regression equation is
PCT_DIABET = 13.9 - 0.0245 PC_FRUVEG
3108 cases used, 32 cases contain missing values
Predictor       Coef    SE Coef       T      P
Constant     13.8867     0.3158   43.97  0.000
PC_FRUVEG  -0.024532   0.001817  -13.50  0.000


S = 1.94954   R-Sq = 5.5%   R-Sq(adj) = 5.5%


Analysis of Variance
Source            DF        SS      MS       F      P
Regression         1    693.17  693.17  182.38  0.000
Residual Error  3106  11804.99    3.80
Total           3107  12498.15
```

For a one pound increase in the per capita consumption of fruits and vegetables, we estimate that the percentage of adults with diabetes will decrease by 0.0245%.
The coefficient $b_0 = 13.9$ states that when the per capita consumption of fruits and vegetables is 0 lbs, the expected value of the percentage of adults with diabetes will be 34.6%.
$R^2 = 5.5\%$. 5.5% of the variation in the percentage of adults with diabetes can be explained by variation in the per capita consumption of fruits and vegetables.

11.96



Correlations: PC_MEAT, PCT_DIABET

```
Pearson correlation of PC_MEAT and PCT_DIABET = 0.205
P-Value = 0.000
```

Regression Analysis: PCT_DIABET versus PC_MEAT

```
The regression equation is
PCT_DIABET = 7.37 + 0.0324 PC_MEAT
3108 cases used, 32 cases contain missing values
Predictor       Coef    SE Coef      T      P
Constant      7.3675     0.1985  37.12  0.000
PC_MEAT     0.032397   0.002775  11.68  0.000


S = 1.96334   R-Sq = 4.2%   R-Sq(adj) = 4.2%


Analysis of Variance
Source            DF        SS      MS       F      P
Regression         1    525.44  525.44  136.31  0.000
Residual Error  3106  11972.72    3.85
Total           3107  12498.15
```

For a one pound increase in the per capita consumption of meat, we estimate that the percentage of adults with diabetes will increase by 0.0324%. .

The coefficient $b_0 = 7.37$ states that when the per capita consumption of fruits and vegetables is 0 lbs, the expected value of the percentage of adults with diabetes will be 7.37%.

$R^2 = 4.2\%$. 4.2% of the variation in the percentage of adults with diabetes can be explained by variation in the per capita consumption of meat.

11.98

Presuming the populations are normally distributed with equal variances, the samples must be independent random samples:

Difference in the cost of healthy and less healthy diet in the first interview

$H_0: \mu_x - \mu_y = 0; H_1: \mu_x - \mu_y > 0$; reject H_0 if $t_{.05} > 1.645$

$n_x = 2173$, $n_y = 2049$, $s_x = 159.34, s_y = 159$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{2172(159.34)^2 + 2048(159)^2}{2173 + 2049 - 2} = 25336.52$$

$$t = \frac{\bar{y} - \bar{x}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{264.86 - 249.76}{\sqrt{\frac{25336.52}{2173} + \frac{25336.52}{2049}}} = 3.08$$

Reject H_0 at 5% level.

```
The regression equation is
HEI2005 = 50.7 + 0.00517 PIR_p

4222 cases used, 238 cases contain missing values


Predictor       Coef    SE Coef       T      P
Constant     50.7101     0.4143  122.41  0.000
PIR_p       0.005169   0.001368    3.78  0.000

S = 14.1611   R-Sq = 0.3%   R-Sq(adj) = 0.3%


Analysis of Variance

Source            DF        SS      MS      F      P
Regression         1    2863.0  2863.0  14.28  0.000
Residual Error  4220  846260.7   200.5
Total           4221  849123.7
```

For each % increase in the poverty income level, we expect the HEI score to increase by .0052, but with an R^2 of .003, it is not a very strong relationship.

Difference in the cost of healthy and less healthy diet in the second interview

$H_0: \mu_y - \mu_x = 0; H_1: \mu_y - \mu_x > 0$; reject H_0 if $t_{.05} > 1.645$

$n_x = 2279$, $n_y = 1644$, $s_x = 158.01, s_y = 159.82$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{2278(158.01)^2 + 1643(159.82)^2}{2279 + 1644 - 2} = 25207.97$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = \frac{269.02 - 249.96}{\sqrt{\frac{25207.97}{2279} + \frac{25207.97}{1644}}} = 3.71$$

Reject H_0 at 5% level

Hence we reject H_0 in both the interviews.

```
The regression equation is
HEI2005 = 52.7 + 0.00620 PIR_p


3923 cases used, 207 cases contain missing values


Predictor       Coef    SE Coef       T      P
Constant     52.7326     0.4461  118.22  0.000
PIR_p       0.006200   0.001459    4.25  0.000


S = 14.5340   R-Sq = 0.5%   R-Sq(adj) = 0.4%


Analysis of Variance

Source            DF        SS      MS      F      P
Regression         1    3812.8  3812.8  18.05  0.000
Residual Error  3921  828258.4   211.2
Total           3922  832071.2
```

For each % increase in the poverty income level, we expect the HEI score to increase by .0062, but with an R^2 of .005, it is not a very strong relationship,

11.100

Regression Analysis: HEI2005 versus p_ate_at_home in the first interview

```
The regression equation is
HEI2005 = 46.9 + 0.0746 p_ate_at_home
Predictor          Coef   SE Coef      T      P
Constant        46.8806    0.4850  96.66  0.000
p_ate_at_home  0.074620  0.006364  11.73  0.000


S = 13.9858   R-Sq = 3.0%   R-Sq(adj) = 3.0%

Analysis of Variance

Source            DF      SS     MS       F      P
Regression         1   26891  26891  137.48  0.000
Residual Error  4458  871990    196
Total           4459  898881
```

For each one % increase in the calories consumed at home, we estimate that the quality of diet will increase by 0.0746 units.

R^2 = 3.0%. 3.0% of the variation in the quality of diet can be explained by variation in the percent of calories consumed at home .
The percent of calories consumed at home has a statistically significant influence on the quality of diet (p = 0.000).

Regression Analysis: HEI2005 versus p_ate_at_home in the second interview

```
The regression equation is
HEI2005 = 49.5 + 0.0687 p_ate_at_home
Predictor         Coef   SE Coef      T      P
Constant       49.5131    0.5417  91.40  0.000
p_ate_at_home 0.068686  0.006973   9.85  0.000


S = 14.3928   R-Sq = 2.3%   R-Sq(adj) = 2.3%

Analysis of Variance

Source            DF      SS     MS      F      P
Regression         1   20097  20097  97.01  0.000
Residual Error  4128  855130    207
Total           4129  875227
```

For each one % increase in the calories consumed at home, we estimate that the quality of diet will increase by 0.0687 units.

$R^2 = 2.3\%$. 2.3% of the variation in the quality of diet can be explained by variation in the percent of calories consumed at home.
The percent of calories consumed at home has significant influence on the quality of diet ($p = 0.000$).

Hence the percent of calories consumed at home has positive effect on the quality of diet ($p = 0.000$) in both the interviews.

Regression Analysis: daily_cost versus p_ate_at_home in the first interview

```
The regression equation is
daily_cost = 7.02 - 0.0262 p_ate_at_home
Predictor           Coef   SE Coef       T      P
Constant          7.0236    0.1021   68.82  0.000
p_ate_at_home  -0.026240  0.001339  -19.60  0.000


S = 2.94275   R-Sq = 7.9%   R-Sq(adj) = 7.9%


Analysis of Variance
Source            DF       SS      MS       F      P
Regression         1   3325.2  3325.2  383.98  0.000
Residual Error  4458  38605.4     8.7
Total           4459  41930.6
```

For each one % increase in the calories consumed at home, we estimate that the daily cost of food will decrease by \$0.0262.

$R^2 = 7.9\%$. 7.9% of the variation in the daily cost of food can be explained by variation in the percent of calories consumed at home .
The percent of calories consumed at home has a statistically significant influence on the daily cost of food ($p = 0.000$).

Regression Analysis: daily_cost versus p_ate_at_home in the second interview

```
The regression equation is
daily_cost = 6.74 - 0.0245 p_ate_at_home
Predictor            Coef    SE Coef       T      P
Constant           6.7434     0.1054   63.95  0.000
p_ate_at_home   -0.024549   0.001357  -18.08  0.000


S = 2.80160   R-Sq = 7.3%   R-Sq(adj) = 7.3%

Analysis of Variance
Source            DF       SS       MS       F      P
Regression         1   2567.1   2567.1  327.07  0.000
Residual Error  4128  32400.5      7.8
Total           4129  34967.6
```

For each one % increase in the calories consumed at home, we estimate that the daily cost of food will decrease by $0.0245.

R^2 = 7.3%. 7.3% of the variation in the daily cost of food can be explained by variation in the percent of calories consumed at home.
The percent of calories consumed at home has a statistically significant influence on the daily cost of food ($p = 0.000$).
Hence the percent of calories consumed at home has negative effect on the daily cost of food ($p = 0.000$) in both the interviews.

Chapter 12:

Multiple Regression

12.2 Given the following estimated linear model: $\hat{y} = 10 + 5x_1 + 4x_2 + 2x_3$.

a. $\hat{y} = 10 + 5(20) + 4(11) + 2(10) = 174$
b. $\hat{y} = 10 + 5(15) + 4(14) + 2(20) = 181$
c. $\hat{y} = 10 + 5(35) + 4(19) + 2(25) = 311$
d. $\hat{y} = 10 + 5(10) + 4(17) + 2(30) = 188$

12.4 Given the following estimated linear model: $\hat{y} = 10 + 2x_1 + 12x_2 + 8x_3$

a. $\hat{y}$ increases by 8
b. $\hat{y}$ increases by 8
c. $\hat{y}$ increases by 24

12.6 The estimated regression slope coefficients are interpreted as follows:

$b_1 = .661$: All else being equal, an increase in the plane's top speed by one mph will increase the expected number of hours in the design effort by an estimated .661 million or 661 thousand worker-hours.

$b_2 = .065$: All else being equal, an increase in the plane's weight by one ton will increase the expected number of hours in the design effort by an estimated .065 million or 65 thousand worker-hours.

$b_3 = -.018$: All else being equal, an increase in the percentage of parts in common with other models will result in a decrease in the expected number of hours in the design effort by an estimated .018 million or 18 thousand worker-hours.

12.8 a. $b_1 = .052$: All else being equal, an increase of one hundred dollars in weekly income results in an estimated .052 quart per week increase in milk consumption. $b_2 = 1.14$: All else being equal, an increase in family size by one person will result in an estimated increase in milk consumption by 1.14 quarts per week.

b. The intercept term b_0 of -.025 is the estimated milk consumption of quarts of milk per week given that the family's weekly income is 0 dollars and there are 0 members in the family. This is likely extrapolating beyond the observed data series and is not a useful interpretation.

12.10 Compute the slope coefficients for the model: $\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i}$

Given that $b_1 = \dfrac{s_y(r_{x_1y} - r_{x_1x_2} r_{x_2y})}{s_{x_1}(1 - r^2_{x_1x_2})}$, $b_2 = \dfrac{s_y(r_{x_2y} - r_{x_1x_2} r_{x_1y})}{s_{x_2}(1 - r^2_{x_1x_2})}$

a. $b_1 = \dfrac{400(.60 - (.50)(.70))}{200(1 - .50^2)} = .667,$

$b_2 = \dfrac{400(.70 - (.50)(.60))}{100(1 - .50^2)} = 2.133$

b. $b_1 = \dfrac{400(-.60 - (-.50)(.70))}{200(1 - (-.50)^2)} = -.667,$

$b_2 = \dfrac{400(.70 - (-.50)(-.60))}{100(1 - (-.50)^2)} = 2.133$

c. $b_1 = \dfrac{400(.40 - (.80)(.45))}{200(1 - (.80)^2)} = .222,$

$b_2 = \dfrac{400(.45 - (.80)(.40))}{100(1 - (.80)^2)} = 1.444$

d. $b_1 = \dfrac{400(.60 - (-.60)(-.50))}{200(1 - (-.60)^2)} = .9375,$

$b_2 = \dfrac{400(-.50 - (-.60)(.60))}{100(1 - (-.60)^2)} = -.875$

12.12 a. Electricity sales as a function of number of customers and price

Regression Analysis: SalesMwh versus Pricelec, Numcust

```
The regression equation is
SalesMwh = - 584616 + 15421 Pricelec + 2.31 Numcust

Predictor      Coef   SE Coef       T      P
Constant    -584616    267660   -2.18  0.033
Pricelec      15421     21103    0.73  0.468
Numcust      2.3068    0.2007   11.49  0.000

S = 64625.1    R-Sq = 81.6%    R-Sq(adj) = 81.0%

Analysis of Variance

Source           DF           SS           MS       F      P
Regression        2  1.20344E+12  6.01719E+11  144.08  0.000
Residual Error   65  2.71466E+11   4176402085
Total            67  1.47490E+12
```

All else being equal, for every one unit increase in the price of electricity, we estimate that sales will increase by 15,421 mwh. Note that this estimated coefficient is not significantly different from zero (p-value = .468).
All else being equal, for every additional residential customer who uses electricity in the heating of their home, we estimate that sales will increase by 2.31 mwh.

b. Electricity sales as a function of number of customers

Regression Analysis: salesmw2 versus numcust2

```
The regression equation is
salesmw2 = - 410202 + 2.20 numcust2
Predictor        Coef      SE Coef          T        P
Constant      -410202       114132      -3.59    0.001
numcust2       2.2027       0.1445      15.25    0.000

S = 66282        R-Sq = 78.9%     R-Sq(adj) = 78.6%
Analysis of Variance
Source              DF          SS          MS          F         P
Regression           1 1.02136E+12 1.02136E+12     232.48     0.000
Residual Error      62 2.72381E+11  4393240914
Total               63 1.29374E+12
```

Regression Analysis: SalesMwh versus Numcust

```
The regression equation is
SalesMwh = - 404100 + 2.19 Numcust

Predictor     Coef  SE Coef      T      P
Constant   -404100   102674  -3.94  0.000
Numcust     2.1947   0.1290  17.02  0.000

S = 64396.5   R-Sq = 81.4%   R-Sq(adj) = 81.2%

Analysis of Variance

Source          DF           SS           MS       F      P
Regression       1  1.20121E+12  1.20121E+12  289.66  0.000
Residual Error  66  2.73696E+11   4146912960
Total           67  1.47490E+12
```

An additional residential customer will add 2.19 mwh to electricity sales.
The two models have roughly equivalent explanatory power; therefore, adding price as a variable does not add a significant amount of explanatory power to the model. There appears to be a problem of high correlation between the independent variables of price and customers.

c. Electricity sales as a function of price and degree days

Regression Analysis: SalesMwh versus Pricelec, Degreday

```
The regression equation is
SalesMwh = 2370438 - 173882 Pricelec + 56.7 Degreday

Predictor     Coef  SE Coef      T      P
Constant   2370438   142346  16.65  0.000
Pricelec   -173882    23870  -7.28  0.000
Degreday     56.69    58.65   0.97  0.337

S = 111735   R-Sq = 45.0%   R-Sq(adj) = 43.3%

Analysis of Variance

Source          DF           SS           MS      F      P
Regression       2  6.63398E+11  3.31699E+11  26.57  0.000
Residual Error  65  8.11506E+11  12484705377
Total           67  1.47490E+12
```

All else being equal, an increase in the price of electricity will reduce electricity sales by 173,882 mwh.

All else being equal, an increase in the degree days (departure from normal weather) by one unit will increase electricity sales by 56.7 mwh.
Note that the coefficient on the price variable is now negative, as expected, and it is significantly different from zero (p-value = .000)

d. Electricity sales as a function of disposable income and degree days

Regression Analysis: SalesMwh versus YD87, Degreday

```
The regression equation is
SalesMwh = 317513 + 318 YD87 + 57.1 Degreday

Predictor      Coef  SE Coef      T      P
Constant     317513    60649   5.24  0.000
YD87         317.91    18.73  16.97  0.000
Degreday      57.06    33.70   1.69  0.095

S = 64613.5   R-Sq = 81.6%   R-Sq(adj) = 81.0%

Analysis of Variance

Source          DF           SS           MS       F      P
Regression       2  1.20353E+12  6.01767E+11  144.14  0.000
Residual Error  65  2.71369E+11   4174903356
Total           67  1.47490E+12
```

All else being equal, an increase in personal disposable income by one unit will increase electricity sales by 318 mwh.
All else being equal, an increase in degree days by one unit will increase electricity sales by 57.1 mwh.

12.14 a. Horsepower as a function of weight, cubic inches of displacement

Regression Analysis: horspwr versus weight, displace

```
The regression equation is
horspwr = 23.5 + 0.0154 weight + 0.157 displace
151 cases used 4 cases contain missing values
Predictor        Coef    SE Coef      T      P    VIF
Constant       23.496      7.341   3.20  0.002
weight       0.015432   0.004538   3.40  0.001    6.0
displace      0.15667    0.03746   4.18  0.000    6.0

S = 13.64      R-Sq = 69.2%     R-Sq(adj) = 68.8%
Analysis of Variance
Source           DF       SS       MS       F      P
Regression        2    61929    30964  166.33  0.000
Residual Error  148    27551      186
Total           150    89480
```

All else being equal, a 100 pound increase in the weight of the car is associated with a 1.54 increase in horsepower of the auto.
All else being equal, a 10 cubic inch increase in the displacement of the engine is associated with a 1.57 increase in the horsepower of the auto.

b. Horsepower as a function of weight, displacement, number of cylinders

Regression Analysis: horspwr versus weight, displace, cylinder

```
The regression equation is
horspwr = 16.7 + 0.0163 weight + 0.105 displace + 2.57 cylinder
151 cases used 4 cases contain missing values
Predictor        Coef     SE Coef        T       P      VIF
Constant       16.703       9.449     1.77   0.079
weight       0.016261    0.004592     3.54   0.001      6.2
displace      0.10527     0.05859     1.80   0.074     14.8
cylinder        2.574       2.258     1.14   0.256      7.8

S = 13.63       R-Sq = 69.5%     R-Sq(adj) = 68.9%
Analysis of Variance
Source            DF        SS        MS        F       P
Regression         3     62170     20723   111.55   0.000
Residual Error   147     27310       186
Total            150     89480
```

All else being equal, a 100 pound increase in the weight of the car is associated with a 1.63 increase in horsepower of the auto.

All else being equal, a 10 cubic inch increase in the displacement of the engine is associated with a 1.05 increase in the horsepower of the auto.

All else being equal, one additional cylinder in the engine is associated with a 2.57 increase in the horsepower of the auto.

Note that adding the independent variable number of cylinders has not added to the explanatory power of the model. R square has increased marginally. Engine displacement is no longer significant at the .05 level (p-value of .074) and the estimated regression slope coefficient on the number of cylinders is not significantly different from zero. This is due to the strong correlation that exists between cubic inches of engine displacement and the number of cylinders.

c. Horsepower as a function of weight, displacement and fuel mileage

Regression Analysis: horspwr versus weight, displace, milpgal

```
The regression equation is
horspwr = 93.6 + 0.00203 weight + 0.165 displace - 1.24 milpgal
150 cases used 5 cases contain missing values
Predictor        Coef     SE Coef        T       P      VIF
Constant        93.57       15.33     6.11   0.000
weight       0.002031    0.004879     0.42   0.678      8.3
displace      0.16475     0.03475     4.74   0.000      6.1
milpgal       -1.2392      0.2474    -5.01   0.000      3.1

S = 12.55       R-Sq = 74.2%     R-Sq(adj) = 73.6%
Analysis of Variance
Source            DF        SS        MS        F       P
Regression         3     66042     22014   139.77   0.000
Residual Error   146     22994       157
Total            149     89036
```

All else being equal, a 100 pound increase in the weight of the car is associated with a .203 increase in horsepower of the auto.

All else being equal, a 10 cubic inch increase in the displacement of the engine is associated with a 1.6475 increase in the horsepower of the auto.

All else being equal, an increase in the fuel mileage of the vehicle by 1 mile per gallon is associated with a reduction in horsepower of 1.2392.

Note that the negative coefficient on fuel mileage indicates the trade-off that is expected between horsepower and fuel mileage. The displacement variable is significantly positive, as expected, however, the weight variable is no longer significant. Again, one would expect high correlation among the independent variables.

d. Horsepower as a function of weight, displacement, mpg and price

Regression Analysis: horspwr versus weight, displace, milpgal, price

```
The regression equation is
horspwr = 98.1 - 0.00032 weight + 0.175 displace - 1.32 milpgal
+0.000138 price
150 cases used 5 cases contain missing values
Predictor          Coef       SE Coef          T         P        VIF
Constant          98.14         16.05       6.11     0.000
weight        -0.000324      0.005462      -0.06     0.953       10.3
displace        0.17533       0.03647       4.81     0.000        6.8
milpgal         -1.3194        0.2613      -5.05     0.000        3.5
price         0.0001379     0.0001438       0.96     0.339        1.3

S = 12.55         R-Sq = 74.3%        R-Sq(adj) = 73.6%
Analysis of Variance
Source             DF           SS          MS         F          P
Regression          4        66187       16547    105.00      0.000
Residual Error    145        22849         158
Total             149        89036
```

All else being equal, a 100 pound increase in the weight of the car is associated with a reduction of .00324 in horsepower of the auto.

All else being equal, a 10 cubic inch increase in the displacement of the engine is associated with a 1.7533 increase in the horsepower of the auto.

All else being equal, an increase in the fuel mileage of the vehicle by 1 mile per gallon is associated with a reduction in horsepower of 1.3194.

All else being equal, an increase in the price of the auto by one dollar will increase fuel mileage by .0001379 mpg.

Engine displacement has a significant positive impact on horsepower, fuel mileage is negatively related to horsepower and price is not significant.

e. Explanatory power has marginally increased from the first model to the last. The estimated coefficient on price is not significantly different from zero. Displacement and fuel mileage have the expected signs. The coefficient on weight has the wrong sign; however, it is not significantly different from zero (p-value of .953).

12.16 From the given Analysis of Variance table,

$$SST = SSR + SSE,\ s_e^2 = \frac{SSE}{n-K-1},\ R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},\ \bar{R}^2 = 1 - \frac{SSE/(n-K-1)}{SST/(n-1)},$$

$n - 1 = (n - K - 1) + (K)$

a. $SSE = 2500,\ s_e^2 = \dfrac{2500}{32-2-1} = 86.207,\ s_e = 9.2848$

b. $SST = SSR + SSE = 7{,}000 + 2{,}500 = 9{,}500$

c. $R^2 = \dfrac{7000}{9500} = 1 - \dfrac{2500}{9500} = .7368,\ \bar{R}^2 = 1 - \dfrac{2500/(29)}{9500/(31)} = .7187$

12.18 From the given Analysis of Variance table,

$$SST = SSR + SSE,\ s_e^2 = \frac{SSE}{n-K-1},\ R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},$$

$$\bar{R}^2 = 1 - \frac{SSE/(n-K-1)}{SST/(n-1)},\ n-1 = (n-K-1) + (K)$$

a. $SSE = 15{,}000,\ s_e^2 = \dfrac{15{,}000}{206-5-1} = 75.0,\ s_e = 8.660$

b. SST = SSR + SSE = 80,000 + 15,000 = 95,000

c. $R^2 = \dfrac{80{,}000}{95{,}000} = 1 - \dfrac{15{,}000}{95{,}000} = .8421,\ \bar{R}^2 = 1 - \dfrac{15{,}000/(200)}{95{,}000/(205)} = .8382$

12.20 a. $R^2 = \dfrac{88.2}{162.1} = .5441$, therefore, 54.41% of the variability in milk consumption can be explained by the variations in weekly income and family size.

b. $\bar{R}^2 = 1 - \dfrac{73.9/(30-3)}{162.1/29} = .5103$

c. $R = \sqrt{.5441} = .7376$. This is the sample correlation between observed and predicted values of milk consumption.

12.22 a.

Regression Analysis: Y profit versus X2 offices

```
The regression equation is
Y profit = 1.55 -0.000120 X2  offices
Predictor          Coef      SE Coef          T        P
Constant         1.5460       0.1048      14.75    0.000
X2  offi    -0.00012033   0.00001434      -8.39    0.000

S = 0.07049      R-Sq = 75.4%      R-Sq(adj) = 74.3%
Analysis of Variance
Source              DF          SS          MS          F        P
Regression           1     0.34973     0.34973      70.38    0.000
Residual Error      23     0.11429     0.00497
Total               24     0.46402
```

b.

Regression Analysis: X1 revenue versus X2 offices

```
The regression equation is
X1 revenue = - 0.078 +0.000543 X2  offices

Predictor          Coef      SE Coef          T        P
Constant        -0.0781       0.2975      -0.26    0.795
X2  offi     0.00054280   0.00004070      13.34    0.000

S = 0.2000       R-Sq = 88.5%      R-Sq(adj) = 88.1%
Analysis of Variance
Source              DF          SS          MS          F        P
Regression           1      7.1166      7.1166     177.84    0.000
Residual Error      23      0.9204      0.0400
Total               24      8.0370
```

c.

Regression Analysis: Y profit versus X1 revenue

```
The regression equation is
Y profit = 1.33 - 0.169 X1 revenue
Predictor         Coef      SE Coef         T        P
Constant        1.3262       0.1386      9.57    0.000
X1 reven      -0.16913      0.03559     -4.75    0.000

S = 0.1009      R-Sq = 49.5%      R-Sq(adj) = 47.4%
Analysis of Variance
Source            DF         SS         MS         F        P
Regression         1    0.22990    0.22990     22.59    0.000
Residual Error    23    0.23412    0.01018
Total             24    0.46402
```

d.

Regression Analysis: X2 offices versus X1 revenue

```
The regression equation is
X2  offices = 957 + 1631 X1 revenue
Predictor         Coef      SE Coef         T        P
Constant         956.9        476.5      2.01    0.057
X1 reven        1631.3        122.3     13.34    0.000

S = 346.8       R-Sq = 88.5%      R-Sq(adj) = 88.1%
Analysis of Variance
Source            DF         SS         MS         F        P
Regression         1   21388013   21388013    177.84    0.000
Residual Error    23    2766147     120267
Total             24   24154159
```

12.24 Given the regression results where the numbers in parentheses are the sample standard error of the coefficient estimates

a. The two-sided 95% confidence intervals for the three regression slope coefficients are given by $b_j \pm t_{n-K-1,\alpha/2} s_{b_j}$

95% CI for $x_1 = 6.8 \pm 2.042\ (3.1)$; .4698 up to 13.1302
95% CI for $x_2 = 6.9 \pm 2.042\ (3.7)$; -.6554 up to 14.4554
95% CI for $x_3 = -7.2 \pm 2.042\ (3.2)$; -13.7344 up to -.6656

b. Testing the hypothesis $H_0 : \beta_j = 0, H_1 : \beta_j > 0$

For x_1: $t = \dfrac{6.8}{3.1} = 2.194$; $t_{30,.05/.01} = 1.697, 2.457$

Therefore, reject H_0 at the 5% level but not at the 1% level

For x_2: $t = \dfrac{6.9}{3.7} = 1.865$; $t_{30,.05/.01} = 1.697, 2.457$

Therefore, reject H_0 at the 5% level but not at the 1% level

For x_3: $t = \dfrac{-7.2}{3.2} = -2.25$; $t_{30,.05/.01} = 1.697, 2.457$

Therefore, do not reject H_0 at either level

12.26 Given the regression results where the numbers in parentheses are the sample standard error of the coefficient estimates

a. The two-sided 95% confidence intervals for the three regression slope coefficients are given by $b_j \pm t_{n-K-1,\alpha/2} s_{b_j}$

95% CI for x_1 = 17.8 ± 2.030 (7.1); 3.387 up to 32.213
95% CI for x_2 = 26.9 ± 2.030 (13.7); -.911 up to 54.711
95% CI for x_3 = -9.2 ± 2.030 (3.8); -16.914 up to -1.486

b. Testing the hypothesis $H_0 : \beta_j = 0, H_1 : \beta_j > 0$

For x_1: $t = \frac{17.8}{7.1} = 2.507$; $t_{35,.05/.01} = 1.690, 2.438$

Therefore, reject H_0 at the 5% level and at the 1% level

For x_2: $t = \frac{26.9}{13.7} = 1.964$; $t_{35,.05/.01} = 1.690, 2.438$

Therefore, reject H_0 at the 5% level but not at the 1% level

For x_3: $t = \frac{-9.2}{3.8} = -2.421$; $t_{35,.05/.01} = 1.690, 2.438$

Therefore, do not reject H_0 at either level

12.28 a. $H_0 : \beta_1 = 0; H_1 : \beta_1 > 0$

$$t = \frac{.052}{.023} = 2.26$$

$t_{27,.025/.01} = 2.052, 2.473$

Therefore, reject H_0 at the 2.5% level but not at the 1% level

b. $t_{27,.05/.025/.005} = 1.703, 2.052, 2.771$

90% CI: 1.14 ± 1.703(.35); .5439 up to 1.7361
95% CI: 1.14 ± 2.052(.35); .4218 up to 1.8582
99% CI: 1.14 ± 2.771(.35); .1701 up to 2.1099

12.30 a. $H_0 : \beta_3 = 0, H_1 : \beta_3 \neq 0$

$$t = \frac{-.000191}{.000446} = -.428$$

$t_{16,.10} = -1.337$

Therefore, do not reject H_0 at the 20% level

b. $H_0 : \beta_1 = \beta_2 = \beta_3 = 0, H_1 :$ At least one $\beta_i \neq 0, (i = 1,2,3)$

$$F = \frac{(n-K-1)}{K}\left[\frac{R^2}{1-R^2}\right] = \frac{16}{3} \times \frac{.71}{1-.71} = 13.057, \quad F_{3,16,.01} = 5.29$$

Therefore, reject H_0 at the 1% level

12.32 a. All else being equal, an extra \$1 in mean per capita personal income leads to an expected extra \$.4 of net revenue per capita from the lottery

b. $b_2 = .8772, s_{b_2} = .3107, n = 29, t_{24,.025} = 2.064$

95% CI: $.8772 \pm 2.064(.3107)$; .2359 up to 1.5185

Therefore the 95% confidence interval for the expected increase in the dollars of net revenue per capita per year generated by the lottery resulting from a one-unit increase in number of hotel, motel, inn, and resort rooms per thousand persons in the country, if the other variables do not change, runs from .2359 up to 1.5185. Also, since the 95% confidence interval does not include 0, we conclude that the coefficient on x_2 in the population regression is statistically significant.

c. $H_0 : \beta_3 = 0, H_1 : \beta_3 < 0$

$$t = \frac{-365.01}{263.88} = -1.383$$

$t_{24,.10/.05}$ = -1.318, -1.711

Therefore, reject H_0 at the 10% level but not at the 5% level

12.34 a. $n = 39, b_5 = .0495, s_{b_1} = .01172, t_{31,.005} = 2.744$

99% CI: $.0495 \pm 2.744(.01172)$; .0173 up to .0817

Therefore, the 99% confidence interval for the expected increase in the number of full-time firefighters resulting from a 1 unit increase in the amount of intergovernmental grants per capita, if the other variables do not change, runs from .0173 up to .0817.

Also, since the 99% confidence interval for β_5 does not include 0, we conclude that this variable is statistically significant.

b. $H_0 : \beta_4 = 0, H_1 : \beta_4 \neq 0$

$$t = \frac{.48122}{.77954} = .617$$

$t_{31,.10}$ = 1.309

Therefore, do not reject H_0 at the 20% level. And conclude that the population density is not statistically significant.

c. $H_0 : \beta_7 = 0, H_1 : \beta_7 \neq 0$

$$t = \frac{.00645}{.00306} = 2.108$$

$t_{31,.025/.01}$ = 2.040, 2.453

Therefore, reject H_0 at the 5% level but not at the 2% level. And conclude that the percentage of the population that is male and between 12 and 21 years of age is statistically significant at the 5% level but not at the 2% level.

12.36 a. SST = 3.881, SSR = 3.549, SSE = .332

$H_0: \beta_1 = \beta_2 = \beta_3 = 0, H_1:$ At least one $\beta_i \neq 0, (i = 1,2,3)$

$$F = \frac{3.549/3}{.332/23} = 81.955$$

$F_{3,23,.01} = 4.76$

Therefore, reject H_0 at the 1% level

b. Analysis of Variance table:

Sources of variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Ratio
Regression	3.549	3	1.183	81.955
Error	.332	23	.014435	
Total	3.881	26		

12.38 a. SST = 162.1, SSR =88.2, SSE = 73.9

$H_0: \beta_1 = \beta_2 = 0, H_1:$ At least one $\beta_i \neq 0, (i = 1,2)$

$$F = \frac{88.2/2}{73.9/27} = 16.113, \quad F_{2,27,.01} = 5.49$$

Therefore, reject H_0 at the 1% level

b. Analysis of Variance table:

Sources of variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Ratio
Regression	88.2	2	44.10	16.113
Error	73.9	27	2.737	
Total	162.1	29		

12.40 $$\frac{(SSE^* - SSE)/K_1}{SSE/(n-K-1)} = \frac{n-K-1}{K_1}\left[\frac{(SSE^* - SSE)/SST}{SSE/SST}\right]$$

$$= \frac{n-K-1}{K_1}\left[\frac{\left(1-R^{*2}\right)-(1-R^2)}{1-R^2}\right] = \frac{n-K-1}{K_1}\left[\frac{R^2 - R^{*2}}{1-R^2}\right]$$

12.42 a. $$\bar{R}^2 = 1 - \frac{SSE/(n-K-1)}{SST/(n-1)} = 1 - \left[\frac{n-1}{n-K-1}(1-R^2)\right]$$

$$= 1 - \frac{n-1}{n-K-1} + \frac{n-1}{n-K-1}R^2$$

$$= \frac{n-1}{n-K-1}R^2 - \frac{K}{n-K-1} = \frac{(n-1)R^2 - K}{n-K-1}$$

b. Since $\bar{R}^2 = \dfrac{(n-1)R^2 - K}{n-K-1}$, then $R^2 = \dfrac{(n-K-1)\bar{R}^2 + K}{n-1}$

c. $$\frac{SSR/K}{SSE/(n-K-1)} = \frac{n-K-1}{K}\left[\frac{SSR/SST}{SSE/SST}\right] = \frac{n-K-1}{K}\left[\frac{R^2}{1-R^2}\right]$$

$$= \frac{n-K-1}{K}\left[\frac{[(n-K-1)\bar{R}^2+K]/(n-1)}{[n-1-(n-K-1)\bar{R}^2-K]/(n-1)}\right]$$

$$= \frac{n-K-1}{K}\left[\frac{(n-K-1)\bar{R}^2+K}{(n-K-1)(1-\bar{R}^2)}\right]$$

$$= \frac{n-K-1}{K}\left[\frac{\bar{R}^2+\dfrac{K}{(n-K-1)}}{(1-\bar{R}^2)}\right]$$

$$= \frac{n-K-1}{K}\left[\frac{\bar{R}^2+A}{(1-\bar{R}^2)}\right] \text{ where } A = \frac{K}{n-K-1}$$

12.44 $\hat{Y} = 7.35 + .653(20) - 1.345(10) + .613(6) = 10.638$ pounds

12.46 $\hat{Y} = 2.0 + .661(1) + .065(7) - .018(50) = 2.216$ million worker hours

12.48 a. mpg as a function of horsepower and weight

Regression Analysis: milpgal versus horspwr, weight

```
The regression equation is
milpgal = 55.8 - 0.105 horspwr - 0.00661 weight
150 cases used 5 cases contain missing values
Predictor        Coef     SE Coef        T       P     VIF
Constant       55.769       1.448    38.51   0.000
horspwr      -0.10489     0.02233    -4.70   0.000   2.918
weight     -0.0066143   0.0009015    -7.34   0.000   2.918

S = 3.901       R-Sq = 72.3%     R-Sq(adj) = 72.0%

Analysis of Variance
Source            DF         SS        MS        F        P
Regression         2     5850.0    2925.0   192.23    0.000
Residual Error   147     2236.8      15.2
Total            149     8086.8

Predicted Values for New Observations

New
Obs     Fit   SE Fit        95% CI              95% PI
  1  21.242    0.975  (19.316, 23.168)  (13.296, 29.188)X

X denotes a point that is an outlier in the predictors.

       Values of Predictors for New Observations

       New
       Obs  horspwr  weight
         1      140    3000
```

b. Adding the number of cylinders

Regression Analysis: milpgal versus horspwr, weight, cylinder

```
The regression equation is
milpgal = 55.9 - 0.117 horspwr - 0.00758 weight + 0.726 cylinder
150 cases used 5 cases contain missing values

Predictor        Coef     SE Coef        T       P      VIF
Constant       55.925       1.443    38.77   0.000
horspwr      -0.11744     0.02344    -5.01   0.000    3.254
weight      -0.007576    0.001066    -7.10   0.000    4.131
cylinder       0.7260      0.4362     1.66   0.098    3.586

S = 3.878          R-Sq = 72.9%      R-Sq(adj) = 72.3%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         3      5891.6      1963.9    130.62    0.000
Residual Error   146      2195.1        15.0
Total            149      8086.8

Predicted Values for New Observations

New
Obs      Fit   SE Fit           95% CI               95% PI
  1   21.112    0.972  (19.190, 23.033)   (13.211, 29.012)

Values of Predictors for New Observations

New
Obs   horspwr   weight   cylinder
  1       140     3000       6.00
```

12.50 Computing values of y_i when x_i = 1, 2, 4, 6, 8, 10

Xi	1	2	4	6	8	10
$y_i = 4x^{1.8}$	4	13.9288	48.5029	100.6311	168.8970	252.3829
$y_i = 1 + 2x_i + 2x_i^2$	5	13	41	85	145	221

12.52 Computing values of y_i when x_i = 1, 2, 4, 6, 8, 10

Xi	1	2	4	6	8	10
$y_i = 3x^{1.2}$	3	6.8922	15.8341	25.7574	36.3772	47.5468
$y_i = 1 + 5x_i - 1.5x_i^2$	4.5	5	-3	-23	-55	-99

12.54 To estimate the function with linear least squares, solve the equation $\beta_1 + \beta_2 = 2$ for β_2. Since $\beta_2 = 2 - \beta_1$, plug into the equation and algebraically manipulate:

$$Y = \beta_o + \beta_1 X_1 + (2 - \beta_1)X_1^2 + \beta_3 X_2$$
$$Y = \beta_o + \beta_1 X_1 + 2X_1^2 - \beta_1 X_1^2 + \beta_3 X_2$$
$$Y = \beta_o + \beta_1[X_1 - X_1^2] + 2X_1^2 + \beta_3 X_2$$
$$Y - 2X_1^2 = \beta_o + \beta_1[X_1 - X_1^2] + \beta_3 X_2$$

Conduct the variable transformations and estimate the model using least squares.

12.56 a. A 1% increase in median income leads to an expected .68% increase in store size.

b. $H_0: \beta_1 = 0, H_1: \beta_1 > 0$, $t = \frac{.68}{.077} = 8.831$. Therefore, reject H_0 at the 1% level

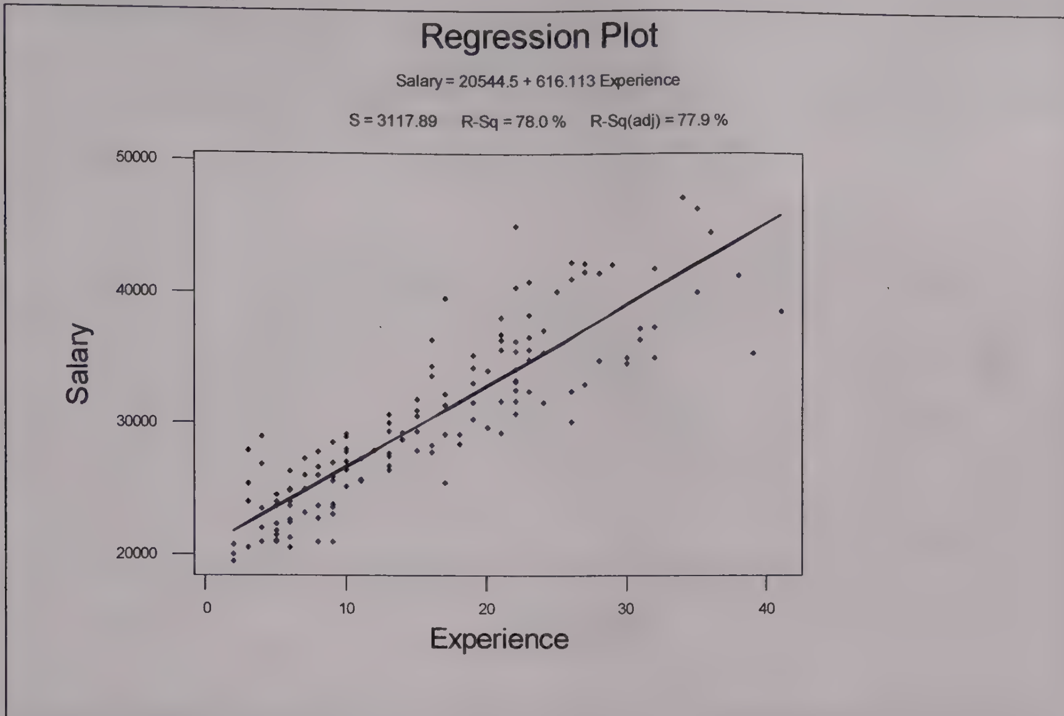
12.58 Estimating a Cobb-Douglas production function with three independent variables:
$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} \varepsilon$ where X_1 = capital, X_2 = labor, and X_3 = basic research
Taking the log of both sides of the equation yields:
$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \beta_3 \log(X_3) + \varepsilon$
Using this form, now regression the log of Y on the logs of the three independent variables and obtain the estimated regression slope coefficients.

Now, Substituting in the restrictions on the coefficients:
$\beta_1 + \beta_2 + \beta_3 = 1, \beta_3 = 1 - \beta_1 - \beta_2$
$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) + \beta_2 \log(X_2) + [1 - \beta_1 - \beta_2] \log(X_3) + \varepsilon$
$\log(Y) = \log(\beta_0) + \beta_1 \log(X_1) + \beta_2 \log(X_2) + \log(X_3) - \beta_1 \log(X_3) - \beta_2 \log(X_3) + \varepsilon$
$\log(Y) - \log(X_3) = \log(\beta_0) + \beta_1 [\log(X_1) - \log(X_3)] + \beta_2 [\log(X_2) - \log(X_3)] + \varepsilon$

$$\log\left[\frac{Y}{X_3}\right] = \log(\beta_0) + \beta_1 \log\left[\frac{X_1}{X_3}\right] + \beta_2 \log\left[\frac{X_2}{X_3}\right] + \varepsilon$$

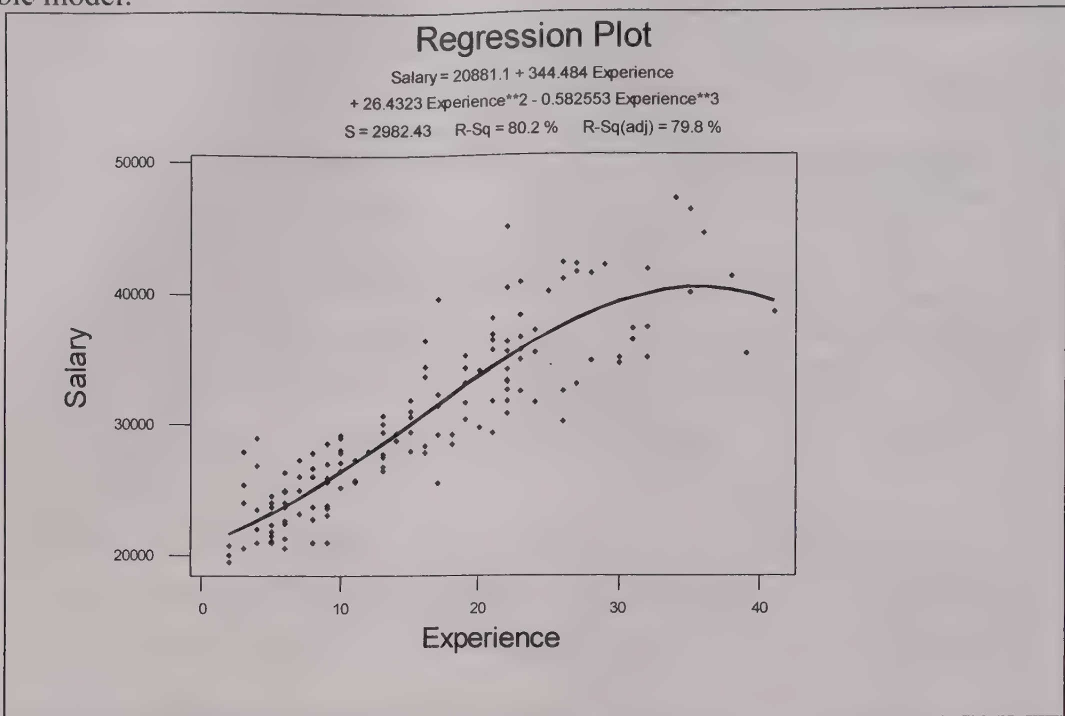
Thus, we see that the β_1 coefficient is obtained by regressing $\log(Y / X_3)$ on $\log(X_1 / X_3)$ and the coefficient β_2 by regressing $\log(Y / X_3)$ on $\log(X_2 / X_3)$. The coefficient β_3 can be found by subtracting β_1 and β_2 from 1.0.

12.60 Linear model:



Quadratic model:

Cubic model:



All three of the models appear to fit the data well. The cubic model appears to fit the data the best as the standard error of the estimate is lowest. In addition, explanatory power is marginally higher for the cubic model than the other models.

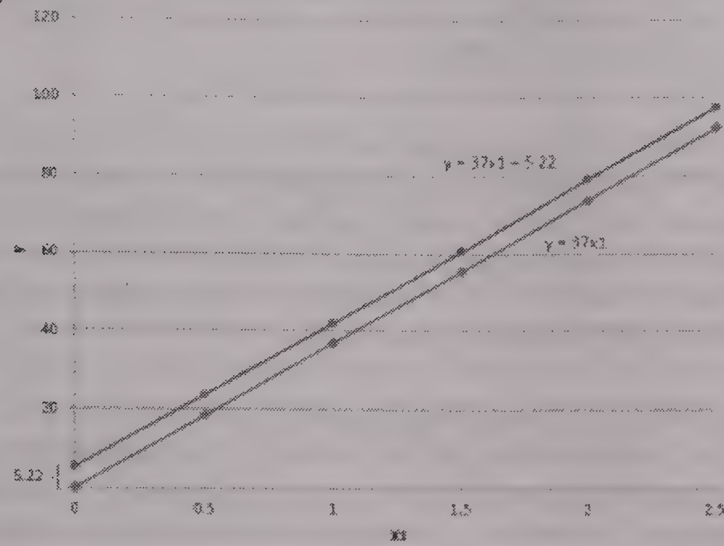
12.62 The model constants when the dummy variable equals 1 are

a. $\hat{y} = 7 + 8x_1$, $b_0 = 7$

b. $\hat{y} = 12 + 6x_1$, $b_0 = 12$

c. $\hat{y} = 7 + 12x_1$, $b_0 = 7$

12.64 The interpretation of the dummy variable is that we can conclude that for a given difference between the spot price in the current year and spot price in the previous year, the difference between the OPEC price in the current year and OPEC price in the previous years is \$5.22 higher in 1974 during the oil embargo than in other years.



Dummy variable x2 is equal to 1 for the year 1974 and 0 otherwise to represent the specific effect of the oil embargo of that year. The graph indicates that the OPEC price in the previous years is \$5.22 higher in 1974 during the oil embargo than in other years keeping other factors constant.

12.66 a. All else being equal, the price-earnings ratio is higher by 1.23 for a regional company than a national company.

b. $H_0: \beta_2 = 0, H_1: \beta_2 \neq 0$, $t = \dfrac{1.23}{.496} = 2.48$, $t_{29,.01/.005} = 2.462,\ 2.756$

Therefore, reject H_0 at the 2% level but not at the 1% level

c. $H_0: \beta_1 = \beta_2 = 0, H_1:$ At least one $\beta_i \neq 0, (i = 1, 2)$

$$F = \frac{n-K-1}{K}\frac{R^2}{1-R^2} = \frac{29}{2} \times \frac{.37}{1-.37} = 8.516,\ F_{2,29,.05} = 3.33$$

Therefore, reject H_0 at the 5% level and conclude that at least one coefficient is not equal to 0. Hence at least one predictor variable is a significant predictor of the price-earnings ratio.

12.68 a. All else being equal, the annual salary of the attorney general is \$5,793 higher if justices of the state supreme court can be removed from office by the governor, judicial review board, or majority vote of the supreme court.

b. All else being equal, the annual salary of the attorney general of the state is \$3,100 lower if the supreme court justices are elected on partisan ballots.

c. $H_0: \beta_5 = 0, H_1: \beta_5 > 0$, $t = \dfrac{5793}{2897} = 1.9996$, $t_{43,.05} = 1.68$

Therefore, reject H_0 at the 5% level

d. $H_0: \beta_6 = 0, H_1: \beta_6 < 0$, $t = \dfrac{-3100}{1761} = -1.76$, $t_{43,.05} = -1.68$

Therefore, reject H_0 at the 5% level

b. $t_{43,.025} = 2.017$

95% CI: $547 \pm 2.017(124.3)$; 296.29 up to 797.71
Therefore, the 95% confidence interval for the expected increase (in thousands of dollars) in annual salary of the attorney general resulting from a $1000 increase of average annual salary of lawyers, if the other variables do not change, runs from 296.29 to 797.71. Also, since the interval does not include zero, the coefficient β_1 is statistically significant.

12.70 34.4% of the variation in a test on understanding business statistics can be explained by which course was taken, the student's GPA, the teacher that taught the course, the gender of the student, the pre-test score, the number of credit hours completed, and the age of the student. The regression model has significant explanatory power:

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0, H_1:$ At least one $\beta_i \neq 0, (i = 1,2,3,4,5,6,7)$

$$F = \frac{n-K-1}{K}\frac{R^2}{1-R^2} = \frac{342}{7} \times \frac{.344}{1-.344} = 25.62$$

Therefore, reject H_0 at the 1% level and conclude that at least one coefficient is not equal to 0. Hence at least one predictor variable is a significant predictor of the business statistics test.

12.72 a. Begin the analysis with the correlation matrix – identify important independent variables as well as correlations between the independent variables

Correlations: Salary, Experience, yearsenior, Gender_1F

```
          Salary Experien yearseni
Experien   0.883
           0.000

yearseni   0.777    0.674
           0.000    0.000

Gender_1 -0.429   -0.378   -0.292
           0.000    0.000    0.000
```

Regression Analysis: Salary versus Experience, yearsenior, Gender_1F

```
The regression equation is
Salary = 22644 + 437 Experience + 415 yearsenior - 1443 Gender_1F
Predictor        Coef     SE Coef        T        P       VIF
Constant      22644.1       521.8    43.40    0.000
Experien       437.10       31.41    13.92    0.000       2.0
yearseni       414.71       55.31     7.50    0.000       1.8
Gender_1      -1443.2       519.8    -2.78    0.006       1.2

S = 2603         R-Sq = 84.9%      R-Sq(adj) = 84.6%

Analysis of Variance
Source            DF          SS          MS         F       P
Regression         3  5559163505  1853054502    273.54   0.000
Residual Error   146   989063178     6774405
Total            149  6548226683
```

84.9% of the variation in annual salary (in dollars) can be explained by the variation in the years of experience, the years of seniority, and the gender of the employee. All of the variables are significant at the .01 level of significance (p-values of .000, .000, and .006 respectively). The F-test of the significance of the overall model shows that we reject H_0 that all of the slope coefficients are jointly equal to zero in favor of H_1 that at least one slope coefficient is not equal to zero. The F-test yielded a p-value of .000.

b. $H_0 : \beta_3 = 0, H_1 : \beta_3 < 0$

$$t = \frac{-1443.2}{519.8} = -2.78, \quad t_{146,.01} = -2.326$$

Therefore, reject H_0 at the 1% level. And conclude that the annual salaries for females are statistically significantly lower than they are for males.

c. Adding an interaction term and testing for the significance of the slope coefficient on the interaction term.

Adding the interaction term of Salary and Gender_1F, the regression equation is obtained as:

Salary = 23336 + 388 Experience + 467 yearsenior - 12468 Gender_1F + 0.425 int. term

$H_0 : \beta_4 = 0, H_1 : \beta_4 < 0$

$$t = \frac{0.4246}{0.1127} = 3.77, \quad t_{145,.10,.05} = -1.282, -1.645$$

Therefore, do not reject H_0 at either level. And conclude that the rate of salary increase for females is not statistically significantly lower than they are for males at either level.

12.74 n = 34 and four independent variables. r = .23.

Correlation between the independent variable and the dependent variable is not necessarily evidence of a small Student's *t* statistic. A high correlation among the *independent* variables could result in a very small Student's *t* statistic as the correlation creates a high variance.

12.76 n = 49 with two independent variables. One of the independent variables has a correlation of .56 with the dependent variable.
Correlation between the independent variable and the dependent variable is not necessarily evidence of a small Student's *t* statistic. A high correlation among the *independent* variables could result in a very small Student's *t* statistic as the correlation creates a high variance.

12.77. Reports can be written by following the extended Case Study on the data file Cotton – see Section 12.9

12.78

31.2% of the variation in the overall opinion of residence hall can be explained by the variation in the satisfaction with roommates, with floor, with hall, and with resident advisor. The F-test of the significance of the overall model shows that we reject H_0 that all of the slope coefficients are jointly equal to zero in favor of H_1 that at least one slope coefficient is not equal to zero. The F-test yielded a p-value of .000.
All of the variables are significant at the .1 level of significance but not at the .05 level of significance except the variable – satisfaction with resident advisor. The variable – satisfaction with resident advisor is significant at all common levels of alpha.

12.80 Begin the analysis by selecting all the variables. Stepwise, delete the predictor variables which are not significant. The final regression model includes the predictor variables per capita disposable income and percent of population in urban areas

Regression Analysis: Fat Rate versus Percap Disp, P Urban

```
The regression equation is
Fat Rate = 2.80 - 0.000033 Percap Disp - 0.00631 P Urban
Predictor          Coef      SE Coef      T      P     VIF
Constant         2.7998       0.2971   9.42  0.000
Percap Disp -0.00003325   0.00001225  -2.72  0.009   1.456
P Urban       -0.006312     0.003588  -1.76  0.085   1.456

S = 0.321204   R-Sq = 32.4%   R-Sq(adj) = 29.6%


Analysis of Variance

Source          DF      SS      MS      F      P
Regression       2  2.3760  1.1880  11.51  0.000
Residual Error  48  4.9523  0.1032
Total           50  7.3282
```

12.82

Regression Analysis: y_manufgrowt versus x1_aggrowth, x2_exportgro, ...

```
The regression equation is
y_manufgrowth = 2.15 + 0.493 x1_aggrowth + 0.270 x2_exportgrowth
          - 0.117 x3_inflation
Predictor        Coef     SE Coef        T       P      VIF
Constant       2.1505      0.9695     2.22   0.032
x1_aggro       0.4934      0.2020     2.44   0.019      1.0
x2_expor      0.26991     0.06494     4.16   0.000      1.0
x3_infla     -0.11709     0.05204    -2.25   0.030      1.0

S = 3.624       R-Sq = 39.3%      R-Sq(adj) = 35.1%
Analysis of Variance
Source            DF        SS        MS       F       P
Regression         3    373.98    124.66    9.49   0.000
Residual Error    44    577.97     13.14
Total             47    951.95

Source       DF     Seq SS
x1_aggro      1      80.47
x2_expor      1     227.02
x3_infla      1      66.50
```

39.3% of the variation in the manufacturing growth can be explained by the variation in agricultural growth, exports growth, and rate of inflation.
The F-test of the significance of the overall model shows that we reject H_0 that all of the slope coefficients are jointly equal to zero in favor of H_1 that at least one slope coefficient is not equal to zero. The F-test yielded a p-value of .000.
All the independent variables are significant at the .1 and .05 levels of significance. Exports growth is significant in explaining the manufacturing growth at all common levels of alpha.
All else being equal, the agricultural growth and the exports growth variables have the expected sign. This is because, as the percentage of agricultural growth and the exports growth increases, the percentage of manufacturing growth also increases. The rate of inflation variable is negative which indicates that higher the inflation rate, lower is the manufacturing growth.

12.84 The analysis of variance table identifies how the total variability of the dependent variable (SST) is split up between the portion of variability that is explained by the regression model (SSR) and the part that is unexplained (SSE). The Coefficient of Determination (R^2) is derived as the ratio of SSR to SST. The analysis of variance table also computes the F statistic for the test of the significance of the overall regression – whether all of the slope coefficients are jointly equal to zero. The associated p-value is also generally reported in this table.

12.86 If one model contains more explanatory variables, then SST remains the same for both models but SSR will be higher for the model with more explanatory variables. Since SST = $SSR_1 + SSE_1$ which is equivalent to $SSR_2 + SSE_2$ and given that $SSR_2 > SSR_1$, then $SSE_1 > SSE_2$. Hence, the coefficient of determination will be higher with a greater number of explanatory variables and the coefficient of determination must be interpreted in conjunction with whether or not the regression slope coefficients on the explanatory variables are significantly different from zero.

12.88

$$\sum e_i = \sum (y_i - \hat{y}_i)$$
$$\sum e_i = \sum (y_i - a - b_1 x_{1i} - b_2 x_{2i})$$
$$\sum e_i = \sum (y_i - \bar{y} + b_1 \bar{x}_{1i} + b_2 \bar{x}_{2i} - b_1 x_{1i} - b_2 x_{2i})$$
$$\sum e_i = n\bar{y} - n\bar{y} + nb_1\bar{x}_1 + nb_2\bar{x}_2 - nb_1\bar{x}_1 - nb_2\bar{x}_2$$
$$\sum e_i = 0$$

12.90a. All else being equal, an increase of one question results in a decrease of 1.834 in expected percentage of responses received. All else being equal, an increase in one word in length of the questionnaire results in a decrease of .016 in expected percentage of responses received.

b. 63.7% of the variability in the percentage of responses received can be explained by the variability in the number of questions asked and the number of words.

c. $H_0: \beta_1 = \beta_2 = 0, H_1:$ At least one $\beta_i \neq 0, (i = 1,2)$

$$F = \frac{n-K-1}{K}\frac{R^2}{1-R^2} = \frac{27}{2} \times \frac{.637}{1-.637} = 23.69$$

$F_{2,27,.01} = 5.49$. Therefore, reject H_0 at the 1% level

d. $t_{27,.005} = 2.771$, 99% CI: $-1.8345 \pm 2.771(.6349)$; -3.5938 up to -.0752

Therefore, the 99% confidence interval for the expected increase in the percentage of responses received resulting from an increase of one question, if the number of words do not change, runs from -3.5938 to -.0752. Also, since the interval does not include zero, the coefficient of number of questions asked is statistically significant.

e. $t = -1.78$, $t_{27,.05/.025} = -1.703, -2.052$.

Therefore, reject H_0 at the 5% level but not at the 2.5% level. And we conclude that the length of questionnaire in number of words is statistically significant in explaining the percentage of responses received at the 5% level but not at the 2.5% level.

12.92

Regression Analysis: y_rating versus x1_expgrade, x2_Numstudents

```
The regression equation is
y_rating = - 0.200 + 1.41 x1_expgrade - 0.0158 x2_Numstudents
Predictor         Coef     SE Coef        T       P       VIF
Constant       -0.2001      0.6968    -0.29   0.777
x1_expgr        1.4117      0.1780     7.93   0.000      1.5
x2_Numst     -0.015791    0.003783    -4.17   0.001      1.5

S = 0.1866      R-Sq = 91.5%      R-Sq(adj) = 90.5%
Analysis of Variance
Source            DF        SS         MS        F        P
Regression         2    6.3375     3.1687    90.99    0.000
Residual Error    17    0.5920     0.0348
Total             19    6.9295
```

91.5% of the variation in the rating can be explained by the linear dependence on the expected grade and the number of students in the class.
The F-test of the significance of the overall model shows that we reject H_0 that all of the slope coefficients are jointly equal to zero in favor of H_1 that at least one slope coefficient is not equal to zero. The F-test yielded a p-value of .000.
The coefficients of expected grade and number of students are significant at common levels of alpha.
All else being equal, a 1 point increase in the expected grade is associated with a 1.41 increase in the rating. The number of students variable is negative which indicates that higher the student count, the lower is the rating.

12.94 a. All else being equal, each extra point in a student's expected score leads to an expected increase of .469 in the actual score.

b. $t_{103,.025} = 1.96$, therefore, the 95% CI: $3.369 \pm 1.96(.456) = 2.4752$ up to 4.2628

Therefore, the 95% confidence interval for the expected increase in a student's actual score resulting from an increase of 1 hour time spent on the course, if the other variables do not change, runs from 2.4752 up to 4.2628. Also, since the interval does not include zero, the coefficient β_2 is statistically significant.

c. $H_0 : \beta_3 = 0, H_1 : \beta_3 \neq 0$, $t = \frac{3.054}{1.457} = 2.096$, $t_{103,.025,.005} = 1.96, 2.58$

Therefore, reject H_0 at the 5% level but not at the 1% level. And we conclude that a student's grade point average is statistically significant in explaining a student's actual score in the examination at the 5% level but not at the 1% level.

d. 68.6% of the variation in the exam scores is explained by their linear dependence on a student's expected score, hours per week spent working on the course, and a student's grade point average.

e. $H_0: \beta_1 = \beta_2 = \beta_3 = 0, H_1:$ At least one $\beta_i \neq 0, (i = 1,2,3)$

$$F = \frac{n-K-1}{K}\frac{R^2}{1-R^2} = \frac{103}{3} \times \frac{.686}{1-.686} = 75.008, F_{3,103,.01} = 3.95$$

Reject H_0 at any common levels of alpha

f. $R = \sqrt{.686} = .82825$. This is the sample correlation between the observed and the predicted values of a student's actual scores in the examination.

g. $\hat{Y} = 2.178 + .469(80) + 3.369(8) + 3.054(3) = 75.812$

12.96 a. $t_{2669,.05} = 1.645$, therefore, the 90% CI: $480.04 \pm 1.645(224.9) = 110.0795$ up to 850.0005

Therefore, the 90% confidence interval for the expected increase in the minutes played in the season resulting from a 1 unit increase of steals per minute, if the other variables do not change, runs from 110.0795 up to 850.0005. Also, since the interval does not include zero, the coefficient β_6 is statistically significant.

b. $t_{2669,.005} = 2.576$, therefore, the 99% CI: $1350.3 \pm 2.576(212.3) = 803.4152$ up to 1897.1848

Therefore, the 99% confidence interval for the expected increase in the minutes played in the season resulting from a 1 unit increase of blocked shots per minute, if the other variables do not change, runs from 803.4152 up to 1897.1848. Also, since the interval does not include zero, the coefficient β_7 is statistically significant.

c. $H_0: \beta_8 = 0, H_1: \beta_8 < 0$, $t = \frac{-891.67}{180.87} = -4.9299$

$t_{2669,.005} = -2.576$, therefore, reject H_0 at the .5% level. And we conclude that the turnovers per minute is statistically significant in explaining the minutes played in the season at all common levels of alpha.

d. $H_0: \beta_9 = 0, H_1: \beta_9 > 0$, $t = \frac{722.95}{110.98} = 6.5142$

$t_{2669,.005} = 2.576$, therefore, reject H_0 at the .5% level. And we conclude that the assists per minute is statistically significant in explaining the minutes played in the season at all common levels of alpha.

e. 52.39% of the variability in minutes played in the season can be explained by the variability in all 9 variables.

f. $R = \sqrt{.5239} = .7238$. This is the sample correlation between the observed and the predicted values of the minutes played in season.

12.98 57.3% of the variation in the female amateur golfers winnings per tournament can be explained by variations in average length of drive, percentage times drive ends, percentage times green reached, percentage times par saved after hitting into sand trap, average number of putts taken on greens reached, average number of putts taken on greens not reached in regulation, and the number of years the golfer has played.

The independent variables that are significant (10% level in fact all common levels of alpha) include average length of drive (X1) and average number of putts taken on greens reached in regulation (X5). The independent variable, average number of putts taken on greens not reached in regulation (X6) is significant at 10% level but not at the 5% level. The other independent variables are not significant at common levels of alpha.

The F-test of the significance of the overall model shows that we reject H_0 in favor of H_1 that at least one slope coefficient is not equal to zero. The F-test yielded a p-value of .000.

The sample correlation between the observed and the predicted values of the female amateur golfers winnings per tournament is .7572.

A report can be written by following the **Cotton** Case Study and testing the significance of the model. See section 12.9

12.100

Correlations: hseval, Comper, Homper, Indper, sizehse, incom72

	hseval	Comper	Homper	Indper	sizehse
Comper	-0.335				
	0.001				
Homper	0.145	-0.499			
	0.171	0.000			
Indper	-0.086	-0.140	-0.564		
	0.419	0.188	0.000		
sizehse	0.542	-0.278	0.274	-0.245	
	0.000	0.008	0.009	0.020	
incom72	0.426	-0.198	-0.083	0.244	0.393
	0.000	0.062	0.438	0.020	0.000

The correlation matrix indicates that the size of the house, income, and percent homeowners have a positive relationship with house value. There is a negative relationship between the percent industrial and percent commercial and the house value.

Regression Analysis: hseval versus Comper, Homper, ...

```
The regression equation is
hseval = - 19.0 - 26.4 Comper - 12.1 Homper - 15.5 Indper + 7.22
sizehse
         + 0.00408 incom72
Predictor         Coef      SE Coef          T         P        VIF
Constant        -19.02        13.20      -1.44     0.153
Comper         -26.393        9.890      -2.67     0.009        2.2
Homper         -12.123        7.508      -1.61     0.110        3.0
Indper         -15.531        8.630      -1.80     0.075        2.6
sizehse          7.219        2.138       3.38     0.001        1.5
incom72       0.004081     0.001555       2.62     0.010        1.4

S = 3.949        R-Sq = 40.1%       R-Sq(adj) = 36.5%
Analysis of Variance
Source            DF          SS          MS         F          P
Regression         5      876.80      175.36     11.25      0.000
Residual Error    84     1309.83       15.59
Total             89     2186.63
```

All variables are conditionally significant with the exception of Indper and Homper. Since Homper has the smaller t-statistic, it is removed:

Regression Analysis: hseval versus Comper, Indper, sizehse, incom72

```
The regression equation is
hseval = - 30.9 - 15.2 Comper - 5.73 Indper + 7.44 sizehse +
0.00418 incom72
Predictor         Coef      SE Coef          T         P        VIF
Constant        -30.88        11.07      -2.79     0.007
Comper         -15.211        7.126      -2.13     0.036        1.1
Indper          -5.735        6.194      -0.93     0.357        1.3
sizehse          7.439        2.154       3.45     0.001        1.5
incom72       0.004175     0.001569       2.66     0.009        1.4

S = 3.986        R-Sq = 38.2%       R-Sq(adj) = 35.3%
Analysis of Variance
Source            DF          SS          MS         F          P
Regression         4      836.15      209.04     13.16      0.000
Residual Error    85     1350.48       15.89
Total             89     2186.63
```

Indper is not significant and is removed:

Regression Analysis: hseval versus Comper, sizehse, incom72

```
The regression equation is
hseval = - 34.2 - 13.9 Comper + 8.27 sizehse + 0.00364 incom72
Predictor         Coef      SE Coef          T         P        VIF
Constant        -34.24        10.44      -3.28     0.002
Comper         -13.881        6.974      -1.99     0.050        1.1
sizehse          8.270        1.957       4.23     0.000        1.2
incom72       0.003636     0.001456       2.50     0.014        1.2

S = 3.983        R-Sq = 37.6%       R-Sq(adj) = 35.4%

Analysis of Variance
Source            DF          SS          MS         F          P
Regression         3      822.53      274.18     17.29      0.000
Residual Error    86     1364.10       15.86
Total             89     2186.63
```

This becomes the final regression model. The three independent variables are conditionally significant in explaining the house value at .05 level and hence, the final regression model, with conditional t-statistics in parentheses under the coefficients, is:

$$\hat{Y} = -34.2 - \underset{(6.97)}{13.9}(Comper) + \underset{(1.96)}{8.27}(sizehse) + \underset{(0.00146)}{0.00364}(incom72)$$

S = 3.983 $R^2 = .376$ n = 90

The selection of a community with the objective of having larger house values would include communities where the percent of commercial property is low, the median rooms per residence is high and the per capita income is high.

12.102 a. Correlation matrix:

Correlations: deaths, Prurpop, Ruspeed, Prsurf

	deaths	Prurpop	Ruspeed
Prurpop	0.594		
	0.000		
Ruspeed	0.305	0.224	
	0.033	0.121	
Prsurf	-0.556	-0.207	-0.232
	0.000	0.153	0.109

Descriptive Statistics: deaths, Prurpop, Prsurf, Ruspeed

Variable	N	Mean	Median	TrMean	StDev	SE Mean
deaths	49	0.1746	0.1780	0.1675	0.0802	0.0115
Prurpop	49	0.4110	0.3689	0.5992	0.2591	0.0370
Prsurf	49	0.7980	0.8630	0.8117	0.1928	0.0275
Ruspeed	49	58.186	58.400	58.222	1.683	0.240

Variable	Minimum	Maximum	Q1	Q3
deaths	0.0569	0.5505	0.1240	0.2050
Prurpop	0.0311	1.0000	0.1887	0.5915
Prsurf	0.2721	1.0000	0.6563	0.9485
Ruspeed	53.500	62.200	57.050	59.150

The proportion of urban population and rural roads that are surfaced are positively related to crash deaths. Average rural speed is positively related, but the relationship is not as strong as the proportion of urban population and surfaced roads. The simple correlation coefficients among the independent variables are relatively low and hence multicollinearity should not be dominant in this model. Note the relatively narrow range for average rural speed. This would indicate that there is not much variability in this independent variable.

b. Multiple regression

Regression Analysis: deaths versus Prurpop, Prsurf, Ruspeed

```
The regression equation is
deaths = -0.0086 - 0.149 Prurpop - 0.181 Prsurf + 0.00457 Ruspeed

Predictor        Coef     SE Coef         T       P       VIF
Constant      -0.0086      .2943     -0.029   0.977
Purbanpo     -0.14946    0.03192      -4.68   0.000       1.1
Prsurf       -0.18058    0.04299      -4.20   0.000       1.1
Ruspeed      0.004569   0.004942       0.92   0.360       1.1

S = 0.05510      R-Sq = 55.8%      R-Sq(adj) = 52.8%

Analysis of Variance
Source            DF        SS         MS         F        P
Regression         3  0.172207   0.057402     18.91    0.000
Residual Error    45  0.136602   0.003036
Total             48  0.308809
```

The model has conditionally significant variables for percent urban population and percent surfaced roads. Since average rural speed is not conditionally significant, it is dropped from the model:

Regression Analysis: deaths versus Prurpop, Prsurf

```
The regression equation is
deaths = 0.26116 + 0.155 Prurpop - 0.188 Prsurf
Predictor        Coef     SE Coef         T       P       VIF
Constant      0.26116    0.03919       6.66   0.000
Prurpop       0.15493    0.03132       4.95   0.000       1.0
Prsurf       -0.18831    0.04210      -4.47   0.000       1.0

S = 0.05501      R-Sq = 54.9%      R-Sq(adj) = 53.0%

Analysis of Variance
Source            DF        SS         MS         F        P
Regression         2  0.169612   0.084806     28.03    0.000
Residual Error    46  0.139197   0.003026
Total             48  0.308809
```

This becomes the final model since both variables are conditionally significant.

c. Conclude that the proportions of urban populations and the percent of rural roads that are surfaced are important independent variables in explaining crash deaths. All else being equal, the higher the proportion of urban population, the higher the crash deaths. All else being equal, increases in the proportion of rural roads that are surfaced will result in lower crash deaths. The average rural speed is not conditionally significant.

12.104 a. Correlation matrix

Correlations: Retsales, Unemploy, PerInc

	Retsales	Unemploy
Unemploy	-0.454	
	0.001	
PerInc	-0.098	-0.031
	0.495	0.828

There is a negative association between the dependent and the independent variables. High correlation among the independent variables does not appear to be a problem since the correlation between the independent variables is low.

Descriptive Statistics: Retsales, Unemploy, PerInc

Variable	N	Mean	SE Mean	TrMean	StDev	Minimum	Q1	Median
Retsales	51	14.099	0.292	14.071	2.087	6.817	12.883	13.817
Unemploy	51	5.341	0.172	5.333	1.230	2.900	4.400	5.400
PerInc	51	32161	800	31645	5712	24317	28172	31029

Variable	Q3	Maximum
Retsales	15.015	20.307
Unemploy	6.400	8.300
PerInc	34867	53448

Regression Analysis: Retsales versus Unemploy, PerInc

```
The regression equation is
Retsales = 19.6 - 0.777 Unemploy - 0.000041 PerInc
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	19.565	1.941	10.08	0.000	
Unemploy	-0.7768	0.2166	-3.59	0.001	1.001
PerInc	-0.00004097	0.00004664	-0.88	0.384	1.001

```
S = 1.88261   R-Sq = 21.9%   R-Sq(adj) = 18.6%
```

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	47.674	23.837	6.73	0.003
Residual Error	48	170.122	3.544		
Total	50	217.796			

The 95% confidence intervals for the regression slope coefficients:

$$\hat{\beta}_1 \pm t(s_{\hat{\beta}_1}): \quad -.7768 +/- 2.011(.2166) = -.7768 +/- .436$$

$$\hat{\beta}_2 \pm t(s_{\hat{\beta}_2}): \quad -.000041 +/- 2.011(.000047) = -.000041 +/- .000095$$

b. All things equal, the conditional effect of a \$1,000 decrease in per capita income on retail sales would be to improve retail sales by \$.041.

c. Adding state population as a predictor yields the following regression results:

Regression Analysis: Retsales versus Unemploy, PerInc, Population

```
The regression equation is
Retsales = 19.4 - 0.732 Unemploy - 0.000038 PerInc - 0.000025 Population

Predictor           Coef      SE Coef      T      P    VIF
Constant          19.384        1.979   9.80  0.000
Unemploy         -0.7318       0.2313  -3.16  0.003  1.126
PerInc       -0.00003822   0.00004720  -0.81  0.422  1.011
Population   -0.00002476   0.00004244  -0.58  0.562  1.133

S = 1.89568   R-Sq = 22.5%   R-Sq(adj) = 17.5%

Analysis of Variance

Source          DF       SS      MS     F      P
Regression       3   48.897  16.299  4.54  0.007
Residual Error  47  168.899   3.594
Total           50  217.796
```

The population variable is not conditionally significant and adds little explanatory power, therefore, it will not improve the multiple regression model.

12.106

Regression analysis to predict Breast Cancer Death Rate:

Correlations: BCncrDthRte, Nurses, Female Smoke, Alcohol B

```
               BCncrDthRte      Nurses  Female Smoke
Nurses               0.313
                     0.025
Female Smoke         0.454       0.167
                     0.001       0.243
Alcohol B           -0.096       0.528        -0.019
                     0.503       0.000         0.893
```

Regression Analysis: BCncrDthRte versus Nurses, Female Smoke, Alcohol B

```
The regression equation is
BCncrDthRte = 0.00983 + 0.000004 Nurses + 0.000194 Female Smoke
              - 0.000169 Alcohol B

Predictor           Coef     SE Coef      T      P    VIF
Constant        0.009828    0.001950   5.04  0.000
Nurses        0.00000426  0.00000149   2.86  0.006  1.449
Female Smoke  0.00019419  0.00006243   3.11  0.003  1.046
Alcohol B    -0.00016866  0.00007826  -2.16  0.036  1.409

S = 0.00147421   R-Sq = 33.0%   R-Sq(adj) = 28.7%

Analysis of Variance

Source          DF           SS           MS     F      P
Regression       3  0.000050366  0.000016789  7.72  0.000
Residual Error  47  0.000102145  0.000002173
Total           50  0.000152511
```

This is the final regression model. All else being equal, an increase in 1 nurse per 100,000 population increases the B cancer death rate by .000004 per 1000 population. All else being equal, a 1% increase in the percentage of female smokers increases the death rate by .000194 per 1000 population. All else being equal, a 1% increase in the percentage of binge drinkers decreases the death rate by .000169 per 1000 population. The *t* statistics indicate that all the three independent variables are significant at 5% level and hence a significant relationship exists. The small standard error of estimate of .0015 indicates a small variation between observed and predicted values. 33% of the variation in the B cancer death rate is explained by the variation in the number of nurses, percent of female smokers, and percent of binge drinkers.

Regression analysis to predict Lung Cancer Death Rate:

Correlations: LCncrDthRte, Nurses, Smoker Per, Alcohol B, Median Incom, ...

```
              LCncrDthRte        Nurses     Smoker Per      Alcohol B
Nurses              0.353
                    0.011
Smoker Per          0.687         0.023
                    0.000         0.870
Alcohol B          -0.064         0.528         -0.175
                    0.654         0.000          0.220
Median Income      -0.561         0.012         -0.670          0.183
                    0.000         0.931          0.000          0.198
Per Fam Pov         0.338        -0.096          0.548         -0.376
                    0.015         0.502          0.000          0.006

            Median Income
Per Fam Pov        -0.748
                    0.000
```

Regression Analysis: LCncrDthRte versus Nurses, Smoker Per, ...

```
The regression equation is
LCncrDthRte = 0.0473 + 0.000034 Nurses + 0.00221 Smoker Per - 0.00101 Alcohol B
              - 0.000001 Median Income - 0.00137 Per Fam Pov


Predictor             Coef      SE Coef       T      P    VIF
Constant           0.04732      0.02542    1.86  0.069
Nurses          0.00003449   0.00000803    4.30  0.000  1.422
Smoker Per       0.0022084    0.0004840    4.56  0.000  1.849
Alcohol B       -0.0010087    0.0004628   -2.18  0.035  1.668
Median Income  -0.00000060   0.00000024   -2.48  0.017  2.985
Per Fam Pov     -0.0013658    0.0006742   -2.03  0.049  2.652


S = 0.00801372   R-Sq = 65.8%   R-Sq(adj) = 62.0%


Analysis of Variance

Source          DF         SS          MS      F      P
Regression       5  0.0055526   0.0011105  17.29  0.000
Residual Error  45  0.0028899   0.0000642
Total           50  0.0084425
```

This is the final regression model. All else being equal, an increase in 1 nurse per 100,000 population increases the L cancer death rate by .000034 per 1000 population.

All else being equal, a 1% increase in the percentage of smokers increases the death rate by .00221. All else being equal, a 1% increase in the percentage of binge drinkers decreases the death rate by .00101. All else being equal, a 1$ increase in the median household income decreases the death rate by .000001. All else being equal, a 1% increase in the percent of families below poverty decreases the death rate by .00137. The *t* statistics indicate that all the independent variables are significant at 5% level and hence a significant relationship exists. The small standard error of estimate of .008 indicates a small variation between observed and predicted values. 65.8% of the variation in the L cancer death rate is explained by the variation in the number of nurses, percent of smokers, percent of binge drinkers, median household income, and percent of families below poverty.

12.108 a. Correlation matrix:

Correlations: EconGPA, sex, Acteng, ACTmath, ACTss, ACTcomp, HSPct

	EconGPA	sex	Acteng	ACTmath	ACTss	ACTcomp
sex	0.187					
	0.049					
Acteng	0.387	0.270				
	0.001	0.021				
ACTmath	0.338	-0.170	0.368			
	0.003	0.151	0.001			
ACTss	0.442	-0.105	0.448	0.439		
	0.000	0.375	0.000	0.000		
ACTcomp	0.474	-0.084	0.650	0.765	0.812	
	0.000	0.478	0.000	0.000	0.000	
HSPct	0.362	0.216	0.173	0.290	0.224	0.230
	0.000	0.026	0.150	0.014	0.060	0.053

There exists a positive relationship between EconGPA and all of the independent variables, which is expected. Note that there is a high correlation between the composite ACT score and the individual components, which is again, as expected. Thus, high correlation among the independent variables is likely to be a serious concern in this regression model.

Regression Analysis: EconGPA versus sex, Acteng, ...

```
The regression equation is
EconGPA = - 0.050 + 0.261 sex + 0.0099 Acteng + 0.0064 ACTmath + 0.0270
ACTss + 0.0419 ACTcomp + 0.00898 HSPct

71 cases used 41 cases contain missing values
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	-0.0504	0.6554	-0.08	0.939	
sex	0.2611	0.1607	1.62	0.109	1.5
Acteng	0.00991	0.02986	0.33	0.741	2.5
ACTmath	0.00643	0.03041	0.21	0.833	4.3
ACTss	0.02696	0.02794	0.96	0.338	4.7
ACTcomp	0.04188	0.07200	0.58	0.563	12.8
HSPct	0.008978	0.005716	1.57	0.121	1.4

S = 0.4971 R-Sq = 34.1% R-Sq(adj) = 27.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	8.1778	1.3630	5.52	0.000
Residual Error	64	15.8166	0.2471		
Total	70	23.9945			

As expected, high correlation among the independent variables is affecting the results. A strategy of dropping the variable with the lowest t-statistic with each successive model causes the dropping of the following variables (in order): 1) ACTmath, 2) ACTeng, 3) ACTss, 4) HSPct. The two variables that remain are the final model of gender and ACTcomp:

Regression Analysis: EconGPA versus sex, ACTcomp

```
The regression equation is
EconGPA = 0.322 + 0.335 sex + 0.0978 ACTcomp

73 cases used 39 cases contain missing values

Predictor        Coef     SE Coef        T        P       VIF
Constant       0.3216      0.5201     0.62    0.538
sex            0.3350      0.1279     2.62    0.011       1.0
ACTcomp       0.09782     0.01989     4.92    0.000       1.0

S = 0.4931      R-Sq = 29.4%     R-Sq(adj) = 27.3%

Analysis of Variance
Source            DF        SS        MS        F        P
Regression         2    7.0705    3.5352    14.54    0.000
Residual Error    70   17.0192    0.2431
Total             72   24.0897
```

Both independent variables are conditionally significant.

b. The model could be used in college admission decisions by creating a predicted GPA in economics based on sex and ACT comp scores. This predicted GPA could then be used with other factors in deciding admission. Note that this model predicts that females will outperform males with equal test scores. Using this model as the only source of information may lead to charges of unequal treatment.

12.110

Correlations: Sale 2 Price, Time Interva, Sale 1 Price, Atlanta, Chicago, ...

```
              Sale 2 Price  Time Interval   Sale 1 Price        Atlanta
Time Interval       -0.029
                      0.07
Sale 1 Price         0.976         -0.151
                     0.000          0.000
Atlanta             -0.105          0.008         -0.076
                     0.000          0.596          0.000
Chicago             -0.047         -0.018         -0.016         -0.333
                     0.003          0.265          0.322          0.000
Dallas              -0.033          0.035         -0.047         -0.333
                     0.038          0.027          0.003          0.000
Oakland              0.185         -0.026          0.140         -0.333
                     0.000          0.105          0.000          0.000

                   Chicago         Dallas
Dallas              -0.333
                     0.000
Oakland             -0.333         -0.333
                     0.000          0.000
```

There exists a negative relationship between Sale 2 price and all of the independent variables except Sale 1 price and Oakland. Note that there is a high correlation between the Sale 1 price and Sale 2 price, as expected. High correlation among the independent variables is likely to be a serious concern in this regression model.

Regression Analysis: Sale 2 Price versus Time Interva, Sale 1 Price, ...

```
* Oakland is highly correlated with other X variables
* Oakland has been removed from the equation.


The regression equation is
Sale 2 Price = 14726 + 1225 Time Interval + 0.978 Sale 1 Price - 16559 Atlanta
               - 16339 Chicago - 8287 Dallas


Predictor            Coef    SE Coef       T      P    VIF
Constant          14725.7      834.7   17.64  0.000
Time Interval     1224.91      28.00   43.75  0.000  1.024
Sale 1 Price     0.977572   0.002770  352.89  0.000  1.044
Atlanta          -16558.5      927.3  -17.86  0.000  1.527
Chicago          -16339.4      923.3  -17.70  0.000  1.514
Dallas            -8286.8      925.4   -8.95  0.000  1.521

S = 20551.2   R-Sq = 97.0%   R-Sq(adj) = 97.0%

Analysis of Variance

Source            DF           SS           MS         F      P
Regression         5  5.46776E+13  1.09355E+13  25892.07  0.000
Residual Error  3994  1.68687E+12    422350080
Total           3999  5.63645E+13
```

97% of the variation in the second or final sales price can be explained by variations in the interval between house sales, and the initial house sales price with adjustments for the four major U.S. market areas.

All the variables are highly significant in explaining the final sales price at all levels of alpha. The time interval and initial sales price have the expected sign in the regression model. The F-test of the significance of the overall model shows that we reject H_0 in favor of H_1 that at least one slope coefficient is not equal to zero. The F-test yielded a p-value of .000.
The sample correlation between the observed and the predicted values of the final sales price is .985.

12.112

a.

To predict the percentage of students who graduate in 4 years from highly ranked private colleges, the following list of potential predictor variables are selected:

1. Undergrad. Enrollment – the undergraduate enrollments in the college
2. Admission Rate – the rate of the admission of students into the college
3. Student/faculty Ratio – the ratio of the students to the faculty
4. Quality Rank – the rank of the quality of the private colleges

b.

Multiple regression using the listed predictor variables:

Regression Analysis: 4-year Grad. versus Undergrad. E, Admission Ra, ...

```
The regression equation is
4-year Grad. Rate = 0.814 - 0.000009 Undergrad. Enrollment
                    - 0.0064 Admission Rate + 0.0154 Student/faculty Ratio
                    - 0.00387 Quality Rank


Predictor                      Coef      SE Coef      T      P    VIF
Constant                    0.81448      0.02972  27.40  0.000
Undergrad. Enrollment   -0.00000919   0.00000175  -5.26  0.000  1.186
Admission Rate             -0.00638      0.06459  -0.10  0.922  3.522
Student/faculty Ratio      0.015371     0.003920   3.92  0.000  2.008
Quality Rank             -0.0038718    0.0004267  -9.07  0.000  3.480


S = 0.0650174   R-Sq = 69.9%   R-Sq(adj) = 68.6%


Analysis of Variance
Source          DF       SS        MS      F      P
Regression       4  0.91267   0.22817  53.98  0.000
Residual Error  93  0.39313   0.00423
Total           97  1.30580
```

c.
The final regression after eliminating the insignificant predictor variable, Admission rate:

Regression Analysis: 4-year Grad. versus Undergrad. E, Student/facu, ...

```
The regression equation is
4-year Grad. Rate = 0.814 - 0.000009 Undergrad. Enrollment
                    + 0.0153 Student/faculty Ratio - 0.00390 Quality Rank


Predictor                     Coef      SE Coef       T      P    VIF
Constant                   0.81407      0.02927   27.81  0.000
Undergrad. Enrollment  -0.00000914   0.00000167   -5.46  0.000  1.100
Student/faculty Ratio     0.015268     0.003760    4.06  0.000  1.867
Quality Rank            -0.0039013    0.0003034  -12.86  0.000  1.778


S = 0.0646740   R-Sq = 69.9%   R-Sq(adj) = 68.9%


Analysis of Variance

Source          DF       SS       MS      F      P
Regression       3  0.91262  0.30421  72.73  0.000
Residual Error  94  0.39318  0.00418
Total           97  1.30580
```

d.
Undergrad.Enrollment and Quality Rank are negatively related to 4-year Grad.Rate. Student/faculty Ratio has the expected sign for the dependent variable.
All else being equal, a 1-unit increase in Undergrad.Enrollment will decrease 4-year Grad Rate by .000009. All else being equal, a 1-unit increase in Student/faculty Ratio will increase 4-year Grad Rate by .0153. All else being equal, a 1-unit increase in Quality Rank will decrease 4-year Grad Rate by .0039.

12.114

a. daycode2 is a dummy variable where first interview is coded 0 and second interview coded 1.

Regression Analysis: HEI2005 versus doc_bp, waistper, ...

```
The regression equation is
HEI2005 = 47.2 - 0.734 doc_bp - 7.15 waistper + 0.0447 BMI
      + 0.522 sr_overweight + 3.64 female + 0.182 age + 2.33 daycode2
```

8217 cases used, 373 cases contain missing values

Predictor	Coef	SE Coef	T	P	VIF
Constant	47.177	1.139	41.42	0.000	
doc_bp	-0.7339	0.3638	-2.02	0.044	1.269
waistper	-7.149	2.377	-3.01	0.003	6.896
BMI	0.04466	0.06199	0.72	0.471	6.332
sr_overweight	0.5219	0.3799	1.37	0.170	1.519
female	3.6403	0.3780	9.63	0.000	1.513
age	0.181604	0.009224	19.69	0.000	1.338
daycode2	2.3331	0.3073	7.59	0.000	1.000

S = 13.9149   R-Sq = 6.9%   R-Sq(adj) = 6.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	117606	16801	86.77	0.000
Residual Error	8209	1589454	194		
Total	8216	1707060			

b. Adding the dummy variable, immigrant to the regression model

Regression Analysis: HEI2005 versus doc_bp, waistper, ...

```
The regression equation is
HEI2005 = 44.0 - 0.352 doc_bp - 4.29 waistper - 0.0134 BMI + 0.823 sr_overweight
          + 3.45 female + 0.184 age + 2.36 daycode2 + 7.38 immigrant
```

8217 cases used, 373 cases contain missing values

Predictor	Coef	SE Coef	T	P	VIF
Constant	44.048	1.126	39.13	0.000	
doc_bp	-0.3517	0.3563	-0.99	0.324	1.273
waistper	-4.291	2.329	-1.84	0.065	6.923
BMI	-0.01344	0.06069	-0.22	0.825	6.347
sr_overweight	0.8234	0.3718	2.21	0.027	1.521
female	3.4510	0.3698	9.33	0.000	1.514
age	0.184361	0.009022	20.43	0.000	1.338
daycode2	2.3583	0.3005	7.85	0.000	1.000
immigrant	7.3772	0.3809	19.37	0.000	1.018

S = 13.6082   R-Sq = 11.0%   R-Sq(adj) = 10.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	187086	23386	126.29	0.000
Residual Error	8208	1519974	185		
Total	8216	1707060			

c. Adding the dummy variable, single to the initial regression model

Regression Analysis: HEI2005 versus doc_bp, waistper, ...

```
The regression equation is
HEI2005 = 48.0 - 0.654 doc_bp - 7.69 waistper + 0.0691 BMI + 0.146 sr_overweight
          + 4.02 female + 0.180 age + 2.31 daycode2 - 2.40 single

8215 cases used, 375 cases contain missing values

Predictor            Coef   SE Coef      T      P    VIF
Constant           48.039     1.141  42.11  0.000
doc_bp            -0.6538    0.3628  -1.80  0.072  1.270
waistper           -7.692     2.370  -3.25  0.001  6.904
BMI               0.06910   0.06187   1.12  0.264  6.352
sr_overweight      0.1463    0.3817   0.38  0.702  1.544
female             4.0179    0.3803  10.57  0.000  1.542
age              0.180269  0.009194  19.61  0.000  1.339
daycode2           2.3059    0.3063   7.53  0.000  1.000
single            -2.4030    0.3197  -7.52  0.000  1.035

S = 13.8664   R-Sq = 7.5%   R-Sq(adj) = 7.4%
Analysis of Variance

Source            DF       SS     MS      F      P
Regression         8   128336  16042  83.43  0.000
Residual Error  8206  1577814    192
Total           8214  1706151
```

d. Adding the dummy variable, fsp to the initial regression model

Regression Analysis: HEI2005 versus doc_bp, waistper, ...

```
The regression equation is
HEI2005 = 47.3 - 0.706 doc_bp - 6.29 waistper + 0.0470 BMI + 0.260 sr_overweight
          + 3.70 female + 0.171 age + 2.30 daycode2 - 3.51 fsp

8114 cases used, 476 cases contain missing values

Predictor            Coef   SE Coef      T      P    VIF
Constant           47.281     1.141  41.44  0.000
doc_bp            -0.7064    0.3642  -1.94  0.052  1.266
waistper           -6.288     2.382  -2.64  0.008  6.897
BMI               0.04699   0.06200   0.76  0.449  6.322
sr_overweight      0.2598    0.3816   0.68  0.496  1.526
female             3.6984    0.3789   9.76  0.000  1.514
age              0.171463  0.009334  18.37  0.000  1.360
daycode2           2.2953    0.3079   7.46  0.000  1.000
fsp               -3.5064    0.4716  -7.43  0.000  1.037

S = 13.8547   R-Sq = 7.4%   R-Sq(adj) = 7.4%

Analysis of Variance

Source            DF       SS     MS      F      P
Regression         8   125197  15650  81.53  0.000
Residual Error  8105  1555766    192
Total           8113  1680963
```

12.116

a. daycode2 is a dummy variable where first interview is coded 0 and second interview coded 1.

Regression Analysis: daily_cost versus doc_bp, waistper, ...

```
The regression equation is
daily_cost = 6.83 - 0.189 doc_bp + 1.42 waistper - 0.0192 BMI
           + 0.107 sr_overweight - 1.35 female - 0.0359 age - 0.205 daycode2


8217 cases used, 373 cases contain missing values


Predictor            Coef    SE Coef       T      P    VIF
Constant           6.8293     0.2347   29.10  0.000
doc_bp           -0.18874    0.07497   -2.52  0.012  1.269
waistper           1.4179     0.4898    2.90  0.004  6.896
BMI              -0.01916    0.01277   -1.50  0.134  6.332
sr_overweight     0.10659    0.07828    1.36  0.173  1.519
female           -1.35316    0.07789  -17.37  0.000  1.513
age             -0.035866   0.001901  -18.87  0.000  1.338
daycode2         -0.20460    0.06331   -3.23  0.001  1.000

S = 2.86735   R-Sq = 9.4%   R-Sq(adj) = 9.3%

Analysis of Variance

Source            DF       SS      MS       F      P
Regression         7   7009.6  1001.4  121.80  0.000
Residual Error  8209  67491.9     8.2
Total           8216  74501.5
```

b. Adding the dummy variable, immigrant to the regression model

Regression Analysis: daily_cost versus doc_bp, waistper, ...

```
The regression equation is
daily_cost = 6.98 - 0.207 doc_bp + 1.28 waistper - 0.0164 BMI
           + 0.0923 sr_overweight - 1.34 female - 0.0360 age - 0.206 daycode2
           - 0.351 immigrant

8217 cases used, 373 cases contain missing values

Predictor            Coef    SE Coef       T      P    VIF
Constant           6.9780     0.2369   29.45  0.000
doc_bp           -0.20691    0.07500   -2.76  0.006  1.273
waistper           1.2821     0.4902    2.62  0.009  6.923
BMI              -0.01639    0.01277   -1.28  0.199  6.347
sr_overweight     0.09226    0.07826    1.18  0.238  1.521
female           -1.34416    0.07783  -17.27  0.000  1.514
age             -0.035997   0.001899  -18.96  0.000  1.338
daycode2         -0.20580    0.06325   -3.25  0.001  1.000
immigrant        -0.35068    0.08016   -4.37  0.000  1.018

S = 2.86419   R-Sq = 9.6%   R-Sq(adj) = 9.5%
Analysis of Variance

Source            DF        SS      MS       F      P
Regression         8   7166.63  895.83  109.20  0.000
Residual Error  8208  67334.92    8.20
Total           8216  74501.55
```

c. Adding the dummy variable, single to the initial regression model

Regression Analysis: daily_cost versus doc_bp, waistper, ...

```
The regression equation is
daily_cost = 6.89 - 0.187 doc_bp + 1.36 waistper - 0.0169 BMI
             + 0.0805 sr_overweight - 1.32 female - 0.0359 age - 0.208 daycode2
             - 0.182 single


8215 cases used, 375 cases contain missing values


Predictor             Coef    SE Coef       T      P    VIF
Constant            6.8942     0.2358   29.24  0.000
doc_bp            -0.18678    0.07498   -2.49  0.013  1.270
waistper            1.3629     0.4899    2.78  0.005  6.904
BMI               -0.01695    0.01279   -1.33  0.185  6.352
sr_overweight      0.08051    0.07889    1.02  0.308  1.544
female            -1.32153    0.07860  -16.81  0.000  1.542
age              -0.035904   0.001900  -18.89  0.000  1.339
daycode2          -0.20834    0.06330   -3.29  0.001  1.000
single            -0.18199    0.06607   -2.75  0.006  1.035


S = 2.86594   R-Sq = 9.5%   R-Sq(adj) = 9.4%


Analysis of Variance

Source             DF        SS      MS       F      P
Regression          8   7066.32  883.29  107.54  0.000
Residual Error   8206  67400.82    8.21
Total            8214  74467.14
```

d. Adding the dummy variable, fsp to the initial regression model

Regression Analysis: daily_cost versus doc_bp, waistper, ...

```
The regression equation is
daily_cost = 6.84 - 0.165 doc_bp + 1.55 waistper - 0.0171 BMI
           + 0.0684 sr_overweight - 1.35 female - 0.0378 age - 0.206 daycode2
           - 0.783 fsp


8114 cases used, 476 cases contain missing values


Predictor          Coef   SE Coef       T      P    VIF
Constant         6.8409    0.2354   29.06  0.000
doc_bp         -0.16534   0.07513   -2.20  0.028  1.266
waistper         1.5483    0.4913    3.15  0.002  6.897
BMI            -0.01707   0.01279   -1.33  0.182  6.322
sr_overweight   0.06845   0.07872    0.87  0.385  1.526
female         -1.34695   0.07817  -17.23  0.000  1.514
age           -0.037831  0.001926  -19.65  0.000  1.360
daycode2       -0.20646   0.06351   -3.25  0.001  1.000
fsp            -0.78343   0.09729   -8.05  0.000  1.037


S = 2.85811   R-Sq = 10.1%   R-Sq(adj) = 10.0%


Analysis of Variance

Source            DF        SS      MS       F      P
Regression         8   7438.61  929.83  113.83  0.000
Residual Error  8105  66208.22    8.17
Total           8113  73646.83
```

Chapter 13:

Additional Topics in Regression Analysis

13.2 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

where Y_i = Wages

X_1 = Years of experience

X_2 = 1 for Germany, 0 otherwise

X_3 = 1 for Great Britain, 0 otherwise

X_4 = 1 for Japan, 0 otherwise

X_5 = 1 for Turkey, 0 otherwise

The excluded category consists of wages in the United States

13.4 a. For any observation, the values of the dummy variables sum to one. Since the equation has an intercept term, there is perfect multicollinearity and the existence of the "dummy variable trap".

b. β_3 measures the expected difference between demand in the first and fourth quarters, all else equal. β_4 measures the expected difference between demand in the second and fourth quarters, all else equal. β_5 measures the expected difference between demand in the third and fourth quarters, all else equal.

13.6 Analyze the correlation matrix first:

Correlations: Time (in sec, Gender_1male, Day Shift, Evening shif, ...

	Time (in seconds	Gender_1male)	Day Shift
Gender_1male)	-0.016		
	0.872		
Day Shift	-0.143	0.127	
	0.136	0.188	
Evening shift	0.032	0.032	-0.470
	0.736	0.744	0.000
Benefits_rate_2	-0.050	0.098	-0.075
	0.606	0.311	0.434
Benefits_rate_3	0.008	0.006	-0.053
	0.933	0.952	0.581
Benefits_rate_4	-0.027	-0.189	0.032
	0.777	0.049	0.738
Benefits_rate_5	0.086	0.113	0.036
	0.372	0.240	0.711

	Evening shift	Benefits_rate_2	Benefits_rate_3	Benefits_rate_4
Benefits_rate_2	-0.057			
	0.552			
Benefits_rate_3	-0.117	-0.030		
	0.222	0.753		
Benefits_rate_4	-0.111	-0.061	-0.203	
	0.250	0.524	0.034	
Benefits_rate_5	0.100	-0.111	-0.366	-0.742
	0.300	0.249	0.000	0.000

Regression Analysis: Time (in sec versus Gender_1male, Day Shift, ...

```
The regression equation is
Time (in seconds) = 250 + 0.74 Gender_1male) - 6.19 Day Shift
                    - 1.09 Evening shift + 1.7 Benefits_rate_2
                    + 13.5 Benefits_rate_3 + 13.2 Benefits_rate_4
                    + 15.2 Benefits_rate_5
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	249.557	9.912	25.18	0.000	
Gender_1male)	0.740	3.495	0.21	0.833	1.082
Day Shift	-6.192	4.045	-1.53	0.129	1.369
Evening shift	-1.086	4.554	-0.24	0.812	1.428
Benefits_rate_2	1.70	20.14	0.08	0.933	1.295
Benefits_rate_3	13.54	10.83	1.25	0.214	3.434
Benefits_rate_4	13.220	9.714	1.36	0.177	6.901
Benefits_rate_5	15.177	9.288	1.63	0.105	7.484

S = 17.6147   R-Sq = 5.2%   R-Sq(adj) = 0.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	1753.4	250.5	0.81	0.583
Residual Error	102	31648.2	310.3		
Total	109	33401.7			

The *F*-test for the significance of the overall regression indicates that the model has no significant explanatory power. None of the independent variables are significantly different from zero. However, it appears that working on different shifts has an expected negative influence on the productivity of the workers. Gender differences have a positive effect on the productivity. And, as expected, the employee attitude toward the present benefits plan provided by the company has an increasingly positive effect on the productivity as the rating for the benefits increases.

13.8 The following variables are defined for the experiment

Y = The number of defective parts per 8 hour work shift

Shift

X_1 = 1 if Day shift, 0 otherwise

X_2 = 1 if Afternoon shift, 0 otherwise

The excluded category is Night shift Material suppliers

X_3 = 1 if Supplier 1, 0 otherwise

X_4 = 1 if Supplier 2, 0 otherwise

X_5 = 1 if Supplier 3, 0 otherwise

The exclkuded category is Supplier 4

X_6 = Production level

X_7 = Number of shift workers

Two series of dummy variables are required to analyze the impact of shifts and material suppliers on the number of defective parts. For each dummy variable, (k–1) categories are required to avoid the 'dummy variable trap.' Interaction terms may be appropriate between the production level and shift.

13.10 a. The following variables are defined for the experiment

Y = Worker compensation – annual average rate of wage increase

X_1 = Years of experience

Job classification group

X_2 = 1 if Administrative, 0 otherwise

X_3 = 1 if Analytical, 0 otherwise

The excluded category is Managerial

X_4 = 1 for MBA, 0 otherwise

Gender

X_5 =1 if Male, 0 otherwise

The excluded category is Female

Race

X_6 = 1 if White, 0 otherwise

X_7 =1 ifBlack, 0 otherwise

The excluded category is Latino

Average annual rate of wage increase can be analyzed with a combination of continuous independent variables and a series of dummy variables. Dummy variables are required to analyze the impact of job classifications on salary. Discrimination can be measured by the size of the dummy variable on gender and on race. For each dummy variable, (k–1) categories are required to avoid the 'dummy variable trap.'

b. Key points would include interpretations of coefficients on the dichotomous variables and the existence, if any, of interaction terms. Tests of significance of the overall regression, t-tests on significance of individual coefficients and model diagnostics would be conducted to provide statistical evidence of wage discrimination.

13.12 a. $H_0 : \beta_2 = 0, \quad H_1 : \beta_2 > 0$

$t = \dfrac{.027}{.0021} = 12.86; \quad t_{21,.01} = 2.518$, therefore, reject H_0 at the 1% level

b. $t_{21,.025} = 2.08$, 95% CI: $.142 \pm 2.08(.047)$, $(.0442, .2398)$

c. Total effect: $\dfrac{.142}{1-.432} = .25$

The expected impact over time of a \$1 increase in disposable income per student is \$.25 increase in clothing expenditures.

13.14

Regression Analysis: Y_money versus X1_income, X2_ir, Y_lagmoney

```
The regression equation is
Y_money = - 2309 + 0.158 X1_income - 14126 X2_ir + 1.06 Y_lagmoney
27 cases used 1 cases contain missing values

Predictor        Coef     SE Coef        T       P
Constant        -2309        1876    -1.23   0.231
X1_incom       0.1584      0.2263     0.70   0.491
X2_ir          -14126        6372    -2.22   0.037
Y_lagmon       1.0631      0.1266     8.40   0.000

S = 456.1       R-Sq = 97.6%      R-Sq(adj) = 97.3%

Analysis of Variance
Source            DF         SS         MS        F        P
Regression         3  194108213   64702738   311.02    0.000
Residual Error    23    4784762     208033
Total             26  198892975

Source       DF      Seq SS
X1_incom      1   167714527
X2_ir         1    11728933
Y_lagmon      1    14664753
Unusual Observations
Obs   X1_incom     Y_money         Fit       SE Fit     Residual      St Resid
 24      17455     24975.2     23990.1        186.1        985.1
2.37R
 25      16620     24736.3     24663.3        322.8         73.0          0.23
X
 26      17779     23407.3     24922.0        189.3      -1514.7          -3.6
Durbin-Watson statistic = 1.65
```

The multiple regression with all of the independent variables indicates that 97.6% of the variation in the quantity of money can be explained by all of the independent variables. The *F*-test for the significance of the overall regression indicates the model has significant explanatory power. However, not all of the independent variables are significantly different from zero. It appears that income is not statistically significant different from zero. Local authority interest rate has an expected negative influence on the quantity of money.

13.16

Regression Analysis: Y_income versus X_money, Y_lagincome

```
The regression equation is
Y_income = 11843 + 0.388 X_money + 0.807 Y_lagincome
19 cases used 1 cases contain missing values

Predictor         Coef      SE Coef          T        P
Constant         11843         5666       2.09    0.053
X_money         0.3875       0.3778       1.03    0.320
Y_laginc        0.8068       0.1801       4.48    0.000

S = 1952        R-Sq = 99.6%      R-Sq(adj) = 99.6%
Analysis of Variance
Source            DF           SS           MS          F        P
Regression         2  15787845901   7893922950    2071.84    0.000
Residual Error    16     60961685      3810105
Total             18  15848807586

Source        DF      Seq SS
X_money        1 15711421835
Y_laginc       1    76424065

Unusual Observations
Obs     X_money    Y_income          Fit       SE Fit     Residual     St Resid
 13       68694      182744       178826          521         3918
2.08R
```

The multiple regression indicates that 99.6% of the variation in income can be explained by all of the independent variables. The *F*-test for the significance of the overall regression indicates the model has significant explanatory power. However, not all the independent variables are significantly different from zero. It appears that money supply is not statistically significantly different from zero. Money supply and the lagged variable of income have an expected positive influence on the income in Canada.

13.18

Regression Analysis: Y_logCons versus X_LogDI, Y_laglogCons

```
The regression equation is
Y_logCons = 0.405 + 0.373 X_LogDI + 0.558 Y_laglogCons

28 cases used 1 cases contain missing values

Predictor        Coef    SE Coef        T        P
Constant       0.4049     0.1051     3.85    0.001
X_LogDI        0.3734     0.1075     3.47    0.002
Y_laglog       0.5577     0.1243     4.49    0.000

S = 0.03023      R-Sq = 99.6%      R-Sq(adj) = 99.6%
Analysis of Variance
Source            DF         SS        MS         F        P
Regression         2     6.1960    3.0980   3389.90    0.000
Residual Error    25     0.0228    0.0009
Total             27     6.2189

Source         DF      Seq SS
X_LogDI         1      6.1776
Y_laglog        1      0.0184

Unusual Observations
Obs    X_LogDI    Y_logCon        Fit     SE Fit    Residual    St Resid
  9       5.84     5.80814    5.72298    0.01074     0.08517
3.01R

Durbin-Watson statistic = 1.63
```

The multiple regression indicates that 99.6% of the variation in the private consumption can be explained by all of the independent variables. The *F*-test for the significance of the overall regression indicates the model has significant explanatory power with all the independent variables significantly different from zero. As expected, disposable income and the lagged variable of private consumption have an expected positive effect on the dependent variable.

13.20 a. In the special case where the sample correlations between X_1 and X_2 is zero, the estimate for β_1 will be the same whether or not X_2 is included in the regression equation. In the simple linear regression of Y on X_1, the intercept term will embody the influence of X_2 on Y, under these special circumstances.

b.

$$b_1 = \frac{\sum(x_{2i}-\bar{x}_2)^2\sum(x_{1i}-\bar{x}_1)(y_{1i}-\bar{y}) - \sum(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)\sum(x_{2i}-\bar{x}_2)(y_i-\bar{y})}{\sum(x_{1i}-\bar{x}_1)^2\sum(x_{2i}-\bar{x}_2)^2 - [\sum(x_{1i}-\bar{x}_1)\sum(x_{2i}-\bar{x}_2)]^2}$$

If the sample correlation between X_1 and X_2 is zero, then $\sum(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)=0$ and the slope coefficient equation can be simplified. The result is

$$b_1 = \frac{\sum(x_{1i}-\bar{x}_1)(y_{1i}-\bar{y})}{\sum(x_{1i}-\bar{x}_1)^2}$$ which is the estimated slope coefficient for the bivariate linear regression of Y on X_1.

13.22

Results for: CITYDATR.XLS

Regression Analysis: hseval versus Comper, Homper, ...

```
The regression equation is
hseval = - 19.0 - 26.4 Comper - 12.1 Homper - 15.5 Indper + 7.22 sizehse
         + 0.00408 incom72
Predictor        Coef      SE Coef        T        P
Constant       -19.02        13.20    -1.44    0.153
Comper        -26.393        9.890    -2.67    0.009
Homper        -12.123        7.508    -1.61    0.110
Indper        -15.531        8.630    -1.80    0.075
sizehse         7.219        2.138     3.38    0.001
incom72      0.004081     0.001555     2.62    0.010

S = 3.949       R-Sq = 40.1%     R-Sq(adj) = 36.5%
Analysis of Variance
Source            DF          SS         MS        F        P
Regression         5      876.80     175.36    11.25    0.000
Residual Error    84     1309.83      15.59
Total             89     2186.63

Source       DF      Seq SS
Comper        1      245.47
Homper        1        1.38
Indper        1      112.83
sizehse       1      409.77
incom72       1      107.36

Unusual Observations
Obs     Comper     hseval        Fit      SE Fit     Residual     St Resid
 23      0.100     20.003     28.296       1.913       -8.294            -
2.40RX
 24      0.103     20.932     29.292       2.487       -8.360            -
2.73RX
 29      0.139     16.498     19.321       1.872       -2.823        -0.81
X
 30      0.141     16.705     19.276       1.859       -2.570        -0.74
X
 75      0.112     35.976     24.513       0.747       11.463
2.96R
 76      0.116     35.736     24.418       0.749       11.317
2.92R
R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.
Durbin-Watson statistic = 1.03
```

Dropping the insignificant independent variables: Homper and Indper yields:

Regression Analysis: hseval versus Comper, sizehse, incom72

```
The regression equation is
hseval = - 34.2 - 13.9 Comper + 8.27 sizehse + 0.00364 incom72

Predictor         Coef      SE Coef          T        P
Constant        -34.24        10.44      -3.28    0.002
Comper         -13.881        6.974      -1.99    0.050
sizehse          8.270        1.957       4.23    0.000
incom72       0.003636     0.001456       2.50    0.014

S = 3.983       R-Sq = 37.6%      R-Sq(adj) = 35.4%
Analysis of Variance
Source            DF          SS          MS         F        P
Regression         3      822.53      274.18     17.29    0.000
Residual Error    86     1364.10       15.86
Total             89     2186.63

Source        DF       Seq SS
Comper         1       245.47
sizehse        1       478.09
incom72        1        98.98

Unusual Observations
Obs      Comper      hseval         Fit        SE Fit      Residual     St Resid
 49       0.282      29.810      23.403         1.576         6.407         1.75
X
 50       0.284      30.061      23.380         1.583         6.681         1.83
X
 75       0.112      35.976      24.708         0.674        11.268
2.87R
 76       0.116      35.736      24.659         0.667        11.077
2.82R

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.
Durbin-Watson statistic = 1.02
```

Excluding median rooms per residence (Sizehse):

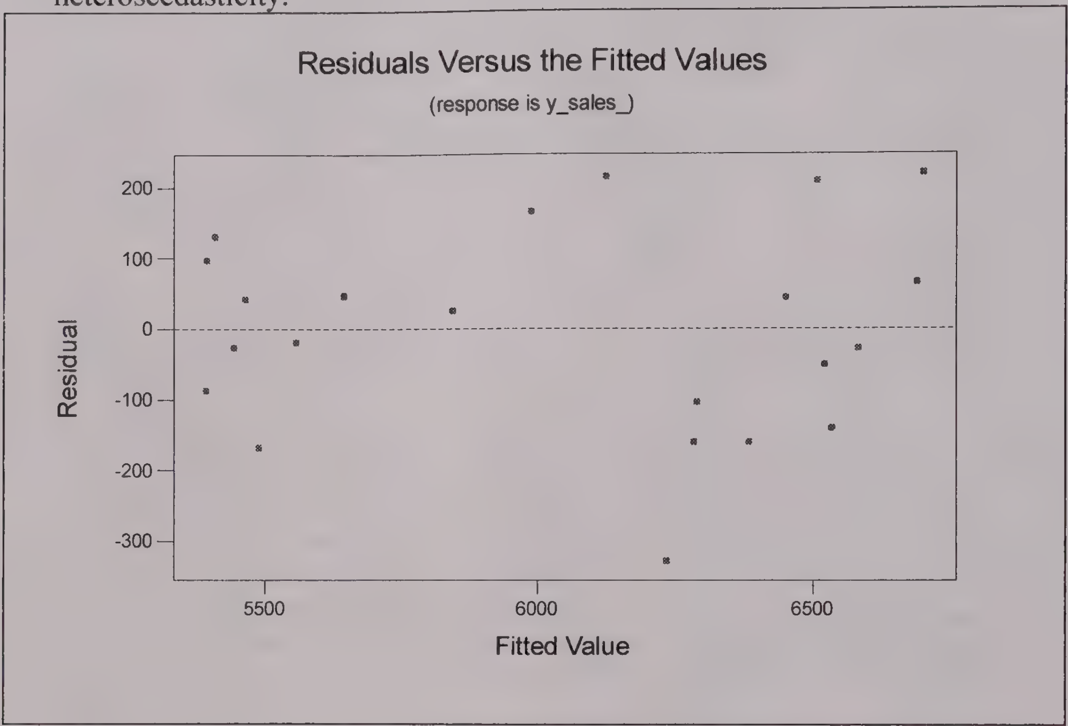
Regression Analysis: hseval versus Comper, incom72

```
The regression equation is
hseval = 4.69 - 20.4 Comper + 0.00585 incom72
Predictor         Coef      SE Coef          T        P
Constant         4.693        5.379       0.87    0.385
Comper         -20.432        7.430      -2.75    0.007
incom72       0.005847     0.001484       3.94    0.000

S = 4.352       R-Sq = 24.7%      R-Sq(adj) = 22.9%
Analysis of Variance
Source            DF          SS          MS         F        P
Regression         2      539.20      269.60     14.24    0.000
Residual Error    87     1647.44       18.94
Total             89     2186.63
Source        DF       Seq SS
Comper         1       245.47
incom72        1       293.73
Durbin-Watson statistic = 0.98
```

Note that the coefficient on percent of commercial property for both of the models is negative; however, it is larger in the second model where the median rooms variable is excluded.

13.24 If Y is, in fact, strongly influenced by X_2, dropping it from the regression equation could lead to serious specification bias. Instead of dropping the variable, it is preferable to acknowledge that, while the group as a whole is clearly influential, the data does not contain information to allow the disentangling of the separate effects of each of the explanatory variables with some degree of precision.

13.26 a. Graphical check for heteroscedasticity shows no evidence of strong heteroscedasticity.



b. The auxiliary regression is $e^2 = -63310.41 + 13.75\hat{y}$

$n = 22$, $R^2 = .06954$, $nR^2 = 1.5299 < 2.71 = \chi^2_{1,.10}$

Therefore, do not reject H_0 that the error terms have constant variance at the 10% level.

13.28 a. The multiple regression of Y on X_1, X_2 and X_3.

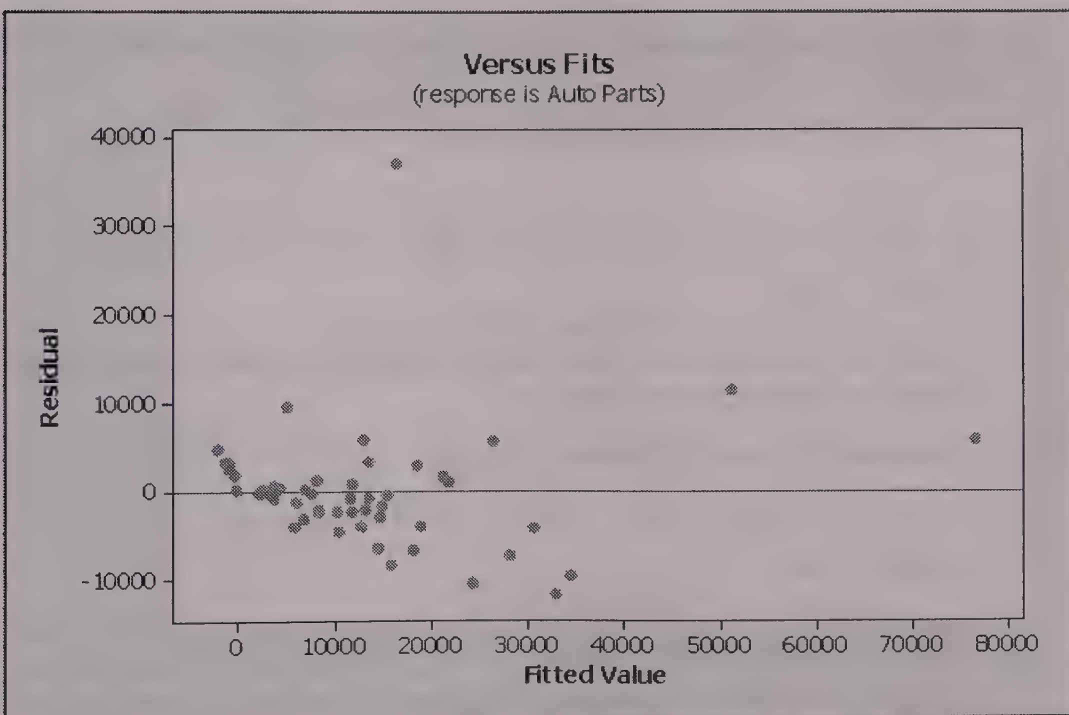
Correlations: Auto Parts, Mfg Pay, Tot Employ, Female Employ

	Auto Parts	Mfg Pay	Tot Employ
Mfg Pay	0.872		
	0.000		
Tot Employ	-0.256	-0.233	
	0.070	0.099	
Female Employ	-0.338	-0.275	0.961
	0.015	0.051	0.000

Regression Analysis: Auto Parts versus Mfg Pay, Tot Employ, Female Emplo

```
The regression equation is
Auto Parts = - 191 + 0.993 Mfg Pay + 2133 Tot Employ - 2302 Female Employ

Predictor          Coef   SE Coef       T       P
Constant           -191     17463   -0.01   0.991
Mfg Pay         0.99253   0.08311   11.94   0.000
Tot Employ       2132.6     906.4    2.35   0.023
Female Employ   -2302.1     854.6   -2.69   0.010

S = 7259.95    R-Sq = 79.5%    R-Sq(adj) = 78.2%

Analysis of Variance

Source          DF           SS          MS      F       P
Regression       3   9636134406  3212044802  60.94   0.000
Residual Error  47   2477219881    52706806
Total           50  12113354287
```

b. Graphical check for heteroscedasticity shows no evidence of strong heteroscedasticity



c. The auxiliary regression is $e^2 = 25046821 + 1731\hat{y}$

$n = 51$, $R^2 = .015$, $nR^2 = .765 < 2.71 = \chi^2_{1,.10}$, therefore, do not reject the H_0 that the error terms have constant variance at the 10% level.

13.30 Given that $Var(\varepsilon_i) = Kx_i^2 \quad (K > 0)$

$$Var(\varepsilon_i / x_i) = \frac{1}{x_i^2} Var(\varepsilon_i) = \frac{1}{x_i^2} Kx_i^2 = K$$

If the squared relationship can be found between the variance of the error terms and X_i such as $Var(\varepsilon_i) = Kx_i^2$, the problem of heteroscedasticity can be removed by dividing both sides of the regression equation by X_i.

13.32 Test for the presence of autocorrelation.

$H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = .50$, $n = 30$, $K = 3$, $\alpha = .05$: $d_L = 1.21$ and $d_U = 1.65$

$\alpha = .01$: $d_L = 1.01$ and $d_U = 1.42$

Reject the null hypothesis based on the Durbin-Watson test at both the 5% and 1% levels.

Estimate of the autocorrelation coefficient: $r = 1 - \frac{d}{2} = 1 - \frac{.5}{2} = .75$

a. $H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = .80$, $n = 30$, $K = 3$, $\alpha = .05$: $d_L = 1.21$ and $d_U = 1.65$

$\alpha = .01$: $d_L = 1.01$ and $d_U = 1.42$

Reject the null hypothesis based on the Durbin-Watson test at both the 5% and 1% levels.

Estimate of the autocorrelation coefficient: $r = 1 - \frac{d}{2} = 1 - \frac{.8}{2} = .60$

b. $H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = 1.10$, $n = 30$, $K = 3$, $\alpha = .05$: $d_L = 1.21$ and $d_U = 1.65$

$\alpha = .01$: $d_L = 1.01$ and $d_U = 1.42$

Reject the null hypothesis based on the Durbin-Watson test at the 5% level. The test is inconclusive at the 1% level.

Estimate of the autocorrelation coefficient: $r = 1 - \frac{d}{2} = 1 - \frac{1.10}{2} = .45$

c. $H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = 1.25$, $n = 30$, $K = 3$, $\alpha = .05$: $d_L = 1.21$ and $d_U = 1.65$

$\alpha = .01$: $d_L = 1.01$ and $d_U = 1.42$

The test is inconclusive at both the 5% level and the 1% level.

d. $H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = 1.70$, $n = 30$, $K = 3$, $\alpha = .05$: $d_L = 1.21$ and $d_U = 1.65$

$\alpha = .01$: $d_L = 1.01$ and $d_U = 1.42$

Do not reject the null hypothesis at either the 5% level or the 1% level. There is insufficient evidence to suggest autocorrelation exists in the residuals.

13.34 a. $n = 30, \quad R^2 = .043$

$nR^2 = 1.29 < 2.71 = \chi^2_{1,.10}$, therefore, do not reject the H_0 that the error terms have constant variance at the 10% level.

b. $H_0 : \rho = 0, H_1 : \rho > 0, d = 1.29, n = 30, K = 4,$

$\alpha = .05 : d_L = 1.14$ and $d_U = 1.74$

$\alpha = .01 : d_L = .94$ and $d_U = 1.51$

The Durbin-Watson test gives inconclusive results at both the 5% and 1% levels.

13.36 The regression model includes the lagged value of the dependent variable as an independent variable.

Regression Analysis: Y_logCons versus X_LogDI, Y_laglogCons

```
The regression equation is
Y_logCons = 0.405 + 0.373 X_LogDI + 0.558 Y_laglogCons

28 cases used 1 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant       0.4049      0.1051       3.85    0.001
X_LogDI        0.3734      0.1075       3.47    0.002
Y_laglog       0.5577      0.1243       4.49    0.000

S = 0.03023      R-Sq = 99.6%      R-Sq(adj) = 99.6%
Analysis of Variance
Source            DF          SS          MS          F        P
Regression         2      6.1960      3.0980    3389.90    0.000
Residual Error    25      0.0228      0.0009
Total             27      6.2189

Source        DF       Seq SS
X_LogDI        1       6.1776
Y_laglog       1       0.0184

Unusual Observations
Obs     X_LogDI    Y_logCon          Fit       SE Fit     Residual     St Resid
  9        5.84     5.80814      5.72298      0.01074      0.08517
3.01R

Durbin-Watson statistic = 1.63
```

The multiple regression indicates that 99.6% of the variation in the private consumption can be explained by all of the independent variables. The *F*-test for the significance of the overall regression indicates the model has significant explanatory power with all the independent variables significantly different from zero. As expected, disposable income and the lagged variable of private consumption have an expected positive effect on the dependent variable.

In the presence of a lagged dependent variable used as an independent variable, the Durbin-Watson statistic is no longer valid. Instead, use of Durbin's h statistic is appropriate:

$$H_0: \rho = 0, H_1: \rho > 0$$

$$r = 1 - \frac{d}{2} = 1 - \frac{1.63}{2} = .185, \quad s_c^2 = (.1243)^2 = .01545$$

$$h = r\sqrt{\frac{n}{1 - n(s_c^2)}} = .185\sqrt{\frac{28}{1 - 28(.01545)}} = 1.30$$

$z_{.10} = 1.28$, therefore, reject H_0 at the 10% level but not at the 5% level

13.38 a.

Results for: Advertising Retail.xls
Regression Analysis: Retail Sales Y(t) versus Advertising X(t)

```
The regression equation is
Retail Sales Y(t) = 2269 + 28.5 Advertising X(t)

Predictor       Coef      SE Coef          T         P
Constant      2269.5        278.4       8.15     0.000
Advertis      28.504        2.087      13.66     0.000

S = 161.7        R-Sq = 90.3%      R-Sq(adj) = 89.8%
Analysis of Variance
Source              DF         SS          MS          F         P
Regression           1    4874833     4874833     186.52     0.000
Residual Error      20     522728       26136
Total               21    5397561
Durbin-Watson statistic = 1.12
```

b. $H_0: \rho = 0, H_1: \rho > 0$

$d = 1.12, n = 22, K = 1$

$\alpha = .05$: $d_L = 1.24$ and $d_U = 1.43$

$\alpha = .01$: $d_L = 1.00$ and $d_U = 1.17$

Reject H_0 at the 5% level, autocorrelation of the residuals exists at the 5% level, test is inconclusive at the 1% level.

c. The fitted regression is

$$y_t - .44y_{t-1} = 1370 + 27.3(x_t - .44x_{t-1})$$

13.40 a. Dummy Variables: Dummy variables are used whenever a factor is not readily quantifiable. For example, if we wished to determine the effect of trade barriers on output growth rates, we could include a dummy variable which takes the value of 1 when trade barriers are imposed and 0 otherwise. This could then be used to distinguish between different trade barrier levels.

b. Lagged dependent variables: Lagged dependent variables are useful when time-series data are analyzed. For example, one might wish to include lagged growth rates in a model used to explain fluctuations in output.

c. Logarithmic transformation: Logarithmic transformations allows inherently linear statistical techniques such as least squares linear regression to estimate non-linear functions. For example, cost functions where cost is some function of output produced is typically nonlinear. The log transformation allows us to express non-linear relationships in linear form and hence use linear estimation techniques to estimate the model.

13.42 The statement is not valid. The summation of several bivariate (simple) linear regressions does not equal the results obtained from a multiple regression. Therefore, while separating the independent variables may give some indication of the statistical significance of the individual effects, they will not provide any information about the influence on the dependent variable when the independent variables are taken together. It is preferable to acknowledge that the group as a whole is clearly influential but the data are not sufficiently informative to allow the disentangling, with any precision, of each independent variable's separate effects.

13.44 a. $H_0 : \beta_1 = 0, \quad H_1 : \beta_1 > 0 , \; t = \dfrac{2.11}{1.79} = 1.179$

Do not reject H_0 at the 10% level since $t < 1.292 \approx t_{84,.1}$

b. $H_0 : \beta_2 = 0, \quad H_1 : \beta_2 > 0 , \; t = \dfrac{.96}{1.94} = .495$

Do not reject H_0 at the 10% level since $t < 1.292 \approx t_{84,.1}$

c. The difference in results is likely due to the existence of multicollinearity between earnings per share (x_1) and funds flow per share (x_2)

13.46 Based on the *t*-ratios, none of the gender dummy variables has been found to be statistically different from zero. We have not found strong evidence to suggest that the gender of the student or the gender of the instructor, or gender taken together has a significant impact on total student score in the intermediate economics courses.

13.48 a. All else being equal, a 1% increase in value of new business orders leads to expected decrease of .82% in number of business failures.

b. $H_0 : \rho = 0, H_1 : \rho > 0 , d = .49, n = 30, K = 3, \alpha = .01$: $d_L = 1.01$ and $d_U = 1.42$

Reject H_0 at the 1% level

c. Given that the residuals are autocorrelated, the hypothesis test results for the influence of short-term interest rates on business failures will not be valid. The model must be reestimated taking into account the autocorrelated errors.

d. $r = 1 - \dfrac{.49}{2} = .755$

13.50 a. 95% CI: $.253 \pm 2.052(.106)$: $.035 < \beta < .471$

b. \$.253 increase in current period, further \$.138 increase next period, \$.075 increase two periods ahead, and so on. Total expected increase of \$.557.

c. $H_0: \rho = 0, H_1: \rho > 0$

Note that due to the presence of a lagged dependent variable used as an independent variable, Durbin's h statistic is relevant.

$$r = 1 - \frac{d}{2} = 1 - \frac{1.86}{2} = .07, \quad s_c^2 = (.134)^2 = .01796$$

$$h = r\sqrt{\frac{n}{1 - ns_c^2}} = .07\sqrt{\frac{29}{1 - 29(.01796)}} = .5446$$

$z_{.10} = 1.28$, therefore, do not reject H_0 at the 10% level

13.52

Regression Analysis: y_log versus x1_log, x2_log

```
The regression equation is
y_log = - 2.14 + 0.909 x1_log + 0.195 x2_log
Predictor        Coef      SE Coef        T       P
Constant      -2.1415       0.2000   -10.71   0.000
x1_log        0.90947      0.03518    25.85   0.000
x2_log        0.19451      0.07126     2.73   0.018

S = 0.07721     R-Sq = 99.6%     R-Sq(adj) = 99.5%
Analysis of Variance
Source            DF        SS        MS        F       P
Regression         2   16.7802    8.3901  1407.52   0.000
Residual Error    12    0.0715    0.0060
Total             14   16.8517

Source        DF     Seq SS
x1_log         1    16.7358
x2_log         1     0.0444
Durbin-Watson statistic = 1.67
```

The multiple regression indicates that 99.6% of the variation in the total government tax revenues other than from oil can be explained by all of the independent variables. The F-test for the significance of the overall regression indicates the model has significant explanatory power with all the independent variables significantly different from zero.

Test for autocorrelation:

$H_0: \rho = 0, H_1: \rho > 0$, $d = 1.67$, $n = 15$, $K = 2$, $\alpha = .05$: $d_L = .95$ and $d_U = 1.54$

Do not reject H_0 at the 5% level

13.54

Regression Analysis: y_log versus x1_log, x2_log, y_laglog_1

```
The regression equation is
y_log = 0.435 - 0.101 x1_log + 0.237 x2_log + 0.666 y_laglog_1
34 cases used 1 cases contain missing values

Predictor        Coef     SE Coef        T        P
Constant       0.4352      0.4360     1.00    0.326
x1_log       -0.10116     0.03822    -2.65    0.013
x2_log         0.2365      0.1017     2.32    0.027
y_laglog       0.6658      0.1174     5.67    0.000

S = 0.04039     R-Sq = 75.1%     R-Sq(adj) = 72.6%
Analysis of Variance
Source            DF         SS          MS        F        P
Regression         3   0.147751    0.049250    30.18    0.000
Residual Error    30   0.048952    0.001632
Total             33   0.196704

Source        DF     Seq SS
x1_log         1   0.001144
x2_log         1   0.094096
y_laglog       1   0.052511
Unusual Observations
Obs     x1_log      y_log         Fit      SE Fit    Residual    St Resid
20       0.410    4.58497     4.59677     0.02689    -0.01181       -0.39
X
35      -0.119    4.68398     4.59012     0.01650     0.09386       2.55R
R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.
Durbin-Watson statistic = 2.22
```

The multiple regression indicates that 75.1% of the variation in the quantity of imports can be explained by all of the independent variables. The *F*-test for the significance of the overall regression indicates the model has significant explanatory power with all the independent variables significantly different from zero.

Test for autocorrelation: Note that the lagged variable removes one observation from the original 35 observations.

$H_0 : \rho = 0, H_1 : \rho > 0$

Note that due to the presence of a lagged dependent variable used as an independent variable, Durbin's h statistic is relevant.

$$r = 1 - \frac{d}{2} = 1 - \frac{2.22}{2} = -.11, \quad s_c^2 = (.1174)^2 = .01378$$

$$h = r\sqrt{\frac{n}{1 - ns_c^2}} = -.11\sqrt{\frac{34}{1 - 34(.01378)}} = -.880$$

p-value = 2[1-*Fz*(.880))] = .3788, do not reject H_0 at any common level of alpha

13.55 76.6% of the variation in the FDIC examiner work hours can be explained by the variation in total assets of the bank, total number of offices, classified to total loan ratio, management rating, and if the examination was conducted jointly with the state.

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_7 = 0, \quad H_1 : at\ least\ one\ \beta_i \neq 0\ (i = 1,...,7)$$

$F = \frac{83}{7} \times \frac{.766}{1-.766} = 38.814$, $F_{7,83,.01} \approx 2.95$, therefore, reject H_0 at the 1% level

13.56

Regression Analysis: y_log versus x1_log, x2_log, x3_log

```
The regression equation is
y_log = 2.72 - 0.0252 x1_log + 0.315 x2_log + 0.379 x3_log

Predictor        Coef     SE Coef        T       P
Constant      2.71584     0.08821    30.79   0.000
x1_log       -0.02519     0.04049    -0.62   0.543
x2_log        0.31472     0.05689     5.53   0.000
x3_log         0.3788      0.2009     1.89   0.078

S = 0.03611      R-Sq = 91.7%      R-Sq(adj) = 90.2%
Analysis of Variance
Source          DF        SS         MS        F       P
Regression       3   0.231282   0.077094   59.14   0.000
Residual Error  16   0.020859   0.001304
Total           19   0.252140

Source       DF     Seq SS
x1_log        1   0.158727
x2_log        1   0.067919
x3_log        1   0.004636
Durbin-Watson statistic = 1.75
```

The multiple regression indicates that 91.7% of the variation in the days of sick leave per person can be explained by all of the independent variables. The *F*-test for the significance of the overall regression indicates the model has significant explanatory power. However, not all the independent variables are significantly different from zero. Unemployment rate is not statistically significant in explaining the dependent variable.

Test for autocorrelation:

$H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = 1.75$, $n = 20$, $K = 3$, $\alpha = .05$: $d_L = 1.00$ and $d_U = 1.68$

$\alpha = .01$: $d_L = .77$ and $d_U = 1.41$, do not reject H_0 at the 1% level or 5% level

13.58

a.

Regression Analysis: Services versus Disposable Personal Income

```
The regression equation is
Services = 172 + 0.592 Disposable Personal Income


Predictor                        Coef    SE Coef       T      P
Constant                       172.02      22.59    7.62  0.000
Disposable Personal Income   0.591698   0.003089  191.57  0.000

S = 66.7076   R-Sq = 99.7%   R-Sq(adj) = 99.7%

Analysis of Variance

Source           DF          SS         MS         F      P
Regression        1   163302353  163302353  36697.92  0.000
Residual Error  122      542889       4450
Total           123   163845242

Durbin-Watson statistic = 0.310542
```

Test for autocorrelation:
$H_0: \rho = 0, H_1: \rho > 0$, $d = 0.31$, $n = 124$, $K = 1$, $\alpha = .05$: $d_L = 1.65$ and $d_U = 1.69$
$\alpha = .01$: $d_L = 1.52$ and $d_U = 1.56$, reject H_0 at the 1% level and 5% level. There is autocorrelation.

b.

Regression Analysis: Services versus Disposable P, Total Consum, ...

```
The regression equation is
Services = 407 + 0.220 Disposable Personal Income
           + 0.191 Total Consumption (lagged 1 per - 8.51 Prime Rate


123 cases used, 1 cases contain missing values

Predictor                         Coef   SE Coef      T      P
Constant                        406.53     49.33   8.24  0.000
Disposable Personal Income     0.22007   0.05064   4.35  0.000
Total Consumption (lagged 1 per 0.19098  0.02632   7.26  0.000
Prime Rate                      -8.509     2.309  -3.69  0.000

S = 55.5800   R-Sq = 99.8%   R-Sq(adj) = 99.8%

Analysis of Variance

Source           DF          SS        MS         F      P
Regression        3   160361580  53453860  17303.84  0.000
Residual Error  119      367607      3089
Total           122   160729186

Durbin-Watson statistic = 0.121922
```

Test for autocorrelation:
$H_0 : \rho = 0, H_1 : \rho > 0, d = 0.12, n = 123, K = 3, \alpha = .05: d_L = 1.61$ and $d_U = 1.74$
$\alpha = .01: d_L = 1.48$ and $d_U = 1.60$, reject H_0 at the 1% level and 5% level. There is still autocorrelation. The new model reduces but does not eliminate autocorrelation.

13.60 a. Estimate with the statistically significant independent variables:

Regression Analysis: hseval versus sizehse, taxrate, totexp, Comper

```
The regression equation is
hseval = - 23.4 + 9.21 sizehse - 178 taxrate +0.000001 totexp - 20.4 Comper

Predictor          Coef       SE Coef          T        P
Constant        -23.433         8.986      -2.61    0.011
sizehse           9.210         1.564       5.89    0.000
taxrate         -177.53         39.87      -4.45    0.000
totexp       0.00000142    0.00000030       4.80    0.000
Comper          -20.370         6.199      -3.29    0.001

S = 3.400        R-Sq = 55.1%      R-Sq(adj) = 52.9%
Analysis of Variance
Source              DF          SS          MS          F        P
Regression           4     1203.84      300.96      26.03    0.000
Residual Error      85      982.79       11.56
Total               89     2186.63

Source        DF       Seq SS
sizehse        1       643.12
taxrate        1       244.06
totexp         1       191.82
Comper         1       124.84

Unusual Observations
Obs    sizehse      hseval         Fit      SE Fit    Residual    St Resid
 23       5.70      20.003      27.850       0.806      -7.847      -2.38R
 49       5.60      29.810      28.522       1.708       1.288       0.44 X
 50       5.60      30.061      28.178       1.687       1.883       0.64 X
 75       5.70      35.976      24.490       0.535      11.486       3.42R
 76       5.70      35.736      25.093       0.553      10.643       3.17R

R denotes an observation with a large standardized residual
X denotes an observation whose X value gives it large influence.
Durbin-Watson statistic = 1.20
```

Since all of the independent variables are statistically significant, leave all of the independent variables in the regression model.

b. The auxiliary regression is as follows:

Regression Analysis: ResiSq versus FITS1

```
The regression equation is
ResiSq = - 15.1 + 1.24 FITS1

Predictor        Coef     SE Coef         T        P
Constant       -15.09       11.96     -1.26    0.210
FITS1          1.2370      0.5604      2.21    0.030

S = 19.44        R-Sq = 5.2%       R-Sq(adj) = 4.2%
```

$e^2 = -15.1 + 1.24\hat{y}$, $n = 90$, $R^2 = .052$, $nR^2 = 4.68 > 3.84 = \chi^2_{1,.05}$; therefore, reject the null hypothesis that the error terms have constant variance at the 5% level and the economist is correct that heteroscedasticity is likely to be a problem.

c. Use population as the weighting variable. The constant term has been suppressed.

Regression Analysis: hseval/pop versus sizehse/pop, taxrate/pop, ...

```
The regression equation is
hseval/pop = 5.02 sizehse/pop - 191 taxrate/pop + 0.000002 totexp/pop
           - 25.6 comper/pop
Predictor          Coef      SE Coef       T       P
Noconstant
sizehse/pop      5.0208       0.2433   20.63   0.000
taxrate/pop     -191.06        33.12   -5.77   0.000
totexp/pop   0.00000166   0.00000068    2.45   0.016
comper/pop      -25.575        5.119   -5.00   0.000
S = 0.000354555

Analysis of Variance
Source           DF          SS          MS       F      P
Regression        4  0.00044572  0.00011143  886.42  0.000
Residual Error   86  0.00001081  0.00000013
Total            90  0.00045653
```

The partial regression slope coefficients in the weighted least squares model have smaller rates of response than the unweighted model. The coefficients all have the same sign and are significantly different from zero at the .02 level of significance. *F*-calc has increased from 26.03 up to 886.42.

13.62

a.

Regression Analysis: Durable Goods versus Disposable Personal Income

```
The regression equation is
Durable Goods = - 433 + 0.155 Disposable Personal Income

Predictor                       Coef   SE Coef       T      P
Constant                     -432.97     18.40  -23.53  0.000
Disposable Personal Income  0.154915  0.002626   59.00  0.000

S = 64.8051   R-Sq = 96.2%   R-Sq(adj) = 96.2%

Analysis of Variance

Source           DF        SS        MS        F      P
Regression        1  14618294  14618294  3480.79  0.000
Residual Error  138    579559      4200
Total           139  15197853

Durbin-Watson statistic = 0.0814542
```

Test for autocorrelation:

$H_0: \rho = 0$, $H_1: \rho > 0$, $d = 0.08$, $n = 140$, $K = 1$, $\alpha = .05$: $d_L = 1.65$ and $d_U = 1.69$
$\alpha = .01$: $d_L = 1.52$ and $d_U = 1.56$, reject H_0 at the 1% level and 5% level. There is autocorrelation.

b.

Regression Analysis: Durable Good versus Disposable P, Total consum, ...

```
The regression equation is
Durable Goods = 1844 + 0.109 Disposable Personal Income
               + 0.0699 Total consumption (lagged 1) + 0.269 Imports of
Goods
               - 0.0116 Population (midperiod, thousand - 1.75 Prime
Rate

139 cases used, 1 cases contain missing values
```

Predictor	Coef	SE Coef	T	P
Constant	1843.5	191.3	9.64	0.000
Disposable Personal Income	0.10880	0.03228	3.37	0.001
Total consumption (lagged 1)	0.06989	0.01601	4.37	0.000
Imports of Goods	0.26912	0.03255	8.27	0.000
Population (midperiod, thousand	-0.011598	0.001171	-9.90	0.000
Prime Rate	-1.7541	0.9136	-1.92	0.057

```
S = 26.4750   R-Sq = 99.4%   R-Sq(adj) = 99.4%

Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	5	14963287	2992657	4269.58	0.000
Residual Error	133	93223	701		
Total	138	15056510			

```
Durbin-Watson statistic = 0.549966
```

Test for autocorrelation:

$H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = 0.55$, $n = 139$, $K = 5$, $\alpha = .05$: $d_L = 1.57$ and $d_U = 1.78$ $\alpha = .01$: $d_L = 1.44$ and $d_U = 1.65$, reject H_0 at the 1% level and 5% level. There is autocorrelation. This regression model does not solve the problem of autocorrelation.

13.64

a.

Regression Analysis: Veal Retail versus U. S. Popula, Veal Product

```
The regression equation is
Veal Retail = 77.4 - 0.350 U. S. Population + 0.874 Veal Production
```

Predictor	Coef	SE Coef	T	P
Constant	77.43	26.31	2.94	0.004
U. S. Population	-0.34986	0.09056	-3.86	0.000
Veal Production	0.87403	0.01043	83.78	0.000

```
S = 19.4392   R-Sq = 99.8%   R-Sq(adj) = 99.8%

Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	2	13112014	6556007	17349.38	0.000
Residual Error	71	26830	378		
Total	73	13138844			

```
Durbin-Watson statistic = 0.294336
```

Test for autocorrelation:

$H_0 : \rho = 0, H_1 : \rho > 0$, $d = 0.29$, $n = 74$, $K = 2$, $\alpha = .05$: $d_L = 1.57$ and $d_U = 1.68$

$\alpha = .01$: $d_L = 1.42$ and $d_U = 1.53$, reject H_0 at the 1% level and 5% level. There is autocorrelation.

Estimate : $r = 1 - \frac{.29}{2} = .855$

Adjusted for serial correlation:

Regression Analysis: Veal Retail versus U. S. Popula, Veal Product

```
The regression equation is
Veal Retail adj = 4.63 - 0.155 U. S. Population adj_1
                  + 0.871 Veal Production adj

73 cases used, 1 cases contain missing values

Predictor                    Coef   SE Coef      T      P
Constant                    4.632     5.867   0.79  0.433
U. S. Population adj_1    -0.1551    0.1618  -0.96  0.341
Veal Production adj      0.870698  0.009294  93.68  0.000

S = 9.66582   R-Sq = 99.4%   R-Sq(adj) = 99.4%

Analysis of Variance

Source          DF       SS      MS        F      P
Regression       2  1075905  537952  5757.93  0.000
Residual Error  70     6540      93
Total           72  1082445

Durbin-Watson statistic = 1.56978
```

b.

Regression Analysis: Veal Retail versus U. S. Popula, Veal Product

```
The regression equation is
Veal Retail = - 11.5 + 0.0009 U. S. Population + 0.879 Veal Production

Predictor            Coef  SE Coef      T      P
Constant           -11.48    28.81  -0.40  0.693
U. S. Population  0.00087  0.08827   0.01  0.992
Veal Production   0.87922  0.01843  47.70  0.000

S = 4.21761   R-Sq = 99.8%   R-Sq(adj) = 99.8%

Analysis of Variance

Source          DF      SS      MS        F      P
Regression       2  292599  146299  8224.51  0.000
Residual Error  26     462      18
Total           28  293061

Durbin-Watson statistic = 0.563839
```

Test for autocorrelation:

$H_0 : \rho = 0, H_1 : \rho > 0$, $d = 0.56$, $n = 29$, $K = 2$, $\alpha = .05$: $d_L = 1.27$ and $d_U = 1.56$

$\alpha = .01$: $d_L = 1.05$ and $d_U = 1.33$, reject H_0 at the 1% level and 5% level. There is autocorrelation.

$$\text{Estimate : } r = 1 - \frac{.56}{2} = .72$$

Adjusted for serial correlation:

Regression Analysis: Veal Retail versus U. S. Popula, Veal Product

```
The regression equation is
Veal Retail adj = - 8.9 + 0.075 U. S. Population adj + 0.877 Veal
Production adj


28 cases used, 1 cases contain missing values


Predictor                 Coef   SE Coef       T       P
Constant                 -8.94     10.67   -0.84   0.410
U. S. Population adj    0.0749    0.1237    0.61   0.550
Veal Production adj    0.87743   0.01886   46.53   0.000


S = 2.82324   R-Sq = 99.5%   R-Sq(adj) = 99.5%


Analysis of Variance

Source          DF     SS     MS        F      P
Regression       2  40557  20279  2544.16  0.000
Residual Error  25    199      8
Total           27  40757


Durbin-Watson statistic = 1.56227
```

c.

The first regression has a Durbin-Watson d statistic of 1.56978, indicating slight negative autocorrelation.
The second regression for which time includes only data beginning in the year 1980 has a Durbin-Watson d statistic of 1.56227. This also indicates that there is a slight negative autocorrelation.

13.66

Correlation matrix:

Correlations: Physician an, National Hea, Medicare, Hospital Car, ...

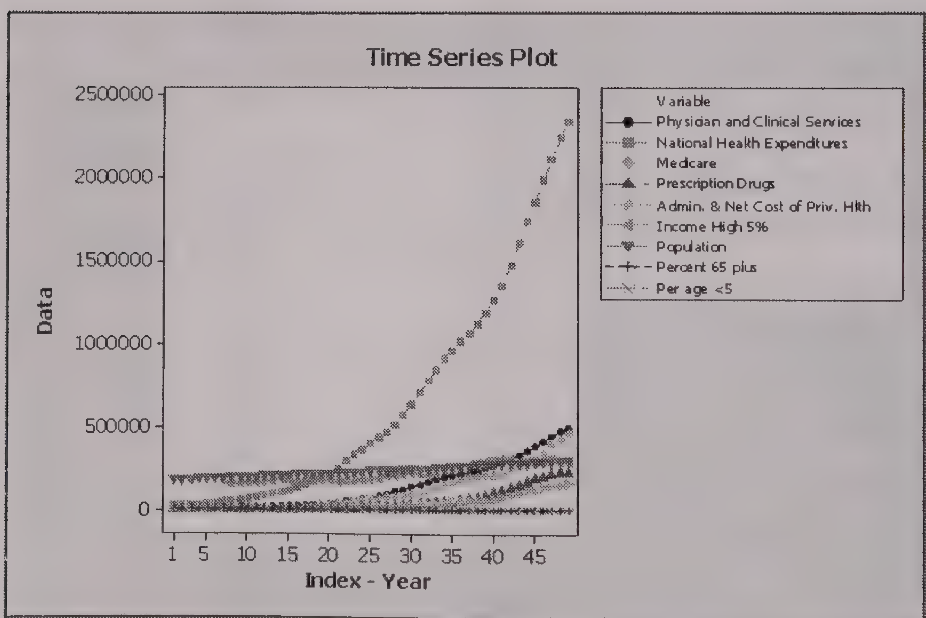
	Physician and Cl	National Health	Medicare
National Health	1.000		
	0.000		
Medicare	0.994	0.995	
	0.000	0.000	
Hospital Care	0.998	0.998	0.992
	0.000	0.000	0.000
Prescription Dru	0.975	0.978	0.977
	0.000	0.000	0.000
Admin. & Net Cos	0.990	0.992	0.990
	0.000	0.000	0.000
Income Low 5th	0.762	0.759	0.734
	0.000	0.000	0.000
Income Median	0.883	0.880	0.854
	0.000	0.000	0.000
Income High 5th	0.945	0.941	0.919
	0.000	0.000	0.000
Income High 5%	0.939	0.935	0.914
	0.000	0.000	0.000
Population	0.953	0.952	0.936
	0.000	0.000	0.000
Unemployment	-0.208	-0.200	-0.208
	0.151	0.168	0.151
Percent 65 plus	0.778	0.773	0.751
	0.000	0.000	0.000
Per age <5	-0.601	-0.606	-0.587
	0.000	0.000	0.000

	Hospital Care	Prescription Dru	Admin. & Net Cos
Prescription Dru	0.964		
	0.000		
Admin. & Net Cos	0.984	0.993	
	0.000	0.000	
Income Low 5th	0.763	0.696	0.708
	0.000	0.000	0.000
Income Median	0.879	0.833	0.840
	0.000	0.000	0.000
Income High 5th	0.944	0.882	0.902
	0.000	0.000	0.000
Income High 5%	0.935	0.880	0.897
	0.000	0.000	0.000
Population	0.966	0.884	0.917
	0.000	0.000	0.000
Unemployment	-0.162	-0.256	-0.212
	0.267	0.076	0.143
Percent 65 plus	0.811	0.639	0.706
	0.000	0.000	0.000
Per age <5	-0.633	-0.526	-0.563
	0.000	0.000	0.000

	Income Low 5th	Income Median	Income High 5th
Income Median	0.949		
	0.000		
Income High 5th	0.892	0.968	
	0.000	0.000	
Income High 5%	0.872	0.953	0.995
	0.000	0.000	0.000
Population	0.839	0.921	0.977
	0.000	0.000	0.000
Unemployment	-0.292	-0.398	-0.334
	0.061	0.009	0.031
Percent 65 plus	0.766	0.776	0.829
	0.000	0.000	0.000
Per age <5	-0.838	-0.765	-0.740
	0.000	0.000	0.000

	Income High 5%	Population	Unemployment
Population	0.962		
	0.000		
Unemployment	-0.379	-0.071	
	0.013	0.627	
Percent 65 plus	0.792	0.914	0.150
	0.000	0.000	0.302
Per age <5	-0.700	-0.788	-0.182
	0.000	0.000	0.210

	Percent 65 plus
Per age <5	-0.817
	0.000



Regression Analysis: Physician an versus National Hea, Medicare, ...

```
The regression equation is
Physician and Clinical Services = 317228 + 0.386 National Health
Expenditures
                                  - 0.228 Medicare - 0.281 Prescription
Drugs
                                  - 0.529 Admin. & Net Cost of Priv. Hlth
                                  + 0.138 Income High 5% - 1.93
Population
                                  + 9164 Percent 65 plus - 5134 Per age
<5


42 cases used, 7 cases contain missing values


Predictor                             Coef   SE Coef       T       P
Constant                            317228     37668    8.42   0.000
National Health Expenditures       0.38618   0.01546   24.98   0.000
Medicare                          -0.22760   0.02487   -9.15   0.000
Prescription Drugs                -0.28050   0.06393   -4.39   0.000
Admin. & Net Cost of Priv. Hlth   -0.52851   0.08505   -6.21   0.000
Income High 5%                     0.13794   0.02394    5.76   0.000
Population                         -1.9325    0.1866  -10.36   0.000
Percent 65 plus                       9164      1715    5.34   0.000
Per age <5                           -5134      1224   -4.19   0.000


S = 1432.46   R-Sq = 100.0%   R-Sq(adj) = 100.0%


Analysis of Variance

Source          DF           SS            MS         F      P
Regression       8  8.82972E+11  1.10372E+11  53788.84  0.000
Residual Error  33     67714065      2051941
Total           41  8.83040E+11


Durbin-Watson statistic = 1.40357
```

Test for autocorrelation:

$H_0 : \rho = 0$, $H_1 : \rho > 0$, $d = 1.40$, $n = 42$, $K = 8$, $\alpha = .05$: $d_L = 1.23$ and $d_U = 1.79$

$\alpha = .01$: $d_L = 1.05$ and $d_U = 1.58$, test is inconclusive at the 1% level and 5% level.

13.68

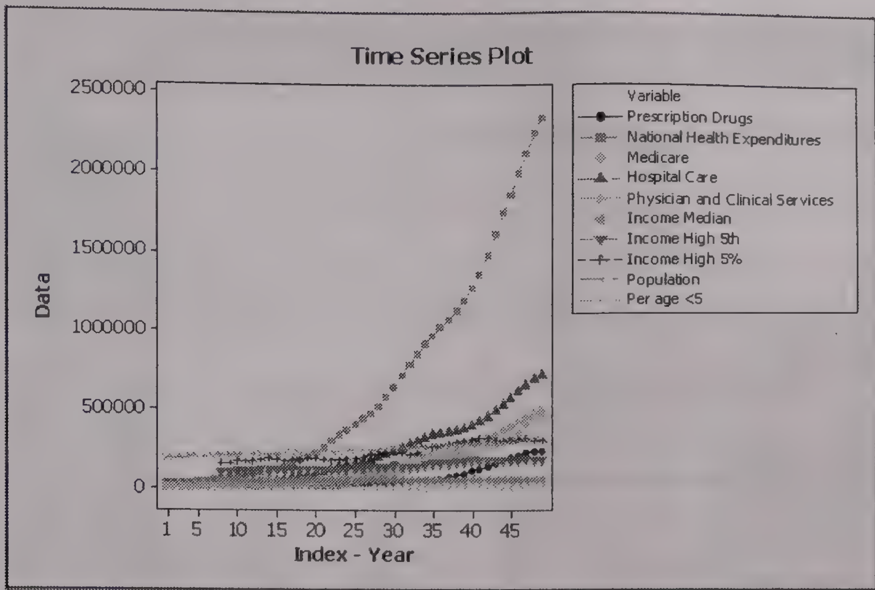
Correlation matrix:

Correlations: Prescription, National Hea, Medicare, Hospital Car, ...

	Prescription Dru	National Health	Medicare
National Health	0.978		
	0.000		
Medicare	0.977	0.995	
	0.000	0.000	
Hospital Care	0.964	0.998	0.992
	0.000	0.000	0.000
Physician and Cl	0.975	1.000	0.994
	0.000	0.000	0.000
Income Median	0.833	0.880	0.854
	0.000	0.000	0.000
Income High 5th	0.882	0.941	0.919
	0.000	0.000	0.000
Income High 5%	0.880	0.935	0.914
	0.000	0.000	0.000
Population	0.884	0.952	0.936
	0.000	0.000	0.000
Per age <5	-0.526	-0.606	-0.587
	0.000	0.000	0.000

	Hospital Care	Physician and Cl	Income Median
Physician and Cl	0.998		
	0.000		
Income Median	0.879	0.883	
	0.000	0.000	
Income High 5th	0.944	0.945	0.968
	0.000	0.000	0.000
Income High 5%	0.935	0.939	0.953
	0.000	0.000	0.000
Population	0.966	0.953	0.921
	0.000	0.000	0.000
Per age <5	-0.633	-0.601	-0.765
	0.000	0.000	0.000

	Income High 5th	Income High 5%	Population
Income High 5%	0.995		
	0.000		
Population	0.977	0.962	
	0.000	0.000	
Per age <5	-0.740	-0.700	-0.788
	0.000	0.000	0.000



Regression Analysis: Prescription versus National Hea, Medicare, ...

```
The regression equation is
Prescription Drugs = 430348 + 0.493 National Health Expenditures
                     - 0.174 Medicare - 0.461 Hospital Care
                     - 0.735 Physician and Clinical Services
                     - 3.27 Income Median + 3.02 Income High 5th
                     - 0.950 Income High 5% - 1.94 Population - 7848 Per age <5

42 cases used, 7 cases contain missing values
```

Predictor	Coef	SE Coef	T	P
Constant	430348	67025	6.42	0.000
National Health Expenditures	0.49320	0.02425	20.34	0.000
Medicare	-0.17437	0.04861	-3.59	0.001
Hospital Care	-0.4613	0.1233	-3.74	0.001
Physician and Clinical Services	-0.7348	0.1655	-4.44	0.000
Income Median	-3.270	1.545	-2.12	0.042
Income High 5th	3.0205	0.9432	3.20	0.003
Income High 5%	-0.9498	0.2791	-3.40	0.002
Population	-1.9379	0.2787	-6.95	0.000
Per age <5	-7848	1857	-4.23	0.000

```
S = 2343.51   R-Sq = 99.9%   R-Sq(adj) = 99.9%

Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	9	2.06652E+11	22961282104	4180.82	0.000
Residual Error	32	175745870	5492058		
Total	41	2.06827E+11			

```
Durbin-Watson statistic = 1.65580
```

Test for autocorrelation:
$H_0: \rho = 0, H_1: \rho > 0$, $d = 1.66$, $n = 42$, $K = 9$, $\alpha = .05$: $d_L = 1.23$ and $d_U = 1.79$
$\alpha = .01$: $d_L = 1.05$ and $d_U = 1.58$, test is inconclusive at the 5% level. H_0 is not rejected at the 5% level.

Chapter 14:

Analysis of Categorical Data

14.2 H_0: Business executives preferences are evenly distributed across 4 airlines.

H_1: otherwise

Airline	Delta	United	Southwest	U.S Airways	Total
Observed Number	60	50	48	42	200
Probability (H_0)	0.25	0.25	0.25	0.25	1
Expected Number	50	50	50	50	200
Chi-square calculation	2	0	0.08	1.28	3.36

Chi-square calculation: $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 3.36$

$\chi^2_{(3,.10)} = 6.25$; Therefore, fail to reject H_0 at the 10% level

14.4 H_0: Quality of the output conforms to the usual pattern.

H_1: otherwise

Electronic component	No faults	1 fault	>1 fault	Total
Observed Number	458	30	12	500
Probability (H_0)	0.93	0.05	0.02	1
Expected Number	465	25	10	500
Chi-square calculation	0.105376344	1	0.4	1.505376

Chi-square calculation: $\chi^2 = \sum \frac{(O_i - E_i)^2}{Ei} = 1.505$

$\chi^2_{(2,.05)} = 5.99$; Therefore, do not reject H_0 at the 5% level

14.6 Student opinion of business courses is the same as that for all courses.

H_1: otherwise

Opinion	Very useful	Somewhat	Worthless	Total
Observed Number	68	18	14	100
Probability (H_0)	0.6	0.2	0.2	1
Expected Number	60	20	20	100
Chi-square calculation	1.066666667	0.2	1.8	3.066667

Chi-square calculation: $\chi^2 = \sum \frac{(O_i - E_i)^2}{Ei} = 3.067$

$\chi^2_{(2,.10)} = 4.61$; Therefore, do not reject H_0 at the 10% level

14.8 Consumer preferences for soft drinks are equally spread across 5 soft drinks.
H_1: otherwise

Drink	A	B	C	D	E	Total
Observed Number	20	25	28	15	27	115
Probability (H_0)	0.2	0.2	0.2	0.2	0.2	1
Expected Number	23	23	23	23	23	115
Chi-square calculation	0.391304	0.173913	1.086957	2.782609	0.695652	5.130435

Chi-square calculation: $\chi^2 = \sum \frac{(O_i - E_i)^2}{Ei} = 5.130$

$\chi^2_{(4,.10)} = 7.78$; Therefore, do not reject H_0 at the 10% level

14.10 statistics professors preferences for software packages are equally divided across 4 packages.
H_1: otherwise

Software	M	E	S	P	Total
Observed Number	100	80	35	35	250
Probability (H_0)	0.25	0.25	0.25	0.25	1
Expected Number	62.5	62.5	62.5	62.5	250
Chi-square calculation	22.5	4.9	12.1	12.1	51.6

Chi-square calculation: $\chi^2 = \sum \frac{(O_i - E_i)^2}{Ei} = 51.6$

$\chi^2_{(3,.005)} = 12.84$; Therefore, reject H_0 at the .5% level

14.12 H_0: population distribution of arrivals per minute is Poisson.
H_1: otherwise

Arrivals	0	1	2	3	4+	Total
Observed Number	10	26	35	24	5	100
Probability (H_0)	0.1496	0.2842	0.27	0.171	0.1252	1
Expected Number	14.96	28.42	27	17.1	12.52	100
Chi-square calculation	1.644492	0.206066	2.37037	2.784211	4.516805	11.52194

Chi-square calculation: $\chi^2 = \sum \frac{(O_i - E_i)^2}{Ei} = 11.52$

$\chi^2_{(3,.01)} = 11.34$, $\chi^2_{(3,.005)} = 12.84$. Reject H_0 at the 1% level but not at the .5% level

14.14 H_0: resistance of electronic components is normally distributed.
H_1: otherwise

$$JB = 100\left[\frac{(.63)^2}{6} + \frac{(3.85-3)^2}{24}\right] = 9.625$$

From Table 14.9 – Significance points of the Jarque-Bera statistic; 5% point (n=100) is 4.29. Therefore, reject H_0 at the 5% level

14.16 H_0: monthly balances for credit card holders of a particular card are normally distributed.
H_1: otherwise

$$JB = 125\left[\frac{(.55)^2}{6} + \frac{(2.77-3)^2}{24}\right] = 6.578$$

From Table 14.9 – Significance points of the Jarque-Bera statistic; 5% point (n = 125) is 4.34. Therefore, reject H_0 at the 5% level

14.18 a. H_0: No association exists between GPA and major.
H_1: otherwise

Chi-Square Test: GPA<3, GPA3+

Expected counts are printed below observed counts

	GPA<3	GPA3+	Total
1	50	35	85
	46.75	38.25	
2	45	30	75
	41.25	33.75	
3	15	25	40
	22.00	18.00	
Total	110	90	200

Chi-Sq = 0.226 + 0.276 + 0.341 + 0.417 + 2.227 + 2.722 = 6.209
DF = 2, P-Value = 0.045

$\chi^2_{(2,.05)} = 5.99$; Therefore, reject H_0 of no association at the 5% level

14.20 Complete the contingency table:

	Method of learning about product		
Age	Friend	Ad	col. total
<21	30	20	50
21-35	60	30	90
35+	18	42	60
row total	108	92	200

H_0: No association exists between the method of learning about the product and the age of the respondent.
H_1: otherwise

Chi-Square Test: Friend, Ad

Expected counts are printed below observed counts

	Friend	Ad	Total
1	30	20	50
	27.00	23.00	
2	60	30	90
	48.60	41.40	
3	18	42	60
	32.40	27.60	
Total	108	92	200

Chi-Sq = 0.333 + 0.391 +2.674 + 3.139 +6.400 + 7.513 = 20.451
DF = 2, P-Value = 0.000

$\chi^2_{(2,.005)} = 10.6$ Therefore, reject H_0 of no association at the .5% level

14.22 $H_0 : P = 0.50$ (there is no preference for one stock over the other)
$H_1 : P \neq 0.50$ (otherwise)
n = 11. For stock 2 and a two-sided test, P(2 ≥ X ≥ 9) = 2P(X ≤ 1) =
2[.0005 + .0054] = .0118
Therefore, reject H_0 at levels of alpha in excess of 1.18%

14.24 $H_0 : P = 0.50$ (grocery store managers are equally divided about customers attitudes about electronic coupons)
$H_1 : P \neq 0.50$ (otherwise)
n = 11. For 8 "yes" answers and a two-sided test,
P(4 ≤ X ≤ 7) = 2P(X ≤ 3) = 2[.0005 + .0054 + .0269 + .0806] = .2268
Therefore, reject H_0 at levels of alpha in excess of 22.68%

14.26 $H_0 : P = 0.50$ (voters are evenly divided)
$H_1 : P \neq 0.50$ (otherwise)
n = 130 – 18 = 112. $\hat{p} = 68/112 = .6071$
$\mu = nP = 112(.5) = 56 \quad \sigma = .5\sqrt{n} = .5\sqrt{112} = 5.2915$
$Z = \frac{S^* - \mu}{\sigma} = \frac{67.5 - 56}{5.2915} = 2.17$, p-value = 2[1 – F_z(2.17)] = 2[1 - .9850] = .030
Therefore, reject H_0 at levels of alpha in excess of .30%

14.28 $H_0 : P = 0.50$ (there is no preference for one energy drink over the other)
$H_1 : P \neq 0.50$ (otherwise)
n = 120 – 2 = 118. $\hat{p} = 65/118 = .5508$
$\mu = nP = 118(.5) = 59 \quad \sigma = .5\sqrt{n} = .5\sqrt{118} = 5.4314$
$Z = \frac{S^* - \mu}{\sigma} = \frac{64.5 - 59}{5.4314} = 1.01$, p-value = 2[1 – F_z(1.01)] = 2[1 - .8444] = .3112
Therefore, reject H_0 at levels of alpha in excess of 31.12%

14.30 H_0: origin of the student has no effect on GPA

H_1: otherwise

Wilcoxon Signed Rank Test:

Pair	In State	Out of State	Difer	Mod	Rank	Negetive	Positive
A	3.4	2.8	0.6	0.6	7		7
B	3	3.1	-0.1	0.1	1	1	
C	2.4	2.7	-0.3	0.3	3	3	
D	3.8	3.3	0.5	0.5	5.5		5.5
E	3.9	3.7	0.2	0.2	2		2
F	2.3	2.8	-0.5	0.5	5.5	5.5	
G	2.6	2.6	0	0			
H	3.7	3.3	0.4	0.4	4		4
						9.5	18.5

Wilcoxon signed rank statistic T is min(T_+, T_-) = min(9.5, 18.5) = 9.5

$n = 7$, $T = 9.5$, $T_{.10} = 6$. Do not reject H_0 at any common level of alpha

14.32 H_0: Total quality management has no impact on job satisfaction of employees

H_1: Total quality management has a positive impact on job satisfaction

$T = 169$, $\mu_T = 232.5$, $\sigma^2_T = 30(31)(61)/24 = 2363.75$

$z = \frac{169 - 232.5}{\sqrt{2363.75}} = -1.31$, p-value = 1-$F_Z$(1.31) = 1- .9049 = .0951

Therefore, reject H_0 at levels in excess of 9.51%

14.34 H_0: directors' ownership is the same for firms with and without an audit committee

H_1: directors' ownership is higher for firms without an audit committee

$z = \frac{U - \mu_U}{\sigma_U} = 2.12$, p-value = 1-$F_Z$(2.12) = 1- .9830 = .0170

Therefore, reject H_0 at levels in excess of 1.70%

14.36 H_0: starting salaries are equal for both types of schools

H_1: otherwise (two-tailed)

Sum of ranks for technical college = 153.5

$R_1 = 153.5, n_1 = 12, n_2 = 10 \qquad U = 12(10) + 12(13)/2 - 153.5 = 44.5$

$\mu_U = 120/2 = 60, \sigma^2_U = 120(23)/12 = 230$

$z = \frac{44.5 - 60}{\sqrt{230}} = -1.02$, p-value = 2[1-$F_Z$(1.02)] = 2[1-.8461] = .3078

Therefore, reject H_0 at levels in excess of 30.78%

14.38 H_0: no difference in rankings by gender

H_1: otherwise

Sum of ranks for males = 242

$T = 242, n_1 = 15, n_2 = 15,\ E(T) = \mu_T = n_1(n_1 + n_2 + 1)/2 = 232.5$

$Var(T) = \sigma_T^2 = n_1 n_2(n_1 + n_2 + 1)/12 = 581.25,\ z = \dfrac{242 - 232.5}{\sqrt{581.25}} = .394$

p-value = 2[1-F_Z(.39)] = 2[1- .6517] = .6966

Therefore, reject H_0 at levels in excess of 69.66%

14.40 H_0: no difference between faculty and students for salary of the football coach

H_1: students would propose a higher salary than would faculty (one-tailed)

Sum of ranks of faculty = 2024

$T = 2042, n_1 = 50, n_2 = 50,\ E(T) = \mu_T = n_1(n_1 + n_2 + 1)/2 = 2525$

$Var(T) = \sigma_T^2 = n_1 n_2(n_1 + n_2 + 1)/12 = 21041.667,\ z = \dfrac{2024 - 2525}{\sqrt{21041.6667}} = -3.454$

p-value = [1-F_Z(3.45)] = [1- .9997] = .0003

Therefore, reject H_0 at levels in excess of 0.03%

14.42 a. Obtain rankings of the two variables

RankExam	RankProject
6	5.5
1	3.0
4	2.0
5	5.5
10	9.0
2	1.0
3	8.0
7	4.0
9	10.0
8	7.0

Therefore, pearson correlation between rankings of variables is the spearman rank correlation coefficient:

Correlations: RankExam, RankProject

Pearson correlation of RankExam and RankProject = 0.717

b. H_0: no association between scores on the exam vs. on the project

H_1: an association exists (two-tailed test)

$n = 10, r_{s.025} = .648, r_{s.010} = .745$. Therefore, reject H_0 of no association between the two variables at the .05 level but not at .02 level (two-tailed test)

14.44 A time series contains 16 observations, the probability that the number of runs

a. is at most 5

Since $n \leq 20$, using Appendix Table 14

P(the number of runs is at most 5) = 0.032

b. exceeds 12

Since $n \leq 20$, using Appendix Table 14,

P(the number of runs exceeds 12) = $1 - P(R \leq 12) = 1 - 0.968 =$ 0.032

14.46 A time series contains 50 observations, the probability that the number of runs

a. is no more than 14 (less than 15)

$$Z = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}} = Z = \frac{15 - \frac{50}{2} - 1}{\sqrt{\frac{50^2 - 2(50)}{4(50-1)}}} = -3.14, \ P(Z < -3.14) = .0008$$

b. is fewer than 16

$$Z = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}} = Z = \frac{16 - \frac{50}{2} - 1}{\sqrt{\frac{50^2 - 2(50)}{4(50-1)}}} = -2.86, \ P(Z < -2.86) = .0021$$

c. is greater than 28

$$Z = \frac{R - \frac{n}{2} - 1}{\sqrt{\frac{n^2 - 2n}{4(n-1)}}} = Z = \frac{28 - \frac{50}{2} - 1}{\sqrt{\frac{50^2 - 2(50)}{4(50-1)}}} = .57$$

$P(Z > .57) = 1 - .7157 = 0.2843$.

14.48 Runs test on Value of the exchange rate

Runs Test: Value

```
Value    K = 99.8500
The observed number of runs = 6
The expected number of runs = 7.0000
6 Observations above K, 6 below
* N Small -- The following approximation may be invalid
     The test is significant at  0.5448
     Cannot reject at alpha = 0.05
```

Do not reject Ho that the data series is random. There is no evidence of nonrandom patterns in the data.

14.50 Runs test on Stock Market Return

Runs Test: Return

```
Return   K =     17.5000
The observed number of runs =    9
The expected number of runs =    8.0000
7 Observations above K    7 below
* N Small -- The following approximation may be invalid
    The test is significant at  0.5780
```

Do not reject Ho that the data series is random. There is no evidence of nonrandom patterns in the data.

14.52 H_0: No association exists between write-downs of assets and merger activity

H_1: otherwise

Chi-Square Test: Yes, No

Expected counts are printed below observed counts

	Yes	No	Total
1	32	48	80
	28.15	51.85	
2	25	57	82
	28.85	53.15	
Total	57	105	162

Chi-Sq = 0.527 + 0.286 + 0.514 + 0.279 = 1.607
DF = 1, P-Value = 0.205

$\chi^2_{(1,.10)} = 2.71$; Therefore, do not reject H_0 at the 10% level

14.54 H_0: No association exists between personnel rating and college major

H_1: otherwise

Chi-Square Test: Excellent, Strong, Average

Expected counts are printed below observed counts

	Excellent	Strong	Average	Total
1	21	18	10	49
	19.11	18.42	11.47	
2	19	15	5	39
	15.21	14.66	9.13	
3	10	5	5	20
	7.80	7.52	4.68	
4	5	15	13	33
	12.87	12.40	7.72	
Total	55	53	33	141

Chi-Sq = 0.186 + 0.010 + 0.188 + 0.943 + 0.008 + 1.867 + 0.620 + 0.843 + 0.022 + 4.814 + 0.543 + 3.605 = 13.648
DF = 6, P-Value = 0.034
1 cells with expected counts less than 5.0

$\chi^2_{(6,.05)} = 12.59$; Therefore, reject H_0 at the 5% level

14.56 H_0: No association exists between graduate studies and college major

H_1: otherwise

Chi-Square Test: Business, Law, Theology

Expected counts are printed below observed counts

	Business	Law	Theology	Total
1	30	20	10	60
	18.00	27.00	15.00	
2	6	34	20	60
	18.00	27.00	15.00	
Total	36	54	30	120

Chi-Sq = 8.000 + 1.815 + 1.667 + 8.000 + 1.815 + 1.667 = 22.963

DF = 2, P-Value = 0.000

$\chi^2_{(2,.005)} = 10.60$; Therefore, reject H_0 at the .5% level

14.58 H_0: No association exists between primary election candidate preferences and voting district

H_1: otherwise

Chi-Square Test: A, B, C, D

Expected counts are printed below observed counts

	A	B	C	D	Total
1	52	34	80	34	200
	46.46	31.69	92.00	29.85	
2	33	15	78	24	150
	34.85	23.77	69.00	22.38	
3	66	54	141	39	300
	69.69	47.54	138.00	44.77	
Total	151	103	299	97	650

Chi-Sq = 0.660 + 0.168 + 1.565 + 0.578 + 0.098 + 3.235 + 1.174 + 0.117 + 0.196 + 0.878 + 0.065 + 0.743 = 9.478

DF = 6, P-Value = 0.148

$\chi^2_{(6,.10)} = 10.64$; Therefore, do not reject H_0 at the 10% level

14.60 H_0: No association exists between years of experience and parts produced per hour

H_1: otherwise

Chi-Square Test: Subgroup1, Subgroup2, Subgroup3

Expected counts are printed below observed counts

	Subgroup1	Subgroup2	Subgroup3	Total
1	10	30	10	50
	10.00	20.00	20.00	
2	10	20	20	50
	10.00	20.00	20.00	
3	10	10	30	50
	10.00	20.00	20.00	
Total	30	60	60	150

Chi-Sq = 0.000 + 5.000 + 5.000 + 0.000 + 0.000 + 0.000 + 0.000 + 5.000 + 5.000 = 20.000

DF = 4, P-Value = 0.000

$\chi^2_{(4,.005)} = 14.86$; Therefore, reject H_0 at the .5% level

14.62 a. H_0: No association exists between package weight and package source

H_1: otherwise

Chi-Square Test: <3lb, 4-10lb, 11-75lb

Expected counts are printed below observed counts

	<3 lb	4-10lb	11-75lb	Total
1	40	40	20	100
	37.85	36.62	25.54	
2	119	63	18	200
	75.69	73.23	51.08	
3	18	71	111	200
	75.69	73.23	51.08	
4	69	64	17	150
	56.77	54.92	38.31	
Total	246	238	166	650

Chi-Sq = 0.123 + 0.313 + 1.201 + 24.779 + 1.429 + 21.420 + 43.973 + 0.068 + 70.301 + 2.635 + 1.500 + 11.852 = 179.594
DF = 6, P-Value = 0.000

$\chi^2_{(6,.005)} = 18.55$; Therefore, reject H_0 at the .5% level

b. The combinations with the largest percentage gap between observed and expected frequencies are 1) between factories and 11-75 pound packages, and 2) between factories and under 3 pound packages.

14.64 H_0: No association exists between the respondents' opinion and age regarding the bail out of the automobile industry

H_1: otherwise

Chi-Square Test: Opposed, Favor

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

      Opposed  Favor  Total
  1        60     20     80
        61.47  18.53
        0.035  0.116

  2       132     22    154
       118.33  35.67
        1.580  5.240

  3        80     40    120
        92.20  27.80
        1.615  5.358

Total     272     82    354

Chi-Sq = 13.944, DF = 2, P-Value = 0.001
```

$\chi^2_{(2,.1)} = 4.61$; Therefore, reject H_0 at the 10% level

14.66 H_0: No association exists between reason for moving to Florida and industry type

H_1: otherwise

Chi-Square Test: Manufacturing, Retail, Tourism

Expected counts are printed below observed counts

	Manufacturing	Retail	Tourism	Total
1	53	25	10	88
	42.04	28.31	17.66	
2	67	36	20	123
	58.76	39.56	24.68	
3	30	40	33	103
	49.20	33.13	20.67	
Total	150	101	63	314

Chi-Sq = 2.858 + 0.386 + 3.320 + 1.156 + 0.321 + 0.887 + 7.495 + 1.424 + 7.362 = 25.210

DF = 4, P-Value = 0.000

$\chi^2_{(4,.005)} = 14.86$; Therefore, reject H_0 at the .5% level

14.68 H_0: No association exists between opinions on stricter advertising controls of weight loss products and usage of quick weight reduction product

H_1: otherwise

Chi-Square Test: Yes, No

Expected counts are printed below observed counts

	Yes	No	Total
1	85	40	125
	64.25	60.75	
2	25	64	89
	45.75	43.25	
Total	110	104	214

Chi-Sq = 6.700 + 7.086 + 9.410 + 9.952 = 33.148

DF = 1, P-Value = 0.000

$\chi^2_{(1,.005)} = 7.88$; Therefore, reject H_0 at the .5% level

14.70 H_0: No difference in current and past customer preferences

H_1: otherwise

	A	B	C	D
observed frequency	56	70	28	126
expected frequency	56	92.4	56	75.6
$(O_i - E_i)^2 / E_i$	0	5.43	14	33.6

Chi-square Test Statistic = 53.03

$\chi^2_{(3,.005)} = 12.84$; Therefore, reject H_0 at the .5% level

14.72 Nonparametric tests make no assumption about the behavior of the population distribution. The advantages of the tests are less restrictive assumptions, easily calculated tests that can be used on nominal or ordinal data. And less weight is placed on outliers by nonparametric tests.

14.74 $H_0 : P = 0.50$ (the yen is a poor investment)

$H_1 : P \neq 0.50$ (the yen is an excellent investment)

n = 13. For 8 "excellent" and a one-sided test,

P(5 ≥ X ≥ 8) = 2P(X ≤ 5) = 2[.0001 + .0016 + .0095 + .0349 + .0873+.1571]=.581

Therefore, reject H_0 at levels of alpha in excess of 58.1%

14.76 $H_0 : P = 0.50$ (more professors believe analytical skills have deteriorated)

$H_1 : P > 0.50$ (more professors believe that analytical skills have improved)

$n = 120 - 37 = 83,\ \hat{p} = 48/83 = .5783$

$\mu = nP = 83(.5) = 41.5 \quad \sigma = .5\sqrt{n} = .5\sqrt{83} = 4.5552$

$Z = \dfrac{S^* - \mu}{\sigma} = \dfrac{47.5 - 41.5}{4.5552} = 1.32$, p-value= 1 – $F_Z(1.32)$ = 1 - .9066 = .0934

Therefore, reject H_0 at levels of alpha in excess of 9.34%

14.78 H_0: no preference for the management candidates

H_1: otherwise (two-tailed test)

$n = 6, \quad T = \min(T_+, T_-) = 8, \quad T_{0.10} = 4$. From Appendix Table 10, $T_{0.10} = 4$

Therefore, do not reject H_0 at the 20% level

Chapter 15:

Analysis of Variance

15.2 Given the Analysis of Variance table, compute mean squares for between nd for within groups. Compute the F ratio and test the hypothesis that the group means are equal.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,\ H_1 : otherwise$

$$MSG = \frac{SSG}{k-1},\ MSW = \frac{SSW}{n-k},\ F = \frac{MSG}{MSW}$$

$$MSG = \frac{879}{3},\ MSW = \frac{798}{16},\ F = \frac{293}{49.875} = 5.875$$

$F_{3,16,.05} = 3.24,\ F_{3,16,.01} = 5.29$, Therefore, reject H_0 at the 1% level.

15.4 a. $\bar{x}_1 = 62, \bar{x}_2 = 53, \bar{x}_3 = 52$

n=16, SSW = 1028 + 1044 + 1536 = 3608

SSG = $7(62 - 56.0625)^2 + 6(53 - 56.0625)^2 + 6(52 - 56.0625)^2 =$ 340.9375

SST = 3948.9375

b. Complete the anova table

One-way ANOVA: SodaSales versus CanColor

```
Analysis of Variance for SodaSale
Source      DF        SS        MS        F        P
CanColor     2       341       170     0.61    0.556
Error       13      3608       278
Total       15      3949
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level        N      Mean     StDev -+---------+---------+---------+-----
1            6     62.00     14.34           (------------*-----------)
2            5     53.00     16.16    (------------*-------------)
3            5     52.00     19.60   (------------*-------------)
                                   -+---------+---------+---------+-----
Pooled StDev =     16.66           36        48        60        72
```

$H_0 : \mu_1 = \mu_2 = \mu_3,\ H_1 : otherwise$

$F_{2,13,.05} = 3.81$, do not reject H_0 at the 5% level.

15.6a.

MINITAB Output Display:

One-way Analysis of Variance

```
Analysis of Variance
Source      DF        SS         MS         F        P
Factor       2     354.1      177.1     10.45    0.001
Error       15     254.2       16.9
Total       17     608.3
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level        N      Mean     StDev  ---------+---------+---------+------
SupplierA    6    32.000     3.347                   (------*------)
SupplierB    6    24.333     5.007    (------*------)
SupplierC    6    34.833     3.817                         (------*-----)
                                    ---------+---------+---------+------
Pooled StDev =    4.116                    25.0      30.0      35.0
```

b. Assume $\alpha = 0.05$. Reject H_0 and conclude that the population mean numbers of parts per shipments not conforming to standards are not the same for all three suppliers.

c. Assume $\alpha = 0.05$.

$$MSD(k) = q\frac{S_p}{\sqrt{n}} \text{ with } S_p = \sqrt{MSW}$$

$$MSD(3) = 3.67\frac{\sqrt{16.9}}{\sqrt{18}} = 3.56$$

The mean difference for Supplier A and B is

32- 24.333 = 22.33

The mean difference for Supplier A and C is

32 – 34.833 = -2.83

The mean difference for Supplier B and C is

24.333 – 34.833 = -10.50

The Supplier B mean is significantly different from both Supplier A and Supplier C, but the later two are not different.

15.8 a.

MINITAB Output Display:

One-way Analysis of Variance

```
Analysis of Variance
Source      DF        SS        MS        F       P
Factor       2        89        44     0.28   0.756
Error       18      2813       156
Total       20      2901
                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level        N      Mean     StDev  --+---------+---------+---------+---
Freshman     7     71.71     13.16  (-------------*--------------)
Sophomor     7     75.29     11.19       (--------------*-------------)
Juniors      7     76.57     13.05         (-------------*-------------)
                                    --+---------+---------+---------+---
Pooled StDev =     12.50           63.0      70.0      77.0      84.0
```

b. Assume $\alpha = 0.05$. Fail to reject H_0 and conclude that there is insufficient evidence that the three population mean scores are not equal.

c. Assume $\alpha = 0.05$.

$$MSD(k) = q\frac{S_p}{\sqrt{n}} \text{ with } S_p = \sqrt{MSW}$$

$$MSD(3) = 3.61\frac{\sqrt{156}}{\sqrt{21}} = 9.84$$

The mean difference for Freshman and Sophomor is

$71.71 - 75.29 = -3.58$

The mean difference for Freshman and Juniors is

$71.71 - 76.57 = 0.94$

The mean difference for Sophomor and Juniors is

$75.29 - 76.57 = -1.28$

None of the subgroup means are significantly different from each other.

15.10 a. $\bar{x}_1 = 11.3333, \bar{x}_2 = 12.5, \bar{x}_3 = 8$, set out the anova table

One-way ANOVA: Time versus Rank

```
Analysis of Variance for Time
Source     DF        SS        MS        F        P
Rank        2     51.40     25.70     3.27    0.074
Error      12     94.33      7.86
Total      14    145.73
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ---+---------+---------+---------+---
1           6    11.333     3.011              (--------*--------)
2           4    12.500     3.317                (----------*----------)
3           5     8.000     2.000   (--------*--------)
                                   ---+---------+---------+---------+---
Pooled StDev =    2.804               6.0       9.0      12.0      15.0
```

b. $H_0 : \mu_1 = \mu_2 = \mu_3, H_1 : otherwise$

$F_{2,12,.05} = 3.89$, do not reject H_0 at the 5% level.

15.12 MINITAB Output Display:

One-way Analysis of Variance

```
Analysis of Variance
Source      DF        SS        MS        F        P
Factor       2     48.96     24.48     4.07    0.039
Error       15     90.13      6.01
Total       17    139.09
                                    Individual 95% CIs For Mean
                                    Based on Pooled StDev
Level        N      Mean     StDev  --+---------+---------+---------+---
Confessi     6    10.402     2.268                  (--------*-------)
People W     6     7.045     2.182    (-------*--------)
Newsweek     6     6.777     2.850   (-------*--------)
                                    --+---------+---------+---------+---
Pooled StDev =     2.451              5.0       7.5      10.0      12.5
```

Assume $\alpha = 0.05$. Reject H_0 and conclude that the population mean fog indices are not the same for all three magazines.

$MSD(k) = q\frac{S_p}{\sqrt{n}}$ with $S_p = \sqrt{MSW}$

$MSD(3) = 3.67\frac{\sqrt{6.01}}{\sqrt{18}} = 2.12$

The mean difference for Service True Confessions and People Weekly is
$10.402 - 7.045 = 3.357$

The mean difference for True Confessions and Newsweek is
$10.402 - 6.777 = 3.625$

The mean difference for People Weekly and Newsweek is
$7.045 - 6.777 = .268$

The *True Confessions* mean is significantly different from both *People Weekly* and *Newsweek*, but the later two are not different.

15.14 a. $\hat{\mu} = 8.0744$

b. $\hat{G}_1 = 10.4017 - 8.0744 = 2.3273$, $\hat{G}_2 = 7.045 - 8.0744 = -1.0294$

$\hat{G}_3 = 6.7767 - 8.0744 = -1.300$

c. $\hat{\varepsilon}_{13} = 11.15 - 10.4017 = .7483$

15.16

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,\ H_1 : otherwise$

$R_1 = 49,\quad R_2 = 84,\quad R_3 = 76,\quad R_4 = 81$

$$W = \frac{12}{23(24)}\left[(49^2/4) + (84^2/6) + (76^2/7) + (81^2/6)\right] - 3(24) = 8.32$$

$\chi^2_{(3,.05)} = 7.81$, therefore, reject H_0 at the 5% level

15.18 $H_0 : \mu_1 = \mu_2 = \mu_3,\ H_1 : otherwise$

$R_1 = 61,\quad R_2 = 37,\quad R_3 = 38$

$$W = \frac{12}{16(17)}\left[(3721/6) + (1396/5) + (1444/5)\right] - 3(17) = 1.18$$

$\chi^2_{(2,.10)} = 4.61$, therefore, do not reject H_0 at the 10% level

Using Minitab:

Kruskal-Wallis Test: SodaSales versus CanColor

```
Kruskal-Wallis Test on SodaSale
CanColor     N     Median     Ave Rank         Z
1            6      60.00         10.2      1.08
2            5      52.00          7.4     -0.62
3            5      53.00          7.6     -0.51
Overall     16                     8.5
H = 1.18   DF = 2   P = 0.554
H = 1.19   DF = 2   P = 0.553 (adjusted for ties)
```

15.20 $H_0 : \mu_1 = \mu_2 = \mu_3,\ H_1 : otherwise$

$R_1 = 63.5,\quad R_2 = 26,\quad R_3 = 81.5$

$$W = \frac{12}{18(19)}\left[(4032.25 + 676 + 6642.25)/6\right] - 3(19) = 9.3772$$

$\chi^2_{(2,.01)} = 9.21$, therefore, reject H_0 at the 1% level

Using Minitab:

Kruskal-Wallis Test: Nonconforming versus Supplier

```
Kruskal-Wallis Test on Nonconfo
Supplier     N     Median     Ave Rank         Z
1            6      32.00         10.6      0.61
2            6      24.50          4.3     -2.90
3            6      35.00         13.6      2.29
Overall     18                     9.5
H = 9.38   DF = 2   P = 0.009
H = 9.47   DF = 2   P = 0.009 (adjusted for ties)
```

15.22 $H_0: \mu_1 = \mu_2 = \mu_3, H_1: otherwise$

$R_1 = 66, \quad R_2 = 79.5, \quad R_3 = 85.5$

$$W = \frac{12}{21(22)}\left[(4356 + 6320.25 + 7310.25)/7\right] - 3(22) = .7403$$

$\chi^2_{(2,.10)} = 4.61$, therefore, do not reject H_0 at the 10% level

15.24 $H_0: \mu_1 = \mu_2 = \mu_3, H_1: otherwise$

$R_1 = 54.5, \quad R_2 = 43.5, \quad R_3 = 22$

$$W = \frac{12}{15(16)}\left[(2970.25/6) + (1892.25/4) + (484/5)\right] - 3(16) = 5.2452$$

$\chi^2_{(2,.10)} = 4.61$, therefore, reject H_0 at the 10% level

15.26 a. The null hypothesis tests the equality of the population mean ratings across the classes

b. $H_0: \mu_1 = \mu_2 = \mu_3, H_1: otherwise$

$W = .15$, $\chi^2_{(2,.10)} = 4.61$, therefore, do not reject H_0 at the 10% level

15.28

$$MSG = \frac{SSG}{K-1}, \; MSB = \frac{SSB}{H-1}, \; MSE = \frac{SSE}{(K-1)(H-1)}$$

Test of K population group means all the same:

Reject H_0 if $\frac{MSG}{MSE} > F_{K-1,(K-1)(H-1),\alpha}$

Test of H population block means all the same:

Reject H_0 if $\frac{MSB}{MSE} > F_{H-1,(K-1)(H-1),\alpha}$

$$MSG = \frac{380}{6} = 63.333, \; MSB = \frac{232}{5} = 46.4, \; MSE = \frac{387}{(6)(5)} = 12.90$$

$\frac{MSB}{MSE} = \frac{46.4}{12.90} = 3.597$, $F_{5,30,.05} = 2.53$, $F_{5,30,.01} = 3.70$

Reject at the 5% level, do not reject H_0 at the 1% level that the block means differ.

$\frac{MSG}{MSE} = \frac{63.333}{12.90} = 4.91$, $F_{6,30,.05} = 2.42$, $F_{6,30,.01} = 3.47$

Reject H_0 at the 1% level. Evidence suggests the group means differ

15.30 a. two-way ANOVA table:

Two-way ANOVA: earngrowth versus OilCo, Analyst

Analysis of Variance for earngrow

Source	DF	SS	MS	F	P
OilCo	4	3.30	0.83	0.31	0.866
Analyst	3	31.35	10.45	3.93	0.036
Error	12	31.90	2.66		
Total	19	66.55			

```
                              Individual 95% CI
OilCo           Mean      ------+---------+---------+---------+-----
1              10.00            (--------------*--------------)
2               9.50      (--------------*--------------)
3              10.25              (--------------*--------------)
4              10.75                  (--------------*--------------)
5              10.25              (--------------*--------------)
                          ------+---------+---------+---------+-----
                              8.40      9.60     10.80     12.00
                              Individual 95% CI
Analyst_        Mean      -----+---------+---------+---------+------
A               9.80            (----------*----------)
B               9.80            (----------*----------)
C               8.80      (----------*----------)
D              12.20                       (----------*----------)
                          -----+---------+---------+---------+------
                             8.00      9.60     11.20     12.80
```

b. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5, H_1 : Otherwise$

$F_{4,12,.05} = 3.26 > .31$, therefore, do not reject H_0 at the 5% level

15.32 a. two-way ANOVA table:

Two-way ANOVA: sales versus Quarter, soup

Analysis of Variance for sales

Source	DF	SS	MS	F	P
Quarter_	3	615.0	205.0	2.10	0.202
Soup	2	6.2	3.1	0.03	0.969
Error	6	586.5	97.7		
Total	11	1207.7			

```
                              Individual 95% CI
Quarter_        Mean      ----------+---------+---------+---------+-
1               56.3          (-----------*-----------)
2               67.3                   (-----------*-----------)
3               66.7                   (-----------*----------)
4               50.3      (-----------*-----------)
                          ----------+---------+---------+---------+-
                                  48.0      60.0      72.0      84.0
                              Individual 95% CI
Soup            Mean      ---+---------+---------+---------+--------
A               60.3        (----------------*----------------)
B               59.3      (----------------*----------------)
C               61.0         (----------------*----------------)
                          ---+---------+---------+---------+--------
                           49.0      56.0      63.0      70.0
```

b. $H_0 : \mu_1 = \mu_2 = \mu_3, H_1 : otherwise$

$F_{2,6,.05} = 5.14 > .03$, therefore, do not reject H_0 at the 5% level

15.34 a. two-way ANOVA table:

Two-way ANOVA: Ratings versus Exam, Text

```
Analysis of Variance for Ratings
Source       DF        SS        MS        F        P
Exam          2    0.2022    0.1011     5.20    0.077
Text          2    0.4356    0.2178    11.20    0.023
Error         4    0.0778    0.0194
Total         8    0.7156
                         Individual 95% CI
Exam           Mean   ----------+---------+---------+---------+-
Essays        4.633   (-----------*----------)
MC            5.000                      (----------*----------)
Mix           4.833             (-----------*----------)
                      ----------+---------+---------+---------+-
                              4.600     4.800     5.000     5.200
                         Individual 95% CI
Text           Mean   ---+---------+---------+---------+--------
A             4.667    (--------*--------)
B             5.133                      (--------*--------)
C             4.667    (--------*--------)
                      ---+---------+---------+---------+--------
                       4.500     4.750     5.000     5.250
```

b. $H_0: \mu_1 = \mu_2 = \mu_3, H_1: otherwise$ [texts]

$F_{2,4,.05} = 6.94 < 11.20$, therefore, reject H_0 at the 5% level

c. $H_0: \mu_1 = \mu_2 = \mu_3, H_1: otherwise$ [exam type]

$F_{2,4,.05} = 6.94 > 5.20$, therefore, do not reject H_0 at the 5% level

15.36

$\hat{G}_3 = -.1556$

$\hat{B}_1 = .1778$

$\hat{\varepsilon}_{31} = .0556$

15.38 a. complete the ANOVA table:

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Fertilizers	135.6	3	45.20	6.0916
Soil Types	81.7	5	16.34	2.2022
Error	111.3	15	7.42	
Total	328.6	23		

b. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4, H_1: otherwise$ [fertilizers]

$F_{3,15,.01} = 5.42 < 6.0916$, therefore, reject H_0 at the 1% level

c. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6, H_1: Otherwise$ [soil types]

$F_{5,15,.05} = 2.90 > 2.2022$, therefore, do not reject H_0 at the 5% level

15.40 Given, say, ten pairs observations, the F statistic would have 1, 9 degrees of freedom. The test is in the form of a two-tailed test. With alpha = .05, the critical value of F would be 5.12. For a matched –pairs test, the degrees of freedom would be 9, and the area in each tail would be .025. The critical value for t would be 2.262 (which is the square root of the F statistic of 5.12). Therefore, the two tests are equivalent.

15.42

$$MSG = \frac{SSG}{K-1}, MSB = \frac{SSB}{H-1}, MSI = \frac{SSI}{(K-1)(H-1)}, MSE = \frac{SSE}{KH(L-1)}, F = \frac{MSG}{MSE}, \frac{MSB}{MSE}, \frac{MSI}{MSE}$$

$$MSG_A = \frac{86}{4} = 21.5, MSG_B = \frac{75}{5} = 15, MSI = \frac{75}{20} = 3.75, MSE = \frac{300}{90} = 3.33$$

$$F\ Ratio: Interaction = \frac{MSI}{MSE} = \frac{3.75}{3.33} = 1.125,$$

We wil use Excel function "F.INV.RT(probability,deg_freedom1,deg_freedom2)" to calculate Critical values.

$F_{20,90,.05} \approx 1.69$, $F_{20,90,.01} \approx 2.09$ Do not reject H_0 at the 5% level. No significant interaction exists between treatment groups A and B. Therefore, go on to test the main effects of each treatment group.

$$F\ Ratio: Treatment\ A = \frac{MSG_A}{MSE} = \frac{21.5}{3.33} = 6.456, F_{4,90,.05} \approx 2.47, F_{4,90,.01} \approx 3.53$$

Reject H_0 at the 1% level, there is a significant main effect for group A.

$$F\ Ratio: Treatment\ B = \frac{MSG_B}{MSE} = \frac{15}{3.33} = 4.505,\ F_{5,90,.05} \approx 2.32, F_{5,90,.01} \approx 3.23$$

Reject H_0 at the 1% level, there is a significant main effect for group B.

15.44 a. ANOVA table:

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Contestant	364.50	21	17.3571	19.2724
Judges	.81	8	.1013	.1124
Interaction	4.94	168	.0294	.0326
Error	1069.94	1188	.9006	
Total	1440.19	1385		

H_0: Mean value for all 22 contestants is the same
H_1: Otherwise

$F_{21,1188,.01} \approx 1.88 < 19.2724$, therefore, reject H_0 at the 1% level

H_0: Mean value for all 9 judges is the same
H_1: Otherwise

$F_{8,1188,.05} \approx 1.94 > .1124$, therefore, do not reject H_0 at the 5% level

H_0: No interaction exists between contestants and judges
H_1: Otherwise

$F_{168,1188,.05} \approx 1.22 > .0326$, therefore, do not reject H_0 at the 5% level

15.46 a. ANOVA table:

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Test type	57.5556	2	28.7778	4.7091
Subject	389.0000	3	129.6667	21.2182
Interaction	586.0000	6	97.66667	15.9818
Error	146.6667	24	6.1111	
Total	1179.2223	35		

b. H_0: No interaction exists between subject type and test type

H_1: Otherwise

$F_{6,24,.01} = 3.67 < 15.9818$, therefore, reject H_0 at the 1% level

15.48 a. The implied assumption is that there is no interaction effect between student year and dormitory ratings

b. Using Minitab:

General Linear Model: Ratings_48 versus Dorm_48, Year_48

```
Factor      Type Levels Values
Dorm_48    fixed      4 A B C D
Year_48    fixed      4 1 2 3 4

Analysis of Variance for Ratings_, using Adjusted SS for Tests
Source       DF      Seq SS     Adj SS     Adj MS        F      P
Dorm_48       3      20.344     20.344      6.781     4.91  0.008
Year_48       3      10.594     10.594      3.531     2.56  0.078
Error        25      34.531     34.531      1.381
Total        31      65.469
```

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Dorm	20.344	3	6.781	4.91
Year	10.594	3	3.531	2.56
Error	34.531	25	1.381	
Total	65.469	31		

c. H_0: Mean ratings for all 4 dormitories is the same

H_1: Otherwise

$F_{3,25,.01} = 4.68 < 4.91$, therefore, reject H_0 at the 1% level

d. H_0: Mean ratings for all 4 student years is the same

H_1: Otherwise

$F_{3,25,.05} = 2.99 > 2.56$, therefore, do not reject H_0 at the 5% level

15.50

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Color	243.250	2	121.625	11.3140
Region	354.000	3	118.000	10.9767
Interaction	189.750	6	31.625	2.9419
Error	129.000	12	10.750	
Total	916.000	23		

H_0: No interaction exists between region and can color,
H_1: Otherwise

$F_{6,12,.01} = 4.82 > 2.9419$, therefore, do not reject H_0 at the 1% level

Main effects are significant for both Color and Region

15.52 One-way ANOVA examines the effect of a single factor (having three or more conditions). Two-way ANOVA recognizes situations in which more than one factor may be significant.
Example of One-way ANOVA
The data obtained on yield of different varieties of the same crop can be analyzed using One-way ANOVA.
Example of Two-way ANOVA
The data on number of students in a University categorized into different groups- Freshmen, Sophomores, and Juniors and different Races- black, white, and Hispanic can be analyzed using Two-way ANOVA.

15.54

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Between	5165	2	2582.5	21.4848
Within	120802	1005	120.201	
Total	125967	1007		

$F_{2,1005,.01} = 4.61 < 21.4848$, therefore, reject H_0 at the 1% level

15.56 a. Use result from Equation 15.2 to find SSG, then compute MSG and finally, find SSW using the fact that SSW = $(n-k)\dfrac{MSG}{F}$

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Between	221.3400	3	73.7800	25.6
Within	374.6640	130	2.8820	
Total	596.0040	133		

b. H_0: Mean salaries are the same for managers in all 4 groups
H_1: Otherwise
$F_{3,130,.01} \approx 3.95 < 25.6$, therefore, reject H_0 at the 1% level

15.58

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Between	11438.3028	2	5719.1514	.7856
Within	109200.000	15	7280.000	
Total	120638.3028	17		

H_0: Mean sales levels are the same for all three periods
H_1: Otherwise
$F_{2,15,.05} = 3.68 > .7856$, therefore, do not reject H_0 at the 5% level

15.60

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,\ H_1 : otherwise$$

$$R_1 = 48.5, \quad R_2 = 55, \quad R_3 = 74, \quad R_4 = 32.5$$

$$W = \frac{12}{20(21)}\left[(2352.25 + 3025 + 5476 + 1056.25)/5\right] - 3(21) = 5.0543$$

$\chi^2_{(3,.10)} = 6.25$, therefore, do not reject H_0 at the 10% level

15.62 a. $$SSW = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$$

$$= \sum_{i=1}^{K}\left[\sum_{j=1}^{n_i} x_{ij}^2 - 2n_i\bar{x}_i^2 + n_i\bar{x}_i^2\right]$$

$$= \sum_{i=1}^{K}\sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{K} n_i\bar{x}_i^2$$

b. $$SSG = \sum_{i=1}^{K} n_i(\bar{x}_i - \bar{\bar{x}})^2$$

$$= \sum_{i=1}^{K} n_i\bar{x}_i^2 - 2\bar{\bar{x}}\sum_{i=1}^{k} n_i\bar{x}_i + n\bar{\bar{x}}^2$$

$$= \sum_{i=1}^{K} n_i\bar{x}_i^2 - 2n\bar{\bar{x}}^2 + n\bar{\bar{x}}^2$$

$$= \sum_{i=1}^{K} n_i\bar{x}_i^2 - n\bar{\bar{x}}^2$$

c. $SST = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{\bar{x}})^2$

$$= \sum_{i=1}^{K}\sum_{j=1}^{n_i}x^2_{ij} - 2\bar{\bar{x}}\sum_{i=1}^{K}\sum_{j=1}^{n_i}x_{ij} + \sum_{i=1}^{K}\sum_{j=1}^{n_i}\bar{\bar{x}}^2$$

$$= \sum_{i=1}^{K}\sum_{j=1}^{n_i}x^2_{ij} - 2\bar{\bar{x}}n\bar{\bar{x}} + n\bar{\bar{x}}^2$$

$$= \sum_{i=1}^{K}\sum_{j=1}^{n_i}x^2_{ij} - n\bar{\bar{x}}^2$$

15.64

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Consumers	37571.5	124	302.996	1.3488
Brands	32987.3	2	16493.65	73.4226
Error	55710.7	248	224.6399	
Total	126269.5	374		

H_0: Mean perception levels are the same for all three brands
H_1: Otherwise
$F_{2,248,.01} \approx 4.79 < 73.4226$, therefore, reject H_0 at the 1% level

15.66 Using Minitab:

Two-way ANOVA: GPA versus SAT, Income

```
Source     DF        SS         MS      F      P
SAT         2  0.826667  0.413333  24.80  0.006
Income      2  0.006667  0.003333   0.20  0.826
Error       4  0.066667  0.016667
Total       8  0.900000
S = 0.1291   R-Sq = 92.59%   R-Sq(adj) = 85.19%
                    Individual 95% CIs For Mean Based on
                    Pooled StDev
SAT        Mean    +---------+---------+---------+---------
High    3.36667                       (------*------)
Mod     2.90000    (------*------)
VeryH   3.63333                                 (------*------)
                   +---------+---------+---------+---------
                 2.70      3.00      3.30      3.60
                       Individual 95% CIs For Mean Based on
                       Pooled StDev
Income        Mean  ------+---------+---------+---------+---
High       3.33333      (-------------*-------------)
Low        3.26667  (-------------*-------------)
Mod        3.30000    (-------------*-------------)
                    ------+---------+---------+---------+---
                        3.15      3.30      3.45      3.60
```

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Income	.0067	2	.0033	.2000
SAT Score	.8267	2	.4133	24.8000
Error	.0667	4	.0167	
Total	.9000	8		

H_0: Mean GPAs are the same for all three income groups
H_1: Otherwise
$F_{2,4,.05} = 6.94 > .2000$, therefore, do not reject H_0 at the 5% level

H_0: Mean GPAs are the same for all three SAT score groups
H_1: Otherwise
$F_{2,4,.01} = 18.0 < 24.8$, therefore, reject H_0 at the 1% level

Descriptive Statistics: GPA

```
Variable    Mean  SE Mean  StDev  Variance  Minimum     Q1  Median     Q3
GPA        3.300    0.112  0.335     0.112    2.800  2.950   3.400  3.600

Variable  Maximum    IQR
GPA         3.700  0.650
```

15.68 a. $\hat{\mu} = 3.3$

b. $\hat{G}_2 = 0.0$

c. $\hat{B}_2 = .0667$

d. $\hat{\varepsilon}_{22} = .1333$

15.70 ANOVA table:

Source of Variation	Sum of Squares	df	Mean square	F Ratio
Prices	.178	2	.0890	.0944
Countries	4.365	2	2.1825	2.3151
Interaction	1.262	4	.3155	.3347
Error	93.330	99	.9427	
Total	99.135	107		

H_0: Mean quality ratings for all three prices levels is the same
H_1: Otherwise
$F_{2,99,.05} \approx 3.07 > .0944$; therefore, do not reject H_0 at the 5% level

H_0: Mean quality ratings for all three countries is the same
H_1: Otherwise
$F_{2,99,.05} \approx 3.07 > 2.3151$; therefore, do not reject H_0 at the 5% level

H_0: No interaction exists between price and country
H_1: Otherwise
$F_{4,99,.05} \approx 2.45 > .3347$, therefore, do not reject H_0 at the 5% level

15.72

Data

Sat	High	Moderate	Low
Very	3.7	3.6	3.6
Very	3.9	3.7	3.8
High	3.4	3.5	3.2
High	3.2	3.6	3.4
Moderate	2.9	2.8	3
Moderate	**2.7**	**3**	**2.8**

a. ANOVA table:

Two-way ANOVA: GPA versus SAT score, Income

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample(sat)	2.201111	2	1.100556	66.03333	4.18E-06	4.256495
Columns(Income group)	0.017778	2	0.008889	0.533333	0.604096	4.256495
Interaction	0.102222	4	0.025556	1.533333	0.272407	3.633089
Within	0.15	9	0.016667			
Total	2.471111	17				

a. H_0: The population mean grade point averages are the same for all three income groups H_1: Otherwise

p-value $0.6041 > 0.05$, therefore, do not reject H_0 at the 5% level

b. H_0: Mean GPAs for all three SAT score groups is the same

H_1: Otherwise

p-value $= 0.000 < 0.1$; therefore, reject H_0 at the 1% level.

c. H_0: There is no interaction between income group and SAT score.

H_1: Otherwise

p-value $= 0.2724 > 0.05$ therefore, do not reject H_0 at the 5% level.

Chapter 16:

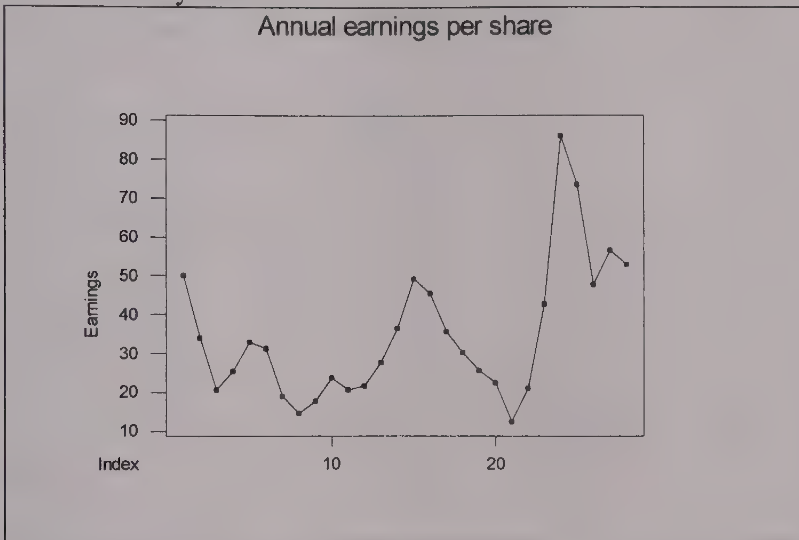
Time-Series Analysis and Forecasting

16.2 a. Runs test on Earnings per share

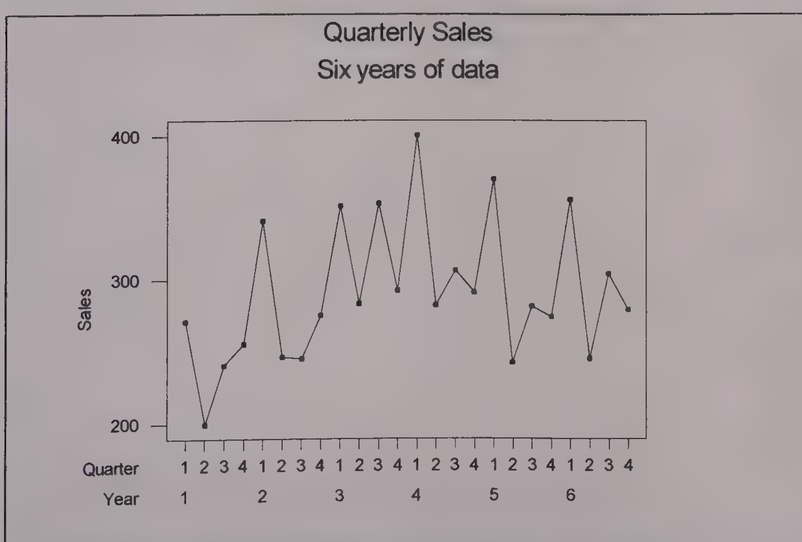
Runs Test: Earnings

```
Earnings   K =    30.7000
   The observed number of runs =   7
    The expected number of runs =  15.0000
    14 Observations above K   14 below
              The test is significant at  0.0021
```

b. From the time series plot below, cyclical behavior tends to dominate with upward trend for the most recent 7 years.



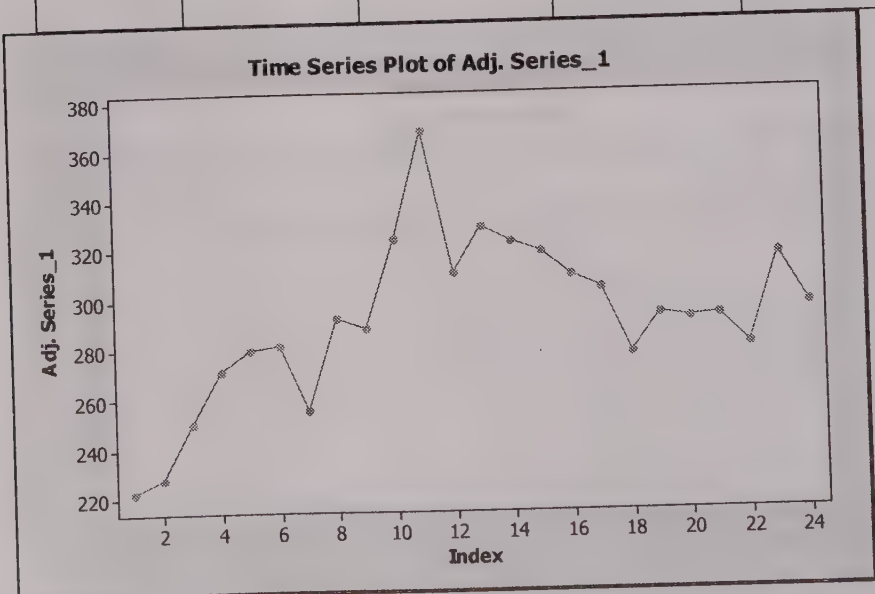
16.4 a. Time series plot – Quarterly Sales



The data exhibits seasonality and a quadratic trend

b.

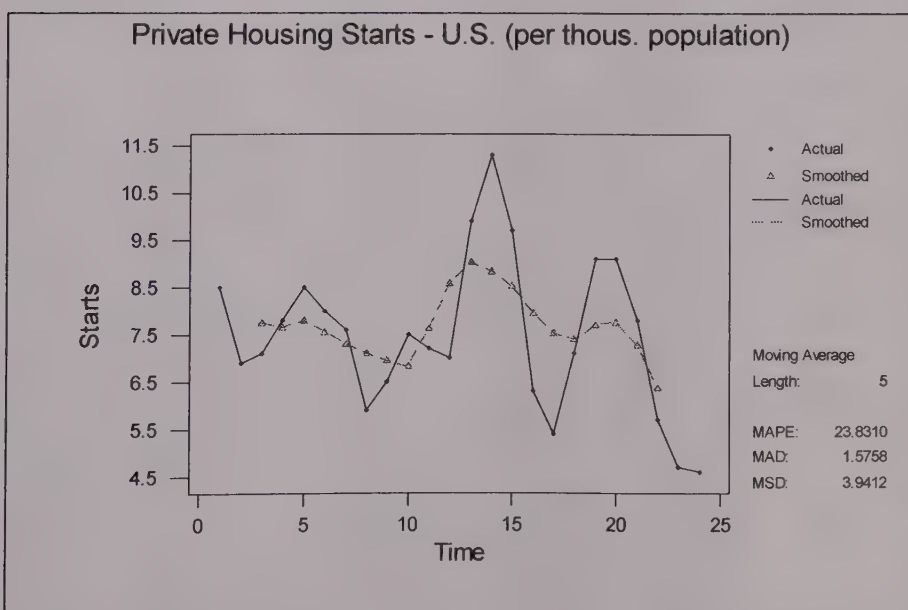
Period	4 Period MA	$100\frac{X_t}{X_t^*}$	Seas. Factor	Adj. Series
1-1			122.110	221.930
2			87.483	227.474
3	250.000	96.000	96.129	249.664
4	264.625	96.363	94.278	270.477
2-1	271.125	125.772		279.255
2	274.250	89.699		281.199
3	278.000	88.130		254.566
4	283.875	96.874		291.691
3-1	302.000	116.225		287.445
2	317.625	89.099		323.493
3	326.000	108.282		367.214
4	332.125	87.919		309.723
4-1	326.125	122.959		328.391
2	320.125	88.091		322.350
3	316.125	96.797		318.322
4	307.250	94.711		308.663
5-1	299.125	123.694		303.004
2	293.875	82.348		276.626
3	290.000	96.897		292.315
4	288.625	94.933		290.631
6-1	291.875	121.970		291.539
2	295.375	82.945		280.056
3				316.241
4				295.934



The seasonally adjusted data exhibits irregular patterns.

16.6 Housing Starts – Annual, per thousand population, U.S.

Starts	5ptMA
1	*
2	*
3	7.76
4	7.66
5	7.80
6	7.56
7	7.30
8	7.10
9	6.94
10	6.82
11	7.62
12	8.58
13	9.02
14	8.84
15	8.52
16	7.96
17	7.52
18	7.40
19	7.70
20	7.76
21	7.28
22	6.38
23	*
24	*



Strong cyclical behavior is evident in the smoothed series

16.8 $$x^*_{t+1} - x^*_t = \frac{1}{2m+1}\sum_{j=-m}^{m} x_{t+j+1} - \frac{1}{2m+1}\sum_{j=-m}^{m} x_{t+j} = \frac{1}{2m+1}(x_{t+m+1} - x_{t-m})$$

$$x^*_{t+1} = x^*_t + \frac{x_{t+m+1} - x_{t-m}}{2m+1}$$

Using this formula, one may avoid summing 2m+1 terms when calculating each moving average value

16.10 a. $$x^*_t = \frac{x^*_{t-.5} + x^*_{t+.5}}{2} = \frac{\sum_{j=-(s/2)+1}^{s/2}(x_{t+j-1} + x_{t+j})}{2s}$$

$$= \frac{x_{t-(s/2)} + 2(x_{t-(s/2)+1} + \cdots + x_{t+(s/2)-1}) + x_{t+(s/2)}}{2s}$$

b. $$x^*_{t+1} = \frac{x_{t-(s/2)} + 2(x_{t-(s/2)+1} + \cdots + x_{t+(s/2)-1}) + x_{t+(s/2)}}{2s} + x^*_t - x^*_t$$

$$= x^*_t + \frac{x_{t+(s/2)+1} + x_{t+(s/2)} - x_{t-(s/2)+1} - x_{t-(s/2)}}{2s}$$

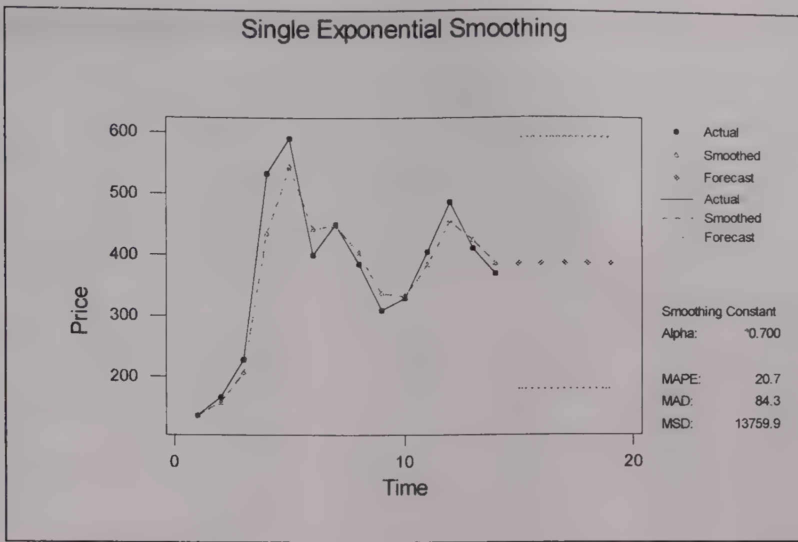
16.12 Use smoothing constant of .7 (alpha of .3) in Minitab. Set initial smoothing value at the average of the first '1' observations

Single Exponential Smoothing

```
Data          Price
Length        14.0000
NMissing      0
Smoothing Constant
Alpha: 0.7
Accuracy Measures
MAPE:      20.7
MAD:       84.3
MSD:    13759.9
```

Row	Time	Price	SMOO2	FITS2	Error
1	1	135	135.000	135.000	0.000
2	2	166	156.700	135.000	31.000
3	3	227	205.910	156.700	70.300
4	4	533	434.873	205.910	327.090
5	5	591	544.162	434.873	156.127
6	6	399	442.549	544.162	-145.162
7	7	450	447.765	442.549	7.451
8	8	385	403.829	447.765	-62.765
9	9	308	336.749	403.829	-95.829
10	10	329	331.325	336.749	-7.749
11	11	405	382.897	331.325	73.675
12	12	486	455.069	382.897	103.103
13	13	410	423.521	455.069	-45.069
14	14	369	385.356	423.521	-54.521

Row	Period	Forecast	Lower	Upper
1	15	385.356	178.884	591.828
2	16	385.356	178.884	591.828
3	17	385.356	178.884	591.828
4	18	385.356	178.884	591.828
5	19	385.356	178.884	591.828



16.14a. Obtain the smoothed series $\hat{x}_t$ as shown below.

$\hat{x}_1 = x_1$ and $\hat{x}_t = (1-\alpha)\hat{x}_{t-1} + \alpha x_t$

YEAR	Smoothed	Smoothed	Smoothed	Smoothed
1	50.20	50.20	50.20	50.20
2	37.16	40.42	43.68	46.94
3	23.91	28.53	34.45	41.67
4	25.10	26.65	30.83	38.42
5	31.34	30.40	31.66	37.31
6	31.31	30.94	31.51	36.11
7	21.30	23.66	26.43	32.65
8	15.86	18.16	21.66	29.02
9	17.17	17.76	19.99	26.72
10	22.31	21.27	21.44	26.09
11	20.94	20.87	21.10	24.99
12	21.47	21.31	21.30	24.32
13	26.37	25.08	23.82	24.97
14	34.47	31.93	28.89	27.28
15	46.33	42.35	37.06	31.68
16	45.59	44.18	40.39	34.43
17	37.60	39.03	38.48	34.66
18	31.60	33.67	35.13	33.75
19	26.72	28.77	31.28	32.10
20	23.26	24.95	27.73	30.16
21	14.49	17.36	21.56	26.59
22	19.62	19.48	21.29	25.45
23	38.16	33.47	29.90	28.92
24	76.67	65.17	52.46	40.40
25	74.21	70.23	60.91	47.04
26	53.00	56.71	55.63	47.17
27	55.88	56.64	56.02	49.06
28	53.66	54.52	54.85	49.86

b. The squared error, $e_t^2 = (x_t - \hat{x}_{t-1})^2$, for each forecast for each value of α is shown below.

YEAR	e^2	e^2	e^2	e^2
1				
2	265.69	265.69	265.69	265.69
3	274.23	392.83	532.69	693.80
4	2.21	9.78	81.87	264.78
5	60.80	39.05	4.29	30.44
6	0.00	0.81	0.13	36.17
7	156.45	147.38	161.66	299.68
8	46.26	83.83	142.29	329.39
9	2.69	0.44	17.28	132.69
10	41.32	34.05	13.00	9.71
11	2.94	0.44	0.70	30.17
12	0.43	0.54	0.25	11.52
13	37.59	39.61	39.68	10.79
14	102.54	130.36	160.76	132.89
15	219.79	301.61	416.47	484.98
16	0.87	9.28	69.63	188.18
17	99.74	73.64	22.98	1.38
18	56.21	79.79	70.16	20.80
19	37.20	66.80	92.65	68.04
20	18.66	40.57	78.77	94.07
21	120.21	159.96	237.94	318.94
22	41.05	12.54	0.43	32.34
23	537.38	543.65	462.55	301.03
24	2317.10	2790.64	3181.43	3292.49
25	9.44	71.08	447.01	1102.52
26	703.02	507.50	174.62	0.44
27	12.94	0.01	0.94	88.94
28	7.73	12.56	8.51	16.36
Total	5174.49	5814.44	6684.38	8258.23

Use the forecast with smoothing constant $\alpha = 0.8$ since it minimizes the sum of squared forecast errors.

16.16 If alpha is 1.0, then the forecast will always be equal to the first observation.

$\hat{X}_{t+h} = X_1$

16.18 Double Exponential Smoothing

```
Data          INDEX
Length        15.0000
NMissing      0

Smoothing Constants
Alpha (level): 0.7
Gamma (trend): 0.5

Accuracy Measures
MAPE:    65.69
MAD:     16.64
MSD:   1063.04
```

Row	Period	Forecast	Lower	Upper
1	16	127.69	86.93	168.46
2	17	142.91	91.44	194.38
3	18	158.12	95.02	221.22
4	19	173.33	98.10	248.57
5	20	188.54	100.89	276.20

Using Minitab, the forecasts for the next 5 years are 127.69, 142.91, 158.12, 173.33, and 188.54.

16.20 Double Exponential Smoothing

```
Data          FoodPrice
Length        14.0000
NMissing      0
Smoothing Constants
Alpha (level): 0.5
Gamma (trend): 0.5
Accuracy Measures
MAPE: 0.250818
MAD:  0.303748
MSD:  0.129733
```

Row	Time	FoodPrice	Smooth	Predict	Error
1	1	116.6	116.659	116.717	-0.117143
2	2	117.1	117.174	117.248	-0.147527
3	3	117.8	117.763	117.726	0.074162
4	4	118.9	118.617	118.334	0.566466
5	5	119.5	119.414	119.329	0.171002
6	6	120.3	120.235	120.169	0.130519
7	7	120.6	120.811	121.022	-0.422352
8	8	120.8	121.147	121.493	-0.693200
9	9	121.2	121.428	121.655	-0.455323
10	10	122.1	121.961	121.823	0.277445
11	11	122.6	122.513	122.426	0.174469
12	12	123.6	123.310	123.021	0.579363
13	13	124.2	124.082	123.963	0.236969
14	14	125.0	124.897	124.793	0.206530

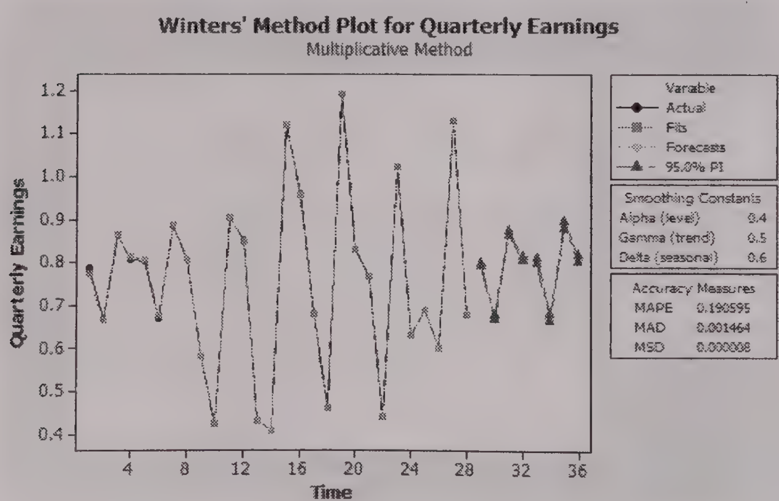
Row	Period	Forecast	Lower	Upper
1	15	125.660	124.916	126.405
2	16	126.424	125.580	127.268
3	17	127.187	126.234	128.141

16.22

Winters' Method for Quarterly Earnings

```
Data    Quarterly Earnings
Length  28


Smoothing Constants

Alpha (level)     0.4
Gamma (trend)     0.5
Delta (seasonal)  0.6


Accuracy Measures

MAPE  0.190595
MAD   0.001464
MSD   0.000008


Forecasts

Period  Forecast     Lower     Upper
29      0.795987  0.792399  0.799575
30      0.674085  0.669967  0.678203
31      0.868721  0.863958  0.873484
32      0.811278  0.805797  0.816759
33      0.805946  0.799698  0.812194
34      0.673472  0.666424  0.680519
35      0.890112  0.882242  0.897982
36      0.810194  0.801485  0.818902
```

Using Minitab, the forecasts for the next 8 quarters are 0.795, 0.674, 0.868, 0.811, 0.805, 0.673, 0.890, and 0.810. The graph of the data and the forecasts is shown below



16.24 **Regression Analysis: EARNINGS versus EARNINGSlag1**

```
The regression equation is
EARNINGS = 9.65 + 0.721 EARNINGSlag1
27 cases used, 1 cases contain missing values
Predictor        Coef   SE Coef      T      P
Constant        9.646     5.387   1.79  0.085
EARNINGSlag1   0.7211    0.1403   5.14  0.000

S = 12.7116   R-Sq = 51.4%   R-Sq(adj) = 49.4%

Analysis of Variance
Source          DF      SS      MS      F      P
Regression       1  4267.3  4267.3  26.41  0.000
Residual Error  25  4039.6   161.6
Total           26  8306.9
```

The first-order autoregressive model is

$\hat{y}_t = 9.646 + 0.7211 y_{t-1} + a_t$

$y_{29} = 9.646 + .7211(53.1) = 47.9364$

$y_{30} = 9.646 + .7211(47.9364) = 44.2129$

$y_{31} = 9.646 + .7211(44.2129) = 41.5280$

$y_{32} = 9.646 + .7211(41.5280) = 39.5918$

16.26 4th order model:

Regression Analysis: Starts versus Startlag1, Startlag2, ...

```
The regression equation is
Starts = 4.45 + 1.25 Startlag1 - 1.10 Startlag2 + 0.313 Startlag3
          - 0.064 Startlag4
20 cases used 4 cases contain missing values
Predictor        Coef     SE Coef        T        P
Constant        4.449      2.884      1.54    0.144
Startlag1      1.2517     0.2783      4.50    0.000
Startlag2     -1.0950     0.4182     -2.62    0.019
Startlag3      0.3131     0.4193      0.75    0.467
Startlag4     -0.0641     0.2935     -0.22    0.830

S = 1.042       R-Sq = 73.2%     R-Sq(adj) = 66.1%
```

$T - statistic\ for\ \phi_4 = -.218$. Fail to reject H_0 at the 10% level

3rd order model:

Regression Analysis: Starts versus Startlag1, Startlag2, ...

```
The regression equation is
Starts = 4.10 + 1.24 Startlag1 - 1.03 Startlag2 + 0.245 Startlag3
21 cases used 3 cases contain missing values
Predictor        Coef     SE Coef        T        P
Constant        4.100      2.183      1.88    0.078
Startlag1      1.2375     0.2509      4.93    0.000
Startlag2     -1.0325     0.2900     -3.56    0.002
Startlag3      0.2450     0.2695      0.91    0.376

S = 0.9808      R-Sq = 73.2%     R-Sq(adj) = 68.4%
```

$T - statistic\ for\ \phi_3 = .909$. Fail to reject H_0 at the 10% level

2nd order model:

Regression Analysis: Starts versus Startlag1, Startlag2

```
The regression equation is
Starts = 5.73 + 1.03 Startlag1 - 0.788 Startlag2
22 cases used 2 cases contain missing values
Predictor        Coef     SE Coef        T       P
Constant        5.728       1.245     4.60   0.000
Startlag       1.0332      0.1573     6.57   0.000
Startlag      -0.7877      0.1705    -4.62   0.000

S = 0.9765      R-Sq = 70.3%      R-Sq(adj) = 67.2%
```

$T-statistic\ for\ \phi_2 = -4.621$. Reject H_0 at the 10% level

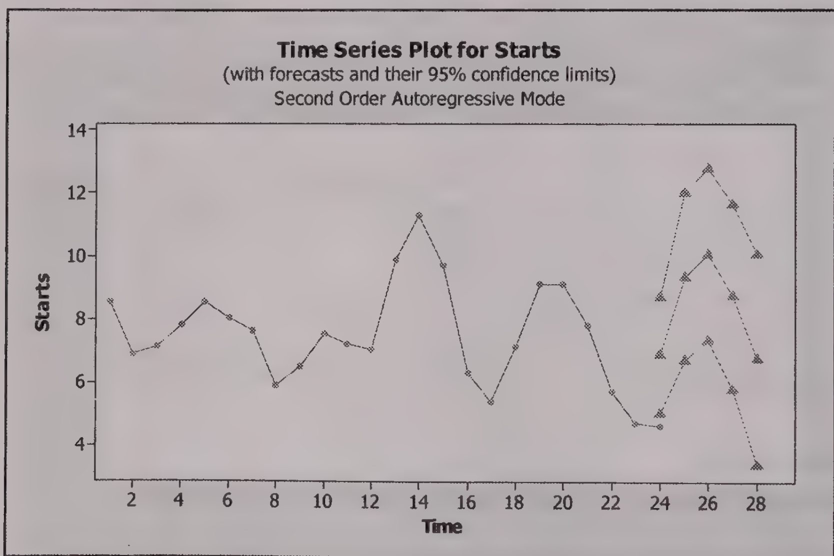
1st order model:

Regression Analysis: Starts versus Startlag1

```
The regression equation is
Starts = 2.61 + 0.633 Startlag1
23 cases used 1 cases contain missing values

Predictor        Coef     SE Coef        T       P
Constant        2.614       1.451     1.80   0.086
Startlag       0.6333      0.1874     3.38   0.003

S = 1.376       R-Sq = 35.2%      R-Sq(adj) = 32.1%
```

Forecasts from the second order model:

$\hat{y}_{25} = 6.776,\ \hat{y}_{26} = 9.103,\ \hat{y}_{27} = 9.792,\ \hat{y}_{28} = 8.670,\ \hat{y}_{29} = 6.968$



No difference in model selection with 10% or 5% significance level.

16.28

a. $T-statistic\ for\ \phi_2 = 2.72$, Reject H_0 at the 10% level

Hence the 3rd order model is selected.

b. $\hat{y}_{66} = 131.573$, $\hat{y}_{67} = 133.913$, $\hat{y}_{68} = 134.452$

16.30 For $h=1$: $\hat{x}_{n+1} = x_n$

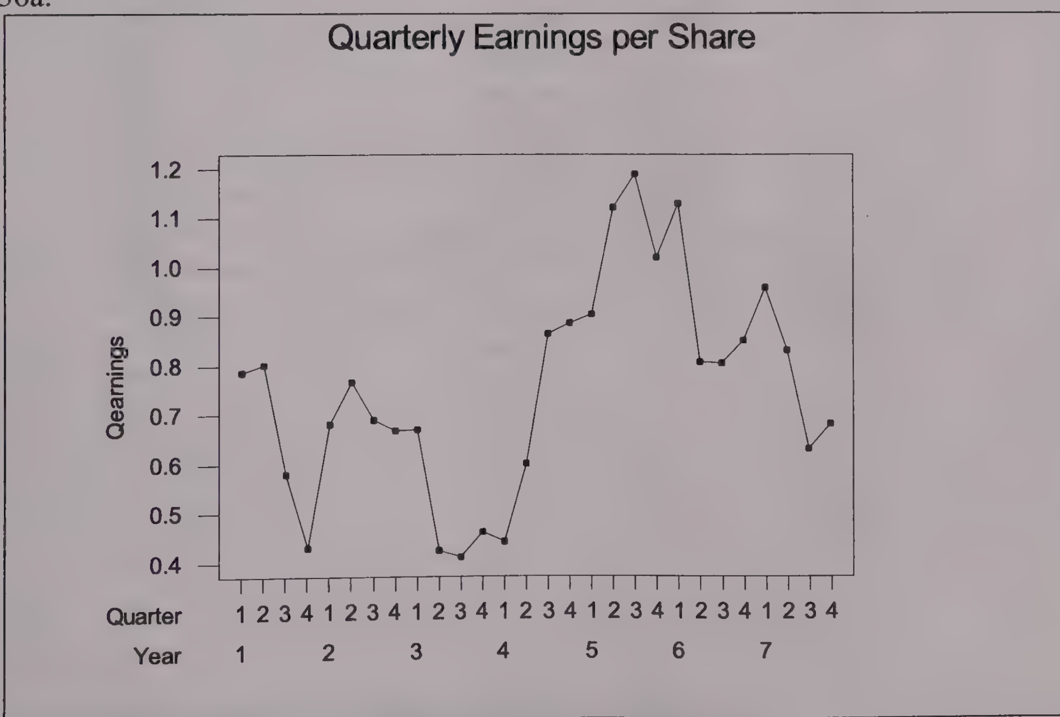
For $h=2$: $\hat{x}_{n+2} = \hat{x}_{n+1} = x_n$

Thus, for h: $\hat{x}_{n+h} = \hat{x}_{n+h-1} = \cdots = x_n$

16.32 Values of a time series can be made up of trend, seasonal, cyclical and random components. For example, total retail sales in the U.S. tend to have a seasonal component with much larger retail sales in the November/December months than others due to gift buying for the holiday season. Strong trend components tend to exist in total personal income in the U.S. since the factors that influence income change slowly over time. Unemployment rates in the U.S. show strong cyclical behavior due to the ups and downs of the business cycle.

16.34 This will very likely not be a successful strategy since the manager could use the trend and seasonal patterns that exist in the monthly data to generate more accurate forecasts.

16.36a.



No evidence of a strong seasonal pattern is shown.

b. Seasonally adjusted data:

Period	4 Period MA	$100\frac{X_t}{X_t^*}$	Seas. Factor	Adj. Series
1-1			89.385	.879
2			61.016	1.095
3	.783	110.217	140.489	.614
4	.786	102.770	109.110	.740
2-1	.788	101.744		.897
2	.791	87.730		1.098
3	.763	116.046		.630
4	.704	114.367		.738
3-1	.675	85.730		.648
2	.684	61.887		.693
3	.671	134.800		.644
4	.650	130.873		.780
4-1	.676	63.657		.481
2	.716	57.133		.670
3	.761	147.272		.797
4	.798	120.031		.878
5-1	.813	83.615		.761
2	.806	57.072		.754
3	.801	148.611		.847
4	.809	102.596		.761
6-1	.785	97.549		.857
2	.739	59.540		.721
3	.705	144.784		.726
4	.715	88.112		.577
7-1	.749	92.154		.772
2	.769	78.049		.983
3				.804
4				.623

16.38

Winters' Method for Quarterly Earnings

```
Multiplicative Method


Data    Quarterly Earnings
Length  28

Smoothing Constants
Alpha (level)     0.6
Gamma (trend)     0.6
Delta (seasonal)  0.8


Accuracy Measures
MAPE  0.127452
MAD   0.001008
MSD   0.000007


Forecasts

Period  Forecast     Lower     Upper
29      0.795523  0.793053  0.797992
30      0.672838  0.669713  0.675962
31      0.867631  0.863710  0.871552
```

16.40

Regression Analysis for the period 1980, first quarter, through 2000, fourth quarter : Prime Rate versus primeratelag1

```
The regression equation is
Prime Rate = 0.741 + 0.917 primeratelag1
83 cases used, 1 cases contain missing values
Predictor          Coef   SE Coef      T      P
Constant         0.7412    0.3801   1.95  0.055
primeratelag1   0.91661   0.03661  25.04  0.000

S = 1.05843   R-Sq = 88.6%   R-Sq(adj) = 88.4%
```

Years	Actual values	Predicted values
2001-I	8.623333	9.448995
2001-II	7.34	9.402243
2001-III	6.566667	9.35939
2001-IV	5.156667	9.320111
2002-I	4.75	9.284107
2002-II	4.75	9.251105
2002-III	4.75	9.220855
2002-IV	4.45	9.193128
2003-I	4.25	9.167713
2003-II	4.24	9.144418
2003-III	4	9.123065
2003-IV	4	9.103492

There is difference in the actual value and predicted value of prime interest rate in years 2001-2003. The difference is due to the use of predicted values of the previous quarters to obtain the predicted values of the subsequent quarters.

Regression Analysis for the period 1980, first quarter, through 2007, fourth quarter: Prime Rate versus primeratelag1

```
The regression equation is
Prime Rate = 0.500 + 0.935 primeratelag1
111 cases used, 1 cases contain missing values
Predictor         Coef  SE Coef      T      P
Constant        0.5004   0.2600   1.93  0.057
primeratelag1  0.93502  0.02728  34.27  0.000

S = 0.953416   R-Sq = 91.5%   R-Sq(adj) = 91.4%
```

Years	Actual values	Predicted values
2008-I	6.213333	7.534836
2008-II	5.08	7.545622
2008-III	5	7.555708
2008-IV	4.056667	7.565138
2009-I	3.25	7.573955
2009-II	3.25	7.5822
2009-III	3.25	7.589908
2009-IV	3.25	7.597116

There is difference in the actual value and predicted value of prime interest rate in years 2008-2009. The difference is due to the use of predicted values of the previous quarters to obtain the predicted values of the subsequent quarters.

16.42

Regression Analysis for the period 1965, first quarter, through 2000, fourth quarter: Fixed investment versus Fixedinvstlag1

```
The regression equation is
Fixed investment = - 8.23 + 1.02 Fixedinvstlag1
143 cases used, 1 cases contain missing values
Predictor           Coef  SE Coef       T      P
Constant          -8.233    3.916   -2.10  0.037
Fixedinvstlag1   1.02179  0.00415  246.20  0.000

S = 18.2512   R-Sq = 99.8%   R-Sq(adj) = 99.8%
```

Years	Actual values	Predicted values
2001-I	1921.861	1967.206
2001-II	1891.799	2001.838
2001-III	1872.035	2037.225
2001-IV	1825.069	2073.383
2002-I	1808.845	2110.33
2002-II	1801.582	2148.081
2002-III	1798.418	2186.654
2002-IV	1783.867	2226.068
2003-I	1784.023	2266.341
2003-II	1829.625	2307.492
2003-III	1888.403	2349.539
2003-IV	1922.142	2392.503

There is difference in the actual value and predicted value of Fixed investment in years 2001-2003. The difference is due to the use of predicted values of the previous quarters to obtain the predicted values of the subsequent quarters.

Regression Analysis for the period 1965, first quarter, through 2007, fourth quarter: Fixed investment versus Fixedinvstlag1

```
The regression equation is
Fixed investment = 3.48 + 1.01 Fixedinvstlag1


171 cases used, 1 cases contain missing values


Predictor           Coef  SE Coef       T      P
Constant           3.475    3.638    0.96  0.341
Fixedinvstlag1   1.00621  0.00308  326.55  0.000


S = 21.6004   R-Sq = 99.8%   R-Sq(adj) = 99.8%
```

Years	Actual values	Predicted values
2008-I	2082.02	2131.7
2008-II	2057.927	2148.413
2008-III	1993.946	2165.229
2008-IV	1856.041	2182.15
2009-I	1663.524	2199.176
2009-II	1619.512	2216.308
2009-III	1622.145	2233.547
2009-IV	1616.745	2250.892

There is difference in the actual value and predicted value of Fixed investment in years 2008-2009. The difference is due to the use of predicted values of the previous quarters to obtain the predicted values of the subsequent quarters.

Chapter 17:

Additional Topics in Sampling

17.2 a. $\bar{x}_3 = 43.3$, $\hat{\sigma}^2_{\bar{x}_3} = \frac{s_3^2}{n_3}\frac{N_3 - n_3}{N_3 - 1} = \frac{(12.3)^2}{50}\frac{208-50}{208-1} = 2.31$

90% confidence interval: $43.3 \pm 1.645\sqrt{2.31}$: 40.799 up to 45.800

b. $\bar{x}_{st} = \frac{1}{N}\sum_{j=1}^{k} N_j\bar{x}_j = \frac{152(27.6)+127(39.2)+208(43.3)}{487} = 37.3306$

c. $\hat{\sigma}^2_{\bar{x}_1} = \frac{(7.1)^2}{40}\frac{152-40}{152-1} = .9350$, $\hat{\sigma}^2_{\bar{x}_2} = \frac{(9.9)^2}{40}\frac{127-40}{127-1} = 1.692$

$$\hat{\sigma}^2_{\bar{x}_{st}} = \frac{(152)^2(.9350)+(127)^2(1.692)+(208)^2(2.31)}{(487)^2} = .6275$$

90% confidence interval: $37.3306 \pm 1.645\sqrt{.6275}$: 36.0275 up to 38.6337

95% confidence interval: $37.3306 \pm 1.96\sqrt{.6275}$: 35.7799 up to 38.8833

17.4 a. $\hat{\sigma}^2_{\bar{x}_1} = \frac{s_1^2}{n_1}\frac{N_1 - n_1}{N_1 - 1} = \frac{(1.04)^2}{50}\frac{632-50}{632-1} = .02$; $3.12 \pm 1.96\sqrt{.02}$:
2.843 up to 3.397

b. $\hat{\sigma}^2_{\bar{x}_2} = \frac{(.86)^2}{50}\frac{529-50}{529} = .013$; $3.37 \pm 1.96\sqrt{.013}$: 3.147 up to 3.593

c. $\hat{\sigma}^2_{\bar{x}_{st}} = \frac{(632)^2(.0199)+(529)^2(.0134)}{(1161)^2} = .0086$; $\bar{x}_{st} = 3.2339$

$3.2339 \pm 1.96\sqrt{.0086}$: 3.0519 up to 3.4159

17.6 a. $\bar{x}_{st} = \frac{237(120)+198(150)+131(180)}{450} = 181.6$ $N\bar{x}_{st}$

---

b. $\hat{\sigma}^2_{\bar{x}_1} = \frac{s_1^2}{n_1}\frac{N_1 - n_1}{N_1} = \frac{93^2}{40}\frac{120-40}{120-1} = 145.3610$,

$\hat{\sigma}^2_{\bar{x}_2} = \frac{64^2}{45}\frac{150-45}{150-1} = 64.143$, $\hat{\sigma}^2_{\bar{x}_3} = \frac{47^2}{50}\frac{180-50}{180-1} = 32.086$

$$\hat{\sigma}^2_{\bar{x}_{st}} = \frac{(120)^2(145.3610)+(150)^2(64.143)+(180)^2(32.086)}{(450)^2} = 22.5975$$

95% confidence interval: $181.6(450) \pm 1.96\sqrt{22.5975}\,(450)$:
$77{,}527.2479 < N\mu < 85{,}912.7521$

17.8 a. $\hat{p}_{st} = [100\frac{6}{25} + 50\frac{14}{25}]/150 = .3467$

b. $\hat{\sigma}^2_{\hat{p}_1} = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1 - 1}\frac{N_1 - n_1}{N_1} = \frac{.24(.76)}{25-1}\frac{100-25}{100-1} = .0057$

$\hat{\sigma}^2_{\hat{p}_2} = \frac{.56(.44)}{25-1}\frac{50-25}{50-1} = .0052$,

$\hat{\sigma}^2_{\hat{p}_{st}} = \frac{(100)^2(.0056) + (50)^2(.0051)}{(150)^2} = .0031$

90% confidence interval: $.3467 \pm 1.645\sqrt{.0031}$: .2550 up to .4383

95% confidence interval: $.3467 \pm 1.96\sqrt{.0031}$: .2375 up to .4559

17.10 a. $n_3 = \frac{208}{487}130 = 55.52 = 56$ observations

b. $n_3 = \left[\frac{208(12.3)}{152(7.1) + 127(9.9) + 208(12.3)}\right]130 = 67.95 = 68$ observations

17.12 a. $n_1 = \frac{632}{1161}100 = 54.43 = 55$ observations

b. $n_1 = \left[\frac{632(1.04)}{632(1.04) + 529(.86)}\right]100 = 59.09 = 60$ observations

17.14 a. $n_2 = \frac{1031}{1395}100 = 73.91 = 74$ observations

b. $n_2 = \left[\frac{1031(219.9)}{364(87.3) + 1031(219.9)}\right]100 = 87.71 = 88$ observations

17.16 Proportional allocation: $\sigma_{\bar{x}} = \frac{500}{1.96} = 255.1020$

$\sum N_j \sigma^2{}_j = 1150(4000)^2 + 2120(6000)^2 + 930(800000)^2 = 15424 \times 10^7$

$n = \frac{15424 \times 10^7}{4200(255.1020)^2 + 15424 \times 10^7 / 4200} = 497.47 = 498$ observations

Optimal allocation:

$\sum N_j \sigma_j = 1150(4000) + 2120(6000) + 930(8000) = 24760000$

$n = \frac{(24760000)^2 / 4200}{4200(255.1020)^2 + 15424 \times 10^7 / 4200} = 470.78 = 471$ observations

17.18 a. $\bar{x}_c = \dfrac{69(83)+75(64)+\cdots+71(98)}{497} = 91.6761$

b.

$$\hat{\sigma}^2_{\bar{X}_c} = \frac{(52-8)}{52(8)(62.125)^2}\,\frac{(69)^2(83-91.67605634)^2+\cdots+(71)^2(98-91.67605634)^2}{8-1}$$
$$= 66.409$$

99% confidence interval: $91.6761 \pm 2.58\sqrt{66.4090}$
70.6920 up to 112.6602

17.20 a. $\hat{p}_c = \dfrac{24+\cdots+34}{497} = .4507$

b.

$$\hat{\sigma}^2_{\hat{p}_c} = \frac{52-8}{52(8)(62.125)^2}\,\frac{(69)^2\left(\frac{24}{69}-.4507\right)^2+\cdots+(71)^2\left(\frac{34}{71}-.4507\right)^2}{8-1} = .0013$$

95% confidence interval: $.4507 \pm 1.96\sqrt{.0013}$: .38 up to .5214

17.22 $\sigma_{\bar{x}} = \dfrac{5000}{1.645} = 3039.5$, $n = \dfrac{720(37600)^2}{719(3039)^2+(37600)^2} = 126.34 = 127$
observations.
Additional sample observations needed are 127-20 = 107

17.24 $\sigma_{\bar{x}} = \dfrac{20}{1.96} = 10.2$

$$n = \frac{(100(105)+180(162)+200(183))^2/480}{480(10.2)^2+(100(105)^2+180(162)^2+200(183)^2)/480} = 159.35 = 160$$
observations. Additional sample observations needed: 160-30 = 130

17.26 a. $\hat{p} = \dfrac{38}{61} = .623$, $\hat{\sigma}^2_{\hat{p}} = \dfrac{.623(.377)}{61-1}\dfrac{100-61}{100-1} = .0015$

90% confidence interval: $.623 \pm 1.645\sqrt{.0015}$: .559 up to .687

b. If the sample information is not randomly selected, the resulting conclusions may be biased

17.28 a. $\bar{x}_{st} = \dfrac{120(1.6)+180(.74)}{300} = \dfrac{325.2}{300} = 1.084$

$$\hat{\sigma}^2_{\bar{x}_1} = \frac{s^2}{n}\frac{N-n}{N-1} = \frac{(.98)^2}{20}\frac{120-20}{120-1} = .0403$$

$$\hat{\sigma}^2_{\bar{x}_2} = \frac{(.56)^2}{20}\frac{180-20}{180-1} = .0140,$$

$$\hat{\sigma}^2_{\bar{x}_{st}} = \frac{(120)^2(.0403)+(180)^2(.0140)}{(300)^2} = .0114$$

95% confidence interval for mean number of errors per page in the book: $1.084 \pm 1.96\sqrt{.0114}$: .8744 up to 1.2936

b. $N\bar{x}_{st} - Z_{\alpha/2}N\hat{\sigma}_{\bar{x}_{st}} < N\mu < N\bar{x}_{st} + Z_{\alpha/2}N\hat{\sigma}_{\bar{x}_{st}}$

where, $N\bar{x}_{st} = (120)(1.6)+(180)(.74) = 325.2$

$N^2\hat{\sigma}^2_{\bar{x}_{st}} = (120)^2.0400 + (180)^2.0140 = 1{,}029.600$

$N\hat{\sigma}_{\bar{x}_{st}} = \sqrt{1029.600} = 32.08738$, $325.2 \pm 2.58(32.08738)$

99% confidence interval for the total number of errors in the book: 242.446 up to 407.9854 or from 243 total errors up to 408 total errors.

17.30 a. $n_1 = \dfrac{352}{970}80 = 29.03 = 30$ observations

b. $n_1 = \left[\dfrac{352(4.9)}{352(4.9)+287(6.4)+331(7.6)}\right]80 = 22.7 = 23$ observations

17.32

The population of students whose views are to be canvassed could be divided into two strata—business majors and nonbusiness majors. Less straightforward stratification is also possible. Suppose that, on some other topic, you believe that gender and class year (senior, junior, sophomore, or first-year) are both potentially relevant. In that case, to satisfy the requirement that the strata be mutually exclusive and collectively exhaustive, eight strata—senior women, senior men, and so on—are needed.

ISBN-13: 978-0-13-274568-0
ISBN-10: 0-13-274568-2

www.pearsonhighered.com

PEARSON

ALWAYS LEARNING