INSTRUCTOR'S SOLUTIONS MANUAL

David S. Rubin
University of North Carolina at Chapel Hill

STATISTICS FOR MANAGEMENT

Seventh Edition

Richard I. Levin
David S. Rubin

Prentice Hall, Upper Saddle River, New Jersey 07458

Acquisitions editor: Tom Tucker
Associate editor: Audrey Regan
Project editor: Richard Bretan
Manufacturing buyer: Paul Smolenski

A Simon & Schuster Company
Upper Saddle River, New Jersey 07458

Printed in the United States of America

10 9 8 7 6 5 4 3 2

ISBN 0-13-619552-0

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

CONTENTS

NOTE TO INSTRUCTORS

From personal experience, I realize that one of the greatest frustrations faced by instructors and students of elementary statistics courses is encountering mistakes in the selected brief solutions in the back of the text and in the complete solutions worked out in the Instructor's Manual. To minimize that frustration, every effort has been to make these solutions as error-free as possible. However, it would be both immodest and foolish to assert that both my dedicated typist and I have made no errors in assembling this volume.

Fortunately, the wonders of modern technology have enabled us to place the entire manuscript on a small number of disks, and it is easy to make corrections. I view you, my fellow instructors, as partners in this task, and I seek the following help from you. If, as I am afraid inevitably must happen, you encounter any errors in these solutions, please drop me a note at the address below. In turn, I will correct these solutions and send you new pages to replace the incorrect ones in this manual. Those corrected pages will also be sent to Prentice-Hall for circulation to all adopters of the text.

I thank you for your understanding, forebearance, and assistance.

David S. Rubin
CB #3490, McColl Building
University of North Carolina
Chapel Hill, NC 27599–3490
david_rubin@unc.edu

WORKED OUT

ANSWERS

TO ALL

EXERCISES

CHAPTER 1

INTRODUCTION

There are

NO

problems

in

Chapter 1

Please move on

to

Chapter 2

CHAPTER 2

ARRANGING DATA TO CONVEY MEANING: TABLES AND GRAPHS

2-1 It is very unlikely that the company making this claim actually interviewed every doctor in the United States (or in the world, for that matter) before coming to this conclusion. Instead, they most likely interviewed a specific number of doctors living in different parts of the country, residing in both large cities and small towns, having different educational backgrounds, etc. From this sample then, they drew the conclusion stated in the problem.

2-2 Since the Department of Commerce keeps statistics on all the cars sold in the U.S., this conclusion is drawn from a <u>population</u>.

2-3 From the strength of the data, it would seem that the president is premature in his actions. We are not given any information about any favorable letters concerning the new amplifier, but even without any other letters, three complaints out of 10,000 sold is <u>not</u> a bad record. Certainly, the letters which are mentioned come from a <u>biased source</u>. There is also no supporting evidence for these letters. The company most probably did extensive testing of the new products, and the complaints may, in fact, <u>contradict</u> this test evidence. As already mentioned, there is little other <u>support</u> for these three complaints. Most importantly, the size of the sample is very small, considering the number of units sold. We can conclude that the president may have reached a false conclusion concerning the new amplifier, given the data which he has received.

2-4 On the basis of German history since the end of World War II, and given the bias produced by his own strong belief in the validity of Communism, Ulbricht was unable to foresee the possibility of the changes that resulted from Gorbachev's hands-off policy toward the eastern European satellite nations.

2-5 Although the information given in the text is limited we can still make a determination for most of the tests. As far as can be determined, the sampling would be <u>unbiased</u>, provided that the sample gallons are taken from the larger population of gallons of water on a random basis. We do not have the information to determine if this set of data <u>supports</u> similar data collected on other days, but we can conclude that most likely there is no <u>missing evidence</u>--given that the gallons are selected at random. The <u>size</u> of the sample is probably sufficient to honestly represent the true characteristics of the actual population. Finally, we can use this data to draw <u>honest conclusions</u> about the general population.

2-6 The point of this section is to show that the raw data are not in a useful form. We cannot draw any conclusions from these data in their current form. We would first need to do a certain amount of rearranging, such as listing the grades from highest to lowest or determining the most frequent grade pair, before we could begin to make meaningful observations about the problem at hand.

2-7 Yes, this is an example of raw data in that the information that the manager receives is unaltered by statistical methods. The manager might want to determine volumes, or compare these figures with those of previous months to determine the trend in sales. These operations would be examples of some statistical methods which could be applied to this raw data.

2-8 Here is a clear-cut case of data which has already undergone statistical analysis. In this case, the raw data would be a list of sample units indicating whether or not they were defective. The quality control [illegible] which has already performed an analysis on these data to [illegible] Anytime you are given "averages" or "rates" or "highest/lowest" figures [illegible] with data which have already been statistically reduced.

2-9 In ascending order, the data array is

217	308	321	349	360	427	468	555	586	588
634	641	648	669	722	752	766	823	847	904

Seven stores are not breaking even; five will get bonuses.

2-10 In addition to the 7 stores with under 475 service actions which are not breaking even, another 6 stores fall on the "store watch list."

2-11

Class	2.0 - 2.9	3.0 - 3.9	4.0 - 4.9	5.0 - 5.9	6.0 - 6.9
Frequency	5	6	4	2	3

Almost all of the jobs are done in under six hours, and fully 75% of them are done in five hours or less. Since 15% take more than 6.0 hours, there is a minor productivity problem here that merits some attention.

2-12

5 Intervals

Class	Frequency	Relative Frequency
15 - 25	3	0.0667
26 - 36	4	0.0889
37 - 47	12	0.2667
48 - 58	18	0.4000
59 - 69	8	0.1778
	45	1.0000

11 Intervals

Class	Frequency	Relative Frequency
15 - 19	3	0.0667
20 - 24	0	0.0000
25 - 29	2	0.0444
30 - 34	2	0.0444
35 - 39	5	0.1111
40 - 44	3	0.0667
45 - 49	11	0.2444
50 - 54	3	0.0667
55 - 59	8	0.1778
60 - 64	3	0.1111
65 - 69	3	0.0667
	45	1.0000

a) No: we can see from either distribution that more than 10% of the motorists drive at 55 mph or more. (5 intervals ⇒ over 17.78%; 11 intervals ⇒ 35.56%)

b) Either can be used; the 11-interval distribution gives a more precise answer.

c) The 5-interval distribution shows that 66.67% of the motorists drive between 37 and 58 mph inclusive.

2-13 Table 2-2 arranged from highest to lowest:

lbs/sq in	lbs/sq in	lbs/sq in	lbs/sq in
2509.5	2502.5	2500.0	2496.9
2508.4	2502.3	2499.9	2496.9
2508.2	2502.2	2499.7	2496.7
2508.1	2502.0	2499.2	2495.3
2506.4	2501.3	2498.4	2493.8
2505.0	2500.8	2498.3	2493.4
2504.1	2500.8	2498.1	2491.6
2503.7	2500.8	2497.8	2491.3
2503.2	2500.7	2497.8	2490.5
2502.5	2500.2	2497.1	2490.4

a) 19 samples have a strength less than 2500.00 lbs/sq in [illegible] samples have a strength between 2497 and 2504 lbs/sq in

c) 2502.5: twice; 2500.8: 3 times; 2497.8: twice, 2496.9: twice

2-14 Sorting the data in ascending order by number of channels purchased gives us:

Channels purchased	17	18	22	25	28	29	39	42	43	76	84	96	104
Hours watched	19	16	13	14	13	7	9	12	16	8	4	6	6

This shows that hours watching television decreases as number of channels purchased increases. Sorting by number of hours watched leads to the same conclusion.

2-15 a)

Sample	Pollution Rating (ppm)
3	68.4
9	60.0
4	54.2
8	52.7
2	51.7
5	49.9
12	49.1
10	46.1
7	39.8
11	38.5
1	37.2
6	33.4

b)

Class	Frequency
30.0 - 39.9	4
40.0 - 49.9	3
50.0 - 59.9	3
60.0 - 69.9	2

c) 8 samples have excessive pollution.

d) The largest distance between successive samples is 8.4 ppm, between samples 3 and 9.

2-16 a) First we will construct the raw data from the information in the problem.

Spread	SAT Differential	Spread	SAT Differential
− 1.1	− 120	− 0.2	20
0.1	− 20	0.1	− 10
− 0.5	− 30	0.1	20
− 0.5	− 90	0.3	50
0.2	0	0.4	60
0.1	− 10	0.6	60
0.3	60	0.0	10
1.3	140	− 0.6	− 120
− 0.2	0	− 0.7	− 100
− 0.1	− 10	0.8	150

Now, we will array these data by highest to lowest spreads.

Spread	SAT Differential	Spread	SAT Differential
1.3	140	0.1	− 20
0.8	150	0.0	10
0.6	60	− 0.1	− 10
0.4	60	− 0.2	0
0.3	60	− 0.2	0
0.3	50	− 0.5	− 30
0.2	20	− 0.5	− 90
0.1	20	− 0.6	− 120
0.1	− 10	− 0.7	− 100
0.1	− 10	− 1.1	− 120

b) From the array we can see that the most common spread is 0.1, which occurs four times.

c) For a spread of 0.1, the most common SAT differential is − 10, which occurs twice out of the four data points.

[illegible]	Frequency	Relative Frequencies
1 to 2	2	.10
3 to 4	4	.20
5 to 6	6	.30
7 to 8	6	.30
9 to 10	2	.10
	20	1.00

c) In order to be able to make comparisons of the frequency distributions

d)

"Change" Classes	Frequency	Relative Frequencies
− 5 to − 4	1	.05
− 3 to − 2	0	.00
− 1 to 0	5	.25
1 to 2	8	.40
3 to 4	5	.25
5 to 6	1	.05
	20	1.00

e) Sales appear to have increased, but the apparent increase could be due to other factors we don't know about, so we can't say for sure that the new slogan has helped.

2-21 a) It is hard to tell anything from the raw data.

b)

Class	10-19	20-29	30-39	40-49	50-59	60-69
Frequency	3	7	7	7	3	3

Most video recorders are bought by people between 20 and 50, so marketing efforts should be aimed at that age group.

2-22

Class	< 25	25-34	35-44	45-54	≥55
Frequency	6	9	7	3	5
Relative Frequency	.200	.300	.233	.100	.167

a) Now we see that most purchasers are under 45.

b) We can be more precise: about 75% of the purchasers are under 45.

2-23

Class		Frequency
0.00 -	0.99	50
1.00 -	9.99	1000
10.00 -	18.99	3100
19.00 -	27.99	2590
28.00 -	36.99	2090
37.00 -	45.99	800
46.00 -	48.99	90

2-24 a) No. Since most of the runners will take between 5 and 6 minutes per mile, almost all of the observations will fall into the two lower classes. As a result, the coach won't get much information from this distribution.

b) Five classes with midpoints at 25, 27, 29, 31, and 33 would be much more helpful.

2-25 The intervals have unequal sizes. The frequencies sum to 140, so either 40 observations are missing or the total (180) is incorrect. The last interval is not needed, because its frequency is 0.

2-26 To construct a closed classification, we must insure that the list is all-inclusive. A distribution containing the following categories would meet this requirement: Single, Married, Divorced, Separated, Widowed.

d) The SAT differential appears to be a good indicator of [illegible]: students with high SAT differentials do better in college than in high school, and students with large negative differentials do worse.

2-17

Class	Frequency	Relative Frequency
2 - 6	3	0.1667
7 - 11	2	0.1111
12 - 16	4	0.2222
17 - 21	4	0.2222
22 - 26	3	0.1667
27 - 31	2	0.1111
	18	1.0000

a) From either distribution, we can see that only half of the shifts had 16 or fewer burgers wasted. The goal has not been met.

b) From the relative frequency distribution, we can see immediately that 21 or fewer burgers are wasted in 72.22% of the shifts. The same results can be obtained from the frequency distribution.

2-18

Classes (lbs/sq in)	Relative Frequencies
2490.0 - 2493.9	.150
2494.0 - 2497.9	.175
2498.0 - 2501.9	.325
2502.0 - 2505.9	.225
2506.0 - 2509.9	.125
	1.000

The greatest number of these samples (32.5%) fell into the class 2498.0 - 2501.9 lbs/sq in. This would have been somewhat difficult to see from the data in the table.

2-19 a)

− 0.5	− 0.4	− 0.3	− 0.3	− 0.3	− 0.1	0.0	0.0	0.0	0.1
0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3
0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.7	0.7	1.0

b)

Class	− 0.5 to − 0.2	− 0.1 to 0.2	0.3 to 0.6	0.7 to 1.0
Frequency	5	12	10	3
Relative Frequency	.167	.400	.333	.100

c) From the data array, we see that 3 did not change, whereas 29 increased (that is, changed) less than 1%; hence 29 did not change or increased less than 1.0%. (If we interpret "increased less than 1%" to mean "increased, but the increase was less than 1.0%, then the answer is 3 did not change, 20 increased less than 1.0%; hence 23 did not change or increased less than 1.0%.)

d) In theory, the data could have been measured to any degree of precision, so theoretically they are continuous. Since, in practice, they were measured to the nearest tenth of a percent, they should be treated as discrete data.

2-20 a)

"Before" Classes	Frequency	Relative Frequencies
1 to 2	5	.25
3 to 4	6	.30
5 to 6	7	.35
7 to 8	2	.10
9 to 10	0	.00
	20	1.00

To construct an open-ended classification, we can use any of the categories listed above plus the category "other". Since the problem asks for only three categories, we can choose any two of the above. The logical choice would be to use the following: Single, Married, Other

Notice that as you open the distribution you lose completeness, but you gain simplicity.

2-27 a) Qualitative, discrete, open-ended: A listing of industries is obviously non-numerical and would therefore be discrete. Because of the large number of industries and their variability, the distribution would most likely be open-ended.

b) Quantitative, discrete, open-ended: The closing bid is obviously numerical and would be discrete. Again, the amount of data would probably warrant an open-ended distribution. (Prices are usually restricted to 1/8's).

c) Quantitative, discrete, open-ended: The change is also numerical and discrete. Again the quantity of data would probably warrant an open-ended distribution. (Prices are usually restricted to 1/8's.)

The answer to part (c) would change to qualitative, discrete and closed. The only choices would be "higher", "lower", or "unchanged". These choices form a closed system; there are no others. These choices are also non-numerical and therefore discrete.

2-28 The classes run from 85 to 114 and from 115 to 144, with additional open-ended classes at the bottom (84 and under) and at the top (145 and over). Since the group wants to highlight the noisy flights, this distribution is inadequate, because the 115 to 144 class includes noise levels on both sides of the 140 decibel limit.

2-29 Probably not. She would like to further subdivide the noisy flights.

2-30 Looking at the data, we see that 22 days is the minimum and 51 days is the maximum. Using equation 2-1, the intervals will be

$$\frac{51 - 22}{10} = 2.9, \text{ or approximately 3 days wide for part (a)}$$

and

$$\frac{51 - 22}{5} = 5.8, \text{ or approximately 6 days wide for part (b).}$$

a)

Waiting Time (days)	Frequency
22 - 24	3
25 - 27	3
28 - 30	6
31 - 33	12
34 - 36	8
37 - 39	6
40 - 42	5
43 - 45	4
46 - 48	2
49 - 51	1
	50

b)

Waiting Time (days)	Frequency
22 - 27	6
28 - 33	18
34 - 39	14
40 - 45	9
46 - 51	3
	50

With ten divisions the 31-33 interval has the highest number of observations with 12; however, when only five divisions are used, then the 28-33 interval has the most observations with 18.

c) Yes, since he wants to know the relative proportions at each level.

2-31 a) Arranging the data into a frequency distribution by one-sale intervals, we can see that the most frequent numbers of sales are 5, 8, and 10.

Sales	5	6	7	8	9	10	11	12	13	14	15
Frequency	7	4	4	8	3	6	2	2	1	2	1

b) Since we cannot form an evenly-spaced distribution having each of these numbers as marks, we will choose 5 and 8, since these are the most frequent. With these two marks, our intervals will have to be 3 units wide and start at 4.

Sales	4-6	7-9	10-12	13-15	Total
Frequency	11	15	10	4	40
Relative Frequency	.275	.375	.250	.100	1.00

c) Only (b) helps, since he needs frequency information, not information about what the classes are.

2-32 a) Discrete and closed

b) Discrete and closed

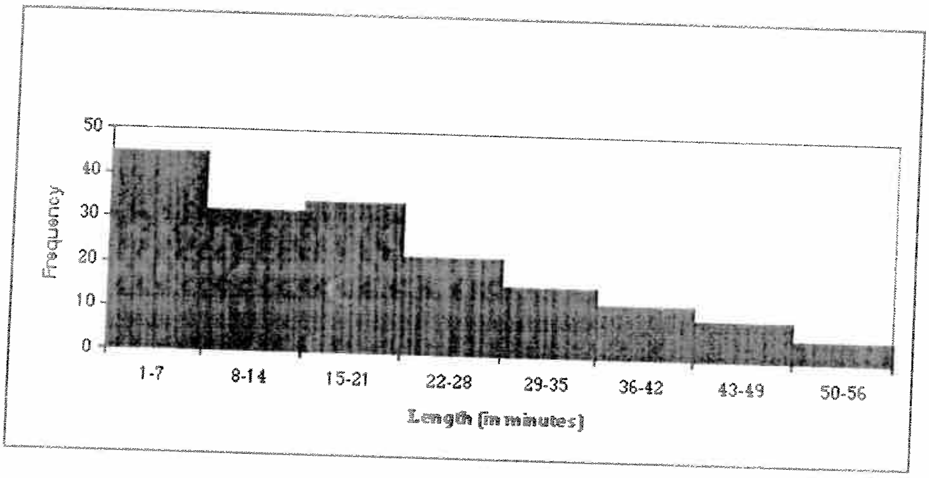
c) Flavor is qualitative, amount is quantitative

d) He should collect data on how often the stores ran out of the various flavors and how much of each flavor was left over.

2-33 a) 75 to 94, 95 to 114, 115 to 134, 135 to 154, 155 to 174, and 175 to 194

b) 23 to 30, 31 to 38, 39 to 46, 47 to 54, 55 to 62, and 63 to 70

2-34

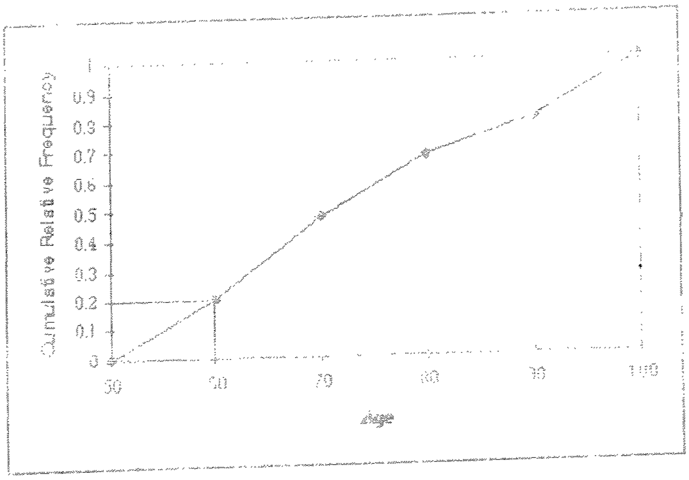


a) With one minor exception (the 15-21 minute interval), the frequencies decrease as the length of the calls increases.

b) Let the people with the longest calls go first. Whenever a phone becomes available, let a person in the highest class with calls still remaining make the next call.

c) Yes. If we allowed all the 1-7 minute calls to be made first and then worked upward, the 50-56 minute calls would have to wait until the shorter calls were made. Using the order suggested in (b) can save time by allowing shorter calls to be made while longer calls are in progress.

Class	50 - 59	60 - [illegible]	[illegible]	[illegible]	[illegible]
Frequency	5	7	5	[illegible]	5
Cumulative Frequency	5	12	17	20	25
Cumulative Relative Frequency	.20	[illegible]	.68	.80	1.00

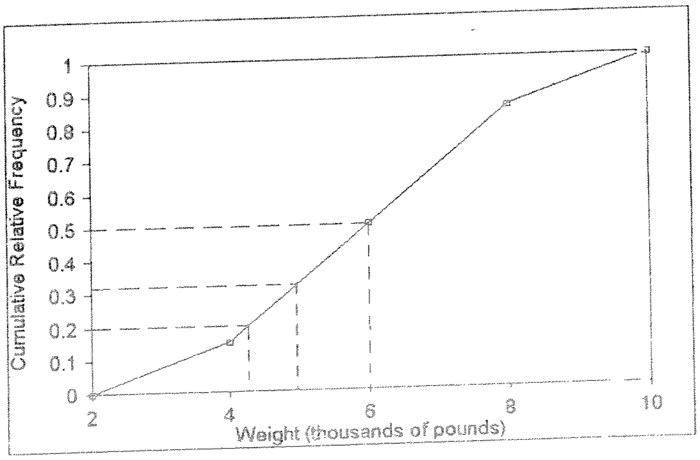


a) About 80% will be eligible.

b) The fee will have to be increased to about $60.

2-36

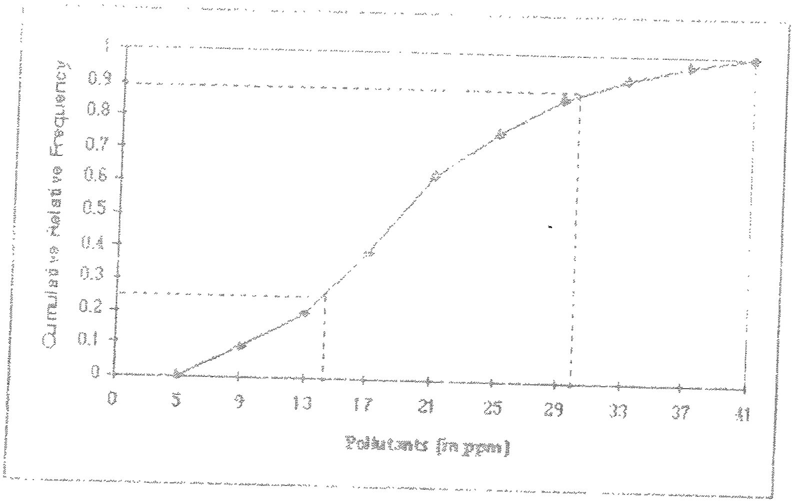
Class	Frequency	Cumulative Relative Frequency
2000 - 3999	3	.15
4000 - 5999	7	50
6000 - 7999	7	.85
8000 - 9999	3	1.00



a) About 65% of all trips break even by catching at least 5000 pounds.

b) The middle value is approximately 6000 pounds.

c) 80% of the catches are greater than approximately 4300 pounds.

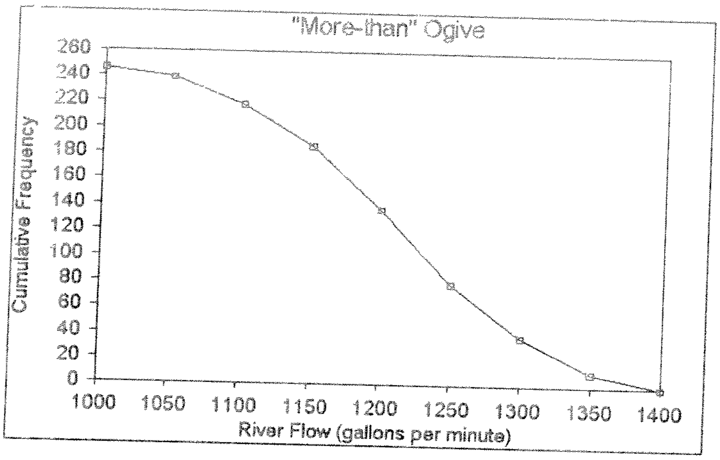


a) 25% of the observations fall below about 14 ppm.

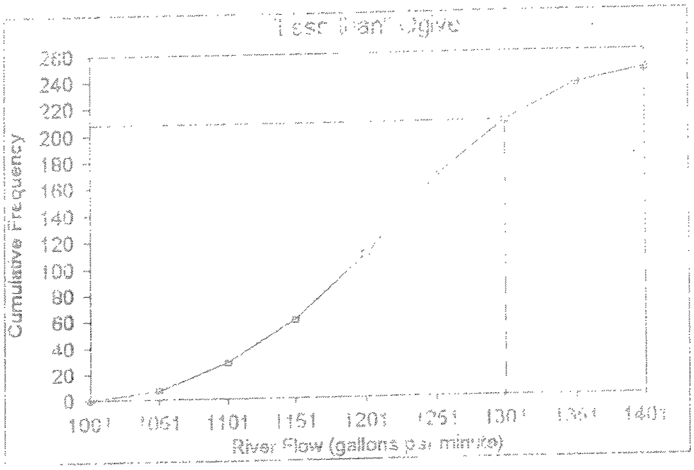
b) Approximately 11% of the sites will be heavily monitored.

2-38 a)

River Flow (>)	1000	1050	1100	1150	1200	1250	1300	1350	1400
Cum. Frequency	246	239	218	186	137	79	38	11	0



Cum. Frequency



c) About 85%.

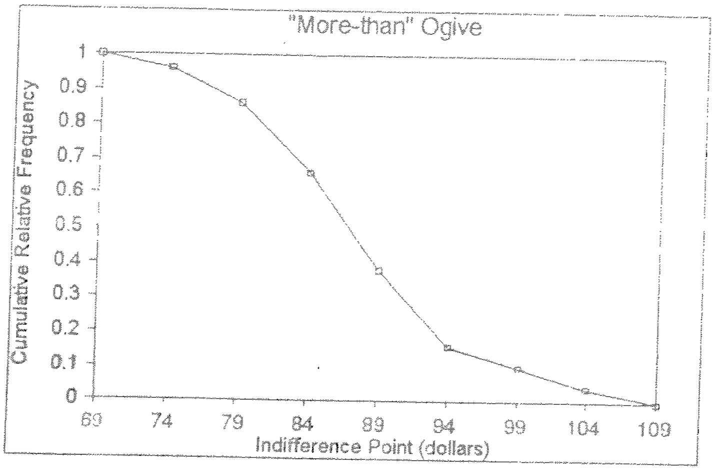
2-39 First, we will construct a relative frequency distribution from the data given in the problem, adding an additional class to each end.

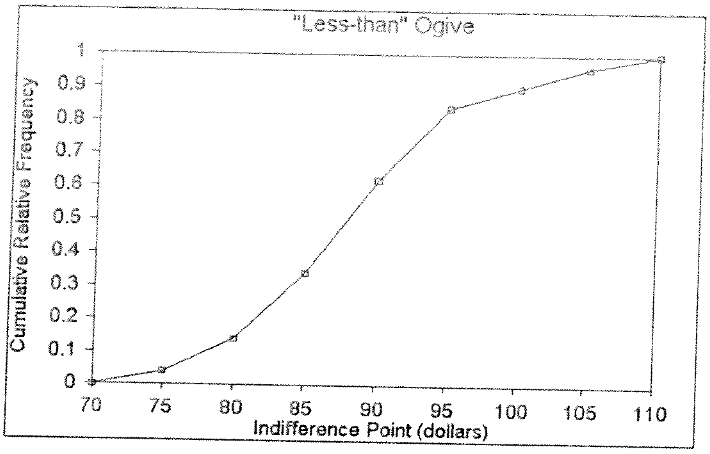
Indifference Point	Relative Frequency
\$ 65 - 69	0
70 - 74	.04
75 - 79	.10
80 - 84	.20
85 - 89	.28
90 - 94	.22
95 - 99	.06
100 - 104	.06
105 - 109	.04
110 - 114	0
	1.00

a) Now we will use this relative frequency distribution to construct the cumulative frequency distribution.

Indifference Point (more than)	Cumulative Frequency	Indifference Point (less than)	Cumulative Frequency
\$ 69	1.00	\$ 70	0
74	.96	75	.04
79	.86	80	.14
84	.66	85	.34
89	.38	90	.62
94	.16	95	.84
99	.10	100	.90
104	.04	105	.96
109	0	110	1.00

b)





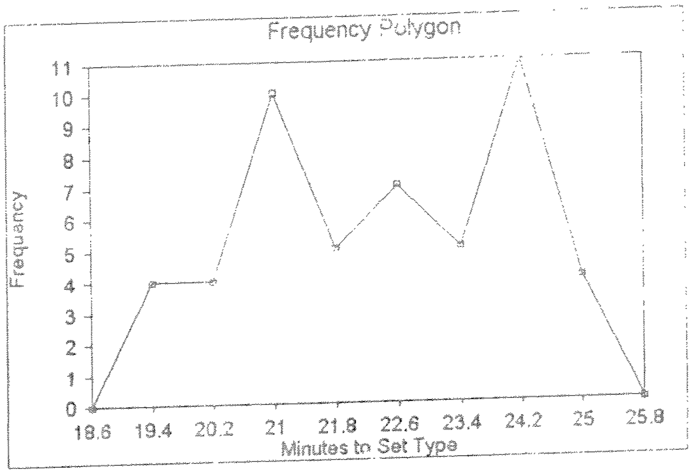
2-40 a)

19.0	20.7	21.8	23.1	24.1
19.5	20.8	21.9	23.3	24.1
19.5	20.9	22.0	23.5	24.2
19.7	20.9	22.2	23.6	24.2
19.8	20.9	22.5	23.7	24.2
19.9	21.1	22.7	23.8	24.3
20.1	21.2	22.8	23.8	25.0
20.3	21.3	22.8	23.8	25.0
20.7	21.5	22.8	23.9	25.1
20.7	21.6	22.9	23.9	25.3

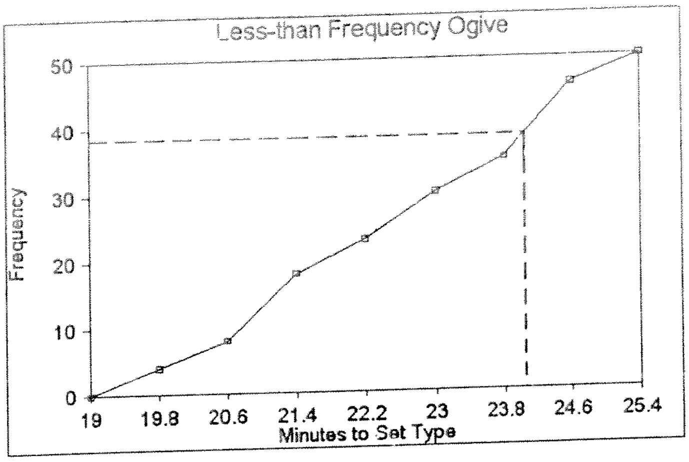
b)

Minutes to Set Type	Frequency	Minutes to Set Type ($\leq$)	Frequency
19.0 - 19.7	4	19.0	0
19.8 - 20.5	4	19.8	4
20.6 - 21.3	10	20.6	8
21.4 - 22.1	5	21.4	18
22.2 - 22.9	7	22.2	23
23.0 - 23.7	5	23.0	30
23.8 - 24.5	11	23.8	35
24.6 - 25.3	4	24.6	46
	50	25.4	50

c)



d)

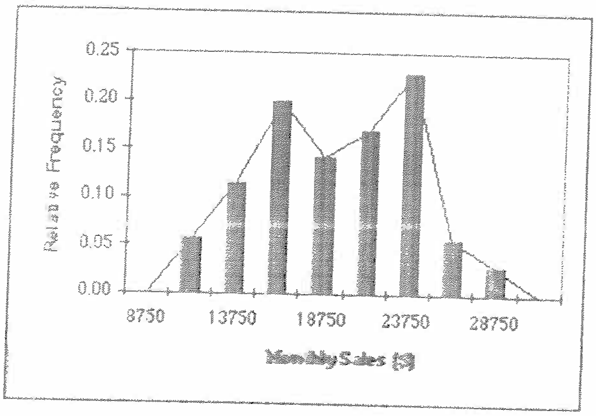


e) About 78% of the time.

2-41 a)

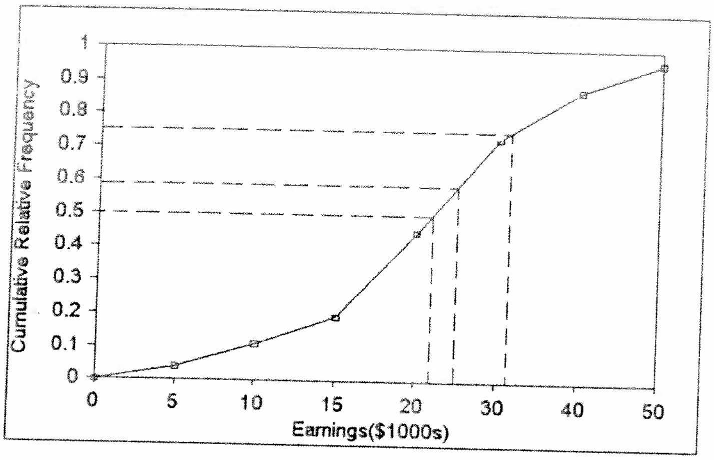
Monthly Sales	Frequency	Relative Frequency
\$10,000 - 12,499	2	.057
12,500 - 14,999	4	.114
15,000 - 17,449	7	.200
17,500 - 19,999	5	.143
20,000 - 22,499	6	.171
22,500 - 24,999	8	.229
25,000 - 27,449	2	.057
27,500 - 29,999	1	.029
	35	1.000

b)



2-42

Class ($)	Cumulative Relative Frequency	Class ($)	Cumulative Relative Frequency
< 5,001	.038	20,001 - 30,000	.731
5,001 - 10,000	.108	30,001 - 40,000	.877
10,001 - 15,000	.192	40,001 - 50,000	.946
15,001 - 16,000	.446	> 50,000	1.000



a) About 42%.

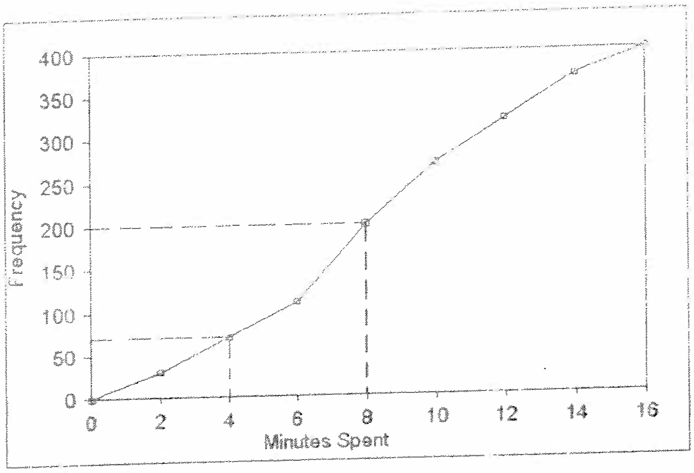
b) About $22,000.

c) About $31,000.

2-43 a)

Minutes Spent (<)	0	2	4	6	8	10	12	14	16
Cumulative Frequency	0	30	70	110	200	270	320	370	400

b)



c) At least $\frac{280}{400}$, or 72.5%, and at most $\frac{330}{400}$, or 82.5%, took longer than 4 minutes.

2-44 By grouping those of the same educational level together, we can more clearly see group differences associated with educational level.

Educational Level	Salary range
Did not finish high school	\$14,400 - 17,600
High school graduates	17,000 - 30,400
One or more years of college	14,400 - 22,400
College graduates	19,600 - 34,400
Master's degree	23,200 - 36,200
PhD degrees	29,000 - 64,000
Doctors & lawyers	52,000 - 100,000

2-45 No, because it's been analyzed to some extent to get averages for each week's attendance and to get percentages of full attendance. Raw data would be the actual number of absences for each day or week of the time period.

2-46 a)

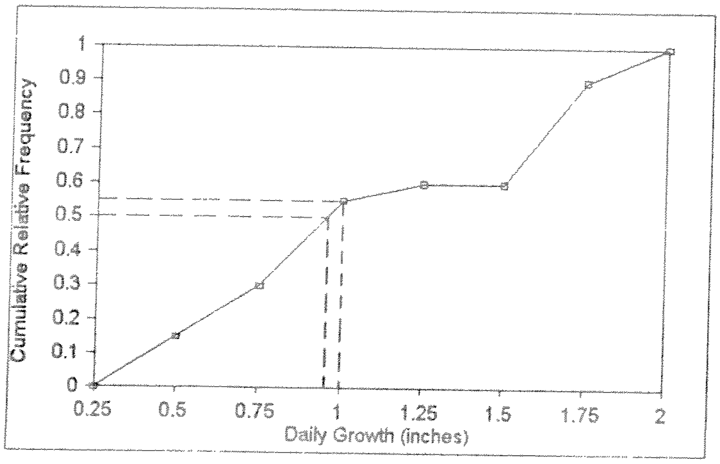
1.9	1.8	1.7	1.6	1.5	1.5	1.5	1.5	1.2	0.9
0.9	0.9	0.9	0.8	0.7	0.7	0.5	0.4	0.4	0.3

b)

Class (inches)	Relative Frequency	Cumulative Relative Frequency
0.000 - 0.249	.00	.00
0.250 - 0.499	.15	.15
0.500 - 0.749	.15	.30
0.750 - 0.999	.25	.55
1.000 - 1.249	.05	.60
1.250 - 1.499	.00	.60
1.500 - 1.749	.30	.90
1.750 - 1.999	.10	1.00

c) The data are distinctly bimodal, with modal classes 0.750 - 0.999 and 1.500 - 1.749.

d)



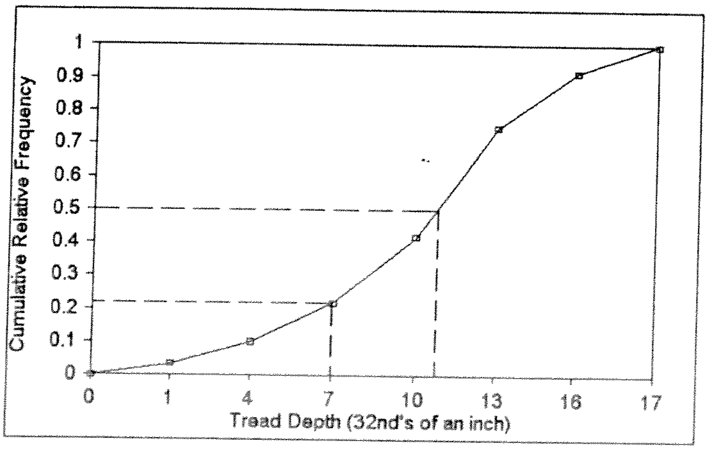
About 45% grew more than 1.0 inches per week.

e) About .95 inches.

2-47 To calculate the answer, we must first construct a relative frequency distribution and a "less than" cumulative distribution.

Tread Depth (inches)	Relative Frequency	Tread Depth (inches)	Cumulative Relative Frequency
16/32	.083	0/32	0
13/32 - 15/32	.167	1/32	.033
10/32 - 12/32	.333	4/32	.100
7/32 - 9/32	.200	7/32	.217
4/32 - 6/32	.117	10/32	.417
1/32 - 3/32	.067	13/32	.750
0/32	.033	16/32	.917
	1.000	17/32	1.000

Next we will graph the cumulative data in an ogive.

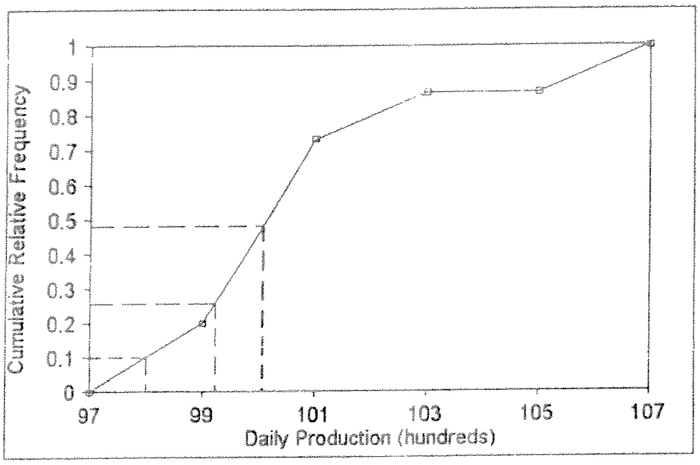


a) Approximately 11/32 of an inch.

b) Approximately 22% are unsafe.

2-48

Class (units)		Cumulative Relative Frequency
9700 -	9899	.200
9900 -	10099	.733
10100 -	10299	.867
10300 -	10499	.867
10500 -	10699	1.000



a) About 50% (7 or 8 items) exceeded the breakeven point.

b) About 9,900 units.

c) About 9,800 units.

2-49 a)

1	3	4	5	7	11	12	14	15	16
16	17	18	20	21	24	25	26	27	29

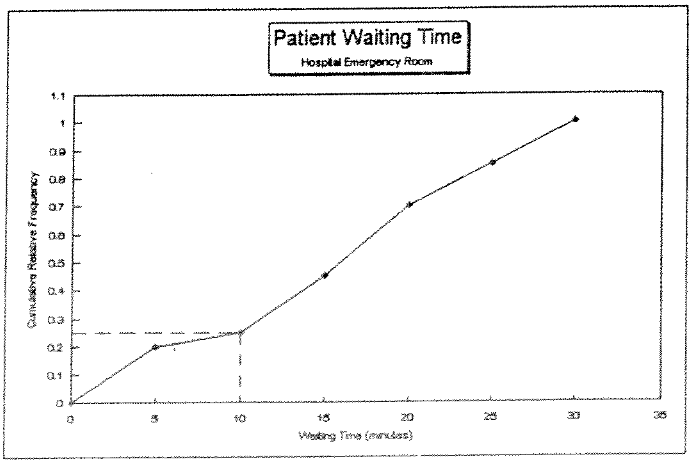
The data range from 1 to 29 minutes.

b)

Classes (min.)	1 - 5	6 - 10	11 - 15	16 - 20	21 - 25	26 - 30	Total
Frequency	4	1	4	5	3	3	20

There is a group of quite short waits (5 minutes or less), with most of the data fairly uniformly distributed between 11 and 30 minutes.

c) From the ogive below, we see that 75% of the patients should expect to wait for more than 10 minutes.



2-50 It tells you what fraction of the observations fit into each class. This makes it easier to compare samples or populations of different sizes.

2-51 a) For the first 10 elements, we get different samples if we sample across the first row of the array (sample "R") or down the first column (sample"C").

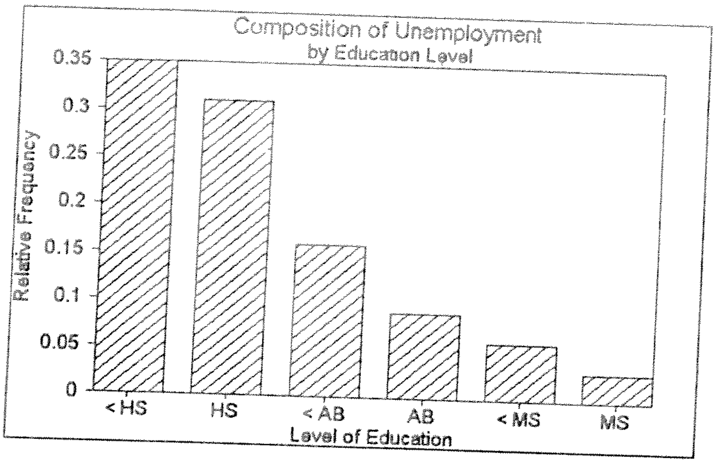
1st ten (R):	226	210	222	215	201	198	233	175	191	175
1st ten (C):	226	174	189	217	180	264	233	155	220	207
Largest ten:	267	264	259	258	257	252	248	245	244	243

b) No, the sample of the first ten elements is more representative because the items in it more adequately represent the distribution of items in the whole population in terms of their values.

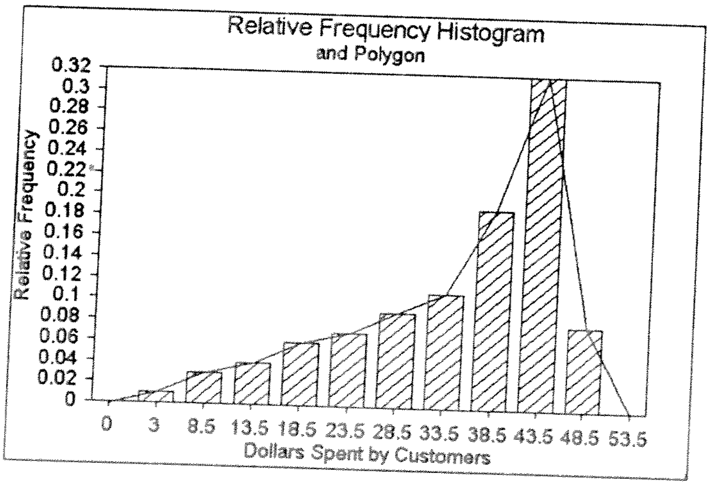
c) Only if all the players had the same weight, which would rarely happen.

2-52 $\frac{2000}{2000+8000} = .2$, so there should be $.2 \times 250 = 50$ women and 200 men.

2-53



2-54

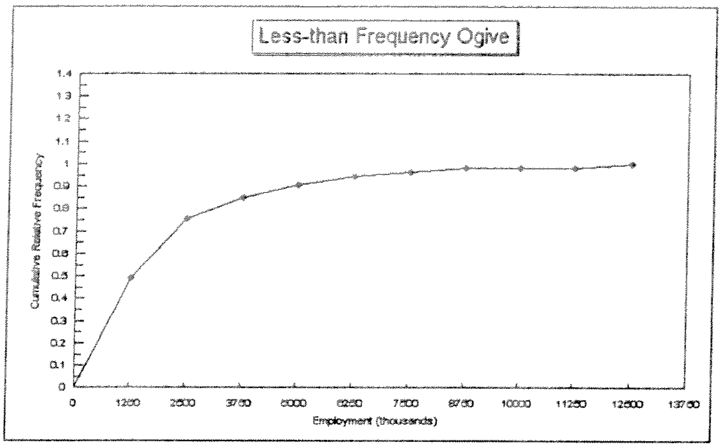


2-55 a,b)

Employment (thousands)	Frequency	Relative Frequency	Cumulative Relative Frequency
0.00 - 1,249.90	26	0.4906	0.4906
1,250.00 - 2,499.90	14	0.2642	0.7547
2,500.00 - 3,749.90	5	0.0943	0.8491
3,750.00 - 4,999.90	3	0.0566	0.9057
5,000.00 - 6,249.90	2	0.0377	0.9434
6,250.00 - 7,499.90	1	0.0189	0.9623
7,500.00 - 8,749.90	1	0.0189	0.9811
8,750.00 - 9,999.90	0	0.0000	0.9811
10,000.00 - 11,249.90	0	0.0000	0.9811
11,250.00 - 12,499.90	1	0.0189	1.0000

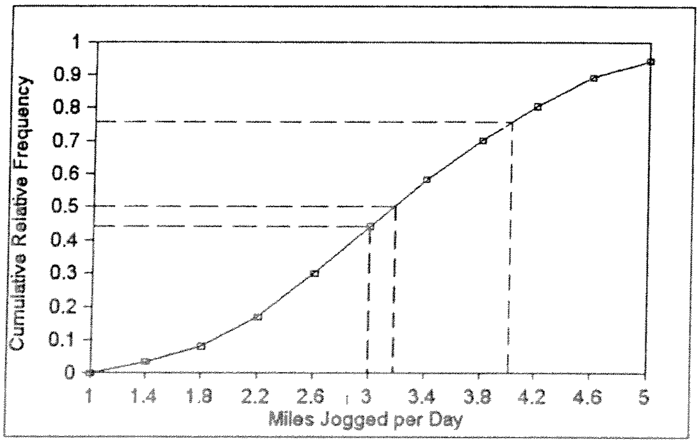
c) These are discrete data, since the number of workers must be a whole number.

d) See part (a) for the cumulative relative frequencies.



e) About 20% have nonfarm employment greater than 3 million.

2-56



About 75% of the joggers average 4 or fewer miles per day.

2-57 Using the ogive from the previous problem:

a) About 3.20 miles per day.

b) About 56% run 3.0 miles per day.

2-58 Since the Ivy League schools have more male than female undergraduates, this is not a representative sample.

2-59

Age Group	Relative Proportion		Sample Size		# in Sample
12 - 17	.17	×	3000	=	510
18 - 23	.31	×	3000	=	930
24 - 29	.27	×	3000	=	810
30 - 35	.21	×	3000	=	630
36 +	.04	×	3000	=	120
					3000

2-60 a)

Journal Number	Frequency	Relative Frequency
1	1	.0417
2	2	.0833
3	2	.0833
4	0	.0000
5	2	.0833
6	2	.0833
7	3	.1250
8	1	.0417
9	2	.0833
10	1	.0417
11	2	.0833
12	2	.0833
13	0	.0000
14	2	.0833
15	2	.0833
	24	.9998

NOTE: The relative frequency column does not total 1.0000 due to rounding errors.

b)

Branch	Frequency	Relative Frequency
North	11	.4583
West	8	.3333
South	5	.2083
	24	.9999

c)

# of Publications	Frequency	Relative Frequency
1 - 3	6	.2500
4 - 6	5	.2083
7 - 9	4	.1667
10 - 12	4	.1667
13 - 15	2	.0833
16 - 18	2	.0833
19 - 21	1	.0417
	24	1.0000

d) Although the faculty use of journals is widespread, it is clear that the North branch accounts for most of the publications. Well over half of the faculty (about 62%) publish 9 or fewer articles.

2-61 Only 880 of the 2000 questionnaires were returned.

1) With only a 44% response rate, the data may be biased. We have no way of telling if those who respond are representative of the population as a whole.

2) So far as we know, there is no other evidence available.

3) Yes, we are missing the data from the 56% who didn't reply.

4) 880; No, they didn't represent those individuals who chose not to respond.

5) No conclusions are given in the example.

2-62 1) Quantitative, discrete, open-ended: in all likelihood there would be open-ended classes like under 20 and over 50.

2) Quantitative, discrete, open-ended: people will report incomes to the nearest thousand dollars, and for very high and low incomes, the company won't care about precise values.

3) Qualititative, discrete, closed: the only possibilities are single, married, divorced, and widowed.

4) and 5) Both of these distributions would be qualitative and discrete. However, in contrast to 3), the answers could be so varied that it is unlikely that the company would choose to list all possibilities. Instead, the most frequent responses would be used and "other" included to cover all other possibilities. Therefore, these distributions would most likely be open-ended.

2-63 Rounding to avoid inconvenient values like $2.995, the class marks for the first nine classes are $3, $8.50, $13.50, $18.50, $23.50, $28.50, $33.50, $38.50, and $43.50. The open uppermost class has no class mark.

2-64 Group I

None	Mild	Moderate	Severe	or	None	- 3
None	Mild	Moderate	Severe		Mild	- 7
None	Mild	Moderate	Severe		Moderate	- 5
	Mild	Moderate			Severe	- 3
	Mild	Moderate				
	Mild					
	Mild					

Group II

None	Mild	Moderate	Severe	or	None	- 1
	Mild	Moderate	Severe		Mild	- 4
	Mild	Moderate	Severe		Moderate	- 8
	Mild	Moderate	Severe		Severe	- 5
		Moderate	Severe			
		Moderate				
		Moderate				
		Moderate				

Displayed this way, it is much easier to compare the two groups.

2-65 Yes and no. It is not raw data, because the scores are composites of other values. The raw data would be the individual scores on tests, homework and papers. However, if someone were interested in doing some analysis on final grades, it would be raw data for that purpose.

2-66 The problem asks for a distribution by specialty, not combinations. Therefore, the only categories would be accounting, marketing, statistics, finance, and no publications. The faculty who have more than one specialty will be double counted in this particular distribution.

Specialty	Frequency	Relative Frequency
Accounting	17	.140
Marketing	41	.339
Statistics	40	.331
Finance	22	.182
No Publications	1	.008
	121	1.000

2-67 a)

$ 0 -	$ 4999
5000 -	9999
10000 -	14999
15000 -	19999
20000 -	24999
25000 -	29999
30000 -	34999

If the student elected to start with $0 - $5000 as the first interval, the solution is given on the right.

$ 0 -	$ 5000
5001 -	10000
10001 -	15000
15001 -	20000
20001 -	25000
25001 -	30000
30,001 -	35000

b)

< $10000	
10000 -	14999
15000 -	19999
20000 -	24999
25000 -	29999
≥$30000	

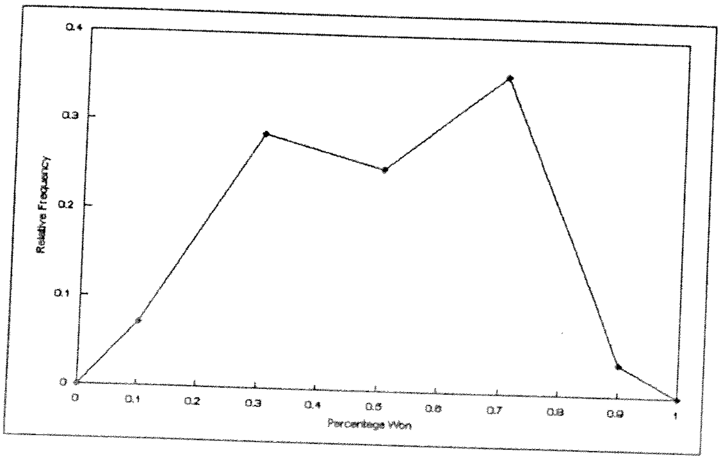
or

≤ $10000	
10001 -	15000
15001 -	20000
20001 -	25000
25001 -	30000
>$30000	

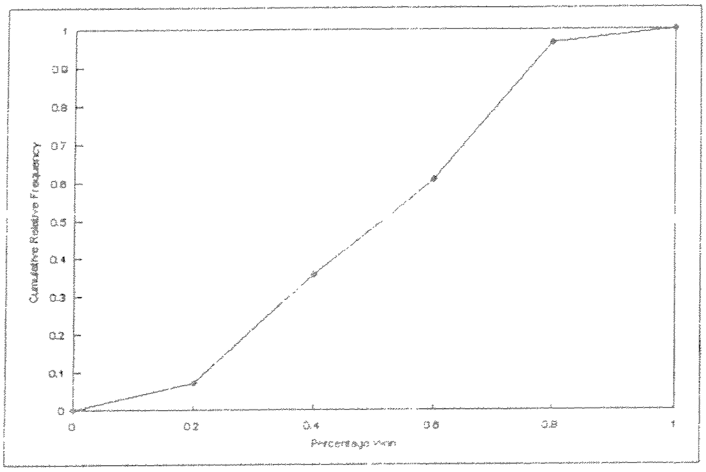
2-68 a, b)

Percentage won	Frequency	Relative Frequency	Cumulative Relative Frequency
0.000 - 0.199	2	0.0714	0.0714
0.200 - 0.399	8	0.2857	0.3571
0.400 - 0.599	7	0.2500	0.6071
0.600 - 0.799	10	0.3571	0.9643
0.800 - 0.999	1	0.0357	1.0000

c)



d) See part (a) for the cumulative relative frequencies.



e) The one team in the 0.800 - 0.999 class gets a playoff berth, as do 5 of the 10 teams in the 0.600 - 0.799 class.

2-69 Since we know that 40 salespeople were used for the table, the total in the frequency column must be 40. Also, we know that the relative frequencies must always sum to 1.000.

For any row where the relative frequency is given but the number of observations is not, we need only to multiply the total observations (40) by the relative frequencies to get the number of observations. For example, in row 1 we have $40 \times .075 = 3$.

Similarly, if we are given the number of observations for any row, we need only to divide this number by the total observations (40) to determine the relative frequency. For example, in row 3 we have $4 \div 40 = .100$.

Finally, we are left with row 4 where neither the frequency nor the relative frequency is given. However, the number of observations for row 4 must be eight for the column to sum to 40. Also, we know that the relative frequency column must total 1.00, and so its entry in row 4 must be .20. The completed table is given below.

Class	Frequency	Relative Frequency
0 - 10	3	.075
11 - 20	1	.025
21 - 30	4	.100
31 - 40	8	.200
41 - 50	2	.050
51 - 60	7	.175
61 - 70	9	.225
71 - 80	5	.125
81 - 90	0	.000
91 - 100	1	.025
	40	1.000

2-70 a) 139 129 128 126 121 119 119 116 115 114 113 113 112 111 110
110 108 107 105 102 101 100 99 99 97 93 93 91 87 84
80 75 72 66 60

b) 16 exceeded the limit of 108 minutes, 18 were under, 1 was exactly at the limit.

c)

Class (min.)	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139
Frequency	2	2	3	6	6	11	4	1
Relative Freq.	.057	.057	.086	.171	.171	.314	.114	.029

d) If 108 minutes is typical, then about half should be above 108 and half below. The data support this. Since we don't know how much downtime per shift is viewed as excessive, we cannot tell if Cline should be concerned or not.

2-71 a)

Class (tons)	281-305	306-330	331-355	356-380	381-405	406-430
Frequency	3	3	3	15	8	3
Relative Frequency	.086	.086	.086	.429	.229	.086

b) 6 (or 17.1%) produced less than expected; 11 (or 31.4%) produced more.

c) Even though about half of the crews have higher than standard downtimes, only 17% don't produce as expected, so downtime doesn't seem to be a big problem.

2-72 a) 151.1 147.8 145.7 142.3 142.0 142.0 141.2 141.1 140.9 138.7
138.2 137.4 134.9 133.3 133.0 130.8 129.8 128.9 126.3 125.7
125.7 125.2 125.0 119.9 118.6

b) 23/25 (or 92%) withstood 120,000 pounds of force; only 1/25 (or 4%) withstood 150,000 pounds of force.

c) 16/25 (or 64%) would have failed. They should stop ordering from this supplier until it can improve the strength of its bolts.

2-73 1. The classes beginning at 5.00% and 10.00% should begin at 5.01% and 10.01%, respectively, so they don't overlap the classes immediately below.

2. The class beginning at 20.01% should end at 22.50%, so it doesn't overlap the class immediately above.

3. The 15.01-17.50% class is missing.

4. The 22.51-25.50% class is too wide. It should end at 25.00%, and the open-ended class above it should be 25.01% and greater.

2-74 a) 9, 8, 7, 7, 6, 6, 5, 5, 5, 5, 4, 3, 3, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1

b) 1.

Class (sales)	1	2	3	4	5	6	7	8	9
Frequency	10	3	4	1	4	2	2	1	1
Relative Frequency	.357	.107	.143	.036	.143	.071	.071	.036	.036

2.

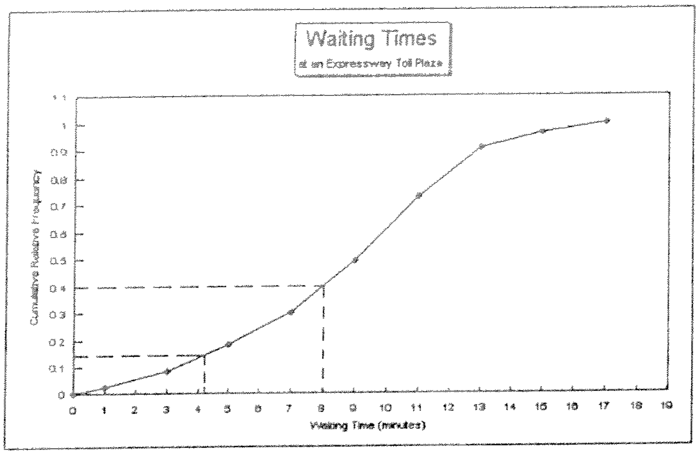
Class (sales)	1-3	4-6	7-9
Frequency	17	7	4
Relative Frequency	.607	.250	.143

Both distributions are skewed: many countries have relatively few sales, and then the distribution tails off to the right.

2-75 a)

Class (minutes)	Cumulative Frequency	Cumulative Relative Frequency
0.00 - 0.99	75	.025
1.00 - 2.99	258	.086
3.00 - 4.99	552	.184
5.00 - 6.99	902	.301
7.00 - 8.99	1482	.494
9.00 - 10.99	2191	.730
11.00 - 12.99	2730	.910
13.00 - 14.99	2894	.965
15.00 - 16.99	3000	1.000

b)



About 85% waited more than 4 minutes; about 60% waited more than 8 minutes.

2-76 a) 4600 - 5199, 5200 - 5799, 5800 - 6399, 6400 - 6999, 7000 - 7599, and 7600 - 8199.

b) 0.00 - 1.39, 1.40 - 2.79, 2.80 - 4.19, 4.20 - 5.59, 5.60 - 6.99, and 7.00 - 8.39.

2-77

Class (inches^3)	Cumulative Frequency	Cumulative Relative Frequency
101 - 150	1	.013
151 - 200	8	.103
201 - 250	15	.192
251 - 300	23	.295
301 - 350	40	.513
351 - 400	56	.718
401 - 450	71	.910
451 - 500	78	1.000

a) 70% are larger than about 300 cubic inches.

b) The middle value was approximately 350 cubic inches.

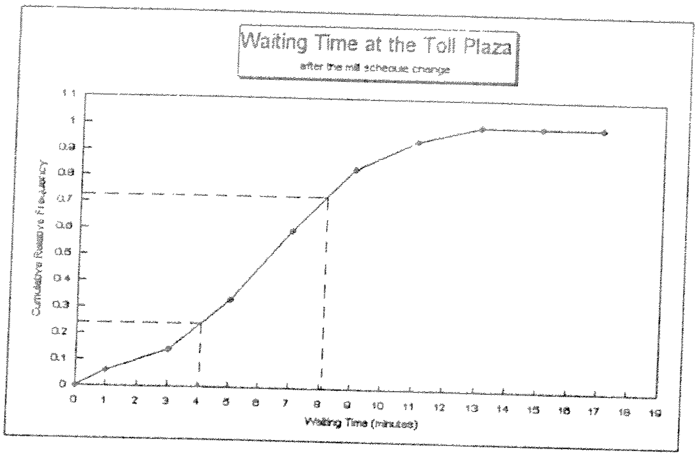
c) About 28% are over 400 cubic inches and won't be able to use the system.

2-78 a) 8/61 = 13.1% b) 29/61 = 47.5%

2-79 a)

Class (minutes)	Cumulative Frequency	Cumulative Relative Frequency
0.00 - 0.99	177	.059
1.00 - 2.99	415	.138
3.00 - 4.99	993	.331
5.00 - 6.99	1793	.598
7.00 - 8.99	2506	.835
9.00 - 10.99	2832	.944
11.00 - 12.99	2991	.997
13.00 - 14.99	3000	1.000
15.00 - 16.99	3000	1.000

b)



About 77% waited more than 4 minutes; about 28% waited more than 8 minutes.

c) Waiting times have clearly been reduced.

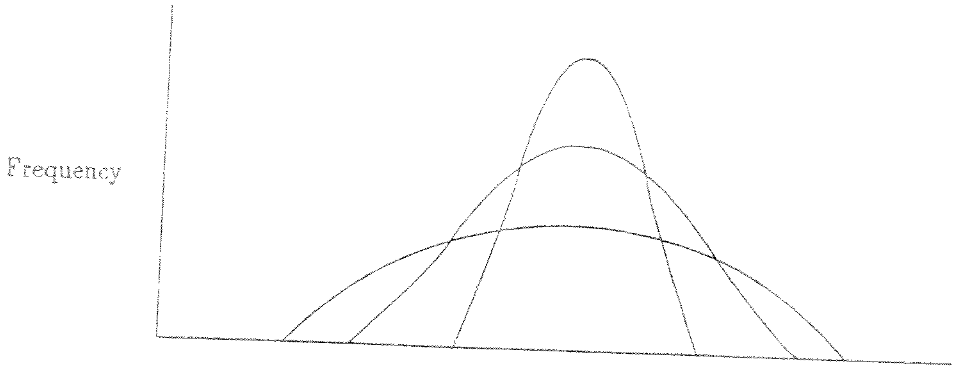
2-80 A histogram will further highlight the pattern of low numbers of customers from 11 p.m. to 6 a.m., then increasing fairly steadily until 11 a.m. or noon, remaining reasonably steady until 5 p.m., and then decreasing fairly steadily until 11 p.m. One limitation: national data may not apply to Utah. For example, if there were many factories close by, the number of customers at Fresh Foods might reflect changes in work shifts at those factories.

2-81 a) For each market, combine the domestic and foreign equities. Then a histogram will clearly show the dominance of the New York, London, and Tokyo markets. (You might even combine the NYSE and Nasdaq markets to further emphasize the dominance of New York.)

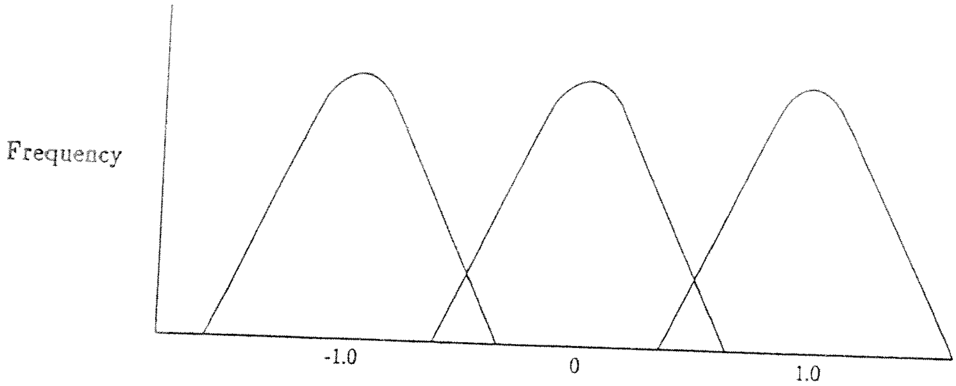
b) For the two markets, compute the relative frequencies of the domestic and foreign equities. Construct a relative frequency histogram (like Figure 2-16), with bars for New York and London divided horizontally to show the proportions of domestic and foreign equities in each market. This will clearly show that the New York market is predominantly in domestic equities, while the London market is predominantly in foreign equities.

MEASURES OF CENTRAL TENDENCY AND DISPERSION IN FREQUENCY DISTRIBUTIONS

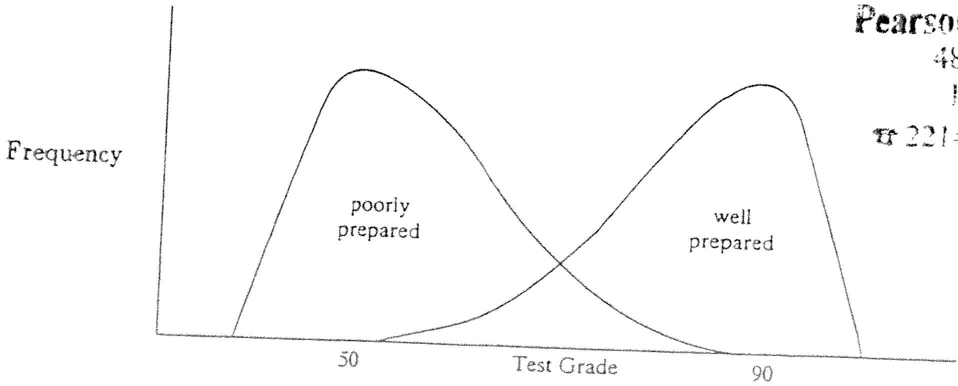
3-1



3-2



3-3



3-4 a) B b) A c) A d) B e) B f) A g) neither

3-5 The scores are concentrated at the lower end of the distribution for test A, indicating it is more difficult.

3-6 $\bar{x} = (8 + 5 + 9 + 10 + 9 + 12 + 7 + 12 + 13 + 7 + 8)/11 = 100/11 = 9.091$
Since this is over 9, they do not qualify.

Pearson Education(s) Pte Ltd.
482, F. I. E. Patparganj
Delhi-110092, India
☎ 22146067-70, Fax: 22146071

3-7 a) $\bar{x} = \frac{\sum x}{n} = \frac{118,600}{11} = \$10,782$

Since this is less than $12,500, they do qualify.

b) They already qualify, so it does not have to fall.

c) Average income can rise by 12500 − 10782 = $1718 and Child-Care will remain eligible.

3-8 a)

Age	40 - 49	50 - 59	60 - 69	70 - 79	80 - 89
Frequency	4	4	3	2	7

b) $\bar{x} = \frac{\sum(f \times x)}{n} = \frac{4(45) + 4(55) + 3(65) + 2(75) + 7(85)}{20} = \frac{1340}{20} = 67$

c) $\bar{x} = \frac{\sum x}{n} = \frac{1335}{20} = 66.75$

d) As expected, they are close, but not exactly the same.

3-9

Class	Frequency (f)	Midpoint (x)	$f \times x$	u	$u \times f$
20 - 29	6	24.5	147.0	− 5	− 30
30 - 39	16	34.5	552.0	− 4	− 64
40 - 49	21	44.5	934.5	− 3	− 63
50 - 59	29	54.5	1580.5	− 2	− 58
60 - 69	25	64.5	1612.5	− 1	− 25
70 - 79	22	74.5	1639.0	0	0
80 - 89	11	84.5	929.5	1	11
90 - 99	7	94.5	661.5	2	14
100 - 109	4	104.5	418.0	3	12
110 - 119	0	114.5	0.0	3	0
120 - 129	2	124.5	249.0	5	10
	143		8723.5		− 193

a) $\bar{x} = \frac{\sum(f \times x)}{n} = \frac{8723.5}{143} = 61.0035$ seconds per customer

b) $\bar{x} = x_0 + w\frac{\sum(u \times x)}{n} = 74.5 + \frac{10(-193)}{143} = 61.0035$ seconds per customer

3-10 For the first six months, the average is

(234 + 216 + 195 + 400 + 315 + 274)/6 = 1634/6 = 272.33 animals.

For the entire year, the average is

(1634 + 302 + 291 + 275 + 300 + 375 + 450)/12 = 3627/12 = 302.25 animals.

Since the criterion is not met for the first six months, the owner will not build the new store.

3-11 $\bar{x} = \frac{\sum x}{n} = \frac{53.04}{18} = 2.947$ owners.

Since this is below 2.96 owners, the machine should be recalibrated.

3-12 $\bar{x} = \frac{\sum x}{n} = \frac{465.9}{20} = 23.295$.

Since this exceeds 23, the manager should be concerned.

3-13 a) $\bar{x} = \frac{\sum x}{n} = \frac{9.866}{5} = \$1.9732$ million

b) $\bar{x} = \frac{\sum x}{n} = \frac{9.536}{5} = \$1.9072$ million

c) We can either average all 10 balances, or else just average the two weekly averages. In either case, the result will be

$$x = \frac{1.9732 + 1.9072}{2} = \$1.9402 \text{ million}$$

d) Since $1.9402 < 1.9700$, National does not qualify for the special interest rates.

e, f) $10(1.9700 - 1.9402) = 0.298$, i.e., the last day's investment would have to increase by \$298,000 to qualify for the special rates.

3-14

	Q_1	Q_2	Q_3	Q_4	(b)
Y_1	10	5	25	15	13.75
Y_2	20	10	20	10	15.00
Y_3	30	15	45	50	35.00
(a)	20	10	30	25	21.25 (c)

(All figures in thousands of dollars)

3-15 a) $\bar{x} = \frac{\sum x}{n} = \frac{136{,}000}{5} = \$27{,}200$ in 1988-1992

b) $\bar{x} = \frac{\sum x}{n} = \frac{110{,}000}{5} = \$22{,}000$ in 1983-1987

c) $\bar{x} = \frac{\sum x}{n} = \frac{84{,}000}{5} = \$16{,}800$ in 1978-1982

d) The averages indicate an increasing trend. However, the increase is less than the increase in the cost of living over the same period. It looks like she has kept the the budget down.

3-16

Student 1: $(.20 \times 85) + (.10 \times 89) + (.10 \times 94) + (.25 \times 87) + (.35 \times 90) = 88.55$
Student 2: $(.20 \times 78) + (.10 \times 84) + (.10 \times 88) + (.25 \times 91) + (.35 \times 92) = 87.75$
Student 3: $(.20 \times 94) + (.10 \times 88) + (.10 \times 93) + (.25 \times 86) + (.35 \times 89) = 89.55$
Student 4: $(.20 \times 82) + (.10 \times 79) + (.10 \times 88) + (.25 \times 84) + (.35 \times 93) = 86.65$
Student 5: $(.20 \times 95) + (.10 \times 90) + (.10 \times 92) + (.25 \times 82) + (.35 \times 88) = 88.50$

3-17 a) The average cost per case is

$$\bar{x}_w = \frac{\sum(w \times x)}{\sum w} = \frac{6(28)+4(36)+8(16)+3(18)+1(6)}{6+4+8+3+1} = \frac{500}{22} = \$22.73$$

b) With 24 tapes per case, the average cost per tape is \$22.73/24 = \$0.95.

c) Although he will be losing money on the Performance High-Grade tapes (which cost him \$1.50 each), on average he will be making 30 cents per tape if he sells them at \$1.25 apiece.

d) The average cost per case would be unchanged. The average cost per tape would be halved. Selling the tapes at \$1.25 apiece would now be an even better deal for Jim.

3-18

$$\bar{x}_w = \frac{\sum(w \times x)}{\sum w} = \frac{0(897) + 1(1082) + 2(1325) + 3(814) + 4(307) + 5(253) + 6(198)}{897 + 1082 + 1325 + 814 + 307 + 253 + 198}$$

$$= \frac{9855}{4876} = 2.021 \text{ times}$$

3-19 The average rate of change is

$$\bar{x}_w = \frac{\sum(w \times x)}{\sum w} = \frac{72(11.5)+62(6.4)+48(6.4)+89(-9.7)+94(-18.2)}{72+62+48+89+94} = \frac{-1042.1}{365}$$

$$= -2.86\%$$

3-20 $$\bar{x}_w = \frac{\sum(w \times x)}{\sum w}$$

$$= \frac{16400(.05)+24100(.08)+77600(.13)+1900(.17)+1300(.35)+750(.40)+800(.45)}{16400+24100+77600+1900+1300+750+800}$$

$$= \frac{14274}{122850} = \$0.1162 \text{ per ounce}$$

3-21 $$\bar{x}_w = \frac{\sum(w \times x)}{\sum w} = \frac{8000(75)+14000(40)+24000(30)+35000(15)}{8000+14000+24000+350000}$$

$$= \frac{2{,}405{,}000}{81{,}000} = \$29.69/\text{hour}$$

However, they should only cite this as an average rate for clients who use the four professional categories for approximately 10%, 17%, 30%, and 43% of the total hours billed.

3-22 $\text{GM} = \sqrt[5]{1.05(1.105)(1.09)(1.06)(1.075)} = \sqrt[5]{1.441094314} = 1.0758172$

So the average increase is about 7.58% per year.

3-23 $\text{GM} = \sqrt[9]{1.11(1.09)(1.07)(1.08)(0.96)(1.14)(1.11)(0.97)(1.06)} = \sqrt[9]{1.74635959} = 1.0639$

So the average annual percentage change in net worth is about 6.39%. The estimated change from 1996 to 1998 is $(1.0639)^2 - 1 = 0.1319$, that is, 13.19%.

3-24 $\text{GM} = \sqrt[4]{17630/12500} = \sqrt[4]{1.4104} = 1.0897718$

So the average increase is 8.98% per year. In 1996, the estimated production is $17630(1.0898)^3 = 22819$ units.

3-25 Since $\text{GM} = 1.24 = \sqrt[6]{1.19(1.35)(1.23)(1.19)(1.30)x}$,

$$x = \frac{(1.24)^6}{(1.19)(1.35)(1.23)(1.19)(1.30)} = 1.189$$

3-26 $\text{GM} = \sqrt{1.15/1.00} = 1.0723805.$

So the price has been increasing by about 7.24% per week.

3-27 $\text{GM} = \sqrt[7]{2.49/2.30} = \sqrt[7]{1.08260870} = 1.0114$

So the price has been increasing at an average rate of 1.14% per week.

3-28 $\text{GM} = \sqrt[4]{66/55} = \sqrt[4]{1.2} = 1.0466351$

So the average increase is about 4.66% per year. In three more years, the estimated cost is $66(1.0466)^3 =$ \$75.66.

3-29 a) $GM = \sqrt[4]{1.05(1.1)(1.03)(1.06)} = \sqrt[4]{1.261029} = 1.0597$, so from 1992 to 1995, the average rate of increase was 5.97% per year.

b) $GM = \sqrt[6]{.96(1.05)(1.1)(1.03)(1.06)(.95)} = \sqrt[6]{1.150058448} = 1.0236$, so from 1991 to 1996, the average rate of increase was 2.36% per year.

c) The new penal code does appear to be having an effect on the rate of growth of the prison population, which has increased from 2% to 2.4% per year.

3-30 We first arrange the mileages in ascending order:

210	447	450	469	488	559	560	589	657	689
756	775	788	789	810	876	890	943	987	1450

a) median $= \dfrac{689 + 756}{2} = 722.5$ miles, the average of items 10 and 11

b) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{14182}{20} = 709.1$ miles

c) In this instance, both are equally good since they are so close to each other.

3-31 We first arrange the prices in ascending order:

8 12 14 15 22 24 25 26 28 28 29 29 31 32 32 33 35

The median is the 9th ((17+1)/2) of these 17 numbers, or 28 channels.

b) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{423}{17} = 24.88$ channels

c) Noting that only 6 of the 17 observations fall below the mean, we can see that the data are skewed to the right. Hence the median is the better measure of central tendency for these data.

3-32

Class	Frequency	Cumulative Frequency	Class	Frequency	Cumulative Frequency
10 - 19.5	8	8	60 - 69.5	52	181
20 - 29.5	15	23	70 - 79.5	84	265
30 - 39.5	23	46	80 - 89.5	97	362
40 - 49.5	37	83	90 - 99.5	16	378
50 - 59.5	46	129	≥100	5	383

a) The median is the (383 + 1)/2 = 192nd item.

b) The median class is 70 – 79.5.

c) The step width in the median class = 10/84 = 0.1190

d) median = 70 + 10(.1190) = 71.190

e) $\tilde{m} = \left(\dfrac{(n+1)/2-(F+1)}{f_m}\right)w + L_m = \left(\dfrac{192-82}{84}\right)10 + 70 = 71.1905$

The answers in (d) and (e) differ slightly because of rounding

3-33 a) Since there are 48 observations, the median falls midway between the 24th and 25th in the 50 - 74.9 pounds class.

$$\tilde{m} = \left(\frac{(n+1)/2-(F+1)}{f_m}\right) w + L_m = \left(\frac{49/2 - 19}{16}\right) 25 + 50 = 58.59 \text{ pounds}$$

b) $\bar{x} = \frac{\sum(f \times x)}{n} = \frac{5(12.5) + 13(37.5) + 16(62.5) + 8(87.5) + 6(112.5)}{5 + 13 + 16 + 8 + 6}$

$= \frac{2925}{48} = 60.9375$ pounds

c) Since the data are slightly skewed, the median is a bit better than the mean, but there isn't much difference.

3-34 We first arrange the times in ascending order:

17	19	21	22	22	28	29	29	29	30
32	33	33	34	34	39	41	43	44	52

The median time is $(30 + 32)/2 = 31$ minutes, the average of the 10th and 11th items. This is close enough to the target of 30 minutes to conclude that excessive speeds have not been a problem.

3-35 First, we must arrange the data in descending order:

10.500	10.125	9.500	9.000	7.500	6.000	5.875
10.500	10.000	9.375	8.500	7.000	6.000	5.750
10.250	9.875	9.250	8.250	6.500	6.000	5.500
10.250	9.875	9.250	8.000	6.250	5.875	5.375
10.125	9.500	9.125	7.875	6.250	5.875	5.250

The median is $(35 + 1)/2 =$ the 18th element. For our array, the 18th element is 8.250 yards. The 150 jobs should require about 1240 yards of material, i.e., 150(8.250).

3-36 The median is the $(4723 + 1)/2 = 2362$nd element, which occurs as the 907th element in the 750 - 999.99 class. The distance between elements in this class is $250/1776 = .14077$, so the median is $750 + 906(.14077) = \$877.54$. Using equation 3-8, we have

$$\tilde{m} = \left(\frac{(n+1)/2 - (F+1)}{f_m}\right) w + L_m = \left(\frac{(4723+1)/2 - (1456)}{1776}\right) 250 + 750$$

$$= \$877.54$$

3-37 The median is halfway between "mildly disagree" and "agree somewhat."

3-38

Books checked out (x)	0	1	2	3	4	5	6	7	
Frequency (f)	3	3	7	3	2	1	0	1	
$f \times x$	0	3	14	9	8	5	0	7	$\sum f \times x = 46$

a) The mode is 2 books.
b) The mean is 46/20 = 4.3 books.
c) Since the distribution is skewed, the mode is the better measure of central tendency.

3-39 $\text{Mo} = L_{Mo} + \frac{d_1}{d_1 + d_2} w = 62 + \left(\frac{29}{29+3}\right) 5 = 66.53$

3-40 a) Brunette b) A c) Wednesday and Saturday

3-41 a)

Class	Frequency
66 - 87	4
88 - 109	8
110 - 131	3
132 - 153	6
154 - 175	3
176 - 197	2
198 - 219	0
220 - 241	1
	27

b) $\text{Mo} = L_{Mo} + \dfrac{d_1}{d_1 + d_2} w = 88 + \left(\dfrac{4}{4+5}\right) 22 = 97.78$ apartments

c) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{3407}{27} = 126.19$ apartments

d) Since this distribution is skewed, the mode is a better measure of central tendency.

3-42 $\text{Mo} = L_{Mo} + \dfrac{d_1}{d_1 + d_2} w = 750 + \left(\dfrac{710}{710 + 284}\right) 250 = \$928.57$

3-43 a) $\bar{x} = \dfrac{\sum(f \times x)}{n} = \dfrac{2(.5) + 4(1.5) + 6(2.5) + 7(3.5) + 5(4.5) + 3(5.5) + 1(6.5)}{2 + 4 + 6 + 7 + 5 + 3 + 1}$

$= \dfrac{92}{28} = 3.28\dot{6}$ days

b) $\text{Mo} = L_{Mo} + \dfrac{d_1}{d_1 + d_2} w = 3 + \left(\dfrac{1}{1+2}\right) 1 = 3.333$ days

c) Since there are 28 observations, the median falls midway between the 14th and 15th in the 3 - 3.99 days class.

$$\tilde{m} = \left(\frac{(n+1)/2 - (F+1)}{f_m}\right) w + L_m = \left(\frac{29/2 - 13}{7}\right) 1 + 3 = 3.214 \text{ days}$$

d) Since the distribution is slightly skewed, the median may be a bit better than the others as a measure of central tendency, but the values are so close that there really isn't much difference.

3-44 a) The modal class is \$1000 - \$1499

b) $L_{Mo} = 1000$ $\quad d_1 = 96$ $\quad d_2 = 104$ $\quad w = 500$

$\text{Mo} = L_{Mo} + \dfrac{d_1}{d_1 + d_2} w = 1000 + \left(\dfrac{96}{96 + 104}\right) 500 = \$1240$

c) $.90(1240) = \$1116$. All 535 applicants who earn below \$1000 qualify. If we assume the applicants in the 1000 - 1499 class are evenly distributed within that class, then about $(116/500)(400) = 93$ of them qualify as well. Thus about 628 applicants are qualified.

3-45 A, because the values tend to cluster closer to the mean

3-46 C

3-47 B, because it spreads the scores more widely

3-48 A, because it has less variability

3-49 Each of the groups has a different range of likely values. Because the chairmen of major congressional committees are usually the most senior, this distribution would probably have the tightest spread and closely resemble Curve A. Curve B would best fit the distribution of ages of newly elected members. This distribution would be centered around a younger age than the one for committee chairmen, but it would have more spread because of more variability in the ages of first-term congressmen. Finally, when you consider the ages of all members of Congress, the re- sulting distribution would not be strongly centered because it would span all ages. Therefore, this distribution would more nearly fit Curve C.

3-50 There are many ways that the concept may be involved. Certainly, the FTC would be examining the price variability for the industry and comparing the result to that of the suspect companies. The agency might examine price distributions for similar products, for the same products in a city, or for the same products in different cities. If the variability was significantly different in any of these cases, this result might constitute evidence of a conspiracy to set prices at the same levels.

3-51 a) C. These numbers could fall between 0 and 3000 per player.

b) A. The federal government has a fairly uniform salary structure.

c) B. Some variation, but within a relatively moderate range.

d) B or C. Private industry is more flexible than government in part b.

e) A. The range here is much smaller than in part c.

f) A or B. Some variation, but in a much smaller range than in part a.

3-52 First we arrange the data in increasing order:

33	45	52	54	55	- 1st quartile
61	66	68	69	72	- 2nd quartile
74	75	76	77	84	- 3rd quartile
91	91	93	97	99	- 4th quartile

Interquartile range $= Q_3 - Q_1 = 84 - 55 = 29$

3-53 First we arrange the data in increasing order:

2145	2200	2228	2268	2549	- 1st quartile
2598	2653	2668	2697	2841	- 2nd quartile
3249	3268	3362	3469	3661	- 3rd quartile
3692	3812	3842	3891	3897	- 4th quartile

a) range $= 3897 - 2145 = 1752$

b) 20th percentile = 4th element = 2268
80th percentile = 16th element = 3692
Interfractile range = 3692 - 2268 = 1424

c) Interquartile range $= Q_3 - Q_1 = 3661 - 2549 = 1112$

3-54 Placing the 30 temperatures in ascending order

69	72	82	84	84	86	87	87	88	88	88	88	89	89	89
92	92	94	94	94	94	95	96	97	98	99	99	102	102	105

we see that the 70th percentile (the 21st observation) is 94 degrees.

3-55 The range is 185 − 51 = \$134. It is not particularly useful because all the rest of the data falls between 83 and 157 dollars. The range greatly overstates the typical variability because it is determined by two outliers in the data set.

3-56 First we arrange the data in increasing order:

.10	.12	.23	.32	.45 - 1st quartile
.48	.50	.51	.53	.58 - 2nd quartile
.59	.66	.67	.69	.77 - 3rd quartile
.89	.95	.99	1.10	1.20 - 4th quartile

Range = 1.20 − .10 = 1.10 minutes
Interquartile range = $Q_3 - Q_1$ = .77 − .45 = .32 minutes

3-57 First we arrange the data in increasing order:

28	31	34	35	35	37	38	38	38	38	40	41 - 1st quartile
41	42	43	43	45	45	46	46	47	48	49	49 - 2nd quartile
49	49	50	51	52	52	52	52	55	55	57	57 - 3rd quartile
58	60	60	61	61	64	65	66	68	69	69	72 - 4th quartile

Interquartile range = $Q3 - Q1$ = 57 − 41 = \$16
Since 16 > 14, their hopes were not confirmed.

3-58 Range = highest − lowest = 502.6 − 6.3 = 496.3 megabytes
Interquartile range = $Q_3 - Q_1$ = 405.6 − 29.5 = 376.5 megabytes.

3-59 Since there are 30 total observations, we determine where the percentiles fall:

Percentile	Observation	Interfractile Range
20	10	---
40	13	3
60	15	2
80	17	2

3-60 First we arrange the data in increasing order:

43 57 104 162 201 220 253 302 380 467 500 633

a) 2nd decile = 57 (2nd smallest data value)
8th decile = 467 (10th smallest data value)
Interfractile range = 467 − 57 = 410

b) Median = (220 + 253)/2 = 236.5
Q_1 = 104 (3rd smallest value)
Q_3 = 380 (9th smallest value)

c) Interquartile range = $Q_3 - Q_1$ = 380 − 104 = 276

3-61 $N = 24, \sum x = 180, \sum x^2 = 1356.4$

$$\mu = \frac{\sum x}{N} = \frac{180}{24} = 7.5 \text{ ounces, meeting the chef's standard.}$$

$$\sigma^2 = \frac{\sum x^2}{N} - \mu^2 = \frac{1356.4}{24} - (7.5)^2 = 0.2667, \text{ so } \mu = \sqrt{0.2667} = 0.5164 \text{ ounces.}$$

This is slightly more variable than desired, so she should reject this batch, unless she is willing to relax her standards a bit.

3-62

x	17	21	18	27	17	21	20	22	18	23
x^2	289	441	324	729	289	441	400	484	324	529

$\sum x = 204 \quad \sum x^2 = 4250$

$\bar{x} = \sum x/n = 204/10 = 20.4$

$s^2 = (\sum x^2 - n\bar{x}^2)/(n - 1) = (4250 - 10(20.4)^2)/9 = 9.8222$

$s = \sqrt{s^2} = \sqrt{9.8222} = 3.1340$

This exceeds the allowable variability of 3 boats per day; she should be concerned.

3-63 $\mu = 66.8 \qquad \sigma^2 = 12.60 \qquad \sigma = \sqrt{12.60} = 3.55 \qquad n = 60$

a) 75% between $\mu \pm 2\sigma$: $66.8 \pm 2(3.55) = 66.8 \pm 7.1 = (59.7, 73.9)$

b) $59.7 = \mu - 2\sigma$, $73.9 = \mu + 2\sigma$
In a symmetrical, bell-shaped distribution $\mu \pm 2\sigma$ should include 95% of the observations and 95% of 60 is approximately 57.

c) (1) $z = \dfrac{61.45 - 66.8}{3.55} = -1.51$

(2) $z = \dfrac{75.37 - 66.8}{3.55} = 2.41$

(3) $z = \dfrac{84.65 - 66.8}{3.55} = 5.03$

(4) $z = \dfrac{51.50 - 66.8}{3.55} = -4.31$

3-64

Class	Midpoint (x)	Frequency (f)	$f \times x$	$x - \mu$	$(x - \mu)^2$	$f(x - \mu)^2$
0 - 199	100	10	1000	−490	240100	2401000
200 - 399	300	13	3900	−290	84100	1093300
400 - 599	500	17	8500	− 90	8100	137700
600 - 799	700	42	29400	110	12100	508200
800 - 999	900	18	16200	310	96100	1729800
		100	59000			5870000

$\mu = \dfrac{\sum f \times x}{N} = \dfrac{59000}{100} = 590$ checks per day

$\sigma^2 = \dfrac{\sum f(x - \mu)^2}{N} = \dfrac{5870000}{100} = 58700$

$\sigma = \sqrt{58700} = 242.28$ checks per day. Since this is $>$ 200, Hank should worry.

3-65

Observation (x)	$x - \mu$	$(x - \mu)^2$
5.25	−3.50	12.2500
7.50	−1.25	1.5625
8.75	0.00	0.0000
9.50	0.75	0.5625
10.50	1.75	3.0625
11.00	2.25	5.0625
52.50		22.5000

$\mu = \frac{\sum x}{N} = \frac{52.5}{6} = 8.75\%$ $\sigma^2 = \frac{\sum(x-\mu)^2}{N} = \frac{22.5}{6} = 3.75$, so $\sigma = \sqrt{3.75} = 1.94\%$

We can see that the only effect of the constant added to our original distribution will be to increase the mean by the amount of the constant. The variance and standard deviation remain the same. The constant has no effect on the original "shape" of our distribution; it simply "shifts" the distribution. Since the variance is a measure of the shape, we can predict this result without redoing the computation.

3-66

Class	Midpoint (x)	Frequency (f)	$f \times x$	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
1 - 3	2	18	36	−5.715	32.6612	587.9021
4 - 6	5	90	450	−2.715	7.3712	663.4102
7 - 9	8	44	352	0.285	0.0812	3.5739
10 - 12	11	21	231	3.285	10.7912	226.6157
13 - 15	14	9	126	6.285	39.5012	355.5110
16 - 18	17	9	153	9.285	86.2112	775.9010
19 - 21	20	4	80	12.285	150.9212	603.6849
22 - 24	23	5	115	15.285	233.6312	1168.1561
		200	1543			4384.7550

a) $\bar{x} = \frac{\sum f \times x}{n} = \frac{1543}{200} = 7.715$ days

$s^2 = \frac{\sum f(x-\bar{x})^2}{n-1} = \frac{4384.755}{199} = 22.0339$, so $s = \sqrt{22.0339} = 4.69$ days

b) The interval 0 to 17 is roughly the mean $\pm$ 2 standard deviations, so about 75% of he data, or $.75(200) = 150$ observations should fall in the interval. In fact something between 182 and 191 of the observations are in the interval.

c) About 95% or $.95(200) = 191$ stays can be expected to fall in the interval from 0 to 17.

3-67 a)

Class	Midpoint (x)	Frequency (f)	$f \times x$	x^2	$f \times x^2$
11.0 - 11.9	11.5	2	23.0	132.25	264.50
12.0 - 12.9	12.5	2	25.0	156.25	312.50
13.0 - 13.9	13.5	8	108.0	182.25	1458.00
14.0 - 14.9	14.5	10	145.0	210.25	2102.50
15.0 - 15.9	15.5	11	170.5	240.25	2642.75
16.0 - 16.9	16.5	8	132.0	272.25	2178.00
17.0 - 17.9	17.5	3	52.5	306.25	918.75
18.0 - 18.9	18.5	1	18.5	342.25	342.25
		45	674.5		10219.25

$\bar{x} = \frac{\sum(f \times x)}{n} = \frac{674.5}{45} = 14.9889\%$

$s^2 = \frac{\sum(f \times x^2)}{n-1} - \frac{n\bar{x}^2}{n-1} = \frac{10219.25}{44} - \frac{45(14.9889)^2}{44} = 2.4828$

$s = \sqrt{2.4828} = 1.5757\%$

b) We can expect at least 75% of the observations to fall in the interval:

$\bar{x} \pm 2s = 14.9889 \pm 2(1.5757) = 14.9889 \pm 3.1514 = (11.8375, 18.1403)$

All of the observations in the 2nd through 7th classes fall in that interval, and some of those in the first and last classes may also fall in it. Hence, at least 42/45 or 93.33% of the observations fall between 11.8375 and 18.1403.

c) We can expect roughly 68% of the observations to fall in the interval:

$$\bar{x} \pm s = 14.9889 \pm 1.5757 = (13.4132, 16.5646)$$

All of the observations in the 14.0–14.9 and 15.0–15.9 classes fall in that interval, and roughly half of those in the 13.0–13.9 and 16.0–16.9 classes are also in that interval. Hence about 29/45 or 64.44% of the observations fall between 13.4132 and 16.5646.

3-68 First, we will calculate the standard deviation for the distribution:

$$\sigma = \sqrt{\sigma^2} = \sqrt{49{,}729} = 223$$

A production of 11,175 loaves is one standard deviation below the mean (11,398 − 11,175) = 223. Assuming that the distribution is symmetrical, we know that within ± one standard deviation from μ fall about 68% of all observations. The interval from the mean to one standard deviation below the mean would contain about 34% (68% ÷ 2) of the data. Therefore, 50% − 34% = 16% (or approximately 5 weeks) of the data would be below 11,175 loaves.

A similar argument can be made for two standard deviations above the mean. Again, assuming a symmetrical distribution, 47.5% of the data (95% ÷ 2) would be contained within this range. This leaves 50% − 47.5% = 2.5%, or about one week (.025 × 32 = .8) with production above 11,844 loaves.

3-69 Washington: $\dfrac{1100 - 1500}{400} = -1.00$ New York: $\dfrac{3200 - 3760}{622} = -0.90$

Durham: $\dfrac{500 - 850}{95} = -3.68$

The Durham raise is the farthest from its mean and therefore is the lowest of the three relative to its mean and standard deviation.

3-70 The key to this problem is to express the actual deviations in terms of standard scores. For example, the first product response latency is off

$$\frac{2.495 - 2.500}{.004} = \frac{-.005}{.004} = -1.25 \text{ standard scores.}$$

For the other products,

$$\text{II: } \frac{2.790 - 2.800}{.006} = \frac{-.010}{.006} = -1.67 \qquad \text{III: } \frac{3.900 - 3.700}{.09} = \frac{.20}{.09} = 2.22$$

Disregarding the signs, we can see that product III produced the response latency with the largest deviation.

3-71 a)

x	33	50	22	27	48	$\sum x$	= 180
x^2	1089	2500	484	729	2304	$\sum x^2$	= 7106

$$\bar{x} = \frac{\sum x}{n} = \frac{180}{5} = 36 \text{ patients}$$

$$s^2 = \left(\frac{1}{n-1}\right)\left(\sum x^2 - n\bar{x}^2\right) = \frac{7106 - 5(36)^2}{4} = 156.5 \text{ patients squared,}$$

so, $s = \sqrt{156.5} = 12.51$ patients.

The middle 75% should be in the interval $\bar{x} \pm 2s = 36 \pm 2(12.51) = (11, 61)$ patients.

b)

x	34	31	37	36	27	$\sum x$	= 165
x^2	1156	961	1369	1296	729	$\sum x^2$	= 5511

$$\bar{x} = \frac{\sum x}{n} = \frac{165}{5} = 33 \text{ patients}$$

$$s^2 = \left(\frac{1}{n-1}\right)\left(\sum x^2 - n\bar{x}^2\right) = \frac{5511 - 5(33)^2}{4} = 16.5 \text{ patients squared,}$$

so, $s = \sqrt{16.5} = 4.06$ patients.

With such a small sample, it is hard to say that the distribution is symmetric and bell-shaped, but if it were, the middle 68% of the data would be expected to fall in the interval $\bar{x} \pm s = 33 \pm 4.06 = (29, 37)$ patients.

3-72 $\sigma^2 = 100{,}000{,}000$ dollars squared, so $\sigma = \$10{,}000$; $\mu = \$390{,}000$.
The desired interval is: $\mu \pm 2\sigma = \$390{,}000 \pm 2(\$10{,}000) = (\$370{,}000, \$410{,}000)$.

3-73 Standard deviation units from the mean (or z-scores):

Child: $\frac{90 - 110}{\sqrt{81}} = \frac{-20}{9} = -2.22$ Young Adult: $\frac{92 - 90}{\sqrt{64}} = \frac{2}{8} = 0.25$

Adult: $\frac{100 - 95}{\sqrt{49}} = \frac{5}{7} = 0.71$ Elderly: $\frac{98 - 90}{\sqrt{121}} = \frac{8}{11} = 0.73$

Thus, the elderly patient has the highest z-score.

3-74 Bullets: $CV = \frac{\sigma}{\mu}(100) = \frac{18(100)}{224} = 8.04\%$

Trailblazers: $CV = \frac{\sigma}{\mu}(100) = \frac{12(100)}{195} = 6.15\%$

The Bullets have the greater relative dispersion.

3-75

Bulb	1	2	3
μ	1470	1400	1350
σ	$\sqrt{156}$	$\sqrt{81}$	6
CV=100(σ/μ)	0.85%	0.64%	0.44%

Bulb 3 was best and bulb 1 was worst.

3-76 Regular MBA: $\bar{x} = 24.8$, $s = 2.486$, $CV = (s/\bar{x})(100) = \frac{2.486\,(100)}{24.8} = 10.02\%$

Evening MBA: $\bar{x} = 30.4$, $s = 2.875$, $CV = (s/\bar{x})(100) = \frac{2.875\,(100)}{30.4} = 9.46\%$

There is not much difference between the two groups.

3-77 a)

Salesperson	$\bar{x}$	s	CV = $(s/\bar{x})(100)$
Patricia	88	12.67	14.4%
John	83.8	6.02	7.2%
Frank	104.2	16.35	15.7%

John is the most consistent, both in terms of standard deviation and coefficient of variation.

b) If the averages are close to each other, then a measure of consistency could be an appropriate way to distinguish between the salespersons' levels of performance. In this case, however, since Frank is more effective than John in every year, arguing John's greater consistency is not particularly convincing.

c) When the means are so very different, no measure of consistency is very appropriate.

3-78 Company 1: $CV = \frac{\sigma}{\mu}(100) = \frac{5.3(100)}{28} = 18.93\%$

Company 2: $CV = \frac{\sigma}{\mu}(100) = \frac{4.8(100)}{37.8} = 12.70\%$

Company 1 pursued the riskier strategy.

3-79 Machine 1: $CV = \frac{\sigma}{\mu}(100) = \frac{5.2(100)}{100} = 5.20\%$

Machine 2: $CV = \frac{\sigma}{\mu}(100) = \frac{8.6(100)}{180} = 4.78\%$

Machine 1 is the less accurate of the two from a relative standpoint.

3-80

Employee	John	Jeff	Mary	Tammy
$\bar{x}$	66.67	67.40	71.83	61.20
s	3.78	1.14	9.70	4.66
$CV=100(\sigma/\bar{x})$	5.67%	1.69%	13.50%	7.61%

Jeff is the best employee, since he has the lowest coefficient of variation.

3-81

Grade	$\bar{x}$	s
Regular	87.8	5.17
Extra	89.8	2.28
Super	87.8	5.26

The grading makes little sense. In terms of both average rate of germination and variability of that rate, the Super grade closely resembles the Regular grade, and both are inferior to the Extra grade.

3-82 Configuration 1: $CV = \frac{\sigma}{\mu}(100) = \frac{4.8(100)}{34.8} = 13.79\%$

Configuration 2: $CV = \frac{\sigma}{\mu}(100) = \frac{7.5(100)}{25.5} = 29.41\%$

Configuration 3: $CV = \frac{\sigma}{\mu}(100) = \frac{3.8(100)}{37.5} = 10.13\%$

The third configuration has the least relative variation.

3-83

Value	3.97	3.98	3.99	4.00	4.01	4.02	4.03
Frequency	1	2	4	4	4	4	1

$$\bar{x} = \frac{\sum x}{n} = \frac{80.04}{20} = 4.002 \text{ ounces}$$

$$s = \sqrt{\frac{1}{n-1}\left(\frac{\sum(x-\bar{x})^2}{n-1}\right)} = \sqrt{\frac{1}{19}\left(320.325 - 20(4.002)^2\right)} = .016 \text{ ounces}$$

If the population is approximately bell-shaped, then 95% of the packets will have weights in the interval

$$\bar{x} \pm 2s = 4.002 \pm 2(.016) = (3.970, 4.034) \text{ ounces}$$

3-84 This statement is incorrect, because it completely ignores the variability in yards gained per carry. If the Raiders gain 85 yards once in every 200 carries, but gain 1.382 yards in the other 199 carries, then they average 3.6 yards a carry, but will rarely get a first down.

3-85 A person making this statement has missed the point of variability. By definition, you always have an equal chance of falling above or below the median, regardless of the variability. We are not considering the average result when we consider variability, however. Instead, we are looking at what a single outcome might be. Even though the outcome might be the same on average, you would not want to be the "one guy in a thousand" who happens to get the lowest possible result. The variability of a distribution describes this risk of having an outcome other than the mean value. You might say that variability tells us how bad or how good the outcome might be and gives us a relative measure of how likely these extreme values are.

3-86 Since military manpower levels, staffing, and salaries are known with a fair amount of certainty, we could probably assign Curve A to the distribution of outcomes of actual officer salaries. Food purchases are also known with a fair amount of certainty, but prices are less certain. Thus, the distribution of outcomes of actual food purchases would probably fit Curve B. Finally, Curve C would be the best fit for the distribution of outcomes of actual aircraft maintenance since this expense would be fairly uncertain.

3-87 We should use the line with the higher probability of having a breaking strength above 25 pounds, and this is equivalent to chosing the line with the more negative z-score.

Now $z_{40} = \dfrac{25-40}{\sigma_{40}} = \dfrac{-15}{\sigma_{40}}$ and $z_{30} = \dfrac{25-30}{\sigma_{30}} = \dfrac{-5}{\sigma_{30}}$

Hence $z_{30} < z_{40}$ provided $\dfrac{-5}{\sigma_{30}} < \dfrac{-15}{\sigma_{40}}$, i.e., provided $\sigma_{30} < \sigma_{40}/3$.

Since we are told that σ_{30} is quite small and σ_{40} is quite large, our condition probably holds, and so the 30-pound-test line should be used.

3-88 Perhaps the company has changed its hiring policy; specifically, it may now be hiring some less experienced sales reps and providing them with on-the-job training. This action, however, would also reduce the mean; thus hiring less experienced sales reps could not change the variation without also changing the mean. Perhaps the company is simultaneously hiring some highly experienced sales reps along with the less experienced ones. The higher scores of the former would balance out the lower scores of the latter, leaving the mean relatively unchanged while the variation increased.

3-89 a)

x	x^2
200	40000
156	24336
231	53361
222	49284
96	9216
289	83521
126	15876
308	94864
1628	370458

$\mu = \dfrac{\sum x}{N} = \dfrac{1628}{8} = 203.5$ cars

range $= 308 - 96 = 212$ cars

interquartile range $= Q_3 - Q_1 = 231 - 126 = 105$

standard deviation: $\sigma = \sqrt{\dfrac{\sum x^2}{N} - \mu^2} = \sqrt{\dfrac{370458}{8} - (203.5)^2} = 69.96$ cars

b) The standard deviation and interquartile range are less affected by the outliers at 96 an 308, so of the three they are the better measures of variability.

3-90 The later period will show both a higher mean and a higher variability.

3-91 Line 1: $CV = \frac{\sigma}{\mu}(100) = \frac{1050(100)}{11350} = 9.25\%$

Line 2: $CV = \frac{\sigma}{\mu}(100) = \frac{1010(100)}{9935} = 10.17\%$

Line 2 has the greater relative dispersion.

3-92 a)

x	17	21	44	50	79	86	140	178	203
x^2	289	441	1936	2500	6241	7396	19600	31684	41209

$\sum x = 818$ $\sum x^2 = 111{,}296$

$$\mu = \frac{\sum x}{N} = \frac{818}{9} = 90.8889$$

b) The median is the (9+1)/2 = 5th element, that is, 79.

c) Since each observation appears only once, there is no mode.

d) Since the data tend to cluster at the lower end of the range and the mean is pulled up by the the three high observations, the median is the better measure of central tendency.

e) $\sigma^2 = \sum x^2/N - \mu^2 = 111,296/9 - 90.8889^2 = 4105.43$ $\sigma = \sqrt{4105.43} = 64.074$

3-93

x	$x - \bar{x}$	$(x - \bar{x})^2$	x	$x - \bar{x}$	$(x - \bar{x})^2$
88	− 55.1	3036.01	140	− 3.1	9.61
90	− 53.1	2819.61	144	− 0.9	0.81
99	− 44.1	1944.81	145	1.9	3.61
100	− 43.1	1857.61	156	12.9	166.41
101	− 42.1	1772.41	165	21.9	479.61
130	− 13.1	171.61	169	25.9	670.81
130	− 13.1	171.61	188	44.9	2016.01
130	− 13.1	171.61	192	48.9	2391.21
132	− 11.1	123.21	208	64.9	4212.01
139	− 4.1	16.81	216	72.9	5314.41
			2862		27349.80

$$\bar{x} = \frac{2862}{20} = 143.1 \text{ pounds}$$

range = 216 − 88 = 128 pounds

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{27349.80}{19} = 1439.46 \text{ pounds}^2$$

$$s = \sqrt{1439.46} = 37.94 \text{ pounds}$$

The data are reasonably well spread out over the range, so it is not a bad measure of the variability, but even in this case the range is strongly affected by the extreme values.

3-94 Formula #1: $CV = \frac{\sigma}{\mu}(100) = \frac{35(100)}{700} = 5.0\%$

Formula #2: $CV = \frac{\sigma}{\mu}(100) = \frac{16(100)}{300} = 5.3\%$

Formula #2 is relatively less accurate because it has the greater coefficient of variation.

3-95

x	x^2
1.66	2.7556
1.68	2.8224
1.69	2.8561
1.71	2.9241
1.73	2.9929
1.77	3.1329
1.83	3.3489
1.89	3.5721
13.96	24.4050

$$\mu = \frac{\sum x}{N} = \frac{13.96}{8} = \$1.745$$

Interquartile range $= Q_3 - Q_1 = 1.77 - 1.69 = \$0.08$

$$\sigma^2 = \frac{\sum x^2}{N} - \mu^2 = \frac{24.405}{8} - (1.745)^2 = .0056 \text{ dollars squared}$$

$$\sigma = \sqrt{.0056} = \$0.0748$$

However you measure it, the variation in average price of heating fuel across the eight states is quite small.

3-96 a) Neither the variance nor the standard deviation would be a good measure of the variability.

b) The data clearly came from two distinct populations: one for weekdays (Sunday through Thursday), the other for weekends (Friday and Saturday). This fact would be lost if we lumped all seven days together to compute a single measure of variability.

3-97 a)

x	$x-\bar{x}$	$(x-\bar{x})^2$	x	$x-\bar{x}$	$(x-\bar{x})^2$
134	−9	81	145	2	4
136	−7	49	146	3	9
137	−6	36	146	3	9
138	−5	25	146	3	9
138	−5	25	147	4	16
143	0	0	148	5	25
144	1	0	153	10	100
144	1	1	2145		390

$$\bar{x} = \frac{\sum x}{n} = \frac{2145}{15} = 143$$

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{390}{14}} = 5.28$$

b) The given interval is $\bar{x} \pm 2s$, which should contain about 75% of the data, i.e., about .75(15) = 11.25 observations. In fact, all of the observations fall between 132.44 and 153.56.

3-98 To find the deciles, we arrange the 40 observations in ascending order:

	9.1	10.4	11.9	12.8	13.7	14.6	15.8	16.9	17.8	18.8
	9.3	10.6	12.1	13.0	13.9	14.7	16.0	17.1	18.0	19.0
	9.6	10.9	12.4	13.3	14.2	15.0	16.3	17.4	18.3	19.3
	9.9	11.2	12.7	13.6	14.5	15.3	16.6	17.7	18.6	19.6
deciles	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th

Eighty percent of the trucks delivered fewer than 17.8 tons.

3-99 a)

x	$x - \bar{x}$	$(x - \bar{x})^2$
20,100	– 6,851	46,936,201
24,500	– 2,451	6,007,401
31,600	4,649	21,613,201
28,400	1,449	2,099,601
49,500	22,549	508,457,401
19,350	– 7,601	57,775,201
25,600	– 1,351	1,825,201
30,600	3,649	13,315,201
11,300	15,651	244,953,801
28,560	1,609	2,588,881
269,510		905,572,090

$$\bar{x} = \frac{\sum x}{n} = \frac{269,510}{10} = 26,951$$

range = 49,500 – 11,300 = 38,200

$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{905,572,090}{9} = 100,619,121.1$$

$$s = \sqrt{100,619,121.1} = 10,030.91$$

b) The standard deviation is a good measure of the variability in the data.

c) If the data are viewed in isolation, no other measure is needed. However, to compare the variability in the Eagles' attendance figures with those of other teams, we might want to look at the coefficient of variation.

d) $\text{CV} = (s/\bar{x})(100) = \dfrac{10,031\,(100)}{26,951} = 37.22\%$

3-100 a) Range will give the least information because it considers only the highest and lowest observations. Standard deviation considers all of the data.

b) The range is certainly the easier to compute of the two measures.

c) In this case, since the data are fairly evenly spread out between the low of 210 days and the high of 231 days, the range is a reasonable measure of the variability in the data, so it may not be necessary to consider one of the other measures.

3-101 a) L companies:

Class	Midpoint (x)	Frequency (f)	$f \times x$	x^2	$f \times x^2$
-1.25 to -1.01	-1.125	1	-1.125	1.265625	1.265625
-1.00 to -0.76	-0.875	1	-0.875	0.765625	0.765625
-0.75 to -0.51	-0.625	1	-0.625	0.390625	0.390625
-0.50 to -0.26	-0.375	7	-2.625	0.140625	0.984375
-0.25 to -0.01	-0.125	19	-2.375	0.015625	0.296875
0.00	0.000	14	0.000	0.000000	0.000000
0.01 to 0.25	0.125	21	2.625	0.015625	0.328125
0.26 to 0.50	0.375	5	1.875	0.140625	0.703125
0.51 to 0.75	0.625	3	1.875	0.390625	1.171875
0.76 to 1.00	0.875	2	1.750	0.765625	1.531250
1.01 to 1.25	1.125	1	1.125	1.265625	1.265625
		75	1.625		8.703125

$\mu = \sum x/N = 1.625/75 = 0.0217$

R companies:

Class	Midpoint (x)	Frequency (f)	$f \times x$	x^2	$f \times x^2$
-1.25 to -1.01	-1.125	1	-1.125	1.265625	1.265625
-1.00 to -0.76	-0.875	1	-0.875	0.765625	0.765625
-0.75 to -0.51	-0.625	0	0.000	0.390625	0.000000
-0.50 to -0.26	-0.375	5	-1.875	0.140625	0.703125
-0.25 to -0.01	-0.125	20	-2.500	0.015625	0.312500
0.00	0.000	20	0.000	0.000000	0.000000
0.01 to 0.25	0.125	14	1.750	0.015625	0.218750
0.26 to 0.50	0.375	8	3.000	0.140625	1.125000
0.51 to 0.75	0.625	1	0.625	0.390625	0.390625
0.76 to 1.00	0.875	4	3.500	0.765625	3.062500
1.01 to 1.25	1.125	0	0.000	1.265625	0.000000
		74	2.500		7.843750

$\mu = \sum x/\text{N} = 2.500/74 = 0.0338$

b) L companies: $\tilde{m} = \left(\dfrac{(n+1)/2 - (F+1)}{f_m}\right)w + L_m$

$$= \left(\frac{(75+1)/2 - (29+1)}{14}\right)(0) + 0 = 0$$

R companies: $\tilde{m} = \left(\dfrac{(n+1)/2 - (F+1)}{f_m}\right)w + L_m$

$$= \left(\frac{(74+1)/2 - (27+1)}{20}\right)(0) + 0 = 0$$

c) L companies: $\text{M}_0 = \text{L}_{\text{M}_0} + \dfrac{d_1}{d_1+d_2}w = 0.01 + 7\,(0.25) = 0.861$

R companies: Taking the class from -0.25 to -0.01 as the modal class,

$$\text{M}_0 = \text{L}_{\text{M}_0} + \frac{d_1}{d_1+d_2}w = -0.25 + \frac{15}{15+0}(0.25) = 0$$

Taking the class at 0.00 as the modal class gives the same result, since $\text{L}_{\text{M}_0} = w = 0$ for that class.

d) For each distribution, the median is probably the best measure of central tendency. But all three measures are close to each other, so it doesn't make much difference which one you use.

e) L companies: $\sigma^2 = \sum fx^2/\text{N} - \mu^2 = 8.7031/75 - 0.0217^2 = 0.1156$

$\sigma = \sqrt{\sigma^2} = 0.1156 = 0.3400$

R companies: $\sigma^2 = \sum fx^2/\text{N} - \mu^2 = 7.8438/74 - 0.0338^2 = 0.1049$

$\sigma^2 = \sqrt{\sigma^2} = 0.1049 = 0.3239$

f) L companies: CV $= \sigma(100)/\mu = 0.3400\,(100)/0.0217 = 1567$
R companies: CV $= \sigma(100)/\mu = 0.3239\,(100)/0.0338 = 958$
Hence the price changes of the R companies show less relative variability.

3-102 a) Listing the data in ascending order:

0 2 4 4 5 6 7 8 10 11 14 19 21 29

we see that the median is 7.5 days, halfway between the 7th and 8th observations.

b) $\mu = \sum x/N = 140/14 = 10$ days

3-103 a) 140 days/38 machines = 3.68 days/machine

b)

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Downtime ($f \cdot x$)	2	19	14	21	5	7	11	8	29	6	0	4	4	10
# of pieces (f)	1	3	1	4	2	1	1	5	8	2	2	6	1	1
Downtime/piece	2	6.33	14	5.25	2.5	7	11	1.6	3.625	3	0	0.67	4	10

c) Seven groups (2, 3, 4, 6, 7, 13, and 14) had higher than average downtime per piece of machinery.

3-104 The weekly news magazines would probably have the highest average readerships, the medical journals the smallest average readerships, with the monthly magazines somewhere in the middle.

Monthly magazines and medical journals, with many low circulation items and few high circulation items are likely to be skewed to the right. There are only a few weekly news magazines, so it's difficult to assess the skewness of this distribution.

3-105 The Federal tax payments have the largest, the North Carolina tax payments the middle, and the airport tax payments the smallest central tendency.

The Federal and North Carolina tax payments are likely to be skewed to the right, since both have a large number of low- or middle-income taxpayers and a smaller number of high-income taxpayers. The airport tax payments will not be skewed, since the tax is applied at a uniform rate per passenger.

3-106 a) Listing the data in ascending order:

4.77 4.89 4.91 5.02 5.05 5.22 5.24 5.27
5.75 5.99 6.01 6.02 6.05 6.11 6.11 6.11

we see that the median is 5.51 mpg, halfway between the 8th and 9th observations.

b) $x = \sum x/n = 88.52/16 = 5.5325$ mpg

c)

Class (mpg)	4.77-5.03	5.04-5.30	5.31-5.57	5.58-5.84	5.85-6.11
Frequency	4	4	0	1	7

The modal class is 5.85-6.11 mpg.

d) It depends. If she is ordering fuel for only one car, she should be cautious and use use the modal value. If she is ordering fuel for several cars running in the same race, the mean or median is probably ok.

3-107 The median. The data will be skewed to the right, with the modal class determined by the large number of low-income taxpayers. The mean will be pulled to the right by the relatively small number of taxpayers with extremely high incomes.

3-108 a) Listing the data in ascending order:

21 23 25 26 26 32 33 33 33 34 36 37 37 37 37 43 45 47 47 56

we see that the median is 35 bulbs, halfway between the 10th and 11th observations.

$\bar{x} = \sum x/n = 708/20 = 35.4$ bulbs

b) Since the mean is bigger than the median, the distribution is skewed to the right.

3-109 a) Listing the data in ascending order:

										$\sum x$	$\sum x/9$
A	14	15	16	16	17	20	30	31	53	212	23.56
B	17	18	21	22	23	26	27	28	39	221	24.56
C	15	16	16	17	18	19	20	31	42	194	21.56

we see that the medians are the fifth element in each group: 17, 23, and 18 hours. The means are 23.56, 24.56, and 21.56 hours. Design B is best; both the mean and the median are highest.

3-110 a)

Class (mm)	Frequency	Cumulative Frequency
$\leq$ 1.00	12	12
1.01-1.50	129	141
1.51-2.00	186	327
2.01-2.50	275	602
2.51-3.00	341	943
3.01-3.50	422	1365
3.51-4.00	6287	7652
4.01-4.50	8163	15815
4.51-5.00	6212	22027
5.01-5.50	2416	24443
$\geq$5.51	1019	25462

The modal class is 4.01-4.50mm, which is also the median class. The median value is approximately

$$\tilde{m} = \left(\frac{(n+1)/2 - (F+1)}{f_m}\right) w + L_m = \left(\frac{25463/2 - 7653}{8163}\right)(.5) + 4.01 = 4.32\text{mm}$$

b) The coarse 3.5 mm screen suffices to remove at least half of the debris.

3-111 a) 365(16,500) = 6,022,500 $\neq$ 3,976,951, so the reported average is <u>not</u> an arithmetic mean.

b) No, scheduling 100 people <u>each</u> day does not make sense. Some days will attract a heavy volume of calls, where as others (like the day after Thanksgiving, for example) will have light volume.

3-112 $\bar{x} = (7.41 + 7.24 + 5.15 + 5.09 + 4.61 + 2.77 + 2.67 + 2.00 + 0.14)/9 = \$4.12$

median = 5th observation = \$4.61

The median is better, since the mean is distorted by the observation for Southwest (\$0.14), which is clearly an outlier.

3-113 Yes. The disparity between the median and the mean indicates that the distribution of investable funds is highly skewed, so there are many families with significant amounts of funds available to invest. Merrill Lynch should target its marketing to reach that group. In fact, the Merrill Lynch study cited in the problem reported further that the least wealthy 40% of the population had <u>no</u> investable cash at all, so that efforts aimed at the general public will be wasted on at least 40% of the families reached by those efforts.

CHAPTER 4

PROBABILITY I: INTRODUCTORY IDEAS

4-1 The fact that the probability of the death of each policy holder is certain is not the key to life insurance premiums. Whereas with fire insurance the question is whether a given house will burn down or not, the question with life insurance is what proportion of the policyholders will die at a given age. The student should realize that probability theory is used to determine the timing of the deaths. Probability theory allows the insurance company to determine the timing of the future cash flows and, thus, the premiums which will be needed to generate these cash flows.

4-2 The FDA conducted extensive experiments, exposing some laboratory animals to saccharin and keeping other animals (the "control group") free of saccharin intake. These experiments indicated that, with all other factors held constant (as much as possible), animals exposed to saccharin were more likely to develop cancer than those not exposed. This is an example of sampling from a larger population (the population of all laboratory animals, or, in a larger sense, the entire animal kingdom).

The warning statement is very much a probability statement for at least two reasons. First, it addresses the likelihood of cancer in laboratory animals under two sets of conditions, with the conclusion that the likelihood of cancer is higher under conditions including exposure to saccharin. Second, it claims that humans and laboratory animals are likely to react similarly to saccharin intake. Clearly, the FDA deems this likelihood to be high enough to merit the use of the warning statement.

4-3 This question involves the concept that any alternative can be assigned some probability. The student should realize that any decision involves at least a subjective estimate of the probabilities involved. These estimates may take the form of "gut feel" or "intuition," but estimates are made just the same. Otherwise, decisions would be impossible. Therefore, there is no such thing as a risk which is not "calculated" in some way.

4-4 This decision involves estimates of consumer preference, brand loyalty, competitor response, and numerous other factors, which all involve uncertainty. Thus the only estimates possible are probability-based.

4-5 a) Mutually exclusive (b), (d)

b) a, b, d

4-6 a) (ball, strike) (ball, ball) (strike, strike) (strike, ball)

b)

(ball, ball, ball)	(strike, ball, ball)
(ball, ball, strike)	(strike, ball, strike)
(ball, strike, strike)	(strike, strike, ball)
(ball, strike, ball)	(strike, strike, strike)

4-7 There are 54 possible elements, having the form (d, c):

d is the number of spots on the die and ranges from 1 to 6
c is the number of pips on the card and ranges from 2 to 10

4-8

Sum	Ways to get this sum (card, die)	Probability
0	impossible	0
2	(1, 1)	1/54
3	(1, 2), (2, 1)	2/54
8	(2, 6), (3, 5), ..., (7, 1)	6/54
9	(3, 6), (4, 5), ..., (8, 1)	6/54
12	(6, 6), (7, 5), ..., (10, 2)	5/54
14	(8, 6), (9, 5), (10, 4)	3/54
16	(10, 6)	1/54

4-9 a) Only two events would apply to this decision: Royal wins or Royal loses.

b) The list in part (a) is collectively exhaustive and the events are mutually exclusive.

c) Knowing nothing about the decision, you would be forced to assign a probability of 1/2 (or .50) to each event (i.e., they are equally likely).

4-10 a) The segments are collectively exhaustive, because they are considered to be the only ones worthy of special campaigns. They are not mutually exclusive, because more than one campaign can be funded, depending on the total amount spent.

b) The list of collectively exhaustive and mutually exclusive events for the spending decision (with A = minorities, B = business people, etc.) is:

A only	A,C only	B,E only	A,C,E only
B only	A,D only	C,D only	A,D,E only
C only	A,E only	C,E only	C,D,E only
D only	B,C only	D,E only	
E only	B,D only		

c) Yes. The new list, which would again be both collectively exhaustive and mutually exclusive, is:

B,C only	A,C,D only
B,E only	A,D,E only

4-11 a) P(seven) = 4/52 = 1/13

b) P(black card) = 26/52 = 1/2

c) P(ace or king) = 8/52 = 2/13

d) P(black two or three) = 4/52 = 1/13

e) P(red face card) = 6/52 = 3/26

f) classical probability

4-12 a) 6/26

b) 5/26

c) 1/2

d) 1/4

4-13 a) P($5000 - $9999) = 25/300 = 1/12

b) P(less than $15000) = (15 + 25 + 35)/300 = 1/4

c) P(more than \$20000) = (70 + 30)/300 = 1/3

d) P(\$15000 - \$19999) = 125/300 = 5/12

4-14 If we let $P(50 - 74\%) = P$, then we know that:

$P(75 - 99\%) = 1/2\ P$ and $P(25 - 49\%) = 2/5\ P$

Further, $P(0 - 24\%) = 0$ and $P(100\%) = 1/20 = .05$

Thus, $2/5\ P + P + 1/2\ P = 1.9P = .95$, which means that $P = .5$

Therefore,
$P(0 - 24\%) = 0$
$P(25 - 49\%) = .20$
$P(50 - 74\%) = .50$
$P(75 - 99\%) = .25$
$P(100\%) = .05$

4-15 $$\text{P(copier out of service)} = \frac{\#\text{ days out of service}}{\text{total days available for use}}$$

$$= \frac{51 + 43 + 2 + 31 + 13}{260 + 260 + 260 + 260 + 260} = .1077$$

4-16 a) subjective
b) relative frequency or subjective
c) classical
d) relative frequency
e) classical
f) relative frequency or subjective

4-17 $P(A) = 11/60$ $P(B) = 7/60$ $P(A \text{ or } B) = (11 + 7)/60 = .3$

4-18 $P(A) = 21/100$ $P(B) = 29/100$ $P(C) = 38/100$

$P(A \text{ or } B) = 45/100$ $P(A \text{ or } C) = 50/100$ $P(B \text{ but not } (A \text{or } C)) = 20/100$

4-19 $N = 75$ blue = 35, red = 40; 25 blue swirled, 30 red swirled; 10 blue clear, 10 red clear

a) P(blue) = 35/75 = 7/15

b) P(clear) = (10 + 10)/75 = 4/15

c) P(blue and swirled) = 25/75 = 1/3

d) P(red and clear) = 10/75 = 2/15

e) P(swirled) = (25 + 30)/75 = 11/15 (or 1 − P(clear) = 1 − 4/15 = 11/15)

4-20 a) Comparing the two expressions shows that when A and B are mutually exclusive, P(A and B) = 0. This is in line with the definition of mutually exclusive, which states that two events are mutually exclusive if they cannot both occur at the same time.

b) The correct expression would be:

$$\begin{aligned} P(A \text{ or } B \text{ or } C) = {} & P(A) + P(B) + P(C) - P(A \text{ and } B) \\ & - P(A \text{ and } C) - P(B \text{ and } C) + P(A, B, \text{ and } C) \end{aligned}$$

c) If A and B are mutually exclusive, then by definition P(A and B) = 0. Also, if A and B cannot occur simultaneously, then A, B, and C cannot occur simultaneously. Therefore, P(A, B, and C) = 0. The correct expression becomes:

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(A \text{ and } C) - P(B \text{ and } C)$$

d) If A and C are also mutually exclusive, then P(A and C) = 0. Therefore, the expression would be:

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(B \text{ and } C)$$

e) Finally, if all the events are mutually exclusive of each other, then

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

4-21 a) P(PC fails) = P(light pen or keyboard fails)
= P(light pen fails) + P(keyboard fails) — P(both fail)
= 0.025 + 0.150 — 0.005 = 0.17
Hence P(PC works) = 1 — 0.17 = 0.83

b) P(PC or mainframe fails) = P(PC fails) + P(mainframe fails) — P(both fail)
= 0.17 + 0.25 — 0 = 0.42

4-22 D = disk-drive failure K = keyboard failure

a) P(D or K) = P(D) + P(K) − P(D and K)
= P(D) + 3P(D) − .05 = 4P(D) − .05 = .20
Thus, 4P(D) = .25, so P(D) = .0625

b) P(K) = 2P(D) = .125
Thus, P(D or K) = P(D) + P(K) − .05 = .0625 + .125 − .05 = .1375
This means that the computer is 86.25% resistant to disk-drive and/or keyboard failure.

4-23 P(interior flaw) = 10/1000 = .01, P(casing flaw) = 8/1000 = .008, and
P(interior and casing flaw) = 5/1000 = .005
Thus, P(flaw) = P(interior flaw or casing flaw) = .01 + .008 − .005 = .013

4-24 a) $P(\text{Boy}_2 \mid \text{Girl}_1) = 1/2$ b) $P(\text{Girl}_2 \mid \text{Girl}_1) = 1/2$

4-25 a) $P(7_1) = P(7) = 1/6$; $P(11_2) = P(11) = 1/18$
Thus, $P(7_1 \text{ and } 11_2) = P(7_1)P(11_2) = 1/108$

b) P(1st + 2nd = 21)
$= P[(9_1 \text{ and } 12_2) \text{ or } (12_1 \text{ and } 9_2) \text{ or } (11_1 \text{ and } 10_2) \text{ or } (10_1 \text{ and } 11_2)]$
$= P(9_1 \text{ and } 12_2) + P(12_1 \text{ and } 9_2) + P(11_1 \text{ and } 10_2) + P(10_1 \text{ and } 11_2)$
$= 2P(9_1 \text{ and } 12_2) + 2P(10_1 \text{ and } 11_2)$
$= 2(1/9)(1/36) + 2(1/12)(1/18) = 5/324$

c) $P(\text{1st} + \text{2nd} + \text{3rd} = 6) = P(2_1 \text{ and } 2_2 \text{ and } 2_3) = (1/36)^3 = 1/46656$

4-26 a) 6/32 b) 6/32 c) 1/32

4-27 a) P(John survives first draw) = 3/4
P(John survives second draw | John survives first draw) = 2/3
P(John survives first two draws) = (3/4)(2/3) = 1/2

b) P(Paul wins game) = (3/4)(2/3)(1/2) = 1/4

4-28 a) P(A passes | B fails) = P(A passes) = .02

b) P(B passes | A passes) = P(B passes) = .07

c) P(A and B pass) = P(A passes)P(B passes) = (.02)(.07) = .0014

4-29 a) P(2 and 3 broken | 1 broken) $= \text{P(2 and 3 broken)} = \text{P(2 broken)P(3 broken)}$
$= (.04)(.04) = .0016$

b) P(1, 2, 3, and 4 broken) $= \text{P(1 broken)P(2 broken)P(3 broken)P(4 broken)}$
$= (.04)^4 = .00000256 > .0000002 = 1/5{,}000{,}000$

Thus, the statement is false.

4-30 a) P(GL, DH, and DC approve) $= (.85)(.80)(.82) = .5576$

b) P(GL and DH approve, DC doesn't approve) $= (.85)(.80)(.18) = .1224$

4-31 a) 6/30 b) 12/30 c) 6/30

4-32 a) P(1, 2, 3, 4) $= (.75)(.82)(.87)(.9) = .481545$

b) P(1, not 2, not 3, 4) $= (.75)(.18)(.13)(.9) = .015795$

c) P(one noticed) $=$ P(1, not 2, not 3, not 4) + P(not 1, 2, not 3, not 4)
+ P(not 1, not 2, 3, not 4) + P(not 1, not 2, not 3, 4)
$= (.75)(.18)(.13)(.1) + (.25)(.82)(.13)(.1)$
$+ (.25)(.18)(.87)(.1) + (.25)(.18)(.13)(.9)$
$= .01360$

d) P(not 1, not 2, not 3, not 4) $= (.25)(.18)(.13)(.1) = .000585$

e) P(not 3, not 4) $= (.13)(.1) = .013$

4-33 a) P(neither A nor B) $= 1 - \text{P(A or B)} = 1 - .47 = .53$

b) P(A and B) $= \text{P(A)} + \text{P(B)} - \text{P(A or B)} = .39 + .21 - .47 = .13$

c) $\text{P(B | A)} = \dfrac{\text{P(A and B)}}{\text{P(B)}} = \dfrac{.13}{.39} = \dfrac{1}{3}$

d) $\text{P(A | B)} = \dfrac{\text{P(A and B)}}{\text{P(B)}} = \dfrac{.13}{.21} = \dfrac{13}{21}$

4-34 $\text{P(A | C)} = \dfrac{\text{P(A and C)}}{\text{P(C)}} = \dfrac{1/7}{1/3} = \dfrac{3}{7}$

$\text{P(C | A)} = \dfrac{\text{P(A and C)}}{\text{P(A)}} = \dfrac{1/7}{3/14} = \dfrac{2}{3}$

$\text{P(B and C)} = \text{P(B | C)P(C)} = (5/21)(1/3) = 5/63$

$\text{P(C | B)} = \dfrac{\text{P(B and C)}}{\text{P(B)}} = \dfrac{5/63}{1/6} = \dfrac{10}{21}$

4-35 The assignment is not consistent. Computing P(A and B), we find that
$\text{P(A and B)} = \text{P(A | B)P(B)} = \text{P(A)P(B)} = (.65)(.80) = .52$
and $\text{P(A and B)} = \text{P(B | A)P(A)} = (.85)(.65) = .5525 \neq .52$

4-36 $\text{P(alcoholic | male)} = \dfrac{\text{P(male and alcoholic)}}{\text{P(male)}} = \dfrac{.21}{.59} = .356$

4-37 a) $\text{P(A | N)} = \dfrac{\text{P(A and N)}}{\text{P(N)}} = \dfrac{.37}{.6} = .617$

b) $\text{P(N | A)} = \dfrac{\text{P(A and N)}}{\text{P(A)}} = \dfrac{.37}{.52} = .712$

4-38 H = hurricane in eastern Gulf, F = hurricane hits Florida

a) P(H and F) = P(F | H)P(H) = (.76)(.85) = .646

b) P(F | H) = (3/4)(.76) = .57
P(H and F) = P(F | H)P(H) = (.57)(.85) = .4845
[Note: .4845 = (3/4)(.646)]

4-39 R = WTR completes research on time, I = investigation occurs, and
C = Litre awarded contract

a) P(R) = .80 Let a = P(I)

$$P(C) = P(C\mid R, \text{not } I)P(R, \text{not } I) + P(C\mid \text{not } R, I)P(\text{not } R, I) + P(C\mid R, I)P(R, I) + P(C \mid \text{not } R, \text{not } I)P(\text{not } R, \text{not } I)$$

$$= (.67)(.80)(1-a) + (.72)(.20)(a) + (.58)(.80)(a) + (.85)(.20)(1-a)$$

$$= .706 - .098a$$

P(C) $\geq$ .65 means .706 − .098a $\geq$ .65
or .098a $\leq$ .056
or a $\leq$ .571

b) P(I) = .70 Let b = P(R)
As in part (a):

$$P(C) = (.67)(b)(.30) + (.72)(1-b)(.70) + (.58)(b)(.70) + (.85)(1-b)(.30)$$

$$= .759 - .152b$$

P(C) $\geq$ .65 means .759 − .152b $\geq$ .65
or .152b $\leq$ .109
or b $\leq$ .717

c) P(I) = .75 P(R)= .85
P(C) = (.67)(.85)(.25) + (.72)(.15)(.75) + (.58)(.85)(.75) + (.85)(.15)(.25) = .625

4-40 P(upgrade and favorable evaluation)
= P(favorable evaluation) P(upgrade | favorable evaluation) = .65(.85) = .5525

4-41 a) P(graduate) = (24 + 61 + 20)/350 = 105/350

b) $$P(\text{periodicals} \mid \text{graduate}) = \frac{P(\text{periodicals and graduate})}{P(\text{graduate})} = \frac{(61/350)}{(105/350)} = 61/105$$

c) $$P(\text{faculty} \mid \text{reference}) = \frac{P(\text{faculty and reference})}{P(\text{reference})} = \frac{(16/350)}{(84/350)} = 16/84$$

d) P(undergraduate and books) = 72/350

4-42 A = pilots strike D = drivers strike

a) P(A and D) = P(A | D)P(D) = (.90)(.65) = .585

b) P(D | A) = P(A and D)/P(A) = .585/.75 = .78

4-43

Event	P(Event)	P(X | Event)	P(X and Event)	P(Event | X)
A	.20	.75	.15	.15/.60 = .25
B	.65	.60	.39	.39/.60 = .65
C	.15	.40	.06	.06/.60 = .10
			P(X) = .60	

Thus, P(A | X) = .25, P(B | X) = .65, and P(C | X) = .10.
In addition, P(Y) = 1 — P(X) = 1 — .60 = .40.

4-44 R = received payment

Event	P(Event)	P(R | Event)	P(R and Event)	P(Event | R)
Personal Call	.7	.75	.525	.525/.71 = .739 (a)
Phone Call	.2	.60	.120	.120/.71 = .169 (b)
Letter	.1	.65	.065	.065/.71 = .092 (c)
			P(R) = .710	

4-45 F = favorable ruling
B = choose Baltimore
A = choose Atlanta
C = choose Cleveland

Event	P(Event)	P(F | Event)	P(F and Event)	P(Event | F)
A	.40	.45	.1800	.1800/.4775 = .3770
B	.35	.60	.2100	.2100/.4775 = .4398
C	.25	.35	.0875	.0875/.4775 = .1832
			P(F) = .4775	

Since P(B | F) is the largest of the three, they most likely chose Baltimore.

4-46

Event	P(Event)	P(Storm | Event)	P(Storm & Event)	P(Event | Storm)
Dry	0.20	0.30	0.06	0.06/0.61 = 0.0984
Moist	0.45	0 .60	0.27	0.27/0.61 = 0.4426
Wet	0.35	0 .80	0.28	0.28/0.61 = 0.4590
			P(Storm) = 0.61	

The probability of a thunderstorm is 0.61. The probability of moist conditions, given that the picnic was cancelled (i.e., given that there was a thunderstorm) is 0.4426.

4-47 R = radiation leak occurs
M = mechanical failure occurs
F = fire occurs
H = human error occurs

Event	P(Event)	P(R | Event)	P(R and Event)	P(Event | R)
F	$\frac{.0010}{.20} = .005$	.20	.0010	$\frac{.0010}{.0037} = .2703$
M	$\frac{.0015}{.50} = .003$	.50	.0015	$\frac{.0015}{.0037} = .4054$
H	$\frac{.0012}{.10} = .012$	.10	.0012	$\frac{.0012}{.0037} = .3243$
			P(R) = .0037	

a) P(F) = .005 P(M) = .003 P(H) = .012

b) P(F | R) = .2703 P(M | R) = .4054 P(H | R) = .3243

c) P(R) = .0037

4-48 G = play on grass T = play on turf I = incur knee injury

$P(I \mid T) = (1.5)P(I \mid G) = .42 \Rightarrow P(I \mid G) = .42/1.5 = .28$

a) $P(I) = P(I \text{ and } T) + P(I \text{ and } G) = (.42)(.4) + (.28)(.6) = .336$

b) $P(G \mid I) = P(I \text{ and } G)/P(I) = (.28)(.6)/.336 = .5$

4-49 a) P(offensive line | foot injury) = 32/90 = 16/45

b) P(defensive line | foot injury) = 38/90 = 19/45

c) P(offensive backfield | foot injury) = 11/90

d) P(defensive backfield | foot injury) = 9/90 = 1/10

4-50 L_{10}/G_{10} = lose/gain more than 10 seats

L_6/G_6 = lose/gain 6 - 10 seats

S =lose or gain 5 or fewer sets

R_2/F_2 = unemployment rises/falls by 2 percent or more

U = unemployment changes by less than 2 percent

a)

Event	P(Event)	$P(G_6 \mid \text{Event})$	$P(G_6 \text{ and Event})$	$P(\text{Event} \mid G_6)$
R_2	.25	.15	.0375	.0375/.3150
U	.45	.35	.1575	.1575/.3150
F_2	.30	.40	.1200	.1200/.2150 = .3810
			$P(G_6)$ = .3150	

b)

Event	P(Event)	P(S | Event)	P(S and Event)	P(Event | S)
R_2	.25	.15	.0375	.0375/.1350
U	.45	.15	.0675	.0675/.1350 = .5000
F_2	.30	.10	.0300	.0300/.1350
			P(S) = .1350	

4-51 F = flop M = moderate H = hit

RL = rating of ≤ 3 RM = rating $\geq$ 3 and $\leq$ 7 RH = rating of $\geq$ 7

a) Prior probabilities on film performance before first screening:

P(H) = .60 P(M) = .25 P(F) = .15

Event	P(Event)	P(RM | Event)	P(RM and Event)	P(Event | RM)
H	.60	.30	.1800	.1800/.345 = .522
M	.25	.45	.1125	.1125/.345 = .326
F	.15	.35	.0525	.0525/.345 = .152
			P(RM) = .3450	

b) Prior probabilities on film performance before second screening:

P(H) = .522 P(M) = .326 P(F) = .152

Event	P(Event)	P(RL | Event)	P(RL and Event)	P(Event | RL)
H	.522	.10	.0522	.0522/.2097 = .2489
M	.326	.25	.0815	.0815/.2097
F	.152	.50	.0760	.0760/.2097
			P(RL) = .2097	

4-52 The difference in rates would seem to suggest that the risk or probability of dying is greater as one gets older (common sense tells us that). Thus, this type of protection costs more because the company will have a greater probability of having to pay the claim before it has had the chance to collect very much in premiums. On the other hand, the higher rates for young drivers suggests that young drivers have a greater probability of having an accident, making it necessary for them to pay more for this protection.

4-53 c

4-54 Using past data on the rate of restaurant failures in an area or in the nation, econometricians calculate the relative frequency of restaurant failures and use this as an estimate of the chances of any given restaurant failing in this time period.

4-55 It is obvious that a gambling house relies on the fact that the majority of the time the odds, or probabilities, are in its favor. The chances of a player continually getting very good hands in a poker game, or a long string of successful rolls at a dice table are slim. Hence, even if one player wins for a while, his chances of losing in the long run are greater than his chances of winning. In addition, for every player whose luck goes against the odds, allowing him to win, there will be many who lose.

4-56 a) 1/5 b) 2/5 c) 3/5 d) classical

4-57 G = landing gear failure C = crash on landing

a) $P(G) = .12$ $P(\text{not } G) = .88$

Event	P(Event)	P(C | Event)	P(C and Event)	P(Event | C)
G	.12	.55	.0660	.0660/.1188 = .5556
not G	.88	.06	.0528	.0528/.1188
			P(C) = .1188	

b) $P(G) = .03$ $P(\text{not } G) = .97$

Event	P(Event)	P(C | Event)	P(C and Event)	P(Event | C)
G	.03	.55	.0165	.0165/.0747 = .2209
not G	.97	.06	.0582	.0582/.0747
			P(C) = .0747	

4-58 a) No. The item "he is nominated for vice-president" cannot occur during midterm elections.

b) No. He can, for example, win his party's re-election nomination and be re-elected.
Yes. Winning and losing his party's re-election nomination, for example are mutually exclusive.

c) No. Another possible event is "he is not re-elected."

4-59 a) $P(\text{ROE} > 16 \mid E/A < 7) = \dfrac{P(\text{ROE}>16 \text{ and } E/A<7)}{P(E/A<7)} = 5/13 = 0.4167$

b) $P(14 \le \text{ROE} \le 16 \mid E/A>7) = \dfrac{P(14 \le \text{ROE} \le 16 \text{ and } E/A>7)}{P(E/A>7)} = 5/13 = 0.3846$

c) $P(\text{NI}>50\text{m} \mid \text{TA}>2\text{b}) = \dfrac{P(\text{NI}>50\text{m and TA}>2\text{b})}{P(\text{TA}>2\text{b})} = 1/2 = 0.5$

d) $P(\text{ROE} > 15) = \frac{15}{25} = 0.6$

e) $P(\text{ROE} > 15 \text{ AND TA } \ge 2) = P(\text{ROE} > 15) \times P(\text{TA}>2)$

$$= \frac{15}{25}\left(\frac{2}{25}\right) = 0.6(0.8) = 0.048$$

f) $P(\text{ROE} > 20 \mid \text{TA} \ge 1\text{b}) = P(\text{ROE}>20) = \frac{5}{25} = 0.2$

4-60 a) No. For example, the firm might have won 5 other bigger contracts.

b) No. The cousin need not be the uncle's child.

c) Yes.

d) No. The promotion may come before the discovery of embezzlement.

4-61 N/n = major/minor crime in northern neighborhood
S/s = major/minor crime in southern neighborhood

a) $P(N \text{ or } n) = P(N) + P(n) - P(N \text{ and } n) = P(N) + P(n) - P(N)P(n)$
$= .478 + .602 - (.478)(.602) = .792$
Thus, $P(\text{not } N \text{ and not } n) = 1 - .792 = .208$

b) $P(S \text{ or } s) = P(S) + P(s) - P(S)P(s)$
$= .350 + .523 - (.350)(.523) = .690$

c) Let N = a crime is committed in the northern neighborhood
S = a crime is committed in the southern neighborhood
Then from a) and b), we know that: $P(N) = .792$ and $P(S) = .690$
$P(\text{no crime in either neighborhood}) = 1 - P(N \text{ or } S)$
$P(N \text{ or } S) = P(N) + P(S) - P(N \text{ and } S) = P(N) + P(S) - P(N)P(S)$
$= .792 + .690 - (.792)(.690) = .936$
Thus, $P(\text{no crime in either neighborhood}) = 1 - .936 = .064$

4-62 a) $P(\text{plant will be a polluter}) = \dfrac{\text{observed \# of polluting plants}}{\text{total \# of observations}}$
$= \frac{2}{6} = .3333$

b) This probability was determined using the relative frequency of occurrence method.

c) The calculated probability is an inaccurate estimate of the true probability. The number of observations is extremely small, and the observations were made for plants which are not exactly like the proposed plant. Therefore, the probability estimate actually takes on many of the characteristics of a subjective estimate.

4-63 a) The percentages given were determined from analysis of the frequency of past occurrences.

b)

Fraction of questionnaires mailed		1.000
Less: mistake in address	.0130	
lost or damaged	.0280	
unforwarded (.19 × .52 =)	.0988	
		.1398
Fraction of questionnaires reaching respondents		.8602
Times: responses from questionnaires received		× .15
Fraction of replies from original mailing		.1290

4-64 E = Engulf and Devour take over
R = R. A. Venns takes over

a) $P(E) = 7/20 = .35$ $\quad P(R) = 6/15 = .4$ $\quad P(E \text{ and } R) = 0$
$P(E \text{ or } R) = P(E) + P(R) - P(E \text{ and } R) = .35 + .4 - 0 = .75$

b) There is not sufficient information to answer the question.

4-65 X = entering patient needs X-ray
I = entering patient has sufficient insurance

The necessary additional assumption is that X and I are independent, in which case:

$P(X \text{ and } I) = P(X)P(I) = (.23)(.72) = .1656$

4-66 $A0/B0$ = Flight 100/200 on time
$A5/B5$ = Flight 100/200 5 minutes late/early
$A10/B10$ = Flight 100/200 10 minutes late/early

a) $P(\text{collision}) = P(A5 \text{ and } B5) + P(A0 \text{ and } B10) + P(A10 \text{ and } B0)$
$= (.03)(.02) + (.95)(.01) + (.02)(.97)$
$= .0295 > .025$; diversion required

b) $P(\text{collision} \mid A5) = P(B5 \mid A5) = P(B5) = .02 < .025$; no diversion required

c) $P(\text{collision} \mid B5) = P(A5 \mid B5) = P(A5) = .03 > .025$; diversion required

4-67 NSF = insufficient funds CB = cash back

a) $P(NSF) = 0.12$, or 120 per 1000

b) $P(CB) = 0.1$, or 100 per 1000

c) $P(NSF \text{ and } CB) = P(NSF)\,P(CB \mid NSF) = 0.12(0.5) = 0.06$, or 60 per 1000

d) $P(NSF \text{ or } CB) = P(NSF) + P(CB) - P(NSF \text{ and } CB)$
$= 0.12 + 0.10 - 0.06 = 0.16$, or 160 per 1000

4-68 a) $P(\text{Family}) = \frac{120}{300} = 0.4$

b) $P(\text{Spouse or other relative}) = \frac{6 + 15}{300} = \frac{21}{300} = 0.07$

4-69 a) possibly dependent (they may, e.g., have some identical components)

b) independent

c) possibly independent

d) dependent

e) dependent (Christian Scientists, e.g., do not permit organ donation)

4-70 B = passenger bumped
K = passenger takes Kansas City flight
A = passenger takes Atlanta flight
D = passenger takes Detroit flight

Event	P(Event)	P(B | Event)	P(B and Event)	P(Event | B)
A	.55	.07	.0385	.0385/.0670 = .5746 (a)
K	.20	.08	.0160	.0160/.0670 = .2388 (b)
D	.25	.05	.0125	.0125/.0670 = .1866 (c)
			P(S) = .0670	

4-71 a-d) First, we must find the probabilities of consumer and government sales increases for years 1, 2, 3, and 4. We do this by subtracting the cumulative probability for the preceding year from the cumulative probability for the year of interest.

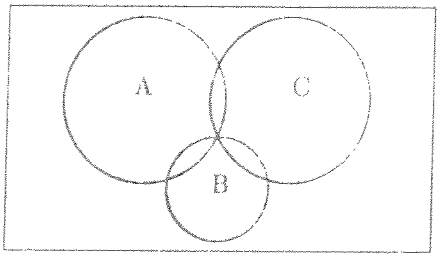
Year	Probability of Increase in Consumer Sales	Probability of Increase in Government Sales
1	.05	.08
2	.03	.07
3	.04	.10
4	.04	.07

Since consumer and government sales increases are assumed to be mutually exclusive, we need only add the two probabilities for each year to get the probability of plant expansion.

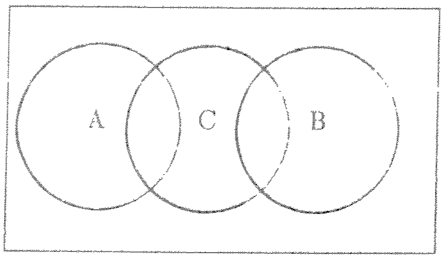
Year 1: $P(\text{expand}) = .05 + .08 = .13$
2: $P(\text{expand}) = .03 + .07 = .10$
3: $P(\text{expand}) = .04 + .10 = .14$
4: $P(\text{expand}) = .04 + .07 = .11$

e) P(expand) = P(expand in year 1) + P(expand in year 2) + P(expand in year 3) + P(expand in year 4) = .13 + .10 + .14 + .11 = .48

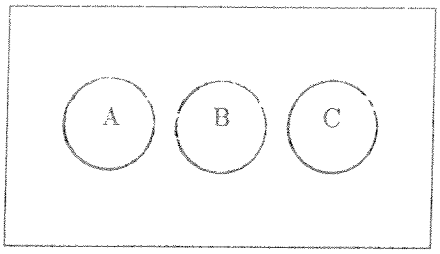
4-72 a)



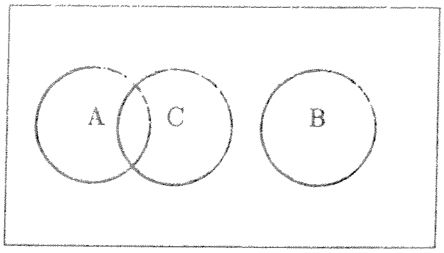
b)



c)



d)



4-73 L = lose item T = item sent by truck R = item sent by rail

If P(L | T) = .035 and P(L | R) = .02, then

Event	P(Event)	P(L | Event)	P(L and Event)	P(Event | L)
T	.40	.035	.014	.014/.026
R	.60	.020	.012	.012/.026
			P(L) = .026	

The lost comics were most likely carried by truck.

If P(L | T) = P(L | R) = .02, then

Event	P(Event)	P(L | Event)	P(L and Event)	P(Event | L)
T	.40	.02	.008	.008/.02
R	.60	.02	.012	.012/.02
			P(L) = .020	

Now, the lost comics were most likely carried by rail.

4-74 a) E1 = engine 1 fails E2 = engine 2 fails

P(E1 and E2) = P(E1)P(E2 | E1) = .05(.10) = .005

b) B = recalled for brakes S = steering flaw

Since the two events are statistically independent,

P(B and S) = P(B)P(S) = (.15)(.02) = .003

c) F = files on return C = cheats on return

P(C and F) = P(C | F)P(F) = (.25)(.70) = .175

4-75 O = out-of-town client L = local client S = house was sold

Event	P(Event)	P(S | Event)	P(S and Event)	P(Event | S)
O	.4	.075	.0300	.0300/.0618 = .4854
L	.6	.053	.0318	.0318/.0618 = .5146
			P(S) = .0618	

Thus, the house was more likely to have been shown to a local client.

4-76 a) P(neutral | Chapel Hill) = 3/15 = .2
P(strongly opposed | Chapel Hill) = 2/15 = .1333

b) P(strongly supports) = (6 + 3 + 1)/45 = 10/45 = .222

c) P(neutral or slightly opposed | Raleigh or Lumberton) = (4 + 3 + 3 + 5)/30
= 15/30 = .5

4-77 a) $P(\text{old Republican}) = \frac{166}{435} = 0.3816$

b) P(new Not Republican) = 1 - P(new Republican)

$$= 1 - \frac{175}{435} = 1 - 0.4023$$

$$= 0.5977$$

c) No. At least 9 incumbent Democrats were not re-elected. However, for example, it is possible that 12 incumbent Democrats were not re-elected, but 3 new Democrats were elected, giving the Democrats a net loss of 9 seats.

4-78 a) P(damaged) = 565/10000 = .0565
P(overripe) = 1135/10000 = .1135

b) P(Ecuador or Honduras) = 1

c) P(Honduras | overripe) = 295/1135 = .2599

d) P(damaged or overripe) = P(damaged) + P(overripe) − P(damaged and overripe)
If damaged and overripe are mutually exclusive, then
P(damaged and overripe) = 0, so
P(damaged or overripe) = .0565 + .1135 = .1700
However, if they are not exclusive, then we can't find P(damaged or overripe) since we don't know P(damaged and overripe).

4-79 a) P(no offer in 3 interviews) = $(1 - .07)^3 = (.93)^3 = .8044$

b) P(at least one offer in 9 interviews) = 1 − P(no offers in 9 interviews)
$= 1 - (.93)^9 = .4796$

c) Y = offer, N = no offer
P(NNYNY) = .93(.93)(.07)(.93)(.07) = .0039

4-80 a) P(8-ball first) = 1/15 = .0667

b) P(8-ball first, second or third) = 3/15 = .2000

c) P(8-ball last) = 1/15 = .0667

4-81 a) P(pump A fails) = .03 + .08 = .11
P(pump B fails) = .02 + .11 = .13
They should use pump A.

b) Now they should use pump B, because only failures caused by seized bearings will be of concern, and P(seized bearings | B) is less than P(seized bearings | A).

4-82 a) P(death | vaccine) = .02(.03) + .98(.0005) = .0011
P(death | no vaccine) = .3(.04) = .0120

b) P(death) = P(death | vaccine) P(vaccine) + P(death | no vaccine) P(no vaccine)
= .0011(.25) + .0120(.75) = .0093

4-83 a) Let $x = P(\text{no tear} \mid \text{normal})$. Then

$$P(\text{tear}) = P(\text{tear} \mid \text{normal})P(\text{normal}) + P(\text{tear} \mid \text{fast})P(\text{fast})$$
$$.112 = (1-x)(.4) + 2(1-x)(.6) = 1.6 - 1.6x$$
$$1.6x = 1.6 - .112 = 1.488$$
$$x = 1.488/1.6 = 0.93$$

b)

Event	P(Event)	P(Tear | Event)	P(Tear & Event)	P(Event | Tear)
Normal	.4	.1	.04	.04/.16 = .25
Fast	.6	.2	.12	.12/.16 = .75
		P(Tear) =	.16	

Thus, P(Normal | Tear) = .25

4-84

Event	P(Event)	P(4 tears | Event)	P(4 tears & Event)
Normal	.4	$(.07)^4 = .00002401$	.000009604
Fast	.6	$(.14)^4 = .00038416$	.000230496
		P(4 tears) =	.000240100

Thus, P(Fast | 4 tears) = .000230496/.00024010 = .96

4-85 Assuming that profitability is independent from airline to airline, the probability of all five privately owned lines having profits while all five state-owned lines have losses is $(0.5)^{10} = 0.0001$.

4-86 a) $6/50,000 + 6/50,000 - (6/50,000)^2 = 0.00024$, or about once in 4167 six-hour flights.

b) $6/50,000$

c) $(6/50,000)^2$

CHAPTER 5

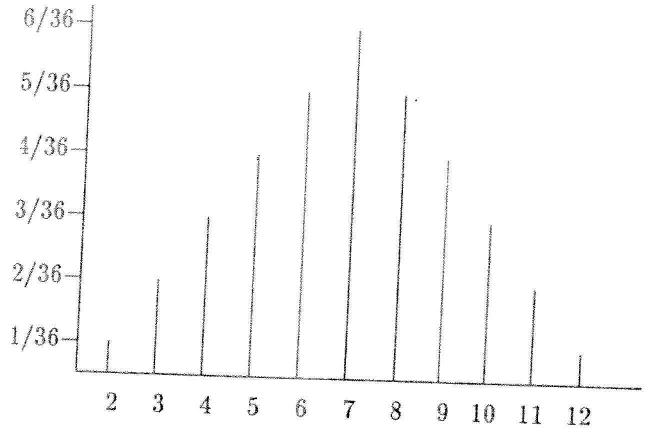
PROBABILITY II: DISTRIBUTIONS

5-1

Value	2	4	5	7	8
Probability	.30	.20	.05	.35	.10

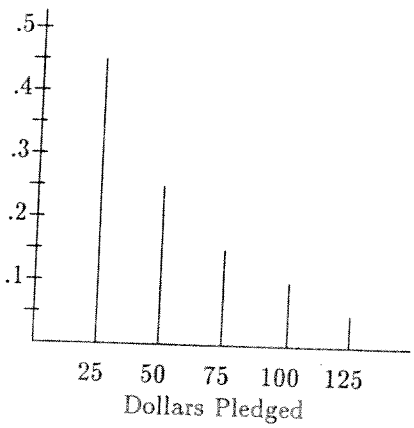
5-2

Total	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



5-3 (a), (b), (e)

5-4



5-5

Combination #	1	2	3	4
Probability	.4	.2	.2	.2

a) P(V-8) = P(#1) + P(#3) + P(#4) = .4 + .2 + .2 = .8

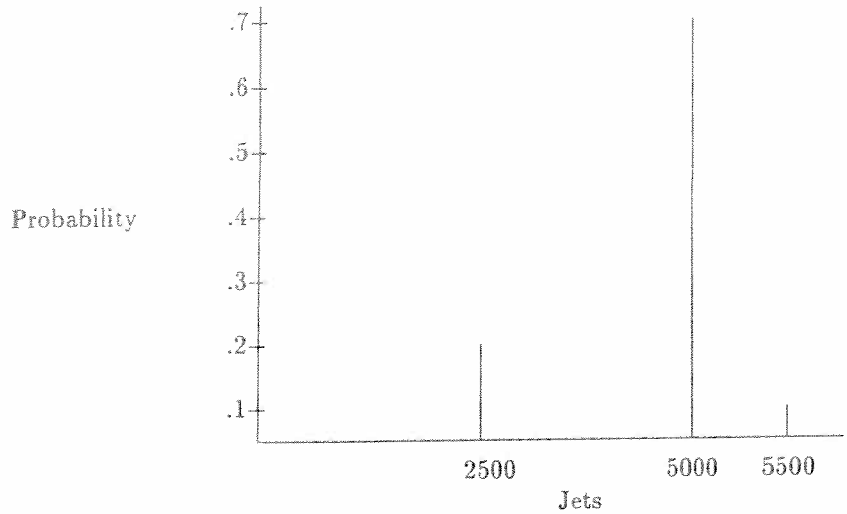
b)

Number of V-8 Engines	Events*	Probability
0	(N,N)	(.2)(.2) = .04
1	(N,Y),(Y,N)	(.2)(.8) + (.8)(.2) = .32
2	(Y,Y)	(.8)(.8) = .64

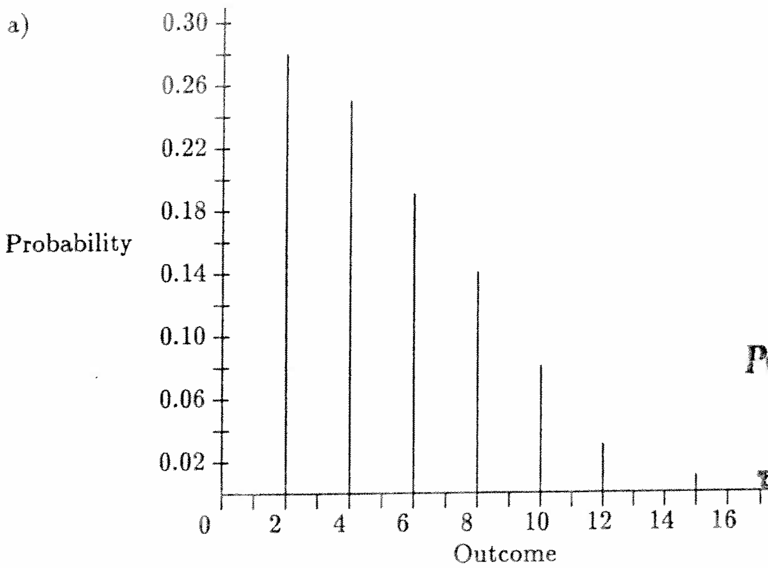
* (Y,N) indicates that customer 1 orders a V-8 engine but 2 does not, and similarly for the other events.

5-6

# of Jets Sold	Probability
2500	.2
5000	.7
5500	.1



5-7 a)



b)

Outcome (1)	Frequency (2)	P(Outcome) (3)	(1) × (3)
2	24	24/85 = 0.2824	0.5648
4	22	22/85 = 0.2588	1.0352
6	16	16/85 = 0.1882	1.1292
8	12	12/85 = 0.1412	1.1296
10	7	7/85 = 0.0824	0.8240
12	3	3/85 = 0.0353	0.4236
15	1	1/85 = 0.0118	0.1770
	85	85/85 = 1.0000	5.2834 = expected outcome

5-8 a)

Outcome (1)	P(Outcome) (2)
\$ 8,000	.05
9,000	.15
10,000	.25
11,000	.30
12,000	.20
13,000	.05
	1.00

b)

(1) × (2)
400
1350
2500
3300
2400
650
\$10,600 = expected value

5-9 a)

Outcome (1)	Frequency (2)	P(Outcome) (3)	(1) × (3)
0	25	.05	0.00
15	125	.25	3.75
30	75	.15	4.50
45	175	.35	15.75
60	75	.15	9.00
75	25	.05	3.75
	500	1.00	36.75

b) Expected value of an outcome = 36.75.

5-10 The expected cost is

$$0(0.35) + 50(0.25) + 100(0.15) + 150(0.10) + 200(0.08) + 250(0.05) + 300(0.02) = \$77$$

so Jim should not pay \$100 for the warranty.

5-11

# of Accidents (1)	Frequency (2)	Probability (3)	(1) × (3)
0	1	.0417	.0000
2	1	.0417	.0834
3	1	.0417	.1251
4	2	.0833	.3332
6	1	.0417	.2502
7	3	.1250	.8750
8	5	.2083	1.6664
9	2	.0833	.7497
10	5	.2083	2.0830
12	2	.0833	.9996
14	1	.0417	.5838
	24	1.0000	7.7494

The expected number of accidents per month is 7.7494. Since this is greater than 7, Mr. Opsine should recommend the installment of a traffic light.

5-12

Months to Settle (1)	Probability (2)	(1) × (2)	Months to Settle (1)	Probability (2)	(1) × (2)
1	.02	.01	11	.05	.55
2	.02	.04	12	.06	.72
3	.01	.03	13	.07	.91
4	.02	.08	14	.08	1.12
5	.02	.10	15	.10	1.50
6	.03	.18	16	.09	1.44
7	.04	.28	17	.08	1.36

8	.03	.24	18	.08	1.44
9	.04	.36	19	.07	1.33
10	.04	.40	20	.06	1.20
				1.00	13.29

The expected number of months to settle = 13.29.

5-13 a)

Number of Fires (1)	Frequency (2)	Probability (3)	(1) × (3)
0	1	.0417	.0000
1	1	.0417	.0417
2	3	.1250	.2500
4	2	.0833	.3332
5	3	.1250	.6250
8	3	.1250	1.0000
10	5	.2083	2.0830
15	2	.0833	1.2495
20	1	.0417	.8340
25	2	.0833	2.0825
30	1	.0417	1.2510
	24	1.0000	9.7499

The expected number of fires per month = 9.7499.

b)

Number of Fires (1)	Frequency (2)	Probability (3)	(1) × (3)
10	1	.1667	1.667
15	1	.1667	2.500
20	1	.1667	3.333
25	2	.3333	8.333
30	1	.1667	5.000
	6	1.0000	20.833

The expected number of fires per winter month = 20.8333.

5-14 Truck Division:

# of Letters Lost (1)	Frequency (2)	Probability (3)	(1) × (3)
0	1	.0833	.0000
1	2	.1667	.1667
2	2	.1667	.3333
3	2	.1667	.5000
4	2	.1667	.6667
5	2	.1667	.8333
6	0	.0000	.0000
7	1	.0833	.5833
	12	1.0000	3.0833

Air Division:

# of Letters Lost (1)	Frequency (2)	Probability (3)	(1) × (3)
0	2	.1667	.0000
1	1	.0833	.0833
2	2	.1667	.3333
3	1	.0833	.2500

4	3	.2500	1.0000
5	1	.0833	.4167
6	1	.0833	.5000
7	1	.0833	.5833
	12	1.0000	3.1667

He investigates the air division.

5-15

Loss Table

States of Nature

	25000	40000	55000	70000	
Probability	.10	.30	.45	.15	
Action					Expected loss
25000	0	15000	30000	45000	24750.0
40000	3750	0	15000	30000	11625.0
55000	7500	3750	0	15000	4125.0 ←
70000	11250	7500	3750	0	5062.5

Ordering 55,000 programs will minimize expected losses.

5-16 Cost = 6300/2 = \$3150 per car

Revenue = 35 − 2.50 = \$32.50 per car per day

Revenue = 32.50 × 312 = \$10,140 per year

10140 − 3150 = \$6990 = opportunity cost per car

Loss Table

States of Nature

	13	14	15	16	17	18	
Probability	.08	.15	.22	.25	.21	.09	
Action							Expected Loss
13	0	6690	13980	20970	27960	34950	18383.70
14	3150	0	6990	13980	20970	27960	12204.90
15	6300	3150	0	6990	13980	20970	7547.10
16	9450	6300	3150	0	6990	13980	5120.10 ←
17	12600	9450	6300	3150	0	6990	5228.10
18	15750	12600	9450	6300	3150	0	7465.50

The optimal number of cars is 16.

5-17 a) If all three seats are released, the expected revenue is
$250(3) = \$750$

b) If all three seats are released, the expected net revenue is
$750 - 0.45(0) - 0.30(150) - 0.15(300) - 0.10(450) = \$615$

c) If two seats are released, the expected net revenue is
$500 + 0.45(0) + 0.30(475) + 0.15(475 - 150) + 0.10(475 - 300) = \$708.75$

d) If one seat is released, the expected net revenue is
$250 + 0.45(0) + 0.30(475) + 0.15(950) + 0.10(950 - 150) = \$615$

If no seats are released, the expected net revenue is
$0 + 0.45(0) + 0.30(475) + 0.15(950) + 0.10(1425) = \$427.50$

Therefore, releasing two seats will maximize expected net revenue.

5-18 binomial ($n = 7$, $p = .2$)

a) $P(r = 5) = \left(\frac{7!}{5!2!}\right)(.2)^5(.8)^2 = .0043$

b) $P(r > 2) = 1 - P(r \leq 2) = 1 - [P(r=0) + P(r=1) + P(r=2)]$

$$= 1 - \left(\frac{7!}{0!7!}\right)(.2)^0(.8)^7 - \left(\frac{7!}{1!6!}\right)(.2)^1(.8)^6 - \left(\frac{7!}{2!5!}\right)(.2)^2(.8)^5$$

$$= 1 - .2097 - .3670 - .2753 = .1480$$

c) $P(r < 8) = 1$

d) $P(r \geq 4) = P(r > 2) - P(r=3) = .1480 - \left(\frac{7!}{3!4!}\right)(.2)^3(.8)^4$

$$= .1480 - .1147 = .0333$$

5-19 binomial ($n = 15$, $p = 0.2$)

a) $P(r=6) = 0.0430$

b) $P(r \geq 11) = 0.0000$

c) $P(r \leq 4) = 0.0352 + 0.1319 + 0.2309 + 0.2501 + 0.1876 = 0.8357$

5-20

	n	p	$\mu = np$	$\sigma = \sqrt{npq}$
a)	15	.20	3.00	1.549
b)	8	.42	3.36	1.396
c)	72	.26	4.32	2.015
d)	29	.49	14.21	2.692
e)	642	.21	134.82	10.320

5-21 binomial, $n = 8$

$P(r \geq 1) = 1 - P(r=0)$

	p	$P(r=0)$	$P(r \geq 1)$
a)	.1	$\left(\frac{8!}{0!8!}\right)(.1)^0(.9)^8 = .4305$	.5695
b)	.3	$\left(\frac{8!}{0!8!}\right)(.3)^0(.7)^8 = .0576$	.9424
c)	.6	$\left(\frac{8!}{0!8!}\right)(.6)^0(.4)^8 = .0007$	.9993
d)	.4	$\left(\frac{8!}{0!8!}\right)(.4)^0(.6)^8 = .0168$	.9832

5-22 a) P(more than 2 flaws) = 1 − P(0 flaws) − P(1 flaw) − P(2 flaws)

$$= 1 - \left(\frac{10!}{0!10!}\right)(.02)^0(.98)^{10} - \left(\frac{10!}{1!9!}\right)(.02)^1(.98)^9 - \left(\frac{10!}{2!8!}\right)(.02)^2(.98)^8$$

$$= 1 - .8171 - .1667 - .0153 = .0009$$

b) P(0 flaws) $= \left(\frac{10!}{0!10!}\right)(.02)^0(.98)^{10} = .8171$

5-23 a) P(more than 8 have jobs) $= P(r=9) + P(r=10)$

$$= \left(\frac{10!}{9!1!}\right)(.40)^9(.60)^1 + \left(\frac{10!}{10!0!}\right)(.40)^{10}(.60)^0$$

$$= .0016 + .0001 = .0017$$

b) P(3 have jobs) $= \left(\frac{10!}{3!7!}\right)(.40)^3(.60)^7 = .2150$

5-24 binomial, $n = 15$, p to be determined

a) claim 1: $P(r \geq 4) = .9095 \Rightarrow p = .40$
claim 2: $np > 7 \Rightarrow p > 7/15 = .467$
The claims are not consistent.

b) No. The assistants claim that $p = .40$

5-25 a) $n = 20$, $p = 0.45$ $\quad P(r = 8) = \left(\frac{20!}{8!12!}\right)(0.45)^8(.55)^{12} = 0.1623$

b) $n = 20$, $p = 0.1$ $\quad P(r = 1) = \left(\frac{20!}{1!19!}\right)(0.1)^1(0.9)^{19} = 0.2702$

c) $n = 20$, $p = 0.2$ $\quad P(r \leq 2) = \left(\frac{20!}{0!20!}\right)(.2)^0(.8)^{20} + \left(\frac{20!}{1!19!}\right)(.2)^1(.8)^{19} + \left(\frac{20!}{2!18!}\right)(.2)^2(.8)^{18}$

$= 0.0115 + 0.0576 + 0.1369 = 0.2060$

d) $n = 20$, $p = 0.1$ $\quad P(r \geq 2) = 1 - P(r = 0) - P(r = 1)$

$= 1 - \left(\frac{20!}{0!20!}\right)(0.1)^0(0.9)^{20} - 0.2702$

$= 1 - 0.1216 - 0.2702 = 0.6082$

5-26 Let r be a binomial random variable, with $n = 15$ and $p = 0.3$, representing the number of browsing customers who buy something.

a) $P(r \geq 1) = .9953$

b) $P(r \geq 4) = .7031$

c) $P(r = 0) = 1 - P(r \geq 1) = 1 - .9953 = .0047$

d) $P(r \leq 4) = 1 - P(r \geq 5) = 1 - .4845 = .5155$

5-27 binomial, $n = 28$, $p = .025$; $\lambda = np = .7$; $e^{-.7} = .49659$

a) $P(r \geq 3) = 1 - P(r \leq 2) = 1 - P(r = 0) - P(r = 1) - P(r = 2)$

$$= 1 - \frac{e^{-.7}(.7)^0}{0!} - \frac{e^{-.7}(.7)^1}{1!} - \frac{e^{-.7}(.7)^2}{2!}$$

$= 1 - .49659 - .34761 - .12166 = .03414$

b) $P(r < 5) = P(r \leq 2) + P(r = 3) + P(r = 4)$

$$= .96586 + \frac{e^{-.7}(.7)^3}{3!} + \frac{e^{-.7}(.7)^4}{4!} = .96586 + .02839 + .00497 = .99922$$

c) $P(r = 9) = \frac{e^{-.7}(.7)^9}{9!} = 0.0000$

5-28 $\lambda = 4$, $e^{-4} = .0183$

a) $P(x = 0) = \frac{e^{-4}(4)^0}{0!} = .0183$

b) $P(x = 2) = \frac{e^{-4}(4)^2}{2!} = .1465$

c) $P(x = 4) = \dfrac{e^{-4}(4)^4}{4!} = .1954$

d) $P(x \geq 5) = 1 - [P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)]$

$$= 1 - \left[\frac{e^{-4}(4)^0}{0!} + \frac{e^{-4}(4)^1}{1!} + \frac{e^{-4}(4)^2}{2!} + \frac{e^{-4}(4)^3}{3!} + \frac{e^{-4}(4)^4}{4!}\right]$$

$$= 1 - [.0183 + .0733 + .1465 + .1954 + .1954] = .3711$$

5-29 binomial, $n = 25$, $p = 0.032$; $\lambda = np = 0.8$; $e^{-0.8} = 0.44933$

a) $P(r = 3) = \dfrac{e^{-0.8}(0.8)^3}{3!} = 0.0383$

b) $P(r = 5) = \dfrac{e^{-0.8}(0.8)^5}{5!} = 0.0012$

c) $P(r \leq 2) = e^{-0.8}\left(\dfrac{(0.8)^0}{0!} + \dfrac{(0.8)^1}{1!} + \dfrac{(0.8)^2}{2!}\right) = 0.9526$

5-30 Using Appendix Table 4(b),

a) $P(x \leq 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$
$= .0022 + .0137 + .0417 + .0848 = .1424$

b) $P(x \geq 2) = 1 - P(x = 0) - P(x = 1) = 1 - .0022 - .0137 = .9841$

c) $P(x = 6) = .1605$

d) $P(1 \leq x \leq 4) = P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)$
$= .0137 + .0417 + .0848 + .1294 = .2696$

5-31 $\lambda = 8$; $e^{-8} = .000335$

P(Ima plays) $= P(x \leq 5)$

$$= \frac{e^{-8}(8)^0}{0!} + \frac{e^{-8}(8)^1}{1!} + \frac{e^{-8}(8)^2}{2!} + \frac{e^{-8}(8)^3}{3!} + \frac{e^{-8}(8)^4}{4!} + \frac{e^{-8}(8)^5}{5!}$$

$= .000335 + .002684 + .010735 + .028626 + .057252 + .091604$
$= .191236$

5-32 $P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) \quad = e^{-4.1}\left[1 + \dfrac{4.1}{1!} + \dfrac{(4.1)^2}{2!} + \dfrac{(4.1)^3}{3!}\right]$

$= .01657(1 + 4.1 + 8.405 + 11.4868) = .4141$

So the probability of 4 or more stops per hour is $1 - .4141 = .5859$, and so Ford should move the employee.

5-33 $\lambda = 5$; $e^{-5} = .006738$

$P(r > 3) = 1 - P(r \leq 3) = 1 - P(r = 0) - P(r = 1) - P(r = 2) - P(r = 3)$

$$= 1 - \frac{e^{-5}(5)^0}{0!} - \frac{e^{-5}(5)^1}{1!} - \frac{e^{-5}(5)^2}{2!} - \frac{e^{-5}(5)^3}{3!}$$

$$= 1 - .006738 - .033690 - .084224 - .140374 = .734974$$

Since this exceeds .7, the funds will be allocated.

5-34 The mean number of malfunctions per 50 calculators is $50(.04) = 2$, so

a) $P(\geq 3 \text{ malfunctions}) = 1 - P(0 \text{ malfunctions}) - P(1 \text{ malfunction}) - P(2 \text{ malfunctions})$

$$= 1 - \frac{e^{-2}(2)^0}{0!} - \frac{e^{-2}(2)^1}{1!} - \frac{e^{-2}(2)^0}{2!}$$

$$= 1 - .135335 - .270671 - .270671 = .323323$$

b) $P(0 \text{ malfunctions}) = .135335$

5-35 binomial, $n = 80$, $p = .04$; $\lambda = np = 3.2$, $e^{-3.2} = .04076$

a) $P(7) = \dfrac{e^{-3.2}(3.2)^7}{7!} = .02779$

b) $P(0) = \dfrac{e^{-3.2}(3.2)^0}{0!} = .04076$

5-36 binomial, $n = 1000$, $p = .005$; $\lambda = np = 5$, $e^{-5} = .00674$

a) $P(0) = \dfrac{e^{-5}(5)^0}{0!} = .00674$

b) $P(10) = \dfrac{e^{-5}(5)^{10}}{10!} = .01813$

c) $P(15) = \dfrac{e^{-5}(5)^{15}}{15!} = .00016$

5-37 $\mu = 6.4 \quad \sigma = 2.7$

a) $P(4.0 < x < 5.0) = P\left(\frac{4 - 6.4}{2.7} < z < \frac{5 - 6.4}{2.7}\right) = P(-.89 < z < -.52)$
$= .3133 - .1985 = .1148$

b) $P(x > 2.0) = P\left(z > \frac{2.0 - 6.4}{2.7}\right) = P(z > -1.63) = .5 + .4484 = .9484$

c) $P(x < 7.2) = P\left(z < \frac{7.2 - 6.4}{2.7}\right) = P(z < .30) = .5 + .1179 = .6179$

d) $P(x < 3.0 \text{ or } x > 9.0) = P\left(z < \frac{3.0 - 6.4}{2.7}\right) + P\left(z > \frac{9.0 - 6.4}{2.7}\right)$
$= P(z < -1.26) + P(z > .96)$
$= (.5 - .3962) + (.5 - .3315) = .2723$

5-38 $\mu = np = 50(.25) = 12.5$, $\sigma = \sqrt{npq} = \sqrt{50(.25)(.75)} = 3.062$

a) $P(r > 10) = P\left(z > \frac{10.5 - 12.5}{3.062}\right) = P(z > -.65) = .5 + .2422 = .7422$

b) $P(r < 18) = P\left(z < \frac{17.5 - 12.5}{3.062}\right) = P(z < 1.63) = .5 + .4484 = .9484$

c) $P(r > 21) = P\left(z > \frac{21.5 - 12.5}{3.062}\right) = P(z > 2.94) = .5 - .4984 = .0016$

d) $P(9 < r < 14) = P\left(\frac{9.5 - 12.5}{3.062} < z < \frac{13.5 - 12.5}{3.062}\right)$
$= P(-.98 < z < .33) = .3365 + .1293 = .4658$

5-39 μ unknown, $\sigma = 50$

a) $P(x > 21) = P\left(\frac{x-\mu}{5.0} > \frac{21-\mu}{5.0}\right) = .14 = P(z > 1.08)$

Thus, $\frac{21-\mu}{5.0} = 1.08 \Rightarrow 21 - \mu = 5.4 \Rightarrow \mu = 15.6$

b) $P(x < a) = P\left(z < \frac{a-15.6}{5}\right) = .04 = P(z < -1.75)$

Thus, $\frac{a-15.6}{5} = -1.75 \Rightarrow a - 15.6 = -8.75 \Rightarrow a = 6.85$

5-40 a) $\mu = 35(.15) = 5.25, \quad \sigma = \sqrt{35(.15)(.85)} = 2.112$

$$P(7 \leq r \leq 10) = P\left(\frac{6.5-5.25}{2.112} \leq z \leq \frac{10.5-5.25}{2.112}\right)$$
$$= P(0.59 \leq z \leq 2.49) = .4936 - .2224 = .2712$$

b) $\mu = 29(.25) = 7.25, \quad \sigma = \sqrt{29(.25)(.75)} = 2.332$

$P(r \geq 9) = P\left(z \geq \frac{8.5-7.25}{2.332}\right) = P(z \geq 0.54) = .5 - .2054 = .2946$

c) $\mu = 84(.42) = 35.28, \quad \sigma = \sqrt{84(.42)(.58)} = 4.524$

$P(r \leq 40) = P\left(z \leq \frac{40.5-35.28}{4.524}\right) = P(z \leq 1.15) = .5 + .3749 = .8749$

d) $\mu = 63(.11) = 6.93, \quad \sigma = \sqrt{63(.11)(.89)} = 2.483$

$P(r \geq 10) = P\left(z \geq \frac{9.5-6.93}{2.483}\right) = P(z \geq 1.04) = .5 - .3508 = .1492$

e) $\mu = 18(.67) = 12.06, \quad \sigma = \sqrt{18(.67)(.33)} = 1.995$

$$P(9 \leq r \leq 12) = P\left(\frac{8.5-12.06}{1.995} \leq z \leq \frac{12.5-12.06}{1.995}\right)$$
$$= P(-1.78 \leq z \leq 0.22) = .4625 + .0871 = .5496$$

5-41 $\mu = 100$, σ unknown

a) $P(x < 115) = P\left(z < \frac{115-100}{\sigma}\right) = .90 = P(z < 1.28)$

Therefore, $\frac{15}{\sigma} = 1.28 \Rightarrow 15 = 1.28\sigma \Rightarrow \sigma = 11.72$

b) $P(x > a) = P\left(z > \frac{a-100}{11.72}\right) = .05 = P(z > 1.64)$

Thus, $\frac{a-100}{11.72} = 1.64 \Rightarrow a - 100 = 19.22 \Rightarrow a = 119.22$

The lowest suitable stock level is 120 tubes.

5-42 $\mu = 5.07, \sigma = .07$

$$P(5.05 \leq x \leq 5.15) = P\left(\frac{5.05-5.07}{.07} \leq z \leq \frac{5.15-5.07}{.07}\right)$$
$$= P(-.29 \leq z \leq 1.14) = .1141 + .3729 = .4870$$

5-43 a) $\mu = 95$, σ unknown

$P(x > 110) = P\left(z > \frac{110-95}{\sigma}\right) = .25 = P(z > .67)$

Thus, $\frac{15}{\sigma} = .67 \Rightarrow 15 = .67\sigma \Rightarrow \sigma = 22.39$

b) $P(x > a) = P\left(z > \frac{a-95}{22.39}\right) = .2 = P(z > .84)$

Thus, $\frac{a-95}{22.39} = .84 \Rightarrow a - 95 = 18.81 \Rightarrow a = 113.81$

The manager should order at least 114 cartridges per week.

5-44 $\mu = 4.02$, $\sigma = .08$ (since 68% of the probability in a normal distribution falls in the interval $\mu \pm \sigma$)

$P(3.9 < x < 4.1) = P\left(\frac{3.9-4.02}{.08} < z < \frac{4.1-4.2}{.08}\right)$

$= P(-1.5 < z < 1) = .4332 + .3413 = .7745 < .8$

The prototype does not satisfy the medical standards.

5-45 $\mu = 7$

a) $\sigma = .75$

$P(6.25 \leq x \leq 7.75) = P\left(\frac{6.25-7}{.75} \leq z \leq \frac{7.75-7}{.75}\right)$

$= P(-1 \leq z \leq 1) = .3413 + .3413 = .6826$

b) $\sigma = .875$

$P(6.25 \leq x \leq 7.75) = P\left(\frac{6.25-7}{.875} \leq z \leq \frac{7.75-7}{.875}\right)$

$= P(-.86 \leq z \leq .86) = .3051 + .3051 = .6102$

5-46 $\mu = 44$, $\sigma = 12$

a) $P(33 \leq x \leq 42) = P\left(\frac{33-44}{12} \leq z \leq \frac{42-44}{12}\right)$

$= P(-.92 \leq z \leq -.17) = .3212 - .0675 = .2537$

b) $P(x < 30) = P\left(z < \frac{30-44}{12}\right) = P(z < -1.17) = .5 - .3790 = .1210$

c) $P(x < 25 \text{ or } x > 60) = P\left(z < \frac{25-44}{12}\right) + P\left(z > \frac{60-44}{12}\right)$

$= P(z < -1.58) + P(z > 1.33)$

$= (.5 - .4429) + (.5 - .4082) = .1489$

5-47 $\mu = np = 200(.05) = 10$, $\sigma = \sqrt{npq} = \sqrt{200(.05)(.95)} = 3.08$

$P(7 \leq r \leq 18) = P\left(\frac{6.5-10}{3.08} \leq z \leq \frac{18.5-10}{3.08}\right)$

$= P(-1.14 \leq z \leq 2.76) = .3729 + .4971 = .8700$

5-48 a) $P(x > 72000) = P\left(z \geq \frac{72000-67000}{4000}\right) = P(z \geq 1.25) = .5 - .3944 = .1056$

b) The expected loss to future business is \$5000(.1056) = \$528. Since this can be avoided if he hires the new employees for \$200, he should hire them

5-49 $\mu = 11$, $\sigma = 2.4$

$P(x < 12) = P\left(z < \frac{12-11}{2.4}\right) = P(z < .42) = .5 + .1628 = .6628$

Since .6628(19) = 12.59, 12 or 13 of the 19 books will be completed in less than a year.

5-50 a) $\mu = 325$, $\sigma = 60$, given Estimate I is accurate

$P(x > 350 \mid EI) = P\left(z > \frac{350-325}{60}\right) = P(z > .42) = .5 - .1628 = .3372$

b) $\mu = 300$, $\sigma = 50$, given Estimate II is accurate

$$P(x > 350 \mid EII) = P\left(z > \frac{350 - 300}{50}\right) = P(z > 1) = .5 - .3413 = .1587$$

c,d) Let X denote $x > 350$.

Event	P(Event)	P(X | Event)	P(X and Event)	P(Event | X)
EI	.5	.3372	.16860	.16860/.24795 = .68 (c)
EII	.5	.1587	.07935	.07935/.24795 = .32 (d)
			P(X) = .24795	

5-51 We seek c such that $P(x \le c) = .95$, when x is normal with $\mu = 73$ and $\sigma = 6$.

$$P(x \le c) = P\left(z \le \frac{c - 73}{6}\right) = .95 = P(z \le 1.65)$$

Thus, $\frac{c - 73}{6} = 1.65 \Rightarrow c - 73 = 9.9 \Rightarrow c = 82.9$

Nobb should manufacture doors of height 82.9 inches.

5-52 a) normal b) Poisson c) binomial d) normal

5-53 A fixed number of independent trials with 2 possible outcomes implies binomial. The Poisson distribution can be used to describe the number of occurrences of certain events (such as calls through a telephone exchange) during a fixed time interval. Both are for discrete variables. The normal distribution is for continuous variables or the case where a discrete random variable takes on so many values that its distribution may be approximated by a continuous function.

5-54 A random variable is considered to be discrete if it can assume only a limited number of values. In other words, all the values could be listed. A continuous random variable can assume any value within a given range. It is impossible, therefore, to list all the possible values of a continuous random variable. Often a discrete random variable can assume so many values that we assume it is continuous for ease of calculation.

Just as there are discrete and continuous random variables, so also are there discrete and continuous probability functions, each associated with the appropriate type of random variable.

5-55 One solution is to use the services of a professional statistician who is trained to make these kinds of decisions. As an alternative, it is reasonable to expect that someone who completes this book will be able to employ the tests described to determine whether an observed distribution can be described by one of the probability distributions we will study.

5-56 Since $n = 200$ and $p = .03$, we can approximate the binomial distribution for the number of checks that bounce by a Poisson distribution with $\lambda = np = 200(.03) = 6$.

a) $P(r = 10) = \frac{6^{10} e^{-6}}{10!} = .0413$

b) $P(r = 5) = \frac{6^5 e^{-6}}{5!} = .1606$

5-57 Clearly, there are two possible outcomes to her inspection of the sides of beef: she will find that each side is either contaminated or uncontaminated. However, these "trials" are far from independent. Two different sides of beef might come from animals in the same herd, or even from the same animal. Moreover, probability of contamination might be dependent on the origin of the beef, regardless of the average probability value.

5-58

Number of Cases Researched (1)	Frequency (2)	Probability (3)	(1) × (3)
4	3	.3	1.2
5	3	.3	1.5
6	2	.2	1.2
7	1	.1	.7
8	1	.1	.8

Expected number of cases = 5.4

5.4 × 2 interns per case = 10.8, or 11, interns to hire. (But if each case requires two interns for the whole summer, then perhaps 12 should be hired.)

5-59 a) continuous b) discrete c) continuous

5-60 a) normal b) binomial c) Poisson d) normal

5-61

Loss Table

	States of Nature				
	2	3	4	5	
Probability	.20	.25	.40	.15	
Action					Expected loss
2	0	2800	5600	8400	4200
3	350	0	2800	5600	2030
4	700	350	0	2800	647.5
5	1050	700	350	0	525 ←

They should bake 5 batches each morning, and their expected daily profit is:

$$.20(4550) + .25(7700) + .40(10850) + .15(14000) = \$9275$$

5-62 Since $n = 200$ and $p = .02$, we can approximate the binomial distribution for the number of flights more than 10 minutes early or late by a Poisson distribution with $\lambda = np = 200(.02) = 4$.

$$P(r = 0) = \frac{4^0 e^{-4}}{0!} = e^{-4} = .01832$$

$$P(r = 10) = \frac{4^{10} e^{-4}}{10!} = \frac{1048576(.018316)}{3628800} = .00529$$

5-63 a) Let X be a Poisson random variable with $\lambda = 7$

$$P(X = 6) = \frac{7^6 e^{-7}}{6!} = \frac{117649(.000912)}{720} = .1490$$

b) Let Y be a Poisson random variable with $\lambda = 5$

$$P(Y = 8) = \frac{5^8 e^{-5}}{8!} = \frac{390625(.00674)}{40320} = .0653$$

5-64 Let r be a binomial random variable with $n = 15$ and $p = .25$, representing the number of applicants who were rejected.

a) $P(r = 4) = .2252$

b) $P(r = 8) = .0131$

c) $P(r < 3) = .0134 + .0668 + .1559 = .2361$

d) $P(r > 5) = 1 - .0134 - .0668 - .1559 - .2552 - .2552 - .1651 = .0884$

5-65 Let r be a binomial random variable with $n = 12$ and $p = .44$, representing the number who have typhus.

a) $P(r \geq 6) = 1 - .0010 - .0090 - .0388 - .1015 - .1794 - .2256 = .4447$

b) $P(r \leq 7) = 1 - .0684 - .0239 - .0056 - .0008 - .0001 = .9012$

c) $P(r \geq 9) = .0239 + .0056 + .0008 + .0001 = .0304$

5-66 Let r be a binomial random variable with $n = 15$ and $p = .12$, representing the number who repeat the course.

a) $P(r < 6) = 1 - .0047 - .0008 - .0001 = .9944$

b) $P(r = 5) = .0208$

c) $P(r \leq 2) = .1470 + .3006 + .2870 = .7346$

5-67 a) $\dfrac{5(166)(165)(164)(163)(323)}{489(488)(487)(486)(485)} = .0432$

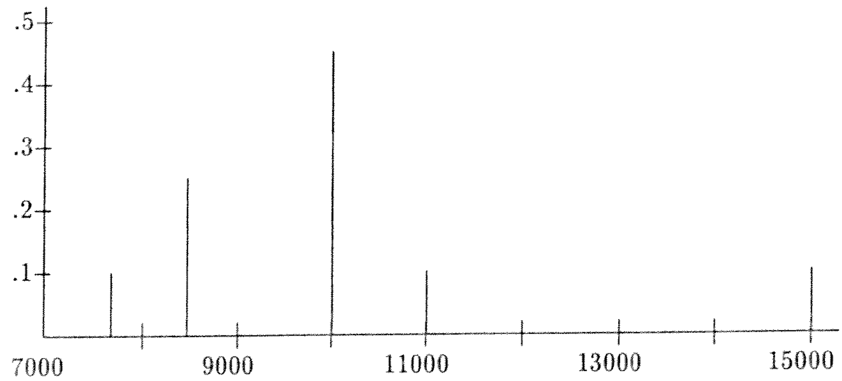
b.1) $\dfrac{4(34)(132)(131)(130)}{166(165)(164)(163)} = .4175$

b.2) $\dfrac{132(131)(130)(129)}{166(165)(164)(163)} = .3961$

b.3) $1 - \dfrac{138(137)(136)}{166(165)(164)} = .4276$

5-68 a)

Year	1992	1993	1994	1995	1996
Number of Copies	15,000	8,500	10,000	7,700	11,000
Probability	.10	.25	.45	.10	.10



b) She should order the maximum number of pamphlets which could be demanded in 1989, i.e., 15,000.

5-69

Weekly Demand (lbs) (1)	Probability (2)	(1) × (2)
2500	.30	750
3500	.45	1575
4500	.20	900
5500	.05	275
	1.00	3500

a) Expected weekly demand from past data = 3500 lbs.

b) Cost − Revenue = (3500 × \$4) − (2500 × \$5) = \$1500 loss

5-70

# of Coats Sold (1)	Probability (2)	(a) Profit (3)	(2) × (3)	(b) Profit (4)	(2) × (3)
8	.10	1160	116	920	92
10	.20	1240	248	1080	216
12	.25	1320	330	1240	310
14	.45	1400	630	1400	630
	1.00		1324		1248

a) Heidi's expected profit is \$1324.

b) Heidi's expected profit will fall to \$1248.

5-71 Let x be a normal random variable with $\mu = 5650$ and $\sigma = 120$, representing monthly mileage.

$$P(x \geq 5500) = P\left(z \geq \frac{5500 - 5650}{120}\right) = P(z \geq -1.25) = .5 + .3944 = .8944$$

5-72 a) Yes: knowing that the average number of customers varies from day to day enables the manager to know how inconvenient the ARM maintenance will be, on average.

b) Yes: only the Friday data are relevant if the maintenance is rescheduled for Friday.

5-73 Let X_A be a normal random variable with $\mu = 42$ and $\sigma = 4$;
Let X_B be a normal random variable with $\mu = 38$ and $\sigma = 7$.

a) $P(X_A > 45) = P\left(z > \frac{45 - 42}{4}\right) = P(z > .75) = .5 - .2734 = .2266$

$P(X_B > 45) = P\left(z > \frac{45 - 38}{7}\right) = P(z > 1) = .5 - .3413 = .1587$

She should choose car A.

b) $P(X_A < 39) = P\left(z < \frac{39 - 42}{4}\right) = P(z < -.75) = .5 - .2734 = .2266$

$P(X_B < 39) = P\left(z < \frac{39 - 38}{7}\right) = P(z < +.14) = .5 + .0557 = .5557$

She should choose car A.

5-74 Let x be a normal random variable with $\mu = 16{,}050$ and $\sigma = 2500$, representing the number of fans attending.

a) $P(x > 20{,}000) = P\left(z > \frac{20000 - 16050}{2500}\right) = P(z > 1.58) = .5 - .4429 = .0571$

b) $P(x < 10{,}000) = P\left(z < \frac{10000 - 16050}{2500}\right) = P(z < -2.42) = .5 - .4922 = .0078$

c) $P(14{,}000 < x < 17{,}500) = P\left(\frac{14000 - 16050}{2500} < z < \frac{17500 - 16050}{2500}\right)$

$= P(-.82 < z < .58) = .2939 + .2190 = .5129$

5-75 a)

	Expected Accidents/Week
No Action	0(.05) + 1(.10) + 2(.20) + 3(.25) + 4(.15) + 5(.10) = 2.80
Lighting	0(.10) + 1(.20) + 2(.30) + 3(.25) + 4(.10) + 5(.05) = 2.20
Bike Lanes	0(.20) + 1(.20) + 2(.20) + 3(.30) + 4(.05) + 5(.05) = 1.95

He should add bike lanes for the greatest reduction in expected number of accidents.

	$P(> 3$ accidents)	$P(\geq 3$ accidents)
No Action	$.15 + .10 = .25$	$.40 + .25 = .65$
Lighting	$.10 + .50 = .15$	$.25 + .15 = .40$
Bike Lanes	$.05 + .05 = .10$	$.30 + .10 = .40$

b) Adding bike lanes also minimizes the probability of more than 3 accidents.

c) In this case, the two actions lead to equal reduction in the probability of 3 or more accidents, so he is indifferent.

5-76 Let x be a normal random variable with $\mu = 24$ and $\sigma = 7.5$, representing the length of a lease.

a) $P(x \geq 28) = P\left(z \geq \frac{28-24}{7.5}\right) = P(z \geq .53) = .5 - .2019 = .2981$

b) $P(x < 12) = P\left(z < \frac{12-24}{7.5}\right) = P(z < -1.60) = .5 - .4452 = .0548$

5-77 Let x be a normal random variable with $\mu = 320$ and $\sigma = 25$, representing the level of smoke which activates the detector.

a) $P(x \leq 4(82)) = P\left(z \leq \frac{328-320}{25}\right) = P(z \leq .32) = .5 + .1255 = .6255$

b) $P(x \leq 3(82)) = P\left(z \leq \frac{246-320}{25}\right) = P(z \leq -2.96) = .5 - .4985 = .0015$

5-78 Since we are not given the number of graduates from each school, we will treat each school as one observation. Thus we are computing statistics and probabilities per school, rather than per graduate.

a,c)

Post-MBA Salaries x	x^2	Job Offers x
89.93	8087.4049	3.02
84.64	7163.9296	2.96
83.21	6923.9041	2.92
100.80	10160.6400	3.47
102.63	10532.9169	3.60
67.82	4599.5524	2.68
58.52	3424.5904	2.45
100.48	10096.2304	2.43
74.01	5477.4801	2.74
80.50	6480.2500	3.25
70.49	4968.8401	2.78
74.28	5517.5184	2.69
95.41	9103.0681	2.40
69.89	4884.6121	2.69
71.97	5179.6809	2.40
70.66	4992.8356	2.12
61.89	3830.3721	2.58
69.88	4883.2144	3.09
71.97	5179.6809	2.34
54.72	2994.2784	2.19
1553.70	124480.9998	54.80

For the post-MBA salaries:

$\mu = \sum x/N = 1553.70/20 = 77.6850$

$\sigma^2 = \sum x^2/N - \mu^2$
$= 124480.9998/20 - 77.6850^2 = 189.0908$

$\sigma = \sqrt{\sigma^2} = \sqrt{189.0908} = 13.7510$

b) 1. $P(x > 100) = P(z > (100 - 77.685)/13.751) = P(z > 1.62) = .5 - .4474 = .0526$

2. $P(x \leq 60) = P(z \leq (60 - 77.685)/13.751) = P(z \leq -1.29) = .5 + .4015 = .0985$

3. $P(75 \leq x \leq 95) = P((75 - 77.685)/13.751 \leq z \leq (95 - 77.685)/13.751)$
$= P(-0.20 \leq z \leq 1.26) = .0793 + .3962 = .4755$

c) $\lambda = \sum x/N = 54.8/20 = 2.74$

d) 1. $P(x < 2) = P(x{=}0) + P(x{=}1)$

$$= e^{-2.74}\left(\frac{(2.74)^0}{0!} + \frac{(2.74)^1)}{1!}\right)$$

$= .06457(1 + 2.74) = .2415$

2. $P(x{=}2 \text{ or } x{=}3) = P(x{=}2) + P(x{=}3)$

$$= e^{-2.74}\left(\frac{(2.74)^2}{2!} + \frac{(2.74)^3)}{3!}\right)$$

$= .06457(3.7538 + 3.4285) = .4638$

3. $P(x > 3) = 1 - P(x \leq 3) = 1 - .2415 - .4638 = .2947$

5-79 $\mu = np = 12(.44) = 5.28$, $\sigma = \sqrt{npq} = \sqrt{12(.44)(.56)} = 1.720$

a) $P(r \geq 6) = P\left(z \geq \frac{5.5 - 5.28}{1.720}\right) = P(z \geq .13) = .5 - .0517 = .4483$
(exact answer = .4447; 0.8% error)

b) $P(r \leq 7) = P\left(z \leq \frac{7.5 - 5.28}{1.720}\right) = P(z \geq 1.29) = .5 + .4015 = .9015$
(exact answer = .9012; 0.03% error)

c) $P(r \geq 9) = P\left(z \geq \frac{8.5 - 5.28}{1.720}\right) = P(z \geq 1.87) = .5 - .4693 = .0307$
(exact answer = .0304; 1% error)

5-80 $\mu = np = 15(.12) = 1.8$, $\sigma = \sqrt{npq} = \sqrt{15(.12)(.88)} = 1.259$

a) $P(r < 6) = P\left(z < \frac{5.5 - 1.8}{1.259}\right) = P(z < 2.94) = .5 + .4984 = .9984$
(exact answer = .9943; 0.4% error)

b) $P(r = 5) = P\left(\frac{4.5 - 1.8}{1.259} < z < \frac{5.5 - 1.8}{1.259}\right) = P(2.14 < z < 2.94)$
$= .4984 - .4838 = .0146$ (exact answer = .0208; 30% error)

c) $P(r < 3) = P\left(z < \frac{2.5 - 1.8}{1.259}\right) = P(z < .56) = .5 + .2123 = .7123$
(exact answer = .7346; 3% error)

In contrast to problem 79, where $np = 5.28 > 5$ and $nq = 6.72 > 5$ and the largest percentage error was only 1%, here where $np < 5$, we encounter larger errors.

5-81 a) Expected number of pages
$= 100(.05) + 300(.10) + 500(.25) + 700(.25) + 900(.20) + 1100(.15)$
$= 5 + 30 + 125 + 175 + 180 + 165 = 680$
He won't have enough boxes if purchase is based on a 600-page expected length.

b) Now expected number of pages = $680 - 200(.25) = 680 - 50 = 630$

5-82 Opportunity loss = 0.70 - 0.35 = \$0.35 per rose = \$350 per thousand roses
Obsolescence loss = 0.35 - 0.10 = \$0.25 per rose = \$250 per thousand roses
Actions and states of nature are given in thousands of roses.

Loss Table

	States of Nature				
	15	20	25	30	
Probability	.10	.30	.40	.20	
Action					Expected loss
15	0	1750	3500	5250	2975
20	1250	0	1750	3500	1525
25	2500	1250	0	1750	975 ←
30	3750	2500	1250	0	1625

The optimal order of roses to produce is 25 thousand.

5-83 a) $n = 3,\ p = 116/400 = .29$ $\quad P(r=2) = \frac{3!}{2!1!}(.29)^2(.71)^1 = .1791$

b) $n = 13,\ p = 116/400 = .29$ $\quad P(r=4) = \frac{13!}{4!9!}(29)^4(.71)^9 = .2319$

5-84 $n = 4,\ p = .11$

$$P(r=0) = \frac{4!}{0!4!}(.11)^0(.89)^4 = .62742$$
$$P(r=1) = \frac{4!}{1!3!}(.11)^1(.89)^3 = .31019$$
$$P(r=2) = \frac{4!}{2!2!}(.11)^2(.89)^2 = .05751$$
$$P(r=3) = \frac{4!}{3!1!}(.11)^3(.89)^1 = .00474$$
$$P(r=4) = \frac{4!}{4!0!}(.11)^4(.89)^0 = .00015$$

5-85 a) $n = 25,\ p = .05$ $\quad$ Binomial: $P(r=7) = \frac{25!}{7!18!}(.05)^7(.95)^{18} = .000149$

$np = 25(.05) = 1.25$ $\quad$ Poisson: $P(r=7) = \frac{(1.25)^7 e^{-1.25}}{7!} = .000271$

b) Binomial: $P(r=2) = \frac{25!}{2!23!}(.05)^2(.95)^{23} = .2305$

Poisson: $P(r=2) = \frac{(1.25)^2 e^{-1.25}}{2!} = .2238$

5-86 a) Assuming that approvals are independent from loan to loan, and that all loans have the same .8 probability of approval, then

$\mu = np = 1460(.8) = 1168, \quad \sigma = \sqrt{npq} = \sqrt{1460(.8)(.2)} = 15.28$

b) $\mu = 1327(.77) = 1021.79, \quad \sigma = \sqrt{1327(.77)(.23)} = 15.33$

5-87 Using Appendix Table 3, with $n = 10$ and $p = .4$, we find

a) $P(r = 6) = .1115$

b) $P(r > 3) \doteq 1 - .0060 - .0403 - .1209 - .2150 = .6178$

c) $P(2 < r < 6) = .2150 + .2508 + .2007 = .6665$

5-88
$$P(r \le 102 \mid n = 300,\ p = .4) = P\left(z \le \frac{102.5 - 300(.4)}{\sqrt{300(.4)(.6)}}\right)$$
$$= P(z \le -2.06) = .5 - .4803 = .0197$$

This small probability makes it unlikely that the 40% estimate is correct. The estimate appears to be overly optimistic.

5-89 Krista's candidate's support goes up to $p = .34 + .25(.50) = .465$. Getting the support of between 51% and 55% of a sample of $n = 250$ registered voters, means being supported by between 127.5 and 137.5 of those voters. Hence, with

$np = 250(.465) = 116.25$, and $\sqrt{npq} = \sqrt{250(.465)(.535)} = 7.8863$, we have

$$P(127.5 \le r \le 137.5) = P\left(\frac{127.5 - 116.25}{7.8863} \le z \le \frac{137.5 - 116.25}{7.8863}\right)$$
$$= P(1.43 \le z \le 2.69) = .4964 - .4236 = .0728$$

5-90 a) $p = P(\text{pay decrease}) = 10/39 = .2564$

a.1) $P(x = 5) = \frac{6!}{5!1!}(.2564)^5(.7436)^1 = .0049$

a.2) If at least five got raises, then at most one saw a drecrease:
$$P(x \le 1) = \frac{6!}{0!6!}(.2564)^0(.7436)^6 + \frac{6!}{1!5!}(.2564)^1(.7436)^5$$
$$= .1691 + .3498 = .5189$$

a.3) If fewer than four got raises, then at least three saw decreases:
$$P(x \ge 3) = 1 - P(x \le 2) = 1 - .5189 - \frac{6!}{2!4!}(.2564)^2(.7436)^4$$
$$= 1 - .5189 - .3015 = .1796$$

b) Using a calculator, $\bar{x} = 10.977$, $s = 20.912$

c.1) $P(x \ge 25) = P\left(z \ge \frac{25 - 10.977}{20.912}\right) = P(z \ge 0.67) = .5 - .2486 = .2514$

c.2) $P(x \le 5) = P\left(z \le \frac{5 - 10.977}{20.912}\right) = P(z \le -0.29) = .5 - .1141 = .3859$

c.3) $P(-15 \le x \le 15) = P\left(\frac{-15 - 10.977}{20.912} \le z \le \frac{15 - 10.977}{20.912}\right) = P(-1.24 \le z \le 0.19)$
$$= .3925 + .0753 = .4678$$

CHAPTER 6

SAMPLING AND SAMPLING DISTRIBUTIONS

6-1 The major drawback of judgmental sampling is the lack of a measure of the validity of the sample as representative of the population.

6-2 They are not necessarily mutually exclusive. For example, a situation could arise in which the history (or lack of history) of the population of interest indicates no clear-cut trends. Thus, the personal opinion of the individual conducting the study could be that probability sampling is best.

6-3 Sampling is less costly, less time-consuming, more convenient, and more efficient than complete enumeration in determining something about a population in general.

6-4 Probability samples involve more rational analysis and planning at the beginning of a study and usually take more time and money than judgment samples.

6-5 It is probably both. It is a parameter for the population of shareholders attending the meeting. However, unless all shareholders were at the meeting, it is only a sample statistic for the population of all shareholders.

6-6 From what we have been told in the problem, Jean's position is apparently quite defensible. Perhaps what makes statistical sampling unique is that it permits statistical inference to be made about a population and its parameters. This is apparently what Jean has done. There are no hard and fast rules as to the size of the sample that must be drawn before inferences can be made. Specifically there is nothing magic about the 50% mark. Common sense would seem to point out that gathering data from 50% of some populations might tend to be almost as difficult as gathering data from the entire population--for instance, the population of the United States, or the world. The defense for Jean's position lies in empirical evidence and some explanation and reasoning with the project leaders, educating them about the abilities of statistical inference.

6-7 If the deceased victims have relatives or friends who were told of the heater's behavior by the deceased, and if they could be located, the census could be completed. Otherwise, only a census of surviving victims or some other reduced population is possible.

6-8 In (b), the distributions have greater between-group variance and less within-group variance than in (a).

6-9 To make life easy, we'll ignore pages 1000–1026. Starting at the 4th line of the 2nd column, we'll read groups of 3 digits to identify the selected pages. When we reach the end of the 4th line of the 5th column, we'll go to the 5th line in the 1st column. The selected pages and numbers of words in italics are given below. (We did not count symbols or numbers as words.)

page:	543	581	078	896	708	529	131	291	265	730	489	003	130	500	995
words:	8	1	6	0	4	0	0	0	0	12	0	32	0	2	0

In all likelihood, your results were different, because you used a different way to pick your pages from Table 6-3.

6-10 Assuming a non-leap year:

1/6	1/24	2/11	3/1	3/19	4/6	4/24	5/12	5/30	6/17
7/5	7/23	8/10	8/28	9/15	10/3	10/21	11/8	11/26	12/14

6-11 (c)

6-12 The probability that a 4, 7, or 2 will appear is .10, since each digit is equally likely to be selected by a random number generator. Since there are 115 ten digit numbers, we would expect to see each number appear 11.5 times. Actually, the numbers appear as follows:

Number	4	7	2
Appearances	16	13	10

Assuming that our table is random, we would expect, in a larger version of this table, the number of appearances of any digit in any given position to be close to one tenth of the possible appearances. The large deviations from 12.5 in this case are due to insufficient sample size.

6-13 Since there is so much variation within each group, stratified sampling would not be recommended. Since the groups are very different from each other, cluster sampling is also inappropriate. Random sampling is reasonable, because a list of all the subscribers is readily available. Depending on the order in which the subscribers are listed, systematic sampling may or may not be advisable.

6-14 No. Between noon and 5 pm on a weekday, no one will be at home if both parents work, and their children are of pre-school age. Thus, the heaviest users of daycare centers are excluded from a poll on daycare centers.

6-15 Very possibly not. The population that Peterson is interested in is all of Piedmont's customers. The intersection of the set of customers and the set of persons with telephones may be large enough for Peterson to make inferences, but quite probably it is not. Peterson's poll might (1) include homes without electricity and (2) definitely does not include customers who do not have phones. If a large percentage of the population of the service area does not have telephone service, it would be very dangerous for Peterson to infer too much from his sample. In short, for his purposes, the sample is probably not random.

6-16 Every seventh is better. If every fifth lot were inspected, then the output of only two machines would be inspected. If every seventh lot is inspected, then each machine's output will be inspected.

6-17 Yes. Systematic sampling is particularly useful when the population being sampled is randomly distributed. In this case, accidents are filed by date. For the purpose that the department has in mind, such a filing is random. The department has no control over when the accidents happen so there seems little chance that there is any underlying systematic distribution; hence the results obtained from the sample will be representative of the population.

6-18 We need to remember that when we use stratified sampling, we are only acknowledging that the population is already divided into groups of different sizes. However, these groups are required to be relatively homogeneous before we can use this sampling method. In the case under question, Mary proposes dividing the population into groups such as urban, suburban, and rural; examination of the differences among these groups should convince you that they are reasonably homogeneous, and thus stratified sampling would work.

6-19 a) The only way to guarantee that there will be no sample error is to sample the entire population.

b) No sample size will guarantee a zero standard error of the mean, unless each member of the population has the same value.

6-20 Sampling error, the chance error which is due to the particular elements selected as a sample of the population.

6-21 Your box fell short of the company's claim by 0.1 cups, which is only 0.5 standard deviations. Since this is so close to the claimed average, you have no reason to doubt the validity of the claim.

6-22 A sample mean overestimating the true mean is no better than one under- estimating the true mean. In this case, 30 cents is closer than 35 cents to the true mean, in absolute differences:

$$|30 - 31.4| = 31.4 - 30 = 1.4 < |35 - 31.4| = 35 - 31.4 = 3.6$$

6-23 No. In a sampling distribution of the mean, all samples are of the same size. It is very unlikely that she interviewed the same number of people each day, unless she did so intentionally. In any case, the sampling distribution in question is the distribution of the means of all possible samples of the given size, and she certainly has not looked at all such samples.

6-24 Average weekly sales have decreased from 3538 cartons to 3462 cartons.

6-25 No. The standard deviation of the mean of all possible samples is referred to as the standard error of the mean. It is a measure of the dispersion of the theoretical sampling distribution rather than, as the secretary assumes, an allowable deviation from the population mean.

6-26 The information gathered concerns mean customer satisfaction for groups of 30 customers, not for single customers, so it is a sample from the sampling distribution of the mean of samples of size 30 drawn from the customer population. It is not a sample from the customer population.

6-27 $n = 16 \qquad \mu = 150 \qquad \sigma^2 = 256 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 16/\sqrt{16} = 4$

a) $P(\bar{x} < 160) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{160-150}{4}\right) = P(z < 2.5) = .5 + .4938 = .9938$

b) $P(\bar{x} > 142) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{142-150}{4}\right) = P(z > -2) = .5 + .4772 = .9772$

$n = 9 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 16/\sqrt{9} = 5.33$

c) $P(\bar{x} < 160) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{160-150}{5.33}\right) = P(z < 1.88) = .5 + .4699 = .9699$

d) $P(\bar{x} > 142) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{142-150}{5.33}\right) = P(z > -1.5) = .5 + .4332 = .9332$

6-28 a) $n = 19 \qquad \mu = 18 \qquad \sigma = 4.8 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 4.8/\sqrt{19} = 1.101$

$$P(16 < \bar{x} < 20) = P\left(\frac{16-18}{1.101} < z < \frac{20-18}{1.101}\right) = P(-1.82 < z < 1.82) = 2(.4656) = .9312$$

b) The same as (a), since the distribution is continuous

c) $n = 48, \qquad \mu = 18 \qquad \sigma = 4.8 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 4.8/\sqrt{48} = 0.693$

$$P(16 < \bar{x} < 20) = P\left(\frac{16-18}{0.693} < z < \frac{20-18}{0.693}\right) = P(-2.89 < z < 2.89) = 2(.4981) = .9962$$

6-29 $\mu = 56 \qquad \sigma = 21 \qquad P(z > -1.28) = .90$

Want $P(\bar{x} > 52) = .90 \Rightarrow \dfrac{52 - 56}{21/\sqrt{n}} \le -1.28 \Rightarrow \sqrt{n} \ge 6.72 \Rightarrow n \ge 45.16$

A sample size of 46 will be sufficient.

6-30 $\mu = 375 \qquad \sigma = 48 \qquad P(-1.96 < z < 1.96) = .95$

Want $P(370 < \bar{x} < 380) \ge .95 \Rightarrow P\left(0 < z < \dfrac{380 - 375}{48/\sqrt{n}}\right) \ge .475 \Rightarrow \dfrac{5\sqrt{n}}{48} \ge 1.96$

$\Rightarrow \sqrt{n} \ge 18.816 \Rightarrow n \ge 354.04$

A sample size of 355 will be sufficient.

6-31 $\mu = 50 \qquad \sigma^2 = 9 \qquad \sigma = 3$

a) $n = 4 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 3/\sqrt{4} = 1.5$

$P(\bar{x} \ge 48) = P\left(\dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} \ge \dfrac{48 - 50}{1.5}\right) = P(z \ge -1.33) = .5 + .4082 = .9082$

b) The probability will rise, because additional sampling decreases the standard error of the sampling distribution. Thus, $P(48 \le \bar{x} \le 50)$ increases, meaning that $P(\bar{x} \ge 48) = .5 + P(48 \le \bar{x} \le 50)$ increases.

6-32 $\mu = \$62{,}000 \qquad \sigma = \$4{,}200$

a) $n = 1 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = \$4{,}200$

$P(\bar{x} \ge 65{,}000) = P\left(\dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} \ge \dfrac{65{,}000 - 62{,}000}{4{,}200}\right) = P(z \ge .71)$

$= .5 - .2611 = .2389$

b) The probability is less. More sampling decreases the standard error of the sampling distribution of the mean.

$n = 2 \qquad \sigma_{\bar{x}} = \$4200/\sqrt{2} = \$2970$

$P(\bar{x} \ge 65000) = P\left(\dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} \ge \dfrac{65000 - 62000}{2970}\right) = P(z \ge 1.01)$

$= .5 - .3438 = .1562$

The decrease is .0827, or 35%.

6-33 The sample size of 75 is large enough to use the central limit theorem.

$\mu = 86 \qquad \sigma = 16 \qquad n = 75 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 16/\sqrt{75} = 1.848$

$P(\bar{x} < 84) = P\left(\dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} < \dfrac{84 - 86}{1.848}\right) = P(z < -1.08) = .5 - .3599 = .1401$

$P(\bar{x} > 90) = P\left(\dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} > \dfrac{90 - 86}{1.848}\right) = P(z > 2.16) = .5 - .4846 = .0154$

Thus, $P(\bar{x} < 84 \text{ or } \bar{x} > 90) = .1401 + .0154 = .1555$

6-34 a) $\mu = 4300 \qquad \sigma = 730 \qquad n = 3 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 730/\sqrt{3} = 421.5$

For a set of monitors to last 13000 hours, they must each last 4333.33 hours on average.

$$P(\bar{x} \geq 4333.33) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \geq \frac{4333.33-4300}{421.5}\right)$$
$$= P(z > .08) = .5 - .0319 = .4681$$

b) For the set to last at most 12630 hours, the average life cannot exceed 4210 hours.
$$P(\bar{x} \leq 4210) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \leq \frac{4210-4300}{421.5}\right) = P(z \leq -.21) = .5 - .0832 = .4168$$

6-35 $\mu = 64$ $\sigma = \sqrt{17.6} = 4.195$ $n = 35$ $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 4.195/\sqrt{35} = 0.709$

a) $P(\bar{x} > 72) = P\left(z > \frac{72-64}{0.709}\right) = P(z > 11.28) \approx 0$

b) $P(64 < \bar{x} < 72) = P(0 < z < 11.28) \approx 0.5$

c) $P(\bar{x} = 64) = P(z = 0) = 0$

d) $P(\bar{x} > 94) = P\left(z > \frac{72-64}{0.709}\right) = P(z > 42.31) \approx 0$

e) No. A sample average of 100 ppm is so far above what is expected, and hence so extremely unlikely, that it casts very strong doubt on the reliability of the study.

6-36 The sample size of 48 is large enough to use the central limit theorem.

$\mu = 110$ $\sigma = 64$ $n = 48$ $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 64/\sqrt{48} = 9.238$

$$P(\bar{x} < 120) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{120-110}{9.238}\right) = P(z < 1.08) = .5 + .3599 = .8599 > .80$$

The overhaul will not be ordered.

6-37 $\mu = 7500$ $\sigma = 3300$ $n = 1800$ $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 3300/\sqrt{1800} = 77.782$

a) $P(\bar{x} > 7700) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{7700-7500}{77.782}\right) = P(z > 2.57) = .5 - .4949 = .0051$

b) $P(\bar{x} < 7400) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{7400-7500}{77.782}\right) = P(z < -1.29) = .5 - .4015 = .0985$

c) $P(7275 < \bar{x} < 7650) = P\left(\frac{7275-7500}{77.782} < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{7650-7500}{77.782}\right)$
$$= P(-2.89 < z < 1.93) = .4981 + .4732 = .9713$$

6-38 $\mu = 120$ $\sigma = 12$ $n = 60$

a) $\mu_{\bar{x}} = \mu = 120$

b) $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 12/\sqrt{60} = 1.549$

c) $P(\bar{x} > 123.8) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{123.8-120}{1.549}\right) = P(z > 2.45) = .5 - .4929 = .0071$

d) $P(117 < \bar{x} < 122) = P\left(\frac{117-120}{1.549} < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{122-120}{1.549}\right)$
$$= P(-1.94 < z < 1.29) = .4738 + .4015 = .8753$$

6-39 $\mu = 168$ $\sigma = \sqrt{361} = 19$ $n = 25$ $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 19/\sqrt{25} = 3.8$

a) P(total weight > 4250) $= P(\bar{x} > 170) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{170-168}{3.8}\right)$
$$= P(z > 0.53) = .5 - .2019 = .2981$$

b) Note: $.05 = P(z > 1.64)$

We want to find c such that P(total weight $> c) = .05$

$$\Rightarrow P\left(\bar{x} > \frac{c}{25}\right) = .05 \Rightarrow P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} > \frac{\frac{c}{25} - 168}{\frac{19}{5}}\right) = .05$$

$$\Rightarrow P\left(z > \frac{c - 4200}{95}\right) = .05 \Rightarrow \frac{c - 4200}{95} = 1.64 \Rightarrow c - 4200 = 155.8$$

$\Rightarrow c = 4355.8$ is the 95th percentile of the distribution of total weight of ferry passengers.

The ferry is not complying with safety regulations.

6-40 $N = 75$ $n = 32$ $\mu = 364$ $\sigma = \sqrt{18} = 4.243$

a) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}} = \frac{4.243}{\sqrt{32}}\sqrt{\frac{75-32}{75-1}} = 0.572$

b) $P(363 \le \bar{x} \le 366) = P\left(\frac{363-364}{0.572} \le z \le \frac{366-364}{0.572}\right)$

$= P(-1.75 \le z \le 3.50) = .4599 + .5 = .9599$

c) With replacement, $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 4.243/\sqrt{32} = 0.750$

6-41 $N = 80$ $\mu = 22$ $\sigma = 3.2$ $n = 25$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}} = \frac{3.2}{5}\sqrt{\frac{55}{79}} = .5340$

$P(21 < \bar{x} < 23.5) = P\left(\frac{21-22}{.5340} < z < \frac{23.5-22}{.5340}\right)$

$= P(-1.87 < z < 2.81) = .4693 + .4975 = .9668$

6-42 $N = 80$ $\mu = 8.2$ $\sigma = 2.1$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}}$

a) $n = 16$ $\sigma_{\bar{x}} = \frac{2.1}{4}\sqrt{\frac{64}{79}} = .4725$

b) $n = 25$ $\sigma_{\bar{x}} = \frac{2.1}{5}\sqrt{\frac{55}{79}} = .3504$

c) $n = 49$ $\sigma_{\bar{x}} = \frac{2.1}{7}\sqrt{\frac{31}{79}} = .1879$

6-43 μ unknown $\sigma = 216.4$

a) $n = 800$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{216.4}{\sqrt{800}} = 7.651$

$P(\mu < \bar{x} < \mu + 300) = P\left(0 < z < \frac{300}{7.651}\right) = P(0 < z < 39.21) \approx 0.5$

b) $0.95 = P(-1.96 < z < 1.96) = P\left(\frac{-100}{\sigma_{\bar{x}}} < z < \frac{100}{\sigma_{\bar{x}}}\right)$

Hence $1.96 = \frac{100}{216.4/\sqrt{n}} = 0.4621\sqrt{n}$, so $n = \left(\frac{1.96}{0.4621}\right)^2 = 17.99$

They should sample at least 18 tires.

6-44 $N = 45$ $n = 9$ $\mu = 225{,}000$ $\sigma = 39{,}000$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{39{,}000}{3}\sqrt{\frac{36}{44}} = 11{,}759$$

$$P(9\bar{x} \geq 2{,}100{,}000) = P\left(\bar{x} \geq \frac{2{,}100{,}000}{9} = 233{,}333.33\right)$$

$$= P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \geq \frac{233{,}333.33 - 225{,}000}{11{,}759}\right)$$

$$= P(z \geq .71) = .5 - .2611 = .2389$$

6-45 $\sigma = 150$

We want to find n such that $\frac{150}{\sqrt{n}} = \sigma_{\bar{x}} \leq 25 \Rightarrow \sqrt{n} \geq 6 \Rightarrow n \geq 36$

She should take at least 36 readings.

6-46 With $250(.36) = 90$ contributors, the average donation must be between \$1222.22 and \$1333.33 for the total to be between \$112,950 and \$116,100. Each 4% gift has $\mu = .04(32{,}000) = \$1{,}280$ and $\sigma = .04(9{,}600) = \$384$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{384}{\sqrt{90}}\sqrt{\frac{160}{249}} = 32.4467$$

$$P(1222.22 < \bar{x} < 1333.33) = P\left(\frac{1222.22 - 1280}{32.4467} < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < \frac{1333.33 - 1280}{32.4467}\right)$$

$$= P(-1.78 < z < 1.64) = .4625 + .4495 = .9120$$

6-47 $\mu = 96$ $\sigma = 7$ $n = 6$ $\sigma_{\bar{x}} = \sigma/n = 7/\sqrt{6} = 2.858$

$$P(\bar{x} \geq 98) = P\left(\frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \geq \frac{98-96}{2.858}\right) = P(z \geq .70) = .5 - .2580 = .2420$$

(The population is the large set of planes on which the de-icing system can be installed, not just the 30 planes on which it has been installed so far. Thus, it is incorrect to use the finite population multiplier for this problem.)

6-48 $N = 145$ $n = 36$ $\sigma = 1200$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1200}{\sqrt{36}}\sqrt{\frac{109}{144}} = 174.01$

$$P(\mu - 200 \leq \bar{x} < \mu + 200) = P\left(\frac{-200}{174.01} \leq \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \leq \frac{200}{174.01}\right)$$

$$= P(-1.15 \leq z \leq 1.15) = .3749 + .3749 = .7498$$

6-49 $\mu = 310$ $\sigma = 150$

We want to find an n large enough so that

$$\sigma_{\bar{x}} = 150/\sqrt{n} \leq (.015)(310) = 4.65 \quad \Rightarrow 150 \leq 4.65\sqrt{n}$$

$$\Rightarrow \sqrt{n} \geq 32.258 \Rightarrow n \geq 1040.58$$

Thus, a sample size of at least 1041 is needed.

6-50 Judgmental, because the sample (those skates inspected) is determined by whether the skate is Crash's size.

6-51 Proponents of random sampling often argue that it is better than judgmental sampling because more statistical inferences can be made and the information obtained is more reliable. In many situations, these arguments are true. In this case, the opposite may be true. Judgmental sampling seems to have worked well. Given the cost of the proposed alternative, it may be better to leave well enough alone.

6-52 Stratified sampling is used when we have reason to believe that we can divide the population into groups which are relatively homogeneous. In this instance, the manager feels that the residents fall into various age and income levels. Since he is trying to measure residents' attitudes, stratified sampling may be his best method, especially since he feels residents in different age and income groups may have different attitudes.

6-53 No. It must be noted that the employees of different departments probably have varying backgrounds and concerns. For these reasons, cluster sampling will not give representative results to the management. The problems in a research and development department may be entirely different from those in an assembly department. Similarly, cultural differences may be prevalent in the different departments. The company would obtain better and more representative results if it incorporated input from every department into its plan.

6-54 $\mu = 26 \qquad \sigma = 5.65 \qquad \sigma_{\bar{x}} = 5.65/\sqrt{n}$

We want to find n such that: $P(25 \le \bar{x} \le 27) \ge .9544 = P(-2 \le z \le 2)$

$$\Rightarrow P\left(\frac{25-26}{5.65/\sqrt{n}} \le \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \le \frac{27-26}{5.65/\sqrt{n}}\right) \ge P(-2 \le z \le 2)$$

$$\Rightarrow P(-.177\sqrt{n} \le z \le .177\sqrt{n}) \ge P(-2 \le z \le 2)$$

$$\Rightarrow .177\sqrt{n} \ge 2 \Rightarrow \sqrt{n} \ge 11.3 \Rightarrow n \ge 127.7$$

Thus, a sample of at least 128 customers is needed.

6-55 a)

State	Rate x	x^2	State	Rate x	x^2
Alabama	7.5	56.25	Montana	7.3	53.29
Alaska	10.1	102.01	Nebraska	2.8	7.84
Arizona	8.4	70.56	Nevada	6.8	46.24
Arkansas	7.0	49.00	New Hampshire	7.5	56.25
California	8.7	75.69	New Jersey	7.5	56.25
Colorada	6.3	39.69	New Mexico	7.6	57.76
Connecticut	7.4	54.76	New York	8.5	72.25
Delaware	6.4	40.96	North Carolina	6.4	40.96
District of Columbia	8.2	67.24	North Dakota	5.3	28.09
Florida	8.1	65.61	Ohio	7.8	60.84
Georgia	6.3	39.69	Oklahoma	6.8	46.24
Hawaii	3.5	12.25	Oregon	8.6	73.96
Idaho	7.8	60.84	Pennsylvania	7.6	57.76
Illinois	8.2	67.24	Rhode Island	8.9	79.21
Indiana	6.3	39.69	South Carolina	7.1	50.41
Iowa	5.3	28.09	South Dakota	4.0	16.00
Kansas	3.6	12.96	Tennessee	7.0	49.00
Kentucky	7.0	49.00	Texas	7.4	54.76

Louisiana	6.9	47.61	Utah	5.0	25.00
Maine	8.4	70.56	Vermont	7.1	50.41
Maryland	7.4	54.76	Virginia	6.8	46.24
Massachusetts	10.0	100.00	Washington	8.3	68.89
Michigan	10.0	100.00	West Virginia	12.9	168.49
Minnesota	6.3	39.69	Wisconsin	5.7	32.49
Mississippi	8.1	65.61	Wyoming	7.5	56.25
Missouri	5.6	31.36		367.0	2793.92

$\mu = \sum x/N = 367.0/51 = 7.196$

$\sigma^2 = \sum x^2/N - \mu^2 = 2793.92/51 - 7.196^2 = 3.000$

$\sigma = \sqrt{\sigma^2} = \sqrt{3.000} = 1.732$

b) $\bar{x} = \sum x/n = (7.5 + 3.6 + 10.0 + 2.8 + 6.4)/5 = 6.06$

c) $\mu_{\bar{x}} = \mu = 7.196, \quad \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} \times \sqrt{\dfrac{N-n}{N-1}} = \dfrac{1.732}{\sqrt{5}} \times \sqrt{\dfrac{51-5}{51-1}} = 0.743$

d) Probably not. The sample size is small (n=5), and the population itself, with N=51, is too small to be considered even approximately normal.

e) $P(5.9 \le \bar{x} \le 6.5) = P((5.9 - 7.196)/0.743 \le z \le (6.5 - 7.196)/0.743$
$= P(-1.74 \le z \le -0.94) = 0.4591 - 0.3264 = 0.1327$

6-56 In this situation, Fargo Lanna is not constrained by (1) cost, (2) time, (3) destruction of population members, or (4) accessibility of the population. Nevertheless, sampling would be entirely appropriate for this company's purpose. It will be able to obtain the same information without expending nearly the effort required to poll all employees. The company may get the same job done by using a sample, and will be able to reassign the clerical staff earlier.

6-57 There is no truth to what Simmons said. We would expect that the sample mean would overstate the population mean as often as it would understate it. Sample variation works in both ways, and to assume that the sample mean always understates is to be guilty of an error in thinking.

6-58 Again Simmons is incorrect. A sampling distribution of means is a frequency distribution of the means of all possible samples. It is not in any sense a graph of the individual observations in sample contributions.

6-59 $\sigma = 275 \qquad n = 25 \;\Rightarrow\; \sigma_{\bar{x}} = 275/\sqrt{25} = 55$

If $= \sigma_{\bar{x}} = 55/2 = 27.5$, there must be a sample size n^* such that

$\sigma_{\bar{x}} = 275/\sqrt{n^*} = 27.5 \;\Rightarrow\; \sqrt{n^*} = 275/27.5 \;\Rightarrow\; \sqrt{n^*} = 10 \;\Rightarrow\; n^* = 100$

Sampling another 75 women costs $1600 but increases the benefits by $1720. She should increase the sample size.

6-60 a)

Job Title	Wage x	x^2	Job Title	Wage x	x^2
Assembler A	10.72	114.9184	Packaging-wrapping packer	9.04	81.7216
Assembler B	9.13	83.3569	Packer - heavy	10.08	101.6064
Assembler C	7.98	63.6804	Packer - light	8.82	77.7924
Carpenter, maintenance	13.58	184.4164	Painter - maintenance	12.72	161.7984
Chemical compounder	12.64	159.7696	Painter - spray	9.78	95.6484
Chemical mixer	11.19	125.2161	Plastic injection molder	9.72	94.4784
Degreaser operator	9.11	82.9921	Polisher & buffer A	10.24	104.8576
Drill press operator A	12.01	144.2401	Polisher & buffer B	9.59	91.9681
Drill press operator B	9.89	97.8121	Punch press die setter A	12.80	163.8400
Drill press operator C	9.51	90.4401	Punch press die setter B	11.31	127.9161
Electrician	15.37	236.2369	Punch press - heavy	9.75	95.0625
Grinding machine operator A	12.92	166.9264	Punch press - light	8.91	79.3881
Grinding machine operator B	9.89	97.8121	Punch press - setup/operate	11.32	128.1424
Group leader A	13.55	183.6025	Receiving clerk	9.98	99.6004
Group leader B	11.28	127.2384	Screw machine operator	16.01	256.3201
Guard-watchman	9.86	97.2196	Screw machine setup	12.40	153.7600
Inspector A	11.55	133.4025	Shear operator	10.45	109.2025
Inspector B	10.11	102.2121	Shipper/receiver	9.73	94.6729
Inspector C	8.57	73.4449	Shipping clerk	10.03	100.6009
Janitor - heavy	9.19	84.4561	Solderer A	5.69	32.3761
Janitor - light	8.26	68.2276	Solderer B	9.88	97.6144
Laborer-dock hand	9.26	85.7476	Stationary engineer	16.52	272.9104
Lathe operator-turret A	12.66	160.2756	Stock person A	9.71	94.2841
Lathe operator-turret B	10.62	112.7844	Stock person B	8.86	78.4996
Lift truck operator	10.52	110.6704	Test analyzer - junior	10.04	100.8016
Machine operator	9.82	96.4324	Test analyzer - senior	11.73	137.5929
Maintenance machinist A	15.31	234.3961	Tool & die maker A	17.66	311.8756
Maintenance machinist B	14.42	207.9364	Tool & die maker B	15.49	239.9401
Maintenance machinist C	12.07	145.6849	Tool & die maker C	11.72	137.3584
Maintenance person A	14.13	199.6569	Tool room machinist	13.55	183.6025
Maintenance person B	11.19	125.2161	Trucker, hand	9.00	81.0000
Modelmaker	16.35	267.3225	Warehouse person	9.87	97.4169
Numerical control A	13.99	195.7201	Welder - arc acetylene	12.69	161.0361
Numerical control B	10.69	114.2761	Welder - spot	10.01	100.2001
Packager	7.92	62.7264	Wirer A	10.67	113.8489
Packaging-wrapping operator	9.93	98.6049	Wirer B	8.81	77.6161
				799.77	9271.4231

$$\mu = \sum x/N = 799.77/72 = \$11.108$$

$$\sigma^2 = \sum x^2/N - \mu^2 = 9271.4231/72 - 11.108^2 = 5.3821$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{5.3821} = \$2.320$$

b) $\bar{x} = \sum x/n = (7.98+15.37+15.37+9.86+10.52+16.35+12.80+10.45+10.67)/9$
$= \$12.152$

c) $\mu_{\bar{x}} = \mu = 11.108, \ \sigma_{\bar{x}} = \sigma/\sqrt{n} = 2.320/\sqrt{9} = \$0.773$

d) Probably not. The sample size is small (n=9), and the population, with 15 of its 72 members between \$9.50 and \$10.00 and a distinct skewness to the right, cannot be considered even approximately normal.

e) $P(10.5 \leq x \leq 11.7) = P((10.5 - 11.108)/0.773 \leq z \leq (11.7 - 11.108)/0.773)$
$= P(-0.79 \leq z \leq 0.77) = 0.2852 + 0.2794 = 0.5646$

6-61 $\mu = 42 \qquad \sigma = 11 \qquad n = 5 \qquad \sigma_{\bar{x}} = \sigma/n = 11/\sqrt{5} = 4.919$

$$P(\bar{x} > 50) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} > \frac{50 - 42}{4.919}\right) = P(z > 1.63) = .5 - .4484 = .0516$$

6-62 $\mu = 41 \qquad \sigma = 8 \qquad n = 6 \qquad \sigma_{\bar{x}} = \sigma/n = 8/\sqrt{6} = 3.266$

$$P(\bar{x} < 50) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} < \frac{50 - 41}{3.266}\right) = P(z < 2.76) = .5 + .4971 = .9971$$

6-63 $\mu = 5.8 \qquad \sigma = .8 \qquad n = 20 \qquad \sigma_{\bar{x}} = \sigma/n = .8/\sqrt{20} = .1789$

$$P(5.5 < \bar{x} < 6.2) = P\left(\frac{5.5 - 5.8}{.1789} < \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} < \frac{6.2 - 5.8}{.1789}\right)$$

$$= P(-1.68 < z < 2.24) = .4535 + .4875 = .9410$$

6-64 Cost = Benefit

$$\Rightarrow 4n = \frac{5249}{\sigma_{\bar{x}}} = \frac{5249}{\frac{\sigma}{\sqrt{n}}} = \frac{5249\sqrt{n}}{265} = 19.81\sqrt{n} \Rightarrow \sqrt{n} = \frac{19.81}{4} = 4.95 \Rightarrow n = 24.5$$

Thus, she should sample at least 25 detectors.

6-65 $N = 70 \qquad n = 15 \qquad \mu = 18 \qquad \sigma = 4$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}} = \frac{4}{\sqrt{15}}\sqrt{\frac{55}{69}} = .9221$$

a) $P(\bar{x} < 15.5) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} < \frac{15.5 - 18}{.9221}\right) = P(z < -2.71) = .5 - .4966 = .0034$

b) $P(\bar{x} > 20) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} > \frac{20 - 18}{.9221}\right) = P(z > 2.17) = .5 - .4850 = .0150$

6-66 a) Enumeration

b) Finite population

6-67 a) No. Four in ten Californians who were interviewed expressed this view.

b) Probably stratified, as a result of dividing the state's population into various demographic strata.

c) Strictly speaking, no; but as a practical matter, 32 million could be treated as infinite relative to the size of the sample in most polls.

CHAPTER 7

ESTIMATION

7-1 Estimation, hypothesis testing

7-2 a) Measuring an entire population may be impossible or infeasible, because of time and cost considerations, among other reasons.

b) A sample yields only an estimate and is subject to sampling errors.

7-3 a) A point estimate is one value which is either right or wrong, and it gives no idea of the error which might be involved.

b) An estimate of the error which might be involved is included with the point estimate.

7-4 a) An estimator is a sample statistic used to estimate a population parameter.

b) An estimate is a specific numerical value for an estimator, resulting from the particular sample which is observed.

7-5 **Unbiasedness.** An estimator is unbiased if it tends to assume values above the population parameter being estimated as frequently and to the same extent that it assumes values below the parameter being estimated.

Efficiency. An estimator becomes more efficient as the size of the standard error, or variation, becomes smaller.

Consistency. An estimator is consistent if, as the sample size increases, the value of the statistic becomes closer to the value of the population parameter.

Sufficiency. An estimator is sufficient if it makes so much use of the information in the sample that no other estimator would add information about the population parameter being estimated.

7-6 It assures us that the estimator becomes more reliable with larger samples.

7-7 $\bar{x} = \frac{\sum x}{n} = \frac{5.9}{16} = 0.369$ inches

7-8 Using a calculator, $\bar{x} = 296.583$ people, $s = 40.751$ people

7-9 $\sum x^2 = 1172.64$ $\quad \sum x = 179.4$ $\quad n = 30$ $\quad \bar{x} = 179.4/30 = 5.98$

$\sum(x-\bar{x})^2 = \sum x^2 - n\bar{x}^2 = 1172.64 - 30(5.98)^2 = 99.828$

a) $s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{99.828}{29} = 3.4423, \quad s = \sqrt{3.4423} = 1.855$ thousands of dollars

b) $\sigma^2 = \frac{\sum(x-\bar{x})^2}{N} = \frac{99.828}{30} = 3.3276, \quad \sigma = \sqrt{3.3276} = 1.824$ thousands of dollars

7-10 $n = 400$ $\quad x = 184$ $\quad \bar{p} = 184/400 = .46$

7-11 Using a calculator:

a) $\overline{x} = 16.933$ dollars

b) $s^2 = 125.924(\text{dollars})^2$

c) No, because of the rounding we cannot estimate the average length of a call. $16.933/3 = 5.644$ minutes is an unbiased estimate of the average amount of time billed per call. It is an overestimate of the average length of a call.

7-12 $\sigma = 1.4 \qquad n = 60 \qquad \overline{x} = 6.2$

a) $\sigma_{\overline{x}} = \sigma/\sqrt{n} = 1.4/\sqrt{60} = .181$

b) $\overline{x} \pm \sigma_{\overline{x}} = 6.2 \pm .181 = (6.019,\ 6.381)$

7-13 $\sigma = 1.65 \qquad n = 32 \qquad \overline{x} = 34.8$

a) $\sigma_{\overline{x}} = \sigma/\sqrt{n} = 1.65/\sqrt{32} = .292$

b) $\overline{x} \pm 3\sigma_{\overline{x}} = 34.8 \pm 3(0.292) = 34.8 \pm .876 = (33.924,\ 35.676)$

7-14 $\sigma = 0.8 \qquad n = 421 \qquad \overline{x} = 6.2$

a) $\sigma_{\overline{x}} = \sigma/\sqrt{n} = 0.8/\sqrt{421} = .0390$ pounds

b) $\overline{x} \pm 2\sigma_{\overline{x}} = 14.2 \pm 2(.039) = 14.2 \pm .078 = (14.122,\ 14.278)$

7-15 $\sigma = 3.76 \qquad n = 30 \qquad \overline{x} = 71 \qquad \sigma_{\overline{x}} = \sigma/\sqrt{n} = 3.76/\sqrt{30} = .686$

a) $\overline{x} \pm \sigma_{\overline{x}} = 71 \pm .686 = (70.31,\ 71.69)$ customers

b) $\overline{x} \pm 3\sigma_{\overline{x}} = 71 \pm 3(.686) = 71 \pm 2.058 = (68.94,\ 73.06)$ customers

7-16 $\sigma = .9 \qquad n = 75 \qquad \overline{x} = 7 \qquad \sigma_{\overline{x}} = \sigma/\sqrt{n} = .9/\sqrt{75} = .104$

$\overline{x} \pm 2\sigma_{\overline{x}} = 7 \pm 2(.104) = 7 \pm .208 = (6.79,\ 7.21)$

7-17 $\sigma^2 = 131 \qquad n = 61 \qquad \overline{x} = 894 \qquad \sigma_{\overline{x}} = \sigma/\sqrt{n} = \sqrt{131}/\sqrt{61} = 1.465$

a) $\overline{x} \pm \sigma_{\overline{x}} = 894 \pm 1.465 = (892.535,\ 895.465)$ kwh

b) $\overline{x} \pm 3\sigma_{\overline{x}} = 894 \pm 3(1.465) = 894 \pm 4.395 = (889.605,\ 898.395)$ kwh

c) Multiplying the usage interval in (a) by \$0.12/kwh gives
$(0.12(892.535),\ 0.12(895.465)) = (\$107.10,\ \$107.46)$

7-18 $\sigma = 8.3 \qquad N = 621 \qquad n = 76 \qquad \overline{x} = 29.8$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{8.3}{\sqrt{6}}\sqrt{\frac{621-76}{621-1}} = .893 \text{ students}$$

a) $\overline{x} \pm 2\sigma_{\overline{x}} = 29.8 \pm 2(.893) = 29.8 \pm 1.786 = (28.01,\ 31.59)$ students

b) We cannot be 95.5% certain that the average class size in Foresight County is less than that of Hindsight County. Therefore, we cannot conclude that Dee has met her goal.

7-19 The confidence level associated with an interval estimate is the fraction of the time that such an interval will contain the value (population parameter) being estimated.

7-20 The confidence interval is the range of an estimate, i.e., the interval between and including the upper confidence limit and the lower confidence limit.

7-21 $\bar{x} \pm 1.28\sigma_{\bar{x}}$

7-22 a) High confidence levels produce wide intervals, so we sacrifice precision to gain confidence.

b) Narrow intervals result from low confidence levels, so we sacrifice confidence to gain precision.

7-23 $\sigma = 27 \qquad n = 50 \qquad \bar{x} = 86 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 27/\sqrt{50} = 3.818$

a) $\bar{x} \pm 2\sigma_{\bar{x}} = 86 \pm 2(3.818) = 86 \pm 7.636 = (78.36, 93.64)$

b) Now $\sigma_{\bar{x}} = 27/\sqrt{5000} = .3818$, so $\bar{x} \pm 2\sigma_{\bar{x}} = 86 \pm 0.7636 = (85.24, 86.76)$

c) It is less costly to sample the 50 items necessary to get estimate (a); on the other hand, estimate (b) is much more precise.

7-24 No. It is based on the process of estimation and the expected results if the sampling process is repeated many times.

7-25 a) $\bar{x} \pm .84\sigma_{\bar{x}}$ b) $\bar{x} \pm 1.04\sigma_{\bar{x}}$ c) $\bar{x} \pm 1.75\sigma_{\bar{x}}$ d) $\bar{x} \pm 2.05\sigma_{\bar{x}}$

7-26 a) $\sigma = 5/2 \qquad x = 25$
$x \pm 1.96\sigma = 25 \pm 1.96(5/2) = 25 \pm 4.9 = (20.1, 29.9)$ minutes

b) $\sigma = 5/3 \qquad x = 15$
$x \pm 1.96\sigma = 15 \pm 1.96(5/3) = 15 \pm 3.267 = (11.73, 18.27)$ minutes

c) $\sigma = 5/5 \qquad x = 38$
$x \pm 1.96\sigma = 38 \pm 1.96(1) = (36.04, 39.96)$ minutes

d) $\sigma = 5/1 \qquad x = 20$
$x \pm 1.96\sigma = 20 \pm 1.96(5) = 20 \pm 9.8 = (10.2, 29.8)$ minutes

A confidence interval will contain the true population mean with the indicated percent confidence. A probability interval is one in which the value of the next observation will fall with the indicated probability.

7-27 $\sigma = 30 \qquad n = 40 \qquad \bar{x} = 1416$

a) $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 30/\sqrt{40} = 4.743$

b) $\bar{x} \pm 1.64\sigma_{\bar{x}} = 1416 \pm 1.64(4.743) = 1416 \pm 7.779 = (1408.22, 1423.78)$ hours

7-28 $\sigma = 13.7 \qquad n = 250 \qquad \bar{x} = 112.4 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 13.7/\sqrt{250} = 0.866$

a) $\bar{x} \pm 1.96\sigma_{\bar{x}} = 112.4 \pm 1.96(.866) = 112.4 \pm 1.697 = (110.70, 114.10)$

b) $\bar{x} \pm 2.58\sigma_{\bar{x}} = 112.4 \pm 2.58(.866) = 112.4 \pm 2.234 = (110.17, 114.63)$

7-29 $s = 2.3 \qquad n = 48 \qquad N = 430 \qquad \bar{x} = 64.5$

a) $\hat{\sigma}_{\bar{x}} = \dfrac{s}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}} = \dfrac{2.3}{\sqrt{48}}\sqrt{\dfrac{430-48}{430-1}} = 0.313$ inches

b) $\bar{x} \pm 1.645\hat{\sigma}_{\bar{x}} = 64.5 \pm 1.645(0.313)$
$= 4.3 \pm 0.515 = (63.985, 65.015)$ inches

7-30 $s = 1.2 \quad n = 40 \quad N = 700 \quad \bar{x} = 4.3$

a) $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1.2}{\sqrt{40}}\sqrt{\frac{700-40}{700-1}} = 0.184$

b) $\bar{x} \pm 1.64\hat{\sigma}_{\bar{x}} = 4.3 \pm 1.64(0.184) = 4.3 \pm 0.302 = (4.00, 4.60)$ typos per page.

7-31 $s = 1.8 \quad n = 84 \quad \bar{x} = 11.6$
($N = 95,000$, so we ignore the finite population multiplier)

a) $\hat{\sigma} = s = 1.8$ hours

b) $\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 1.8/\sqrt{84} = 0.196$

c) $\bar{x} \pm 2.33\hat{\sigma}_{\bar{x}} = 11.6 \pm 2.33(0.196)$
$= 11.6 \pm 0.457 = (11.143, 12.057)$ hours

7-32 $s = 3.2 \quad n = 45 \quad \bar{x} = 24.3 \quad \hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 3.2/\sqrt{45} = 0.477$

$\bar{x} \pm 1.96\hat{\sigma}_{\bar{x}} = 24.3 \pm 1.96(0.477) = 24.3 \pm .935 = (23.37, 25.24)$ minutes

7-33 $s = 30 \quad N = 2500 \quad n = 42 \quad \bar{x} = 24.3 \quad \frac{n}{N} = \frac{42}{2500} = .017 < .05$

a) $\hat{\sigma} = s = 30$ oranges

b) $\hat{\sigma}_{\bar{x}} = \hat{\sigma}/\sqrt{n} = 30/\sqrt{42} = 4.629$ oranges

c) $\bar{x} \pm 2.33\hat{\sigma}_{\bar{x}} = 525 \pm 2.33(4.629) = 525 \pm 10.79 = (514.21, 535.79)$ oranges

d) He can be more than 98% confident that the yield now is lower than 5 years ago, which strongly suggests frost damage to the trees.

7-34 . $\hat{\sigma} = 41000 \quad N = 12368 \quad n = 750 \quad \bar{x} = 250000 \quad \frac{n}{N} = \frac{750}{12368} = .061 > .05$

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{41000}{\sqrt{750}}\sqrt{\frac{12368-750}{12368-1}} = 1451.1$$

$\bar{x} \pm 1.64\hat{\sigma}_{\bar{x}} = 250000 \pm 1.64(1451.1) = 250000 \pm 2380 = (\$247{,}620; \$252{,}380)$

7-35 $n = 200 \quad \bar{p} = ..05$

a) $\hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{.05(.95)}{200}} = .0154$

b) $\bar{p} \pm 2.33\hat{\sigma}_{\bar{p}} = .05 \pm 2.33(.0154) = .5 \pm .0359 = (.014, .086)$

7-36 $n = 55 \quad \bar{p} = .1818$

a) $\hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{.1818(.8182)}{55}} = .0520$

b) $\bar{p} \pm 1.645\hat{\sigma}_{\bar{p}} = .1818 \pm 1.645(.0520) = .1818 \pm .0855 = (.096, .267)$

7-37 $n = 1500 \quad \bar{p} = 956/1500 = .637$

a) $\hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{.637(.363)}{1500}} = .0124$

b) $\bar{p} \pm 2.05\hat{\sigma}_{\bar{p}} = .637 \pm 2.05(.0124) = .637 \pm .0254 = (.612, .662)$

7-38 $n = 200$ $\quad \overline{p} = 174/200 = .87$

a) $\hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}} = \sqrt{\frac{.87(.13)}{200}} = .0238$

b) $\overline{p} \pm 2.33\hat{\sigma}_{\overline{p}} = .87 \pm 2.33(.0238) = .87 \pm .0555 = (.815, .925)$

7-39 $n = 95$ $\quad \overline{p} = .8$

a) $\hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}} = \sqrt{\frac{.8(.2)}{95}} = .0410$

b) $\overline{p} \pm 1.96\hat{\sigma}_{\overline{p}} = .8 \pm 1.96(.0410) = .8 \pm .0804 = (.720, .880)$

7-40 $N = 3000$ $\quad n = 150$ $\quad \frac{n}{N} = \frac{150}{3000} = .05$ $\quad \overline{p} = .6$

$\hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{.6(.4)}{150}}\sqrt{\frac{3000-150}{3000-1}} = .0390$

a) $\overline{p} \pm 1.96\hat{\sigma}_{\overline{p}} = .6 \pm 1.96(.0390) = .6 \pm .0764 = (.524, .676)$

b) $3000 \times (.524, .676) = (1572, 2028)$ accounts

7-41 $N = 1500$ $\quad n = 95$ $\quad \frac{n}{N} = \frac{95}{1500} = .0633$ $\quad \overline{p} = .31$

$\hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{.31(.69)}{95}}\sqrt{\frac{1500-95}{1500-1}} = .0459$

$\overline{p} \pm 2.33\hat{\sigma}_{\overline{p}} = .3 \pm 2.33(.0459) = .31 \pm .1069 = (.203, .417)$

7-42 $n = 45$ $\quad \overline{p} = .6$ $\quad \hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}} = \sqrt{\frac{.6(.4)}{45}} = .0730$

$\overline{p} \pm 2.05\hat{\sigma}_{\overline{p}} = .6 \pm 2.05(.0730) = .6 \pm .1497 = (.450, .750)$

7-43 $n = 800$ $\quad \overline{p} = .25$ $\quad \hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}} = \sqrt{\frac{.25(.75)}{800}} = .0153$

$\overline{p} \pm 1.64\hat{\sigma}_{\overline{p}} = .25 \pm 1.64(.0153) = .25 \pm .0251 = (.225, .275)$

7-44 a) 1.761 b) 2.571 c) 2.878 d) 2.492 e) 3.250 f) 2.704

7-45 a) 95% b) 90% c) 99%

7-46 $s = 10$ $\quad n = 12$ $\quad \overline{x} = 62$ $\quad \hat{\sigma}_{\overline{x}} = s/\sqrt{n} = 10/\sqrt{12} = 2.887$

$\overline{x} \pm t_{11,.025}\hat{\sigma}_{\overline{x}} = 62 \pm 2.201(2.887) = 62 \pm 6.354 = (55.65, 68.35)$

7-47 $n = 8$ $\quad \sum x = 649.3$ $\quad \sum x^2 = 52884.59$

a) $\overline{x} = \frac{\sum x}{n} = \frac{649.3}{8} = 81.1625$

b) $s^2 = \left(\frac{1}{n-1}\right)(\sum x^2 - n\overline{x}^2) = \frac{1}{7}\left(52884.59 - 8(81.1625)^2\right) = 26.5398,$

$\hat{\sigma} = s = \sqrt{26.5398} = 5.1517$

c) $\overline{x} \pm t_{7,.01}\hat{\sigma}_{\overline{x}} = 81.1625 \pm 2.998(5.1517/\sqrt{8}) = 81.1625 \pm 5.4606 = (75.702, 86.623)$

7-48 $s = 6.2 \qquad n = 21 \qquad \bar{x} = 72 \qquad \hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 6.2/\sqrt{21} = 1.353$

$\bar{x} \pm t_{20,.01}\hat{\sigma}_{\bar{x}} = 72 \pm 2.528(1.353) = 72 \pm 3.420 = (68.58, 75.42)$

7-49 $s = 0.42 \qquad n = 12 \qquad \bar{x} = 3.6 \qquad \hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 0.42/\sqrt{12} = 0.121$

$\bar{x} \pm t_{11,.05}\hat{\sigma}_{\bar{x}} = 3.6 \pm 1.796(0.121)$

$= 3.6 \pm 0.217 = (3.383, 3.817)$ errors

7-50 $s = 9 \qquad n = 9 \qquad \bar{x} = 31 \qquad \hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 9/\sqrt{9} = 3$

$\bar{x} \pm t_{3,.05}\hat{\sigma}_{\bar{x}} = 31 \pm 1.860(3) = 31 \pm 5.58 = (25.42, 36.58)$ accidents

7-51 $50 = 1.96\sigma/\sqrt{n} = 1.96(78)/\sqrt{n}$, so $n = \left(\frac{(1.96)78}{50}\right)^2 = 9.35$, i.e., $n \geq 10$

7-52 $p = .7 \qquad q = .3$

$.02 = 1.64\sqrt{\frac{pq}{n}} = 1.64\sqrt{\frac{.7(.3)}{n}}$, so $n = \left(\frac{1.64\sqrt{.7(.3)}}{.02}\right)^2 = 1412.04$, i.e., $n \geq 1413$

7-53 $.50 = 2.57\sigma/\sqrt{n} = 2.57(8.6)/\sqrt{n}$, so $n = \left(\frac{2.57(8.6)}{.50}\right)^2 = 1953.99$, i.e., $n \geq 1954$

7-54 $p = .5 \qquad q = .5$

$.05 = 1.96\sqrt{\frac{pq}{n}} = 1.96\sqrt{\frac{.5(.5)}{n}}$, so $n = \left(\frac{1.96\sqrt{.5(.5)}}{.05}\right)^2 = 384.16$, i.e., $n \geq 385$

With $p = .75$ (or $p = .25$), $n = \left(\frac{1.96\sqrt{.25(.75)}}{.05}\right)^2 = 288.12$, i.e., $n \geq 289$

7-55 $6\sigma = 550$, so $\sigma = 550/6 = 91.67$

$30 = \frac{2.33\sigma}{\sqrt{n}} = \frac{2.33(91.67)}{\sqrt{n}}$, so $n = \left(\frac{2.33(91.67)}{30}\right)^2 = 50.69$, i.e., $n \geq 51$

7-56 $.5 = 1.96\sigma/\sqrt{n} = 1.96(1.2)/\sqrt{n}$, so $n = \left(\frac{1.96(1.2)}{.5}\right)^2 = 22.1$, i.e., $n \geq 23$

7-57 Assume $p = q = .5$

$.02 = 1.645\sqrt{\frac{pq}{n}} = 1.645\sqrt{\frac{.5(.5)}{n}}$, so $n = \left(\frac{1.645(.5)}{.02}\right)^2 = 1691.27$, i.e., $n \geq 1692$

So take a sample of at least 1692 students.

7-58 $4 = 2.05\sigma/\sqrt{n} = 2.05(15)/\sqrt{n}$, *so* $n = \left(\frac{2.05(15)}{4}\right)^2 = 59.09$, i.e., $n \geq 60$

7-59 $s = .04 \qquad n = 42 \qquad \bar{x} = 1.12 \qquad \hat{\sigma}_{\bar{x}} = s/\sqrt{n} = .04/\sqrt{42} = .00617$

$\bar{x} \pm 3\hat{\sigma}_{\bar{x}} = 1.12 \pm 3(.00617) = 1.12 \pm .0185 = (\$1.102, \$1.139)$

7-60 An interval estimate gives an indication of possible error through the extent of its range and through the probability associated with the interval. A point estimate is only a single number, and thus one needs additional information to determine its reliability.

7-61 The size of a statistic's standard error is important in indicating the statistic's reliability as an estimator of a population parameter. The smaller the standard error, the more likely it is that the statistic falls close to the true value of the population parameter. This relates to the efficiency of an estimator.

7-62 Assume $p = q = .5$

$.01 = 1.96\sqrt{\frac{pq}{n}} = 1.96\sqrt{\frac{.5(.5)}{n}}$, so $n = \left(\frac{1.96(.5)}{.01}\right)^2 = 9604$

7-63 The 95% confidence interval has a higher probability of including the true mean, but it is also over 12 points wider, which makes it less precise. The 75% confidence interval narrows down the range of the estimate, but does not have as great a probability of including the true mean value.

7-64 $s = .3$ $\quad n = 57$ $\quad \bar{x} = 23.2$

a) $\hat{\sigma} = s = .3$ mph

b) $\hat{\sigma}_{\bar{x}} = \hat{\sigma}/\sqrt{n} = .3/\sqrt{57} = .0397$ mph

c) $\bar{x} \pm 1.96\hat{\sigma}_{\bar{x}} = 23.2 \pm 1.96(.0397) = 23.2 \pm .0778 = (23.122, 23.278)$ mph

7-65 $\sigma = 2.6$ $\quad n = 32$ $\quad \bar{x} = 8$

a) $\frac{8.3864 - 7.6136}{2} = .3864 = \frac{z\sigma}{\sqrt{n}} = \frac{2.6z}{\sqrt{32}}$, so $z = \frac{.3864\sqrt{32}}{2.6} = 0.84$

$P(\mid z \mid \leq 0.84) = 2(.2995) = 0.5990$, so it is a 60% confidence interval.

b) $\frac{9.15 - 6.85}{2} = 1.15 = \frac{z\sigma}{\sqrt{n}} = \frac{2.6z}{\sqrt{32}}$, so $z = \frac{1.15\sqrt{32}}{2.6} = 2.50$

$P(\mid z \mid \leq 2.50) = 2(.4938) = 0.9876$, so it is a 98.76% confidence interval.

c) $\frac{8.805 - 7.195}{2} = 0.805 = \frac{z\sigma}{\sqrt{n}} = \frac{2.6z}{\sqrt{32}}$, so $z = \frac{0.805\sqrt{32}}{2.6} = .1.75$

$P(\mid z \mid \leq 1.75) = 2(.4599) = 0.9198$, so it is a 92% confidence interval.

7-66 $\bar{x}$ is called the "best" estimator of the population mean because it exhibits all the qualities of good estimators we have discussed. It is unbiased, consistent, relatively efficient, and sufficient.

7-67 $s = .0130$ $\quad n = 41$ $\quad \bar{x} = .032$

a) $\hat{\mu} = \bar{x} = .032$

b) $\hat{\sigma} = s = .0130$

c) $\bar{x} \pm 1.64\hat{\sigma}_{\bar{x}} = .032 \pm 1.64(.0130)/\sqrt{41} = .032 \pm .0033 = (.0287, .0353)$

7-68 a) $2(.3944) = .7888$, i.e., 78.88%

b) $2(.4918) = .9836$, i.e., 98.36%

c) $2(.4535) = .9070$, i.e., 90.70%

7-69 $.3 = 2.33\sigma/\sqrt{n} = 2.33(1.1)/\sqrt{n}$, so $n = \left(\frac{2.33(1.1)}{.3}\right)^2 = 72.99$, i.e., $n \geq 73$

7-70 $p = .85$

$.05 = 2.33\sqrt{\frac{pq}{n}} = 2.33\sqrt{\frac{.85(.15)}{n}}$, so $n = \left(\frac{2.33\sqrt{.85(.15)}}{.05}\right)^2 = 276.87$, i.e., $n \geq 277$

To be completely conservative, assume $p = .5$, so $n = \left(\frac{2.33\sqrt{.5(.5)}}{.05}\right)^2 = 542.89$, i.e., $n \geq 543$.

7-71 Using a calculator, $\bar{x} = 57.741$ and $s = 4.110$

$n = 27 \quad \hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 4.11\sqrt{27} = 0.791$

$\bar{x} \pm t_{26,.025}\hat{\sigma}_{\bar{x}} = 57.741 \pm 2.056(0.791)$
$= 57.741 \pm 1.626 = (56.11,\ 59.37)$ patrons

7-72 $n = 27 \quad \bar{p} = 19/27 = 0.7037$

$\hat{\sigma}_{\bar{p}} = \sqrt{pq/n} = \sqrt{.7037(.2963)/27} = 0.791$

$\bar{p} \pm 1.96\hat{\sigma}_{\bar{p}} = 0.7037 \pm 1.96(0.0879)$
$= 0.7037 \pm 0.1723 = (0.5314,\ 0.8760)$

Since the entire interval is above 0.50, they can be more than 95% confident of breaking even at least half of the time. They should stay open on weeknights.

7-73 a) $\bar{x} = .0020$ (calculated in a LOTUS 1-2-3 spreadsheet)

b) $s = .0552$ (calculated in a LOTUS 1-2-3 spreadsheet)

c) $n = 35, \quad \hat{\sigma}_{\bar{x}} = s/\sqrt{n} = .0552/\sqrt{35}$

$\bar{x} \pm 1.96\ \hat{\sigma}_{\bar{x}} = -.0020 \pm 1.96(.0552)$
$= -.0020 \pm .1082 = (-.0133,\ .0231)$ dollars

Since $n > 30$, no assumption about the shape of the distribution is necessary.

7-74 a) $s = 3.56\%$ (calculated in a LOTUS 1-2-3 spreadsheet)

b) $.5 = 2.58\sigma/\sqrt{n} = 2.58(3.56)/\sqrt{n}$, so $n = (2.58(3.56)/.5)^2 = 337.44$, i.e., $n \geq 338$.

7-75 $n = 35, \quad \bar{p} = 19/35 = .543, \quad \hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}\bar{q}}{n}} = \sqrt{\frac{.543(.457)}{35}} = .0842$

$p \pm 2.33\hat{\sigma}_{\bar{p}} = .543 \pm 2.33(.0842) = .543 \pm .1962 = (.347,\ .739)$

7-76 $n = 19, \quad \bar{x} = 2.88, \quad s = 3.30$ (calculated in a LOTUS 1-2-3 spreadsheet)

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 3.30/\sqrt{19} = .7571$

$x \pm t_{18,.025}\hat{\sigma}_{\bar{x}} = 2.88 \pm 2.101(.7571) = 2.88 \pm 1.591 = (1.29\%,\ 4.47\%)$

Since n < 30, we must assume normality so that the t distribution can be used to form the confidence interval.

7-77 $n = 52 \qquad N = 900 \qquad \bar{p} = .35 \qquad \frac{n}{N} = \frac{52}{900} = .058 > .05$

a) $\hat{\sigma}_{\bar{p}} = \sqrt{\frac{\bar{p}\bar{q}}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{.35(.65)}{52}}\sqrt{\frac{900-52}{900-1}} = .0642$

b) $\bar{p} \pm 1.64\hat{\sigma}_{\bar{p}} = .35 \pm 1.64(.0642) = .35 \pm .1053 = (.245,\ .455)$

7-78 $\sigma = .2 \qquad n = 105 \qquad \bar{x} = 3.2$

a) $\sigma_{\bar{x}} = \sigma/\sqrt{n} = .2/\sqrt{105} = .0195$ apples

b) $\bar{x} \pm \sigma_{\bar{x}} = 3.2 \pm .0195 = (3.181,\ 3.219)$ apples

7-79 $\sigma = 1.2 \qquad n = 60 \qquad \bar{x} = 4.1$

a) $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.2/\sqrt{60} = .1549$ passengers

b) $\bar{x} \pm 1.96\sigma_{\bar{x}} = 4.1 \pm 1.96(.1549) = 4.1 \pm .3036 = (3.796,\ 4.404)$ passengers

7-80 $s = 107.10 \qquad n = 200 \qquad \bar{x} = 425.39$

a) $\hat{\mu} = \bar{x} = \$425.39$
$\hat{\sigma} = s = \$107.10$

b) $\bar{x} \pm 1.96\hat{\sigma}_{\bar{x}} = .032 \pm 1.96\hat{\sigma}/\sqrt{n} \quad = 425.39 \pm 1.96(107.10)/\sqrt{200}$
$= 425.39 \pm 14.843 = (\$410.55,\ \$440.23)$

7-81 $\sigma = 2.5 \qquad n = 49 \qquad \bar{x} = 15.2 \qquad \sigma_{\bar{x}} = \sigma/\sqrt{n} = 2.5/\sqrt{49} = .3571$ minutes

a) $\bar{x} \pm 1.64\sigma_{\bar{x}} = 15.2 \pm 1.64(.3571) = 15.2 \pm .586 = (14.61,\ 15.79)$ minutes

b) $\bar{x} \pm 2.57\sigma_{\bar{x}} = 15.2 \pm 2.57(.3571) = 15.2 \pm .918 = (14.28,\ 16.12)$ minutes

7-82 $\sigma = 1.4 \qquad n = 200 \qquad \bar{x} = 4.6$

a) $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.4/\sqrt{200} = .0990$

b) $\mu \pm \sigma_{\bar{x}} = 5.2 \pm .099 = (5.10,\ 5.30)$

7-83 $n = 7 \qquad \sum x = 17 \qquad \sum x^2 = 43.56$

a) $\bar{x} = \frac{\sum x}{n} = \frac{17}{7} = 2.4286$ minutes

$s^2 = \left(\frac{1}{n-1}\right)\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{6}\left(43.56 - 7(2.4286)^2\right) = .3789$

$s = \sqrt{.3789} = .6155$ minutes

b) $\hat{\sigma} = s = .6155$ minutes

c) $\bar{x} \pm t_{6,.01}\hat{\sigma}_{\bar{x}} = \bar{x} \pm 3.143\hat{\sigma}/\sqrt{n} \quad = 2.4286 \pm 3.143(.6155)/\sqrt{7}$
$= 2.4286 \pm .7312 = (1.697,\ 3.160)$ minutes

7-84 $n = 120 \qquad \overline{p} = .3333 \qquad \hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}} = \sqrt{\frac{.3333(.6667)}{120}} = .0430$

$\overline{p} \pm 1.96\hat{\sigma}_{\overline{p}} = .3333 \pm 1.96(.0430) = .3333 \pm .0843 = (.249, .418)$

7-85 $n = 90 \qquad \overline{p} = 79/90 = .8778 \qquad \hat{\sigma}_{\overline{p}} = \sqrt{\frac{\overline{p}\overline{q}}{n}} = \sqrt{\frac{.8778(.1222)}{90}} = .0345$

$\overline{p} \pm 2.33\hat{\sigma}_{\overline{p}} = .8778 \pm 2.33(.0345) = .8778 \pm .0804 = (.797, .958)$

7-86 $s = .6 \qquad n = 186 \qquad \overline{x} = 66.3$

a) $\hat{\sigma}_{\overline{x}} = s/\sqrt{n} = .6/\sqrt{186} = .0440$ mph

b) $\overline{x} \pm 2\hat{\sigma}_{\overline{x}} = 66.3 \pm 2(.0440) = 66.3 \pm .088 = (66.212, 66.388)$ mph

c) Yes, since the entire interval lies below 67 mph

7-87 $s = 3.6 \qquad n = 9 \qquad \overline{x} = 18.3 \qquad \hat{\sigma}_{\overline{x}} = s/\sqrt{n} = 3.6/\sqrt{9} = 1.20$

$\overline{x} \pm t_{8,\,.025}\hat{\sigma}_{\overline{x}} = 18.3 \pm 2.306(1.2) = 18.3 \pm 2.767 = (\$15.53, \$21.07)$

7-88 $6\sigma = 140 - 80 = 60$, so $\sigma = 10$

$5 = 1.64\sigma/\sqrt{n} = 1.64(10)/\sqrt{n}$, so $n = \left(\frac{1.64(10)}{5}\right)^2 = 10.75$, i.e., $n \geq 11$

CHAPTER 8

TESTING HYPOTHESES: ONE-SAMPLE TESTS

8-1 We must deal with uncertainty in our decisions because we rarely know the values for population characteristics. In addition, there are other circumstances unknown to us in any given situation. Statistical analysis can reduce uncertainty, but not eliminate it entirely.

8-2 Theoretically, one could toss a coin a large number of times and see if the proportion of heads was very different from .5. Similarly by recording the outcomes of many dice rolls, one could see if the proportion of each side was very different from 1/6. You would need a large number of trials for each of these samples.

8-3 Yes, it is possible that a false hypothesis will be accepted. If the observed value does not differ enough from the hypothesized one, the hypothesis will be accepted. Acceptance of a hypothesis is based on probability and therefore we can never be absolutely certain that our decision is correct.

8-4
a) Assume hypothesis about population
b) Collect sample data
c) Calculate a sample statistic
d) Use sample statistic (c) to evaluate hypothesis (a)

8-5 There is always a statistical possibility that a sample does <u>not</u> accurately represent the population from which it has been drawn.

8-6 We mean that we would not have reasonably expected to find that particular sample if in fact the hypothesis had been true.

8-7 The level of probability, or certainty, depends upon how accurate our answer needs to be.

8-8 $P(|z| \geq 1.75) = 2(.5 - .4599) = .0802$

8-9 The z value which leaves $(100-98)\% = 2\%$ in the tails is ± 2.33, so the interval should be the hypothesized value ± 2.33 standard errors.

8-10 $\sigma = 6000 \qquad n = 64 \qquad \bar{x} = 26100 \qquad \mu_{H_0} = 28500$

$$\mu \pm 2\sigma_{\bar{x}} = \mu \pm 2\sigma/\sqrt{n} = 28500 \pm 2(6000)/\sqrt{64}$$
$$= 28500 \pm 1500 = (27000, 30000)$$

Since $\bar{x} = 26100 < 27000$, Ned should not purchase the Stalwarts. Depending on how long ago Ned's last purchase was, it may no longer be reasonable to suppose that $\sigma = 6000$. If σ has increased sufficiently, Ned's decision could change.

8-11 $\sigma = 12.6$ $n = 81$ $\bar{x} = 27.2$ $\mu_{H_0} = 23.9$

$\mu \pm 2\sigma_{\bar{x}} = \mu \pm 2\sigma/\sqrt{n} = 23.9 \pm 2(12.6)/\sqrt{81} = 23.9 + 2.8 = (21.1, 26.7)$

Since $\bar{x} > 26.7$, it is reasonable to conclude that Computer World's subscribers are different from average personal computer owners.

8-12 $\sigma = 0.2$ $n = 42$ $\bar{x} = 2.2$ $\mu_{H_0} = 2.5$

$\mu \pm 2.5\sigma_{\bar{x}} = \mu \pm 2.5\sigma/\sqrt{n} = 2.5 \pm 2.5(0.2)/\sqrt{42} = 2.5 + 0.077 = (2.423, 2.577)$

Since $\bar{x} = 2.2 < 2.423$, it is unreasonable to see such sample results if μ really is 2.5 quarts; the store's claim is not correct.

8-13 $H_0 : \mu = 45$ $H_1 : \mu > 45$

8-14 A null hypothesis represents the hypothesis you are trying to reject. Alternative hypotheses represent all other possibilities.

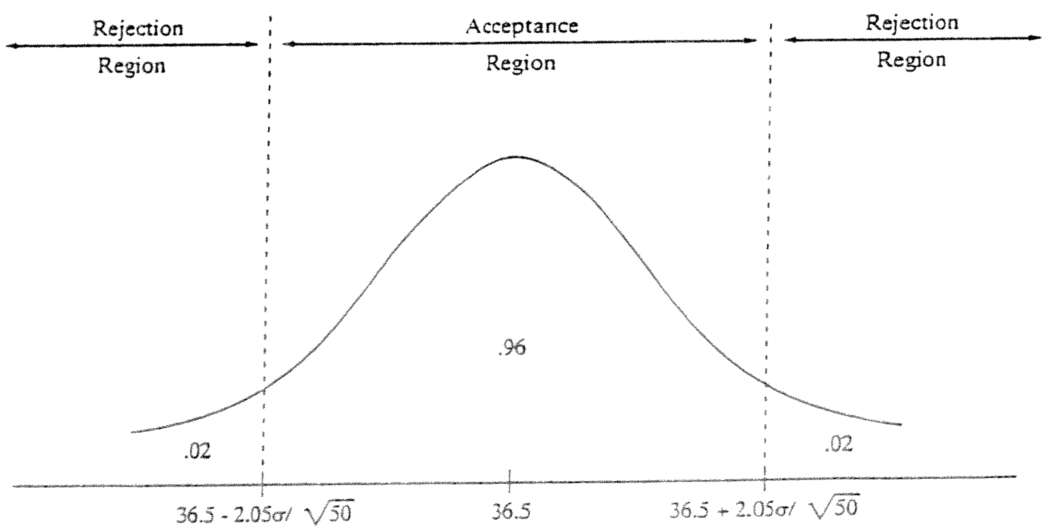
8-15 The significance level refers to the probability that the sample statistic lies outside a specified interval around the hypothesized parameter, assuming the hypothesis is true. Or, the significance level represents the risk of rejecting the null hypothesis when it is true.

8-16 Type I: the probability that we will reject the null hypothesis when in fact it is true.
Type II: the probability that we will accept the null hypothesis when in fact it is false.

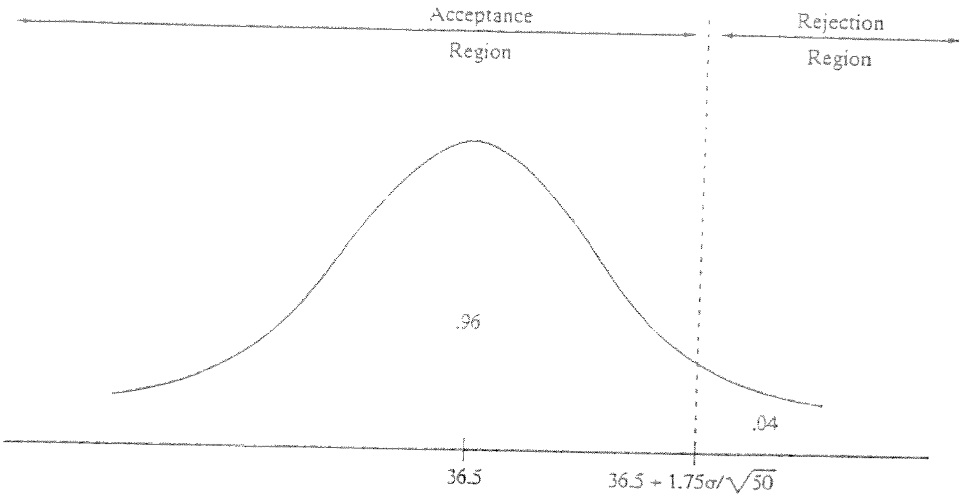
8-17 We would prefer to commit a Type II error and let a guilty person go free than to sentence an innocent individual for a crime he or she didn't commit.

8-18 The significance level of a test is the probability of a Type I error, i.e., the probability that we will reject the null hypothesis when it is, in fact, true. This is because it indicates the percentage of sample means that fall outside the limits of what we will accept as confirming the null hypothesis. Hence, if a sample mean falls outside these limits but is truly from the hypothesized population, it will lead to a Type I error.

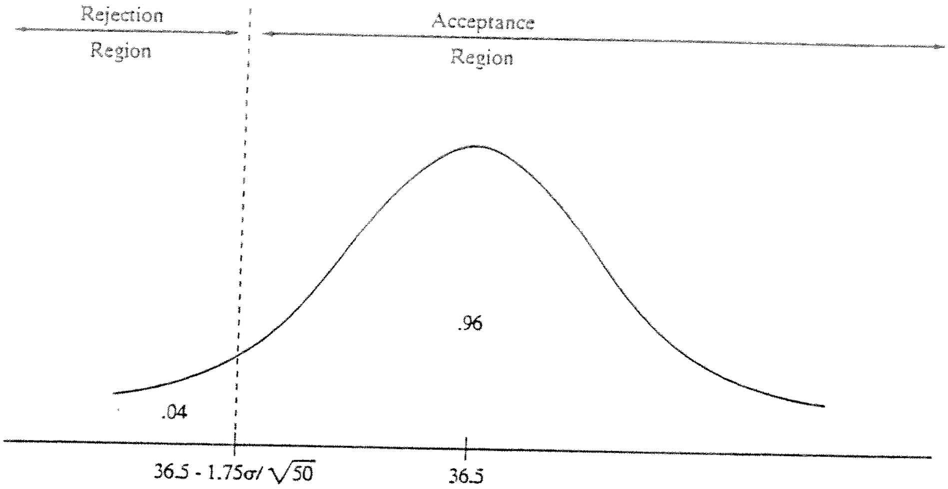
8-19 a)



b)



c)



8-20 a) t with 34 df (so we use the normal table) b) normal c) normal

d) t with 28 df e) t with 23 df

8-21 a) A Type I error (rejecting H_0 when it is true) will release a bad battery for sale, which could have very serious consequences for the patient who buys it. On the other hand, a Type II error (accepting H_0 when it is false) will cause the company to discard a perfectly good battery, which presumably does not cost very much to replace. Thus it would appear that the engineer would be reluctant to make Type I errors, but should not be too concerned about making Type II errors.

b) A low significance level (i.e., a small value of α) would be appropriate.

8-22 A two-tailed test of a hypothesis will reject the null hypothesis if the sample mean is significantly higher <u>or</u> lower than the hypothesized population mean. Thus a two-tailed test is appropriate when we are testing whether the population mean is different from some hypothesized value. A one-tailed test, on the other hand, would be used when we are testing whether the population mean is <u>lower than</u> or <u>higher than</u> some hypothesized value.

8-23 We would use a lower-tailed test if the hypotheses to be tested are:

$H_0: \mu =$ some stipulated value

$H_1: \mu <$ some stipulated value

We would use an upper-tailed test if the hypotheses to be tested are:

$H_0: \mu =$ some stipulated value

$H_1: \mu >$ some stipulated value

8-24 They should perform a lower-tailed test, with H_0: $\mu = 3124$, H_1: $\mu < 3124$.

8-25 a) $H_0: \mu = 78$ $H_1: \mu > 78$ b) $H_0: \mu = 15$ $H_1: \mu > 15$

8-26 $\sigma = 5.75$ $n = 25$ $\bar{x} = 42.95$

$H_0: \mu = 44.95$ $H_1: \mu < 44.95$ $\alpha = .02$

The lower limit of the acceptance region is $z_L = -2.05$, or

$$\bar{x}_L = \mu - z_{.02}\sigma/\sqrt{n} = 44.95 - 2.05(5.75)/\sqrt{25} = \$42.59$$

Since $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{42.95-44.95}{5.75/\sqrt{25}} = -1.74 > -2.05$ (or $\bar{x} > 42.59$), we cannot reject H_0.

Atlas should not believe that the average retail price has decreased.

8-27 $\sigma = 9.73$ $n = 30$ $\bar{x} = 11.77$

$H_0: \mu = 14.35$ $H_1: \mu \neq 14.35$ $\alpha = .05$

The limits of the acceptance region are $z_{CRIT} = \pm 1.96$, or

$$\bar{x}_{CRIT} = \mu \pm z_{.025}\sigma/\sqrt{n} = 14.35 \pm 1.96(9.73)/\sqrt{30} = (10.87,\ 17.85)$$

Since $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{11.77-14.35}{9.73/\sqrt{30}} = -1.45$, it and $\bar{x}$ are in the acceptance region, so we do not reject H_0. The mean P/E ratio in 1986 is not significantly different from its previous value.

8-28 $\sigma = 18.4$ $n = 20$ $\bar{x} = 954$

$H_0: \mu = 960$ $H_1: \mu \neq 960$ $\alpha = .05$

The limits of the acceptance region are $z_{\text{CRIT}} = \pm 1.96$, or

$$\bar{x}_{\text{CRIT}} = \mu \pm z_{.025}\sigma/\sqrt{n} = 960 \pm 1.96(18.4)/\sqrt{20} = (951.94,\ 968.06) \text{ lumens}$$

Since $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{954-960}{18.4/\sqrt{20}} = -1.46 > -1.96$ (or $\bar{x} > 951.94$), we do not reject H_0.

The average light output is not significantly different from the hypothesized value.

8-29 $\sigma = 72.6$ $n = 30$ $\bar{x} = 912.1$

$H_0: \mu = 984.7$ $H_1: \mu < 984.7$ $\alpha = .02$

The lower limit of the acceptance region is $z_L = -2.05$, or

$$\bar{x}_L = \mu - z_{.002}\sigma/\sqrt{n} = 984.7 - 2.05(72.6)/\sqrt{30} = 957.53 \text{ pounds}$$

Since $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{912.1-984.7}{72.6/\sqrt{30}} = -5.48$, it and $\bar{x}$ are in the rejection region, so we reject H_0. Hot chocolate sales have decreased significantly.

8-30 $\sigma = 52$ $n = 121$ $\bar{x} = 151$

$H_0: \mu = 144$ $H_1: \mu > 144$ $\alpha = .10$

The upper limit of the acceptance region is $z_U = 1.28$, or

$$\bar{x}_U = \mu + z_{.10}\sigma/\sqrt{n} = 144 + 1.28(52)/\sqrt{121} = \$150$$

Since $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{151-144}{52/\sqrt{121}} = 1.48 > 1.28$, (or $\bar{x} > 150$), we should reject H_0. Joel's clients' average commission is significantly higher than the industry average.

8-31 $\sigma = 16$ $n = 64$ $\bar{x} = 30.3$

$H_0: \mu = 28$ $H_1: \mu > 28$ $\alpha = .05$

The upper limit of the acceptance region is $z_U = 1.64$, or

$$\bar{x}_U = \mu + z_{.05}\sigma/\sqrt{n} = 28 + 1.64(16)/\sqrt{64} = 31.28 \text{ million dollars}$$

Since $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{30.3-28}{16/\sqrt{64}} = 1.15 < 1.64$ (or $\bar{x} < 31.28$), we do not reject H_0. The Custom's Commissioner need not be concerned that smuggling has increased above its historic level.

8-32 $\sigma = 0.10$ $n = 15$ $\bar{x} = 0.33$

$H_0: \mu = 0.57$ $H_1: \mu < 0.57$ $\alpha = .01$

The lower limit of the acceptance region is $z_L = -2.33$, or

$$\bar{x}_L = \mu - z_{.01}\sigma/\sqrt{n} = 0.57 - 2.33(0.10)/\sqrt{15} = 0.51\%$$

Since $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{0.33-0.57}{0.10/\sqrt{15}} = -9.30 < -2.33$ (or $\bar{x} < 0.51$), we should reject H_0. The rate of growth has decreased significantly, and we infer that this was because of the oil embargo and its consequences.

8-33 $\sigma = .018$ $n = 1$ $\bar{x} = .306$

$H_0: \mu = .343$ $H_1: \mu < .343$ $\alpha = .02$

The lower limit of the acceptance region is $z_L = -2.05$, or

$$\bar{x}_L = \mu - z_{.02}\sigma/\sqrt{n} = .343 - 2.05(.018)/\sqrt{1} = .3061$$

Joe's average last year falls below the acceptance region by .0001 ($z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{.306-.343}{.018/\sqrt{1}} = -2.06$), so strictly speaking, the owner should reject H_0 and cut Joe's salary. However, this is so close that the owner really does have a difficult call.

8-34 From exercise 8-31, we have $\sigma = 16$, $n = 64$, and $\bar{x}_{CRIT} = 31.28$.

a) $P(\bar{x} > 31.28 \mid \mu = 28) = P\left(z > \dfrac{31.28-28}{16/\sqrt{64}}\right) = P(z > 1.64) = .5 - .4495 = .0505$

b) $P(\bar{x} > 31.28 \mid \mu = 29) = P\left(z > \dfrac{31.28-29}{16/\sqrt{64}}\right) = P(z > 1.14) = .5 - .3729 = .1271$

c) $P(\bar{x} > 31.28 \mid \mu = 30) = P\left(z > \dfrac{31.28-30}{16/\sqrt{64}}\right) = P(z > .64) = .5 - .2389 = .2611$

8-35 From exercise 8-30, we have $\sigma = 52$, $n = 121$, and $\bar{x}_U = 150$.

a) $P(\bar{x} > 150 \mid \mu = 140) = P\left(z > \dfrac{150-140}{52/\sqrt{121}}\right) = P(z > 2.12) = .5 - .4830 = .0170$

b) $P(\bar{x} > 150 \mid \mu = 160) = P\left(z > \dfrac{150 - 160}{52/\sqrt{121}}\right) = P(z > -2.12) = .5 + .4830 = .9830$

c) $P(\bar{x} > 150 \mid \mu = 175) = P\left(z > \dfrac{150 - 175}{52/\sqrt{121}}\right) = P(z > -5.29) \approx .5 + .5000 = 1.000$

8-36 From exercise 8-31, we have $\sigma = 16$, $n = 64$, $H_0 : \mu = 28$, $H_1 : \mu > 28$
At $\alpha = .02$, the upper limit of the acceptance region is

$$\mu + z_{.02}\sigma/\sqrt{n} = 28 + 2.05(16)/\sqrt{64} = 32.1 \text{ million dollars}$$

a) $P(\bar{x} > 32.1 \mid \mu = 28) = P\left(z > \dfrac{32.1 - 28}{16/\sqrt{64}}\right) = P(z > 2.05) = .5 - .4798 = .0202$

b) $P(\bar{x} > 32.1 \mid \mu = 29) = P\left(z > \dfrac{32.1 - 29}{16/\sqrt{64}}\right) = P(z > 1.55) = .5 - .4394 = .0606$

c) $P(\bar{x} > 32.1 \mid \mu = 30) = P\left(z > \dfrac{32.1 - 30}{16/\sqrt{64}}\right) = P(z > 1.05) = .5 - .3531 = .1469$

8-37 From exercise 8-30, we have $\sigma = 52$, $n = 121$, $H_0 : \mu = 144$, $H_1 : \mu > 144$
At $\alpha = .05$, the upper limit of the acceptance region is

$$\mu + z_{.05}\sigma/\sqrt{n} = 144 + 1.645(52)/\sqrt{121} = \$151.78$$

a) $P(\bar{x} > 151.78 \mid \mu = 140) = P\left(z > \dfrac{151.78 - 140}{52/\sqrt{121}}\right) = P(z > 2.49)$
$= .5 - .4936 = .0064$

b) $P(\bar{x} > 151.78 \mid \mu = 160) = P\left(z > \dfrac{151.78 - 160}{52/\sqrt{121}}\right) = P(z > -1.74)$
$= .5 + .4591 = .9591$

c) $P(\bar{x} > 151.78 \mid \mu = 175) = P\left(z > \dfrac{151.78 - 175}{52/\sqrt{121}}\right) = P(z > -4.91)$
$\approx .5 + .5000 = 1.000$

8-38 $n = 85$ $\bar{p} = .1412$
$H_0 : p = .19$ $H_1 : p < .19$ $\alpha = .04$
The lower limit of the acceptance region is $z_L = -1.75$, or

$$\bar{p}_L = p - z_{.04}\sqrt{\frac{pq}{n}} = .19 - 1.75\sqrt{\frac{.19(.81)}{85}} = .1155$$

Since $z = \dfrac{\bar{p} - p}{\sqrt{pq/n}} = \dfrac{.1412 - .19}{\sqrt{.19(.81)/85}} = -1.15 > -1.75$ (or $\bar{p} > .1155$), we don't reject H_0. Grant's western distribution is not significantly worse than its eastern distribution.

8-39 $n = 350$ $\bar{p} = .39$
$H_0 : p = .41$ $H_1 : p \neq .41$ $\alpha = .02$
The limits of the acceptance region are $z_{CRIT} = \pm 2.33$, or

$$\bar{p}_{CRIT} = p \pm z_{.01}\sqrt{\frac{pq}{n}} = .41 \pm 2.33\sqrt{\frac{.39(.61)}{350}} = (.3493, .4707)$$

Since $z = \dfrac{\bar{p} - p}{\sqrt{pq/n}} = \dfrac{.39 - .41}{\sqrt{.41(.59)/350}} = -0.76 > -2.33$ (or $.3493 < \bar{p} < .4707$), we don't reject H_0. The proportion of loans made to women has not changed significantly.

8-40 a) $n = 180$ $\quad \overline{p} = 17/180 = .0944$
$H_0 : p = .151$ $\quad H_1 : p < .151$ $\quad \alpha = .05$

The lower limit of the acceptance region is $z_L = -1.64$, or

$$\overline{p}_L = p - z_{.05}\sqrt{\frac{pq}{n}} = .151 - 1.64\sqrt{\frac{.151(.849)}{180}} = .1072$$

Since $z = \dfrac{\overline{p} - p}{\sqrt{pq/n}} = \dfrac{.0944 - .151}{\sqrt{.151(.849)/180}} = -2.12 < -1.64$ (or $\overline{p} < .1072$), they should reject H_0. Spray users are significantly less susceptible to colds.

b) At $\alpha = .02$, the lower limit of the acceptance region is $z_L = -2.05$, or

$$\overline{p}_L = p - z_{.02}\sqrt{\frac{pq}{n}} = .151 - 2.05\sqrt{\frac{.151(.849)}{180}} = 0.0963$$

Again $z < -2.05$ ($\overline{p} < .0963$), so the same conclusion holds.

c) Not necessarily. Although the users of the spray seem to be significantly less susceptible to colds, we do not know that other relevant factors have been controlled for in the experiment, nor have we been told anything about potential side-effects of the spray which might counterbalance its effectiveness in reducing susceptibility to colds.

8-41 $n = 175$ $\quad \overline{p} = 101/175 = .5771$
$H_0 : p = .5$ $\quad H_1 : p > .5$ $\quad \alpha = .01$
The upper limit of the acceptance region is $z_U = 2.33$, or

$$\overline{p}_U = p + z_{.01}\sqrt{\frac{pq}{n}} = .5 + 2.33\sqrt{\frac{.5(.5)}{175}} = .5881$$

Since $z = \dfrac{\overline{p} - p}{\sqrt{pq/n}} = \dfrac{.5771 - .5}{\sqrt{.5(.5)/175}} = 2.04 < 2.33$ (or $\overline{p} = .5771 < .5881$), we do not reject H_0.

The data do not provide significant support for the theory.

8-42 $n = 3000$ $\quad \overline{p} = 950/3000 = .3167$
$H_0 : p = .35$ $\quad H_1 : p < .35$ $\quad \alpha = .05$
The lower limit of the acceptance region is $z_L = -1.645$, or

$$\overline{p}_L = p - z_{.05}\sqrt{\frac{pq}{n}} = .35 - 1.645\sqrt{\frac{.35(.65)}{3000}} = .3357$$

Since $z = \dfrac{\overline{p} - p}{\sqrt{pq/n}} = \dfrac{.3167 - .35}{\sqrt{.35(.65)/3000}} = -3.82 < -1.645$ (or $\overline{p} < .3357$), we should reject H_0.

The proportion of skeptical people is significantly less than it was last year.

8-43 $n = 187$ $\quad \overline{p} = 157/187 = .8396$
$H_0 : p = .86$ $\quad H_1 : p \neq .86$ $\quad \alpha = .01$
The limits of the acceptance region are $z_{CRIT} = \pm 2.58$, or

$$\overline{p}_{CRIT} = p \pm z_{.005}\sqrt{\frac{pq}{n}} = .86 \pm 2.58\sqrt{\frac{.86(.14)}{187}} = (.7945, .9255)$$

Since $z = \dfrac{\overline{p} - p}{\sqrt{pq/n}} = \dfrac{.8396 - .86}{\sqrt{.86(.14)/187}} = -.80 > -2.58$ (or $\overline{p} = .8396$), we do not reject H_0.

Rick's claim is valid. The proportion of totally satisfied customers has not changed significantly.

8-44 $s = 8.4$ $\quad n = 6$ $\quad \overline{x} = 94.3$
$H_0 : \mu = 100$ $\quad H_1 : \mu < 100$ $\quad \alpha = .05$

The lower limit of the acceptance region is $t_L = -t_{5,\ .05} = -2.015$, or

$$\bar{x}_L = \mu - t_{5,\ .05}s/\sqrt{n} = 100 - 2.015(8.4)/\sqrt{6} = 93.09$$

Since $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{94.3 - 100}{8.4/\sqrt{6}} = -1.662 > -2.015$ (or $\bar{x} > 93.09$), we do not reject H_0.

8-45 $s^2 = 4.2$ $(s = 2.049)$ $n = 25$ $\bar{x} = 52$
$H_0 : \mu = 65$ $H_1 : \mu \neq 65$ $\alpha = .01$
The limits of the acceptance region are $t_{CRIT} = \pm t_{24,\ .005} = \pm 2.797$, or

$$\bar{x}_{CRIT} = \mu \pm t_{24,\ .005}s/\sqrt{n} = 65 \pm 2.797(2.049)/\sqrt{25} = (63.85,\ 66.15)$$

Since $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{52 - 65}{2.049/\sqrt{25}} = -31.7 < -2.797$ (or $\bar{x} < 63.85$), we reject H_0.

8-46 $s = 49{,}000$ $n = 12$ $\bar{x} = 780{,}000$
$H_0 : \mu = 825{,}000$ $H_1 : \mu < 825{,}000$ $\alpha = .05$
The lower limit of the acceptance region is $t_L = -t_{11,\ .05} = -1.796$

$$\bar{x}_L = \mu - t_{11,\ .05}s/\sqrt{n} = 825{,}000 - 1.796(49{,}000)/\sqrt{12} = \$799{,}595$$

Since $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{780{,}000 - 825{,}000}{49{,}000/\sqrt{12}} = -3.181 < -1.796$ (or $\bar{x} < 799{,}595$), we reject H_0. The average appraised value of homes in the area is significantly less than \$825,000.

8-47 $s = 42$ $n = 60$ $\bar{x} = 101$
$H_0 : \mu = 75$ $H_1 : \mu > 75$ $\alpha = .02$
The upper limit of the acceptance region is $t_U = t_{59,\ .02} = 2.05$, or

$$\bar{x}_U = \mu + t_{59,\ .02}s/\sqrt{n} = 75 + 2.05(42)/\sqrt{60} = 86.12$$

(with 59 *df*, we use the normal table)

Since $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{101 - 75}{42/\sqrt{60}} = 4.795 > 2.05$ (or $\bar{x} > 86.12$), we reject H_0. On average, the women in the program have blood pressures significantly higher than the recommended level.

8-48 $s^2 = 16.2$ $(s = 4.025)$ $n = 95$ $\bar{x} = 7.2$
$H_0 : \mu = 8.1$ $H_1 : \mu < 8.1$ $\alpha = .01$
The lower limit of the acceptance region is $t_L = -t_{94,\ .01} = -2.33$, or

$$\bar{x}_L = \mu - t_{94,\ .01}s/\sqrt{n} = 8.1 - 2.33(4.025)/\sqrt{95} = 7.14 \text{ hours}$$

(with 94 *df*, we use the normal table)

Since $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{7.2 - 8.1}{4.025/\sqrt{95}} = -2.179 > -2.33$ (or $\bar{x} > 7.14$), we do not reject H_0. The new terminals are not significantly easier to learn to operate.

8-49 $s = 4.65$ $n = 13$ $\bar{x} = 21.60$
$H_0 : \mu = 18$ $H_1 : \mu > 18$ $\alpha = .01$
The upper limit of the acceptance region is $t_U = t_{12,\ .01} = 2.681$, or

$$\bar{x}_U = \mu + t_{12,\ .01}s/\sqrt{n} = 18 + 2.681(4.65)/\sqrt{13} = \$21.46$$

Since $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}} = \dfrac{21.60 - 18}{4.65/\sqrt{13}} = 2.791 > 2.681$ (or $\overline{x} > 21.46$), we reject H_0. It appears that a budget crisis was unlikely.

8-50 $s = 2.7$ $n = 18$ $\overline{x} = 12.4$
$H_0 : \mu = 10$ $H_1 : \mu \neq 10$ $\alpha = .01$
The limits of the acceptance region are $t_{crit} = \pm t_{17,\ .005} = \pm 2.898$, or

$$\overline{x}_{CRIT} = \mu \pm t_{17,.005}s/\sqrt{n} = 10 \pm 2.898(2.7)/\sqrt{18} = (8.16,\ 11.84) \text{ pounds}$$

Since $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}} = \dfrac{12.4 - 10}{2.7/\sqrt{18}} = 3.771 > 2.898$ (or $\overline{x} > 11.84$), we reject H_0. The claim does not appear to be valid.

8-51 $s = 41.3$ $n = 16$ $\overline{x} = 5040$
At \$2.50 per widget, there must be more than $12500/2.5 = 5000$ widgets per batch on average to be profitable. Hence our hypotheses are $H_0 : \mu = 5000$ vs. $H_1 : \mu > 5000$. Testing at $\alpha = .025$, the upper limit of the acceptance region is $t_U = t_{15,.025} = 2.131$, or

$$\overline{x}_U = \mu + t_{15,\ .025}s/\sqrt{n} = 5000 + 2.131(41.3)/\sqrt{16} = 5022.0$$

Since $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}} = \dfrac{5040 - 5000}{41.3/\sqrt{16}} = 3.874 > 2.131$ (or $\overline{x} > 5022$), we reject H_0. XCO should conclude that its widget operation is profitable.

8-52 a) Let p_{NY} = proportion of homeless people in New York City
p_{DC} = proportion of homeless people in Washington, D.C.
$H_0 : p_{NY} = p_{DC}$ $H_1 : p_{NY} \neq p_{DC}$

b) Let μ_A = mean sales after the promotion
μ_B = mean sales before the promotion
$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A > \mu_B$

c) Let μ = average annual snowfall in the 1980's
H_0: $\mu = 8$ H_1: $\mu \neq 8$

d) Let μ = average mpg for the model
$H_0 : \mu = 34$ $H_1 : \mu < 34$

8-53 $H_0 : \mu = 28$ $H_1 : \mu < 28$
A Type I error would not be serious for the patient. A particular battery would not be installed even though it would last long enough. A Type II error could be serious. After installation, this battery might run down before the scheduled operation, thus endangering the patient's life.

8-54 No, because each of the sample means is equally distant from the hypothesized mean and, therefore, equally likely to lead to rejection or acceptance of the null hypothesis.

8-55 $s = 1.5$ $n = 23$ $\overline{x} = 4.3$
$H_0 : \mu = 3.4$ $H_1 : \mu > 3.4$ $\alpha = .01$
The upper limit of the acceptance region is $t_U = t_{22,\ .01} = 2.508$, or

$$\overline{x}_U = \mu + t_{22,\ .01}s/\sqrt{n} = 3.4 + 2.508(1.5)/\sqrt{23} = 4.18$$

Since $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}} = \dfrac{4.3 - 3.4}{1.5/\sqrt{23}} = 2.877 > 2.508$ (or $\overline{x} = 4.3 > 4.18$), we should reject H_0. The average number of passengers has increased significantly.

8-56 From exercise 8-28, we have $\sigma = 2100$, $n = 25$, $H_0 : \mu = 14500$, $H_1 : \mu < 14500$
At $\alpha = .10$, the lower limit of the acceptance region is

$$\mu - z_{.10}\sigma/\sqrt{n} = 14500 - 1.28(2100)/\sqrt{25} = 13962.4 \text{ hours}$$

a) $P(\bar{x} < 13962.4 \mid \mu = 14000) = P\left(z < \dfrac{13962.4 - 14000}{2100/\sqrt{25}}\right)$
$= P(z < -0.09) = .5 - .0359 = .4641$

b) $P(\bar{x} < 13962.4 \mid \mu = 13500) = P\left(z < \dfrac{13962.4 - 13500}{2100/\sqrt{25}}\right)$
$= P(z < 1.10) = .5 + .3643 = .8643$

c) $P(\bar{x} < 13962.4 \mid \mu = 13000) = P\left(z < \dfrac{13962.4 - 13000}{2100/\sqrt{25}}\right)$
$= P(z < 2.29) = .5 + .4890 = .9890$

8-57 $n = 120$ $\quad \bar{p} = 16/20 = .1333$
$H_0 : p = .05$ $\quad H_1 : p > .05$ $\quad \alpha = .01$
The upper limit of the acceptance region is $z_U = 2.33$, or

$$\bar{p}_U = p + z_{.01}\sqrt{\frac{pq}{n}} = .05 + 2.33\sqrt{\frac{.05(.95)}{120}} = .0964$$

Since $z = \dfrac{\bar{p} - p}{\sqrt{pq/n}} = \dfrac{.1333 - .05}{\sqrt{.05(.95)/120}} = 4.19 > 2.33$ (or $\bar{p} > .0964$), we reject H_0. Significantly more stocks than usual did set new highs that day.

8-58 $n = 8000$ $\quad \bar{p} = 18/8000 = .00225$
$H_0 : p = .003$ $\quad H_1 : p < .003$ $\quad \alpha = .10$
The lower limit of the acceptance region is $z_L = -1.28$, or

$$\bar{p}_L = p - z_{.10}\sqrt{\frac{pq}{n}} = .003 - 1.28\sqrt{\frac{.003(.997)}{8000}} = .00222$$

Since $z = \dfrac{\bar{p} - p}{\sqrt{pq/n}} = \dfrac{.00225 - .003}{\sqrt{.003(.997)/8000}} = -1.23 > -1.28$ (or $\bar{p} > .00222$), we do not reject H_0.

The new procedures have not significantly reduced the fraction of lost mail.

8-59 a) $2P(z > 2.15) = 2(.5000 - .4842) = .0316$

b) $P(z > 1.6) = (.5000 - .4452) = .0548$

c) $P(z < -2.33) = (.5000 - .4901) = .0099$

8-60 Let p = the proportion of closed-end equity funds selling at a discount.
$n = 15$ $\quad \bar{p} = 6/15 = 0.4$
$H_0 : p = 0.5$ $\quad H_1 : p < 0.5$ $\quad \alpha = .01$
The lower limit of the acceptance region is $z_L = -2.33$, or

$$\bar{p}_L = p - z_{.01}\sqrt{\frac{pq}{n}} = 0.5 - 2.33\sqrt{\frac{.5(.5)}{15}} = 0.1992$$

Since $z = \dfrac{\bar{p} - p}{\sqrt{pq/n}} = \dfrac{.4 - .5}{\sqrt{.5(.5)/15}} = -0.77 > -2.33$ (or $\bar{p} > 0.1992$), we do not reject H_0.

The proportion of closed-end equity funds selling at a discount is not significantly less than the proportion selling at a premium.

8-61 $s = 5.54$ $\quad n = 15$ $\quad \overline{x} = 1.47$

$H_0 : \mu = 5$ $\quad H_1 : \mu \neq 5$ $\quad \alpha = .05$

The limits of the acceptance region are $t_{crit} = \pm t_{14,\ .025} = \pm 2.145$, or

$$\overline{x}_{CRIT} = \mu \pm t_{14,.025}\, s/\sqrt{n} = 5 \pm 2.145(5.54)/\sqrt{15} = (1.93,\ 8.07)$$

Since $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}} = \dfrac{1.17 - 5}{5.54/\sqrt{15}} = -2.468 < -2.145$ (or $\overline{x} = 1.47 < 1.93$), we reject H_0. The sample does not support his theory.

8-62 a) P(accept | H_0 true) $= .85 \Rightarrow \overline{x} \pm 1.44\sigma_{\overline{x}}$

b) P(accept | H_0 true) $= .98 \Rightarrow \overline{x} \pm 2.33\sigma_{\overline{x}}$

8-63 $\sigma = 13.5$ $\quad n = 70$ $\quad \overline{x} = 86.3$

$H_0 : \mu = 84$ $\quad H_1 : \mu > 84$ $\quad \alpha = .05$

The upper limit of the acceptance region is

$$\overline{x}_U = \mu + z_{.05}\sigma/\sqrt{n} = 84 + 1.64(13.5)/\sqrt{70} = 86.65^\circ \text{ F}$$

Since $\overline{x} < 86.65^\circ$ F, we do not reject H_0. The mean temperature is not significantly higher than 84° F, so the plant should not be cited.

8-64 $\sigma = 1.5$ $\quad n = 200$ $\quad \overline{x} = 31.7$

$H_0 : \mu = 32$ $\quad H_1 : \mu < 32$ $\quad \alpha = .02$

The lower limit of the acceptance region is $z_L = -2.05$, or

$$\overline{x}_L = \mu - z_{.02}\sigma/\sqrt{n} = 32 - 2.05(1.5)/\sqrt{200} = 31.7826 \text{ ounces}$$

Since $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{31.7 - 32}{1.5/\sqrt{200}} = -2.83 < -2.05$ (or $\overline{x} < 31.7826$), we reject H_0. There is significant evidence that the bottles are being underfilled.

8-65 $\sigma = 68$ $\quad n = 90$ $\quad \overline{x} = 218.77$

$H_0 : \mu = 235$ $\quad H_1 : \mu \neq 235$

$$P(\overline{x} \leq 218.77 \text{ or } \overline{x} \geq 251.23 \mid H_0) = 2P\left(z \geq \frac{251.23 - 235}{68/\sqrt{90}}\right)$$

$$= 2P(z \geq 2.26) = 2(.5 - .4881) = .0238$$

Thus, for all $\alpha \leq .0238$, we would accept H_0 and conclude that the average fare is not significantly different from \$235.

8-66 $N = 2400$ $\quad n = 300$ $\quad \overline{p} = 57/300 = .19$

$H_0 : p = .15$ $\quad H_1 : p > .15$ $\quad \alpha = .05$

The upper limit of the acceptance region is $z_U = 1.64$, or

$$\overline{p}_U = p + z_{.05}\sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)} = .15 + 1.64\sqrt{\frac{.15(.85)}{300}\left(\frac{2100}{2399}\right)} = .1816$$

Since $z = \dfrac{\overline{p} - p}{\sqrt{\dfrac{pq}{n}\left(\dfrac{N-n}{N-1}\right)}} = \dfrac{.19 - .15}{\sqrt{\dfrac{.15(.85)}{300}\left(\dfrac{2100}{2399}\right)}} = 2.07 > 1.64$ (or $\overline{p} > .1816$), they should reject H_0 and open the store.

8-67 $\sigma = 2.4$ $\quad n = 16$ $\quad \overline{x} = 28.25$

$H_0 : \mu = 32$ $\quad H_1 : \mu < 32$ $\quad \alpha = .10$

The lower limit of the acceptance region is $z_L = -z_{.10} = -1.28$, or

$$\bar{x}_L = \mu - z_{.10}s/\sqrt{n} = 32 - 1.28(2.4)/\sqrt{16} = 31.23$$

Since $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{28.25 - 32}{2.4/\sqrt{16}} = -6.25 < -1.28$ (or $\bar{x} < 31.23$), we should reject H_0. We should agree with the tax collector — sales have decreased significantly.

8-68 $n = 250$ $\quad \bar{p} = 194/250 = .7760$

$H_0: p = .72 \quad H_1: p > .72 \quad \alpha = .02$

The upper limit of the acceptance region is $z_U = 2.05$, or

$$\bar{p}_U = p + z_{.02}\sqrt{\frac{pq}{n}} = .72 + 2.05\sqrt{\frac{.72(.28)}{250}} = .7782$$

Since $z = \dfrac{\bar{p} - p}{\sqrt{pq/n}} = \dfrac{.7760 - .72}{\sqrt{.72(.28)/250}} = 1.97 < 2.05$ (or $\bar{p} = .7760 < .7782$), we do not reject H_0.

The survey does not support the editor's belief.

8-69 $\sigma = 1250 \quad n = 29 \quad \bar{x} = 23000 \quad \mu_{H_0} = 23500$

$$\mu \pm 2\sigma_{\bar{x}} = \mu \pm 2\sigma/\sqrt{n} = 23500 \pm 2(1250)/\sqrt{29}$$
$$= 23500 \pm 464.24 = (23035.76, 23964.24)$$

Since $\bar{x} < 23035.76$, the company's claim of equal average pay for men and women is not supported by the data.

8-70 $s = 19.48 \quad n = 18 \quad \bar{x} = 87.61$

$H_0: \mu = 77.38 \quad H_1: \mu > 77.38 \quad \alpha = .025$

The upper limit of the acceptance region is $t_U = t_{17,\ .025} = 2.110$, or

$$\bar{x}_U = \mu + t_{17,.025}s/\sqrt{n} = 77.38 + 2.110(19.48)/\sqrt{18} = \$87.07$$

Since $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{87.61 - 77.38}{19.48/\sqrt{18}} = 2.28 > 2.110$ (or $\bar{x} > 87.07$), we reject H_0 and conclude that

Drive-a-Lemon's average total charge is significantly higher than the average total charge at the major national chains. However, this need not indicate that Drive-a-Lemon's rates are not lower than the rates of the major chains. For example, suppose Drive-a-Lemon has offices only in New York and Chicago. The \$77.38 average total charge established by the survey includes charges incurred in other areas which are significantly less expensive than New York and Chicago, and these charges could account for the low average. To validate its claim, Drive-a-Lemon should look at average total charges for the national chains on rentals made in Drive-a-Lemon's service area.

8-71 $s = 9.96 \quad n = 20 \quad \bar{x} = 56$

$H_0: \mu = 50 \quad H_1: \mu > 50 \quad \alpha = .05$

The upper limit of the acceptance region is $t_U = t_{19,\ .05} = 1.729$, or

$$\bar{x}_U = \mu + t_{19,.05}\ s/\sqrt{n} = 50 + 1.729(9.96)/\sqrt{20} = 53.58 \text{ years old}$$

Since $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{56 - 50}{9.96/\sqrt{20}} = 2.694 > 1.729$ (or $\bar{x} = 56 > 53.58$), we reject H_0. The average

age of CEOs of closely held NC corporations is significantly greater than 50.

8-72 $n = 20 \quad \bar{p} = 12/20 = .6$

$H_0: p = .5 \quad H_1: p \neq .5 \quad \alpha = .10$

The limits of the acceptance region are $z_{CRIT} = \pm 1.64$, or

$$\overline{p}_{CRIT} = p \pm z_{.05}\sqrt{\frac{pq}{n}} = .5 \pm 1.64\sqrt{\frac{.5(.5)}{20}} = (0.3166,\ 0.6834)$$

Since $z = \frac{\overline{p}-p}{\sqrt{pq/n}} = \frac{.6-.5}{\sqrt{.5(.5)/20}} = 0.8944 < 1.64$ (or $\overline{p} = .6 < 0.6834$), we fail to reject H_0.

The proportion of CEOs' families with more than two children is not significantly different from 0.5.

8-73 $n = 20$ $\overline{p} = 18/20 = .9$

$H_0 : p = .65$ $H_1 : p > .65$ $\alpha = .02$

The upper limit of the acceptance region is $z_U = 2.05$, or

$$\overline{p}_U = p + z_{.02}\sqrt{\frac{pq}{n}} = .65 + 2.05\sqrt{\frac{.65(.35)}{20}} = 0.8686$$

Since $z = \frac{\overline{p}-p}{\sqrt{pq/n}} = \frac{.9-.65}{\sqrt{.65(.35)/20}} = 2.34 > 2.05$ (or $\overline{p} > 0.8686$), we reject H_0. The proportion of married CEOs is significantly greater than that of the general population.

8-74 From exercise 8-26, we have $\sigma = 5.75$, $n = 25$, and $\overline{x}_{CRIT} = 42.59$.

a) $P(\overline{x} \le 42.59 \mid \mu = 41.95) = P\left(z \le \frac{42.59 - 41.95}{5.75/\sqrt{25}}\right)$

$= P(z \le .56) = .5 + .2123 = .7123$

b) $P(\overline{x} \le 42.59 \mid \mu = 42.95) = P\left(z \le \frac{42.59 - 42.95}{5.75/\sqrt{25}}\right)$

$= P(z \le -.31) = .5 - .1217 = .3783$

c) $P(\overline{x} \le 42.59 \mid \mu = 43.95) = P\left(z \le \frac{42.59 - 43.95}{5.75/\sqrt{25}}\right)$

$= P(z \le -1.18) = 5 - .3810 = .1190$

8-75 $N = 2500$ $n = 250$ $\overline{p} = .13$

$H_0 : p = .18$ $H_1 : p \ne .18$ $\alpha = .05$

$$\hat{\sigma}_{\overline{p}} = \sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)} = \sqrt{\frac{.18(.82)}{250}\left(\frac{2250}{2499}\right)} = .0231$$

The limits of the acceptance region are $z_{CRIT} = \pm 1.96$, or

$$\overline{p}_{CRIT} = p \pm z_{.025}\hat{\sigma}_{\overline{p}} = .18 \pm 1.96(.0231) = (.1347,\ .2253)$$

Since $z = \frac{\overline{p}-p}{\hat{\sigma}_{\overline{p}}} = \frac{.13-.18}{.0231} = -2.16 < -1.96$ (or $\overline{p} < .1347$), we reject H_0. Her belief is not reasonable.

8-76 From exercise 8-28, we have $\sigma = 2100$, $n = 25$, and $\overline{x}_{CRIT} = 13521.4$.

a) $P(\overline{x} \le 13521.4 \mid \mu = 14000) = P\left(z \le \frac{13521.4 - 14000}{2100/\sqrt{25}}\right)$

$= P(z \le -1.14) = .5 - .3729 = .1271$

b) $P(\overline{x} \le 13521.4 \mid \mu = 13500) = P\left(z \le \frac{13521.4 - 13500}{2100/\sqrt{25}}\right)$

$= P(z \le .05) = .5 + .0199 = .5199$

c) $P(\bar{x} \leq 13521.4 \mid \mu = 13000) \quad = P\left(z \leq \frac{13521.4 - 13000}{2100/\sqrt{25}}\right)$

$= P(z \leq 1.24) = 5 + .3925 = .8925$

8-77 $\quad n = 60 \qquad \bar{p} = .75$

$H_0 : p = .85 \qquad H_1 : p < .85 \qquad \alpha = .04$

The lower limit of the acceptance region is $z_L = -1.75$, or

$$\bar{p}_L = p - z_{.04}\sqrt{\frac{pq}{n}} = .85 - 1.75\sqrt{\frac{.85(.15)}{60}} = .7693$$

Since $z = \frac{\bar{p} - p}{\sqrt{pq/n}} = \frac{.75 - .85}{\sqrt{.85(.15)/60}} = -2.17 < -1.75$ (or $\bar{p} < .7693$), we reject H_0. Her accuracy is significantly less than the asserted 85%.

8-78 From exercise 8-26, we have $\sigma = 5.75$, $n = 25$, $H_0 : \mu = 44.95$, $H_1 : \mu < 44.95$

At $\alpha = .05$, the lower limit of the acceptance region is

$$\mu - z_{.05}\sigma/\sqrt{n} = 44.95 - 1.64(5.75)/\sqrt{25} = \$43.06$$

a) $P(\bar{x} \leq 43.06 \mid \mu = 41.95) \quad = P\left(z \leq \frac{43.06 - 41.95}{5.75/\sqrt{25}}\right)$

$= P(z \leq .97) = .5 + .3340 = .8340$

b) $P(\bar{x} \leq 43.06 \mid \mu = 42.95) \quad = P\left(z \leq \frac{43.06 - 42.95}{5.75/\sqrt{25}}\right)$

$= P(z \leq .10) = .5 + .0398 = .5398$

c) $P(\bar{x} \leq 43.06 \mid \mu = 43.95) \quad = P\left(z \leq \frac{43.06 - 43.95}{5.75/\sqrt{25}}\right)$

$= P(z \leq -.77) = 5 - .2794 = .2206$

8-79 $\quad n = 1500 \qquad \bar{p} = 295/1500 = .1967$

$H_0 : p = .18 \qquad H_1 : p > .18 \qquad \alpha = .02$

The upper limit of the acceptance region is $z_U = 2.05$, or

$$\bar{p}_U = p + z_{.02}\sqrt{\frac{pq}{n}} = .18 + 2.05\sqrt{\frac{.18(.82)}{1500}} = .2003$$

Since $z = \frac{\bar{p} - p}{\sqrt{pq/n}} = \frac{.1967 - .18}{\sqrt{.18(.82)/1500}} = 1.68 < 2.05$ (or $\bar{p} < .2003$), we do not reject H_0. The proportion of parents who redeem the coupon has not significantly increased. However, there is nothing we can say about the proportion who use the supplement.

8-80 $\quad N = 5000 \qquad n = 500 \qquad \bar{p} = .43$

$H_0 : p = .48 \qquad H_1 : p < .48 \qquad \alpha = .01$

The lower limit of the acceptance region is $z_L = -2.33$, or

$$\bar{p}_L = p - z_{.01}\sqrt{\frac{pq}{n}\left(\frac{N-n}{N-1}\right)} = .48 - 2.33\sqrt{\frac{.48(.52)}{500}\left(\frac{4500}{4999}\right)} = .4306$$

Since $z = \dfrac{\overline{p} - p}{\sqrt{\dfrac{pq}{n}\left(\dfrac{N-n}{N-1}\right)}} = \dfrac{.43 - .48}{\sqrt{\dfrac{.48(.52)}{500}\left(\dfrac{4500}{4999}\right)}} = -2.36 < -2.33$ (or $\overline{p} < .4306$), we reject H_0. The company has fallen significantly below its target of a 48% market share.

8-81 $\sigma = 2.7$ $n = 36$ $\overline{x} = 12.4$

$H_0 : \mu = 11.6$ $H_1 : \mu > 11.6$ $\alpha = .01$

The upper limit of the acceptance region is $z_U = 2.33$, or

$$\overline{x}_U = \mu + z_{.01}\sigma/\sqrt{n} = 11.6 + 2.33(2.7)/\sqrt{36} = 12.6485$$

Since $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{12.4 - 11.6}{2.7/\sqrt{36}} = 1.78 < 2.33$ (or $\overline{x} < 12.6485$), we do not reject H_0. The machines at Casino World do not have a significantly lower payoff frequency.

CHAPTER 9

TESTING HYPOTHESES: TWO - SAMPLE TESTS

9-1 $s_1 = \sqrt{9} = 3 \quad n_1 = 42 \quad \bar{x}_1 = 32.3 \quad s_2 = \sqrt{16} = 4 \quad n_2 = 57 \quad \bar{x}_2 = 34$

a) $\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{9}{42} + \frac{16}{57}} = 0.704$

b) $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 < \mu_2 \quad \alpha = .05$

The lower limit of the acceptance region is $z_L = -1.645$, or

$(\bar{x}_1 - \bar{x}_2)_L = 0 - z_{.05}\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = -1.645(0.704) = -1.16$

Since $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{(32.3 - 34) - 0}{0.704} = -2.41 < -1.645$ (or $\bar{x}_1 - \bar{x}_2 = 32.3 - 34 = -1.7$), we reject H_0. There is sufficient evidence to conclude that the second population has a larger mean.

9-2 Presumably they wish to determine if $\mu_S < \mu_A$, but, of course, even if this is true, they should also be concerned about the relative costs of the two lines.

$s_S = 32 \quad n_S = 150 \quad \bar{x}_S = 198 \quad s_A = 29 \quad n_A = 200 \quad \bar{x}_A = 206$

$H_0 : \mu_S = \mu_A \quad H_1 : \mu_S < \mu_A \quad \alpha = .02$

$\hat{\sigma}_{\bar{x}_S - \bar{x}_A} = \sqrt{\frac{s_S^2}{n_S} + \frac{s_A^2}{n_A}} = \sqrt{\frac{(32)^2}{150} + \frac{(29)^2}{200}} = 3.3214$ chips per hour

The lower limit of the acceptance region is $z_L = -2.05$, or

$(\bar{x}_S - \bar{x}_A)_L = 0 - z_{.02}\hat{\sigma}_{\bar{x}_S - \bar{x}_A} = -2.05(3.3214) = -6.8089$ chips per hour

Since $z = \frac{(\bar{x}_S - \bar{x}_A) - (\bar{\mu}_S - \bar{\mu}_A)}{\hat{\sigma}_{\bar{x}_S - \bar{x}_A}} = \frac{(198 - 206) - 0}{3.3214} = -2.41 < -2.05$

(or $\bar{x}_S - \bar{x}_A = -8 < -6.8089$), we reject H_0. The output from the automatic line is significantly greater than that from the semiautomatic line.

9-3 $s_1 = 1.8 \quad n_1 = 90 \quad \bar{x}_1 = 8.5 \quad s_2 = 2.1 \quad n_2 = 80 \quad \bar{x}_2 = 7.9$

$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 > \mu_2 \quad \alpha = .05$

$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(1.8)^2}{90} + \frac{(2.1)^2}{80}} = .3019$ hours

The upper limit of the acceptance region is $z_U = 1.64$, or

$(\bar{x}_1 - \bar{x}_2)_{CRIT} = 0 + z_{.05}\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = 1.64(.3019) = .4951$ hours

Since $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{(8.5 - 7.9) - 0}{0.3019} = 1.99 > 1.64$ (or $\bar{x}_1 - \bar{x}_2 = .6 > .4951$), we reject H_0. The second drug does provide a significantly shorter period of relief.

9-4 Sample 1 (1992): $s_1 = 0.84 \quad n_1 = 38 \quad \bar{x}_1 = 4.36$
Sample 2 (1993): $s_2 = 0.51 \quad n_2 = 32 \quad \bar{x}_2 = 3.23$

$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 > \mu_2 \qquad \alpha = .05$

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(0.84)^2}{38} + \frac{(0.51)^2}{32}} = 0.1634\%$$

The upper limit of the acceptance region is $z_U = 1.64$, or

$$(\bar{x}_1 - \bar{x}_2)_{CRIT} = 0 + z_{.05}\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = 1.64(.1634) = 0.2680\%$$

Since $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \dfrac{(4.36 - 3.23) - 0}{0.1634} = 6.92 > 1.64$ (or $\bar{x}_1 - \bar{x}_2 = 1.13\% > 0.2680\%$), we reject H_0 and conclude that money market rates declined significantly in 1992.

9-5 $s_B = .039 \qquad n_B = 75 \qquad \bar{x}_B = 1.059 \qquad s_A = .068 \qquad n_A = 50 \qquad \bar{x}_A = 1.089$

$H_0: \mu_B = \mu_A \qquad H_1: \mu_B \neq \mu_A \qquad \alpha = .02$

$$\hat{\sigma}_{\bar{x}_B-\bar{x}_A} = \sqrt{\frac{s_B^2}{n_B} + \frac{s_A^2}{n_A}} = \sqrt{\frac{(.039)^2}{75} + \frac{(.068)^2}{50}} = \$0.0106$$

The limits of the acceptance region are $z_{CRIT} = \pm 2.33$, or

$$(\bar{x}_B - \bar{x}_A)_{CRIT} = 0 \pm z_{.01}\hat{\sigma}_{\bar{x}_B-\bar{x}_A} = \pm 2.33(.0106) = \pm \$0.0247$$

Since $z = \dfrac{(\bar{x}_B - \bar{x}_A) - (\bar{\mu}_B - \bar{\mu}_A)}{\hat{\sigma}_{\bar{x}_B - \bar{x}_A}} = \dfrac{(1.059 - 1.089) - 0}{0.0106} = -2.83 < -2.33$ (or $\bar{x}_B - \bar{x}_A = -.030 < -.0247$), we reject H_0. The price has changed significantly.

9-6 Sample 1 (male): $s_1 = 1.84 \qquad n_1 = 38 \qquad \bar{x}_1 = 11.38$

Sample 2 (female): $s_2 = 1.31 \qquad n_2 = 45 \qquad \bar{x}_2 = 8.42$

$H_0: \mu_1 - \mu_2 = 2 \qquad H_1: \mu_1 - \mu_2 > 2 \qquad \alpha = .01$

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(1.84)^2}{38} + \frac{(1.31)^2}{45}} = \$0.3567$$

The upper limit of the acceptance region is $z_U = 2.33$, or

$$(\bar{x}_1 - \bar{x}_2)_U = (\mu_1 - \mu_2)_{H_0} + z_{.01}\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = 2 + 2.33(.3567) = \$2.8311$$

Since $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \dfrac{(11.38 - 8.42) - 2}{0.3567} = 2.69 > 2.33$ (or $\bar{x}_1 - \bar{x}_2 = 2.96 > 2.831$), we reject H_0 and conclude that male operators do earn over \$2.00 more per hour than female operators.

9-7 Sample 1 (Southeast): $s_1 = 0.7 \qquad n_1 = 97 \qquad \bar{x}_1 = 8.8$

Sample 2 (Northeast): $s_2 = 0.6 \qquad n_2 = 84 \qquad \bar{x}_2 = 9.0$

$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2 \qquad \alpha = .05$

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(0.7)^2}{97} + \frac{(0.6)^2}{84}} = 0.0966$$

The limits of the acceptance region are $z_{CRIT} = \pm 1.96$, or

$$(\bar{x}_1 - \bar{x}_2)_{CRIT} = (\mu_1 - \mu_2)_{H_0} \pm z_{.025}\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = 0 \pm 1.96(0.0966) = \pm 0.1893$$

Since $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \dfrac{(8.8 - 9.0) - 0}{0.0966} = -2.07 < -1.96$ (or $\bar{x}_1 - \bar{x}_2 = -0.2 < -0.1893$, we reject H_0 and conclude that the FASB proposal will cause a significantly greater reduction in EPS for high tech firms.

9-8 $s_O = 32.63$ $n_O = 16$ $\bar{x}_O = 688$ $s_N = 24.84$ $n_N = 11$ $\bar{x}_N = 706$
$H_0 : \mu_O = \mu_N$ $H_1 : \mu_O < \mu_N$ $\alpha = .05$

$$s_p = \sqrt{\frac{(n_O - 1)s_O^2 + (n_N - 1)s_N^2}{n_O + n_N - 2}} = \sqrt{\frac{15(32.63)^2 + 10(24.84)^2}{25}} = \$29.7597$$

The lower limit of the acceptance region is $t_L = -t_{25,\ .05} = -1.708$, or

$$(\bar{x}_A - \bar{x}_N)_L = 0 - t_{25,.05}s_p\sqrt{\frac{1}{n_O} + \frac{1}{n_N}} = -1.708(29.7597)\sqrt{\frac{1}{16} + \frac{1}{11}} = -\$19.91$$

Since $t = \dfrac{(\bar{x}_O - \bar{x}_B) - (\bar{\mu}_O - \bar{\mu}_N)}{s_p\sqrt{\frac{1}{n_O} + \frac{1}{n_N}}} = \dfrac{(688 - 706) - 0}{29.7597\sqrt{\frac{1}{16} + \frac{1}{11}}} = -1.544 > -1.708$ (or $\bar{x}_O - \bar{x}_B = -18$ > -19.91), we do not reject H_0. Average daily sales have not increased significantly.

9-9 $s_F^2 = 1.0667$ $n_F = 10$ $\bar{x}_F = 12.8$ $s_M^2 = 1.4107$ $n_M = 8$ $\bar{x}_M = 11.625$
$H_0 : \mu_F = \mu_M$ $H_1 : \mu_F > \mu_M$ $\alpha = .05$

$$s_p = \sqrt{\frac{(n_F - 1)s_F^2 + (n_M - 1)s_M^2}{n_F + n_M - 2}} = \sqrt{\frac{9(1.0667) + 7(1.4107)}{16}} = 1.1033 \text{ accounts}$$

The upper limit of the acceptance region is $t_U = t_{16,\ .05} = 1.746$, or

$$(\bar{x}_F - \bar{x}_M)_U = 0 + t_{16,.05}s_p\sqrt{\frac{1}{n_F} + \frac{1}{n_M}} = 1.746(1.1033)\sqrt{\frac{1}{10} + \frac{1}{8}} = .9138 \text{ accounts}$$

Since $t = \dfrac{(\bar{x}_F - \bar{x}_M) - (\bar{\mu}_F - \bar{\mu}_M)}{s_p\sqrt{\frac{1}{n_F} + \frac{1}{n_M}}} = \dfrac{(12.8 - 11.625) - 0}{1.1033\sqrt{\frac{1}{10} + \frac{1}{8}}} = 2.245 > 1.746$ (or $\bar{x}_F - \bar{x}_M =$ $1.175 > 0.9138$), we reject H_0. The women are significantly more effective than the men in generating new accounts.

9-10 $s_1 = 370$ $n_1 = 9$ $\bar{x}_1 = 2990$ $s_2 = 805$ $n_2 = 6$ $\bar{x}_2 = 3065$
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 < \mu_2$ $\alpha = .05$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{8(370)^2 + 5(805)^2}{13}} = \$577.48$$

The lower limit of the acceptance region is $t_L = -t_{13,\ .05} = -1.771$, or

$$(\bar{x}_1 - \bar{x}_2)_L = 0 - t_{13,.05}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = -1.771(577.48)\sqrt{\frac{1}{9} + \frac{1}{6}} = -\$539.02$$

Since $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{(2990 - 3065) - 0}{577.48\sqrt{\frac{1}{9} + \frac{1}{6}}} = -0.246 > -1.771$ (or $\bar{x}_1 - \bar{x}_2 = -75 >$ -539.02), we do not reject H_0. The pear-shaped stones are not significantly more expensive than the marquise stones.

9-11 Sample 1 (California): $s_C = .39$ $n_C = 11$ $\bar{x}_C = 7.61$
Sample 2 (Pennsylvania): $s_P = .56$ $n_P = 8$ $\bar{x}_P = 7.43$

$H_0: \mu_C = \mu_P \qquad H_1: \mu_C \neq \mu_P \qquad \alpha = .10$

$$s_p = \sqrt{\frac{(n_C - 1)s_C^2 + (n_P - 1)s_P^2}{n_C + n_P - 2}} = \sqrt{\frac{10(.39)^2 + 7(.56)^2}{17}} = .4675\ \%$$

The limits of the acceptance region are $t_{CRIT} = \pm t_{17,\ .05} = \pm 1.7401$, or

$$(\bar{x}_C - \bar{x}_P)_{CRIT} = (\mu_C - \mu_P)_{H_0} \pm t_{17,\ .05} s_p \sqrt{\frac{1}{n_C} + \frac{1}{n_P}} = 0 \pm 1.74(.4675)\sqrt{\frac{1}{11} + \frac{1}{8}} = \pm .378$$

Since $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{(7.61 - 7.43) - 0}{0.4675\sqrt{\frac{1}{11} + \frac{1}{8}}} = 0.8286 < -1.740$ (or $\bar{x}_1 - \bar{x}_2 = 0.18 <$ 0.378), we do not reject H_0. The California and Pennsylvania mean mortgage rates are not significantly different.

9-12 Sample 1 (mail): $s_m = 378 \qquad n_m = 17 \qquad \bar{x}_m = 563$
Sample 2 (electronic): $s_e = 619 \qquad n_e = 13 \qquad \bar{x}_e = 958$
$H_0: \mu_m = \mu_e \qquad H_1: \mu_m < \mu_e \qquad \alpha = .01$

$$s_p = \sqrt{\frac{(n_m - 1)s_m^2 + (n_e - 1)s_e^2}{n_m + n_e - 2}} = \sqrt{\frac{16(378)^2 + 12(619)^2}{28}} = \$495.84$$

The lower limit of the acceptance region is $t_L = -t_{28,\ .01} = -2.467$, or

$$(\bar{x}_m - \bar{x}_e)_L = (\mu_m - \mu_e)_{H_0} - t_{28,\ .01} s_p \sqrt{\frac{1}{n_m} + \frac{1}{n_e}} = 0 - (-2.467)(495.84)\sqrt{\frac{1}{17} + \frac{1}{13}} = -450.69$$

Since $t = \dfrac{(\bar{x}_m - \bar{x}_e) - (\bar{\mu}_m - \bar{\mu}_e)}{S_p\sqrt{\frac{1}{n_m} + \frac{1}{n_e}}} = \dfrac{(563 - 958) - 0}{495.84\sqrt{\frac{1}{17} + \frac{1}{13}}} = -2.162 > -2.162$ (or $\bar{x}_m - \bar{x}_e = -395 >$ -450.69), we do not reject H_0. Refunds filed by mail were not significantly smaller than those filed electronically.

9-13 Sample 1 (day): $s_d = 7.806 \qquad n_d = 9 \qquad \bar{x}_d = 113.51$
Sample 2 (night): $s_n = 9.015 \qquad n_n = 9 \qquad \bar{x}_n = 120.52$
$H_0: \mu_d - \mu_n = 0 \qquad H_1: \mu_d - \mu_n < 0 \qquad \alpha = .01$

$$s_p = \sqrt{\frac{(n_d - 1)s_d^2 + (n_n - 1)s_n^2}{n_d + n_n - 2}} = \sqrt{\frac{8(7.806)^2 + 8(9.015)^2}{16}} = 8.4322$$

The lower limit of the acceptance region is $t_L = -t_{.16,\ .01} = -2.583$, or

$$(\bar{x}_d - \bar{x}_n)_L = (\mu_d - \mu_n)_{H_0} - t_{16,\ .01} s_p \sqrt{\frac{1}{n_d} + \frac{1}{n_n}} = -2.583(8.4322)\sqrt{\frac{1}{9} + \frac{1}{9}} = -10.27$$

Since $t = \dfrac{(\bar{x}_d - \bar{x}_n) - (\bar{\mu}_d - \bar{\mu}_n)}{s_p\sqrt{\frac{1}{n_d} + \frac{1}{n_n}}} = \dfrac{(113.51 - 120.52)}{8.4322\sqrt{\frac{1}{9} + \frac{1}{9}}} = -1.764 > -2.583$ (or $\bar{x}_d - \bar{x}_n =$ $-7.01 > -10.27$), we do not reject H_0. The night shift is not producing significantly more tires.

9-14

Firm	1	2	3	4	5	6	7	8	9
1991	1.38	1.26	3.64	3.50	2.47	3.21	1.05	1.98	2.72
1992	2.48	1.50	4.59	3.06	2.11	2.80	1.59	0.92	0.47
Change	1.10	0.24	0.95	−0.44	−0.36	−0.41	0.54	−1.06	−2.25

a) $\bar{x} = \frac{\sum x}{n} = \frac{-1.69}{9} = -0.1878$

b) $s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{8}\left(9.1391 - 9(-0.1878)^2\right) = 1.1027$

$s = \sqrt{s^2} = 1.0501$
$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 1.0501/\sqrt{9} = 0.3500$

c) $H_0: \mu_{1991} = \mu_{1992}$ $\qquad H_1: \mu_{1991} \neq \mu_{1992}$ $\qquad \alpha = .02$

The limits of the acceptance region are $t_{CRIT} = \pm t_{8,\ .01} = \pm 2.896$, or

$\bar{x}_{CRIT} = 0 \pm t_{8,.01}\hat{\sigma}_{\bar{x}} = \pm\ 2.896(0.3500) = \pm\ 1.0136$

Since $t = \frac{\bar{x}-\mu}{\hat{\sigma}_{\bar{x}}} = \frac{-0.1878-0}{0.3500} = -0.537 > -2.896$ (or $\bar{x} = -0.1878$), we do not reject H_0.

Average earnings per share did not change significantly from 1991 to 1992.

9-15

Supplier	1	2	3	4	5	6	7	8
Before	16	12	18	7	14	19	6	17
After	14	13	12	6	9	15	8	15
Change	−2	1	−6	−1	−5	−4	2	−2

$\bar{x} = \frac{\sum x}{n} = \frac{-17}{8} = -2.125$ items,

$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{7}\left(91 - 8(2.125)^2\right) = 7.8393$, $s = \sqrt{7.8393} = 2.7999$ items

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 2.7999/\sqrt{8} = 0.9899$ items

$H_0: \mu_A = \mu_B$ $\qquad H_1: \mu_A < \mu_B$ $\qquad \alpha = .05$

The lower limit of the acceptance region is $t_L = -t_{7,.05} = -1.895$, or

$\bar{x}_L = 0 - t_{7,\ .05}\hat{\sigma}_{\bar{x}} = -1.895(0.9899) = -1.8759$ items

Since $t = \frac{\bar{x}-\mu}{\hat{\sigma}_{\bar{x}}} = \frac{-2.125-0}{0.9899} = -2.147 < -1.895$ (or $\bar{x} = -2.125$), we reject H_0. The new measures have significantly reduced the number of broken items.

9-16

Pair	1	2	3	4	5	6	7	8	9
Regular	5.7	6.1	5.9	6.2	6.4	5.1	5.9	6.0	5.5
Additive	6.0	6.2	5.8	6.6	6.7	5.3	5.7	6.1	5.9
Difference	0.3	0.1	−0.1	0.4	0.3	0.2	−0.2	0.1	0.4

$\bar{x} = \frac{\sum x}{n} = \frac{1.5}{9} = 0.1667$ mpg

$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{8}\left(0.61 - 9(0.1667)^2\right) = 0.0450$, $s = \sqrt{s^2} = 0.2121$ mpg

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 0.2121/\sqrt{9} = 0.0707$ mpg

$H_0: \mu_A = \mu_R$ $\qquad H_1: \mu_A > \mu_R$ $\qquad \alpha = .01$

The upper limit of the acceptance region is $t_U = t_{8,\ .01} = 2.896$, or

$\bar{x}_U = 0 + t_{8,.01}\hat{\sigma}_{\bar{x}} = 2.896(0.0707) = 0.2047$ mpg

Since $t = \frac{\bar{x}-\mu}{\hat{\sigma}_{\bar{x}}} = \frac{0.1667-0}{0.0707} = 2.358 < 2.896$ (or $\bar{x} = 0.1667$), we do not reject H_0. The additive does not yield significantly better fuel efficiency.

9-17

Before	38	11	34	25	17	38	12	27	32	29
After	45	24	41	39	30	44	30	39	40	41
Change	7	13	7	14	13	6	18	12	8	12

$\bar{x} = \frac{\sum x}{n} = \frac{110}{10} = 11$

$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{9}\left(1344 - 10(11)^2\right) = 14.8889,\ s = \sqrt{s^2} = 3.8586$

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 3.8586/\sqrt{10} = 1.2202$ push-ups

$H_0: \mu_A - \mu_B = 8 \qquad H_1: \mu_A - \mu_B > 8 \qquad \alpha = .025$

The upper limit of the acceptance region is $t_U = t_{9,\ .025} = 2.262$, or

$\bar{x}_U = (\mu_A - \mu_B) + t_{9,\ .025}\hat{\sigma}_{\bar{x}} = 8 + 2.262(1.2202) = 10.76$ push-ups

Since $t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{11 - 8}{1.2202} = 2.459 > 2.262$ (or $\bar{x} = 11$), we reject H_0. The data support the club's claim.

9-18

Employee	1	2	3	4	5	6
Production without music	219	205	226	198	209	216
Production with music	235	186	240	203	221	205
Change in production	16	−19	14	5	12	−11

$\bar{x} = \frac{\sum x}{n} = \frac{17}{6} = 2.8333$

$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{5}\left(1103 - 6(2.8333)^2\right) = 210.9667,\ s = \sqrt{s^2} = 14.5247$

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 14.5247/\sqrt{6} = 5.9297$

$H_0: \mu_{AFTER} = \mu_{BEFORE} \qquad H_1: \mu_{AFTER} \neq \mu_{BEFORE} \qquad \alpha = .02$

The limits of the acceptance region are $t_{CRIT} = \pm t_{5,\ .01} = \pm 3.365$, or

$\bar{x}_{CRIT} = 0 \pm t_{5,\ .01}\hat{\sigma}_{\bar{x}} = \pm\ 3.365(5.9297) = \pm\ 19.95$

Since $t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{2.8333 - 0}{5.9297} = 0.478 < 3.365$ (or $\bar{x} = 2.8333$), we do not reject H_0. The music does not have a significant effect on productivity.

9-19

File	1	2	3	4	5	6	7
Haynes Ultima	9.52	10.17	10.33	10.02	10.72	9.62	9.17
Extel PerFAXtion	10.92	11.46	11.18	12.21	10.42	11.36	10.47
Difference in baud rate	−1.40	−1.29	−0.85	−2.19	0.30	−1.74	−1.30

$\bar{x} = \frac{\sum x}{n} = \frac{-8.47}{7} = -1.21$

$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{6}\left(13.9503 - 7(-1.21)^2\right) = 0.7854$

$s = \sqrt{s^2} = 0.2969$ (baud rate in thousands)

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = .7854/\sqrt{7} = 0.2969$

$H_0: \mu_{HAYNES} = \mu_{EXTEL} \qquad H_1: \mu_{HAYNES} < \mu_{EXTEL} \qquad \alpha = .01$

The lower limit of the acceptance region is $t_L = -t_{6,\ .01} = -3.143$, or

$\bar{x}_L = 0 - t_{6,\ .01}\hat{\sigma}_{\bar{x}} = 0 - 3.143(0.2969) = -\ 0.9332$

Since $t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{-1.21 - 0}{0.2969} = -4.075 < -3.143$ (or $\bar{x} = -1.21$), we reject H_0. The Extel PerFAXtion is significantly faster than the Haynes Ultima.

9-20 $n_1 = 40$ $\bar{p}_1 = .275$ $n_2 = 60$ $\bar{p}_2 = .40$

$H_0: p_1 = p_2$ $H_1: p_1 < p_2$ $\alpha = .10$

$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{40(.275) + 60(.40)}{40 + 60} = 0.35$$

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{.35(.65)\left(\frac{1}{40} + \frac{1}{60}\right)} = 0.0974$$

The lower limit of the acceptance region is $z_L = -z_{.10} = -1.28$, or

$$(\bar{p}_1 - \bar{p}_2)_L = 0 - z_{.10}\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = -1.28(.0974) = -0.1247$$

Since $z = \frac{\bar{p}_1 - \bar{p}_2}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}} = \frac{0.275 - 0.4}{0.0974} = -1.283 < -1.28$ (or $\bar{p}_1 - \bar{p}_2 = -.125 < -.1247$), we reject H_0. The proportion of NYSE stocks that advanced on Friday was significantly smaller than those that advanced on Thursday (but just barely significant).

9-21 Sample 1(30–39): $n_1 = 175$ Sample 2 (40–49): $n_2 = 220$

a) Proportions who rate the program as excellent:

$\bar{p}_1 = 87/175 = .4971$ $\bar{p}_2 = 94/220 = .4273$

$H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$ $\alpha = .05$

$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{175(.4971) + 220(.4273)}{175 + 220} = 0.4582$$

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{.4582(.5418)\left(\frac{1}{175} + \frac{1}{220}\right)} = 0.0505$$

The limits of the acceptance region is $z_{CRIT} = \pm 1.96$, or

$$(\bar{p}_1 - \bar{p}_2)_{CRIT} = 0 \pm z_{.025}\hat{\sigma}_{(\bar{p}_1 - \bar{p}_2)} = \pm 1.96(.0505) = \pm 0.0990$$

Since $z = \frac{\bar{p}_1 - \bar{p}_2}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}} = \frac{0.4971 - 0.4273}{0.0505} = 1.38 < 1.96$ (or $\bar{p}_1 - \bar{p}_2 = 0.4971 - 0.4273 = 0.0698 < 0.0990$), we do not reject H_0. The proportions of people in the two age groups who rate the program as excellent are not significantly different.

b) Proportions who plan to purchase an upgrade:

$\bar{p}_1 = 52/175 = .2971$ $\bar{p}_2 = 37/220 = .1682$

$H_0: p_1 = p_2$ $H_1: p_1 \neq p_2$ $\alpha = .05$

$$\hat{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{175(.2971) + 220(.1682)}{175 + 220} = 0.2253$$

$$\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{.2253(.7747)\left(\frac{1}{175} + \frac{1}{220}\right)} = 0.0423$$

The limits of the acceptance region is $z_{CRIT} = \pm 1.96$, or

$$(\bar{p}_1 - \bar{p}_2)_{CRIT} = 0 \pm z_{.025}\hat{\sigma}_{(\bar{p}_1 - \bar{p}_2)} = \pm 1.96(.0423) = \pm 0.0829$$

Since $z = \frac{\bar{p}_1 - \bar{p}_2}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}} = \frac{0.2971 - 0.1682}{0.0423} = 3.05 > 1.96$ (or $\bar{p}_1 - \bar{p}_2 = 0.2971 - 0.1682 = 0.1289 > 0.0829$), now we do reject H_0. The proportions of people in the two age groups who plan to purchase an upgrade are significantly different.

9-22 $n_1 = 200$ $\quad \overline{p}_1 = .68$ $\quad n_2 = 250$ $\quad \overline{p}_2 = .76$

$H_0: p_1 = p_2$ $\quad H_1 = p_1 < p_2$ $\quad \alpha = .02$

$$\hat{p} = \frac{n_1\overline{p}_1 + n_2\overline{p}_2}{n_1 + n_2} = \frac{200(.68) + 250(.76)}{200 + 250} = 0.7244$$

$$\hat{\sigma}_{\overline{p}_1 - \overline{p}_2} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{.7244(.2756)\left(\frac{1}{200} + \frac{1}{250}\right)} = 0.0424$$

The lower limit of the acceptance region is $z_L = -z_{.02} = -2.05$, or

$$(\overline{p}_1 - \overline{p}_2)_L = 0 - z_{.02}\hat{\sigma}_{\overline{p}_1 - \overline{p}_2} = -2.05(0.0424) = -0.0869$$

Since $z = \dfrac{\overline{p}_1 - \overline{p}_2}{\hat{\sigma}_{\overline{p}_1 - \overline{p}_2}} = \dfrac{0.68 - 0.76}{0.0424} = -1.89 > -2.05$ (or $\overline{p}_1 - \overline{p}_2 = -0.08 > -0.0869$), we do not reject H_0. The less expensive system will be installed.

9-23 $n_T = 120$ $\quad \overline{p}_T = .45$ $\quad n_C = 150$ $\quad \overline{p}_C = .36$

$H_0: p_T = p_C$ $\quad H_1: p_T > p_C$ $\quad \alpha = .01$

$$\hat{p} = \frac{n_T\overline{p}_T + n_C\overline{p}_C}{n_T + n_C} = \frac{120(.45) + 150(.36)}{270} = .4$$

$$\hat{\sigma}_{\overline{p}_T - \overline{p}_C} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_T} + \frac{1}{n_C}\right)} = \sqrt{.4(.6)\left(\frac{1}{120} + \frac{1}{150}\right)} = .06$$

The upper limit of the acceptance region is $z_U = z_{.01} = 2.33$, or

$$(\overline{p}_T - \overline{p}_C)_U = 0 + z_{.01}\hat{\sigma}_{\overline{p}_T - \overline{p}_C} = 2.33(.06) = .1398$$

Since $z = \dfrac{\overline{p}_T - \overline{p}_C}{\hat{\sigma}_{\overline{p}_T - \overline{p}_C}} = \dfrac{0.45 - 0.36}{0.06} = 1.5 < 2.33$ (or $\overline{p}_T - \overline{p}_C = .09 < .1398$), we do not reject H_0. The new drug is not significantly more effective in reducing high blood pressure.

9-24 $n_F = 150$ $\quad \overline{p}_F = .46$ $\quad n_S = 175$ $\quad \overline{p}_S = .40$

$H_0: p_F = p_S$ $\quad H_F: p_F \neq p_S$ $\quad \alpha = .10$

$$\hat{p} = \frac{n_F\overline{p}_F + n_S\overline{p}_S}{n_F + n_S} = \frac{150(.46) + 175(.40)}{150 + 175} = .4277$$

$$\hat{\sigma}_{\overline{p}_F - \overline{p}_S} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_F} + \frac{1}{n_S}\right)} = \sqrt{.4277(.5723)\left(\frac{1}{150} + \frac{1}{175}\right)} = .0551$$

The limits of the acceptance region are $z_{CRIT} = \pm z_{.05} = \pm 1.64$, or

$$(\overline{p}_F - \overline{p}_S)_{CRIT} = 0 \pm z_{.05}\hat{\sigma}_{\overline{p}_F - \overline{p}_S} = \pm 1.64(.0551) = \pm .0904$$

Since $z = \dfrac{\overline{p}_F - \overline{p}_S}{\hat{\sigma}_{\overline{p}_F - \overline{p}_S}} = \dfrac{0.46 - 0.40}{0.0551} = 1.09 < 1.64$ (or $\overline{p}_F - \overline{p}_S = 0.06 < 0.0904$), we do not reject H_0. The proportions of freshmen and sophomores do not differ significantly.

9-25 a) $n_M = 1000$ $\quad \overline{p}_M = 743/1000 = .743$ $\quad n_F = 500$ $\quad \overline{p}_F = 405/500 = .810$

$$\hat{p} = \frac{n_M\overline{p}_M + n_F\overline{p}_F}{n_M + n_F} = \frac{743 + 405}{1500} = .7653$$

b) $\hat{\sigma}_{\overline{p}_M - \overline{p}_F} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_M} + \frac{1}{n_F}\right)} = \sqrt{.7653(.2347)\left(\frac{1}{1000} + \frac{1}{500}\right)} = .0232$

c) $H_0 : p_M = p_F$ $\quad H_1 : p_M \neq p_F$ $\quad \alpha = .05$

The limits of the acceptance region are $z_{CRIT} = z_{.025} \pm 1.96$, or

$$(\bar{p}_M - \bar{p}_F)_{CRIT} = 0 \pm z_{.025}\hat{\sigma}_{\bar{p}_M - \bar{p}_F} = \pm 1.96(.0232) = \pm .0455$$

Since $z = \dfrac{\bar{p}_M - \bar{p}_F}{\hat{\sigma}_{\bar{p}_M - \bar{p}_F}} = \dfrac{0.743 - 0.810}{0.0232} = -2.89 < -1.96$ (or $\bar{p}_M - \bar{p}_F = -.067 < -.0455$), we reject H_0. The proportions of men and women in favor of increased retirement benefits are significantly different.

9-26 $\sigma = 7600$ $\quad n = 64$ $\quad \bar{x} = 38500$

$H_0 : \mu = 40000$ $\quad H_1 : \mu < 40000$

$$P(\bar{x} \le 38500 \mid H_0) = P\left(z \le \frac{38500 - 40000}{7600/\sqrt{64}}\right) = P(z \le -1.58) = .5 - .4429 = .0571$$

9-27 $n = 300$ $\quad \bar{p} = 48/300 = .16$

$H_0 : p = .18$ $\quad H_1 : p < .18$

$$P(\bar{p} \le .16 \mid H_0) = P\left(z \le \frac{.16 - .18}{\sqrt{.18(.82)/300}}\right) = P(z \le -0.90) = .5 - .3159 = .1841$$

9-28 $\sigma = .06$ $\quad n = 25$ $\quad \bar{x} = 4.97$

$H_0 : \mu = 5.00$ $\quad H_1 : \mu \neq 5.00$

$$P(\bar{x} \le 4.97 \text{ or } \bar{x} \ge 5.03 \mid H_0) = 2P\left(z \ge \frac{5.03 - 5.00}{.06/\sqrt{25}}\right)$$

$$= 2P(z \ge 2.5) = 2(.5 - .4938) = .0124$$

Hence, for any significance level greater than .0124 we should reject H_0 (and recalibrate the machine), but if $\alpha < .0124$, we will accept H_0 (and leave the machine as it is currently set).

9-29 $n = 125$ $\quad \bar{p} = 94/125 = .752$

$H_0 : p = .8$ $\quad H_1 : p \neq .8$

$$P(\bar{p} \le .752 \text{ or } \bar{p} \ge .848 \mid H_0) = 2P\left(z \ge \frac{.848 - .8}{\sqrt{.8(.2)/125}}\right)$$

$$= 2P(z \ge 1.34) = 2(.5 - .4099) = .1802$$

9-30 From exercise 9-2, we have $s_1 = 32$, $n_1 = 150$, $\bar{x}_1 = 198$, and $s_2 = 29$, $n_2 = 200$, $\bar{x}_2 = 206$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(32)^2}{150} + \frac{(29)^2}{200}} = 3.3214$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{(198 - 206) - 0}{3.3214} = -2.409$$

$$P(z \le -2.409) = .5 - .4920 = .0080$$

9-31 From exercise 9-3, we have $s_1 = 1.8$, $n_1 = 90$, $\bar{x}_1 = 8.5$, and $s_2 = 2.1$, $n_2 = 80$, $\bar{x}_2 = 7.9$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(1.8)^2}{90} + \frac{(2.1)^2}{80}} = 0.3019$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{(8.5 - 7.9) - 0}{0.3019} = 1.9874$$

$$P(z \ge 1.9874) = .5 - .4767 = 0.0233$$

9-32 From exercise 9-8, we have $t = -1.544$ (which is greater than -1.708), so the probability value is greater than .05 in a lowered-tailed test with 25 degrees of freedom.

9-33 From exercise 9-11, we have $s_C = .39$, $n_C = 11$, $\bar{x}_C = 7.61$, and $s_P = .56$, $n_P = 8$, $\bar{x}_P = 7.43$,

$$t = \frac{(\bar{x}_C - \bar{x}_P) - (\mu_C - \mu_P)_{H_0}}{s_p\sqrt{\frac{1}{n_C} + \frac{1}{n_P}}} = \frac{(7.61 - 7.43) - 0}{0.4675\sqrt{\frac{1}{11} + \frac{1}{8}}} = 0.8286$$

Since $t = 0.8286$ (which is less than 1.740 in a two-tailed test with 17 degrees of freedom), the probability value is greater than 0.10.

9-34 From exercise 9-14, we have $s = 1.0501$, n = 9, $\bar{x} = -0.1878$

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{\text{n}}} = \frac{1.0501}{\sqrt{9}} = 0.35$$

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{-0.1878 - 0}{0.35} = -0.5366$$

Since $t = -0.537$ (which is greater than -1.860 in a two-tailed test with 8 degrees of freedom), the probaility value is greater than 0.10.

9-35 From exercise 9-15, we have $s = 2.799$, $n = 8$, $\bar{x} = -2.125$

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.799}{\sqrt{8}} = 0.9899$$

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{-2.125 - 0}{0.9899} = -2.147$$

Since $t = -2.147$ (which is between -1.895 and -2.365 in a lower-tailed test with 7 degrees of freedom), the probability value is between 0.025 and 0.05.

9-36 From exercise 9-22, we have $\text{n}_1 = 200$, $\bar{p}_1 = .68$, $n_2 = 250$, $\bar{p}_2 = .76$, $\hat{\sigma}_{\bar{p}_1 - \bar{p}_2} = 0\ .0424$

$$z = \frac{\bar{p}_1 - \bar{p}_2}{\hat{\sigma}_{\bar{p}_1 - \bar{p}_2}} = \frac{.68 - .76}{.0424} = -1.887$$

Since $z = -1.887$, in a lower tailed test the probablity value is $\text{P}(z \le -1.89) = .5 - .4706 = 0.0294$

9-37 From exercise 9-25, we have $\bar{p}_M = 0.743$, $\bar{p}_F = 0.810$, $\hat{\sigma}_{\bar{p}_M - \bar{p}_F} = 0.0232$

$$z = \frac{\bar{p}_M - \bar{p}_F}{\hat{\sigma}_{\bar{p}_M - \bar{p}_F}} = \frac{.743 - .810}{.0232} = -2.888$$

Since $z = -2.888$, in a two-tailed test the probability value is
$\text{P}(|z| \ge -2.89) = 2(.5 - 0.4981) = 0.0038$

9-38 $s_S = 8$ $n_S = 40$ $\bar{x}_S = 42$ $s_F = 7$ $n_F = 40$ $\bar{x}_F = 45$

$\text{H}_0 : \mu_S = \mu_F$ $\text{H}_1 : \mu_S \neq \mu_F$ $\alpha = .02$

$$\hat{\sigma}_{\bar{x}_S - \bar{x}_F} = \sqrt{\frac{s_S^2}{n_S} + \frac{s_F^2}{n_F}} = \sqrt{\frac{8^2}{40} + \frac{7^2}{40}} = 1.6808 \text{ pens per store per month}$$

The limits of the acceptance region are $z_{CRIT} = \pm z_{.01} = \pm 2.33$, or

$$(\bar{x}_S - \bar{x}_F)_{CRIT} = 0 \pm z_{.01}\hat{\sigma}_{\bar{x}_S - \bar{x}_F} = \pm\ 2.33(1.6808) = \pm\ 3.9163 \text{ pens per store per month}$$

Since $z = \frac{(\bar{x}_S - \bar{x}_F) - (\bar{\mu}_S - \bar{\mu}_F)}{\hat{\sigma}_{\bar{x}_S - \bar{x}_F}} = \frac{(42-45)-0}{1.6808} = -1.78 > -2.33$ (or $\bar{x}_S - \bar{x}_F = -3 > -3.9163$), we do not reject H_0. The two displays did not result in significantly different sales levels.

9-39 $s_1 = 18.2$ $n_1 = 50$ $\bar{x}_1 = 73.6$ $s_2 = 19.7$ $n_2 = 75$ $\bar{x}_2 = 68.9$

$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$ $\alpha = .10$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(18.2)^2}{50} + \frac{(19.7)^2}{75}} = 3.435\%$$

The limits of the acceptance region are $z_{CRIT} = \pm z_{.05} = \pm 1.96$, or

$$(\bar{x}_1 - \bar{x}_2)_{CRIT} = 0 \pm z_{.05}\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \pm 1.64(3.435) = 5.6334\%$$

Since $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{(73.6 - 68.9) - 0}{3.435} = 1.37 < 1.96$ (or $\bar{x}_1 - \bar{x}_2 = 4.7\%$), we do not reject H_0. The occupancy rate did not change significantly.

9-40 a) Before sample: $n = 11$ $\sum x = 195$ $\sum x^2 = 4195$

$$\bar{x}_B = \frac{\sum x}{n} = \frac{195}{11} = 17.7273 \text{ ounces}$$

$$s_B^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{10}\left(4195 - 11(17.7273)^2\right) = 73.8171$$

After sample: $n = 11$ $\sum x = 241$ $\sum x^2 = 5987$

$$\bar{x}_A = \frac{241}{11} = 21.9091 \text{ ounces}$$

$$s_A^2 = \frac{1}{10}\left(5987 - 11(21.9091)^2\right) = 70.6905$$

$$s_p = \sqrt{\frac{(n_B - 1)s_B^2 + (n_A - 1)s_A^2}{n_B + n_A - 2}} = \sqrt{\frac{10(73.8171) + 10(70.6905)}{20}} = 8.5002 \text{ oz}$$

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = s_p\sqrt{\frac{1}{n_B} + \frac{1}{n_A}} = 8.5002\sqrt{\frac{1}{11} + \frac{1}{11}} = 3.6245 \text{ ounces}$$

$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A > \mu_B$ $\alpha = .05$

The upper limit of the acceptance region is $t_U = t_{20,\ .05} = 1.725$, or

$$(\bar{x}_A - \bar{x}_B)_U = 0 + t_{20,.05}\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = 1.725(3.6245) = 6.2523 \text{ ounces}$$

Since $t = \frac{(\bar{x}_A - \bar{x}_B) - (\bar{\mu}_A - \bar{\mu}_B)}{\hat{\sigma}_{\bar{x}_A - \bar{x}_B}} = \frac{(21.9091 - 17.7273) - 0}{3.6245} = 1.154 < 1.725$ (or $\bar{x}_A - \bar{x}_B = 4.1818 < 6.2523$), we do not reject H_0. The demand has not increased significantly.

b) A better sampling procedure would be to re-interview the same 11 customers who were surveyed before the campaign. Since this would control for other factors, he would expect to see a smaller value of $\hat{\sigma}_{\bar{x}_A - \bar{x}_B}$, so any observed difference would be more likely to be significant.

9-41 $n_B = 200$ $\bar{p}_B = .07$ $n_H = 200$ $\bar{p}_H = .03$

$H_0 : \bar{p}_B = \bar{p}_B$ $H_1 : \bar{p}_B > \bar{p}_H$ $\alpha = .10$

$$\hat{p} = \frac{n_B \bar{p}_B + n_H \bar{p}_H}{n_B + n_H} = \frac{200(.07) + 200(.03)}{200 + 200} = .05$$

$$\hat{\sigma}_{\bar{p}_B - \bar{p}_H} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_B} + \frac{1}{n_H}\right)} = \sqrt{.05(.95)\left(\frac{1}{200} + \frac{1}{200}\right)} = 0.0218$$

The upper limit of the acceptance region is $z_U = z_{.10} = 1.28$, or

$$(\bar{p}_B - \bar{p}_H)_U = 0 + z_{.10}\hat{\sigma}_{\bar{p}_B - \bar{p}_H} = 1.28(.0218) = 0.0279$$

Since $z = \frac{\bar{p}_B - \bar{p}_H}{\hat{\sigma}_{\bar{p}_B - \bar{p}_H}} = \frac{.07 - .03}{.0218} = 1.83 > 1.28$ (or $\bar{p}_B - \bar{p}_H = 0.04 > 0.0279$), we reject H_0. A significantly greater proportion of younger consumers (under age 25) preferred Ben & Jerry's to Haagen-Daz ice cream.

9-42 $n_B = 150$ $\bar{p}_B = .44$ $n_A = 200$ $\bar{p}_A = .52$

$H_0: \bar{p}_A = \bar{p}_B$ $H_1: \bar{p}_A > \bar{p}_B$ $\alpha = .04$

$$\hat{p} = \frac{n_A \bar{p}_A + n_B \bar{p}_B}{n_A + n_B} = \frac{200(.52) + 150(.44)}{200 + 150} = .4857$$

$$\hat{\sigma}_{\bar{p}_A - \bar{p}_B} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} = \sqrt{.4857(.5143)\left(\frac{1}{200} + \frac{1}{150}\right)} = .0540$$

The upper limit of the acceptance region is $z_U = z_{.04} = 1.75$, or

$$(\bar{p}_A - \bar{p}_B)_U = 0 + z_{.04}\hat{\sigma}_{\bar{p}_A - \bar{p}_B} = 1.75(.0540) = .0945$$

Since $z = \frac{\bar{p}_A - \bar{p}_B}{\hat{\sigma}_{\bar{p}_A - \bar{p}_B}} = \frac{.52 - .44}{.0540} = 1.48 < 1.75$ (or $\bar{p}_A - \bar{p}_B = 0.08 < 0.0945$), we do not reject H_0. The campaign was not significantly effective.

9-43 a) Letter sample: $n = 7$ $\sum x = 73$ $\sum x^2 = 783$

$$\bar{x}_L = \frac{\sum x}{n} = \frac{73}{7} = 10.4286 \text{ days}$$

$$s_L^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{6}\left(783 - 7(10.4286)^2\right) = 3.6184$$

Call sample: $n = 7$ $\sum x = 43$ $\sum x^2 = 287$

$$\bar{x}_C = \frac{43}{7} = 6.1429 \text{ days}$$

$$s_C^2 = \frac{1}{6}\left(287 - 7(6.1429)^2\right) = 3.8089$$

$$s_p = \sqrt{\frac{(n_L - 1)s_L^2 + (n_C - 1)s_C^2}{n_L + n_C - 2}} = \sqrt{\frac{6(3.6184) + 6(3.8089)}{12}} = 1.9271 \text{ days}$$

$$\hat{\sigma}_{\bar{x}_L - \bar{x}_C} = s_p\sqrt{\frac{1}{n_L} + \frac{1}{n_C}} = 1.9271\sqrt{\frac{1}{7} + \frac{1}{7}} = 1.0301 \text{ days}$$

$H_0: \mu_L = \mu_C$ $H_1: \mu_L > \mu_C$ $\alpha = .025$

The upper limit of the acceptance region is $t_U = t_{12,\ .025} = 2.179$, or

$$(\bar{x}_L - \bar{x}_C)_U = 0 + t_{12,.025}\hat{\sigma}_{\bar{x}_L - \bar{x}_C} = 2.179(1.0301) = 2.2446 \text{ days}$$

Since $t = \frac{(\bar{x}_L - \bar{x}_C) - (\bar{\mu}_L - \bar{\mu}_C)}{\hat{\sigma}_{\bar{x}_L - \bar{x}_C}} = \frac{(10.4286 - 6.1429) - 0}{1.0301} = 4.160 > 2.179$ (or $\bar{x}_L - \bar{x}_C = 4.2857 > 2.2446$), we reject H_0. Slow accounts are collected significantly more quickly with calls than with letters.

b) However, we cannot conclude that the customers respond more quickly to calls than to letters, because we do not know how much of the 10.4286 day average collection time results from the time between the letter's mailing and its receipt by the customer.

9-44

Buffered Aspirin	16.5	25.5	23.0	14.5	28.0	10.0	21.5	18.5	15.5
Competition	12.0	20.5	25.0	16.5	24.0	11.5	17.0	15.0	13.0
Difference	−4.5	−5.0	2.0	2.0	−4.0	1.5	−4.5	−3.5	−2.5

$$\bar{x} = \frac{\sum x}{n} = \frac{-18.5}{9} = -2.0556$$

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{8}\left(110.25 - 9(-2.0556)^2\right) = 9.0276$$

$s = \sqrt{s^2} = 3.0046$ minutes

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 3.0046/\sqrt{9} = 1.0015$ minutes

$H_0 : \mu_C = \mu_B \qquad H_1 : \mu_C > \mu_B \qquad \alpha = .10$

The limits of the acceptance region are $t_{CRIT} = \pm t_{8,\ .05} = \pm 1.860$, or

$$\bar{x}_{CRIT} = 0 \pm t_{8,\ .05}\hat{\sigma}_{\bar{x}} = \pm\ 1.860(1.0015) = \pm\ 1.8628 \text{ minutes}$$

Since $t = \dfrac{\bar{x}-\mu}{\hat{\sigma}_{\bar{x}}} = \dfrac{-2.0556-0}{1.0015} = 2.053 > 1.860$ (or $\bar{x} = -2.0556 < -1.860$), we do reject H_0.

There is a significant difference in the times the two medications take to reach the bloodstream.

9-45 1992: $n_{92} = 373{,}842 \qquad \bar{p}_{92} = .525$

1993: $n_{93} = 372{,}442 \qquad \bar{p}_{93} = .538$

$H_0 : \bar{p}_{92} = \bar{p}_{93} \qquad H_1 : \bar{p}_{92} < \bar{p}_{93} \qquad \alpha = .05$

$$\hat{p} = \frac{n_{92}\bar{p}_{92} + n_{93}\bar{p}_{93}}{n_{92} + n_{93}} = \frac{373{,}842(.525) + 372{,}442(.538)}{373{,}842 + 372{,}442} = 0.5315$$

$$\hat{\sigma}_{\bar{p}_{92}-\bar{p}_{93}} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_{92}} + \frac{1}{n_{93}}\right)} = \sqrt{.5315(.4685)\left(\frac{1}{373{,}842} + \frac{1}{372{,}442}\right)} = 0.0012$$

The lower limit of the acceptance region is $z_L = -z_{.05} = -1.64$, or

$$(\bar{p}_{92} - \bar{x}_{93})_L = 0 - z_{.05}\hat{\sigma}_{\bar{p}_{92}-\bar{p}_{93}} = 0 - 1.64(.0012) = -0.0019$$

Since $z = \dfrac{\bar{p}_{92} - \bar{p}_{93}}{\hat{\sigma}_{\bar{p}_{92}-\bar{p}_{93}}} = \dfrac{.525 - .538}{.00116} = -11.21 < -1.64$ (or $\bar{p}_{92} - \bar{p}_{93} = -0.013 < -0.0019$), we reject H_0. Market share for imports has increased significantly.

9-46

Old Formula	5	2	5	4	3	6	2	4	2	6	5	7	1	3
New Formula	3	1	5	1	1	4	4	2	5	2	3	3	1	2
Difference	−2	−1	0	−3	−2	−2	2	−2	3	−4	−2	−4	0	−1

$$\bar{x} = \frac{\sum x}{n} = \frac{-18}{14} = -1.2857 \text{ days}$$

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{13}\left(76 - 14(-1.2857)^2\right) = 4.0660,\ s = \sqrt{s^2} = 2.0164$$

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 2.0164/\sqrt{14} = 0.5389$

$H_0 : \mu_{NEW} = \mu_{OLD} \qquad H_1 : \mu_{NEW} < \mu_{OLD} \qquad \alpha = .01$

The lower limit of the acceptance region is $t_L = -t_{13,\ .01} = -2.650$, or

$$\bar{x}_L = 0 - t_{13,\ .01}\hat{\sigma}_{\bar{x}} = -2.650(0.5389) = -1.4281$$

Since $t = \dfrac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \dfrac{-1.2857 - 0}{0.5389} = -2.386 > -2.650$ (or $\bar{x} = -1.2857 > -1.4281$), we do not reject H_0. The new formula is not significantly more effective.

9-47 1 month free: $n = 12$ $\sum x = 1324$ $\sum x^2 = 147582$

$$\bar{x}_1 = \frac{\sum x}{n} = \frac{1324}{12} = 110.3333$$

$$s_1^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{11}\left(147582 - 12(110.3333)^2\right) = 136.4323$$

Low monthly fee: $n = 10$ $\sum x = 1198$ $\sum x^2 = 144792$

$$\bar{x}_L = \frac{1198}{10} = 119.8$$

$$s_L^2 = \frac{1}{9}\left(144792 - 10(119.8)^2\right) = 141.2889$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_L - 1)s_L^2}{n_1 + n_L - 2}} = \sqrt{\frac{11(136.4323) + 9(141.2889)}{20}} = 11.7736$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_L} = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_L}} = 11.7736\sqrt{\frac{1}{12} + \frac{1}{10}} = 5.0412$$

$H_0: \mu_1 = \mu_L$ $H_1: \mu_1 \neq \mu_L$ $\alpha = .10$

The limits of the acceptance region are $t_{CRIT} = \pm t_{20,\ .05} = \pm 1.725$, or

$$(\bar{x}_1 - \bar{x}_L)_{CRIT} = 0 \pm t_{20,.05}\hat{\sigma}_{\bar{x}_1 - \bar{x}_L} = \pm 1.725(5.0412) = \pm 8.6961$$

Since $t = \dfrac{(\bar{x}_1 - \bar{x}_L) - (\bar{\mu}_1 - \bar{\mu}_L)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_L}} = \dfrac{(110.3333 - 119.8) - 0}{5.0412} = -1.878 < -1.725$ (or $\bar{x}_1 - \bar{x}_L = -9.4667 < -8.6961$), we reject H_0. The productivities with the two offers are significantly different.

9-48 $s_M^2 = 2.5714$ $n_M = 8$ $\bar{x}_M = 3.5$ $s_S^2 = 1.9821$ $n_S = 8$ $\bar{x}_S = 5.625$

$H_0: \mu_M = \mu_S$ $H_1: \mu_M < \mu_S$ $\alpha = .025$

$$s_p = \sqrt{\frac{(n_M - 1)s_M^2 + (n_S - 1)s_S^2}{n_M + n_S - 2}} = \sqrt{\frac{7(2.5714) + 7(1.9821)}{14}} = 1.5089$$

The lower limit of the acceptance region is $t_L = -t_{14,\ .025} = -2.145$, or

$$(\bar{x}_M - \bar{x}_S)_{CRIT} = 0 - t_{14,.025}s_p\sqrt{\frac{1}{n_M} + \frac{1}{n_S}} = -2.145(1.5089)\sqrt{\frac{1}{8} + \frac{1}{8}} = -1.6183$$

Since $t = \dfrac{(\bar{x}_M - \bar{x}_S) - (\bar{\mu}_M - \bar{\mu}_S)}{s_p\sqrt{\frac{1}{n_M} + \frac{1}{n_S}}} = \dfrac{(3.5 - 5.625) - 0}{1.5089\sqrt{\frac{1}{8} + \frac{1}{8}}} = -2.817 < -2.145$ (or $\bar{x}_M - \bar{x}_S = 3.5 - 5.625 = -2.125 < -1.6183$, we reject H_0. Severe consequences lead to a significantly greater attribution of responsibility.

9-49 $n_m = 120$ $\bar{p}_m = \dfrac{87}{120} = 0.725$ $n_p = 150$ $\bar{p}_p = \dfrac{89}{150} = 0.5933$

$H_0: \bar{p}_m = \bar{p}_p$ $H_1: \bar{p}_m > \bar{p}_p$ $\alpha = .10$

$$\hat{p} = \frac{n_m\bar{p}_m + n_p\bar{p}_p}{n_m + n_p} = \frac{120(.725) + 150(.5933)}{120 + 150} = 0.6518$$

$$\hat{\sigma}_{\bar{p}_m - \bar{p}_p} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_m} + \frac{1}{n_p}\right)} = \sqrt{.6518(.3482)\left(\frac{1}{120} + \frac{1}{150}\right)} = 0.0583$$

The upper limit of the acceptance region is $z_U = z_{.10} = 1.28$, or

$$(\bar{p}_m - \bar{x}_p)_{CRIT} = 0 + z_{.10}\hat{\sigma}_{\bar{p}_m - \bar{p}_p} = 0 + 1.28(.0583) = 0.0746$$

Since $z = \dfrac{\bar{p}_m - \bar{p}_p}{\hat{\sigma}_{\bar{p}_m - \bar{p}_p}} = \dfrac{.725 - .5933}{.0583} = 2.26 > 1.28$ (or $\bar{p}_m - \bar{p}_p = 0.1317 > 0.0746$), we reject H_0. The purchasing agents are significantly more pessimistic than the macroeconomists.

9-50 Disney: $n = 5$ $\quad \sum x = 143.8$ $\quad \sum x^2 = 4898.06$

$$\bar{x}_D = \frac{\sum x}{n} = \frac{143.8}{5} = 28.76$$

$$s_D^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{4}\left(4898.06 - 5(28.76)^2\right) = 190.593$$

Competition: $n = 11$ $\quad \sum x = 156.7$ $\quad \sum x^2 = 3969.05$

$$\bar{x}_C = \frac{\sum x}{n} = \frac{156.7}{11} = 14.2455$$

$$s_C^2 = \frac{1}{10}\left(3969.05 - 11(14.2455)^2\right) = 173.6773$$

$H_0 : \mu_D = \mu_C$ $\qquad H_1 : \mu_D > \mu_C$ $\qquad \alpha = .05$

$$s_p = \sqrt{\frac{(n_D - 1)s_D^2 + (n_C - 1)s_C^2}{n_D + n_C - 2}} = \sqrt{\frac{4(190.593) + 10(173.6773)}{14}} = 13.3608$$

The upper limit of the acceptance region is $t_U = t_{14,\ .05} = 1.761$, or

$$(\bar{x}_D - \bar{x}_C)_U = (\mu_D - \mu_C)_{H_0} + t_{14,\,.05}s_p\sqrt{\frac{1}{n_D} + \frac{1}{n_C}} = 0 + 1.761(13.361)\sqrt{\frac{1}{5} + \frac{1}{11}} = 12.6903$$

Since $t = \dfrac{(\bar{x}_D - \bar{x}_C) - (\bar{\mu}_D - \bar{\mu}_C)}{s_p\sqrt{\frac{1}{n_D} + \frac{1}{n_C}}} = \dfrac{(28.76 - 14.2455) - 0}{13.361\sqrt{\frac{1}{5} + \frac{1}{11}}} = 2.014 > 1.761$ (or $\bar{x}_D - \bar{x}_C = 14.5145 > 12.6903$), we reject H_0. The Disney films earn significantly more than the competitors' films earn.

9-51 $s_A = 1.13$, $n_A = 26$, $\bar{x}_A = 3.37$, and $s_N = 1.89$, $n_N = 35$, $\bar{x}_N = 3.15$
$H_0 : \mu_A - \mu_N = 2$ $\qquad H_1 : \mu_A - \mu_N > 2$ $\qquad \alpha = .02$

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_N} = \sqrt{\frac{s_A^2}{n_A} + \frac{s_N^2}{n_N}} = \sqrt{\frac{(1.13)^2}{26} + \frac{(1.89)^2}{35}} = 0.3888$$

The upper limit of the acceptance region is $z_U = z_{.02} = 2.05$, or

$$(\bar{x}_A - \bar{x}_N)_U = 0 + z_{.02}\hat{\sigma}_{\bar{x}_A - \bar{x}_N} = 0 + 2.05(0.3888) = 0.7970$$

Since $z = \dfrac{(\bar{x}_A - \bar{x}_N) - (\bar{\mu}_A - \bar{\mu}_N)}{\hat{\sigma}_{\bar{x}_A - \bar{x}_N}} = \dfrac{(3.37 - 3.15) - 0}{0.3888} = 0.57 < 2.05$ (or $\bar{x}_A - \bar{x}_N = 0.22 < 0.7970$), we do not reject H_0. The ASW students do not have significantly higher GPA's than non-ASW students. However, this does not say that the ASW students didn't do better than they would have done had they not enrolled in ASW, so it says nothing about whether Andy should advertise that ASW helps student achievement.

9-52 1995: $n_{95} = 2000$ $\bar{p}_{95} = \frac{58}{2000} = 0.029$

1994: $n_{94} = 2500$ $\bar{p}_{94} = \frac{61}{2500} = 0.0244$

$H_0 : p_{95} = p_{94}$ $H_1 : p_{95} \neq p_{94}$ $\alpha = .01$

$$\bar{p} = \frac{n_{95}\bar{p}_{95} + n_{94}\bar{p}_{94}}{n_{95} + n_{94}} = \frac{2000(.029) + 2500(.0244)}{2000 + 2500} = 0.0264$$

$$\hat{\sigma}_{\bar{p}_{95} - \bar{p}_{94}} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_{95}} + \frac{1}{n_{94}}\right)} = \sqrt{.0264(.9736)\left(\frac{1}{2000} + \frac{1}{2500}\right)} = 0.0048$$

The limits of the acceptance region are $z_{CRIT} = z_{.005} = \pm 2.576$, or

$$(\bar{p}_{95} - \bar{p}_{94})_{CRIT} = 0 \pm z_{.01}\hat{\sigma}_{\bar{p}_{95} - \bar{p}_{94}} = 0 \pm 2.576(.0048) = \pm 0.0124$$

Since $z = \frac{\bar{p}_{95} - \bar{p}_{94}}{\hat{\sigma}_{\bar{p}_{95} - \bar{p}_{94}}} = \frac{.029 - .0244}{.00481} = 0.9563 < 2.576$ (or $\bar{p}_{95} - \bar{p}_{94} = 0.0046 < 0.0124$), we fail to reject H_0. The proportion of 1995 tax returns that were audited was not significantly different than the proportion of 1994 tax returns that were audited.

9-53

Chapel Hill	97.3	108.4	135.7	142.3	151.8	158.5	177.4	183.9	195.2	207.6
Durham Co.	81.5	92.0	115.8	137.8	150.9	149.2	168.2	173.9	175.9	194.4
Difference	15.8	16.4	19.9	4.5	0.9	9.3	9.2	10.0	19.3	13.2

$$\bar{x} = \frac{\sum x}{n} = \frac{118.5}{10} = 11.85$$

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{9}\left(1753.53 - 10(11.85)^2\right) = 38.8116$$

$s = \sqrt{s^2} = 6.2299$

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 6.2299/\sqrt{10} = 1.9701$

$H_0 : \mu_C - \mu_D = 15$ $H_1 : \mu_C - \mu_D \neq 15$ $\alpha = .05$

The limits of the acceptance region are $t_{CRIT} = \pm t_{9,\ .025} = \pm 2.262$, or

$$\bar{x}_{CRIT} = \mu \pm t_{9,\ .025}\hat{\sigma}_{\bar{x}} = 15 \pm 2.262(1.9701) = (10.5436,\ \ 19.4564)$$

Since $t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{11.85 - 15}{1.9700} = -1.599 > -2.262$ (or $\bar{x} = 11.85$), we do not reject H_0. The data support Ellen's claim.

9-54 Differences between prices on 5/21/93 and 5/24/93 yield the following:
$s = 0.55232$, n= 20, $\bar{x} = -0.00625$

$\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 0.55232/\sqrt{20} = 0.1235$

$H_0 : \mu_{21} = \mu_{24}$ $H_1 : \mu_{21} > \mu_{24}$ (α not given)

Since $t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}} = \frac{-0.00625 - 0}{0.1235} = -0.0506$ is so close to 0, we would not reject H_0 at any reasonable significance level. The observed decrease is not significant.

9-55 a) 20 - 29: $n = 7$ $\sum x = 537$ $\sum x^2 = 41475$

$$\bar{x}_1 = \frac{\sum x}{n} = \frac{537}{7} = 76.7143$$

$$s_1^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{6}\left(41475 - 7(76.7143)^2\right) = 46.5689$$

≥ 30: $\quad n = 6 \quad \sum x = 431 \quad \sum x^2 = 31131$

$$\bar{x}_2 = \frac{\sum x}{n} = \frac{431}{6} = 71.8333$$

$$s_2^2 = \frac{1}{5}\left(31131 - 6(71.8333)^2\right) = 34.1724$$

$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2 \qquad \alpha = .05$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{6(46.5689) + 5(34.1724)}{11}} = 6.3980$$

The limits of the acceptance region are $t_{CRIT} = \pm t_{11,\ .025} = \pm 2.201$, or

$$(\bar{x}_1 - \bar{x}_2)_{\text{crit}} = 0 \pm t_{11,.025}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0 \pm 2.201(6.3980)\sqrt{\frac{1}{7} + \frac{1}{6}} = \pm 7.8345$$

Since $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{(76.7143 - 71.8333) - 0}{6.3980\sqrt{\frac{1}{7} + \frac{1}{6}}} = 1.371 < 2.201$ (or $\bar{x}_1 - \bar{x}_2 = 4.8810 < 7.8345$), we do not reject H_0. The show is not significantly different in its appeal to both age groups.

b) It is highly doubtful that Terri's convenience sample of the folks in her office is really representative of the population as a whole. She should not use the results of the office survey to plan the advertising campaign.

9-56 $\quad n_C = 280 \qquad \bar{p}_C = \frac{152}{280} = 0.5429 \qquad n_D = 190 \qquad \bar{p}_D = \frac{81}{190} = 0.4263$

$H_0: \bar{p}_C = \bar{p}_D \qquad H_1: \bar{p}_C > \bar{p}_D \qquad \alpha = 0.02$

$$\hat{p} = \frac{n_C\bar{p}_C + n_D\bar{p}_D}{n_C + n_D} = \frac{280(.5429) + 190(.4263)}{280 + 190} = 0.4958$$

$$\hat{\sigma}_{\bar{p}_C - \bar{p}_D} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_C} + \frac{1}{n_D}\right)} = \sqrt{.4958(.5042)\left(\frac{1}{280} + \frac{1}{190}\right)} = 0.047$$

The upper limit of the acceptance region is $z_U = z_{.02} = 2.05$, or

$$(\bar{p}_C - \bar{p}_D)_U = 0 + z_{.02}\hat{\sigma}_{\bar{p}_C - \bar{p}_D} = 0 + 2.05(.047) = 0.0964$$

Since $z = \dfrac{\bar{p}_C - \bar{p}_D}{\hat{\sigma}_{\bar{p}_C - \bar{p}_D}} = \dfrac{.5429 - .4263}{.0470} = 2.48 > 2.05$ (or $\bar{p}_C - \bar{p}_D = 0.1166 > 0.047$), we reject H_0. Cat owners are significantly more likely to feed their pets premium food than dog owners.

9-57 $\quad s_P = 0.43 \qquad n_P = 36 \qquad \bar{x}_P = 0.98 \qquad s_B = 0.38 \qquad n_B = 42 \qquad \bar{x}_B = 1.07$

$H_0: \mu_P = \mu_B \qquad H_1: \mu_P < \mu_B \qquad \alpha = .01$

$$\hat{\sigma}_{\bar{x}_P - \bar{x}_B} = \sqrt{\frac{s_P^2}{n_P} + \frac{s_B^2}{n_B}} = \sqrt{\frac{.43^2}{36} + \frac{.38^2}{42}} = 0.0926$$

The lower limit of the acceptance region is $z_L = -z_{.01} = -2.33$, or

$$(\bar{x}_P - \bar{x}_B)_L = (\mu_P - \mu_B)_{H_0} - z_{.01}\hat{\sigma}_{\bar{x}_P - \bar{x}_B} = 0 - 2.33(0.0926) = -0.2158$$

Since $z = \frac{(\bar{x}_P - \bar{x}_B) - (\bar{\mu}_P - \bar{\mu}_B)}{\hat{\sigma}_{\bar{x}_P - \bar{x}_B}} = \frac{(0.98 - 1.07) - 0}{0.0926} = -0.97 > -2.33$ (or $\bar{x}_P - \bar{x}_B = -0.09 > -0.2158$), we do not reject H_0. Boston is not significantly more expensive. However, a better test of these hypotheses could be done if they bought the same items in each city and did a paired-sample test. Furthermore, cost of living depends on many more things than small-ticket items in grocery stores, for instance, taxes, rents, cost of gasoline, and utility rates.

9-58 a) $59(.45) = 25.55$ responses, which is impossible!
Note that $26/59 = .4407$, but $27/59 = .4576$.

b) Looking at the greatest difference in response rates (with 26 U.K. responses), we have

$n_{US} = 100 \qquad \bar{p}_{US} = \frac{50}{100} = 00.5 \qquad n_{UK} = 59 \qquad \bar{p}_{UK} = \frac{26}{59} = .4407$

$H_0 : \bar{p}_{US} = \bar{p}_{UK} \qquad H_1 : \bar{p}_{US} \neq \bar{p}_{UK} \qquad$ no α is given

$$\hat{p} = \frac{n_{US}\bar{p}_{US} + n_{UK}\bar{p}_{UK}}{n_{US} + n_{UK}} = \frac{100(0.5) + 59(.4407)}{100 + 59} = 0.4780$$

$$\hat{\sigma}_{\bar{p}_{US} - \bar{p}_{UK}} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_{US}} + \frac{1}{n_{UK}}\right)} = \sqrt{.4780(.5220)\left(\frac{1}{100} + \frac{1}{59}\right)} = 0.0820$$

Now $z = \frac{\bar{p}_{US} - \bar{p}_{UK}}{\hat{\sigma}_{\bar{p}_{US} - \bar{p}_{UK}}} = \frac{0.5 - .4407}{.0820} = 0.7232$, which corresponds to a probability value of $2(0.5 - 0.2642) = 0.4716$, which is very insignificant. The proportions of responses in the two surveys are not significantly different.

9-59 a) Looking at all figures in \$1000 and using a calculator to obtain sample means and variances we have:

Sample 1 (East and South): $s_E^2 = 3172.94 \qquad n_E = 11 \qquad \bar{x}_E = 192.786$

Sample 2 (Midwest and West): $s_W^2 = 1265.42 \qquad n_W = 10 \qquad \bar{x}_W = 167.508$

$H_0 : \mu_E = \mu_W \qquad H_1 : \mu_E \neq \mu_W \qquad$ no α is specified

$$s_p = \sqrt{\frac{(n_E - 1)s_E^2 + (n_W - 1)s_W^2}{n_E + n_W - 2}} = \sqrt{\frac{10(3172.94) + 9(1265.42)}{19}} = 47.64$$

The observed $t = \frac{(\bar{x}_E - \bar{x}_W) - (\bar{\mu}_E - \bar{\mu}_W)}{s_p\sqrt{\frac{1}{n_E} + \frac{1}{n_W}}} = \frac{(192.786 - 167.508) - 0}{47.64\sqrt{\frac{1}{11} + \frac{1}{10}}} = 1.214.$

With 19 degrees of freedom and $\alpha = 0.10$, $t_{CRIT} = \pm 1.729$, so even at a 10% significance level, mean housing prices are not significantly different in the two regions. (Using Minitab, the exact prob value for our test is 0.2395, further strengthening our conclusion.)

b) If we throw out the outliers, we have:

$s_E^2 = 1415.68 \qquad n_E = 10 \qquad \bar{x}_E = 179.648$

$s_W^2 = 693.41 \qquad n_W = 9 \qquad \bar{x}_W = 175.564$

With the two sample means so close to each other, one would guess that the difference is even less significant that it was before. In fact, Minitab now gives 0.7897 as the prob value for our test!

CHAPTER 10

QUALITY AND QUALITY CONTROL

10-1 There are many potential answers to this question. Any example that makes a distinction between the concepts of luxury and quality would be appropriate. For example, many luxury and sports cars have been known to suffer from quality problems which required recalls. Also, products using new electronic technologies (such as PC's using a new and faster computer CPU chip) often suffer from quality problems until the bugs are worked out -- they also tend to be very expensive when first introduced.

10-2 As in 10-1, there are many potential answers to this question. Any example that demonstrates how a low cost item works reliably or lacks defects would be appropriate. Examples include paper clips, shoe laces, toothpicks, coins, or pencils.

10-3 "Quality" is the condition where goods or services conform to requirements such that they are consistent, reliable and lack errors or defects.

10-4 Quality control is an important issue to management because good managers want to keep their customers satisifed. From another perspective, quality impacts the "bottom line" -- profits. Parts that don't conform to requirements end up as scrap or rejects, which is costly. It is also costly to inspect for defects. By controlling the quality of the parts used in production at all stages of the process, costs can be minimized.

10-5 An analysis of the relative merits of "inspection" vs. "prevention" control, would require one to consider the costs of labor for each method (e.g., costs associated with hiring full-time inspectors compared to the costs associated with the preventive practice of employee inspection at each stage of the production process) and the costs of defective products (e.g., recalls, public anger) that result from each type of control.

10-6 "Zero defects" is a goal in the production process, where defects are eliminated at each stage of the production, so that the final product is a perfectly reliable, quality product.

10-7 The Japanese who benefited from Deming's ideas, presented formidable competitive pressure to American companies, especially in the automotive and consumer electronics industries.

10-8 Mechanical devices, such as a robot, are created with a specific and limited range of motion. Materials and parts used to build such machines are chosen to insure uniformity and reliability. Robots are therefore, easily controlled and adjusted. In contrast, humans are capable of a great variety of motions and react to all sorts of stimuli (both related and unrelated to the task at hand) which affects behavior. It is expected that for those reasons, humans would introduce more random variation into the work process.

10-9 When the manager replaces a pitcher in a baseball game, he is probably responding to nonrandom variation in the pitcher's performance. The pitcher may be tiring, causing his pitches to be slower or his curve balls to "break" less -- making it easier for the batter to hit the ball. He may also lose some control resulting in a greater number of "walked" batters. Sometimes pitcher's develop bad habits in their pitching motions that leads to nonrandom, poor performance.

10-10 The manager would be trying to control the amount of time to process each customer. By directing customers with just a few items to the express line, you would in effect increase the consistency in the times taken to process each customer, since the registers would be handling customers with a similar number of items. Without express lines, one particular register would process customers with greatly varying numbers of items which would cause greater variation in processing time.

10-11 1) Individual outliers: In the plastic piping industry, PVC pipes are cut to a certain length in order to be pressed into flanges. On some days (for any of a variety of reasons) the pipes may be cut to lengths that vary considerably from the specified length.

2) Decreasing trend: Clerks who fill orders in a warehouse for shipping often must fill a certain number to meet a daily standard. New clerks may work hard early in the job to insure that they meet the required work standard, and often discover that they surpass the standards. But as they grow comfortable with the job, they may gradually relax their efforts and filled orders will decline. This may also happen when clerks get bored with the job or angry with company.

3) Jumps in the level around which the observations vary: Various supervisors may differ in their ability to motivate workers. When one supervisor replaces another, it is common to see the average performance shift.

4) Cycles: Physically challenging jobs such as truck loading at United Parcel Service may show cycling -- as workers get fatigued, performance declines (fewer boxes per minute loaded). But after a break, performance may increase until the worker tires again, causing another performance decline.

10-12 a) $\overline{\overline{x}} = 16.4$ $\sigma_{\overline{x}} = 1.2$

$\text{CL} = \overline{\overline{x}} = 16.4$

$\text{UCL} = \overline{\overline{x}} + 3\sigma_{\overline{x}} = 16.4 + 3(1.2) = 20.0$

$\text{LCL} = \overline{\overline{x}} - 3\sigma_{\overline{x}} = 16.4 - 3(1.2) = 12.8$

b) $\overline{\overline{x}} = 16.4$ $\overline{R} = 7.6$ $n = 12$ $d_2 = 3.258$

$\text{CL} = \overline{\overline{x}} = 16.4$

$$\text{UCL} = \overline{\overline{x}} + \frac{3(\overline{R})}{d_2\sqrt{n}} = 16.4 + \frac{3(7.6)}{3.258\sqrt{12}} = 18.42$$

$$\text{LCL} = \overline{\overline{x}} - \frac{3(\overline{R})}{d_2\sqrt{n}} = 16.4 - \frac{3(7.6)}{3.258\sqrt{12}} = 14.38$$

c) $\overline{\overline{x}} = 4.1$ $\overline{R} = 1.3$ $n = 8$ $d_2 = 2.847$

$\text{CL} = \overline{\overline{x}} = 4.1$

$$\text{UCL} = \overline{\overline{x}} + \frac{3(\overline{R})}{d_2\sqrt{n}} = 4.1 + \frac{3(1.3)}{2.847\sqrt{8}} = 4.58$$

$$\text{LCL} = \overline{\overline{x}} - \frac{3(\overline{R})}{d_2\sqrt{n}} = 4.1 - \frac{3(1.3)}{2.847\sqrt{8}} = 3.62$$

d) $\overline{\overline{x}} = 141.7$ $\overline{R} = 18.6$ $n = 15$ $d_2 = 3.472$

$\text{CL} = \overline{\overline{x}} = 141.7$

$$\text{UCL} = \overline{\overline{x}} + \frac{3(\overline{R})}{d_2\sqrt{n}} = 141.7 + \frac{3(18.6)}{3.472\sqrt{15}} = 145.85$$

$$\text{LCL} = \overline{\overline{x}} - \frac{3(\overline{R})}{d_2\sqrt{n}} = 141.7 - \frac{3(18.6)}{3.472\sqrt{15}} = 137.55$$

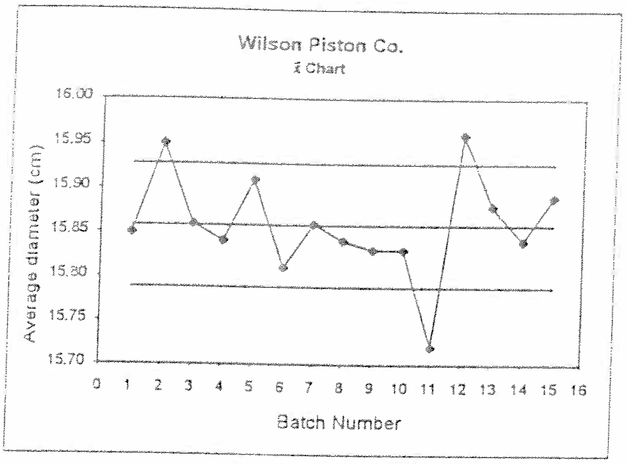
10-13 a) $n = 8$ $k = 15$ $d_2 = 2.847$

$$\bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{237.87}{15} = 15.858 \qquad \bar{R} = \frac{\sum R}{k} = \frac{2.81}{15} = 0.187$$

$$\text{CL} = \bar{\bar{x}} = 15.858$$

$$\text{UCL} = \bar{\bar{x}} + \frac{3(\bar{R})}{d_2\sqrt{n}} = 15.858 + \frac{3(0.187)}{2.847\sqrt{8}} = 15.928$$

$$\text{LCL} = \bar{\bar{x}} - \frac{3(\bar{R})}{d_2\sqrt{n}} = 15.858 - \frac{3(0.187)}{2.847\sqrt{8}} = 15.788$$

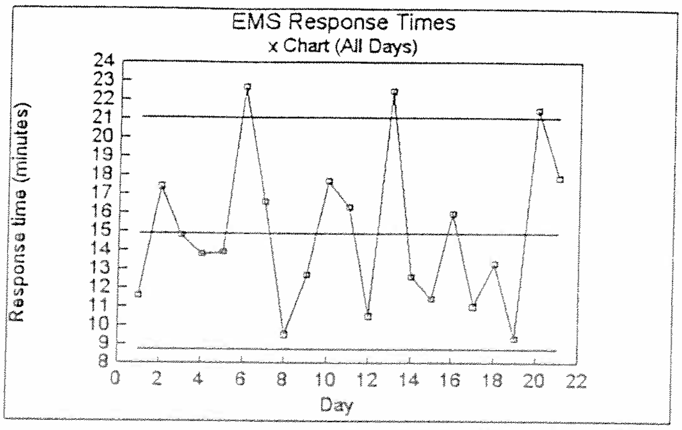


b) The production process appears to be out-of-control. Batches 2, 11 and 12 fall outside of the control limits.

10-14 a) $n = 9$ $k = 21$ $A_2 = 0.337$

$$\bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{312.999}{21} = 14.905 \qquad \bar{R} = \frac{\sum R}{k} = \frac{385.300}{21} = 18.348$$

$$\text{CL} = \bar{\bar{x}} = 14.905$$
$$\text{UCL} = \bar{\bar{x}} + A_2(\bar{R}) = 14.905 + 0.337(18.348) = 21.09$$
$$\text{LCL} = \bar{\bar{x}} - A_2(\bar{R}) = 14.905 - 0.377(18.348) = 8.72$$

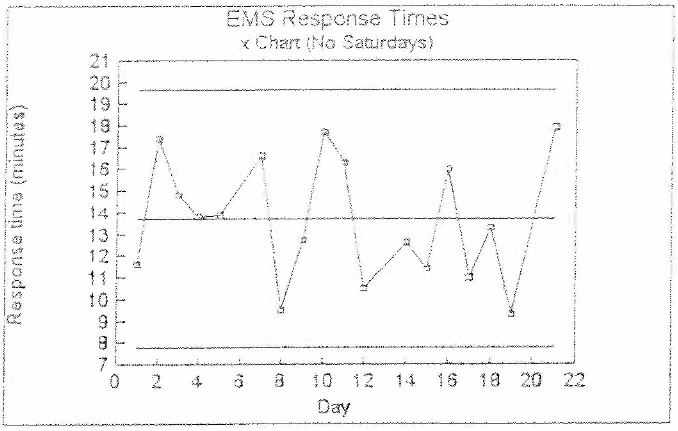


b) It seems that response times are out of control (outliers on the high side) on Saturdays. Dick should investigate whether there are any special circumstances that tend to repeat on Saturdays. For example, there might be more calls coming in on Saturdays which burden the capabilities of the rescue squads. To counteract this problem, Dick might consider increasing the number of squads on call on Saturdays, or he might provide additional training to the crews on Saturdays.

c) No Saturdays:

$$\bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{287.343}{21} = 13.683 \qquad \bar{R} = \frac{\sum R}{k} = \frac{396.600}{21} = 17.6$$

$CL = \bar{\bar{x}} = 13.683$
$UCL = \bar{\bar{x}} + A_2(\bar{R}) = 13.683 + 0.337(17.6) = 19.61$
$LCL = \bar{\bar{x}} - A_2(\bar{R}) = 13.683 - 0.337(17.6) = 7.75$

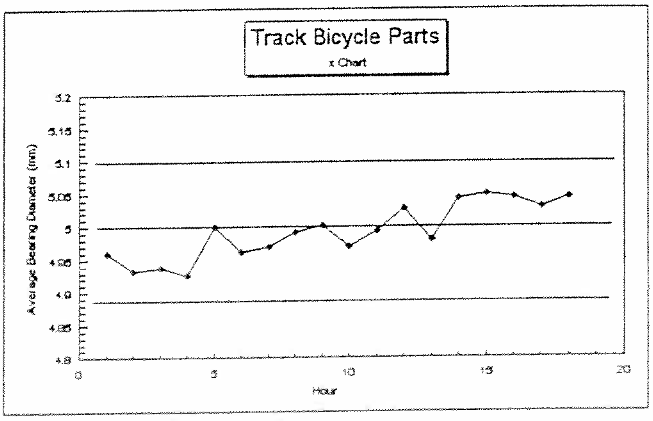


Although some variability remains, the response times are within the control limits when data for Saturdays are removed.

10-15 a) $n = 5 \qquad k = 18 \qquad A_2 = 0.557$

$$\bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{89.82}{18} = 4.99 \qquad \bar{R} = \frac{\sum R}{k} = \frac{3.24}{18} = 0.18$$

$CL = \bar{\bar{x}} = 4.99$
$UCL = \bar{\bar{x}} + A_2(\bar{R}) = 4.99 + 0.557(0.18) = 5.09$
$LCL = \bar{\bar{x}} - A_2(\bar{R}) = 4.99 - 0.557(0.18) = 4.89$

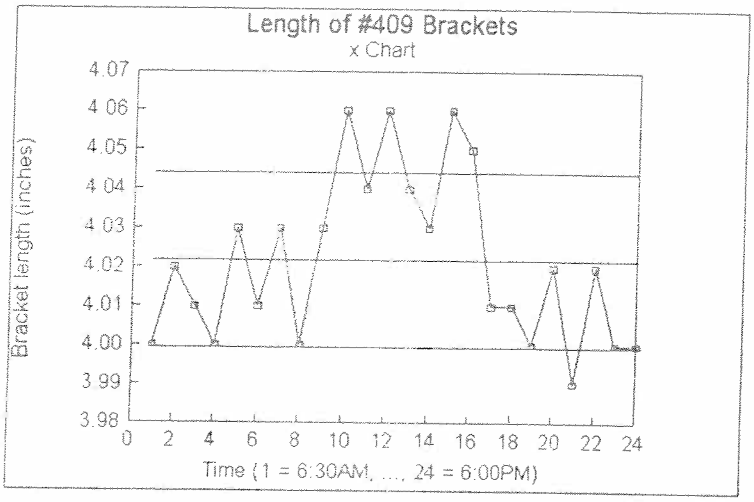


b) Although all observations fall within the control limits, there is a gradual trend in which the diameter of the bearings appears to be increasing, so the process is out-of-control.

10-16 a) $n = 15 \qquad k = 24 \qquad A_2 = 0.233$

$$\bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{96.5208}{24} = 4.0217 \qquad \bar{R} = \frac{\sum R}{k} = \frac{2.4192}{24} = 0.1008$$

$CL = \bar{\bar{x}} = 4.0217$
$UCL = \bar{\bar{x}} + A_2(\bar{R}) = 4.0217 + 0.233(0.1008) = 4.045$
$LCL = \bar{\bar{x}} - A_2(\bar{R}) = 4.0217 - 0.233(0.1008) = 3.998$



b) The three shifts seem to be at different levels, with the bracket lengths in the second shift higher than in the first and third shifts. Silvia should check out the procedures for recalibrating the saw at the beginning of each shift

10-17 a) $n = 3$ $\quad \overline{R} = 3.1$ $\quad D_4 = 2.574$ $\quad D_3 = 0$

$CL = \overline{R} = 3.1$
$UCL = \overline{R}\, D_4 = 3.1(2.574) = 7.98$
$LCL = \overline{R}\, D_3 = 3.1(0) = 0$

b) $n = 19$ $\quad \overline{R} = 6.9$ $\quad D_4 = 0.403$ $\quad D_3 = 1.597$

$CL = \overline{R} = 6.9$
$UCL = \overline{R}\, D_4 = 6.9(1.597) = 11.02$
$LCL = \overline{R}\, D_3 = 6.9(0.403) = 2.78$

c) $n = 8$ $\quad \overline{R} = 18.2$ $\quad D_4 = 1.864$ $\quad D_3 = 0.136$

$CL = \overline{R} = 18.2$
$UCL = \overline{R}\, D_4 = 18.2(1.864) = 33.92$
$LCL = \overline{R}\, D_3 = 18.2(0.136) = 2.48$

d) $n = 24$ $\quad \overline{R} = 1.4$ $\quad D_4 = 1.548$ $\quad D_3 = 0.452$

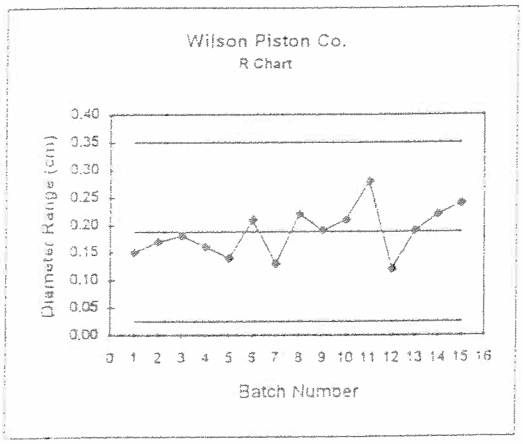
$CL = \overline{R} = 1.4$
$UCL = \overline{R}\, D_4 = 1.4(1.548) = 2.17$
$LCL = \overline{R}\, D_3 = 1.4(0.452) = 0.63$

e) UCL = 9, because it is as far above $\overline{R}$ as the LCL is below $\overline{R}$.

10-18 Since new apprentices are expected to make errors, there should be greater variability in the spindles they produce initially. This variability should decrease as the new apprentice gains experience. Pattern (a) conforms to this expectation.

10-19 $n = 8$ $\quad D_4 = 1.864$ $\quad D_3 = 0.136$

$\overline{R} = 0.187$
$CL = \overline{R} = 0.187$
$UCL = \overline{R}\, D_4 = 0.187(1.864) = 0.349$
$LCL = \overline{R}\, D_3 = 0.187(0.136) = 0.025$

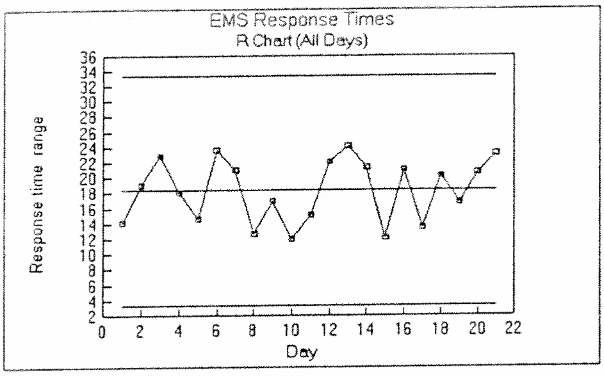


All values of R are within the control limits, and no patterns are apparent, so the variability is in-control.

10-20 a) $n = 9$ $\quad k = 21$ $\quad D_4 = 1.816$ $\quad D_3 = 0.184$

$$\overline{R} = \frac{\sum R}{k} = \frac{\sum 385.3}{21} = 18.348$$

$\text{CL} = \overline{R} = 18.348$
$\text{UCL} = \overline{R}\ D_4 = 18.348(1.816) = 33.32$
$\text{LCL} = \overline{R}\ D_3 = 18.348(0.184) = 3.38$



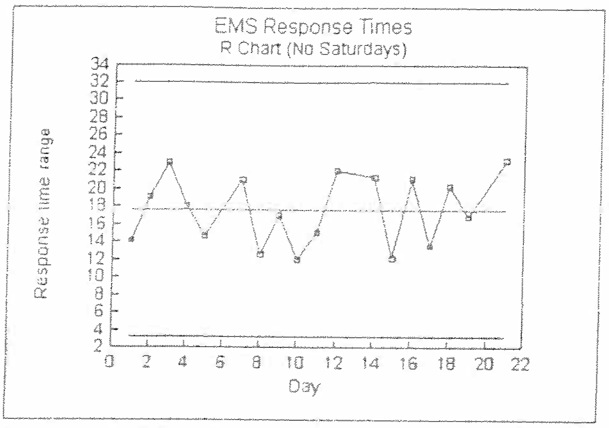
b) The Saturdays are no longer outliers, but they do tend to have higher variability than the other days. This could well arise because of the greater number of calls coming in on Saturdays.

c) No Saturdays:

$n = 9$ $\quad k = 18$ $\quad D_4 = 1.816$ $\quad D_3 = 0.184$

$$\overline{R} = \frac{\sum R}{k} = \frac{\sum 316.8}{18} = 17.6$$

$\text{CL} = \overline{R} = 17.6$
$\text{UCL} = \overline{R}\ D_4 = 17.6(1.816) = 31.96$
$\text{LCL} = \overline{R}\ D_3 = 17.6(0.184) = 3.24$

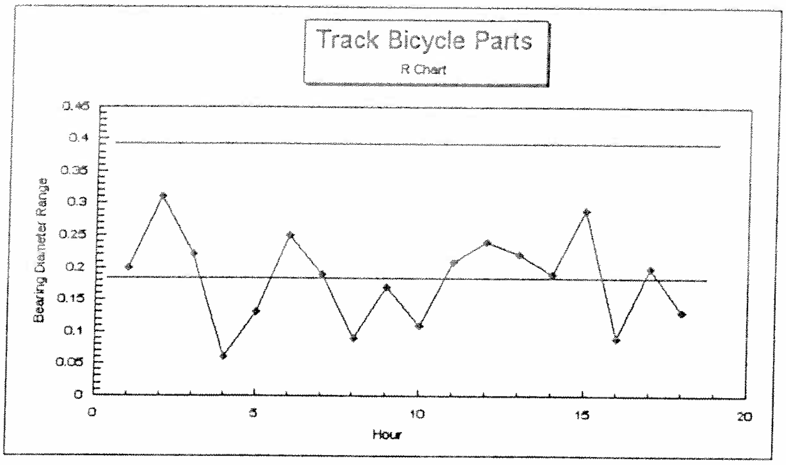


The variability in response times is in-control.

10-21 $n = 5$ $k = 18$ $D_4 = 2.114$ $D_3 = 0$ $\overline{R} = \frac{\sum R}{k} = \frac{3.24}{18} = 0.18$

$CL = \overline{R} = 0.18$ $UCL = \overline{R}\,D_4 = 0.18(2.114) = 0.3798$

$LCL = \overline{R}\,D_3 = 0.18(0) = 0$

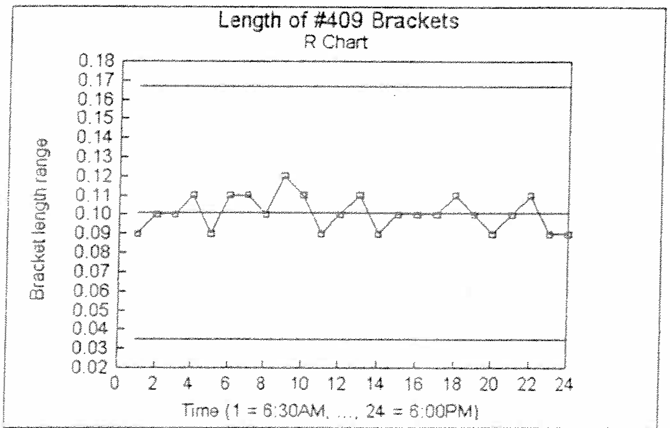


The variability in bearing diameter is in-control.

10-22 $n = 15$ $k = 24$ $D_4 = 1.653$ $D_3 = 0.347$ $\overline{R} = 0.1008$

$CL = \overline{R} = 0.1008$ $UCL = \overline{R}\,D_4 = 0.1008(1.653) = 0.167$

$LCL = \overline{R}\,D_3 = 0.1008(0.347) = 0.035$



The variability in the process appears to be well in control.

10-23 (b) and (c) are attributes. (a) and (d) are not attributes.

10-24 a) $\text{CL} = \overline{\overline{p}} = 0.25$

$$\text{UCL} = \overline{\overline{p}} + 3\sqrt{\frac{pq}{n}} = 0.25 + 3\sqrt{\frac{.25(.75)}{30}} = 0.487$$

$$\text{LCL} = \overline{\overline{p}} - 3\sqrt{\frac{pq}{n}} = 0.25 - 3\sqrt{\frac{.25(.75)}{30}} = 0.013$$

b) $\text{CL} = \overline{\overline{p}} = 0.15$

$$\text{UCL} = \overline{\overline{p}} + 3\sqrt{\frac{pq}{n}} = 0.15 + 3\sqrt{\frac{.15(.85)}{65}} = 0.283$$

$$\text{LCL} = \overline{\overline{p}} - 3\sqrt{\frac{pq}{n}} = 0.15 - 3\sqrt{\frac{.15(.85)}{65}} = 0.017$$

c) $\text{CL} = \overline{\overline{p}} = 0.05$

$$\text{UCL} = p + 3\sqrt{\frac{pq}{n}} = 0.05 + 3\sqrt{\frac{.05(.95)}{82}} = 0.122$$

$$\text{LCL} = p - 3\sqrt{\frac{pq}{n}} = 0.05 - 3\sqrt{\frac{.05(.95)}{82}} < 0, \text{ so the LCL} = 0$$

d) $\text{CL} = p = 0.42$

$$\text{UCL} = p + 3\sqrt{\frac{pq}{n}} = 0.42 + 3\sqrt{\frac{.42(.58)}{97}} = 0.570$$

$$\text{LCL} = p - 3\sqrt{\frac{pq}{n}} = 0.42 - 3\sqrt{\frac{.42(.58)}{97}} = 0.270$$

e) $\text{CL} = \overline{\overline{p}} = 0.63$

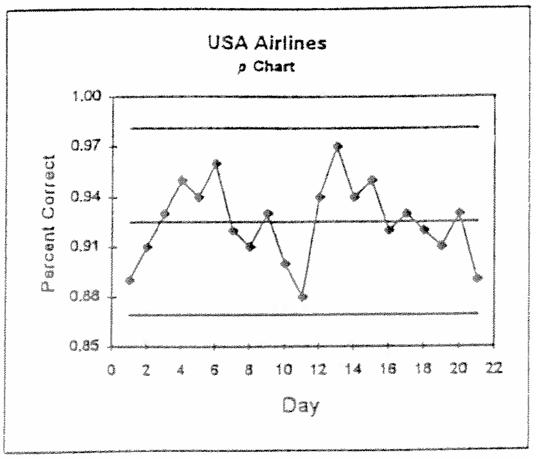
$$\text{UCL} = \overline{\overline{p}} + 3\sqrt{\frac{pq}{n}} = 0.63 + 3\sqrt{\frac{.63(.37)}{124}} = 0.760$$

$$\text{LCL} = \overline{\overline{p}} - 3\sqrt{\frac{pq}{n}} = 0.63 - 3\sqrt{\frac{.63(.37)}{124}} = 0.500$$

10-25 a) $n = 200$ $k = 21$ $\overline{\overline{p}} = \frac{\sum \overline{p}}{k} = \frac{19.42}{21} = 0.925$

$$\text{UCL} = p + 3\sqrt{\frac{pq}{n}} = 0.925 + 3\sqrt{\frac{.925(.075)}{200}} = 0.981$$

$$\text{LCL} = p - 3\sqrt{\frac{pq}{n}} = 0.925 - 3\sqrt{\frac{.925(.075)}{200}} = 0.869$$



b) Although all the observations are within the control limits, the process seems to have a cycle lasting about 10 days—it starts out with a low percentage of luggage delivered correctly, this improves, but then it slips again.

c) Measures should be taken to identify the reason for the cycle. In addition, the overall fraction of luggage being delivered correctly, $\overline{\overline{p}} = 0.925$, is not particularly good, so USA Airlines should try to improve the overall level of the process.

10-26 a) H_0: $p = 0.015$ H_1: $p > 0.015$

$$p = 0.01594 \qquad z = \frac{\overline{p} - \mu_{\overline{p}}}{\sqrt{\frac{pq}{n}}} = \frac{.01594 - .015}{\sqrt{\frac{.015(.985)}{16000}}} = 0.98$$

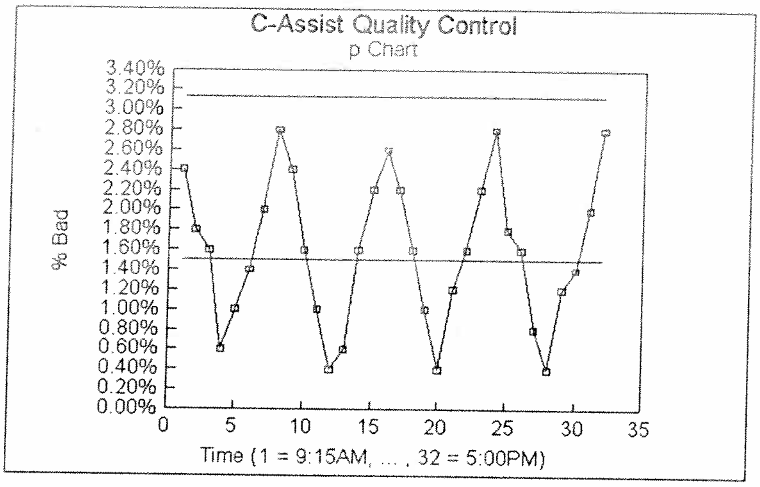
Since the probability value $= 0.5 - 0.3365 = 0.1635$, we accept the H_0. She can be reasonably sure that the proportion of bad capsules is not significantly greater than 1.5%.

b) $n = 500$ $p = 0.015$

$$CL = p = 0.015$$

$$UCL = p + 3\sqrt{\frac{pq}{n}} = 0.015 + 3\sqrt{\frac{.015(.985)}{500}} = 0.0313$$

$$LCL = p - 3\sqrt{\frac{pq}{n}} = 0.015 - 3\sqrt{\frac{.898(.102)}{150}} = 0$$



c) Although, the samples fall within the control limits, the p chart shows a distinct 2-hour cycle in which the samples of capsules approach the upper control limit. Sherry needs to examine the manufacturing process to determine what is causing the cycling.

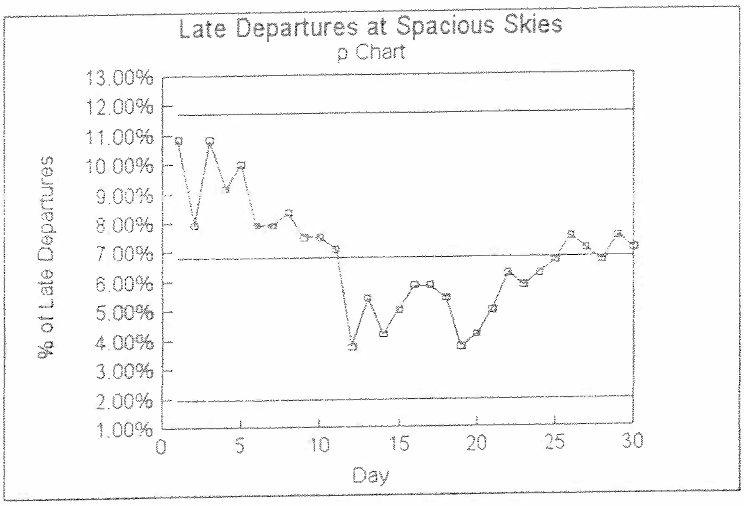
10-27 Andie could use a p chart to determine whether her sampled stocks conform to her belief about stock performance. Since she believes that there is a 50-50 chance of a gain for any given stock, she should should use 0.50 as a standard (the targeted p value). She then would calculate the upper and lower control limits using the following formula: $p \pm 3\sqrt{pq/n}$, which yields upper and lower control limits of 0.65 and 0.35, respectively. The next step is to plot the daily, sample fraction data to determine whether or not her samples conform to her hypothesis.

10-28 $n = 240$ $k = 30$ $\overline{\overline{p}} = \frac{\sum \overline{p}}{k} = \frac{2.043}{30} = 0.0681$

$$CL = \overline{\overline{p}} = 0.0681$$

$$UCL = \overline{\overline{p}} + 3\sqrt{\frac{pq}{n}} = 0.0681 + 3\sqrt{\frac{.0681(.9319)}{240}} = 0.1169$$

$$LCL = \overline{\overline{p}} - 3\sqrt{\frac{pq}{n}} = 0.0681 - 3\sqrt{\frac{.0681(.9319)}{240}} = 0.0193$$

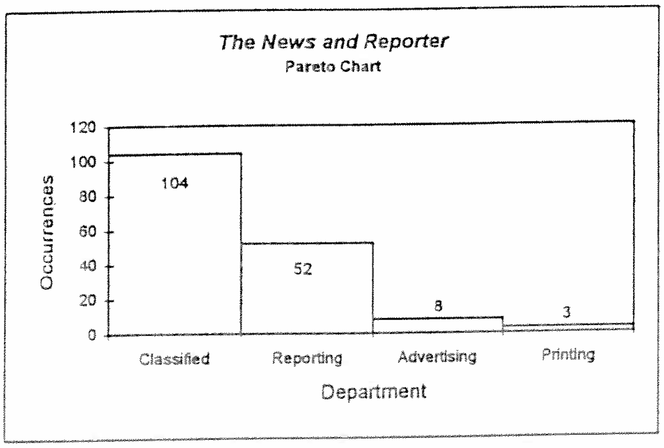


Four weeks ago, the fraction of late departures dropped dramatically, presumably in response to Ross' new procedures. However, in the past two weeks, that fraction has again started drifting upward. If the new procedures aren't being used, he should insist that they be used. If they are being used, he needs to find out why they aren't working.

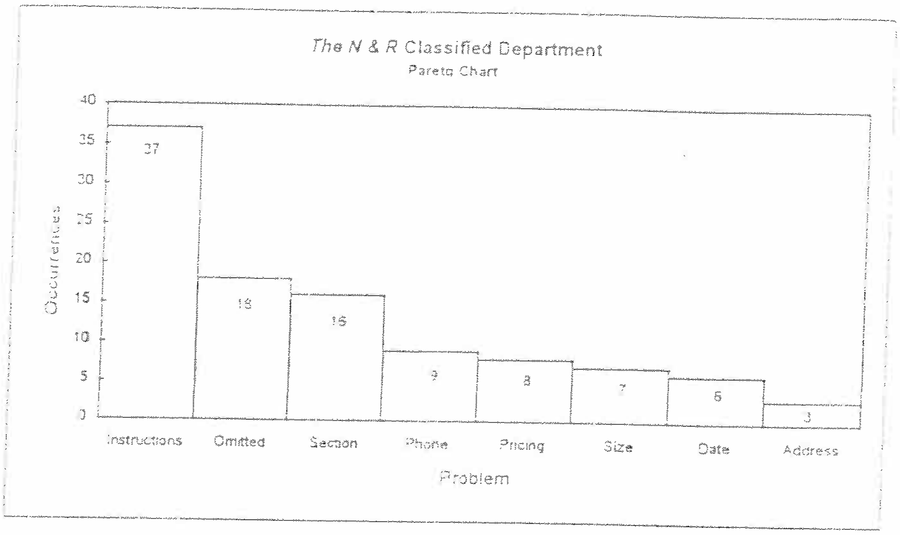
10-29 In most any production process, there are likely to be a great number of potential causes of errors or defects in the final product. Since personnel at all levels contribute to this process, whether directly or indirectly, they do affect the ultimate quality of the product. Employees at all levels of the organization must therefore be involved in eliminating problems.

10-30 After the most important problems are "slain", other problems will likely become "dragons", especially if they are ignored. It should be stressed to Joe, that for TQM to succeed, there must be a focus on "continuous quality improvement."

10-31

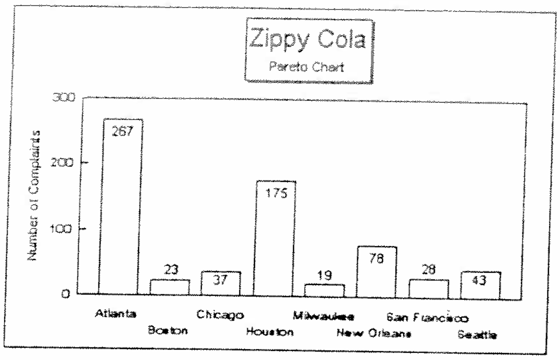


We see from our first chart that the Classified Department needs the most attention. Now let's see what's happening within that department:



Within the Classified Department, the first dragon to slay is the problem of incorrect special instructions.

10-32



The plants in Atlanta and Houston should be visited first.

10-33 Student responses will vary.

10-34 Total inspection is frequently impractical because of the time and cost involved.

10-35 It represents the maximum number of defective pieces in a sample that will be allowed. If there are more defective pieces than c, then the entire lot will be rejected and the supplier will have to absorb the cost.

10-36 AQL = 0.02. Numerical results obtained from an Excel spreadsheet.

a) $n = 175,\ c = 3$: $P(r \geq 4) = 0.46446$
b) $n = 175,\ c = 5$: $P(r \geq 6) = 0.14037$
c) $n = 250,\ c = 3$: $P(r \geq 4) = 0.73781$
d) $n = 250,\ c = 5$: $P(r \geq 6) = 0.38403$

10-37 LTPD = 0.03. Numerical results obtained from an Excel spreadsheet.

a) $n = 175,\ c = 3$: $P(r \leq 3) = 0.22734$
b) $n = 175,\ c = 5$: $P(r \leq 5) = 0.57153$
c) $n = 250,\ c = 3$: $P(r \leq 3) = 0.05652$
d) $n = 250,\ c = 5$: $P(r \leq 5) = 0.23727$

10-38 a) AQL = 0.005
$p = 0.5\%$
The probability of acceptance (Y axis) $\approx$ 0.87
Therefore, the Producer's Risk $\approx 1 - .87 = 0.13$

b) AQL = 0.010
$p = 1.0\%$
The probability of acceptance (Y axis) $\approx$ 0.54
Therefore, the Producer's Risk $\approx 1 - .54 = 0.46$

c) AQL = 0.015
$p = 1.5\%$
The probability of acceptance (Y axis) $\approx$ 0.28
Therefore, the Producer's Risk $\approx 1 - .28 = 0.72$

10-39 a) LTPD = 0.010
$p = 1.0\%$
The probability of acceptance (Y axis) $\approx$ 0.54
Therefore, the Consumer's Risk $\approx$ 0.54

b) LTPD = 0.015
$p = 1.5\%$
The probability of acceptance (Y axis) $\approx$ 0.28
Therefore, the Consumer's Risk $\approx$ 0.28

c) LTPD = 0.020
$p = 2.0\%$
The probability of acceptance (Y axis) $\approx$ 0.12
Therefore, the Consumer's Risk $\approx$ 0.12

10-40 a) H_0: $p = 0.02$ H_1: $p > 0.02$

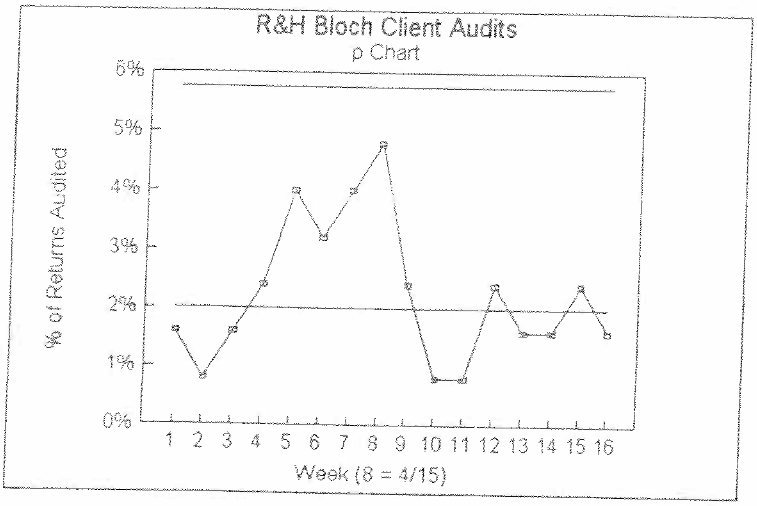
$$p = 0.0225 \qquad z = \frac{\overline{p} - \mu_{\overline{p}}}{\sqrt{\frac{pq}{n}}} = \frac{.0225 - .02}{\sqrt{\frac{.02(.98)}{2000}}} = 0.80$$

Since the probability value $= 0.5 - 0.2881 = 0.2119$, we accept the H_0. She can be reasonably sure that the proportion of audited clients is not significantly greater than 2.0%.

b) $n = 125$ $p = 0.02$
$\text{CL} = p = 0.02$

$$\text{UCL} = p + 3\sqrt{\frac{pq}{n}} = 0.02 + 3\sqrt{\frac{.02(.98)}{125}} = 0.0576$$

$$\text{LCL} = p - 3\sqrt{\frac{pq}{n}} = 0.02 - 3\sqrt{\frac{.02(.98)}{125}} = 0$$



Although, the samples fall within the control limits, the p chart indicates that the percent audited has taken a jump upwards in the four weeks prior to April 15. This may indicate something about the clients who wait until the last minute, or something about how the IRS chooses which returns to audit. In either case, the partners should be aware of this phenomenon.

10-41 When "slaying the dragons", you should be concerned about the "vital few". It has been noted that most problems in complex systems can be attributed to just a few of the potential causes.

10-42 Attributes are categorical variables with only two possible categories.

10-43 As the number of items a teller must process increases, so will the time required to complete the transaction. This is "special cause" variation because it varies systematically with the number of items to be processed.

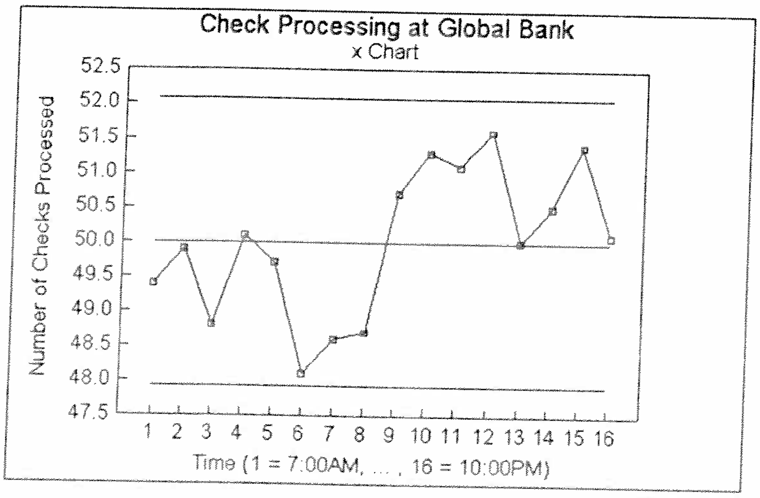
10-44 a) $n = 10 \qquad k = 16 \qquad A_2 = 0.308$

$$\bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{800}{16} = 50 \qquad \bar{R} = \frac{\sum R}{k} = \frac{108}{16} = 6.75$$

$$\text{CL} = \bar{\bar{x}} = 50$$

$$\text{UCL} = \bar{\bar{x}} + A_2(\bar{R}) = 50 + 0.308(6.75) = 52.08$$

$$\text{LCL} = \bar{\bar{x}} - A_2(\bar{R}) = 50 - 0.308(6.75) = 47.92$$



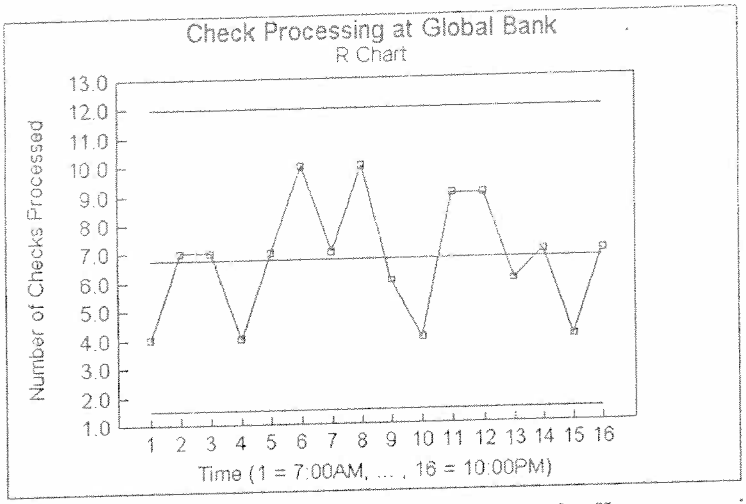
b) The output during the second shift is at a higher level. Shih-Hsing should try to learn why productivity is higher on that shift.

10-45 a) $n = 10$ $\overline{R} = 6.75$ $D_4 = 1.777$ $D_3 = 0.223$

$CL = \overline{R} = 6.75$

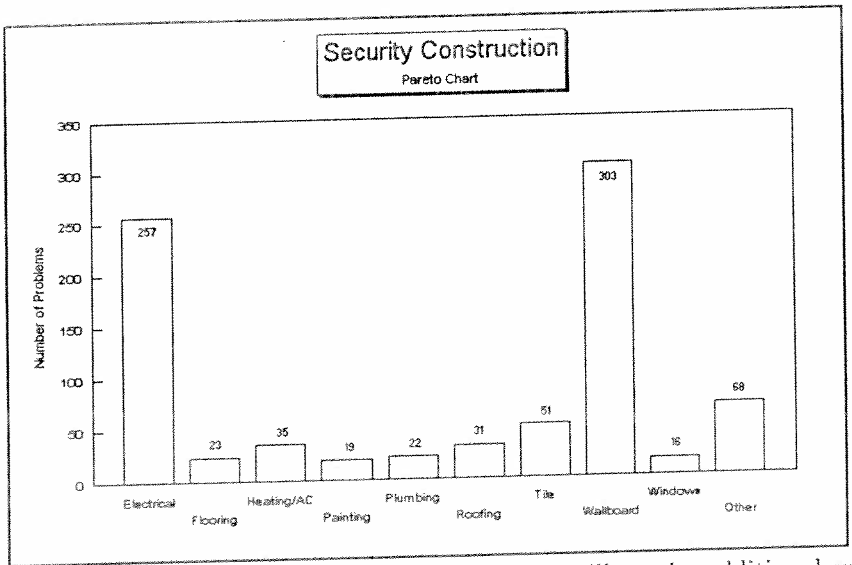
$UCL = \overline{R}\,D_4 = 6.75(1.777) = 11.99$

$LCL = \overline{R}\,D_3 = 6.75(0.223) = 1.51$



The variability in check processing is in-control. However variability does seem to increase around the lunch and dinner hours. Employees might naturally be distracted at these times.

10-46



The wallboard and electrical subcontractors will require additional supervision.

10-47 AQL = 0.01

a) $n = 200$ $c = 1$

$r = 0$: $\dfrac{n!}{r!(n-r)!}\,p^r q^{n-r} = \dfrac{200!}{0!(200)!}\,(.01)^0(.99)^{200} = 1(1)(.1340) = 0.1340$

$r = 1$: $\dfrac{n!}{r!(n-r)!}\,p^r q^{n-r} = \dfrac{200!}{1!(199)!}\,(.01)^1(.99)^{199} = 200(.01)(.1353) = 0.2707$

$1 - .1340 - .2707 = 0.5953$

b) $n = 200 \qquad c = 2$

$r = 2$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{200!}{2!(198)!} (.01)^2(.99)^{198} = 19900(.0001)(.1367) = 0.2720$

$1 - .1340 - .2707 - .2720 = 0.3233$

c) $n = 250 \qquad c = 1$

$r = 0$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{250!}{0!(250)!} (.01)^0(.99)^{250} = 1(1)(.0811) = 0.0811$

$r=1$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{250!}{1!(249)!} (.01)^1(.99)^{249} = 250(.01)(.0819) = 0.2047$

$1 - .0811 - .2047 = 0.7142$

d) $n = 250 \qquad c = 2$

$r = 2$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{250!}{2!(248)!} (.01)^2(.99)^{248} = 31125(.0001)(.0827) = 0.2574$

$1 - .0811 - .2047 - .2574 = 0.4568$

10-48 LTPD = 0.015

a) $n = 200 \qquad c = 1$

$r = 0$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{200!}{0!(200)!} (.015)^0(.985)^{200} = 1(1)(.0487) = 0.0487$

$r = 1$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{200!}{1!(199)!} (.015)^1(.985)^{199} = 200(.015)(.0494) = 0.1482$

$.0487 + .1482 = 0.1969$

b) $n = 200 \qquad c = 2$

$r = 2$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{200!}{2!(198)!} (.015)^2(.985)^{198} = 19900(.000225)(.0502) = 0.2246$

$.0487 + .1482 + .2246 = 0.4215$

c) $n = 250 \qquad c = 1$

$r = 0$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{250!}{0!(250)!} (.015)^0(.985)^{250} = 1(1)(.0229) = 0.0229$

$r = 1$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{250!}{1!(249)!} (.015)^1(.985)^{249} = 250(.015)(.0232) = 0.087$

$.0229 + .087 = 0.1099$

d) $n = 250 \qquad c = 2$

$r = 2$: $\frac{n!}{r!(n-r)!} p^r q^{n-r} = \frac{250!}{2!(248)!} (.015)^2(.985)^{248} = 31125(.000225)(.0236) = 0.1650$

$.0229 + .087 + .1650 = 0.2749$

10-49 Even though customers may be different, they can still place similar demands on a service organization. For example:

- the time to process a bank deposit shouldn't depend on the size of the checks being deposited
- in a well run doctor's office, the time between when patients have appointments and when the doctor gets around to see them shouldn't depend on the reasons for their appointments
- regardless of their size, all windows washed by a cleaning service should be equally clean.

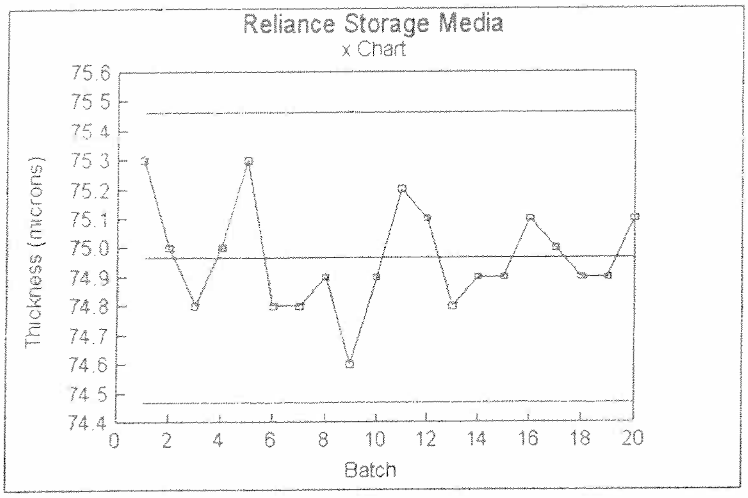
Hence, even in service organizations, variation remains the enemy of quality.

10-50 a) $n = 24 \qquad k = 20 \qquad A_2 = 0.157$

$\bar{\bar{x}} = \frac{\sum \bar{x}}{k} = \frac{1499.3}{20} = 74.965 \qquad \bar{R} = \frac{\sum R}{k} = \frac{63.3}{20} = 3.165$

$CL = \bar{\bar{x}} = 74.965$

$UCL = \overline{\overline{x}} + A_2(\overline{R}) = 74.965 + 0.157(3.165) = 75.462$
$LCL = \overline{\overline{x}} - A_2(\overline{R}) = 74.965 - 0.157(3.165) = 74.468$

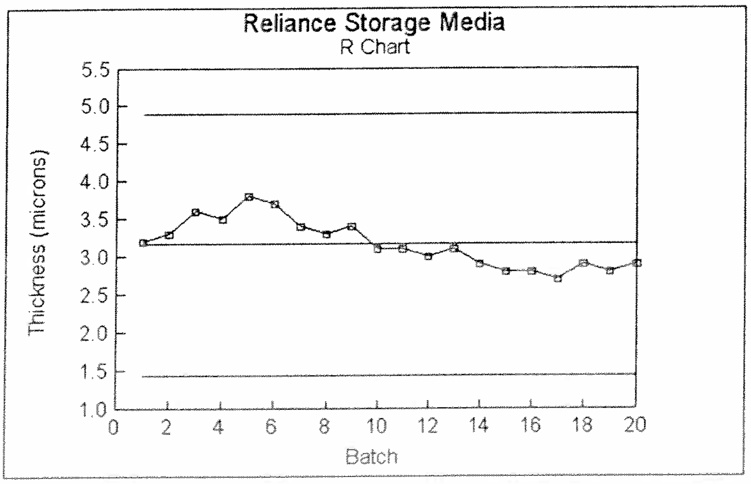


b) The process is in-control. There are no outliers or out-of-control patterns.

c) Yes. The last 10 observations do cluster closer to the center line than do the first 10 observations. Deshawn should be happy to see this pattern, since it indicates the inherent variability of the process has decreased. To the extent this is true, he might want to use the last 10 observations to re-compute the $\overline{x}$ chart. The new chart will have narrower control limits.

10-51 a) $\overline{R} = 3.165$ $D_4 = 1.548$ $D_3 = 0.452$

$CL = \overline{R} = 3.165$
$UCL = \overline{R}\ D_4 = 3.165(1.548) = 4.8994$
$LCL = \overline{R}\ D_3 = 3.165(0.452) = 1.4306$



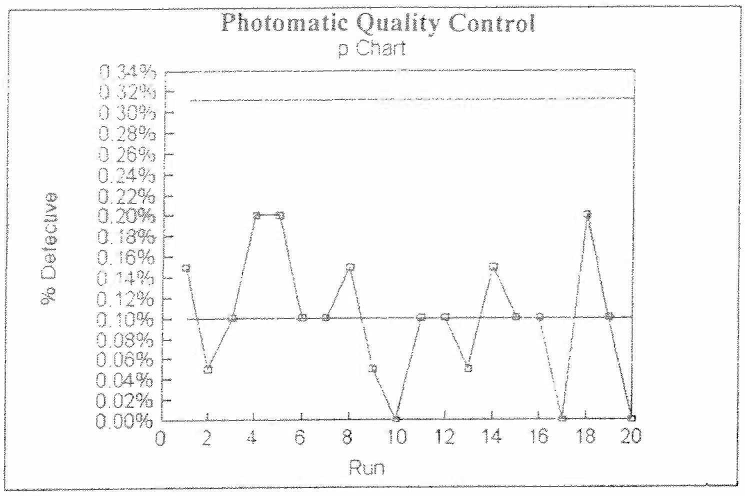
b) No. The pattern shows that variability is decreasing, which means quality is improving.

c) Yes. In the $\overline{x}$ chart, the last 10 observations clustered closer to the center line reflecting decreased variability. The R chart also reflects this pattern, which is good news.

10-52 $n = 2000$ $p = 0.001$
$CL = p = 0.001$

$$UCL = p + 3\sqrt{\frac{pq}{n}} = 0.001 + 3\sqrt{\frac{.001(.999)}{2000}} = 0.0031$$

$$LCL = p - 3\sqrt{\frac{pq}{n}} = 0.001 - 3\sqrt{\frac{.001(.999)}{2000}} = 0$$



The chart shows the process is in-control.

10-53 Suppose the hypotheses are: H_0: the defect rate is acceptable
H_1: the defect rate is unacceptable

Then the producer's risk is the chance that a lot will be rejected even when the actual defect rate falls within acceptable limits, that is, rejecting H_0 when it is true. This corresponds to making a Type I error in hypothesis testing. The consumer's risk is the probability that a bad lot (containing more defective pieces than is acceptable) will be accepted, that is, accepting H_0 when it is false. This corresponds to making a Type II error in hypothesis testing.

10-54 Common variation: density of flour, variability in measuring ingredients, variability in the amount of dough per cracker.
Special cause variation: improper calibration of measuring machinery, drifting termperature in ovens, miscounting by packaging machinery.

10-55 a) AQL = 0.005 $p = 0.5\%$
The probability of acceptance (Y axis) $\approx$ 0.93
Therefore, the Producer's Risk $\approx 1 - .93 = 0.07$

b) AQL = 0.010 $p = 1.0\%$
The probability of acceptance (Y axis) $\approx$ 0.64
Therefore, the Producer's Risk $\approx 1 - .64 = 0.36$

c) AQL = 0.015 $p = 1.5\%$
The probability of acceptance (Y axis) $\approx$ 0.34
Therefore, the Producer's Risk $\approx 1 - .34 = 0.66$

10-56 a) LTPD = 0.010
$p = 1.0\%$
The probability of acceptance (Y axis) $\approx$ 0.64
Therefore, the Consumer's Risk $\approx$ 0.64

b) LTPD = 0.015
$p = 1.5\%$
The probability of acceptance (Y axis) $\approx$ 0.34
Therefore, the Consumer's Risk $\approx$ 0.34

$p = 2.0\%$
The probability of acceptance (Y axis) $\approx$ 0.15
Therefore, the Consumer's Risk $\approx$ 0.15

10-57 Control charts can help Connie identify any nonrandom variation that might suggest grade inflation. She could plot the GPA data as sample averages in an $\bar{x}$ chart. A pattern of outliers above the upper control limit, or an increasing trend, would be consistent with the grade inflation hypothesis. She could also plot the proportion of A & B grades for each year in a p chart. As with the $\bar{x}$ chart, outliers above the upper control limit might indicate grade inflation.

10-58 Acceptance sampling is more effective in the long run than sampling entire batches because it forces the supplier to take responsibility for the quality of the output.

10-59 Since graduation rate is a fraction of the population of all students who start in a degree program, a p-chart is the appropriate tool for monitoring this attribute. However, the chart could be misleading, unless it were clearly specified exactly what graduation rate was being measured. Six-year and four-year graduation rates could give very different impressions of how well students were completing their studies. Furthermore, the graduation rate by itself says nothing about the quality of the curriculum or the instruction at the college.

10-60 a) No. This is a good example of <u>inspection</u> instead of <u>prevention.</u>

b) Major bones could collect causes by parent (failure to make appointment, failure to keep appointment, etc.), by child (illness at time of appointment, allergic reactions, etc.), and by health care professionals (shortage of vaccine, errors in record keeping, etc.).

c) Collecting data from a sample of missed immunizations would enable the HMO to construct a Pareto diagram to help identify the principal causes of the problem.

CHI-SQUARE AND ANALYSIS OF VARIANCE

11-1 To determine if three or more population proportions can be considered equal, or if two attributes of a population are independent

11-2 To determine if three or more population means can be considered equal

11-3 a) Inference on two population variances (use F)

b) analysis of variance

c) chi-square test

d) inference on two population variances (use F)

11-4 a) FALSE; can do inference only on one or two variances

b) TRUE; use analysis of variance

c) TRUE; use a chi-square test

11-5

# of Parameters Involved	Type of Parameter μ	σ	p
1	normal, t	χ^2	normal
2	normal, t	F	normal
3	analysis of variance, F	------	χ^2

11-6 $df = (\text{\# rows} - 1)\ (\text{\# columns} - 1)$

a) $4 \times 3 = 12$ b) $5 \times 1 = 5$ c) $2 \times 6 = 12$ d) $3 \times 3 = 9$

11-7

Purchase Activity	Age Group 18-29	30-39	40-49	50-59	60-69
Frequently					
Observed	12	18	17	22	32
Expected	20.2	20.2	20.2	20.2	20.2
Seldom					
Observed	18	25	29	24	30
Expected	25.2	25.2	25.2	25.2	25.2
Never					
Observed	45	32	29	29	13
Expected	29.6	29.6	29.6	29.6	29.6

11-8 a)

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
12	20.2	−8.2	67.24	3.3287
18	20.2	−2.2	4.84	0.2396
17	20.2	−3.2	10.24	0.5069

22	20.2	1.8	3.24	0.1604
32	20.2	11.8	139.24	6.8931
18	25.2	−7.2	51.84	2.0571
25	25.2	−0.2	0.04	0.0016
29	25.2	3.8	14.44	0.5730
24	25.2	−1.2	1.44	0.0571
30	25.2	4.8	23.04	0.9143
45	29.6	15.4	237.16	8.0122
32	29.6	2.4	5.76	0.1946
29	29.6	−0.6	0.36	0.0122
29	29.6	−0.6	0.36	0.0122
13	29.6	−16.6	275.56	9.3095

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 32.2725$$

b) H_0 : Age group is independent of purchasing plans
H_1 : Age group is related to purchasing plans

c) $df = (3-1) \times (5-1) = 8$ $\qquad \chi^2_{CRIT} = \chi^2_{.01,8} = 20.090$

Thus we reject H_0 because $\chi^2 > \chi^2_{CRIT}$, and we conclude there is a relationship between age and purchasing plans.

11-9

	Weekly Chip Sales High	Medium	Low
at peak			
Observed	20	7	3
Expected	15	9	6
at trough			
Observed	30	40	30
Expected	50	30	20
rising			
Observed	20	8	2
Expected	15	9	6
falling			
Observed	30	5	5
Expected	20	12	8

11-10 a) H_0 : weekly chip sales independent of economy
H_1 : weekly chip sales related to state of economy

b)

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
20	15	5	25	1.667
7	9	−2	4	0.444
3	6	−3	9	1.500
30	50	−20	400	8.000
40	30	10	100	3.333
30	20	10	100	5.000
20	15	5	25	1.667
8	9	−1	1	0.111
2	6	−4	16	2.667

30	20	10	100	5.000
5	12	−7	49	4.083
5	8	−3	9	1.125

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 34.597$$

c) $df = 2 \times 3 = 6$ $\chi^2_{CRIT} = \chi^2_{.10;6} = 10.645$

reject H_0 $\Rightarrow$ silicon chip sales are not independent of U.S. economy

11-11 $H_0 : p_1 = p_2 = p_3$

$H_1 : p_1, p_2$ and p_3 are not all equal

		Firm Asset Size ($1000s) < 500	500-2000	2000+
Debt < equity	f_o	7	10	8
	f_e	6.855	11.290	6.855
Debt > equity	f_o	10	18	9
	f_e	10.145	16.710	10.145

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
7	6.855	0.145	0.021	.003
10	11.290	−1.290	1.664	.147
8	6.855	1.145	1.311	.191
10	10.145	−0.145	0.021	.002
18	16.710	1.290	1.664	.100
9	10.145	−1.145	1.311	.129

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 0.572$$

$df = 1 \times 2 = 2$ $\chi^2_{CRIT} = \chi^2_{.10;2} = 4.605$

do not reject H_0 $\Rightarrow$ capital structures do not differ significantly by size of firm

11-12 H_0 : frequency of readership and level of education are independent

H_1 : they are not independent

	Prof	College	HS	< HS
Never				
Observed	10.000	17.000	11.000	21.000
Expected	18.643	21.275	8.993	10.089
Sometimes				
Observed	12.000	23.000	8.000	5.000
Expected	15.167	17.309	7.316	8.208
AM or PM				
Observed	35.000	38.000	16.000	7.000
Expected	30.335	34.617	14.632	16.416
Both editions				
Observed	28.000	19.000	6.000	13.000
Expected	20.855	23.799	10.059	11.286

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
10	18.643	--8.643	74.701	4.007
17	21.275	−4.275	18.276	0.859
11	8.993	2.007	4.028	0.448
21	10.089	10.911	119.050	11.800
12	15.167	−3.167	10.030	0.661
23	17.309	5.691	32.387	1.871
8	7.316	0.684	0.468	0.064
5	8.208	−3.208	10.291	1.254
35	30.335	4.665	21.762	0.717
38	34.617	3.383	11.445	0.331
16	14.632	1.368	1.871	0.128
7	16.416	−9.416	88.661	5.401
28	20.855	7.145	51.051	2.448
19	23.799	−4.799	23.030	0.968
6	10.059	−4.059	16.475	1.638
13	11.286	1.714	2.938	0.260

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 32.855$$

$df = 3 \times 3 = 9$ $\qquad \chi^2_{CRIT} = \chi^2_{.10;9} = 14.684$

reject H_0 $\Rightarrow$ frequency of readership differs according to education

11-13 The entries in the cells of the following table are f_o, f_e, and $\frac{(f_o - f_e)^2}{f_e}$.

	A	B	C	D	F
< 5 hrs	13.000	10.000	11.000	16.000	5.000
	6.875	10.313	20.625	10.313	6.875
	5.457	0.009	4.492	3.136	0.511
5 - 10 hrs	20.000	27.000	27.000	19.000	2.000
	11.875	17.813	35.625	17.813	11.875
	5.559	4.738	2.088	0.079	8.212
11 - 20 hrs	9.000	20.000	71.000	16.000	32.000
	19.375	29.063	58.125	29.063	19.375
	5.556	0.146	2.852	5.871	8.227
> 20 hrs	8.000	11.000	41.000	24.000	11.000
	11.875	17.813	35.625	17.813	11.875
	1.264	2.606	0.811	2.149	0.064

H_0 : grades independent of hours spent listening to music
H_1 : grades related to hours spent listening to music

$\chi^2_{OBS} = 63.829$ $\qquad df = 3 \times 4 = 12$ $\qquad \chi^2_{CRIT} = \chi^2_{.05;12} = 21.026$

prob-value < 0.001 $\Rightarrow$ reject H_0 $\Rightarrow$ there is a relationship between grades and hours spent listening to music

11-14 a)

Sales(x)		2.6	3.8	5	6.2	7.4
$z = \frac{x-5}{1.5}$		−1.6	−0.8	0	0.8	1.6
Probability	.0548	.1571	.2881	.2881	.1571	.0548

b)

Value	< 2.60	2.60-3.79	3.80-4.99	5.0-6.19	6.20-7.39	≥ 7.40
Observed	6	30	41	52	12	9
Expected	8.220	23.565	43.215	43.215	23.565	8.220
$\frac{(f_o - f_e)^2}{f_e}$	0.600	1.757	0.114	1.786	5.676	0.074

c) $\chi^2 = 0.600 + 1.757 + 0.114 + 1.786 + 5.676 + 0.074 = 10.007$

d) $df = 6 - 1 = 5$ $\qquad \chi^2_{CRIT} = \chi^2_{.10;5} = 9.236$

reject $H_0 \Rightarrow$ no, the distribution is not well described as normal with $\mu = 5$ and $\sigma = 1.5$

11-15 H_0 : Poisson with $\lambda = 5$

H_1 : something else

Test at $\alpha = .05$, with $df = 7$

# of calls/min.	0 or 1	2	3	4	5	6	7+
Poisson prob.	.04043	.08422	.14037	.17547	.17547	.14622	.23782
Observed	19	42	60	89	94	52	80
Expected	17.63	36.72	61.20	76.50	76.50	63.75	103.69
$\frac{(f_o - f_e)^2}{f_e}$	0.1065	0.7592	0.0235	2.0425	4.0037	2.1657	5.4124

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 14.5131 \qquad \chi^2_{CRIT} = \chi^2_{.05,6} = 12.592$$

reject $H_0 \Rightarrow$ the data don't follow a Poisson distribution with $\lambda = 5$

11-16 Let x = sales by Mr. Armstrong

H_0 : x distributed binomial $(n=5, \ p=.4)$

H_1 : x not distributed as above (it could be binomial with a different p or not even binomial at all)

$\alpha = .05$, $df = 4$ after pooling together the last two groups

# of sales/ day	0	1	2	3	4	5	4 or 5
Observed	10	41	60	20	6	3	9
Binomial prob.	.0778	.2592	.3456	.2304	.0768	.0102	.0870
Expected	10.89	36.29	48.38	32.26	10.75	1.43	12.18
$\frac{(f_o - f_e)^2}{f_e}$	.073	.611	2.791	4.659	combine because $f_e < 5$		.830

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 8.964$$

$\chi^2_{CRIT} = \chi^2_{.05;4} = 9.488$, so do not reject $H_0 \Rightarrow$ Mr. Armstrong's sales may be described as binomially distributed, with $n = 5, \ p = .4$

11-17 H_0: Normal with $\mu = 14$ and $\sigma = 5$

H_1: Some other distribution

Test at $\alpha = 0.05$, using 5 equally probable intervals, and hence with $df = 4$.

The 20th, 40th, 60th, and 80th percentiles of any distribution divide that distribution into five equally probable intervals. For the standard normal distribution, those percentiles are $z = -0.84, -0.25, 0.25$, and 0.84. Since $x = \mu + \sigma z$, the corresponding percentiles for a normal distribution with $\mu = 14$ and $\sigma = 5$ are: $14 - 5(0.84) = 9.80$, $14 - 5(0.25) = 12.75$, $14 + 5(0.25) = 15.25$, and $14 + 5(0.84) = 18.20$.

Time Interval	< 9.80	9.80-12.75	12.75-15.25	15.25-18.20	> 18.20
Observed	9	8	6	5	12
Expected	8	8	8	8	8
$\frac{(f_o-f_e)^2}{f_e}$	0.125	0.000	0.500	1.125	2.000

$\chi^2 = 3.75;\ \chi^2_{CRIT} = \chi^2_{.05,4} = 9.488$

So we do not reject H_0. The data are well described by a normal distribution with $\mu = 14$ and $\sigma = 5$.

11-18 a) Dollars(x) 1000 2000

$z = \frac{x - 1500}{600}$ -0.83 0.83

Deposits	0-999	1000-1999	2000 and more
Observed	20	65	25
Normal prob.	.2033	.5934	.2033
Expected	22.363	65.274	22.363
$\frac{(f_o - f_e)^2}{f_e}$	.2497	.0012	.3109

b) $\chi^2 = .2497 + .0012 + .3109 = .5618$

c) H_0 : deposits are normally distributed with $\mu = 1500$, $\sigma = 600$
H_0 : deposits are not distributed as above

d) $\chi^2_{CRIT} = \chi^2_{.10;2} = 4.605 \Rightarrow$ do not reject $H_0 \Rightarrow$ deposits are normally distributed with $\mu = 1500$ and $\sigma = 600$

11-19 Let X = the number of mangled letters received by any person in the test mailing of 2 letters to each person in the sample
H_0 : X is distributed as binomial with $n = 2$ and $p = .4$
H_1 : X is distributed as something else (different p)

# of mangled letters	0	1	2
Binomial prob.	.7225	.2550	.0225
Observed	260	40	10
Expected	223.975	79.05	6.975
$\frac{(f_o - f_e)^2}{f_e}$	5.79	19.29	1.31

$\chi^2 = 5.79 + 19.29 + 1.31 = 26.39$
$df = 3 - 1 = 2$ $\chi^2_{CRIT} = \chi^2_{.10;2} = 4.605$
reject $H_0 \Rightarrow$ the number of mangled letters does not follow a binomial distribution with $n = 2$ and $p = .15$ (from the data, p appears to be smaller)

11-20 H_0 : the winnings are distributed as stated
H_1 : winning probabilities are different

Winnings	\$1	\$100	\$0
Probability	.1	.05	.85
Observed	87	48	865
Expected	100	50	850

$\frac{(f_o - f_e)^2}{f_e}$	1.69	0.08	.2647

$\chi^2 = 1.69 + 0.08 + 0.2647 = 2.0347$ $\chi^2_{CRIT} = \chi^2_{.05;2} = 5.991$

do not reject H_0 $\Rightarrow$ the state's claim is reasonable

11-21 H_0 : the number of broken bone patients is Poisson with $\lambda = 2$
H_1 : not H_0 (either a different λ or a different probability distribution)

# of broken bones	0	1	2	3	4	≥ 5
Poisson prob.	.1355	.2707	.2707	.1804	.0902	.0527
Observed	25	55	65	35	20	10
Expected	28.455	56.847	56.847	37.884	18.942	11.067
$\frac{(f_o - f_e)^2}{f_e}$	.4195	.0600	1.1693	.2196	.0591	.1029

$\chi^2 = 2.0304$ $df = 6 - 1 = 5$ $\chi^2_{CRIT} = \chi^2_{.05;5} = 11.070$

do not reject H_0 $\Rightarrow$ the incidence of broken bone cases can be assumed to follow a Poisson distribution with $\lambda = 2$

11-22 H_0 : binomial with $p = .3$, $n = 3$
H_1 : some other distribution
Test at $\alpha = .05$ with 2 df

# of shifts	0	1	2 or 3
Binomial prob.	.343	.441	.216
Observed	16	27	17
Expected	20.58	26.46	12.96
$\frac{(f_o - f_e)^2}{f_e}$	1.019	0.011	1.259

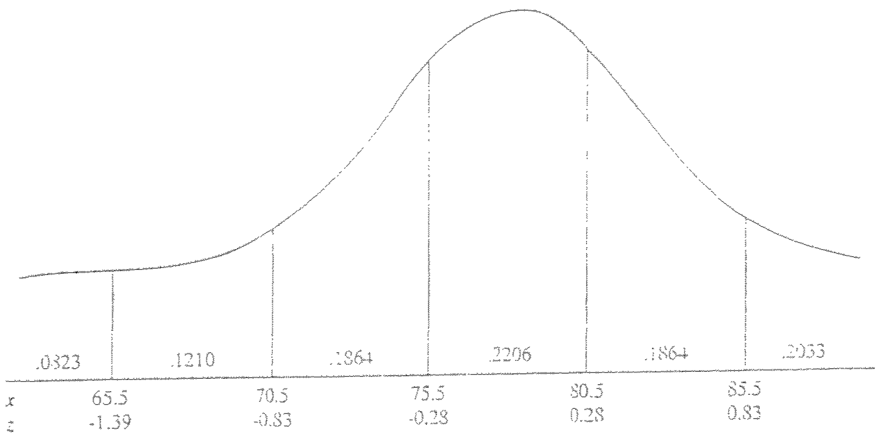
$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 2.289$ $\chi^2_{CRIT} = \chi^2_{.05;2} = 5.991$

accept H_0 $\Rightarrow$ the number of alarms is well described by a binomial distribution with $p = .3$ and $n = 3$.

11-23 a) Performing a t–test on a mean from a small sample is valid only if the data are sampled from a normal population, but in this example, with $n = 200$, the Central Limit Theorem allows us to base the test on μ on the normal distribution.

b) H_0 : the data is normally distributed (any μ or σ)
H_1: the data is not normally distributed

c)



d)

Sales	≤ 65	66-70	71-75	76-80	81-85	≥ 86
Normal prob.	.0823	.1210	.1864	.2206	.1864	.2033
Observed	10	20	40	50	40	40
Expected	16.46	24.20	37.28	44.12	37.28	40.66
$\frac{(f_o - f_e)^2}{f_e}$	2.535	0.729	0.198	0.784	0.198	0.011

$\chi^2 = 4.455$ $\qquad df = 6 - 1 - 2 = 3$ $\qquad \chi^2_{CRIT} = \chi^2_{.05;3} = 7.815$

do not reject $H_0 \Rightarrow$ the assumption of normality is okay

11-24 H_0: the number of customers arriving in 5-minute intervals is distributed Poisson with $\lambda = 3$

H_1: not H_0

Test at $\alpha = .05$ with $df = 7 - 1 = 6$

# of customers	0	1	2	3	4	5	≥ 6
Poisson prob.	.0498	.1494	.2240	.2240	.1680	.1008	.0840
Observed	22	74	115	95	94	80	20
Expected	24.9	74.7	112	112	84	50.4	42
$\frac{(f_o - f_e)^2}{f_e}$	.338	.007	.080	2.580	1.190	17.384	11.524

$\chi^2 = 33.103$ $\qquad \chi^2_{CRIT} = \chi^2_{.05;6} = 12.592$

reject $H_0 \Rightarrow$ Poisson distribution with $\lambda = 3$ is not appropriate

11-25 H_0: Lou's number of hits per game is distributed binomially $(n = 5,\ p = .4)$

H_1: Lou is wrong

Hits	0	1	2	3	4	5
Binomial prob.	.0778	.2592	.3456	23.04	.0768	.0102
Observed	12	38	27	17	5	1
Expected	7.78	25.92	34.56	23.04	7.68	1.02
					combine because $f_e < 5$	
$\frac{(f_o - f_e)^2}{f_e}$	2.289	5.630	1.654	1.583	0.838	

$\chi^2 = 11.994$ $\qquad \chi^2_{CRIT} = \chi^2_{.05;4} = 9.488$

reject $H_0 \Rightarrow$ Lou's number of hits per game is not binomially distributed with $p = .4$ and $n = 5$

11-26 a)

	A	B	C	D	E
	4.4	5.8	4.8	2.9	4.6
	4.6	5.2	5.9	2.7	4.3
	4.5	4.9	4.9	2.9	3.8
	4.1	4.7	4.6	3.9	5.2
	3.8	4.6	4.3	4.3	4.4
$\sum x$	21.4	25.2	24.5	16.7	22.3
n	5	5	5	5	5
$\bar{x}$	4.28	5.04	4.90	3.34	4.46
$\sum x^2$	92.02	127.94	121.51	57.81	100.49
s^2	0.107	0.233	0.365	0.508	0.258

$$\text{grand mean} = \bar{\bar{x}} = \frac{21.4 + 25.2 + 24.5 + 16.7 + 22.3}{25} = 4.404$$

b) $$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{9.0056}{4} = 2.2514$$

c) $$\hat{\sigma}_w^2 = \frac{\sum (n_i-1)s_i^2}{n_T-k} = \frac{5.884}{20} = 0.2942$$

d) $$F = \frac{2.2514}{0.2942} = 7.65$$

$F_{.05}(4,20) = 2.87$, so reject H_0 $\Rightarrow$ the brands produce significantly different amounts of relief

11-27

	n	$\bar{x}$	s^2
Method 1:	6	45.167	30.167
Method 2:	6	48.000	47.600
Method 3:	6	41.667	31.067

$$\text{grand mean} = \bar{\bar{x}} = \frac{6(45.167) + 6(48.000) + 6(41.667)}{18} = 44.944$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{120.766}{2} = 60.383$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i-1)s^2}{n_T-k} = \frac{544.17}{15} = 36.278$$

$$F = \frac{60.383}{36.278} = 1.66$$

With 2,15 *df*, $F_{.05} = 3.68$, so we do not reject H_0 $\Rightarrow$ the methods do not lead to significantly different productivity levels

11-28

	n	$\bar{x}$	s^2
Employee 1	4	14.5	4.333
Employee 2	4	13	8.667
Employee 3	5	13	2.5
Employee 4	6	11.667	3.467

$$\text{grand mean} = \bar{\bar{x}} = \frac{4(14.5) + 4(13) + 5(13) + 6(11.667)}{19} = 12.8947$$

$$\hat{\sigma}_b^2 = \frac{19.4562}{3} = 6.4854$$

$$\hat{\sigma}_w^2 = \frac{66.3333}{15} = 4.4222$$

$$F = \frac{6.4854}{4.4222} = 1.47$$

$F_{.05}(3,15) = 3.29$, so do not reject H_0 $\Rightarrow$ the employees' mean performances are not significantly different

Here : $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1 : at least one pair $\mu_i - \mu_j \neq 0$

11-29

	n	$\bar{x}$	s^2
Sample 1	5	23.6	28.30
Sample 2	6	22.833	27.7667
Sample 3	6	19.0	26.0
Sample 4	5	21.8	14.2

$$\bar{\bar{x}} = \frac{5(23.6) + 6(22.8333) + 6(19) + 5(21.8)}{5 + 6 + 6 + 5} = 21.7273$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{69.53}{3} = 23.1768$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{438.8333}{18} = 24.3796$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{23.1768}{24.3796} = 0.95$$

$F_{CRIT} = F_{.01}(3,18) = 5.09$, so do not reject H_0 $\Rightarrow$ we can conclude that the four samples come from populations with the same mean value

11-30 a)

	n	$\bar{x}$	s^2
Speed 1	5	36	2.5
Speed 2	5	31	7
Speed 3	5	35	10
Speed 4	5	31	10

$$\bar{\bar{x}} = \frac{5(36) + 5(31) + 5(35) + 5(31)}{5 + 5 + 5 + 5} = 33.25$$

b) $$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{103.75}{3} = 34.5833$$

c) $$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{118}{16} = 7.375$$

d) $$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{34.5833}{7.375} = 4.69$$

$F_{CRIT} = F_{.05}(3,16) = 3.24$, so reject H_0 $\Rightarrow$ the different speeds do significantly affect the number of defective clocks

11-31 a) $H_0 : \mu_C = \mu_{\text{liquid H}} = \mu_{\text{solid H}} = \mu$
H_1 : at least one $\mu_i \neq \mu$

b) prob-value = .004 $< \alpha = .05$ $\Rightarrow$ reject H_0

c) There is a significant palatability difference somewhere

11-32

	n	$\bar{x}$	s^2
November	6	46.6667	81.0667
December	6	46.1667	47.7667
January	6	33.5	21.5

$$\bar{\bar{x}} = \frac{6(46.667) + 6(46.1667) + 6(33.5)}{6+6+6} = 42.1111$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{668.1146}{2} = 334.0573$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{751.6667}{15} = 50.1111$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{334.0573}{50.1111} = 6.67$$

$F_{CRIT} = F_{.05}(2,15) = 3.68$, so reject H_0 $\Rightarrow$ the number of shoplifters differs significantly from month to month

11-33

	n	$\bar{x}$	s^2
Section 1	16	86.7250	98.2780
Section 2	8	81.4250	152.8307
Section 3	11	85.3546	114.3467

$$\bar{\bar{x}} = \frac{16(86.7250) + 8(81.4250) + 11(85.3546)}{16+8+11} = 85.0829$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{150.9978}{2} = 75.4989$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{3687.4519}{32} = 115.2329$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{75.4989}{115.2329} = 0.6552$$

$F_{CRIT} = F_{.01}\,(2,32) \approx 5.35$, so do not reject H_0 $\Rightarrow$ average grades do not differ significantly from instructor to instructor

11-34

	n	$\bar{x}$	s^2
Room 1	5	5.7	1.45
Room 2	5	4.1	1.55
Room 3	5	2.4	1.425
Room 4	5	7.6	1.675
Room 5	5	2.2	1.075

$$\bar{\bar{x}} = \frac{5(5.7) + 5(4.1) + 5(2.4) + 5(7.6) + 5(2.2)}{5+5+5+5+5} = 4.4$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{104.3}{4} = 26.075$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{28.7}{20} = 1.435$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{26.075}{1.435} = 18.17$$

$F_{CRIT} = F_{.05}(4,20) = 2.87$, so reject H_0 $\Rightarrow$ at least two clean rooms have significantly different numbers of dust particles

11-35

	n	$\bar{x}$	s^2
Quarter 1	5	49.2	44.2
Quarter 2	5	45.2	21.2
Quarter 3	5	44.0	38.5

$$\bar{\bar{x}} = \frac{5(49.2) + 5(45.2) + 5(44.0)}{5 + 5 + 5} = 46.1333$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k\text{-}1} = \frac{74.1333}{2} = 37.0667 \qquad \hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{415.6}{12} = 34.6333$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{37.0667}{34.6333} = 1.07$$

$F_{CRIT} = F_{.05}(2,12) = 3.89$, so we do not reject H_0 $\Rightarrow$ levels of housing starts don't differ significantly in the three quarters

11-36

	n	$\bar{x}$	s^2
Generic	5	18.0	38.0000
DNA	4	19.5	99.6667
RNA	6	16.5	23.9000
Oops	5	19.4	41.3000

$$\bar{\bar{x}} = \frac{5(18.0) + 4(19.5) + 6(16.5) + 5(19.4)}{5 + 4 + 6 + 5} = 18.2$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{31.5}{3} = 10.5$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{735.7}{16} = 45.9813$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{10.5}{45.9813} = 0.23$$

$F_{CRIT} = F_{.05}(3,16) = 3.24$, so do not reject H_0 $\Rightarrow$ the quantities of jeans sold don't differ across brands

11-37

	n	$\bar{x}$	s^2
Agriculture	5	10	2.5
State	6	11.3333	7.4667
Interior	3	12	16

$$\bar{\bar{x}} = \frac{5(10) + 6(11.3333) + 3(12)}{5 + 6 + 3} = 11$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{8.6667}{2} = 4.3333$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{79.3333}{11} = 7.2121$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{4.3333}{7.2121} = 0.60$$

$F_{CRIT} = F_{.01}(2,11) = 7.21$, so do not reject H_0 $\Rightarrow$ all three departments have the same monthly expenses; there is no reason to single out some departments as usual

11-38

	n	$\bar{x}$	s^2
Restaurant 1	5	4	0.875
Restaurant 2	5	4.1	0.925
Restaurant 3	5	4.6	3.425
Restaurant 4	5	3.6	1.425

a) $\bar{\bar{x}} = \dfrac{5(4) + 5(4.1) + 5(4.6) + 5(3.6)}{5 + 5 + 5 + 5} = 4.075$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{2.5375}{3} = 0.8458$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{26.6}{16} = 1.6625$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{0.8458}{1.6625} = 0.51$$

$F_{CRIT} = F_{.05}(3.16) = 3.24$, so do not reject H_0 $\Rightarrow$ the restaurants do not have significantly different service times

b) Since no group stands out as having a mean greater than any of the others, the owner has no reason to single out any of the manager's restaurants for improvement needed more than the other restaurants.

11-39 For a 90% confidence interval with 19 *df*:

$$\sigma_L^2 = \frac{(n-1)s^2}{\chi_U^2} = \frac{19(12.2)}{30.144} = 7.690$$

$$\sigma_U^2 = \frac{(n-1)s^2}{\chi_L^2} = \frac{19(12.2)}{10.117} = 22.912$$

Thus, the confidence interval is: [7.690, 22.912]

11-40 $H_0 : \sigma = 50$
$H_1 : \sigma \neq 50$

The observed $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2} = \dfrac{29(57)^2}{50^2} = 37.688$

$\chi^2_{.975;29} = 16.047$ $\qquad$ $\chi^2_{.025;29} = 45.722$

Do not reject H_0 $\Rightarrow$ sample standard deviation of 57 is not significantly different from the hypothesized standard deviation of 50

11-41 For a 90% confidence interval with 14 *df*:

$$\sigma_L^2 = \frac{(n-1)s^2}{\chi_U^2} = \frac{14(6.4)^2}{23.685} = 24.211 \qquad \sigma_U^2 = \frac{(n-1)s^2}{\chi_L^2} = \frac{14(6.4)^2}{6.571} = 87.268$$

Thus, the confidence interval is: [24.211, 87.268]

11-42 a) $H_0 : \sigma = 2$ (or $\sigma^2 = 4$)
$H_1 : \sigma < 2$ (or $\sigma^2 < 4$)

b) observed $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2} = \dfrac{29(1.46)^2}{4} = 15.4541$

$\chi^2_{CRIT} = \chi^2_{.99;29} = 14.256$, so do not reject H_0

c) The observed s^2 of 1.46 is not significantly below 2, so the telescope should not be sold.

11-43 H_0: $\sigma^2 = 171$
H_1: $\sigma^2 \neq 171$
The limits of the acceptance region are:

$$\chi^2_{.99;24} = 10.856 \text{ and } \chi^2_{.01;24} = 42.980$$

The observed $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2} = \dfrac{24(196.5)}{171} = 27.579$, so we do not reject H_0; the variability is not significantly different from the previous estimate.

11-44 a) $H_0 : \sigma^2 = 64$
$H_1 : \sigma^2 \neq 64$

b) $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2} = \dfrac{19(28)}{64} = 8.31$

$\chi^2_{.975;19} = 8.907$ $\chi^2_{.025;19} = 32.852$ Reject H_0

c) Thus, six-year olds' attention span is significantly different in variability from five-year olds' attention span.

11-45 For a 90% confidence interval with 29 *df*:

$$\sigma^2_L = \frac{(n-1)s^2}{\chi^2_U} = \frac{29(50)}{42.557} = 34.072 \qquad \sigma^2_U = \frac{(n-1)s^2}{\chi^2_L} = \frac{29(50)}{17.708} = 81.884$$

Thus, the 90% confidence in'erval is [34.072, 81.884].

11-46 $H_0 : \sigma^2 = 80$ (no change)
$H_1 : \sigma^2 < 80$ (the change has reduced variance)

$$\text{observed } \chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{24(28)}{80} = 8.4$$

$$\chi^2_{CRIT} = \chi^2_{.95;24} = 13.848$$

Thus, reject H_0 $\Rightarrow$ the new policy does reduce the variance significantly

11-47 For a 95% confidence interval with 29 *df*:

$$\sigma^2_L = \frac{(n-1)s^2}{\chi^2_U} = \frac{29(15)^2}{45.722} = 142.710$$

$$\sigma^2_U = \frac{(n-1)s^2}{\chi^2_L} = \frac{29(15)^2}{16.047} = 406.618$$

Thus, the confidence interval for σ is $(\sqrt{142.710}, \sqrt{406.618}) = (11.95, 20.16)$

<u>One-tailed test</u>

$H_0 : \sigma = 20$
$H_1 : \sigma < 20$
$\alpha = .01$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{29(15)^2}{(20)^2} = 16.3125$$

$$\chi^2_{CRIT} = \chi^2_{.99;29} = 14.256$$

Therefore, we do not reject H_0 ; Sam should switch distributors.

11-48 $n_1 = 16 \quad s_1^2 = 3.75 \quad n_2 = 10 \quad s_2^2 = 5.38$

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

Test at $\alpha = .10$, with 15,9 df

a) $F = \dfrac{s_1^2}{s_2^2} = \dfrac{3.75}{5.38} = 0.70$

b) $F_{CRIT} = F_{.05}(15{,}9) = 3.01$

c) $F_{.95}(15{,}9) = \dfrac{1}{F_{.05}(9{,}15)} = \dfrac{1}{2.59} = 0.39$

d) do not reject H_0

11-49 Let e = experimental and c = control

$H_0 : \sigma_e^2 = \sigma_c^2$

$H_1 : \sigma_e^2 \neq \sigma_c^2$

$n_e = 25 \quad s_e^2 = 25.8 \quad n_c = 31 \quad s_c^2 = 20.6$

observed $F = \dfrac{s_e^2}{s_c^2} = \dfrac{25.8}{20.6} = 1.25$

$F_{.05}(24{,}30) = 1.89 \qquad F_{.95}(24{,}30) = \dfrac{1}{F_{.05}(30{,}24)} = \dfrac{1}{1.94} = 0.52$

Thus, do not reject H_0 $\Rightarrow$ okay to go ahead and proceed with t-test

11-50 $n_1 = 25 \quad s_1 = 15.0 \quad n_2 = 14 \quad s_2 = 9.7$

$H_0 : \sigma_1 = \sigma_2$ (or $\sigma_1^2 = \sigma_2^2$)
$H_1 : \sigma_1 > \sigma_2$ (or $\sigma_1^2 > \sigma_2^2$)

observed $F = \dfrac{s_1^2}{s_2^2} = \dfrac{15.0^2}{9.7^2} = 2.39$

$F_{CRIT} = F_{.01}(24{,}13) = 3.59$

Thus, do not reject H_0 $\Rightarrow$ we conclude that the variance of the second sample is not significantly smaller

11-51 $H_0 : \sigma^2_{MLPFS} = \sigma^2_{Oppy}$

$H_1 : \sigma^2_{\text{MLPFS}} > \sigma^2_{Oppy}$

$F = \dfrac{s^2_{MLPFS}}{s^2_{Oppy}} = \dfrac{3^2}{2^2} = 2.25$

$F_{CRIT} = F_{.05}(20{,}20) = 2.12$

Reject H_0 $\Rightarrow$ Oppy has a significantly lower variance; the Raj should invest in the Oppy fund

11-52 $n_A = 20 \quad s_A = 0.6 \quad n_B = 25 \quad s_B = 0.9$

$H_0 : \sigma_B^2 = \sigma_A^2$

$H_1 : \sigma_B^2 > \sigma_A^2$

$$\text{observed} \quad F = \frac{s_B^2}{s_A^2} = \frac{0.81}{0.36} = 2.25$$

$$F_{CRIT} = F_{.01}(24,19) = 2.92$$

Thus, we do not reject H_0 $\Rightarrow$ patients at hospital A do not have significantly less variability in their recovery times.

11-53 a) $s_B^2 = 32.33$ $\quad s_R^2 = 136.83$

b) $H_0 : \sigma_B^2 = \sigma_R^2$

$H_1 : \sigma_B^2 \neq \sigma_R^2$

$$\text{observed } F = \frac{s_B^2}{s_R^2} = \frac{32.33}{136.83} = 0.24$$

$$F_{.05}(9,9) = 3.18$$

$$F_{.95}(9,9) = \frac{1}{F_{.05}(9,9)} = \frac{1}{3.18} = 0.31$$

Thus, reject H_0 $\Rightarrow$ the two populations have significantly different variability

11-54 $H_0 : \sigma_{PAL}^2 = \sigma_{CAL}^2$

$H_1 : \sigma_{PAL}^2 > \sigma_{CAL}^2$

$$\text{observed } F = \frac{s_{PAL}^2}{s_{CAL}^2} = \frac{20^2}{10^2} = 4$$

$$F_{CRIT} = F_{.05}(24,24) = 1.98$$

Thus, reject $H_0 \Rightarrow$ PAL's processing speed is significantly more variable than CAL's

11-55 Let group 1 = urban homemakers $\quad n_1 = 70 \quad s_1^2 = 14$

group 2 = rural homemakers $\quad n_2 = 60 \quad s_2^2 = 3.5$

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 > \sigma_2^2$

$$\text{observed } F = \frac{s_1^2}{s_2^2} = \frac{14}{3.5} = 4 \qquad F_{CRIT} = F_{.01}(69,59) = 1.84$$

Thus, reject H_0 $\Rightarrow$ urban homemakers have significantly more variability in grocery shopping patterns

11-56 $H_0 = \sigma_Y^2 = \sigma_G^2$

$H_1 = \sigma_Y^2 \neq \sigma_G^2$

$$\text{observed } F = \frac{s_Y^2}{s_G^2} = \frac{16}{10} = 1.6$$

$$F_{.05}(24,10) = 2.74 \qquad F_{.95}(24,10) = \frac{1}{F_{.05}(10,24)} = \frac{1}{2.25} = 0.44$$

Thus, do not reject H_0 $\Rightarrow$ no significant difference in the variance of ice cream weights between Yum-Yum and Goody

11-57 For a 90% confidence interval with 9 df and $s^2 = 0.84444$:

$$\sigma_L^2 = \frac{(n-1)s^2}{\chi_U^2} = \frac{9(0.84444)}{16.919} = 0.449 \qquad \sigma_U^2 = \frac{(n-1)s^2}{\chi_L^2} = \frac{9(.84444)}{3.325} = 2.286$$

Thus, the confidence interval is [0.449, 2.286]

11-58 H_0: occupation and attitude toward social legislation are independent
H_1: occupation and attitude toward social legislation are dependent

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
19	18.8108	0.1892	0.0358	0.0019
16	15.8919	0.1081	0.0117	0.0007
37	37.2973	−0.2973	0.0884	0.0024
15	21.6847	−6.6847	44.6852	2.0607
22	18.3198	3.6802	13.5439	0.7393
46	42.9955	3.0045	9.0270	0.2100
24	17.5045	6.4955	42.1915	2.4103
11	14.7883	−3.7883	14.3512	0.9704
32	34.7072	−2.7072	7.3289	0.2112

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 6.6069$$

$\chi^2_{CRIT} = \chi^2_{.05;4} = 9.488$

Do not reject H_0 $\Rightarrow$ occupation and attitudes toward social legislation appear to be unrelated

11-59 a) H_0 : preferences for size of car do not depend on the region of the country
H_1 : preferences for size of car do depend on the region of the country

b)

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
105	100	5	25	0.25
120	100	20	400	4.00
105	100	5	25	0.25
70	100	−30	900	9.00
120	125	−5	25	0.20
100	125	−25	625	5.00
130	125	5	25	0.20
150	125	25	625	5.00
25	25	0	0	0.00
30	25	5	25	1.00
15	25	−10	100	4.00
30	25	5	25	1.00

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 29.90$$

$\chi^2_{CRIT} = \chi^2_{.05;6} = 12.592$ at $\alpha = .05$, and $\chi^2_{CRIT} = \chi^2_{.20;6} = 8.558$ at $\alpha = .20$

c) At either significance level, we reject H_0 $\Rightarrow$ region of the country does have a bearing on car-size preferences

11-60 a) normal b) χ^2

c) analysis of variance (F-distribution) d) t-test

11-61 H_0 : the proportion of Gap Kid's stores is the same in the US, UK and Canada
H_1 : the proportion of Gap Kid's stores is not the same in the US, UK, and Canada

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
822	816.8904	5.1096	26.1080	0.0320
20	21.5376	−1.5376	2.3642	0.1182
31	34.6140	−3.6140	13.0610	0.4213
240	245.1096	−5.1096	26.1080	0.1088
8	6.4624	1.5376	2.3642	0.2955
14	10.3860	3.6140	13.0610	0.9329

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 1.9087$$

$\chi^2_{CRIT} = \chi^2_{.01;2} = 9.210$,
so we do not reject H_0 $\Rightarrow$ The proportion of Gap Kids stores appears to be independent of country. The emphasis is the same. This strategy makes sense if there is an equivalent proportion of children in each of the countries.

11-62 H_0 : the proportion of patents granted has not changed over the 10 year period
H_1 : the proportion of patents granted has changed over the 10 year period

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
39223	36641.0241	2581.9759	6666599.548	169.9666
51183	53767.9494	−2584.5374	6681963.400	130.5504
26548	29129.9759	−2581.9759	6666599.548	251.1149
45331	42746.0506	2584.9494	6681963.400	147.4038

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 699.0357$$

$\chi^2_{CRIT} = \chi^2_{.05;1} = 3.841$,

so we reject H_0 $\Rightarrow$ There was a significant change in the proportion of patents awarded over the last 10 years.

11-63 H_0 : the data is normally distributed ($\mu = 182.3$, $\sigma = 57$)
H_1: the data is not normally distributed

Z		−1.4439	−0.5667	0.3105	1.1877	2.0649
Concerts	≤ 100	101-150	151-200	201-250	251-300	≥ 300
Normal prob.	.0749	.2094	.3374	.2613	.0973	.0197
Observed	3	5	17	4	3	1
Expected	2.4717	6.9102	11.1342	8.6229	3.2109	.6501
$\frac{(f_o - f_e)^2}{f_e}$	.1129	.5280	3.0903	2.4784	.0139	.1883

$\chi^2 = 6.4118$ $df = 6 - 1 = 5$ $\chi^2_{CRIT} = \chi^2_{.025;5} = 12.833$
do not reject H_0 $\Rightarrow$ the assumption of normality is okay

11-64 a) t-test (t-distribution) b) F-distribution

c) normal d) χ^2

11-65

	n	$\bar{x}$	s^2
Brand A	6	58.833	6.967
Brand B	5	55.200	5.700
Brand C	4	49.250	4.250
Brand D	5	64.400	12.800

$$\bar{\bar{x}} = \frac{6(58.833) + 5(55.2) + 4(49.25) + 5(64.4)}{6 + 5 + 4 + 5} = 57.4$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{547.217}{3} = 182.4 \qquad \hat{\sigma}_w^2 = \frac{\sum (n_i-1)s_i^2}{n_T - k} = \frac{121.583}{16} = 7.6$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{182.4}{7.6} = 24.00$$

$F_{CRIT} = F_{.05}(3,16) = 3.24$, so reject H_0 $\Rightarrow$ conclude that prices do vary significantly from brand to brand

11-66

	n	$\bar{x}$	s^2
Billboard 1	8	34.0000	94.571
Billboard 2	9	28.5556	64.528
Billboard 3	7	32.7143	100.238

$$\bar{\bar{x}} = \frac{8(34) + 9(28.5556) + 7(32.7143)}{8 + 9 + 7} = 31.5833$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{138.1803}{2} = 69.09 \qquad \hat{\sigma}_w^2 = \frac{\sum (n_i-1)s_i^2}{n_T - k} = \frac{1779.649}{21} = 84.745$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{69.09}{84.745} = 0.82$$

$F_{CRIT} = F_{.05}(2,21) = 3.47$, so do not reject H_0 $\Rightarrow$ the three traffic volumes are not significantly different

11-67 a) $H_0 : \mu_S = \mu_B = \mu_M = \mu$

H_1 : not all $\mu_i = \mu$ (at least one $\mu_i - \mu_j \neq 0$)

b)

	n	$\bar{x}$	s^2
Stocks	6	3.5333	4.0707
Bonds	5	4.1	2.3250
Mutual Funds	4	3.875	2.0692

$$\bar{\bar{x}} = \frac{6(3.5333) + 5(4.1) + 4(3.875)}{6 + 5 + 4} = 3.8133$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{.8965}{2} = 0.4483 \qquad \hat{\sigma}_w^2 = \frac{\sum (n_i-1)s_i^2}{n_T - k} = \frac{35.8611}{12} = 2.9884$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{0.4483}{2.9884} = 0.15 \qquad F_{CRIT} = F_{.05}(2,12) = 3.89, \text{ so accept } H_0$$

c) Thus, the three rates of return are not significantly different.

11-68 a, b)

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
27	22.2	4.8	23.04	1.0378
48	55.5	−7.5	56.25	1.0135
15	12.3	2.7	7.29	0.5927
25	25.16	−0.16	0.0256	0.0010
63	62.9	0.1	0.01	0.0002
14	13.94	0.06	0.0036	0.0003
22	26.64	−4.64	21.5296	0.8082
74	66.6	7.4	54.76	0.8222
12	14.76	−2.76	7.6176	0.5161

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 4.7920$$

c) H_0: church attendance and income level are independent
H_1: church attendance and income level are dependent

d) $\chi^2_{CRIT} = \chi^2_{.05;4} = 9.488$, so accept H_0 $\Rightarrow$ church attendance seems to be unrelated to income level

11-69 H_0 : the three types of ships are equally likely to be on charter
H_1 : any of the three types of ships is less likely to be on charter

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
7	11.8734	−4.8734	23.75	2.0003
7	9.1749	−2.1749	4.7302	0.5156
20	12.9528	7.0472	49.663	3.8342
15	10.1266	4.8734	23.75	2.3453
10	7.8251	2.1749	4.7302	0.6045
4	11.0472	−7.0472	49.663	4.4955

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 13.7954$$

$\chi^2_{CRIT} = \chi^2_{.10;2} = 4.605$,

so reject H_0 $\Rightarrow$ the three types of ships are not equally likely to be on charter

11-70

	n	$\bar{x}$	s^2
Industrial	30	0.2417	0.3932
Transportation	20	0.0813	0.3343
Utility	15	0.2000	0.0844

$$\bar{\bar{x}} = \frac{30(.2417) + 20(.0813) + 15(.2)}{30 + 20 + 15} = 0.1827$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{.3145}{2} = 0.1573$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{19.748}{62} = 0.3185$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{.1573}{.3185} = .4939$$

$F_{CRIT} = F_{.05}\ (2{,}62) = 3.15,$

so do not reject H_0 $\Rightarrow$ the three groups did not have significantly different average changes in share prices on that particular day.

11-71 a)

Type of Car	Age Group 16-21	22-30	31-45	45+	Total
4 × 4 Off Road-Observed	19	23	15	2	59
-Expected	11.933	15.689	16.352	15.026	
Sports Car-Observed	9	14	11	7	41
-Expected	8.292	10.903	11.363	10.442	
Compact-Observed	6	8	7	9	30
-Expected	6.067	7.978	8.315	7.640	
Mid Size-Observed	11	13	19	24	67
-Expected	13.551	17.816	18.569	17.064	
Full Size-Observed	9	13	22	26	70
-Expected	14.157	18.614	19.401	17.828	
TOTAL	54	71	74	68	267

b)

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
19	11.933	7.067	49.942	4.185
23	15.689	7.311	53.451	3.407
15	16.352	−1.352	1.828	0.112
2	15.026	−13.026	169.677	11.292
9	8.292	0.708	0.501	0.060
14	10.903	3.097	9.591	0.880
11	11.363	−0.363	0.132	0.012
7	10.442	−3.442	11.847	1.135
6	6.067	−0.067	0.004	0.001
8	7.978	0.022	0.000	0.000
7	8.315	−1.315	1.729	0.208
9	7.640	1.360	1.850	0.242
11	13.551	−2.551	6.508	0.480
13	17.816	−4.816	23.194	1.302
19	18.569	0.431	0.186	0.010
24	17.064	6.936	48.108	2.819
9	14.157	−5.157	26.595	1.879
13	18.614	−5.614	31.517	1.693
22	19.401	2.599	6.755	0.348
26	17.828	8.172	66.782	3.746

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 33.881$$

c) H_0 : The type of car an individual drives does not depend on (is independent of) the age group to which he belongs.

H_1 : The type of car an individual drives is related to his age.

d) $df = (5-1)(4-1) = 12$. At $\alpha = .01$, the upper limit of the acceptance region is $\chi^2 = 26.217$, so we reject H_0 $\Rightarrow$ the type of car and age group are dependent.

11-72 a) Let x = the number of correct guesses

H_0 : x is distributed binomially ($n = 10$, $p = .5$)

H_1 : x is not distributed as above

b)

# of shifts	0-2	3-5	6-10
Binomial prob.	.0547	.5683	.3770
Observed	50	47	3
Expected	5.47	56.83	37.70
$\frac{(f_o - f_e)^2}{f_e}$	362.5084	1.7003	31.9387

(The last two categories are combined because $f_e = 1.07$ for 9-10 correct guesses.)

$$\chi^2_{OBS} = \sum \frac{(f_0 - f_e)^2}{f_e} = 396.147 \qquad \chi^2_{CRIT} = \chi^2_{.10;2} = 4.605$$

reject H_0 $\Rightarrow$ Swami Zhami's probability of guessing the correct card is not .5

c)

# of correct guesses	0-2	3-10
Binomial probability	.5256	.4744
Observed	50	50
Expected	52.56	47.44
$\frac{(f_o - f_e)^2}{f_e}$	.1247	.1381

(The last three categories are combined because $f_e < 5$.)

$$\chi^2_{OBS} = \sum \frac{(f_0 - f_e)^2}{f_e} = .2628 \qquad \chi^2_{CRIT} = \chi^2_{.10;1} = 2.706$$

do not reject H_0 $\Rightarrow$ Swami Zhami has no psychic power ($p = .25$)

11-73 $H_0 : \sigma^2 = 16^2$

$H_1 : \sigma^2 > 16^2$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{29(400)}{16^2} = 45.3125$$

$$\chi^2_{CRIT} = \chi^2_{.01;29} = 49.588$$

do not reject H_0 $\Rightarrow$ women do not show significantly greater variability on this attitude scale

11-74 H_0 : Jim's errors are $N(0,16)$

H_1 : Jim's errors follow another distribution

Errors	< -6.5		-6.5 to 0.5		0.5 to 6.5		> 6.5l
Z		-1.63		0.13		1.63	
Normal prob.	.0516		.5001		.3967		.0516
Observed	5		45		45		5
Expected	5.16		50.01		39.67		5.16
$\frac{(f_o - f_e)^2}{f_e}$	.0050		.5019		.7161		.0050

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 1.228 \qquad \chi^2_{CRIT} = \chi^2_{.05;3} = 7.815$$

do not reject H_0 $\Rightarrow$ Jim's errors are normally distributed with mean 0 and variance 16

11-75 $n_1 = 18$ $s_1^2 = 23.9$ $n_2 = 18$ $s_2^2 = 81.2$

$$H_0 : \sigma_2^2 = \sigma_1^2$$
$$H_1 : \sigma_2^2 > \sigma_1^2$$

$$F_{OBS} = \frac{s_2^2}{s_1^2} = \frac{81.2}{23.9} = 3.4 \qquad F_{CRIT} = F_{.05}(17,17) \doteq 2.25$$

reject H_0 $\Rightarrow$ the stress group has a significantly higher variance

11-76

	n	$\bar{x}$	s^2
NC	11	32.8727	4.1422
SA	11	34.8000	28.3660
SC	9	30.3889	3.7911

$$\bar{\bar{x}} = \frac{11(32.8727) + 11(34.8) + 9(30.3889)}{11 + 11+9} = 32.8355$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{96.3398}{2} = 48.1699$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{359.2019}{28} = 12.8286$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{48.1699}{12.8286} = 3.7549$$

$F_{CRIT} = F_{.05}\,(2,28) = 3.35,$

so reject H_0 $\Rightarrow$ the average ages in these regions do differ significantly.

11-77

	n	$\bar{x}$	s^2
MA	5	23.88	5.642
MN	5	19.9	61.09
NE	2	23.4	3.38
PA	7	19.7143	11.2781
NC	11	22.9364	18.8286
SA	11	22.7	7.506
SC	9	18	13.95

$$\bar{\bar{x}} = \frac{5(23.88) + 5(19.9) + 2(23.4) + 7(19.7143) + 11(22.9364) + 11(22.7) + 9(18)}{5 + 5 + 2 + 7 + 11 + 11 + 9} = 21.354$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{218.3881}{6} = 36.398$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i - 1)s_i^2}{n_T - k} = \frac{712.9226}{43} = 16.5796$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{36.398}{16.5796} = 2.1953$$

$F_{CRIT} = F_{.01}\,(6,43) = 3.26,$

so do not reject H_0 $\Rightarrow$ the average number of single person households does not differ significantly across the seven regions

11-78 H_0: Normal distribution
H_1: Some other distribution

No μ or σ are specified, so we'll estimate them by $\bar{x} = 1,764,857.8$ and $s = 409,322.2$. As suggested in exercise 11-17, we'll use 5 equally probable intervals, so $df = 5-1-2 = 2$. No significance level is specified.

The 20th, 40th, 60th, and 80th percentiles of any distribution divide that distribution into five equally probable intervals. For the standard normal distribution, those percentiles are $z = -0.84, -0.25, 0.25$, and 0.84. Since $x = \mu + \sigma z$, the corresponding percentiles for a normal distribution with $\mu = 1,764,857.8$ and $\sigma = 409,322.2$ are:

20th percentile: $1,764,857.8 - 409,322.2(0.84) = 1,421,027.2 = A$
40th percentile: $1,764,857.8 - 409,322.2(0.25) = 1,662,527.3 = B$
60th percentile: $1,764,857.8 + 409,322.2(0.25) = 1,867,188.4 = C$
80th percentile: $1,764,857.8 + 409,322.2(0.84) = 2,108,688.4 = D$

Sales interval	< A	A to B	B to C	C to D	> D
Observed	9	11	12	10	8
Expected	10	10	10	10	10
$\frac{(f_o - f_e)^2}{f_e}$	0.1	0.1	0.4	0.0	0.4

$\chi^2 = 1.0$; even with α as large as 0.20 (with $\chi^2_{CRIT} = \chi^2_{.20;2} = 3.219$) we would not reject H_0. Hence, the retail-sales data are well described by a normal distribution.

11-79 H_0 : style preferred is independent of occupation
H_1 : style preferred is related to occupation
The table below contains observed frequency, expected frequency $= \frac{(RT \times CT)}{n}$, and cell chi-square $= \frac{(f_o - f_e)^2}{f_e}$, in that order.

Profession	Style A	Style B	Style C	Style D	
Realtor	5 4.3750 0.0893	7 6.0000 0.1667	6 7.375 0.2564	8 8.250 0.0076	26
Secretary	10 7.5721 0.7785	15 10.3846 2.0513	12 12.7644 0.0458	8 14.2788 2.7610	45
Entrepreneur	8 11.1058 0.8686	12 15.2308 0.6853	21 18.7212 0.2774	25 20.9423 0.7862	66
Account Executive	12 11.9471 0.0002	14 16.3846 0.3471	20 20.1394 0.0010	25 22.5288 0.2711	71
	35	48	59	66	

$\chi^2_{OBS} = 9.3932$ $\quad df = (4-1)(4-1) = 9$ $\quad \chi^2_{CRIT} = \chi^2_{.10;9} = 14.684$

do not reject H_0 $\Rightarrow$ women with different occupations do not respond differently to fashion brands

11-80

	n	$\bar{x}$	s^2
Drug 1	4	245.75	72.9167
Drug 2	4	272.50	41.6667
Drug 3	5	229.00	138.5000
Drug 4	5	243.40	39.3000

$$\bar{\bar{x}} = \frac{4.(245.75) + 4(272.5) + 5(229) + 5(243.4)}{4+4+5+5} = 246.3889$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{4285.3278}{3} = 1428.4426$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i-1)s_i^2}{n_T - k} = \frac{1054.95}{14} = 75.3536$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{1428.4426}{75.3536} = 18.96$$

$F_{CRIT} = F_{.05}(3,14) = 3.34$, so reject H_0 $\Rightarrow$ the four drugs do affect driving skill differently

11-81 $H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

$$F_{OBS} = \frac{984}{1136} = .87$$

F_{CRIT} values are : $F_{.01}(30,40) = 2.20$

$$F_{.99}(30,40) = \frac{1}{F_{.01}(40,30)} = \frac{1}{2.30} = 0.43$$

F_{OBS} lies between these two values, so do not reject H_0. The two mills have about the same variability (or at any rate not significantly different variability).

11-82

	n	$\bar{x}$	s^2
Aircraft Type A	5	7.6	0.345
Aircraft Type B	3	6.8	1.120
Aircraft Type C	6	8.7	0.404

$$\bar{\bar{x}} = \frac{5(7.6) + 3(6.8) + 6(8.7)}{5+3+6} = 7.9$$

$$\hat{\sigma}_b^2 = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1} = \frac{7.92}{2} = 3.96$$

$$\hat{\sigma}_w^2 = \frac{\sum (n_i-1)s_i^2}{n_T - k} = \frac{5.64}{11} = 0.513$$

$$F = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_w^2} = \frac{3.96}{0.513} = 7.72$$

$F_{CRIT} = F_{.01}(2,11) = 7.21$, so reject H_0 $\Rightarrow$ there are significantly different fuel costs per mile among the three types of aircraft

11-83 H_0: $\mu_1 = \mu_2 = ... = \mu_6$ (the six divisions have the same batting averages).

H_1: Not all μ_i are the same (the six divisions have different batting averages).

From Minitab's ONEWAY analysis of variance, we get $F = 0.300$, with a prob-value of 0.9103. Since this is more than our significance level of $\alpha = 0.05$, we do not reject H_0. Batting averages do not differ significantly across the six divisions.

11-84 H_0: $\sigma_A^2 = \sigma_N^2$ (variability the same in both leagues)

H_0: $\sigma_A^2 > \sigma_N^2$ (more variability in the American League)

$n_A = 26$ $\sigma_A^2 = 0.001499$ $n_N = 24$ $\sigma_N^2 = 0.001372$

$F = \dfrac{\sigma_A^2}{\sigma_N^2} = 1.09$, with a prob-value of 0.4171 for this upper-tailed test.

Since this is greater than our significance level of $\alpha = 0.10$, we do not reject H_0. Batting skills are not significantly more variable in the American League. Dick is correct.

11-85 a)

Type of car	Newspaper *N&O*	*CHN*	*VA*	Total
Domestic:				
f_o	543	32	36	611
f_e	528.7248	36.3847	45.8906	
$\frac{(f_o - f_e)^2}{f_e}$	0.3854	0.5284	2.1317	
Foreign				
f_o	576	59	73	708
f_e	612.6631	42.1609	53.1760	
$\frac{(f_o - f_e)^2}{f_e}$	2.1940	6.7256	7.3904	
Light trucks/vans				
f_0	494	20	31	545
f_e	471.6121	32.4544	40.9335	
$\frac{(f_o - f_e)^2}{f_e}$	1.0628	4.7794	2.4106	
Total	1613	111	140	1864

H_0 : Type of car ad and newspaper are independent
H_1 : Type of car ad and newspaper are not independent

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 27.6083 \qquad df = 2 \times 2 = 4 \qquad \chi^2_{CRIT} = \chi^2_{.01;4} = 13.277$$

Reject H_0 $\Rightarrow$ the proportions of the three types of ads vary significantly from paper to paper

b) Not really. Suppose you are interested in buying a foreign car. Although the proportions of the ads vary from paper to paper, and the proportion of foreign ads is least for the *N&O*, still that doesn't suggest not consulting the *N&O*. After all, the *N&O* has over four times as many foreign cars listed as do the other two newspapers taken together.

CHAPTER 12

SIMPLE REGRESSION AND CORRELATION

NOTE: In hand-calculated regressions, all intermediate results are carried to only 4 decimal places, resulting in rounding errors in some of the solutions. For some problems, 4-place solutions from SAS are also given for comparison purposes.

12-1 It is the process of determining the relationship between a dependent and an independent variable.

12-2 An estimating equation is the formula describing the relationship between a dependent variable and one or more independent variables.

12-3 Correlation analysis is the tool used to describe the degree to which the dependent variable is related to the independent variable(s).

12-4 In a "direct (indirect)" relationship the dependent variable increases (decreases) as the independent variable increases.

12-5 A causal relationship is an association in which the change in the dependent variable is caused by the change in the independent variable(s).

12-6 In a linear relationship, the dependent variable changes a constant amount for equal incremental changes in the independent variable(s). In a curvi-linear relationship, the dependent variable does not change at a constant rate with equal incremental changes in the independent variable(s).

12-7 A scatter diagram is constructed, by recording paired points of data on a graph, to determine whether or not there is a relationship between a dependent and an independent variable; if there is a relationship, then the pattern of points will approximate a line.

12-8 It is a process which determines the relationship between a dependent variable and more than one independent variable.

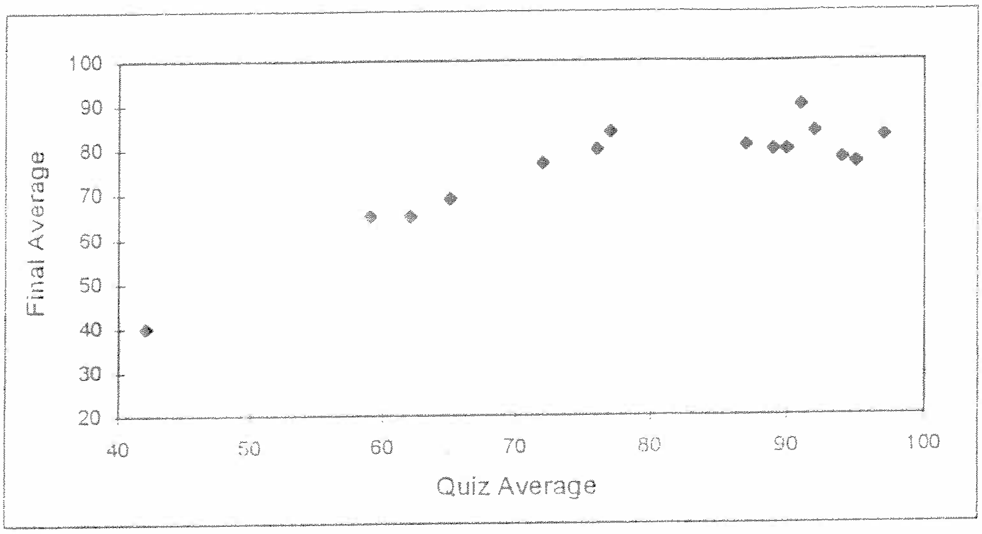
12-9 a) yes, direct and linear

b) yes, inverse and curvilinear

c) yes, inverse and curvilinear

12-10 a) We want to see if Final Average (FA) depends on Quiz Average (QA), so FA is the dependent variable and QA is the independent variable.

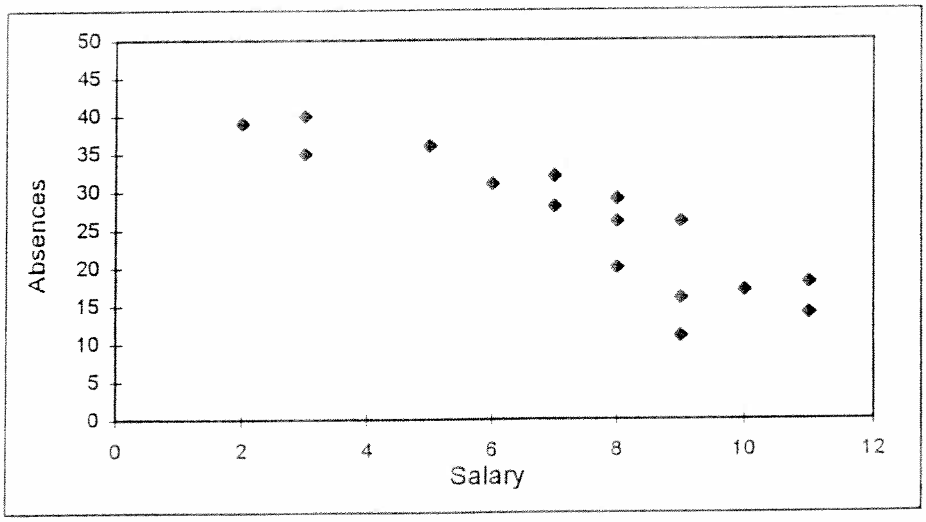
b)



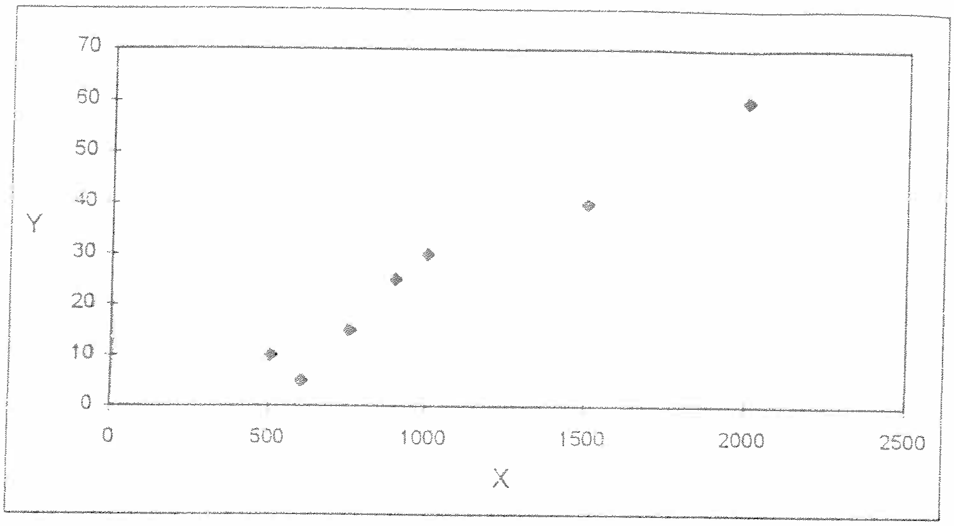
c) curvilinear

d) For the most part, the professor is right. However, for very high quiz averages, the curve appears to be turning down.

12-11

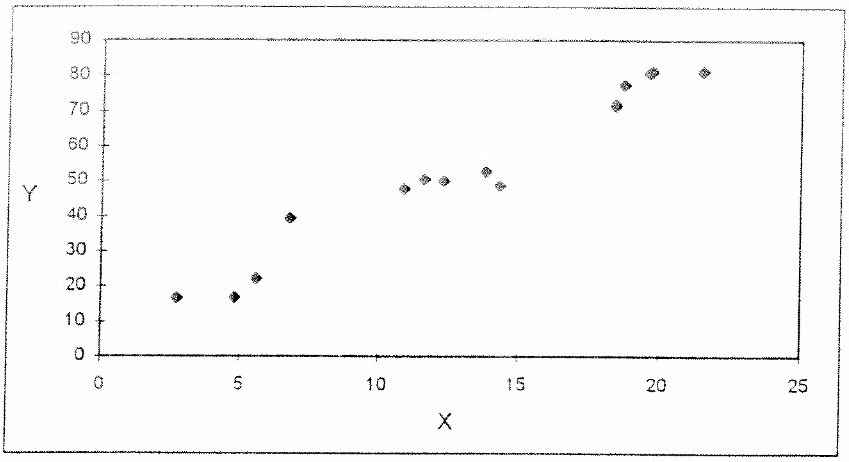


12-12



A direct linear relationship seems to exist. Clearly it is absurd to suggest that the use of facial tissues causes the common cold. (But the opposite may well be true.)

12-13 a)



b)

X	Y	XY	X^2
2.7	16.66	44.982	7.29
4.8	16.92	81.216	23.04
5.6	22.30	124.880	31.36
18.4	71.80	1321.120	338.56
19.6	80.88	1585.248	384.16
21.5	81.40	1750.100	462.25
18.7	77.46	1448.502	349.69
14.3	48.70	696.410	204.49
11.6	50.48	585.568	134.56
10.9	47.82	521.238	118.81
18.4	71.50	1315.600	338.56
19.7	81.26	1600.822	388.09
12.3	50.10	616.230	151.29
6.8	39.40	267.920	46.24
13.8	52.80	728.640	190.44
$\sum X = 199.1$	$\sum Y = 809.48$	$\sum XY = 12688.476$	$\sum X^2 = 3168.83$

$\overline{X} = 199.1/15 = 13.2733$ $\qquad \overline{Y} = 809.48/15 = 53.9653$

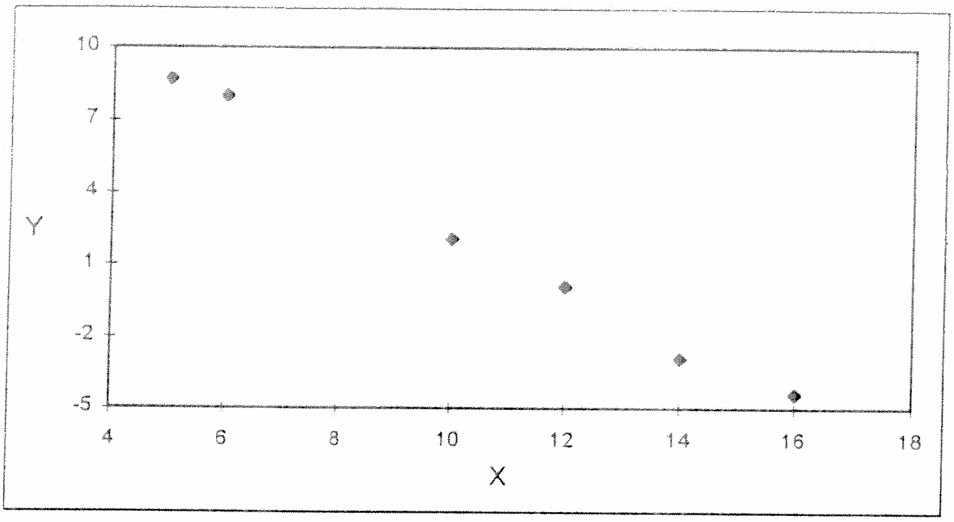
$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{12688.476 - 15(13.2733)(53.9653)}{3168.83 - 15(13.2733)^2} = 3.6950$$

$a = \overline{Y} - b\overline{X} = 53.9653 - 3.6950(13.2733) = 4.9205$

Thus, $\widehat{Y} = 4.9205 + 3.6950X$ $\qquad$ (SAS: $\widehat{Y} = 4.9203 + 3.9650X$)

c) $X = 6.0,\ \widehat{Y} = 4.9203 + 3.6950(6.0) = 27.0903$
$X = 13.4,\ \widehat{Y} = 4.9203 + 3.6950(13.4) = 54.4333$
$X = 20.5,\ \widehat{Y} = 4.9203 + 3.6950(20.5) = 80.6678$

12-14 a)



b)

X	Y	XY	X^2
16	−4.4	−70.4	256
6	8.0	48.0	36
10	2.1	21.0	100
5	8.7	43.5	25
12	0.1	1.2	144
14	−2.9	−40.6	196
$\sum X = 63$	$\sum Y = 11.6$	$\sum XY = 2.7$	$\sum X^2 = 757$

$\overline{X} = \frac{63}{6} = 10.5$ $\qquad \overline{Y} = \frac{11.6}{6} = 1.9333$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{2.7 - 6(10.5)(1.9333)}{757 - 6(10.5)^2} = -1.2471$$

$a = \overline{Y} - b\overline{X} = 1.9333 - (-1.2471)(10.5) = 15.0279$

Thus, $\widehat{Y} = 15.0279 - 1.2471X$ $\qquad$ (SAS: $\widehat{Y} = 15.0281 - 1.2471X$)

c) $X = 5,\ \widehat{Y} = 15.0279 - 1.2471(5) = 8.7924$

$X = 6,\ \widehat{Y} = 15.0279 - 1.2471(6) = 7.5453$

$X = 7,\ \widehat{Y} = 15.0279 - 1.2471(7) = 6.2982$

12-15 a)

X	Y	XY	X^2	Y^2
56	45.0	2520.0	3136	2025.00
48	38.5	1848.0	2304	1482.25
42	34.5	1449.0	1764	1190.25
58	46.1	2673.8	3364	2125.21
40	33.3	1332.0	1600	1108.89
39	32.1	1251.9	1521	1030.41
50	40.4	2020.0	2500	1632.16
$\sum X = 333$	$\sum Y = 269.9$	$\sum XY = 13094.7$	$\sum X^2 = 16189$	$\sum Y^2 = 10594.17$

$\overline{X} = \frac{333}{7} = 47.5714$ $\qquad$ $\overline{Y} = \frac{269.9}{7} = 38.5571$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{13094.7 - 7(47.5714)(38.5571)}{16189 - 7(47.5714)^2} = 0.7339$$

$$a = \overline{Y} - b\overline{X} = 38.5571 - 0.7339(47.5714) = 3.6444$$

Thus, $\widehat{Y} = 3.6444 + 0.7339X$ $\quad$ (SAS: $\widehat{Y} = 3.6467 + 0.7339X$)

b) $$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}} = \sqrt{\frac{10594.17 - (3.6444)(269.9) - 0.7339(13094.7)}{7-2}}$$
$= 0.2631$ (SAS: 0.2602)

c) $X = 44$ $\qquad$ $t = 2.571$ $\qquad$ $\widehat{Y} = 3.6444 + 0.7339(44) = 35.9360$

$\widehat{Y} \pm t(s_e) = 35.9360 \pm 2.571(0.2631) = 35.9360 \pm 0.6764$

12-16 a)

X(housing starts)	Y(appliance sales)	XY	X^2	Y^2
2.0	5.0	10.00	4.00	25.00
2.5	5.5	13.75	6.25	30.25
3.2	6.0	19.20	10.24	36.00
3.6	7.0	25.20	12.96	49.00
3.3	7.2	23.76	10.89	51.84
4.0	7.7	30.80	16.00	59.29
4.2	8.4	35.28	17.64	70.56
4.6	9.0	41.40	21.16	81.00
4.8	9.7	46.56	23.04	94.09
5.0	10.0	50.00	25.00	100.00
$\sum X = 37.2$	$\sum Y = 75.5$	$\sum XY = 295.95$	$\sum x^2 = 147.18$	$\sum Y^2 = 597.03$

$\overline{X} = \frac{37.2}{10} = 3.72$ $\qquad$ $\overline{Y} = \frac{75.5}{10} = 7.55$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{295.95 - 10(3.72)(7.55)}{147.18 - 10(3.72)^2} = 1.7156$$

$$a = \overline{Y} - b\overline{X} = 7.55 - 1.7156(3.72) = 1.1680$$

Thus, $\widehat{Y} = 1.1680 + 1.7156X$ $\quad$ (SAS: $\widehat{Y} = 1.1681 + 1.7156X$)

b) When housing starts go up 1000 units, appliance sales go up 1.7156 thousand units.

c) $$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}} = \sqrt{\frac{597.03 - 1.1680(75.5) - 1.7156(295.95)}{8}}$$
$= 0.3732$ (SAS: 0.3737)

(The standard deviation at the data points around the regression line is about 370 appliances.)

d) $\widehat{Y} = 1.1680 + 1.7156(8) = 14.89$

$\widehat{Y} \pm t_{\alpha/2;n\text{-}2}\, s_e = 14.89 \pm 1.860(.3732) = 14.89 \pm .694 = (14.20, 15.58)$ thousand units.

12-17 a) We are trying to predict the number of lobs not returned based on the height of the opponent. Thus, Y = lobs not returned and X = opponent's height.

b)

X	Y	XY	X^2
5.0	9	45.0	25.00
5.5	6	33.0	30.25
6.0	3	18.0	36.00
6.5	0	0.0	42.25
5.0	7	35.0	25.00
$\sum X = 28.0$	$\sum Y = 25$	$\sum XY = 131.0$	$\sum X^2 = 158.50$

$\overline{X} = \frac{28.0}{5} = 5.6$ $\qquad$ $\overline{Y} = \frac{25.0}{5} = 5.0$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{131.0 - 5(5.6)(5.0)}{158.5 - 5(5.6)^2} = -5.2941$$

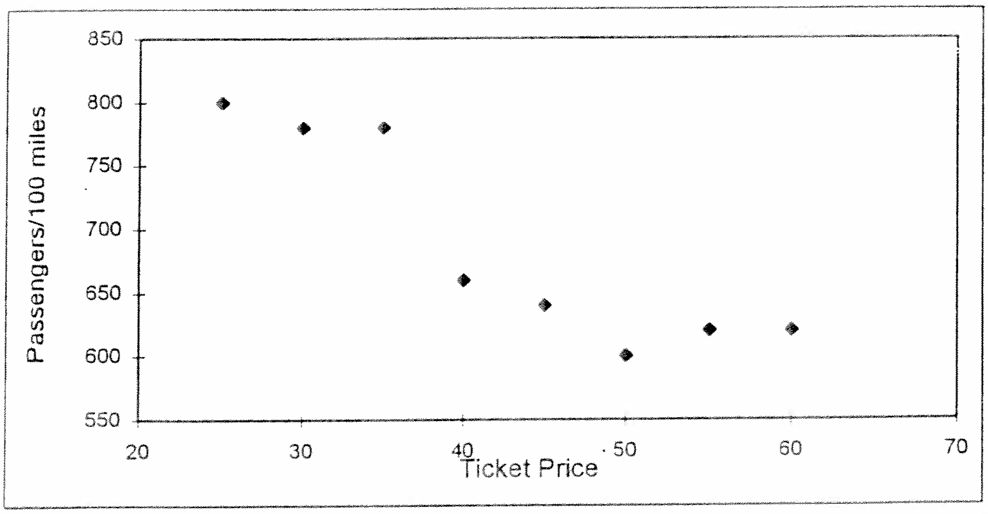
$a = \overline{Y} - b\overline{X} = 5.0 - (-5.2941)(5.6) = 34.6470$

Thus, $\widehat{Y} = 34.6470 - 5.2941X$ (SAS: $\widehat{Y} = 34.6471 - 5.2941X$)

c) $\widehat{Y} = 34.6470 - 5.2941(5.9) = 3.4$

12-18 In this problem, Y = passengers per 100 miles and X = ticket price.

a)



b)

X	Y	XY	X^2	Y^2
25	800	20000	625	640000
30	780	23400	900	608400
35	780	27300	1225	608400
40	660	26400	1600	435600
45	640	28800	2025	409600
50	600	30000	2500	360000
55	620	34100	3025	384400
60	620	37200	3600	384400
$\sum X = 340$	$\sum Y = 5500$	$\sum XY = 227200$	$\sum X^2 = 15500$	$\sum Y^2 = 3830800$

$\overline{X} = 340/8 = 42.5$ $\qquad\qquad$ $\overline{Y} = 5500/8 = 687.5$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{227200 - 8(42.5)(687.5)}{15500 - 8(42.5)^2} = -6.2381$$

$a = \overline{Y} - b\overline{X} = 687.5 - (-6.2381)(42.5) = 952.6193$

Thus, $\widehat{Y} = 952.6193 - 6.2381X$ $\quad$ (SAS: $\widehat{Y} = 952.6190 - 6.2381X$)

c)

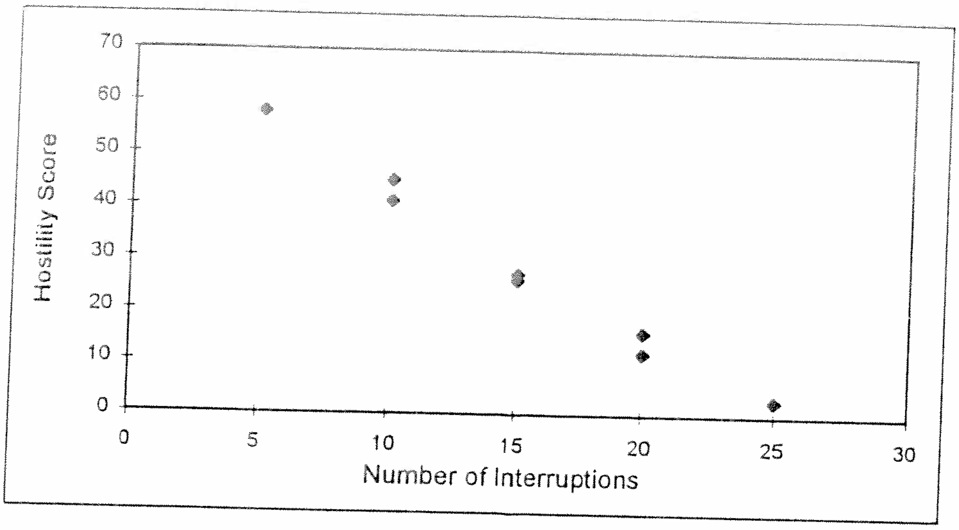
$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{3830800 - 952.6193(5500) - (-6.2381)(227200)}{6}} = 38.0573 \quad \text{(SAS: 38.0580)}$$

$\widehat{Y} = 952.6193 - 6.2381(50) = 640.7143$

$$\begin{aligned}\widehat{Y} \pm t_{\alpha/2;n\text{-}2}s_e &= 640.7143 \pm 2.447(38.0573)\\ &= 640.7143 \pm 93.1262 = (547.5881,\ 733.8405)\end{aligned}$$

12-19 a)



b)

X	Y	XY	X^2
5	58	290	25
10	41	410	100
10	45	450	100
15	27	405	225
15	26	390	225
20	12	240	400
20	16	320	400
25	3	75	625
$\sum X = 120$	$\sum Y = 228$	$\sum XY = 2580$	$\sum X^2 = 2100$

$\overline{X} = \frac{120}{8} = 15$ $\qquad\qquad$ $\overline{Y} = \frac{228}{8} = 28.5$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{2580 - 8(15)(28.5)}{2100 - 8(15)^2} = -2.8$$

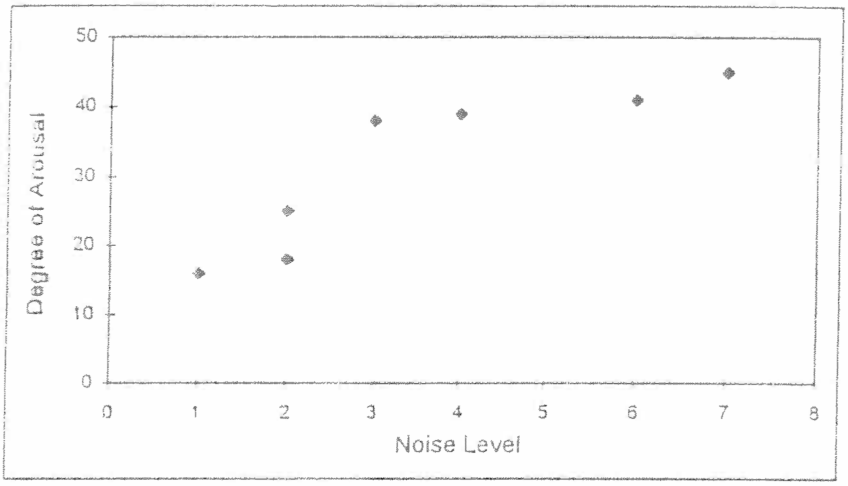
$a = \overline{Y} - b\overline{X} = 28.5 - (-2.8)(15) = 70.5$

Thus, $\widehat{Y} = 70.5 - 2.8X$

c) $\widehat{Y} = 70.5 - 2.8(18) = 20.1$

12-20 In this problem, Y = degree of arousal and X = noise level.

a)



b)

X	Y	XY	X^2
4	39	156	16
3	38	114	9
1	16	16	1
2	18	36	4
6	41	246	36
7	45	315	49
2	25	50	4
3	38	114	9
$\sum X = 28$	$\sum Y = 260$	$\sum XY = 1047$	$\sum X^2 = 128$

$\overline{X} = \frac{28}{8} = 3.5$ $\quad\quad$ $\overline{Y} = \frac{260}{8} = 32.5$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{1047 - 8(3.5)(32.5)}{128 - 8(3.5)^2} = 4.5667$$

$a = \overline{Y} - b\overline{X} = 32.5 - 4.5667(3.5) = 16.5166$

Thus, $\widehat{Y} = 16.5166 + 4.5667X$ (SAS: $\widehat{Y} = 16.5167 - 4.5667X$)

c) $\widehat{Y} = 16.5166 + 4.5667(5) = 39.35$

12-21 a) Here,. Y = units sold and X = test score.

X	Y	XY	X^2
2.6	95	247	6.76
3.7	140	518	13.69
2.4	85	204	5.76
4.5	180	810	20.25
2.6	100	260	6.76
5.0	195	975	25.00
2.8	115	322	7.84
3.0	136	408	9.00
4.0	175	700	16.00
3.4	150	510	11.56
$\sum X = 34.0$	$\sum Y = 1371$	$\sum XY = 4954$	$\sum X^2 = 122.62$

$$\overline{X} = \frac{34.0}{10} = 3.40 \qquad \overline{Y} = \frac{1371}{10} = 137.1$$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{4954 - 10(3.4)(137.1)}{122.62 - 10(3.4)^2} = 41.6809$$

$$a = \overline{Y} - b\overline{X} = 137.1 - 41.6809(3.4) = -4.6151$$

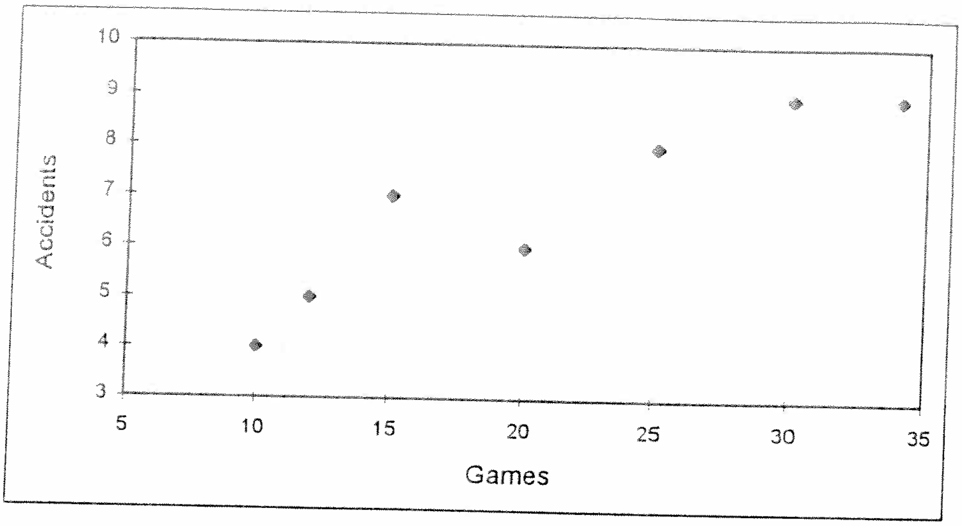
Thus, $\widehat{Y} = -4.6151 + 41.6809X$ (SAS: $\widehat{Y} = -4.6151 + 41.6809X$)

b) Change in Y for a one-unit change in X = slope = b = 41.6809 more units sold.

c) When $\overline{X}$ is plugged into the equation, $\widehat{Y} = \overline{Y} = 137.1$ units sold.

12-22 In this problem, Y = number of accidents per game, X = number of games played.

a)



b)

X	Y	XY	X^2	Y^2
20	6	120	400	36
30	9	270	900	81
10	4	40	100	16
12	5	60	144	25
15	7	105	225	49
25	8	200	625	64
34	9	306	1156	81
$\sum X = 146$	$\sum Y = 48$	$\sum XY = 1101$	$\sum X^2 = 3550$	$\sum Y^2 = 352$

$$\overline{X} = \frac{146}{7} = 20.8571 \qquad \overline{Y} = \frac{48}{7} = 6.8571$$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{1101 - 7(20.8571)(6.8671)}{3550 - 7(20.8571)^2} = 0.1978$$

$$a = \overline{Y} - b\overline{X} = 6.8571 - 0.1978(20.8571) = 2.7316$$

Thus, $\widehat{Y} = 2.7316 + 0.1978X$ (SAS: $\widehat{Y} = 2.7317 + 0.1978X$)

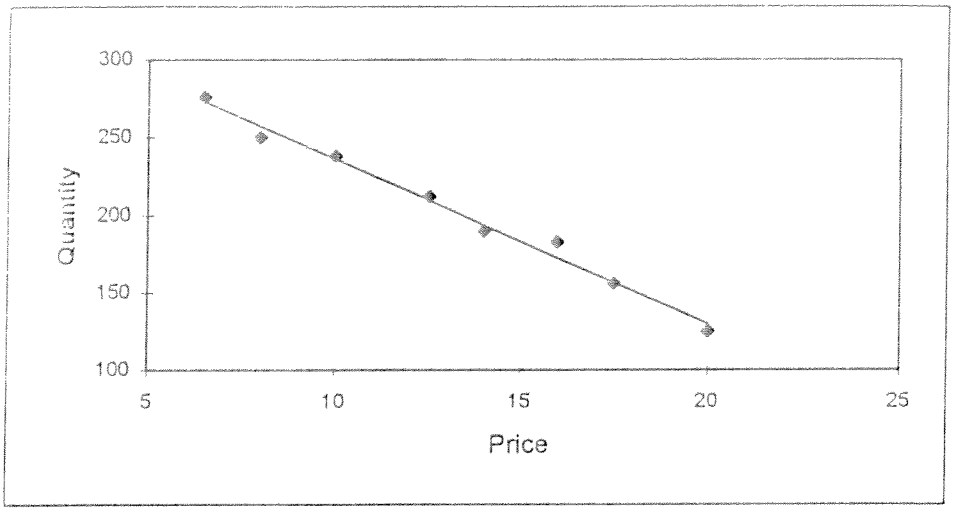
c) $\widehat{Y} = 2.7316 + 0.1978(33) = 9.3$

d) $$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{352 - 2.7317(48) - 0.1978(1101)}{5}} = 0.7875 \quad \text{(SAS: } 0.7882)$$

12-23 In this problem, Y = quantity sold and X = price.

a), c)



b)

X	Y	XY	X^2
20.0	125	2500	400.00
17.5	156	2730	306.25
16.0	183	2928	256.00
14.0	190	2660	196.00
12.5	212	2650	156.25
10.0	238	2380	100.00
8.0	250	2000	64.00
6.5	276	1794	42.25
$\sum X = 104.5$	$\sum Y = 1630$	$\sum XY = 19642$	$\sum X^2 = 1520.75$

$$\overline{X} = \frac{104.5}{8} = 13.0625 \qquad \overline{Y} = \frac{1630}{8} = 203.75$$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{19642 - 8(13.0625)(203.75)}{1520.75 - 8(13.0625)^2} = -10.5952$$

$$a = \overline{Y} - b\overline{X} = 203.75 - (-10.5952)(13.0625) = 342.1498$$

Thus, $\widehat{Y} = 342.1498 - 10.5952X$ (SAS: $\widehat{Y} = \mathit{342.1501 - 10.5952X}$)

12-24 In this problem, Y = percent of pollutants and X = money spent.

a)

X	Y	XY	X^2	Y^2
8.4	35.9	301.56	70.56	1288.81
10.2	31.8	324.36	104.04	1011.24
16.5	24.7	407.55	272.25	610.09
21.7	25.2	546.84	470.89	635.04
9.4	36.8	345.92	88.36	1354.24
8.3	35.8	297.14	68.89	1281.64
11.5	33.4	384.10	132.25	1115.56
18.4	25.4	467.36	338.56	645.16
16.7	31.4	524.38	278.89	985.96
19.3	27.4	528.82	372.49	750.76
28.4	15.8	448.72	806.56	249.64
4.7	31.5	148.05	22.09	992.25
12.3	28.9	355.47	151.29	835.21
$\sum X = 185.8$	$\sum Y = 384.0$	$\sum XY = 5080.27$	$\sum X^2 = 3177.12$	$\sum Y^2 = 11755.60$

$$\overline{X} = \frac{185.8}{13} = 14.2923 \qquad \overline{Y} = \frac{384.0}{13} = 29.5385$$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{5080.27 - 13(14.2923)(29.5385)}{3177.12 - 13(14.2923)^2} = -0.7822$$

$$a = \overline{Y} - b\overline{X} = 29.5385 + 0.7822(14.2923) = 40.7179$$

Thus, $\widehat{Y} = 40.7179 - 0.7822X$ (SAS: $\widehat{Y} = 40.7172 - 0.7821X$)

b) $\widehat{Y} = 40.7179 - 0.7822(20) = 25.0739$, so 25.0739 percent of the emissions will be dangerous pollutants.

c) $$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}} = \sqrt{\frac{11755.6 - 40.7179(384.0) - (-0.7822)(5080.27)}{11}}$$
$= 2.9188$

12-25 a) Positive b) Positive c) Positive d) Zero

12-26 $$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{34.6470(25) + (-5.2941)(131) - 6(5)^2}{175 - 5(5)^2} = 0.9530 \text{ (SAS:0.9529)}$$

$$r = -\sqrt{0.9530} = -0.9762$$

12-27 $$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{952.6193(5500) + (-6.2381)(227200) - 8(687.5)^2}{3830800 - 8(687.5)^2} = 0.8246$$

$$r = -\sqrt{0.8246} = -0.9081$$

12-28

Y^2
3364
1681
2025
729
676
144
256
9
8884

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$= \frac{70.5(228) + (-2.8)(2580) - 8(28.5)^2}{8884 - 8(28.5)^2} = .9858$$

$$r = -\sqrt{.9858} = -.9929$$

12-29 $r^2 = \dfrac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \dfrac{16.5166(260) + 4.5667(1047) - 8(32.5)^2}{9320 - 8(32.5)^2} = 0.7191$

$r = \sqrt{0.7191} = 0.8480$

12-30 $r^2 = \dfrac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$

$= \dfrac{-4.6151(1371) + 41.6809(4954) - 10(137.1)^2}{201121 - 10(137.1)^2} = 0.9269$ (SAS: .9270)

$r = \sqrt{0.9269} = 0.9628$

12-31 In this problem, Y = waiting time and X = number of bankers.

a)

X	Y	XY	X^2	Y^2
2	12.8	25.6	4	163.84
3	11.3	33.9	9	127.69
5	3.2	16.0	25	10.24
4	6.4	25.6	16	40.96
2	11.6	23.2	4	134.56
6	3.2	19.2	36	10.24
1	8.7	8.7	1	75.69
3	10.5	31.5	9	110.25
4	8.2	32.8	16	67.24
3	11.3	33.9	9	127.69
3	9.4	28.2	9	88.36
2	12.8	25.6	4	163.84
4	8.2	32.8	16	67.24
$\sum X = 42$	$\sum Y = 117.6$	$\sum XY = 337.0$	$\sum X^2 = 158$	$\sum Y^2 = 1182.84$

$\overline{X} = 42/13 = 3.2308$ $\overline{Y} = 1.76/13 = 9.0462$

$b = \dfrac{\sum XY - n\overline{XY}}{\sum X^2 - n\overline{X}^2} = \dfrac{337.0 - 13(3.2308)(9.0462)}{158 - 13(3.2308)^2} = -1.9253$

$a = \overline{Y} - b\overline{X} = 9.0462 - (-1.9253)(3.2308) = 15.2665$

Thus, $\widehat{Y} = 15.2665 - 1.9253X$ (SAS: $\widehat{Y} = 15.2648 - 1.9248X$)

b) $r^2 = \dfrac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \dfrac{15.2665(117.6) - 1.9253(337.0) - 13(9.0462)^2}{1187.84 - 13(9.0462)^2} = 0.6667$

$r = -\sqrt{0.6667} = -0.8165$

12-32 a)

X	Y	XY	X^2	Y^2
3	11	33	9	121
7	18	126	49	324
4	9	36	16	81
2	4	8	4	16
0	7	0	0	49
4	6	24	16	36
1	3	3	1	9
2	8	16	4	64
$\sum X = 23$	$\sum Y = 66$	$\sum XY = 246$	$\sum X^2 = 99$	$\sum Y^2 = 700$

$\overline{X} = 23/8 = 2.875 \qquad \overline{Y} = 66/8 = 8.25$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{246 - 8(2.875)(8.25)}{99 - 8(2.875)^2} = 1.7110$$

$a = \overline{Y} - b\overline{X} = 8.25 - 1.7110(2.875) = 3.3309$

Thus, $\widehat{Y} = 3.3309 + 1.7110X$ (SAS: $\widehat{Y} = 3.3308 + 1.7110X$)

b) $$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{3.3309(66) + 1.7110(246) - 8(8.25)^2}{700 - 8(8.25)^2} = 0.6189$$

$r = \sqrt{0.6189} = 0.7867$

12-33 a) $$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{8.516}{\sqrt{327.52}} = 0.4706$$

b) $H_0 : B = 0 \qquad H_1 : B \neq 0 \qquad \alpha = 0.05$

The limits of the acceptance region are $B \pm t(s_b) = 0 \pm 2.069(0.4706)$
$= 0 \pm .97 = [-0.97, 0.97]$

Since $b = 1.12$, we reject $H_0 \Rightarrow$ the coefficient is significantly different from 0.

c) The 95% confidence interval is $b \pm t(s_b) = 1.12 \pm 2.069(0.4706)$
$= 1.12 \pm 0.97 = [0.15, 2.09]$

12-34

X	Y	XY	X^2	Y^2
3.6	12.13	43.668	12.96	147.1369
4.8	14.70	70.560	23.04	216.0900
9.7	22.83	221.451	94.09	521.2089
12.6	28.40	357.840	158.76	806.5600
11.5	28.33	325.795	132.25	802.5889
10.9	27.05	294.845	118.81	731.7025
14.6	33.60	490.560	213.16	1128.9600
18.2	40.80	742.560	331.24	1664.6400
3.7	9.40	34.780	13.69	88.3600
9.8	24.84	243.432	96.04	617.0256
12.4	30.17	374.108	153.76	910.2289
16.9	34.70	586.430	285.61	1204.0900
$\sum X = 128.7$	$\sum Y = 306.95$	$\sum XY = 3786.029$	$\sum X^2 = 1633.41$	$\sum Y^2 = 8838.5920$

$\overline{X} = \frac{128.7}{12} = 10.7250 \qquad \overline{Y} = \frac{306.95}{12} = 25.5792$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{3786.029 - 12(10.7250)(25.5792)}{1633.41 - 12(10.7250)^2} = 1.9517$$

$a = \overline{Y} - b\overline{X} = 25.5792 - 1.9517(10.7250) = 4.6472$ (SAS: $a = 4.6468$, $b = 1.9517$)

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{8838.592 - 4.6472(306.95) - 1.9517(3786.029)}{10}} = 1.5146 \quad \text{(SAS: 1.5141)}$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{1.5146}{\sqrt{1633.41 - 12(10.7250)^2}} = 0.0952$$

$H_0 : B = 1.5 \qquad H_1 : B > 1.5 \qquad \alpha = .05$

The upper limit of the acceptance region is $B + t(s_b) = 1.5 + 1.812(0.0952) = 1.67$
Here, $b = 1.9517 > 1.67 \Rightarrow$ reject $H_0 \Rightarrow$ Ned should advertise.

12-35 $H_0 : B = 3.2$ $H_1 : B \neq 3.2$ $\alpha = .05$

The limits of the acceptance region are $B \pm t(s_b) = 3.2 \pm 2.160(0.18) = 3.2 \pm .39 = [2.81, 3.59]$

Here, $b = 2.9 \Rightarrow$ do not reject $H_0 \Rightarrow$ slope is not significantly different from 3.2

12-36 $H_0 : B = 1.50$ $H_1 : B \neq 1.50$ $\alpha = .05$

The limits of the acceptance region are $B - t(s_b) = 1.5 \pm 2.069(.11) = 1.5 \pm .23 = [1.27, 1.73]$

Here, $b = 1.685 \Rightarrow$ do not reject $H_0 \Rightarrow$ slope has not changed significantly from its past value.

12-37 In this problem, $Y =$ value and $X =$ area.

a)

X	Y	XY	X^2	Y^2
1.1	75	82.5	1.21	5625
1.5	95	142.5	2.25	9025
1.6	110	176.0	2.56	12100
1.6	102	163.2	2.56	10404
1.4	95	133.0	1.96	9025
1.3	87	113.1	1.69	7569
1.1	82	90.2	1.21	6724
1.7	115	195.5	2.89	13225
1.9	122	231.8	3.61	14884
1.5	98	147.0	2.25	9604
1.3	90	117.0	1.69	8100
$\sum X = 16.0$	$\sum Y = 1071$	$\sum XY = 1591.8$	$\sum X^2 = 23.88$	$\sum Y^2 = 106285$

$$\overline{X} = \frac{16.0}{11} = 1.45455 \qquad \overline{Y} = \frac{1071}{11} = 97.36364$$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{1591.8 - 11(1.45455)(97.36364)}{23.88 - 11(1.45455)^2} = 55.9634$$

$$a = \overline{Y} - b\overline{X} = 97.36364 - 55.9634(1.45455) = 15.9621$$

Thus, $\widehat{Y} = 15.9621 + 55.9634X$ (SAS: $\widehat{Y} = 15.9701 + 55.9581X$)

b) $H_0 : B = 50$ $H_1 : B \neq 50$ $\alpha = .10$

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{106285 - 15.9621(1071) - 55.9634(1591.8)}{11-2}} = 3.4488 \quad \text{(SAS: 3.4478)}$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{3.4488}{\sqrt{0.6087}} = 4.4204 \quad \text{(SAS: 4.4244)}$$

The limits of the acceptance region are:

$$B \pm t(s_b) = 50 \pm 1.833(4.4204) = 50 \pm 8.1026 = [41.8974, 58.1026]$$

Here, $b = 55.9634 \Rightarrow$ do not reject $H_0 \Rightarrow$ the standard relationship does seem to hold for this sample.

12-38 a) The 90% confidence interval is $b \pm t(s_b)$ $= .147 \pm 1.746(.032)$
$= .147 \pm .0559 = [.091, .203]$

Since .08 is not contained within this confidence interval, it does appear, at the .10 significance level, that the true slope has changed from the value found in 1969.

b) The 99% confidence interval is $b \pm t(s_b)$ $= .147 \pm 2.921(.032)$
$= .147 \pm .0935 = [.054, .241]$

Since .08 is contained within this confidence interval, it does not appear, at the .01 significance level, that the true slope has changed from the value found in 1969.

12-39 $H_0 : B = 1.5$ $H_1 : B > 1.5$ $\alpha = .05$
The upper limit of the acceptance region is:
$B + t(s_b) = 1.5 + 1.645(0.2) = 1.829$ (using z because $df = 62 > 30$)
Here, $b = 1.8 < 1.829 \Rightarrow$ do not reject $H_0 \Rightarrow$ people are not significantly more talkative.

12-40 $H_0 : B = 0.85$ $H_1 : B \neq 0.85$ $\alpha = .01$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{0.60}{\sqrt{0.25}} = 1.2$$

$B \pm t(s_b) = .85 \pm 2.878(1.2) = .85 \pm 3.45 = [-2.6, 4.3]$
Here, $b = 0.70 \Rightarrow$ do not reject $H_0 \Rightarrow$ the slope has not changed significantly.

12-41 An estimating equation is determined by the relationship between the values over a certain range; therefore, for other values, the relationship between variables may be different, thus leading to incorrect conclusions.

12-42 The coefficient of determination is the fraction of the variation in the dependent variable which is explained by the independent variable. The coefficient of correlation only shows whether the relationship is direct or inverse; it does not tell how much of the variation in Y is explained by X.

12-43 Conditions in which earlier studies were done may change so that the relationship no longer is the same. Other variables may be present to influence the situation and affect the relationship between variables that existed in the past. In addition, assumptions made in earlier studies may be violated in the present due to changes in circumstances.

12-44 Correlation does not imply causality because it is simply a statistical technique applies to a set of numbers thought to show a relationship between variables. The relationship between two variables may be due to an external cause affecting them both. Only by the manipulation of variables and then observation of the results can we infer any kind of causality.

12-45 a)

X	Y	XY	X^2	Y^2
9	36	324	81	1296
7	25	175	49	625
8	33	264	64	1089
4	15	60	16	225
7	28	196	49	784
5	19	95	25	361
5	20	100	25	400
6	22	132	36	484
$\sum X = 51$	$\sum Y = 198$	$\sum XY = 1346$	$\sum X^2 = 345$	$\sum Y^2 = 5264$

$\overline{X} = \frac{51}{8} = 6.375$ $\qquad\qquad$ $\overline{Y} = \frac{198}{8} = 24.75$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{1346 - 8(6.375)(24.75)}{345 - 8(6.375)^2} = 4.2138$$

$a = \overline{Y} - b\overline{X} = 24.75 - 4.2138(6.375) = -2.1130$

Thus, $\widehat{Y} = -2.1130 + 4.2138X$ (SAS: $\hat{Y}$: $-2.1132 + 4.2138X$)

b) $$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{5264 - (-2.1130)(198) - 4.2138(1346)}{8-2}} = 1.3291 \quad \text{(SAS: } 1.3286)$$

c) $$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$= \frac{-2.1130(198) + 4.2138(1346) - 8(24.75)^2}{5264 - 8(24.75)^2} = 0.9708 \quad \text{(SAS: } .9709)$$

12-46 a) Let Y = storks and X = babies

X	Y	XY	X^2	Y^2
27	35	945	729	1225
38	46	1748	1444	2116
13	19	247	169	361
24	32	768	576	1024
6	15	90	36	225
19	31	589	361	961
15	20	300	225	400
$\sum X = 142$	$\sum Y = 198$	$\sum XY = 4687$	$\sum X^2 = 3540$	$\sum Y^2 = 6312$

$\overline{X} = \frac{142}{7} = 20.2857$ $\qquad\qquad$ $\overline{Y} = \frac{198}{7} = 28.2857$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{4687 - 7(20.2857)(28.2857)}{3540 - 7(20.2857)^2} = 1.0167$$

$a = \overline{Y} - b\overline{X} = 28.2857 - 1.0167(20.2857) = 7.6612$ (SAS: 7.6616)

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$= \frac{7.6612(198) + 1.0167(4687) - 7(28.2857)^2}{6312 - 7(28.2857)^2} = 0.9581$$

$r = \sqrt{.9581} = 0.9788$

b) No. The high correlation is spurious. It simply reflects the fact that the number of storks and the number of babies tend to rise together when the population increases. Higher population means more births and more nesting sites for storks.

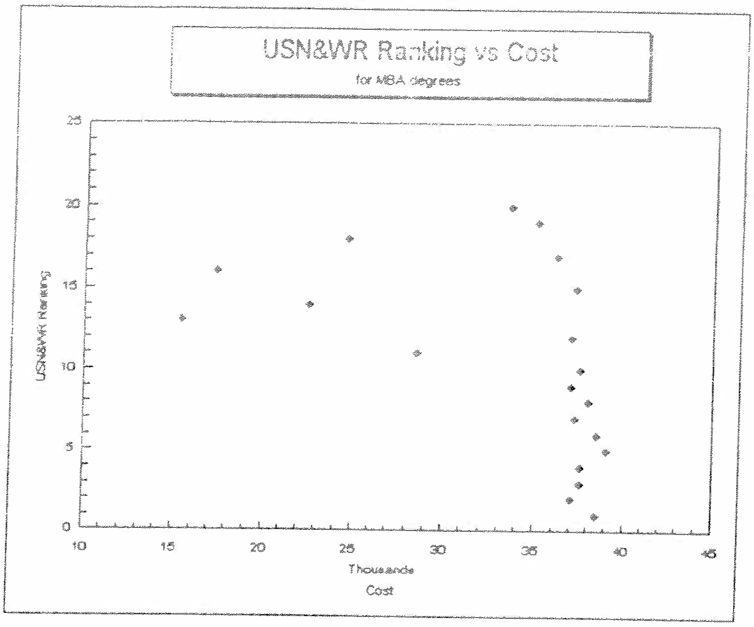
12-47

relationship	estimating	predict
dependent	independent	degree or strength
estimating equation		

12-48 $r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{15.0279(11.6) + (-1.2471)(2.7) - 6(1.9333)^2}{171.88 - 6(1.9333)^2} = 0.9938$

$r = -\sqrt{0.9938} = -0.9969$

12-49



There is a weak curvilinear relationship demonstrating that more expensive schools received higher rankings. The top ten schools all have higher costs. The cheaper schools, however, are not all at the bottom of the curve.

$$b = \frac{\sum XY - n\overline{XY}}{\sum X^2 - n\overline{X}^2} = \frac{6574240 - 20(33257)(10.5)}{23156470000 - 20(33257)^2} = -0.0004$$

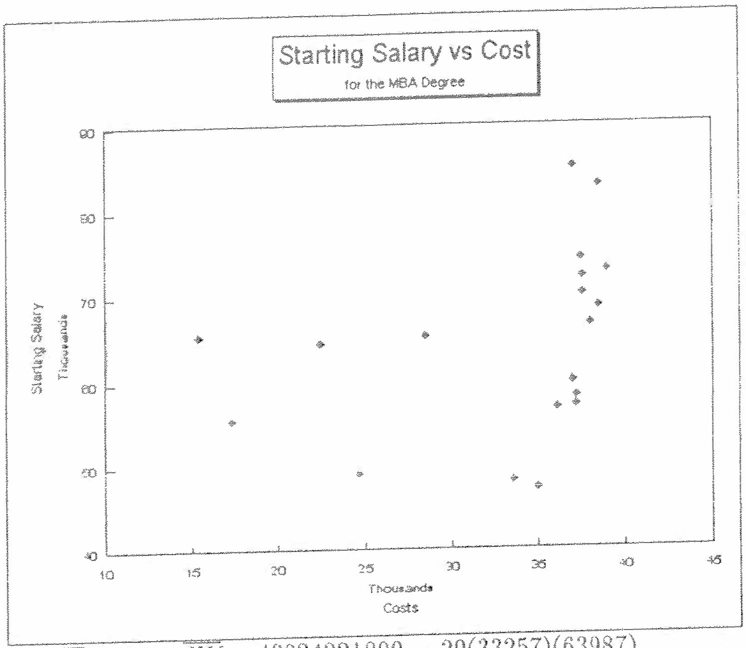
$$a = \overline{Y} - b\overline{X} = 10.5 - (-.0004)(33257) = 23.8028$$

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$r^2 = \frac{23.8028(210) + (-.0004)(6574240) - 20(10.5)^2}{2870 - 20(10.5)^2} = 0.2465$$

$$r = \sqrt{r^2} = \sqrt{0.2465} = 0.4969$$

12-50



$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{43024221800 - 20(33257)(63987)}{23156470000 - 20(33257)^2} = 0.4478$$

$$a = \overline{Y} - b\overline{X} = 63987 - .4478(33257) = 49094.5154$$

$$\widehat{Y} = 49094.5154 + 0.4478X$$

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$s_e = \sqrt{\frac{83999286400 - 49094.5154(1279740) - .4478(43024221800)}{18}} = 10287.0327$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{10287.0327}{\sqrt{23156470000 - 20(33257)^2}} = 0.3196$$

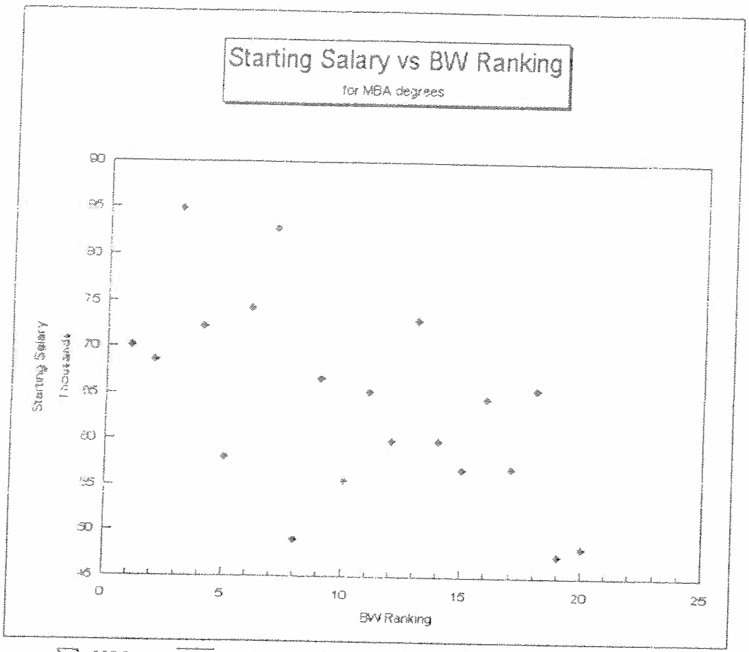
$H_0 : B = 0$ $\qquad$ $H_1 : B > 0$ $\qquad$ $\alpha = .05$

$$t = \frac{b - B_{H_0}}{s_b} = \frac{-0.4478 - 0}{0.3196} = 1.4011$$

$$t_{CRIT} = t_{.05,18} = 2.101$$

Since $t = 1.4011$ (which is less than 2.101), do not reject H_0. Greater school costs are not associated with higher salaries.

12-51



$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{12733360 - 20(10.5)(63987)}{2870 - 20(10.5)^2} = -1058.5113$$

$$a = \overline{Y} - b\overline{X} = 63987 - (-1058.5113)(10.5) = 75101.3687$$

$$\hat{Y} = 75101.3687 - 1058.5113X$$

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$s_e = \sqrt{\frac{83999276400 - 75101.3687(1279740) - (-1058.5113)(12733360)}{18}} = 8716.0652$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{8716.0652}{\sqrt{2870 - 20(10.5)^2}} = 337.9945$$

$H_0 : B = 0$ $\qquad$ $H_1 : B > 0$ $\qquad$ $\alpha = .05$

$$t = \frac{b - B_{H_0}}{s_b} = \frac{-1508.5113 - 0}{337.9945} = -4.4631$$

$$t_{CRIT} = t_{.05,18} = -2.101$$

Since $t = -4.4631$ (which is less than -2.101), reject H_0. We conclude that a higher Business Week ranking is associated with a greater starting salary.

12-52 Business Week overall rankings (From 12-51):
$b = -1508.5113$
$a = 75101.3687$

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$r^2 = \frac{75101.37(1279740) + (-1508.51)(12733360) - 20(63987)^2}{83999276400 - 20(63987)^2} = 0.3527$$

Student rankings:

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{13595200 - 20(11)(63987)}{3286 - 20(11)^2} = -556.513$$

$$a = \overline{Y} - b\overline{X} = 63987 - (-556.513)(11) = 70108.64$$

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$r^2 = \frac{70108.64(1279740) + (-556.513)(13595200) - 20(63987)^2}{83999276400 - 20(63987)^2} = 0.1267$$

Firm rankings:

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{12807650 - 20(10.5)(63987)}{2870 - 20(10.5)^2} = -946.797$$

$$a = \overline{Y} - b\overline{X} = 63987 - (-949.797)(10.5) = 73928.37$$

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$r^2 = \frac{73928.37(1279740) + (-946.797)(12807650) - 20(63987)^2}{83999276400 - 20(63987)^2} = 0.2822$$

The overall ranking explains the largest fraction of the variation in starting salaries.

12-53 In this problem, Y = this year's sales over quota and X = last year's bonus.

a)

X	Y	XY	X^2	Y^2
7.8	64	499.2	60.84	4096
6.9	73	503.7	47.61	5329
6.7	42	281.4	44.89	1764
6.0	49	294.0	36.00	2401
6.9	71	489.9	47.61	5041
5.2	46	239.2	27.04	2116
6.3	32	201.6	39.69	1024
8.4	88	739.2	70.56	7744
7.2	53	381.6	51.84	2809
10.1	84	848.4	102.01	7056
10.8	85	918.0	116.64	7225
7.7	93	716.1	59.29	8649
$\sum X = 90.0$	$\sum Y = 780$	$\sum XY = 6112.3$	$\sum X^2 = 704.02$	$\sum Y^2 = 55254$

$$\overline{X} = \frac{90.0}{12} = 7.5 \qquad \overline{Y} = \frac{780}{12} = 65$$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{6112.3 - 12(7.5)(65)}{704.02 - 12(7.5)^2} = 9.0386$$

$$a = \overline{Y} - b\overline{X} = 65 - 9.0386(7.5) = -2.7895$$

Thus, $\widehat{Y} = -2.7895 + 9.0386X$ (SAS: $\widehat{Y} = -2.7895 + 9.0386X$)

b)

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{55254 - (-2.7895)(780) - (9.0386)(6112.3)}{12-2}} = 14.7756$$

c) $\widehat{Y} = -2.7895 + 9.0386(9.6) = 83.9811$

An approximate 90% confidence interval is:

$$\widehat{Y} \pm t(s_e) = 83.9811 \pm 1.812(14.7756) = 83.9811 \pm 26.7734 = [57.21,\ 110.76]$$

12-54 a) 1, + b) 2, + c) 2, − d) 2, −

12-55 In this problem, Y = cost and X = pounds of material.

X	Y	XY	X^2	Y^2
25	10	250	625	100
20	7	140	400	49
16	5	80	256	25
17	6	102	289	36
19	7	133	361	49
18	6	108	324	36
$\sum X = 115$	$\sum Y = 41$	$\sum XY = 813$	$\sum X^2 = 2255$	$\sum Y^2 = 295$

$$\overline{X} = \frac{115}{6} = 19.1667 \qquad \overline{Y} = \frac{41}{6} = 6.8333$$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{813 - 6(19.1667)(6.8333)}{2255 - 6(19.1667)^2} = 0.5346 \quad \text{(SAS: 0.5344)}$$

$$a = \overline{Y} - b\overline{X} = 6.8333 - .5346(19.1667) = -3.4132 \quad \text{(SAS: } -3.4098\text{)}$$

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}} = \sqrt{\frac{295 - (-3.4132)(41) - 0.5346(813)}{6-2}}$$

$$= 0.2790 \quad \text{(SAS: 0.2805)}$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{0.2790}{\sqrt{50.8257}} = 0.0391 \quad \text{(SAS: 0.0393)}$$

$H_0: B = 0.5$ $\qquad H_1: B > 0.5$ $\qquad \alpha = .05$

The upper limit of the acceptance region is:

$$B + t(s_b) = 0.5 + 2.132(.0391) = 0.5834$$

Here, $b = 0.5346 < 0.5834 \Rightarrow$ do not reject $H_0 \Rightarrow$ slope not significantly greater than 0.5 so no need to adjust assembly line

12-56 $$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{4.9205(809.48) + 3.9650(12688.48) - 15(53.9653)^2}{51156.28 - 15(53.9653)^2} = 0.9613$$

$$r = \sqrt{0.9613} = 0.9805$$

12-57 a

12-58 SALES and POP

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{20994383720 - 50(1764858)(235.324)}{2810359 - 50(235.324)^2} = 5512.5812$$

$$a = \overline{Y} - b\overline{X} = 1764858 - 5512.5812(235.324) = 467615.1589$$

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$r^2 = \frac{467615.1589(88242891) + 5512.5812(2099438720) - 50(1764858)^2}{163945843262647 - 50(1764858)^2} = 0.1536$$

SALES and TM

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{650886990558 - 50(7180.379)(1764858)}{2704284838 - 50(7180.379)^2} = 136.6344$$

$$a = \overline{Y} - b\overline{X} = 1764858 - 136.6344(7180.379) = 783771.369$$

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

$$r^2 = \frac{783771.369(88242891) + 136.6344(650886990558) - 50(1764858)^2}{163945843262647 - 50(1764858)^2} = 0.2874$$

TM explains more of the variation in SALES than does POP.

12-59 SALES and SINGLE

$\overline{X} = 21.354$ $\quad\quad \overline{Y} = 1764857.82$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{1928674572.6 - 50(21.354)(1764857.82)}{23730.97 - 50(21.354)^2} = 47606.23$$

$$a = \overline{Y} - b\overline{X} = 1764857.82 - 47606.23(21.354) = 748274.3846$$

$$\hat{Y} = 748274.3846 + 47606.23X$$

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$s_e = \sqrt{\frac{163945843262647 - 748274.38(88242891) - 47606.23(1928674572.6)}{50-2}} = 356458.95$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{356458.95}{\sqrt{23730.97 - 50(21.354)^2}} = 11680.56$$

$$\hat{Y} = 748274.38 + 47606.23(20) = 1700398.98$$

An approximate 90% confidence interval is:

$$\hat{Y} \pm t(s_e) = 1700398.98 \pm 1.679(356458.95) = 1700398.98 \pm 598494.5771$$
$$\hat{Y} \pm t(s_e) = [1101904.403, 2298893.557]$$

This result would give the company information regarding the approximate amount of sales attributable to single person households. This information can help a "frozen dinner" company decide if the market (singles) was large enough to warrant an introduction of the single-serving product.

12-60 SALES and AGE

$\overline{X} = 32.774$ $\quad\quad \overline{Y} = 1764857.82$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{2900658083 - 50(32.774)(1764857.82)}{54317.81 - 50(32.774)^2} = 14049.8998$$

$$a = \overline{Y} - b\overline{X} = 1764857.82 - 14049.8998(32.774) = 1304386.404$$

$$\hat{Y} = 1304386.404 + 14049.9X$$

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$s_e = \sqrt{\frac{163945843262647 - 1304386.404(88242891) - 14049.8998(2900658083)}{50-2}} = 410514.332$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{410514.322}{\sqrt{54317.81 - 50(32.774)^2}} = 16606.87$$

$H_0 : B = 0 \qquad H_1 : B > 0$

$$t = \frac{b - B_{H_0}}{s_b} = \frac{14049.9 - 0}{16606.87} = 0.846$$

With 48 df, the probability value for the test is greater than 10%, so we would probably accept H_0. Although this appears to indicate that "Business isn't better in communities with lots of older people", it would be erroneous to draw such a conclusion. As we saw in Exercise 12-58, POP explains 15% of the variation in SALES, and a simple regression of SALES on AGE ignores this factor. In order to legitimately draw the suggested conclusion, you should first do a multiple regression analysis.

12-61 $H_0 : B = .94 \qquad H_1 : B < .94 \qquad \alpha = .05$

The lower limit of the acceptance region is:

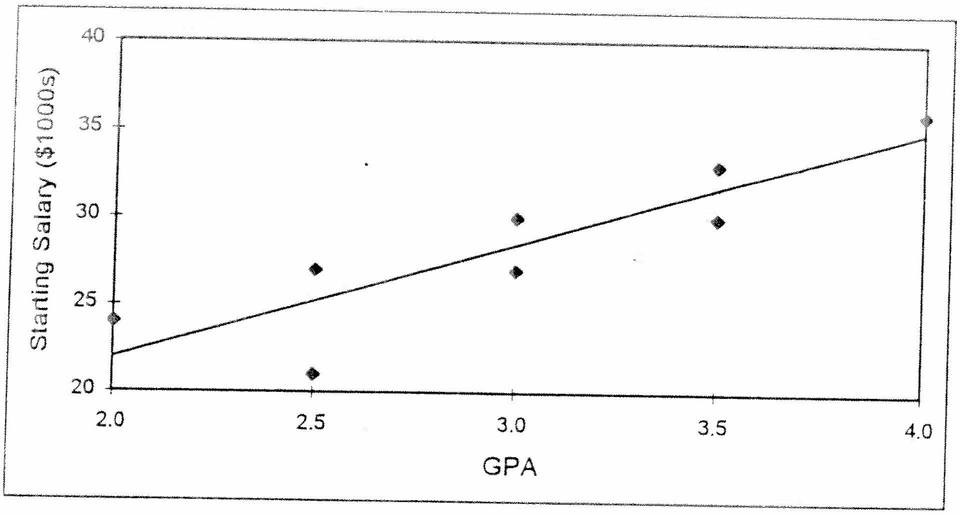
$$B - t(s_b) = .94 - 1.714(.035) = .8800$$

Here, $b = .87 < .8800 \Rightarrow$ reject $H_0 \Rightarrow$ marginal propensity to consume has decreased significantly below the standard .94.

12-62 a

12-63 In this problem, Y = starting salary and X = college GPA.

a)



b)

X	Y	XY	X^2
4.0	36	144.0	16.00
3.0	30	90.0	9.00
3.5	30	105.0	12.25
2.0	24	48.0	4.00
3.0	27	81.0	9.00
3.5	33	115.5	12.25
2.5	21	52.5	6.25
2.5	27	67.5	6.25
$\sum X = 24.0$	$\sum Y = 228$	$\sum XY = 703.5$	$\sum X^2 = 75.00$

$\overline{X} = \frac{24}{8} = 3$ $\qquad\qquad$ $\overline{Y} = \frac{228}{8} = 28.5$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{703.5 - 8(3)(28.5)}{75 - 8(3)^2} = 6.5$$

$a = \overline{Y} - b\overline{X} = 28.5 - 6.5(3) = 9$

Thus, $\widehat{Y} = 9 + 6.5X$

12-64 In this problem, Y = rent and X = number of bedrooms.

a)

X	Y	XY	X^2	Y^2
2	230	460	4	52900
1	190	190	1	36100
3	450	1350	9	202500
2	310	620	4	96100
2	218	436	4	47524
2	185	370	4	34225
2	340	680	4	115600
1	245	245	1	60025
1	125	125	1	15625
2	350	700	4	122500
2	280	560	4	78400
$\sum X = 20$	$\sum Y = 2923$	$\sum XY = 5736$	$\sum X^2 = 40$	$\sum Y^2 = 861499$

$\overline{X} = 20/11 = 1.81818$ $\qquad\qquad$ $\overline{Y} = 2923/11 = 265.72727$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{5736 - 11(1.81818)(265.72727)}{40 - 11(1.81818)^2} = 115.8991$$

$a = \overline{Y} - b\overline{X} = 265.72727 - 115.8991(1.81818) = 55.0018$

Thus, $\widehat{Y} = 55.0018 + 115.8991X$ (SAS: $\widehat{Y} = 55.0 + 115.9X$)

b) $$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{55.0018(2923) + 115.8991(5736) - 11(265.72727)^2}{861499 - 11(265.72727)^2}$$

$$= 0.5762$$

c) $\widehat{Y} = 55.0018 + 115.8991(2) = 286.80$

12-65 In this problem, Y = sales and X = billboard expenditures.

a)

X	Y	XY	X^2	Y^2
25	34	850	625	1156
16	14	224	256	196
42	48	2016	1764	2304
34	32	1088	1156	1024
10	26	260	100	676
21	29	609	441	841
19	20	380	361	400
$\sum X = 167$	$\sum Y = 203$	$\sum XY = 5427$	$\sum X^2 = 4703$	$\sum Y^2 = 6597$

$\overline{X} = 167/7 = 23.8571$ $\qquad\qquad$ $\overline{Y} = 203/7 = 29$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{5427 - 7(23.8571)(29)}{4703 - 7(23.8571)^2} = 0.8124$$

$a = \overline{Y} - b\overline{X} = 29 - 0.8124(23.8571) = 9.6185$

Thus, $\widehat{Y} = 9.6185 + 0.8124X$ (SAS: $\widehat{Y} = 9.6184 + 0.8124X$)

b)

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{6597 - 9.6185(203) - 0.8124(5427)}{7-2}} = 6.8637 \quad \text{(SAS: 6.8638)}$$

c) $\widehat{Y} = 9.6185 + 0.8124(28) = 32.3657$

An approximate 95% confidence interval is:

$$\widehat{Y} \pm t(s_e) = 32.4657 \pm 2.571(6.8637) = 32.4657 \pm 17.6466$$
$$= [14.719,\ 50.012]$$

12-66 $$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{74.3961 - 10(2.571)(2.946)}{68.6733 - 10(2.571)^2} = -0.523$$

$a = \overline{Y} - b\overline{X} = 2.946 - (-0.523)(2.571) = 4.2906$

$\widehat{Y} = 4.2906 - 0.523X$

Answer: (a) Lower price increases sales

12-67 $H_0: B = 4.0$ $\quad H_1: B \neq 4.0$ $\quad \alpha = .05$

The limits of the acceptance region are:

$$B \pm t(s_b) = 4.0 \pm 2.120(.17) = 4.0 \pm .3604 = [3.6396,\ 4.3604]$$

Here, $b = 4.3 \Rightarrow$ accept $H_0 \Rightarrow$ slope has not changed significantly from previous studies.

12-68 In this problem, Y = price and X = size of offering.

a)

X	Y	XY	X^2	Y^2
108.0	12.0	1296.00	11664.00	144.00
4.4	4.0	17.60	19.36	16.00
3.5	5.0	17.50	12.25	25.00
3.6	6.0	21.60	12.96	36.00
39.0	13.0	507.00	1521.00	169.00
68.4	19.0	1299.60	4678.56	361.00
7.5	8.5	63.75	56.25	72.25
5.5	5.0	27.50	30.25	25.00
375.0	15.0	5625.00	140625.00	225.00
12.0	6.0	72.00	144.00	36.00
51.0	12.0	612.00	2601.00	144.00
66.0	12.0	792.00	4356.00	144.00
10.4	6.5	67.60	108.16	42.25
4.0	3.0	12.00	16.00	9.00
$\sum X = 758.3$	$\sum Y = 127$	$\sum XY = 10431.15$	$\sum X^2 = 165844.79$	$\sum Y^2 = 1448.50$

$\overline{X} = 758.3/14 = 54.16429$ $\qquad \overline{Y} = 127/14 = 9.07143$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{10431.15 - 14(54.16429)(9.07143)}{165844.79 - 14(54.16429)^2} = 0.0285$$

$a = \overline{Y} - b\overline{X} = 9.07143 - 0.0285(54.16429) = 7.5277$

Thus, $\widehat{Y} = 7.5277 + 0.0285X$ (SAS: $\widehat{Y} = 7.5294 + 0.0285X$)

b) $r^2 = \dfrac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \dfrac{7.5277(127) + 0.0285(10431.15) - 14(9.07143)^2}{1448.5 - 14(9.07143)^2}$

$= 0.3415$ (SAS: 0.3412)

Since only about a third of the variability in price is explained by the size of the offering, Dave should search for additional explanatory variables.

12-69 In this problem, Y = life and X = daily usage

a) Lithium batteries

X	Y	XY	X^2	Y^2
2.0	3.1	6.20	4.00	9.61
1.5	4.2	6.30	2.25	17.64
1.0	5.1	5.10	1.00	26.01
0.5	6.3	3.15	0.25	36.69
$\sum X = 5.0$	$\sum Y = 18.7$	$\sum XY = 20.75$	$\sum X^2 = 7.50$	$\sum Y^2 = 92.95$

$\overline{X} = \dfrac{5.0}{4} = 1.25$ $\qquad \overline{Y} = \dfrac{18.7}{4} = 4.675$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{20.75 - 4(1.25)(4.675)}{7.50 - 4(1.25)^2} = -2.1$$

$a = \overline{Y} - b\overline{X} = 4.675 - (-2.1)(1.25) = 7.3$

Thus, $\widehat{Y} = 7.3 - 2.1X$

Alkaline batteries

X	Y	XY	X^2	Y^2
2.0	1.3	2.60	4.00	1.69
1.5	1.6	2.40	2.25	2.56
1.0	1.8	1.80	1.00	3.24
0.5	2.2	1.10	0.25	4.84
$\sum X = 5.0$	$\sum Y = 6.9$	$\sum XY = 7.90$	$\sum X^2 = 7.50$	$\sum Y^2 = 12.33$

$\overline{X} = \dfrac{5.0}{4} = 1.25$ $\qquad \overline{Y} = \dfrac{6.9}{4} = 1.725$

$b = \dfrac{7.90 - 4(1.25)(1.725)}{7.50 - 4(1.25)^2} = -0.58$ $\qquad a = 1.725 - (-0.58)(1.25) = 2.45$

Thus, $\widehat{Y} = 2.45 - 0.58X$

b) Lithium batteries

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}} = \sqrt{\frac{92.95 - 7.3(18.7) - (-2.1)(20.75)}{2}} = 0.0866$$

At $X = 1.25$, $\widehat{Y} = 7.3 - 2.1(1.25) = 4.675$, and an approximate 90% confidence interval is:

$$\widehat{Y} \pm t(s_e) = 4.675 \pm 2.920(0.0866) = 4.675 \pm 0.253 = [4.422, 4.928]$$

Alkaline batteries

$$s_e = \sqrt{\frac{12.33 - 2.45(6.9) - (-0.58)(7.90)}{2}} = 0.0592$$

At $X = 1.25$, $\widehat{Y} = 2.45 - 0.58(1.25) = 1.725$, and an approximate 90% confidence interval is:

$$1.725 \pm 2.920(0.0592) = 1.725 \pm 0.173 = [1.552,\ 1.898]$$

Since the entire confidence interval for lithium batteries lies above the upper limit of the interval for alkaline batteries, the company can reasonably claim that lithium batteries have a significantly longer life than alkaline batteries.

12-70 In this problem, Y = adult height (in inches) and X = birth weight (in ounces).

X	Y	XY	X^2	Y^2
88	69	6072	7744	4761
112	72	8064	12544	5184
100	66	6600	10000	4356
120	71	8520	14400	5041
130	73	9490	16900	5329
108	70	7560	11664	4900
$\sum X = 658$	$\sum Y = 421$	$\sum XY = 46306$	$\sum X^2 = 73252$	$\sum Y^2 = 29571$

$\overline{X} = 658/6 = 109.6667$ $\qquad$ $\overline{Y} = 421/6 = 70.1667$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{46306 - 6(109.6667)(70.1667)}{73252 - 6(109.6667)^2} = 0.1249$$

$$a = \overline{Y} - b\overline{X} = 70.1667 - 0.1249(109.6667) = 56.4693$$

Thus, $\widehat{Y} = 56.4693 + 0.1249X$ (SAS: $\widehat{Y} = 56.4667 + 0.1249X$)

$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{56.4693(421) + 0.1249(46306) - 6(70.1667)^2}{29571 - 6(70.1667)^2}$$

$$= 0.5519 \quad \text{(SAS: 0.5524)}$$

Thus, 55% of variation in adult height is explained by birth weight.

12-71 In this problem, Y = junior/senior GPA and X = freshman/sophomore GPA

a)

X	Y	XY	X^2	Y^2
1.7	2.4	4.08	2.89	5.76
3.5	3.7	12.95	12.25	13.69
2.3	2.0	4.60	5.29	4.00
2.6	2.5	6.50	6.76	6.25
3.0	3.2	9.60	9.00	10.24
2.8	3.0	8.40	7.84	9.00
2.4	2.5	6.00	5.76	6.25
1.9	1.8	3.42	3.61	3.24
2.0	2.7	5.40	4.00	7.29
3.1	3.7	11.47	9.61	13.69
$\sum X = 25.3$	$\sum Y = 27.5$	$\sum XY = 72.42$	$\sum X^2 = 67.01$	$\sum Y^2 = 79.41$

$\overline{X} = 25.3/10 = 2.53$ $\qquad$ $\overline{Y} = 27.5/10 = 2.75$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{72.42 - 10(2.53)(2.75)}{67.01 - 10(2.53)^2} = 0.9480$$

$$a = \overline{Y} - b\overline{X} = 2.75 - 0.9480(2.53) = 0.3516$$

Thus, $\widehat{Y} = 0.3516 + 0.9480X$ (SAS: $\widehat{Y} = 0.3515 + 0.9480X$).

b)
$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$
$$= \sqrt{\frac{79.41 - 0.3516(27.5) - 0.9480(72.42)}{8}} = 0.3686 \quad \text{(SAS: 0.3688)}$$

At $X = 2.5$, $\widehat{Y} = 0.3516 + 0.9480(2.5) = 2.7216$, and an approximate 90% confidence interval is:

$$\widehat{Y} \pm ts_e = 2.7216 \pm 1.860(0.3686) = 2.7216 \pm 0.6856 = [2.0360, 3.4072]$$

Since this entire interval is above 2.0, Hoopes should admit this applicant.

12-72 a) Using a calculator, we find $\sum X = 317.271$, $\sum Y = 910.612$, $\sum XY = 29,221.8450$, $\sum X^2 = 10,246.2891$, and $\sum Y^2 = 85,344.9104$. Hence

$\overline{X} = 317.271/10 = 31.7271 \qquad \overline{Y} = 910.612/10 = 91.0612$

$$b = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \frac{29,221.8450 - 10(31.7271)(91.0612)}{10,246.2891 - 10(31.7271)^2} = 1.8356$$

$a = \overline{Y} - b\overline{X} = 91.0612 - 1.8356(31.7172) = 32.8229$

Thus, $\widehat{Y} = 32.8229 + 1.8356X$ (SAS: $\widehat{Y} = 32.8245 + 1.8356X$)

$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$
$$= \sqrt{\frac{85,344.9104 - 32.8229(910.612) - 1.8356(29,221.8450)}{10-2}} = 15.0680$$

$$s_b = \frac{s_e}{\sqrt{\sum X^2 - n\overline{X}^2}} = \frac{15.0680}{\sqrt{10,246.2891 - 10(31.7271)^2}} = 1.1225$$

$H_0: B = 0 \qquad H_1: B \neq 0 \qquad \alpha = 0.05$

The limits of the acceptance region are $B \pm ts_b = 0 \pm 2.306(1.1225) = \pm 2.5885$. Here, $b = 1.8356 < 2.5885$, so do not reject H_0. The Attorney General's salary is not related to the going rate for attorneys in the state.

b)
$$r^2 = \frac{a\sum Y + b\sum XY - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2} = \frac{32.8229(910.612) + 1.8356(29,221.8450) - 10(91.0612)^2}{85,344.9104 - 10(91.0612)^2}$$
$$= 0.2505$$

Thus, 25% of the variation in AG's salaries is accounted for by the going rate for attorneys in the for-profit market.

c) No. First of all, as we've just seen, the relationship between the two is not particularly strong. But even if r^2 had been higher, remember that r^2 measures association, not causality.

12-73 Using a computer package, we find that $r^2 = 0.0863$. Since these two variables are basically not related to each other, the controller should use both of them in setting *per diems* for travel costs.

CHAPTER 13

MULTIPLE REGRESSION AND MODELING TECHNIQUES

13-1 To increase the accuracy of our estimate of the dependent variable by including more than one explanatory variable.

13-2 To enable us to include qualitative factors as explanatory variables in regression models.

13-3 It refers to the fact that there are several explanatory variables.

13-4 Yes. Qualitative factors such as season of the year can be modeled using the techniques of "dummy variables."

13-5 1) Determine the multiple regression equation

2) Examine its standard error of estimate

3) Use multiple correlation analysis to see how well the model describes the data

13-6 No. Conceptually they are quite similar since multiple regression is based on the same assumptions as simple regression. However, they will be more complex in terms of the computations that need to be done.

13-7 a)

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
11.4	4.5	13.2	51.30	150.48	59.40	20.25	174.24	129.96
16.6	8.7	18.7	144.42	310.42	162.69	75.69	349.69	275.56
20.5	12.6	19.8	258.30	405.90	249.48	158.76	392.04	420.25
29.4	19.7	25.4	579.18	746.76	500.38	388.09	645.16	864.36
7.6	2.9	22.8	22.04	173.28	66.12	8.41	519.84	57.76
13.8	6.7	17.8	92.46	245.64	119.26	44.89	316.84	190.44
28.5	17.4	14.6	495.90	416.10	254.04	302.76	213.16	812.25
127.8	72.5	132.3	1643.60	2448.58	1411.37	998.85	2610.97	2750.58

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 127.8 = 7a + 72.5b_1 + 132.3b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \qquad 1643.60 = 72.5a + 998.85b_1 + 1411.37b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \qquad 2448.58 = 132.3a + 1411.37b_1 + 2610.97b_2$$

Solving these equations simultaneously, we get

$$a = 8.1905 \qquad b_1 = 1.3222 \qquad b_2 = -0.1919$$

So the regression equation is $\widehat{Y} = 8.1905 + 1.3222X_1 - 0.1919X_2$

b) With $X_1 = 10.5$ and $X_2 = 13.6$, $\widehat{Y} = 8.1905 + 1.3222(10.5) - 0.1919(13.6) = 19.46$

13-8 a)

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
10	8	4	80	40	32	64	16	100
17	21	9	357	153	189	441	81	289
18	14	11	252	198	154	196	121	324
26	17	20	442	520	340	289	400	676
35	36	13	1260	455	468	1296	169	1225
8	9	28	72	224	252	81	784	64
114	105	85	2463	1590	1435	2367	1571	2678

Equations 11-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 114 = 6a + 105b_1 + 85b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \qquad 2463 = 105a + 2367b_1 + 1435b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \qquad 1590 = 85a + 1435b_1 + 1571b_2$$

Solving these equations simultaneously, we get

$$a = 2.5915 \qquad b_1 = 0.8897 \qquad b_2 = 0.0592$$

So the regression equation is

$$\hat{Y} = 2.5915 + 0.8897X_1 + 0.0592X_2$$

b) With $X_1 = 28$ and $X_2 = 10$,

$$\hat{Y} = 2.5915 + 0.8897(28) + 0.0592(10) = 28.10$$

13-9 a)

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
6	1	3	6	18	3	1	9	36
10	3	−1	30	−10	−3	9	1	100
9	2	4	18	36	8	4	16	81
14	−2	7	−28	98	−14	4	49	196
7	3	2	21	14	6	9	4	49
5	6	−4	30	−20	−24	36	16	25
51	13	11	77	136	−24	63	95	487

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 51 = 6a + 13b_1 + 11b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \qquad 77 = 13a + 63b_1 - 24b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \qquad 136 = 11a - 24b_1 + 95b_2$$

Solving these equations simultaneously, we get

$$a = 12.4247 \qquad b_1 = -1.4874 \qquad b_2 = -0.3828$$

So the regression equation is

$$\hat{Y} = 12.4247 - 1.4874X_1 - 0.3828X_2$$

b) When $X_1 = -1$ and $X_2 = 4$,

$$\hat{Y} = 12.4247 - 1.4874(-1) - 0.3828(4) = 12.38$$

13-10 a) In this problem, Y = sales, X_1 = advertising, X_2 = price.

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
33	3	125	99	4125	375	9	15625	1089
61	6	115	366	7015	690	36	13225	3721
70	10	140	700	9800	1400	100	19600	4900
82	13	130	1066	10660	1690	169	16900	6724
17	9	145	153	2465	1305	81	21025	289
24	6	140	144	3360	840	36	19600	576
287	47	795	2528	37425	6300	431	105975	17299

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 287 = 6a + 47b_1 + 795b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \qquad 2528 = 47a + 431b_1 + 6300b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \qquad 37425 = 795a + 6300b_1 + 105975b_2$$

Solving these equations simultaneously, we get

$a = 219.2306 \qquad b_1 = 6.3815 \qquad b_2 = -1.6708$

So the regression equation is

$$\hat{Y} = 219.2306 + 6.3815X_1 - 1.6708X_2$$

b) When advertising = 7 and price = 132,

$$\hat{Y} = 219.2306 + 6.3815(7) - 1.6708(132) = 43.33 \text{ units}$$

13-11 a) In this problem, Y = weight gain, X_1 = initial weight, X_2 = initial age

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
7	39	8	273	56	312	1521	64	49
6	52	6	312	36	312	2704	36	36
8	49	7	392	56	343	2401	49	64
10	46	12	460	120	552	2116	144	100
9	61	9	549	81	549	3721	81	81
5	35	6	175	30	210	1225	36	25
3	25	7	75	21	175	625	49	9
4	55	4	220	16	220	3025	16	16
52	362	59	2456	416	2673	17338	475	380

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 52 = 8a + 362b_1 + 59b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \qquad 2456 = 362a + 17338b_1 + 2673b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \qquad 416 = 59a + 2673b_1 + 475b_2$$

Solving these equations simultaneously, we get

$a = -4.1917 \qquad b_1 = 0.1048 \qquad b_2 = 0.8065$

So the regression equation is

$$\hat{Y} = -4.1917 + 0.1048X_1 + 0.8065X_2$$

b) When initial weight = 48 pounds and initial age = 9 weeks,

$$\hat{Y} = -4.1917 + 0.1048(48) + 0.8065(9) = 8.10 \text{ pounds}$$

13-12 In this problem, Y = price, X_1 = year, X_2 = miles.

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
2.99	1987	55.6	5941.13	166.244	110477.2	3948169	3091.36	8.9401
6.02	1992	18.4	11991.84	110.768	36652.8	3968064	338.56	36.2404
8.87	1993	21.3	17677.91	188.931	42450.9	3972049	453.69	78.6769
3.92	1988	46.9	7792.96	183.848	93237.2	3952144	2199.61	15.3664
9.55	1994	11.8	19042.70	112.690	23529.2	3976036	139.24	91.2025
9.05	1991	36.4	18018.55	329.420	72472.4	3964081	1324.96	81.9025
9.37	1992	28.2	18665.04	264.234	56174.4	3968064	795.24	87.7969
4.20	1988	44.2	8349.60	185.640	87869.6	3952144	1953.64	17.6400
4.80	1989	34.9	9547.20	167.520	69416.1	3956121	1218.01	23.0400
5.74	1991	26.4	11428.34	151.536	52562.4	3964081	696.96	32.9476
64.51	19905	324.1	128455.30	1860.831	644842.2	39620953	12211.27	473.7533

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 64.51 = 10a + 19905b_1 + 324.1b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \quad 128455.3 = 19905a + 39620953b_1 + 644842.2b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \quad 1860.831 = 324.1a + 644842.2b_1 + 12211.27b_2$$

Solving these equations simultaneously, we get

$a = -4243.1682 \qquad b_1 = 2.1315 \qquad b_2 = 0.2135$

So the regression equation is

$$\hat{Y} = -4243.1682 + 2.1315X_1 + 0.2135X_2$$

b) When year = 1991 and miles = 40.0,

$$\hat{Y} = -4243.1682 + 2.1315(1991) + 0.2135(40) = \$9.188 \text{ (in thousands)}$$

13-13 a) In this problem, Y = percent change in GNP, X_1 = federal deficit, X_2 = mean DJIA

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
2.5	50	950	125	2375	47500	2500	902500	6.25
−1.0	200	700	−200	−700	140000	40000	490000	1.00
4.0	60	1100	240	4400	66000	3600	1210000	16.00
1.0	100	800	100	800	80000	10000	640000	1.00
1.5	90	850	135	1275	76500	8100	722500	2.25
3.0	40	900	120	2700	36000	1600	810000	9.00
11.0	540	5300	520	10850	446000	65800	4775000	35.50

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 11 = 6a + 540b_1 + 5300b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \qquad 520 = 540a + 65800b_1 + 446000b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \qquad 10850 = 5300a + 446400b_1 + 4775000b_2$$

Solving these equations simultaneously, we get

$a = -3.6963 \qquad b_1 = -0.0136 \qquad b_2 = 0.0076$

So the regression equation is

$$\hat{Y} = -3.6963 - 0.0136X_1 + 0.0076X_2$$

b) When the federal deficit = 120 and the mean DJIA = 1000,

$$\hat{Y} = -3.6963 - 0.0136(120) + 0.0076(1000) = 2.272\%$$

13-14 Results taken from computer output:

a) $\widehat{Y} = 34.8079 + 5.2618X_1 - 8.0187X_2 + 6.8084X_3$

b) $s_e = 4.0688$

c) $R^2 = .9834$

d) $\widehat{Y} = 34.8079 + 5.2618(5.8) - 8.0187(4.2) + 6.8084(5.1) = 66.37$

13-15 Results taken from computer output:

a) $\widehat{Y} = 16.4497 - 0.4314X_1 + 0.4000X_2 + 0.3425X_3 + 8.9551X_4$

b) $s_e = 29.5417$

c) $R^2 = 0.8064$

d) $\widehat{Y} = 16.4497 - 0.4314(52.4) + 0.4000(41.6) + 0.3425(35.8) + 8.9551(3) = 49.61$
approximate 95% C.I.:

$$\widehat{Y} \pm t(s_e) = 49.61 \pm 12.706(29.5417) = 49.61 \pm 375.36 = [-325.75,\ 424.97]$$

13-16 Results taken from computer output:

a) $\widehat{Y} = 142.4363 + 3.2741X_1 + 0.5269X_2 - 0.3203X_3$

b) $R^2 = 0.9854 \Rightarrow$ 98.54% of total variation in Y explained by model

c) $\widehat{Y} = 142.4363 + 3.2741(82) + 0.5269(75.0) - 0.3203(10.5) = 453.8$

13-17 Results taken from computer output:

a) Predicted DEMAND = 79.1063 − 4.9281PRICE + 0.0159INCOME + 0.1748SUB

b) $b_{PRICE} = -4.9281 < 0$ is logical: As widgets become more expensive, fewer people buy them.

$b_{INCOME} = 0.0159 > 0$ is logical: As income goes up, people have more money to spend, and can buy more widgets.

$b_{SUB} = 0.1748 > 0$ is logical: As the price of the substitute goes up, people turn to widgets, which are relatively cheaper.

c) $R^2 = 0.9510 \Rightarrow$ 95.10% of the variation in DEMAND is explained by this regression.

d) $s_e = 4.5276$ = the standard deviation of the residuals. This measures the dispersion of the points about the multiple regression plane.

e) Predicted DEMAND = 79.1063 − 4.9281(6) + 0.0159(1200) + 0.1748(17) = 71.59

13-18 Results taken from computer output:

a) Predicted GRADE = −49.95 + 1.07HOURS + 1.36IQ + 2.04BOOKS − 1.80AGE

b) $R^2 = (.857)^2 = .7672$ (i.e., 76.72%)

c) Predicted GRADE = −49.95 + 1.07(5) + 1.36(113) + 2.04(3) − 1.80(21) = 77.40
The grade should be about 77.

13-19 a) Predicted SALES = 175.371 − 0.028AUTOS + 3.775ENTRY + 1.990ANNINC + 212.407DISTANCE

b) s_e = ROOT MSE = 85.5865

c) $R^2 = 0.9579 \Rightarrow$ 95.79% of the variation in SALES is explained by this regression

d) When AUTOS = 100, ENTRY = 50, ANNINC = 20, and DISTANCE = 2,
Predicted SALES = 175.371 − 0.028(100) + 3.775(50) + 1.990(20) + 212.407(2) = 825.9, i.e., $825,900

13-20 Results taken from computer output:

a) Predicted SALESPRICE = −1.381 + 2.852SQFT − 3.713STORIES + 30.285BATHROOMS + 1.172AGE

b) $R^2 = 0.952 \Rightarrow$ 95.2% of the variation in SALESPRICE is explained by this regression.

c) When SQFT = 18, STORIES = 1, BATHROOMS = 1.5, and AGE = 6,
Predicted SALESPRICE = −1.381 + 2.852(18) − 3.713(1) + 30.285(1.5) + 1.172(6) = 98.7, i.e., $98,700

13-21 Results taken from computer output:

a) $\widehat{Y} = -1.049 - 0.028X_1 - 0.051X_2 + 0.890X_3$

b) $R^2 = 0.867 \Rightarrow$ 86.7% of the variation in Y is explained by this regression

c) When $X_1 = 7.1$, $X_2 = 3.50$, $X_3 = 6.0$,
$\widehat{Y} = -1.049 - 0.028(7.1) - 0.051(3.5) + 0.890(6) = 3.91$ million tons

13-22 a) $H_0 : B_{DISPLAY} = 3 \qquad H_1 : B_{DISPLAY} < 3 \qquad \alpha = .05$

The lower limit of the acceptance region is:

$$B_{DISPLAY} - t(S_{b_{DISPLAY}}) = 3 - 1.714(0.844) = 1.485$$

Here, $b_{DISPLAY} = 1.25 < 1.485 \Rightarrow$ reject $H_0 \Rightarrow$ each display ad uses up significantly less than 3 pounds of newsprint (holding all else constant). Mark is wrong.

b) $H_0 : B_{CLASSIFIED} = 0.5 \qquad H_0 : B_{CLASSIFIED} \neq 0.5 \qquad \alpha = .05$

The limits of the acceptance region are:

$$B_{CLASSIFIED} \pm t(s_{b_{CLASSIFIED}}) = 0.5 \pm 2.069(0.126) = 0.5 \pm 0.26 = [0.24, 0.76]$$

Here, $b_{CLASSIFIED}$ =0.251 is within the acceptance region $\Rightarrow$ do not reject H_0
$\Rightarrow$ Mark is correct, each classified ad does use about 0.5 pounds of newsprint.

c) $H_0 : B_{FULLPAGE} = 333.333 \qquad H_1 : B_{FULLPAGE} > 333.333$

Here, $b_{FULLPAGE} = 250.659$ is clearly <u>not</u> significantly above 333.333

$\Rightarrow$ do not reject H_0 (without doing any formal test) $\Rightarrow$ Mark's rates are okay.

13-23 $H_0 : B_i = 0 \qquad H_1 : B_i \neq 0 \qquad \alpha = .10$

Looking at the prob-values:

Variable	Prob-Value	Significant?
HOURS	0.3121	No
IQ	0.0084	Yes
BOOKS	0.2182	No
AGE	0.0319	Yes

Thus, IQ and AGE are the only significant explanatory variables.

13-24 a) $F_{OBS} = \dfrac{SSR/k}{SSE/(n-k-1)} = \dfrac{783.604}{135.893} = 5.77$

b) $F_{CRIT} = F_{.05}(4,7) = 4.12$

c) $F_{OBS} > F_{CRIT}$ $\Rightarrow$ reject H_0 that all B_i's = 0
$\Rightarrow$ Y depends on at least one of the X_i's
$\Rightarrow$ regression is significant as a whole.

13-25 $H_0: B_{DISTANCE} = 0$ $H_1: B_{DISTANCE} \neq 0$ $\alpha = .01$

prob-value = .0001 < α = .01 $\Rightarrow$ reject H_0 $\Rightarrow$ DISTANCE is a significant explanatory variable.

13-26 $H_0: B_1 = B_2 = B_3 = B_4 = 0$ H_1 : at least one $B_i \neq 0$ $\alpha = .05$

prob-value = .0001 < α = .05 $\Rightarrow$ reject H_0 $\Rightarrow$ SALES depends on at least one of the independent variables
$\Rightarrow$ regression is significant as a whole

13-27 Temperature and rainfall must be highly correlated variables. Adding both to the equation as independent variables has led to a situation where multicollinearity is present.

13-28 Juan would almost certainly run into the problem of multicollinearity, since the prime interest rate at banks is dependent upon the Federal Reserve's discount rate, which, for the most part, moves directly with the inflation rate. Thus, the estimates of the parameters in Juan's model would be very unreliable, and the model with both predictor variables included would probably not explain much more of the total variation in consumer demand than either of the simple regression models obtained using just one predictor variable.

13-29 Since $R^2 = 1 - \dfrac{\text{SSE}}{\text{SST}}$, $\text{SST} = \dfrac{\text{SSE}}{1-R^2} = \dfrac{125.4}{1-0.7452} = 492.15$

Since SST = SSR + SSE, SSR = SST − SSE = 492.15 − 125.4 = 366.75

$$\text{Thus } F = \frac{\text{SSR}/k}{\text{SSE}/(n-k-1)} = \frac{366.75/3}{125.4/(18-3-1)} = 13.65$$

$F_{CRIT} = F_{.01}(3,14) = 5.56$

Since $F_{OBS} > F_{CRIT}$, we conclude that the overall regression is significant as a whole.

13-30 Results taken from computer output:

a) Predicted TOURISTS = 5.9188 + 355470RATE − 0.1709PRICE + 0.2426PROMOT + 0.2273TEMP

b) $H_0: B_{RATE} = 0$ $H_1: B_{RATE} \neq 0$ $\alpha = .10$

prob-value = .2809 > α = .10 $\Rightarrow$ do not reject H_0 $\Rightarrow$ the currency exchange rate is not a significant explanatory variable.

c) H_0: $B_{PROMOT} = 0.2$ H_1: $B_{PROMOT} > 0.2$ $\alpha = .05$

The upper limit of the acceptance region is:

$$B_{PROMOT} \pm t(s_{b_{PROMOT}}) = 0.2 + 1.895(0.1628) = 0.5085$$

Here, b_{PROMOT} = 0.2426 is within the acceptance region $\Rightarrow$ do not reject H_0
$\Rightarrow$ change in tourists for 1000 pound increase in promotions is not significantly more than 200.

d) A 95% confidence interval is:

$$b_{TEMP} \pm t(s_{b_{TEMP}}) = 0.2273 \pm 2.365(0.1189)$$
$$= 0.2273 \pm 0.2812 = [-0.0539, 0.5085]$$

13-31 Answers will vary. Possible responses:

a) political party affiliation (Democrat/Republican)

b) athletic events (home or away games)

c) marital status (married/single)

13-32 a) Predicted REVENUE $= a + b_1$FLOW $+ b_2$FLOW2

b) Let CITY = 0 if restaurant is in one city and 1 for restaurant in the other city.
Predicted REVENUE $= a + b_1$FLOW $+ b_2$FLOW^2 $+ b_3$CITY

13-33 a) If a second degree equation is needed, a straight line through the data points will show a group of points at one end lying predominantly above (below) the line, followed by a group of predominantly below (above) the line, and another group predominantly above (below). In other words, deviations from the line will NOT be random.

b) The residuals will exhibit the same non-random patterns as in (a). They will be predominantly positive (negative), then negative (positive), then positive (negative).

13-34 a) $H_0 : B_1 = 0$ $\quad H_1 : B_1 \neq 0$ $\quad \alpha = .05$

The limits of the acceptance region are:

$$B_1 \pm t(s_{b_1}) = 0 \pm 2.110(3.245) = 0 \pm 6.85$$

Here, $b_1 = 2.79$ is in the acceptance region $\Rightarrow$ do not reject H_0
$\Rightarrow$ X_1 is not a significant explanatory variable.

b) $H_0 : B_2 = 0$ $\quad H_1 : B_2 \neq 0$ $\quad \alpha = .05$

The limits of the acceptance region are:

$$B_2 \pm t(s_{b_2}) = 0 \pm 2.110(1.53) = 0 \pm 3.23$$

Here, $b_1 = -3.92 < -3.23$ $\Rightarrow$ reject H_0 $\Rightarrow$ X_2 $(= X_1^2)$ is a significant explanatory variable.

13-35 a) Let $X_2^* = 1$ for positive reaction and 0 for negative reaction.

Thus, $X_2 = 1 - X_2^*$. Substituting into the equation:

$$\hat{Y} = 6.7 + 3.5X_1 + 0.489X_2$$
$$\hat{Y} = 6.7 + 3.5X_1 + 0.489(1-X_2^*) = 6.7 + 3.5X_1 + 0.489 - 0.489X_2^* = 7.189 + 3.5X_1 - 0.489X_2^*$$

b) $H_0 : B_2 = 0$ $\quad H_1 : B_2 \neq 0$ $\quad \alpha = .05$

The limits of the acceptance region are:

$$B_2 \pm t(s_{b_2}) = 0 \pm 2.052(.09) = 0 \pm .18$$

Here, $b_2 = .489 > .18$ $\Rightarrow$ reject H_0 $\Rightarrow$ reaction to penicillin is a significant explanatory variable for systolic blood pressure.

13-36 Results taken from computer output:

a) Predicted DEMAND = $-0.9705 + 4.4146$TIME

b) Predicted DEMAND = $3.4101 + 2.8686$TIME $+ 0.0966$TIME^2
This model is better than the linear model: R^2 has increased from 0.9886 to 0.9956, and both explanatory variables are highly significant (with prob-values of 0.0000 and 0.0009). Furthermore, the residuals in part (a) show a curvilinear pattern, but the residuals for this model are random.

13-37 Results taken from computer output:

a, b) Predicted SALES = 8.2470 + 2.1844PROMO + 1.1247TYPE1 − 3.1110TYPE2

c) H_0: $B_{TYPE1} = 0$ H_1: $B_{TYPE1} \neq 0$ $\alpha = 0.05$

prob-value = $0.4247 > \alpha = 0.05$, so we do not reject H_0. There is no significant difference between the effects of radio and newspaper promotions.

d) H_0: $B_{TYPE2} = 0$ H_1: $B_{TYPE2} \neq 0$ $\alpha = 0.05$

prob-value = $0.0464 < \alpha = 0.05$, so we reject H_0. There is a significant difference between the effects of flyers and newspaper promotions.

e) Predicted SALES = 8.2470 + 2.1844(8.0) + 1.1247(1) − 3.1110(0) = 28.8469
An approximate 90% confidence interval is:

$$\begin{aligned}\hat{Y} \pm ts_e &= 28.8469 \pm 1.860(1.8547)\\ &= 28.8469 \pm 3.4497 = [25.3972,\ 32.3966] \text{ hundreds of dollars}\end{aligned}$$

13-38 a) The judge has spotted the obvious pattern in the residuals.

b) The most direct model is a second-degree equation. Add the square of the number of days in court as an additonal independent variable.

13-39 a)

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
28	74	5	2072	140	370	5476	25	784
33	87	11	2871	363	957	7569	121	1089
21	69	4	1449	84	276	4761	16	441
40	93	9	3720	360	837	8649	81	1600
38	81	7	3078	266	567	6561	49	1444
46	97	10	4462	460	970	9409	100	2116
206	501	46	17652	1673	3977	42425	392	7474

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2 \qquad 206 = 6a + 501b_1 + 46b_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 \qquad 17652 = 501a + 42425b_1 + 3977b_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 \qquad 1673 = 46a + 3977b_1 + 392b_2$$

Solving these equations simultaneously, we get

$$a = -43.6754 \qquad b_1 = 1.0484 \qquad b_2 = -1.2437$$

So the regression equation is

$$\hat{Y} = -43.6754 + 1.0484X_1 - 1.2437X_2$$

b) $\hat{Y} = -43.6754 + 1.0484(83) - 1.2437(7) = 34.64$

13-40 a) Predicted GROWTH = 70.066 + 0.422CREAT + 0.271MECH + 0.745ABST + 0.420MATH

b) $R^2 = 0.9261 \Rightarrow$ 92.61% of the variation in GROWTH is explained by this regression.

c)

Variable	Prob-value	Significant?
CREAT	0.0235	Yes
MECH	0.2284	No
ABST	0.0182	Yes
MATH	0.0001	Yes

Thus, the scores on the creativity, abstract thinking, and mathematical calculation aptitude tests are significant explanatory variables for sales growth.

d) PROB $> F = 0.0001$ very small $\Rightarrow$ yes, overall model is significant as a whole

e) Predicted GROWTH $= 70.066 + 0.422(12) + 0.271(14) + 0.745(18) + 0.420(30) = 104.93$

13-41 a) In this problem, $Y =$ checks, $X_1 =$ age, and $X_2 =$ income.

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
29	37	16.2	1073	469.8	599.4	1369	262.44	841
12	34	25.4	1428	1066.8	863.6	1156	645.16	1764
9	48	12.4	432	111.6	595.2	2304	153.76	81
56	38	25.0	2128	1400.0	950.0	1444	625.00	3136
2	43	8.0	86	16.0	344.0	1849	64.00	4
10	25	18.3	250	183.0	457.5	625	334.89	100
48	33	24.2	1584	1161.6	798.6	1089	585.64	2304
4	45	7.9	180	31.6	355.5	2025	62.41	16
200	303	137.4	7161	4440.4	4963.8	11861	2733.30	8246

Equations 13-2, 3, and 4 become

$$\sum Y = na + b_1 \sum X_1 + b_2 \sum X_2 \qquad 200 = 8a + 303b_1 + 137.4b_2$$

$$\sum X_1 Y = a\sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 \qquad 7161 = 303a + 11861b_1 + 4963.8b_2$$

$$\sum X_2 Y = a\sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 \qquad 4440.4 = 137.4a + 4963.8b_1 + 2733.3b_2$$

Solving these equations simultaneously, we get

$a = -70.665$ $\qquad b_1 = 1.010$ $\qquad b_2 = 3.342$

So the regression equation is

$$\hat{Y} = -70.665 + 1.010X_1 + 3.342X_2$$

b) $\hat{Y} = -70.665 + 1.010(35) + 3.342(22.5) = 40$ checks per month

13-42 Results taken from computer output.

Predicted EATING $= 56177.927 + 506.352$POP
$R^2 = 0.0775$

Predicted EATING $= 22170.308 + 5.029$EBI
$R^2 = 0.258$

EBI accounts for more of the variation in EATING than POP. It explains 25.8% of the variation compared to 7.8% for POP.

13-43 Results taken from computer output.

Predicted EATING $= -71826.1 + 426.501$POP $+ 4.820$EBI
$R^2 = 0.312 = 31.2\%$ of the variation in EATING is explained by this model.

The model is significant at the $\alpha = .05$ level.

13-44 Results taken from computer output.

Predicted EATING $= -104304.6 + 142.356$POP $- 4759.177$SINGLE $+ 4.745$EBI
$R^2 = 0.4419 = 44.19\%$ of the variation in EATING is explained by this model.

$b_{SINGLE} = 4759.177$

b_{SINGLE} is a significant explanatory variable ($t = 3.26$, p $= .0021$)

13-45 Results taken from computer output.

Predicted EATING $= -80458.65 + 5135.893$SINGLE $+ 4.798$EBI

$\widehat{Y} = -80458.65 + 5135.893(20) + 4.798(30000) = 166199.21$ total sales of food & drink

An approximate 90% confidence interval is:

$$\widehat{Y} \pm t(s_e) = 166199.21 \pm 1.679(40550.24) = [98115.357, 234283.063]$$

13-46 a) Predicted ANESTHES $= 90.032 + 99.486$TYPE $+ 21.536$WEIGHT $- 34.461$HOURS

b) $\widehat{Y} = 90.032 + 99.486(1) + 21.536(25) - 34.461(1.5) = 676.23$ milliliters

An approximate 95% confidence interval is:

$$\widehat{Y} \pm t(s_e) = 676.23 \pm 2.262(57.070) = 676.23 \pm 129.09 = [547, 805]$$

c) $H_0 : B_{TYPE} = 0$ $\quad$ $H_1 : B_{TYPE} \neq 0$ $\quad$ $\alpha = .10$

prob-value $= .0435 < \alpha = .10$ $\Rightarrow$ reject H_0

$\Rightarrow$ there is a significant difference in the amount of anesthesia needed for dogs and cats (holding all else constant).

d) PROB $> F = 0.0001$ very small $\Rightarrow$ yes, the regression is significant as a whole.

13-47 Results taken from computer output.

a) Predicted PRICE $= 25.3999 - 0.1040$SIZE $+ 0.0014$AREA $+ 7.6038$VIEW

b) $R^2 = 0.9981 = 99.81\%$ of the variation in PRICE is explained by this regression.

c) The 90% confidence interval is:
$b_{VIEW} \pm t(s_{b_{VIEW}}) = 7.6038 \pm 2.015(0.5029) = 7.6038 \pm 1.0133 = [6.59, 8.62]$

d) Yes. It is highly significant (prob-value = .0045). Notice that SIZE (unsquared) is not a significant variable (prob-value = .2677).

13-48 Results taken from computer output.

a) Predicted PRICE $= 444.7183 - 0.6124$WEIGHT $- 4.3769$SQFT

b) At WEIGHT = 100 ounces and SQFT = 46, the regression suggests a price of
$444.7183 - 0.6124(100) - 4.3769(46) = \$182.14$

13-49 Results taken from computer output.

a) For men, predicted TIME $= -6.3231 - 21.7451$SWIM $+ 6.3472$BIKE $- 4.7468$RUN

For women, predicted TIME $= -7.3332 - 28.2792$SWIM $+ 7.9124$BIKE $- 7.9970$RUN

b) In a triathlon with a 1-mile swim, a 50-mile bike ride, and a 12.5-mile run, the predicted winning times are:

For men: $-6.3231 - 21.7451(1) + 6.3472(50) - 4.7468(12.5)$
$= 229.96$ minutes (i.e., 3:49:58)

For women: $-7.3332 - 28.2792(1) + 7.9124(50) - 7.9970(12.5)$
$= 260.05$ minutes (i.e., 4:20:03)

c) For men, $s_e = 25.5306$, so the lower limit of an approximate 90% confidence interval for predicted time is

$$\hat{Y} - ts_e = 229.96 - 1.943(25.5306) = 180.35 \text{ minutes} \quad \text{(i.e., 3:00:21)}$$

For women, $s_e = 37.6207$, so now the lower limit is

$$\hat{Y} - ts_e = 260.05 - 1.943(37.6207) = 186.95 \text{ minutes} \quad \text{(i.e., 3:06:57)}$$

13-50 Results taken from computer output.

$$\text{Predicted PRICE} = -5.7892 - 7.7134\text{DIV} + 3.8231\text{EPS} + 0.0352\text{SALES} + 0.0396\text{INCOME} - 0.018\text{ASSETS} + 1.5327\text{OLDPR}$$

$R^2 = 0.8043 = 80.43\%$ of the variation in PRICE is explained by this model.

13-51 Results taken from computer output.

$$\text{Predicted PRICE} = -4.0091 - 8.2784\text{DIV} + 4.5006\text{EPS} + 1.4721\text{OLDPR}$$

$R^2 = 0.7947 = 79.47\%$ of the variation in PRICE is explained by this model.

After eliminating SALES, INCOME and ASSETS, the explanatory power of the model was reduced by only 0.96%.

13-52 Results taken from computer output.

$$\text{Predicted PRICE} = -5.9374 - 9.9256\text{DIV} + 4.5839\text{EPS} + 1.4473\text{OLDPR} + 5.1679\text{NY} + 1.2772\text{BANK}$$

$R^2 = 0.8147 = 81.47\%$ of the variation in PRICE is explained by this model.

$H_0 : B_{NY} = 0$ $\qquad$ $H_1 : B_{NY} \neq 0$

Since the prob-value $= 0.137 > \alpha = .10$, we accept H_0. Being listed on the NYSE doesn't appear to have a significant effect on PRICE.

$H_0 : B_{BANK} = 0$ $\qquad$ $H_1 : B_{BANK} \neq 0$

Since the prob-value $= 0.772 > \alpha = .10$, we accept H_0. Neither share prices of banks nor bank holding companies differ significantly from those of other companies in the group.

13-53 Results taken from computer output.

a) H_0: $B_{DIV} = 0$ $\qquad$ $H_1 : B_{DIV} < 0$

$b_{DIV} = -8.2784$
b_{SINGLE} is not significant at α=.05, but it is close ($t = -1.98$, p = .0588)

b) $H_0 : B_{EPS} = 2$ $\qquad$ $H_1 : B_{EPS} > 2$

$$t = \frac{b_{EPS} - B_{EPS_0}}{s_{b_{EPS}}} = \frac{4.5006 - 2}{1.8453} = 1.355, \quad t_{CRIT} = 1.711$$

No. A \$1 increase in EPS does not lead to an increase in share price significantly greater than \$2.

c) $b_{OLDPR} = 1.4721$

$s_b = 0.2212$

$df = n - k - 1 = 28 - 3 - 1 = 24, \quad t_{.02,24} = 2.492$

$b_{OLDPR} \pm t(s_b) = 1.4721 \pm 2.492(.2212) = [0.9209, 2.0233]$

d) Predicted PRICE = $-4.0091 - 8.2784$DIV + 4.5006EPS + 1.4721OLDPR
$\widehat{Y} = -4.0091 - 8.2784(1.51) + 4.5006(4.52)\ 1.4721(40.63) = 63.6447$

The model predicts a share price of \$63.64 on 4/1/93 which is \$8.76 greater than the actual price of \$54.88.

13-54 Results taken from computer output.

a) Predicted REVENUE = 28725.416 − 139.760PROPERTY + 105.176SALES + 56.065GASOLINE

b) Predicted revenues under the two proposals are:

PROPOSAL A: 28725.416 − 139.760(2.763) + 105.176(1) + 56.065(1) = 28500.50

PROPOSAL B: 28725.416 − 139.760(1.639) + 105.176(2.021)) + 56.065(3.3) = 28893.50

If they wish to maximize their revenue, they should adopt proposal B, which yields \$393,000 more than proposal A.

13-55 Results taken from computer output.

a) Predicted PRICE = 8.9087 + 0.00762FRESH + 0.00032PROCESS

b) At FRESH = 980 and PROCESS = 360,
predicted PRICE = 8.9087 + 0.00762(980) + 0.00032(360) = \$16.49

13-56 Results taken from computer output. The PHONES variable was recorded in 100,000's of units. The linear regression equation is:

Predicted PHONES = −6.6325 + 2.6040YEARS, with $r^2 = 0.7951$

The residuals (4.06, 1.77, −0.38, −2.35, −3.50, −3.91, −2.82, 0.51, and 6.62) distinctly show a pattern consistent with a quadratic curve. The quadratic regression equation is:

Predicted PHONES = 3.6280 − 2.9926YEARS + 0.5597(YEARS)^2,

with $r^2 = 0.9836$. This appears to be a much better fit to the data.

13-57 Results taken from computer output.

a)

Variable	Parameter Estimate	Standard Error	T for H_0: Parameter=0	Prob > |T|	Significant?
INTERCEP	373.039158	51.91791040	7.185	0.0880	No
FILL	0.030505	1.11836044	0.027	0.9826	No
WEIGHT	50.493999	7.30263996	6.914	0.0914	No
LOFT	−24.721589	8.21922767	−3.008	0.2043	No
DEGREES	−4.171608	0.38943930	−10.712	0.0593	No

b) PROB > $F = 0.0094$, so at $\alpha = .01$, the regression is significant as a whole.

c) The results in (a) and (b) suggest the presence of multicollinearity. In fact, each pair of independent variables has an r^2 in excess of .85.

13-58 Results taken from computer output:

a) Predicted REVENUE = 8085.6084 + 51.4201STORES − 125.7441SIZE

Since the prob-value for STORES (0.0000) is smaller than that for SIZE (0.0132), the number of stores is more important in determining revenue growth. In fact, larger stores seem to be leading to a decline in revenues. This regression might lead a consultant to emphasize geographic spread.

b) With sales per employee measured in \$1000s,

$$\text{Predicted SALES/EMPLOYEE} = -40841.1511 + 20.8462\text{YEAR} - 5.1665\text{SIZE}$$

Because the coefficient of SIZE is negative, employees are not more productive in larger stores. The positive coefficient of YEARS shows that employee productivity is increasing over time. It appears that sales per employee is growing over time, notwithstanding the adverse effect of larger stores.

13-59 Results taken from computer output:

$$\text{Predicted ROE} = 36.9939 - 1.5736\text{INVENTORY} + 0.5497\text{CLUBS}$$

INVENTORY is a very significant explanatory variable (prob-value = 0.0079), whereas CLUBS is not (prob-value = 0.1745). Looking at the signs of the regression coefficients, it looks like management should decrease inventory and increase the percentage of Sam's Clubs to increase ROE. We must remember, however, that absent other information, we should not assume our regression says anything about causality.

CHAPTER 14

NONPARAMETRIC METHODS

14-1 Parametric tests answer questions about parameters of a population, whereas nonparametric methods answer more general questions about a sample or population, e.g., randomness and rank correlation. Though in some cases the nonparametric tests have implications regarding population parameters, they do not specifically test the parameters' values.

14-2 (b)

14-3 Nonparametric methods do not require restrictive assumptions about the underlying distribution of a population and do not require numbers measured at the interval level. Also, they are computationally simpler.

14-4 Nonparametric methods do not use all the information in the data, since they usually rely on ranks or counts, and they are not as efficient as are parametric tests in estimating and detecting things about parameters.

14-5 Since all the waiting times are greater than or equal to zero, they cannot be normally distributed. In addition, since the waiting times tend to increase throughout the day, their distribution is probably negatively skewed, so in all likelihood, the normal distribution isn't even a good approximation to the time distribution.

14-6 Yes, the company sacrifices a great deal of information by using a ranking test as its decision criterion. If the data were examined by graphing the number of preferences against the combination number, it could be seen that there is a very distinct bi-modal distribution. In this case, the choice of two benefit packages might well be the better alternative. If the company had chosen to use a very simple ranking test for ease of computation, it would have forfeited the information about the two distinct preferences of the participants in the sample.

14-7

Before	98.4	96.6	82.4	96.3	75.4	82.6	81.6	91.4	90.4	92.4
After	82.4	95.4	94.2	97.3	77.5	82.5	81.6	84.5	89.4	90.6
Sign	−	−	+	+	+	−	0	−	−	−

10 responses: 3(+); 6(−); 1(0)

For $n = 9$, $p = .5$, the probability of $\geq$ 3 +'s is .9101 (AppendixTable 3). Since .9101 > .05, do not reject H_0. Employee satisfaction did not increase after the buyout. (In fact, it decreased—but not significantly. The prob-value for testing whether it decreased is the probability of ≥ 6 −'s, which is .2540.)

14-8

Before	33	36	41	32	39	47	34	29	32	34	40	42	33	36	29
After	35	29	38	34	37	47	36	32	30	34	41	38	37	35	28
Sign	+	−	−	+	−	0	+	+	−	0	+	−	+	−	−

15 responses: 6(+); 7(−); 2(0)

For $n = 13$, $p = .5$, the expected number of +'s is 13(.5) = 6.5. We have observed 6 +'s. The probability of being this far or farther away from the expected value is $P(r \geq 6 \text{ or } r \geq 7) = 1$. Since 1 > .05, we accept H_0. There has not been a significant change in collection time.

14-9

Before	After	Sign	Before	After	Sign	Before	After	Sign
29	32	+	44	22	−	33	31	−
34	19	−	41	24	−	34	19	−
32	22	−	23	26	+	20	22	+
19	21	+	34	41	+	21	32	+
31	20	−	25	34	+	22	31	+
22	24	+	42	27	−	45	30	−
28	25	−	20	26	+	43	29	−
31	31	0	25	25	0	31	20	−
32	18	−						

25 responses: 10(+); 13(−); 2(0)

$\overline{p} = \frac{13}{23} = .5652$ $\qquad \sigma_{\overline{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.5(.5)}{23}} = .1043$

$H_0 : p = .5$ (no change) $\qquad H_1 : p > .5$ (productivity has increased)

(A minus sign indicates a smaller task time after the new payment plan, corresponding to increased productivity.)

At $\alpha = .10$, the upper limit of the acceptance region is

$$.5 + 1.28\sigma_{\overline{p}} = .5 + 1.28(.1043) = .63$$

so we accept H_0, since $.5652 < .63$. The plan has not significantly increased productivity.

14-10 a) The meteorologist has a bit of a point, but it is not very strong. Even if 1995 is significantly cooler than 1994, that alone is not strong evidence of a long-run trend toward cooler weather.

b) 15 responses: 5 (+); 9(−); 1(0); (+ indicates a higher 1995 temperature)
For $n = 14$, $p = .5$, the probability of $\geq$ 9 −'s is .2120 (Appendix Table 3).
Since $.2120 > .05$, we must accept H_0. We conclude that there was not any significant cooling from 1994 to 1995.

14-11 a) 11 responses: 7(−); 3(+); 1(0)

b) For $n = 10$, $p = .5$, the probability of $\geq$ 7 −'s is .1719 (Appendix Table 3).
Since $.1719 > .15$, we must accept H_0. We conclude that the level of contamination has not decreased significantly.

14-12 For all parts, H_0 is $p = .5$ (same ideal sizes)
H_1 is $p > .5$ (mother's ideal greater)

(A + response indicates mother's ideal is greater than daughter's.)

a) 13 responses: 6(+); 4(−); 3(0)
For $n = 10$, $p = .5$, the probability of $\geq$ 6 +'s is .3770 (Appendix Table 3).
Since $.3770 \geq .03$, we would accept H_0.

b) $\sigma_{\overline{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.5(.5)}{10}} = .1583$

At $\alpha = .03$, the upper limit of the acceptance region is:

$$.5 + 1.88\sigma_{\overline{p}} = .5 + 1.88(.1583) = .798$$

Since $\overline{p} = .6 < .798$, accept H_0. Daughters don't want significantly smaller families.

c) 143 responses: 66(+); 44(−); 33(0)

$$\sigma_{\overline{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.5(.5)}{110}} = .0477$$

The upper limit of the acceptance region is:

$$.5 + 1.88(.0477) = .590$$

Since $\bar{p} = .6 > .590$, we reject H_0. The ideal family size has decreased.

d) Since $\sigma_{\bar{p}} = \sqrt{pq/n}$, the standard error gets smaller as n gets larger, and we put more trust in the data we have. It is easy to see that in this example, the percentages of smaller, larger, and no difference have not been altered. The only change in the computations of (b) and (c) is in the standard error of the mean. Hence when we took a sample of 143 and got 66 responses showing smaller we could conclude at $\alpha = .03$ that the ideal family size had decreased. With only 13 responses the reliability of our sample was lessened by its small size; hence we could not conclude at a significance level of $\alpha = .03$ that the ideal family size had decreased.

14-13

Before	18.4	16.9	17.4	11.6	10.5	12.7	22.3	18.5	17.5	16.4
After	18.6	16.8	17.3	15.6	19.5	12.6	22.3	16.5	18.0	16.4
Sign	+	−	−	+	+	−	0	−	+	0

Before	15.9	18.6	23.5	18.7	9.4	16.3	18.5	17.4	11.3	8.4
After	17.4	18.6	23.5	18.9	15.6	15.4	17.6	17.4	16.5	13.4
Sign	+	0	0	+	+	−	−	0	+	+

20 responses: 9(+); 6(−); 5(0)

For $n = 15$ $p = .5$, the probability of $\geq$ 9 +'s is .3036 (Appendix Table 3). Since $.3036 > .05$, do not reject H_0. They cannot be 95% confident that average cars sold has increased.

Considering only those salespeople with initial averages below 12, we have 5 responses, all of which are +. For $n = .5$, $p = .5$, the probability of all responses being + is $(.5)^5 = .03125 < .05$. In this case we reject H_0 and conclude that average sales for employees starting below 12 cars per month have increased.

14-14

Men	31	25	38	33	42	40	44	26	43	35
Ranks	4	1	12	6	14	13	16.5	20	15	10
Women	44	30	34	47	35	32	35	47	48	34
Ranks	16.5	3	7.5	18.5	10	5	10	18.5	20	7.5

$n_1 = 10$ $\quad$ $n_2 = 10$ $\quad$ $\alpha = .10$

$R_1 = 93.5$ $\quad$ $R_2 = 116.5$

H_0: $\mu_1 = \mu_2$ $\quad$ H_1: $\mu_1 \neq \mu_2$

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 10(10) + \frac{10(11)}{2} - 93.5 = 61.5$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{10(10)}{2} = 50 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{10(10)(21)}{12}} = 13.23$$

The limits of the acceptance region are:

$$\mu_U \pm 1.96\sigma_U = 50 \pm 1.64(13.23) = 50 \pm 21.70 = (28.30, 71.70)$$

Thus, we accept H_0.

14-15

Brand A	89	90	92	81	76	88	85	95	97	86	100
Ranks	20.5	23	25.5	7.5	2	18.5	12.5	28	30	15.5	31
Brand B	78	93	81	87	89	71	90	96	82	85	
Ranks	3	27	7.5	17	20.5	1	23	29	9	12.5	

Brand C	80	88	86	85	79	80	84	85	90	92
Ranks	5.5	18.5	15.5	12.5	4	5.5	10	12.5	23	25.5

$n_1 = 11$ $\quad n_2 = 10$ $\quad n_3 = 10$ $\quad \alpha = .01$

$R_1 = 214$ $\quad R_2 = 149.5$ $\quad R_3 = 132.5$

H_0: $\mu_A = \mu_B = \mu_C$ $\quad H_1$: the μ's are not all the same

$$K = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{31(32)}\left(\frac{(214)^2}{11} + \frac{(149.5)^2}{10} + \frac{(132.5)^2}{10}\right) - 3(32) = 2.636$$

With $3 - 1 = 2$ degrees of freedom, the upper limit of the acceptance region is $\chi^2_{2,.01} = 9.210$, so we accept H_0. The average retail prices for the three brands are not significantly different.

14-16

Credit cards	78	64	75	45	82	69	60
Ranks	17	8	16	1	18	12	6
Checks	110	70	53	51	61	68	
Ranks	20	13.5	3	2	7	10.5	
Cash	90	68	70	54	74	65	59
Ranks	19	10.5	13.5	4	15	9	5

$n_1 = 7$ $\quad n_2 = 6$ $\quad n_3 = 7$ $\quad \alpha = .05$

$R_1 = 78$ $\quad R_2 = 56$ $\quad R_3 = 76$

H_0: $\mu_1 = \mu_2 = \mu_3$ $\quad H_1$: the μ's are not all the same

$$K = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{20(21)}\left(\frac{78^2}{7} + \frac{56^2}{6} + \frac{76^2}{7}\right) - 3(21) = 0.341$$

With $3 - 1 = 2$ degrees of freedom, the upper limit of the acceptance region is $\chi^2_{2,.05} = 5.991$, so we accept H_0. The average amounts paid by the three methods are not significantly different.

14-17

Men	31	44	25	30	70	63	54	42	36	22	25
Ranks	7	16	4.5	6	23	22	20	15	10.5	2	4.5
Women	38	34	33	47	58	83	18	36	41	37	24
Ranks	13	9	8	17	21	24	1	10.5	14	12	3

$n_1 = 12$ $\quad n_2 = 12$ $\quad \alpha = .10$

$R_1 = 149.5$ $\quad R_2 = 150.5$

H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 12(12) + \frac{12(13)}{2} - 149.5 = 72.5$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{12(12)}{2} = 72 \qquad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{12(12)(25)}{12}} = 17.32$$

The limits of the acceptance region are:

$$\mu_U \pm 1.64\sigma_U = 72 \pm 1.64(17.32) = 72 \pm 28.40 = (43.60, 100.40)$$

Since $U = 72.5$, we accept H_0. There is no significant difference between men and women in hours missed due to illness.

14-18

Old Machine	Ranks	New Machine	Ranks
992	21	965	17
945	15	1054	25
938	12.5	912	11
1027	24	850	2
892	6	796	1
983	20	911	10
1014	22	866	4
1258	26	902	9
966	18	956	16
889	5	900	8
972	19	938	12.5
940	14		
873	3		
1016	23		
897	7		

$n_1 = 15$ $\quad n_2 = 11$ $\quad \alpha = .10$

$R_1 = 235.5$ $\quad R_2 = 115.5$

H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 > \mu_2$

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 15(11) + \frac{11(12)}{2} - 115.5 = 115.5$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{15(11)}{2} = 82.5 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{15(11)(27)}{12}} = 19.27$$

We were asked if output has been reduced. If it has, we should observe a small value of R_2 and, accordingly, a large value of U. Thus an upper-tail test is appropriate.

The upper limit of the acceptance region is:

$$\mu_U + 1.28\,\sigma_U = 82.5 + 1.28(19.27) = 107.2$$

Since $U = 115.5 > 107.2$, we reject H_0. The change has reduced output significantly.

14-19

Visitors north	755	698	725	895	886	794	694	827	814
Ranks	13	5	9	34	32	21	4	29	23
Visitors south	782	724	754	825	815	826	752	784	789
Ranks	18	8	12	27	24	28	11	19	20
Home north	714	758	684	816	856	884	774	812	734
Ranks	6	14	2	25	30	31	15	22	10
Home south	776	824	654	779	898	687	716	889	917
Ranks	16	26	1	17	35	3	7	33	36

All $n_j = 9$ $\quad R_1 = 170$ $\quad R_2 = 167$ $\quad R_3 = 155$ $\quad R_4 = 174$ $\quad \alpha = .10$

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ $\quad H_1$: the μ's are not all the same

$$K = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{36(37)}\left(\frac{170^2}{9} + \frac{167^2}{9} + \frac{155^2}{9} + \frac{174^2}{9}\right) - 3(37) = 0.2012$$

With $4 - 1 = 3$ degrees of freedom, the upper limit of the acceptance region is $\chi^2_{3,.10} = 6.251$, so we do not reject H_0. The average hot dog sales at the four stands are not significantly different.

14-20

Promotion	18	21	23	15	19	26	17	18	22	20	18	21	27
Ranks	8	14	19	1.5	10	23.5	5	8	16.5	11.5	8	14	25

Regular	22	17	15	23	25	20	26	24	16	17	23	21
Ranks	16.5	5	1.5	19	22	11.5	23.5	21	3	5	19	14

$n_1 = 13$ $\quad n_2 = 12$ $\quad \alpha = .05$

$R_1 = 164$ $\quad R_2 = 161$

H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 > \mu_2$

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 13(12) + \frac{12(13)}{2} - 161 = 73$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{12(13)}{2} = 78 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{12(13)(26)}{12}} = 18.38$$

If the promotion produced greater sales, R_2 should be low and U should be high, so an upper tail test is appropriate. The upper limit of the acceptance region is:

$$\mu_U + 1.64\sigma_U = 78 + 1.64(18.38) = 108.14$$

Since $U = 73 < 108.14$, we accept H_0. The promotion does not increase sales.

14-21 a) Given $\sigma_U = 176.4275$ $\quad \mu_U = 1624$ $\quad R_1 = 3255$ $\quad n_2 - n_1 = 2$

$$\mu_U = 1624 = \frac{n_1 n_2}{2} = \frac{n_1(n_1+2)}{2}$$

$$n_1^2 + 2n_1 - 3248 = 0 = (n_1+58)(n_1-56) \Rightarrow n_1 = 56,\ n_2 = 58$$

Now $R_1 + R_2 = 1 + 2 + 3 + \ldots + 112 + 113 + 114 = 6555$

so $R_2 = 6555 - 3255 = 3300$

b) The limits of the acceptance region are:

$$\mu_U \pm 1.96\sigma_U = 1624 \pm 1.96(176.4275) = 1624 \pm 345.80 = (1278.20, 1969.80)$$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 56(58) + \frac{56(57)}{2} - 3255 = 1589$$

Since $1278.20 < 1589 < 1969.80$, we accept H_0. Outputs are not significantly different.

14-22

Rural	3.19	2.05	2.82	2.16	3.84	4.00	2.91	2.75	3.01	1.98	2.58	2.76	2.94
Ranks	23	4	16	7	26	29	18	13	20	3	9	14	19

Urban	3.45	3.16	2.84	2.09	2.11	3.08	3.97	3.85	3.72	2.73	2.81	2.64	1.57	1.87	2.54	2.62
Ranks	24	22	17	5	6	21	28	27	25	12	15	11	1	2	8	10

$n_1 = 13$ $\quad n_2 = 16$ $\quad \alpha = .05$

$R_1 = 201$ $\quad R_2 = 234$

H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 13(16) + \frac{13(14)}{2} - 201 = 98$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{13(16)}{2} = 104 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{13(16)(30)}{12}} = 22.80$$

The limits of the acceptance region are:

$$\mu_U \pm 1.96\sigma_U = 104 \pm 1.96(22.80) = 104 \pm 44.69 = [59.31, 148.69]$$

Since $U = 98$, we do not reject H_0. There is no significant difference in the two groups' first-year GPAs.

14-23

Salesmasters	90	95	105	110	100	75	80	90	105	120
Ranks	5.5	9.5	16	18	13	1	2.5	5.5	16	19.5
Company	80	90	100	120	95	95	90	100	95	105
Ranks	2.5	5.5	13	19.5	9.5	9.5	5.5	13	9.5	16

$n_1 = 10$ $\quad n_2 = 10$ $\quad \alpha = .10$
$R_1 = 106.5$ $\quad R_2 = 103.5$
H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 10(10) + \frac{10(11)}{2} - 106.5 = 48.5$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{10(10)}{2} = 50 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{10(10)(21)}{12}} = 13.23$$

The limits of the acceptance region are:

$$\mu_U \pm 1.64\sigma_U = 50 \pm 1.64(13.23) = 50 \pm 21.70 = (28.30,\ 71.70)$$

Since $U = 48.5$, we accept H_0. Neither method is significantly better than the other.

14-24

$n_1 = \#$ of A's $= 26$ $\quad r = 27$
$n_2 = \#$ of B's $= 22$ $\quad \alpha = .05$

$$\mu_r = \frac{2n_1 n_2}{n_1+n_2} + 1 = \frac{2(26)(22)}{48} + 1 = 24.83$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(26)(22)[2(26)(22) - 26 - 22]}{(48)^2(47)}} = 3.40$$

The limits of the acceptance region are:

$$\mu_r \pm 1.96\sigma_r = 24.83 \pm 1.96(3.40) = 24.83 \pm 6.66 = (18.17,\ 31.49)$$

so we accept H_0. The mix is random.

14-25

$n_1 = \#$ damaged $= 14$ $\quad r = 9$
$n_2 = \#$ acceptable $= 11$ $\quad \alpha = .05$

$$\mu_r = \frac{2n_1 n_2}{n_1+n_2} + 1 = \frac{2(14)(11)}{25} + 1 = 13.32$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(14)(11)[2(14)(11) - 14 - 11]}{(25)^2(24)}} = 2.41$$

The limits of the acceptance region are:

$$\mu_r \pm 1.96\sigma_r = 13.32 \pm 1.96(2.41) = 13.32 \pm 4.72 = (8.60,\ 18.04)$$

so we accept H_0. The damaged pieces occur randomly in the sequence.

14-26

$n_1 = \#$ of men $= 14$ $\quad r = 13$
$n_2 = \#$ of women $= 14$ $\quad \alpha = .05$

$$\mu_r = \frac{2n_1 n_2}{n_1+n_2} + 1 = \frac{2(1)(14)}{28} + 1 = 15$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(14)(14)[2(14)(14) - 14 - 14]}{(28)^2(27)}} = 2.60$$

The limits of the acceptance region are:

$$\mu_r \pm 1.96\sigma_r = 15 \pm 1.96(2.60) = 15 \pm 5.10 = (9.90, 20.10)$$

so we accept H_0. The sequence is random, as we would have expected.

14-27 The mean number of applicants per day is 6.67. Coding the data as A for above the mean and B for below, we get:

3	4	6	8	4	6	7	2	5	7	4	8	4	7	9	5	9	10	5	7	4	9	8	9	11	6	7	5	9	12
B	B	B	A	B	B	A	B	B	A	B	A	B	A	A	B	A	A	B	A	B	A	A	A	A	B	A	B	A	A

$n_1 = \#$ of A's $= 16$ $\qquad r = 20$
$n_2 = \#$ of B's $= 14$ $\qquad \alpha = .10$

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(16)(14)}{30} + 1 = 15.93$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(16)(14)[2(16)(14) - 16 - 14]}{(30)^2(29)}} = 2.68$$

The limits of the acceptance region are:

$$\mu_r \pm 1.64\sigma_r = 15.93 \pm 1.64(2.68) = 15.93 \pm 4.40 = (11.53, 20.33)$$

so we accept H_0. The sequence is random. Since there is no reason to believe that more counseling is needed on particular days of the week, we would have expected the randomness that was observed.

14-28 $n_1 = \#$ of A's $= 15$ $\qquad r = 10$
$n_2 = \#$ of B's $= 16$ $\qquad \alpha = .05$

"Eyeballing" the data shows that the ages of diners are not randomly mixed, but let's see if this is verified by hard statistical analysis.

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(14)(14)}{28} + 1 = 16.484$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(15)(16)[2(15)(16) - 15 - 16]}{(31)^2(30)}} = 2.734$$

If older couples eat earlier there will be too few runs, so a lower tail test is appropriate. The lower limit of the acceptance region is:

$$\mu_r - 1.64\sigma_r = 16.484 - 1.64(2.734) = 12.00$$

so we reject H_0, since $r = 10 < 12$. The pattern is not random.

14-29 $n_1 = \#$ of breakdowns of older presses $= 23$ $\qquad r = 19$
$n_2 = \#$ of breakdowns of newer presses $= 19$ $\qquad \alpha = .05$

a) $$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(23)(19)}{42} + 1 = 21.81$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(23)(19)[2(23)(19) - 23 - 19]}{(42)^2(41)}} = 3.17$$

The limits of the acceptance region are:

$$\mu_r \pm 1.96\sigma_r = 21.81 \pm 1.96(3.17) = 21.81 \pm 6.21 = (15.60, 28.02)$$

so we accept H_0. The pattern is random.

b) No; we would get the same conclusion if we switched 1, 2, 3 with 4, 5, 6. Second, we have in no way tested to see if the older presses break down sooner than the newer ones. The Mann-Whitney test would be more appropriate.

14-30 a) $n_1 = 45$ $\qquad r = 9$
$n_2 = 4$ $\qquad \alpha = .01$

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(45)(4)}{49} + 1 = 8.35$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(45)(4)[2(45)(4) - 45 - 4]}{(49)^2(48)}} = .99$$

The limits of the acceptance region are:

$$\mu_r \pm 2.58\sigma_r = 8.35 \pm 2.58(.99) = 8.35 \pm 2.55 = (5.80,\ 10.90)$$

so we accept H_0. The sample seems to be random.

b) The acceptance region is the same as before, but now $r = 2 < 5.85$, so we reject reject H_0. Of course, this was obvious by inspection.

c) Certainly. A random sample should be roughly composed of 3 times as many analyses done by machine as by hand. If $p = .75$, the probability of observing $\bar{p} \geq 45/49 = .9184$ in a sample of size 49 is very small:

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.75)(.25)}{49}} = .0619, \quad z = \frac{.9184 - .75}{.0619} = 2.72, \quad p(z \geq 2.72) = .0033.$$

Even odder is the particular sequence observed: nine 1's, a 2, nine 1's, a 2, nine 1's, a 2, nine 1's, a 2, nine 1's.

d) The test only looks at the number of runs in the sample. It does not see if the sample proportion is reasonable. Also, no test would pick up the exact pattern that we noted in part (c).

14-31 $n_1 = \#$ of A's $= 23$ $\qquad r = 15$
$n_2 = \#$ of B's $= 22$ $\qquad \alpha = .05$

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(23)(22)}{45} + 1 = 23.49$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(23)(22)[2(23)(22) - 23 - 22]}{(45)^2(44)}} = 3.31$$

The limits of the acceptance region are:

$$\mu_r \pm 1.96\sigma_r = 23.49 \pm 1.96(3.31) = 23.49 \pm 6.49 = (17.00,\ 29.98).$$

Since $r = 15 < 17$, we reject H_0. The sequence is not random.

14-32 $n_1 = \#$ of men $= 29$ $\qquad r = 17$
$n_2 = \#$ of women $= 11$ $\qquad \alpha = .05$

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(29)(11)}{40} + 1 = 16.95$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(29)(11)[2(29)(11) - 29 - 11]}{(40)^2(39)}} = 2.47$$

The limits of the acceptance region are:

$$\mu_r \pm 1.96\sigma_r = 16.95 \pm 1.96(2.47) = 16.95 \pm 4.84 = (12.11, 21.79)$$

so we accept H_0. The sequence is random, as we would have expected.

14-33

X(ranks)	7	5	6	8	3	1	10	4	9	2
Y(ranks)	6	10	2	1	3	7	4	8	9	5
d	1	−5	4	7	0	−6	6	−4	0	−3
d^2	1	25	16	49	0	36	36	16	0	9

$\sum d^2 = 188$ $n = 10$ $\alpha = .05$

H_0: $\rho_s = 0$ H_1: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(188)}{10(99)} = -0.1394$$

From Appendix Table 7, the critical values for r_s are $\pm$ .6364. Since $-0.1394 > -0.6364$, we do not reject H_0. The correlation is not significant.

14-34

Amount of overtime	5	8	2	4	3	7	1	6
Years employed	1	6	4.5	2	7	8	4.5	3
d	4	2	−2.5	2	−4	−1	−3.5	3
d^2	16	4	6.25	4	16	1	12.25	9

$\sum d^2 = 68.5$ $n = 8$ $\alpha = .01$
H_0: $\rho_s = 0$ H_1: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(68.5)}{8(63)} = .1845$$

From Appendix Table 7, the critical values for r_s are $\pm$ .8571.
Since $.1845 < .8571$, we accept H_0. The correlation is not significant.

14-35

Age rank	2	7	6	1	9	10	8	4	5	3
Grievances rank	8.5	1.5	5.5	5.5	3	1.5	5.5	8.5	5.5	10
d	−6.5	5.5	0.5	−4.5	6	8.5	2.5	−4.5	−0.5	−7
d^2	42.25	30.25	0.25	20.25	36	72.25	6.25	20.25	0.25	49

$\sum d^2 = 277$ $n = 10$ $\alpha = .05$
H_0: $\rho_s = 0$ H_1: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(277)}{10(99)} = -.6788$$

We use a lower-tail test to see if the number of grievances decreases with the manager's age. From Appendix Table 7, the critical value for r_s is .5515. Since $-.6788 < -.5515$, we reject H_0. Relationships do improve with age.

14-36

Company	A	B	C	D	E	F	G	H	I	J	K
Ranks of:											
Expenses	1	7	8	11	10	5	6	3	4	2	9
Accidents	10.5	4.5	6	1	4.5	7.5	2.5	10.5	7.5	9	2.5
d	−9.5	2.5	2	10	5.5	−2.5	3.5	−7.5	−3.5	−7	6.5
d^2	90.25	6.25	4	100	30.25	6.25	12.25	56.25	12.25	49	42.25

$\sum d^2 = 409$ $n = 11$ $\alpha = .01$
H_0: $\rho_s = 0$ H_1: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(409)}{11(120)} = -.8591$$

From Appendix Table 7, the critical values for r_s are $\pm$ .7455. Since $-.8591 < -.7455$, we reject H_0. There is a significant correlation.

14-37

Experience rank	1	3	5.5	5.5	8.5	10	3	7	8.5	3
GPA rank	1	1.5	8	6	7	3	4.5	10	9	2
Success rank	4	2	6	5	7	9	1	8	10	3

a) Experience-success

d	−3	1	−0.5	0.5	1.5	1	2	−1	−1.5	0
d^2	9	1	0.25	0.25	2.25	1	4	1	2.25	0

$\sum d^2 = 21$ $\qquad n = 10$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(21)}{10(99)} = .8727$$

b) GPA-success

d	−3	2.5	2	1	0	−6	3.5	2	−1	−1
d^2	9	6.25	4	1	0	36	12.25	4	1	1

$\sum d^2 = 74.5$ $\qquad n = 10$

$$r_s = 1 - \frac{6(74.5)}{10(99)} = .5485$$

Thus years of experience is better than GPA for predicting success. However, we do not know if this difference is significant, because we have not covered a test for comparing two rank correlations.

14-38

Applicant	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Interview 1	1	11	13	2	12	10	3	4	14	5	6	9	7	8
Interview 2	4	12	11	2	14	10	1	3	13	8	6	7	9	5
d	−3	−1	2	0	−2	0	2	1	1	−3	0	2	−2	3
d^2	9	1	4	0	4	0	4	1	1	9	0	4	4	9

$\sum d^2 = 50$ $\qquad n = 14$ $\qquad \alpha = .01$

H_0: $\rho_s = 0$ $\qquad H_1$: $\rho_s > 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(50)}{14(195)} = .8901$$

Critical r_s (one tail) = .6220 < .8901, so we can conclude that the rankings are positively correlated.

14-39

Output Ranks	4	8	5	6.5	9	2	3	1	10	6.5
Days Together	1	2	3	4	5	6	7	8	9	10
d	3	6	2	2.5	4	−4	−4	−7	1	−3.5
d^2	9	36	4	6.25	16	16	16	49	1	12.25

$\sum d^2 = 165.5$ $\qquad n = 10$ $\qquad \alpha = .05$

H_0: $\rho_s = 0$ $\qquad H_1$: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(165.5)}{10(99)} = -.0030$$

From Appendix Table 7, the critical values for r_s are $\pm$ .6364. Since $-.0030 > -.6364$, we accept H_0. There is no significant correlation.

14-40

Individual	Interview Score	Resume Score	d	d^2
1	15	21	−6	36
2	25	9	16	256
3	1	4	−3	9
4	18	27	−9	81
5	11.5	15	−3.5	12.25
6	30	33	−3	9
7	4	11	−7	49
8	23	30	−7	49
9	34.5	6	28.5	812.25
10	10	10	0	0
11	2	7.5	−5.5	30.25
12	20.5	13.5	7	49
13	30	26	4	16
14	5	17.5	12.5	156.25
15	26.5	12	14.5	210.25
16	7	1	6	36
17	23	13.5	9.5	90.25
18	6	16	−10	100
19	15	20	−5	25
20	19	24	−5	25
21	17	7.5	9.5	90.25
22	26.5	5	21.5	462.25
23	3	3	0	0
24	11.5	17.5	−6	36
25	9	2	7	49
26	13	24	−11	121
27	8	19	−11	121
28	34.5	24	10.5	110.25
29	15	32	−17	289
30	23	28	−5	25
31	30	29	1	1
32	20.5	34	−13.5	182.25
33	28	22	6	36
34	32.5	35	−2.5	6.25
35	32.5	31	1.5	2.25

$\sum d^2 = 3583$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(3583)}{35(35^2-1)} = .4982$$

For a one-tail test at the .01 significance level, the critical z value is 2.33.

$$\sigma_{r_s} = \frac{1}{\sqrt{n-1}} = \frac{1}{\sqrt{34}} = .1715$$

H_0: $\rho_s = 0$ $\qquad$ H_1: $\rho_s > 0$

Thus, the critical value for $r_s = 2.33(.1715) = .3996$. Since $.4982 > .3996$, the firm should recommend that the personal interviews no longer be used.

14-41

Salary Rank	8	6	3	4	1	2	5	7	9	10
Age Rank	4	1	5	9	2	6	7	10	3	8
d	4	5	−2	−5	−1	−4	−2	−3	6	2
d^2	16	25	4	25	1	16	4	9	36	4

$\sum d^2 = 140$ $\quad n = 10 \quad \alpha = .05$

$H_0\colon \rho_s = 0 \quad H_1\colon \rho_s > 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(140)}{10(99)} = .1515$$

An upper-tail test is approximate, since we want to see if salary increases with age. From Appendix Table 7, the critical value for $r_s = .5515$. Since $.1515 < .5515$, we accept H_0. Older candidates do not get significantly higher salaries.

14-42

Interval	1000	1200	900	1450	2000	1300	1650	1700	500	2100
Rank	8	7	9	5	2	6	4	3	10	1
Repair time	40	54	41	60	65	50	42	65	43	66
Rank	10	5	9	4	2.5	6	8	2.5	7	1
d	−2	2	0	1	−0.5	0	−4	0.5	3	0
d^2	4	4	0	1	0.25	0	16	0.25	9	0

$\sum d^2 = 34.5$ $\quad n = 10 \quad \alpha = .10$

$H_0\colon \rho_s = 0 \quad H_1\colon \rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(34.5)}{10(99)} = .7909$$

From Appendix Table 7, the critical values for r_s are $\pm .5515$. Since $.7909 > .5515$, we reject H_0. The rank correlation is significant.

14-43 $\quad \lambda = 3, \ e^{-\lambda} = .049787$

x	f_o	cum. f_o	F_o	F_e	$\lvert F_e - F_o \rvert$
0	6	6	.0600	.0498	.0102
1	18	24	.2400	.1991	.0409
2	30	54	.5400	.4232	.1168
3	24	78	.7800	.6472	.1328 ←
4	11	89	.8900	.8153	.0747
5	2	91	.9100	.9161	.0061
≥ 6	9	100	1.0000	1.0000	.0000

$D_n = .1328$; $D_{\text{table}} = \dfrac{1.36}{\sqrt{n}} = \dfrac{1.36}{\sqrt{100}} = .1360$; $D_n < D_{\text{table}}$, so accept H_0.

The data are well described by a Poisson distribution with $\lambda = 3$.

14-44 a) With $\mu = 98.6$ and $\sigma = 3.78$,

$$P(x < 92.0) = P\left(z < \frac{92.0-98.6}{3.78}\right) = P(z < -1.75) = 0.5 - 0.4599 = 0.0401$$

$$P(92.0 \le x < 96.0) = P\left(-1.75 \le z < \frac{96.0-98.6}{3.78}\right)$$
$$= P(-1.75 \le z < -0.69) = 0.4599 - 0.2549 = 0.2050$$

$$P(96.0 \le x < 100.0) = P\left(-0.69 \le z < \frac{100.0-98.6}{3.78}\right)$$
$$= P(-0.69 \le z < 0.37) = 0.2549 + 0.1443 = 0.3992$$

$$P(100.0 \le x < 104.0) = P\left(0.37 \le z < \frac{104.0-98.6}{3.78}\right)$$
$$= P(0.37 \le z < 1.43) = 0.4236 - 0.1443 = 0.2793$$

$$P(104.0 \le x) = P(1.43 \le z) = 0.5 - 0.4236 = .0764$$

b, c) $n = 69 + 408 + 842 + 621 + 137 = 2077$

For each interval, f_e is 2077 times the probability found in part (a).

Interval	f_o	cum. f_o	F_o	f_e	cum. f_e	F_e	$\lvert F_e - F_o \rvert$
< 92	69	69	0.0332	83.29	83.29	0.0401	0.0069
92 – 95.99	408	477	0.2297	425.78	509.07	0.2451	0.0154 ←
96 – 99.99	842	1319	0.6351	829.14	1338.21	0.6443	0.0092
100 – 103.99	621	1940	0.9340	580.11	1918.32	0.9236	0.0104
≥ 104	137	2077	1.0000	158.68	2077.00	1.0000	0.0000

$D_n = 0.0154$; $D_{\text{table}} = \dfrac{1.22}{\sqrt{n}} = \dfrac{1.22}{\sqrt{2077}} = 0.0268$; $D_n < D_{\text{table}}$, so do not reject Ho. The data are described by a normal distribution, with $\mu = 98.6$ and $\sigma = 3.78$.

14-45

x	f_o	cum. f_o	F_o	f_e	cum. f_e	F_e	$\lvert F_e - F_o \rvert$
1 (51-60)	30	30	.0250	40	40	.0333	.0083
2 (61-70)	100	130	.1083	170	210	.1750	.0667
3 (71-80)	440	570	.4750	500	710	.5917	.1167 ←
4 (81-90)	500	1070	.8917	390	1100	.9167	.0250
5 (91-100)	130	1200	1.0000	100	1200	1.0000	.0000

a) K-S statistic is .1167.

b) $D_{\text{table}} = \dfrac{1.22}{\sqrt{n}} = \dfrac{1.22}{\sqrt{1200}} = .0352$; $D_n > D_{\text{table}}$, so reject H_0.
The data are not well described by the suggested distribution.

14-46

Class	f_o	cum. f_o	F_o	f_e	cum. f_e	F_e	$\lvert F_e - F_o \rvert$
25-30	9	9	.072	6	6	.048	.024
31-36	22	31	.248	17	23	.184	.064 ←
37-42	25	56	.448	32	55	.440	.008
43-48	30	86	.688	35	90	.720	.032
49-54	21	107	.856	18	108	.864	.008
55-60	12	119	.952	13	121	.968	.004
61-66	6	125	1.000	4	125	1.000	.000

$D_n = .064$; $D_{\text{table}} = \dfrac{1.22}{\sqrt{n}} = \dfrac{1.22}{\sqrt{125}} = .1091$; $D_n < D_{\text{table}}$, so accept H_0.
The data are well described by the suggested distribution.

14-47

x	f_o	cum. f_o	F_o	F_e	$\lvert F_e - F_o \rvert$
0	25	25	.0980	.0152	.0828
1	32	57	.2235	.1024	.1211
2	61	118	.4627	.3164	.1463 ←
3	47	165	.6471	.6083	.0388
4	39	204	.8000	.8471	.0471
5	21	225	.8824	.9643	.0819
6	18	243	.9529	.9963	.0434
7	12	255	1.0000	1.0000	.0000

$D_n = .1463$; $D_{\text{table}} = \dfrac{1.36}{\sqrt{n}} = \dfrac{1.36}{\sqrt{255}} = .0852$; $D_n > D_{\text{table}}$, so reject H_0. The data are not well described by the binomial distribution with $n = .45$ and $p = .7$.

14-48 $\lambda = 1$, $e^{-\lambda} = .367879$

x	f_o	cum. f_o	F_o	F_e	$\lvert F_e - F_o \rvert$
0	25	25	.1250	.3679	.2429
1	45	70	.3500	.7358	.3858 ←
2	67	137	.6850	.9197	.2347
3	43	180	.9000	.9810	.0810
≥ 4	20	200	1.0000	1.0000	.0000

$D_n = .3858$; $D_{\text{table}} = \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{200}} = .0962$; $D_n > D_{\text{table}}$, so reject H_0.

The data are not well described by a Poisson distribution with $\lambda = 1$.

14-49 a) $n_1 = \#$ of W's $= 21$ $\qquad r = 7$

$n_2 = \#$ of L's $= 4$ $\qquad \alpha = .10$

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(21)(4)}{25} + 1 = 7.72$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(21)(4)[2(21)(4) - 21 - 4]}{(25)^2(24)}} = 1.27$$

The limits of the acceptance region are:

$$\mu_r \pm 1.64\sigma_r = 7.72 \pm 1.64(1.27) = 7.72 \pm 2.08 = (5.64,\ 9.80)$$

so we accept H_0. The sequence is random.

b) Even though there are a very large number of W's, our test still shows the data to be randomly distributed. The test then confines itself only to the occurrences listed and does not make any inferences about the underlying population distribution. In other words, our procedure tests only how items are distributed in the sample and ignores frequency of occurrences.

14-50

Before Data	27	15	20	24	13	18	30	46	15	29	17	21	18
Ranks	22	7.5	15.5	19	5	12.5	25	26	7.5	24	11	17	12.5
After Data	26	23	19	12	25	9	16	12	28	20	16	14	11
Ranks	21	18	14	3.5	20	1	9.5	3.5	23	15.5	9.5	6	2

$n_1 = 13$ $\qquad n_2 = 13$ $\qquad \alpha = .02$

$R_1 = 204.5$ $\qquad R_2 = 146.5$

H_0: $\mu_1 = \mu_2$ $\qquad H_1$: $\mu_1 > \mu_2$

$$U = n_1n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 13(13) + \frac{13(14)}{2} - 146.5 = 113.5$$

$$\mu_U = \frac{n_1n_2}{2} = \frac{13(13)}{2} = 84.5 \qquad \sigma_U = \sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{13(13)(27)}{12}} = 19.52$$

If complaints declined, R_2 should be small and hence U should be large. An upper-tail test is appropriate. The upper limit of the acceptance region is:

$$\mu_U + 2.05\sigma_U = 84.5 + 2.05(19.5) = 124.47$$

so we accept H_0, since $U = 113.5 < 134.47$. There has not been a significant reduction.

14-51 a)

Before Data	22	18	19	20	31	22	25	19	22	24	18	16	14	28	23	15	16
Ranks	15	8	11	13	28	15	20.5	11	15	18	8	4	1	23.5	17	2	4
After Data	25	28	18	30	33	25	29	32	19	16	30	33	17	25			
Ranks	20.5	23.5	8	26.5	30.5	20.5	25	29	11	4	26.5	30.5	6	20.5			

$n_1 = 17$ $\quad n_2 = 14$ $\quad \alpha = .10$
$R_1 = 214$ $\quad R_2 = 282$
$H_0\colon\ \mu_1 = \mu_2$ $\quad H_1\colon\ \mu_1 < \mu_2$

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 17(14) + \frac{14(15)}{2} - 282 = 61$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{17(14)}{2} = 119 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{17(14)(32)}{12}} = 25.19$$

If ratings improved, R_2 should be large and so U should be small. A lower-tail test is appropriate. The lower limit of the acceptance region is:

$$\mu_U - 1.28\sigma_U = 119 - 1.28(25.19) = 86.76$$

so we reject H_0, since $U = 61 < 86.76$. The improvement is significant.

b) Possibly, but in order to give a definitive answer, we would have to analyze other variables that might affect the ratings--e.g., the quality of acting, the competition's shows, the weather, etc. A change in any of these could also have caused the ratings to change. To help formulate our opinion, a multiple regression analysis would probably be in order.

14-52

Infantry	72	80	86	90	95	92	88	96	91	82
Ranks	2	6.5	13	15.5	19	18	14	20	17	9.5
Transport	80	79	90	82	81	84	78	74	85	71
Ranks	6.5	5	15.5	9.5	8	11	4	3	12	1

$n_1 = 10$ $\quad n_2 = 10$ $\quad \alpha = .05$
$R_1 = 134.5$ $\quad R_2 = 75.5$
$H_0\colon\ \mu_1 = \mu_2$ $\quad H_1\colon\ \mu_1 > \mu_2$

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 10(10) + \frac{10(11)}{2} - 75.5 = 79.5$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{10(10)}{2} = 50 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{10(10)(21)}{12}} = 13.23$$

If ratings are lower in the transport command, R_2 should be low and hence U should be high, so an upper-tail test is appropriate. The upper limit of the acceptance region is:

$$\mu_U + 1.64\sigma_U = 50 + 1.64(13.23) = 71.70$$

so we reject H_0, since $U = 79.5 > 71.70$. The transport command has significantly lower ratings.

14-53

Students	2	7	8	9	6	1	3	4	10	5
Firms	1	4	3	2	6	10	7	8	5	9
Sign	−	−	−	−	0	+	+	+	−	+

10 responses: 4(+); 5(−); 1(0)

For $n = 9$, $p = .5$, the probability of ≥ 5 −'s is .5001 (Appendix table 3). Since .5001 is > .10, we accept H_0. We conclude that there is no significant difference between the firm and student rankings.

14-54

Bus. Wk.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
USN&WR	4	6	2	3	7	10	1	18	8	16	11	9	5	12	17	14	15	13	19	20
Sign	+	+	−	−	+	+	−	+	−	−	0	−	−	−	+	−	−	−	0	0

20 responses: 6(+); 11(−); 3(0)

$\overline{p} = \frac{6}{17} = .3529 \qquad \sigma_{\overline{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.5(.5)}{17}} = 0.1213$

H_0: $p = .5$ (no difference) $\qquad$ H_1: $p \neq .5$

At $\alpha = .10$, the limits of the acceptance region are:

$$.5 \pm 1.645\sigma_{\overline{p}} = .5 \pm 1.645(.1213) = [.3005, .6995]$$

so we accept H_0, since $.3011 < .3529 < .6995$. There is no significant difference between the magazine rankings.

14-55 32 responses: 22(+); 8(−); 2(0). (A + denotes higher spending after the legislation was put into effect.)

$\overline{p} = \frac{22}{30} = .7333 \qquad \sigma_{\overline{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.5(.5)}{30}} = .0913$

$H_0 : p = .5 \qquad H_1 : p > .5$

At $\alpha = .03$, the upper limit of the acceptance region is:

$$.5 + 1.88\sigma_{\overline{p}} = .5 + 1.88(.0913) = .6716$$

so we reject H_0, since $.7333 > .6716$. Therefore the tax reduction has achieved its desired goals.

14-56 Although historical data enable him to know what sort of weather to expect at any season of the year, the weather conditions that actually occur on any given day are quite random.

14-57 $n_1 = \#$ of A's $= 28 \qquad r = 28$

$n_2 = \#$ of B's $= 24 \qquad \alpha = .05$

$$\mu_r = \frac{2n_1n_2}{n_1 + n_2} + 1 = \frac{2(28)(24)}{52} + 1 = 26.85$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} = \sqrt{\frac{2(28)(24)[2(28)(24) - 28 - 24]}{(52)^2(51)}} = 3.55$$

The limits of the acceptance region are:

$$\mu_r \pm 1.96\sigma_r = 26.85 \pm 1.96(3.55) = 26.85 \pm 6.96 = (19.89, 33.81)$$

so we accept H_0. The rainfall is random.

14-58

Letters	35	85	90	92	88	46	78	57	85	67
Ranks	1	10.5	18	19	16	4	11	5	13.5	7
Brochures	42	74	82	87	45	73	89	75	60	94
Ranks	2	9	12	15	3	8	17	10	6	20

$n_1 = 10 \qquad n_2 = 10 \qquad \alpha = .15$

$R_1 = 108 \qquad R_2 = 102$

H_0: $\mu_1 = \mu_2 \qquad H_1$: $\mu_1 < \mu_2$

$$U = n_1n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 10(10) + \frac{10(11)}{2} - 102 = 53$$

$$\mu_U = \frac{n_1n_2}{2} = \frac{10(10)}{2} = 50 \qquad \sigma_U = \sqrt{\frac{n_1n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{10(10)(21)}{12}} = 13.23$$

If brochures are more effective, then R_2 should be high, and hence U should be low, so a lower-tail test is appropriate. Since $U > \mu_U$, we must accept H_0. Her hunch is not supported by the data.

14-59

Advertising Rank	11	3	9	6	2	10	1	4	7	5	8
Sales Rank	10	1	9	5	2	11	3	8	6	4	7
d	1	2	0	1	0	-1	-2	-4	1	1	1
d^2	1	4	0	1	0	1	4	16	1	1	1

$\sum d^2 = 30$ $\quad n = 11$ $\quad \alpha = .05$

H_0: $\rho_s = 0$ $\quad H_1$: $\rho_s > 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(30)}{11(120)} = .8636$$

An upper-tail test is appropriate, since we want to see if a significantly positive rank correlation exists. From Appendix Table 7, the critical value for r_s is .5273. Since .8636 > .5273, we reject H_0. The rank correlation is significantly positive.

14-60

Size	R	5 Year Ave.	R	d	d^2	1992	R	d	d^2
21.05	1	11.24	5	-4	16	9.51	15	-10	100
14.03	2	9.50	10	-8	64	11.08	9	1	1
9.48	3	8.99	14	-11	121	11.35	8	6	36
8.23	4	7.00	17	-13	169	9.53	14	3	9
5.77	5	8.73	15	-10	100	10.87	10	5	25
5.64	6	11.57	4	2	4	16.33	1	3	9
5.62	7	9.38	11	-4	16	15.11	2	9	81
5.10	8	9.34	12	-4	16	11.44	6	6	36
4.98	9	11.07	6	3	9	5.77	18	-12	144
4.80	10	9.59	8	2	4	14.71	3	5	25
4.67	11	10.03	7	4	16	11.42	7	0	0
4.66	12	14.70	1	11	121	8.55	16	-15	225
4.65	13	7.29	16	-3	9	12.45	4	12	144
4.60	14	9.06	13	1	1	11.59	5	8	64
4.47	15	6.25	18	-3	9	2.02	19	-1	1
4.40	16	9.52	9	7	49	10.84	11	-2	4
4.29	17	11.80	3	14	144	10.51	12	-9	81
4.02	18	5.47	19	-1	1	7.00	17	2	4
4.01	19	14.55	2	17	289	1.24	20	-18	324
3.97	20	4.78	20	0	0	9.92	13	7	49
				$\sum d^2 =$	1210			$\sum d^2 =$	1362

a) H_0: $\rho_s = 0$ $\quad H_1$: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(1210)}{20(399)} = 0.0902$$

From Appendix Table 7, the critical values for r_s are $\pm$ 0.2977. Since the probability value is greater than 0.20, we accept H_0. There is not a significant relationship between fund size and the annualized 5 year total return.

b) H_0: $\rho_s = 0$ $\quad H_1$: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(1362)}{20(399)} = -0.0241$$

So again we accept H_0. There is not a significant relationship between the 1992 total return and the annualized 5 year return.

14-61 1992 Total Returns:

Fidelity\20th	9.51	15.11	12.45	2.02	10.84	10.51	1.24						
Ranks	6	19	17	2	10	9	1						
Others	11.08	11.35	9.53	10.87	16.33	11.44	5.77	14.71	11.42	8.55	11.59	7.00	9.92
Ranks	12	13	7	11	20	15	3	18	14	5	16	4	8

$n_1 = 7$ $\quad n_2 = 13$ $\quad \alpha = .10$

$R_1 = 64$ $\quad R_2 = 146$

$H_0\!: \mu_1 = \mu_2$ $\quad H_1\!: \mu_1 \neq \mu_2$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 7(13) + \frac{7(8)}{2} - 64 = 55$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{7(13)}{2} = 45.5 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{7(13)(21)}{12}} = 12.6194$$

The limits of the acceptance region are:

$$\mu_U \pm 1.64\sigma_U = 45.5 \pm 1.64(12.6194) = (24.80,\ 66.2)$$

Since $U = 55 < 66.2$, we accept H_0. There is no significant difference in 1992 total returns when comparing the combined Fidelity/Twentieth Century funds with the other funds.

5 Year Average Total Returns:

Fidelity\20th	11.24	9.38	7.29	6.25	9.52	11.80	14.55						
Ranks	16	10	5	3	12	18	1						
Others	9.50	8.99	7.00	8.73	11.57	9.34	11.07	9.59	10.03	14.70	9.06	5.47	4.78
Ranks	11	7	4	6	17	9	15	13	14	20	8	2	1

$n_1 = 7$ $\quad n_2 = 13$ $\quad \alpha = .10$

$R_1 = 83$ $\quad R_2 = 127$

$H_0\!: \mu_1 = \mu_2$ $\quad H_1\!: \mu_1 \neq \mu_2$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 7(13) + \frac{7(8)}{2} - 83 = 36$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{7(13)}{2} = 45.5 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{7(13)(21)}{12}} = 12.6194$$

The limits of the acceptance region are:

$$\mu_U \pm 1.64\sigma_U = 45.5 \pm 1.64(12.6194) = (24.80,\ 66.2)$$

Since $U = 36 < 66.2$, we accept H_0. There is no significant difference in the five year average total returns when comparing the combined Fidelity/Twentieth Century funds with the other funds.

32

Graphite	110	120	130	110	100	105	110	130	145	125
Ranks	9.5	13.5	16.5	9.5	2.5	6	9.5	16.5	20	15
Bronze	100	110	135	105	105	100	100	115	135	120
Ranks	2.5	9.5	18.5	6	6	2.5	2.5	12	18.5	13.5

$n_1 = 10$ $\quad n_2 = 10$ $\quad \alpha = .05$

$R_1 = 118.5$ $\quad R_2 = 91.5$

$H_0\!: \mu_1 = \mu_2$ $\quad H_1\!: \mu_1 \neq \mu_2$

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 10(10) + \frac{10(11)}{2} - 91.5 = 63.5$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{10(10)}{2} = 50 \qquad \sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{10(10)(21)}{12}} = 13.23$$

The limits of the acceptance region are:

$$\mu_U \pm 1.96\sigma_U = 50 \pm 1.96(13.23) = 50 \pm 25.93 = (24.07, 75.93)$$

Since $U = 63.5$, we accept H_0. The stopping distances are not significantly different.

14-63 The quality of the information received must first be evaluated before it can be determined what will be "lost" by using distribution-free tests. In the example we have non-experts, i.e., customers, attempting to rate on a scale of one to ten how they feel about the restaurants. Given a scale of such a broad range and given no established norms as to how the form will be filled out, the actual numbers supplied by the customers take on a very small significance. As the magazine workers point out, what is really important is the ranking of the data. Therefore, using a ranking test as opposed to a test which uses the real data will not in fact sacrifice any significant information.

Accepting the hypothesis that little will be sacrificed, it might be advisable to challenge the argument that statistical tests on real data are sharper than nonparametric tests. The caveat, now that we realize the quality of the input data, is to not attempt to infer too much from our data by applying high-powered statistical tests. The magazine workers' arguments have, therefore, considerable merit.

14-64

Raw Score	63	59	50	60	66	57	76	81	58	65
Rank	8	5	1	6	10	3	14	17	4	9
Interview #	1	2	3	4	5	6	7	8	9	10
d	7	3	−2	2	5	−3	7	9	−5	1
d^2	49	9	4	4	25	9	49	81	25	1

Raw Score	77	61	53	74	82	70	75	90	80	89
Rank	15	7	2	12	18	11	13	20	16	19
Interview #	11	12	13	14	15	16	17	18	19	20
d	4	−5	−11	−2	3	−5	−4	2	−3	−1
d^2	16	25	121	4	9	25	16	4	9	1

$\sum d^2 = 486$ $\quad n = 20$ $\quad \alpha = .02$

H_0: $\rho_s = 0$ $\quad H_1$: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(486)}{20(399)} = .6346$$

From Appendix Table 7, the critical values for r_s are $\pm$ .5203, so we reject H_0, since .6346 > .5203. This supports her suspicion.

14-65 a)

# of Accidents	0	1	2	3	4	5	6
Frequency	5	7	9	9	8	3	1
Cum. Frequency	5	12	21	30	38	41	42

The median is halfway between the 21st and 22nd elements in an ordered array of the data, so 2.5 is the median. Coding the data as A for the above median and B for the below, we get:

5 3 4 2 6 4 3 3 2 4 5 3 4 4 3 3 3 4 0 5 4 2 0 1
A A A B A A A A B A A A A A A A A A B A A B B B

3 2 1 1 0 2 4 3 2 1 1 2 2 1 0 0 1 2
A B B B B B A A B B B B B B B B B B

n_1 = # of A's = 21 $\quad r = 12$

n_2 = # of B's = 21 $\quad \alpha = .03$

$$\mu_r = \frac{2n_1n_2}{n_1 + n_2} + 1 = \frac{2(21)(21)}{42} + 1 = 22$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} = \sqrt{\frac{2(21)(21)[2(21)(21) - 21 - 21]}{(42)^2(41)}} = 3.20$$

The limits of the acceptance region are:

$$\mu_r \pm 2.17\sigma_r = 22 \pm 2.17(3.20) = 22 \pm 6.94 = (15.06,\ 28.94)$$

so we reject H_0, since $12 < 15.06$. The sequence is not random.

b) The number of accidents seems to have declined since the light was installed, but we can only infer relationship, not causality. We don't know what other factors may also be involved. It should be pointed out that the runs test does not allow us to conclude that accidents have declined, only that the sequence of number of accidents per month is not random. A linear regression of number of accidents on time would give a better test.

14-66

x	f_o	cum. f_o	F_o	F_e	$\lvert F_e - F_o \rvert$
0	5	5	.0556	.1785	.1229 ←
1	35	40	.4444	.5630	.1186
2	30	70	.7778	.8735	.0957
3	13	83	.9222	.9850	.0628
4	7	90	1.0000	1.0000	.0000

$D_n = .1229$; $D_{\text{table}} = \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{90}} = .1434$; $D_n < D_{\text{table}}$, so accept H_0.

The data are well described by a binomial distribution with $n = 4$ and $p = .35$.

14-67

U.S.	Ranks	Non-U.S.	Ranks	Non-U.S.	Ranks	Non-U.S.	Ranks
1978	27	1982	39.5	1977	23.5	1981	35.5
1978	27	1982	39.5	1975	17	1982	39.5
1983	43	1975	17	1975	17	1973	8.5
1982	39.5	1975	17	1985	45.5	1975	17
1969	3	1990	63	1985	45.5	1974	12
1968	1.5	1990	63	1986	48.5	1974	12
1968	1.5	1973	8.5	1986	48.5	1989	57.5
1974	12	1973	8.5	1986	48.5	1990	63
1973	8.5	1981	35.5	1987	51.5	1972	6
1977	23.5	1983	43	1989	57.5	1989	57.5
1977	23.5	1983	43	1988	53	1989	57.5
1978	27	1989	57.5	1989	57.5	1976	21
1977	23.5	1989	57.5	1989	57.5	1975	17
1971	5	1980	32	1979	29.5	1975	17
1970	4	1980	32	1981	35.5	1986	48.5
						1987	51.5
						1980	32
						1981	35.5
						1979	29.5

$n_1 = 15$ $\quad n_2 = 49$ $\quad \alpha = .10$

$R_1 = 269.5$ $\quad R_2 = 1810.5$

H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$

$$U = n_1n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 15(49) + \frac{15(16)}{2} - 269.5 = 585.5$$

$$\mu_U = \frac{n_1n_2}{2} = \frac{15(49)}{2} = 367.5 \qquad \sigma_U = \sqrt{\frac{n_1n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{15(49)(65)}{12}} = 63.0971$$

The limits of the acceptance region are:

$$\mu_U \pm 1.64\sigma_U = 367.5 \pm 1.64(63.0971) = (264.021,\ 470.9792)$$

Since $U = 585.5 > 470.9792$, we reject H_0. There is a significant difference between the ages of the U.S. and Non-U.S. fleets.

14-68

Bulk	Ranks	Tanker	Ranks	PPC	Ranks
1978	27	1974	12	1983	43
1978	27	1973	8.5	1982	39.5
1982	39.5	1977	23.5	1969	3
1982	39.5	1977	23.5	1968	1.5
1975	17	1978	27	1968	1.5
1975	17	1977	23.5	1986	48.5
1990	63	1971	5	1986	48.5
1990	63	1970	4	1986	48.5
1973	8.5	1973	8.5	1987	51.5
1973	8.5	1975	17	1989	57.5
1981	35.5	1974	12	1988	53
1983	43	1974	12	1989	57.5
1983	43	1989	57.5	1989	57.5
1989	57.5	1990	63	1979	29.5
1989	57.5	1972	6	1981	35.5
1980	32	1989	57.5	1981	35.5
1980	32	1989	57.5	1982	39.5
1977	23.5	1976	21		
1975	17	1975	17		
1975	17	1975	17		
1985	45.5	1986	48.5		
1985	45.5	1987	51.5		
		1980	32		
		1981	35.5		
		1979	29.5		

$n_1 = 22$ $\quad n_2 = 25$ $\quad n_3 = 17$

$R_1 = 759$ $\quad R_2 = 670$ $\quad R_3 = 651$

H_0: $\mu_1 = \mu_2 = \mu_3$ $\quad H_1$: the μ's are not all the same

$$K = \frac{12}{n(n+1)}\sum \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{64(65)}\left(\frac{759^2}{22} + \frac{670^2}{25} + \frac{651^2}{17}\right) - 3(65) = 4.2432$$

With $3 - 1 = 2$ degrees of freedom, the upper limit of the acceptance region is $\chi^2_{2,.10} = 4.605$. The probability value is $> .10$, so we accept H_0. The average ages of the three types of vessels are not significantly different.

14-69

Hours Rank	6	2	3	4	1	10	5	8	7	9
Age Rank	4.5	8.5	2.5	8.5	4.5	6	8.5	8.5	2.5	1
Cost Rank	7	2	3	4	1	10	5	9	8	6

a) Hours-cost

d	−1	0	0	0	0	0	0	−1	−1	3
d_2	1	0	0	0	0	0	0	1	1	9

$\sum d^2 = 12$ $\quad n = 10$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(12)}{10(99)} = .9273$$

b) Age-cost

d	−2.5	6.5	−0.5	4.5	3.5	−4	3.5	−0.5	−5.5	−5
d^2	6.25	42.25	0.25	20.25	12.25	16	12.25	0.25	30.25	25

$\sum d^2 = 165$ $\qquad n = 10 \qquad r_s = 1 - \frac{6(165)}{10(99)} = 0$

Thus, hours in salt water is better than age of engine for predicting repair cost. However, we do not know if this difference is significant because we have not covered a test for comparing two rank correlations.

14-70

Region	US	Latin America	Africa	Europe	USSR	India	China
Population rank	1	3	5	4	2	6	7
Energy rank	7	3	2	6	5	1	4
d	−6	0	3	−2	−3	5	3
d^2	36	0	9	4	9	25	9

$\sum d^2 = 92$ $\qquad n = 10 \qquad \alpha = .10$

H_0: $\rho_s = 0$ $\qquad H_1$: $\rho_s < 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(92)}{7(48)} = -.6429$$

From Appendix Table 7, the critical value for r_s is −.5357, so we reject H_0, since −.6429 < −.5357. SavEnergy's claim is supported by the data.

14-71

Large	1.2	1.3	1.4	1.5	1.5	1.5	1.6	1.8						
Ranks	4.5	9	13	17	17	17	22	29						
Midsize	1.1	1.2	1.2	1.2	1.3	1.3	1.3	1.3	1.4	1.4	1.5	1.6	1.6	
Ranks	1.5	4.5	4.5	4.5	9	9	9	9	13	13	17	22	22	
Midsize	1.6	1.7	1.7	1.8	1.9	2.0	2.3	2.3	2.4	2.5	2.6	2.9		
Ranks	22	26	26	29	31	33.5	37	37	39	40.5	42.5	45		
Small	1.1	1.5	1.6	1.7	1.8	2.0	2.0	2.0	2.3	2.5	2.6	2.8	3.2	4.1
Ranks	1.5	17	22	26	29	33.5	33.5	33.5	37	40.5	42.5	44	46	47

$n_1 = 8 \qquad n_2 = 7 \qquad n_3 = 14 \qquad \alpha = .05$

$R_1 = 128.5 \qquad R_2 = 546.5 \qquad R_3 = 453$

H_0: $\mu_1 = \mu_2 = \mu_3$ $\qquad H_1$: the μ's are not all the same

$$K = \frac{12}{n(n+1)}\sum \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{47(48)}\left(\frac{128.5^2}{8} + \frac{546.5^2}{25} + \frac{453^2}{14}\right) - 3(48) = 8.491$$

With $3 - 1 = 2$ degrees of freedom, the upper limit of the acceptance region is $\chi^2_{2,.05} = 5.991$, so we reject H_0. The average death rates for the three sizes of four-door cars are significantly different.

14-72

American	46	3	7	21	10	10	1	6	7	17	7	16	6	1	0	10
National	22	0	5	19	7	2	1	4	5	14	3	18	4	4	1	4
Sign	−	−	−	−	−	−	0	−	−	−	−	+	−	+	+	−

16 responses: 3(+); 12(−); 1(0)

H_0: $p = .5$ $\qquad H_1$: $p > .5$ (Here p is the probability of getting a −.)

For $n = 15$ and $p = .5$, the probability of $\geq$ 12 −'s is .0176 (Appendix Table 3). Since $.0176 < .05$, we reject H_0; American League players do suffer more injuries.

14-73

Solar activity rank	15	11	5	3	6	12	10	7	3	1	8.5	14	13	8.5	3
East wind rank	11.5	3	9	11.5	7.5	4	6	14	13	10	7.5	1	5	2	15
d	3.5	8	−4	−8.5	−1.5	8	4	−7	−10	−9	1	13	8	6.5	−12
d^2	12.25	64	16	72.25	2.25	64	16	49	100	81	1	169	64	42.25	144

$\sum d^2 = 897$ $\quad n = 15 \quad \alpha = .05$

H_0: $\rho_s = 0$ $\quad H_1$: $\rho_s < 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(897)}{15(224)} = -.6018$$

From Appendix Table 7, the critical value for r_s is −.4429, so we reject H_0, since $-.6018 < -.4429$. The hypothesized relationship for east winds is supported by the data.

Solar activity rank	15	11	5	3	6	12	10	7	3	1	8.5	14	13	8.5	3
West wind rank	14	12	6.5	2.5	4	13	15	6.5	2.5	1	8.5	11	10	8.5	5
d	1	−1	−1.5	.5	2	−1	−5	.5	.5	0	0	3	3	0	−2
d^2	1	1	2.25	.25	4	1	25	.25	.25	0	0	9	9	0	4

$\sum d^2 = 57$ $\quad n = 15 \quad \alpha = .05$

H_0: $\rho_s = 0$ $\quad H_1$: $\rho_s > 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(57)}{15(224)} = .8982$$

From Appendix Table 7, the critical value for r_s is .4429, so we reject H_0, since $.8982 > .4429$. The hypothesized relationship for west winds is also supported by the data.

14-74 a) $\frac{1500}{9000} = 0.1667$ $\quad 0.1667(42) = 7$ in each group

b)

f_o	f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$\frac{(f_0 - f_e)^2}{f_e}$
10	7	3	9	1.2857
9	7	2	4	0.5714
9	7	2	4	0.0373
1	7	−6	36	5.1429
9	7	2	4	0.5714
4	7	−3	9	1.2857

$$\chi^2 = \sum \frac{(f_0 - f_e)}{f_e} = 9.4285$$

$\chi^2_{.10,5} = 9.236$

Since $\chi^2 = 9.4285$, we reject the H_0. There is reason to believe that the called bonds were not selected randomly.

c)

Class	f_o	cum. f_o	F_o	f_e	cum. f_e	F_e	$\mid F_e - F_o \mid$
1-1500	10	10	.2381	7	7	.1667	.0714
1501-3000	9	19	.4524	7	14	.3333	.1191
3001-4500	9	28	.6667	7	21	.5000	.1667←
4501-6000	1	29	.6905	7	28	.6667	.0238
6001-7500	9	38	.9048	7	35	.8333	.0715
7501-9000	4	42	1.0000	7	42	1.0000	.0000

$D_n = .1667$; $\quad D_{\text{table}} = \frac{1.14}{\sqrt{n}} = \frac{1.14}{\sqrt{42}} = .1759$

Since the probability value is > .15 ($D_n < D_{\text{table}}$), we accept H_0. The called bonds appear to have been selected randomly.

d) The Kolmogorov-Smirnov test suggested that the bonds were selected randomly, whereas the χ^2 test indicated the opposite result (at α=.10). Since the χ^2 test was barely significant (.10) and the K-S test is more powerful, we should conclude that the bonds were randomly selected.

14-75 $n_1 = \text{Up} = 44$ $\quad r = 30$
$n_2 = \text{Down} = 26$

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(44)(26)}{70} + 1 = 33.6857$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(44)(26)[2(44)(26) - 44 - 26]}{(70)^2(69)}} = 3.8742$$

The limits of the acceptance region are:

$$\mu_r \pm 1.64\sigma_r = 33.69 \pm 1.64(3.8742) = 33.69 \pm 6.3537 = (27.34, 40.04).$$

Since $r = 30 < 40.04$, we accept H_0. The sequence is random.

14-76 $\lambda = 6$, $e^{-\lambda} = .002479$

x	f_0	cum. f_0	F_0	F_e	$/F_e - F_0/$
0	0	0	.0000	.0025	.0025
1	5	5	.1000	.0174	.0826
2	3	8	.1600	.0620	.0980
3	2	10	.2000	.1512	.0488
4	6	16	.3200	.2851	.0349
5	6	22	.4400	.4457	.0057
6	2	24	.4800	.6063	.1263
7	6	30	.6000	.7440	.1440 ←
8	10	40	.8000	.8473	.0473
9	4	44	.8800	.9161	.0361
10	4	48	.9600	.9574	.0026
≥ 11	2	50	1.0000	1.0000	.0000

$D_n = .1440$; $D_{\text{table}} = 1.36/\sqrt{n} = 1.36/\sqrt{50} = .1923$; $D_n < D_{\text{table}}$, so accept H_0. The data are well described by a Poisson distribution with $\lambda = 6$.

14-77 $n_1 = \#\text{ of men} = 67$ $\quad r = 37$
$n_2 = \#\text{ of women} = 26$ $\quad \alpha = .20$

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(67)(26)}{93} + 1 = 38.46$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(67)(26)[2(67)(26) - 67 - 26]}{93^2(92)}} = 3.85$$

The limits of the acceptance region are:

$$\mu_r \pm 1.28\sigma_r = 38.46 \pm 1.28(3.85) = 38.46 \pm 4.93 = (33.53, 43.39)$$

so we accept H_0. The men and women finished randomly.

14-78

A rank	B rank	d=A−B	d^2	C rank	d=A−C	d^2	D rank	d=A−D	d^2
1	2	−1	1	5	−4	16	3	−2	4
2	4	−2	4	3	−1	1	4	−2	4
3	6	−3	9	7	−4	16	5	−2	4
4	3	1	1	2	2	4	1	3	9
5	10	−5	25	11	−6	36	9	−4	16
6	5	1	1	6	0	0	2	4	16
7	22	−15	225	17	−10	100	24	−17	289
8	9	−1	1	4	4	16	15	−7	49
9	18	−9	81	19	−10	100	19	−10	100
10	19	−9	81	30	−20	400	17	−7	49
11	20	−9	81	9	2	4	22	−11	121
12	7	5	25	15	−3	9	11	1	1
13	12	1	1	14	−1	1	13	0	0
14	25	−11	121	12	2	4	28	−14	196
15	16	−1	1	24	−9	81	14	1	1
16	17	−1	1	16	0	0	18	−2	4
17	8	9	81	8	9	81	7	10	100
18	11	7	49	10	8	64	6	12	144
19	24	−5	25	23	−4	16	20	−1	1
20	1	19	361	1	19	361	8	12	144
21	26	−5	25	20	1	1	30	−9	81
22	29	−7	49	27	−5	25	27	−5	25
23	13	10	100	18	5	25	10	13	169
24	28	−4	16	21	3	9	26	−2	4
25	30	−5	25	29	−4	16	29	−4	16
26	14	12	144	13	13	169	23	3	9
27	23	4	16	26	1	1	21	6	36
28	27	1	1	28	0	0	25	3	9
29	15	14	196	22	7	49	12	17	289
30	21	9	81	25	5	25	16	14	196
		$\Sigma d^2 =$	1828			1630			2086

$n = 30$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(1828)}{30(899)} = .5933 \text{ for the League of Women Voters}$$

$$= 1 - \frac{6(1630)}{30(899)} = .6374 \text{ for college students}$$

$$= 1 - \frac{6(2086)}{30(899)} = .5359 \text{ for civic club members}$$

The college students seem to have the most accurate perception of the risks. However, we do not know if the observed differences in rank correlation coefficients are significant, because we have not covered tests for comparing rank correlations.

14-79

Drug	9.0	6.3	2.9	1.4	0.9	0.9	0.6	4.6	2.3	0.9	0.5	0.0	1.0
Control	18.1	3.8	5.8	1.0	0.6	0.2	0.0	2.7	3.5	0.5	0.5	0.2	1.4
Sign	+	−	+	−	−	−	−	−	+	−	0	+	+

13 responses: 5(+); 7(−); 1(0)

H_0: $p = .5$ $\quad$ H_1: $p \neq .5$ (Here p is the probability of getting a +.)

For $n = 12$ and $p = .5$, the expected number of +'s is $12(.5) = 6$. We have observed 5 +'s. The probability of being this far or farther away from the expected value is $P(r \leq 5$ or $r \geq 7) = .7744$ (Appendix Table 3). Since $.7744 > .10$, we accept H_o. There is not a significant difference in the incidence of side effects in the two groups.

14-80

1995 Rank	1	2	3	4	5	6	7	8	9	10
1996 Rank	1	2	8	5	4	10	3	7	6	9
d	0	0	−5	−1	1	−4	4	1	3	1
d^2	0	0	25	1	1	16	16	1	9	1

$\sum d^2 = 70$ $\quad$ $n = 10$ $\quad$ $\alpha = .10$

H_0: $\rho_s = 0$ $\quad$ H_1: $\rho_s \neq 0$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(70)}{10(99)} = .5758$$

A two-tailed test is appropriate, since we want to see if the rankings have changed. From Appendix Table 7, the critical values for r_s are ± 0.5515, so we reject H_0. The rankings have not changed significantly.

CHAPTER 15

TIME SERIES

15-1 Since decisions deal with future actions, the more we know about the future (through forecasting) the better our decisions.

15-2 To determine what patterns exist within the data over the period examined.

15-3 We can project past patterns in order to predict the future.

15-4 Demands for services such as water and sewer would perhaps not be met; adjustment of the tax rate to provide for municipal services might lag behind the actual demand for those services. Extra resources would likely be needed to allow a smooth municipal operation in a situation where forecasting is so poor.

15-5 The secular trend is the overall direction of the time series. Cyclical fluctuations are the periodic changes above and below the trend line which occur over one to fifteen year cycles. Seasonal variations are the periodic changes above and below the trend line which occur during the course of one year. Irregular variations are unpredictable fluctuations which are not explained by the other three components.

15-6 Seasonal

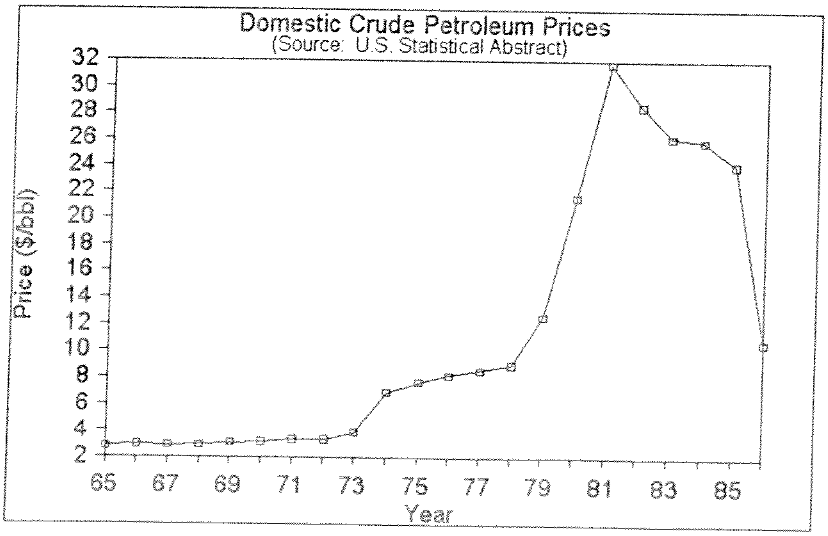
15-7 This allows us to examine each component and to produce a model with which we can predict the future movements of each.

15-8 Cyclical variation

15-9 As an irregular variation

15-10 Secular trend

15-11



As can be seen from the graph above, crude oil prices in 1970-73 continued to follow the gently increasing trend they had followed since 1965 (actually, since the early 1950's). After the Arab embargo of 1973, an irregular variation in 1974 shifted the trend line upward, but the trend then continued through 1978. From 1978 to 1986, we see a cycle whose rise corresponds to OPEC's assertion of monopoly prices and whose fall reflects both the development of non-OPEC supplies and falling demand due to conservation.

15-12 a)

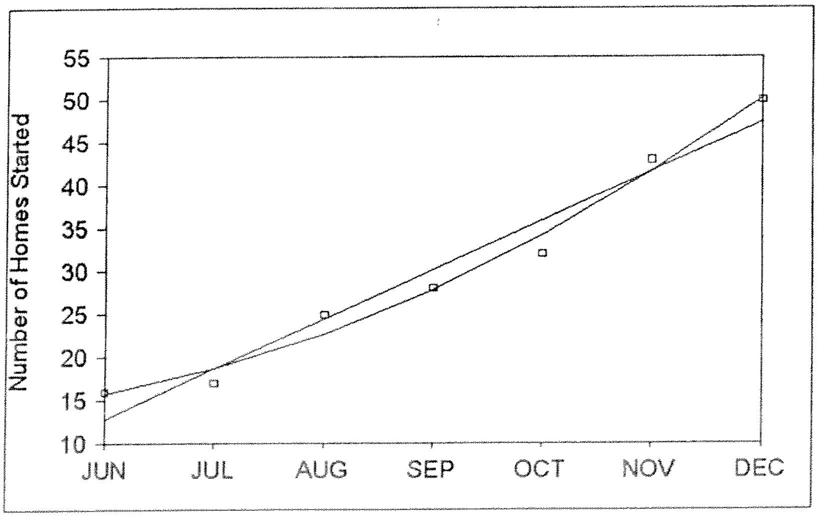
Year	x	Y	xY	x^2
1986	−5	6.4	−32.0	25
1987	−4	11.3	−45.2	16
1988	−3	14.7	−44.1	9
1989	−2	18.4	−36.8	4
1990	−1	19.6	−19.6	1
1991	0	25.7	0.0	0
1992	1	32.5	32.5	1
1993	2	48.7	97.4	4
1994	3	55.4	166.2	9
1995	4	75.7	302.8	16
1996	5	94.3	471.5	25
	0	402.7	892.7	110

$$a = \overline{Y} = \frac{402.7}{11} = 36.6091 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{892.7}{110} = 8.1155$$

$\widehat{Y} = 36.6091 + 8.1155x$ (where 1991 = 0 and x units = 1 year)

b) 1997: $\widehat{Y} = 36.6091 + 8.1155(6) = 85.3$ homes
1998: $\widehat{Y} = 36.6091 + 8.1155(7) = 93.4$ homes
1999: $\widehat{Y} = 36.6091 + 8.1155(8) = 101.5$ homes

15-13 a)



b)

x	Y	xY	x^2	x^2Y	x^4
−3	16	−48	9	144	81
−2	17	−34	4	68	16
−1	25	−25	1	25	1
0	28	0	0	0	0
1	32	32	1	32	1
2	43	86	4	172	16
3	50	150	9	450	81
0	211	161	28	891	196

$$a = \overline{Y} = \frac{211}{7} = 30.1429 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{161}{28} = 5.75$$

$\widehat{Y} = 30.1429 + 5.75x$ (where September = 0 and x units = one month)

c) Equations 15.6 and 15.7 become:

$$\sum Y = na + c\sum x^2 \qquad 211 = 7a + 28c$$

$$\sum x^2 Y = a\sum x^2 + c\sum x^4 \qquad 891 = 28a + 196c$$

Solving these simultaneously, we get:

$a = 27.9048 \qquad b = 5.75 \qquad c = 0.5595$

$\widehat{Y} = 27.9048 + 5.75x + 0.5595x^2$

d) i) $\widehat{Y} = 30.1429 + 5.75(6) = 64.6$

ii) $\widehat{Y} = 27.9048 + 5.75(6) + 0.5595(36) = 82.6$

15-14

Year	x	Y	xY	x^2	x^2Y	x^4
1989	−7	82.4	−576.8	49	4037.6	2401
1990	−5	125.7	−628.5	25	3142.5	625
1991	−3	276.9	−830.7	9	2492.1	81
1992	−1	342.5	−342.5	1	342.5	1
1993	1	543.6	543.6	1	543.6	1
1994	3	691.5	2074.5	9	6223.5	81
1995	5	782.4	3912.0	25	19560.0	625
1996	7	889.5	6226.5	49	43585.5	2401
	0	3734.5	10378.1	168	79927.3	6216

a) $a = \overline{Y} = \frac{3734.5}{8} = 466.8125 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{10378.1}{168} = 61.7744$

$\widehat{Y} = 466.8125 + 61.7744x$ (where 1992.5 = 0 and x units = .5 year)

b) Equations 15.6 and 15.7 become:

$$\sum Y = na + c\sum x^2 \qquad 3734.5 = 8a + 168c$$

$$\sum x^2 Y = a\sum x^2 + c\sum x^4 \qquad 79927.3 = 168a + 6216c$$

Solving these simultaneously, we get:

$a = 455.0719, \ c = 0.5591$

$\widehat{Y} = 455.0719 + 61.7744x + 0.5591x^2$

c) Linear forecast: $\widehat{Y} = 466.8125 + 61.7744(11) = 1146.33$ thousand mice

Quadratic forecast: $\widehat{Y} = 455.0719 + 61.7744(11) + 0.5591(121) = 1202.24$ thousand mice

15-15 a)

Year	x	Y	xY	x^2
1991	−2	4620	−9240	4
1992	−1	4910	−4910	1
1993	0	5490	0	0
1994	1	5730	5730	1
1995	2	5990	11980	4
	0	26740	3560	10

$a = \overline{Y} = \frac{27640}{5} = 5348 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{3560}{10} = 356$

$\widehat{Y} = 5348 + 356x$ (where 1993 = 0 and x units = one year)

b) $\widehat{Y} = 5348 + 356(3) = 6416$ or \$6,416,000

15-16

Year	x	Y	xY	x^2	x^2Y	x^4
1968	−7	5	−35	49	245	2401
1970	−6	5	−30	36	180	1296
1972	−5	8	−40	25	200	625
1974	−4	8	−32	16	128	256
1976	−3	10	−30	9	90	81
1978	−2	13	−26	4	52	16
1980	−1	15	−15	1	15	1
1982	0	18	0	0	0	0
1984	1	20	20	1	20	1
1986	2	22	44	4	88	16
1988	3	25	75	9	225	81
1990	4	25	100	16	400	256
1992	5	29	145	25	725	625
1994	6	29	174	36	1044	1296
1996	7	32	224	49	1568	2401
	0	264	574	280	4980	9352

a) $a = \overline{Y} = \frac{264}{15} = 17.6$ $\qquad b = \frac{\sum xY}{\sum x^2} = \frac{574}{280} = 2.05$

$\widehat{Y} = 17.6 + 2.05x$ (where 1982 = 0 and x units = two years)

b) Equations 15.6 and 15.7 become:

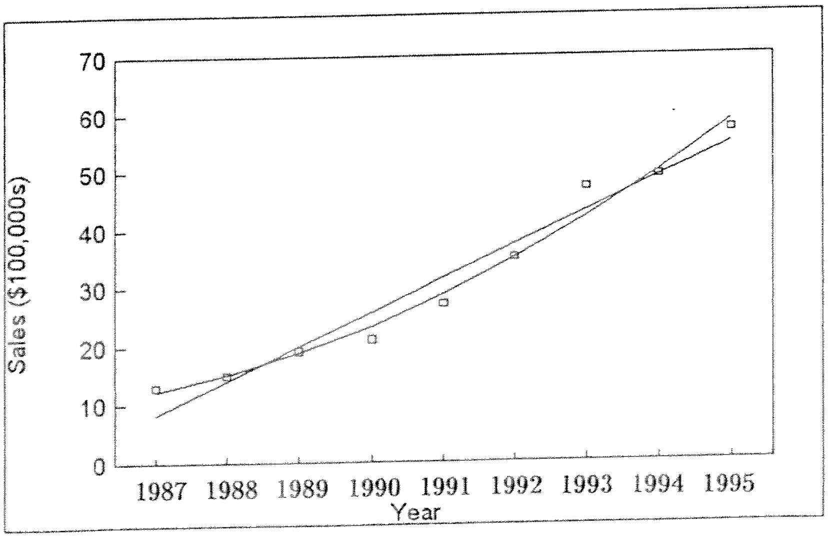
$$\sum Y = na + c\sum x^2 \qquad 264 = 15a + 280c$$
$$\sum x^2 Y = a\sum x^2 + c\sum x^4 \qquad 4980 = 280a + 9352c$$

Solving these simultaneously, we get:

$$a = 17.3647,\ c = 0.0126$$
$$\widehat{Y} = 17.3647 + 2.0500x + 0.0126x^2$$

c) Political resistance to increased rates makes it unlikely that the quadratic trend would continue to be a good predictor.

15-17 a)



b)

Year	x	Y	xY	x^2	x^2Y	x^4
1987	−4	13	−52	16	208	256
1988	−3	15	−45	9	135	81
1989	−2	19	−38	4	76	16
1990	−1	21	−21	1	21	1
1991	0	27	0	0	0	0
1992	1	35	35	1	35	1
1993	2	47	94	4	188	16
1994	3	49	147	9	441	81
1995	4	57	228	16	912	256
	0	283	348	60	2016	708

$a = \overline{Y} = \frac{283}{9} = 31.4444$ $\qquad b = \frac{\sum xY}{\sum x^2} = \frac{348}{60} = 5.8000$

$\widehat{Y} = 31.4444 + 5.8000x$ (where 1991 = 0 and x units = one year)

c) Equations 15.6 and 15.7 become:

$$\sum Y = na + c\sum x^2 \qquad 283 = 9a + 60c$$

$$\sum x^2 Y = a\sum x^2 + c\sum x^4 \qquad 2016 = 60a + 708c$$

Solving these simultaneously, we get:

$a = 28.6451$ $\qquad b = 5.8000$ $\qquad c = 0.4199$

$\widehat{Y} = 28.6451 + 5.8000x + 0.4199x^2$

d) It appears to favor (c).

15-18 a) Since the rate of increase in the pollution rating is itself increasing, a second degree trend would fit the data better than a linear trend.

b) However, as the air gets more polluted and citizens get more concerned, actions will be taken to control pollution, so the predictions of the second-degree trend will in all likelihood be too dire.

c) Since public or political action will likely reduce pollution, neither a linear nor a second-degree estimating equation will be accurate.

15-19 a)

Year	x	Y	xY	x^2
1987	−4	175	−700	16
1988	−3	190	−570	9
1989	−2	185	−370	4
1990	−1	195	−195	1
1991	0	180	0	0
1992	1	200	200	1
1993	2	185	370	4
1994	3	190	570	9
1995	4	205	820	16
	0	1705	125	60

$a = \overline{Y} = \frac{1705}{9} = 189.4444$ $\qquad b = \frac{\sum xY}{\sum x^2} = \frac{125}{60} = 2.0833$

$\widehat{Y} = 189.4444 + 2.0833x$ (where 1991 = 0 and x units = one year)

b) In 1996, $\widehat{Y} = 189.4444 + 2.0833(5) = 200$ fatalities.

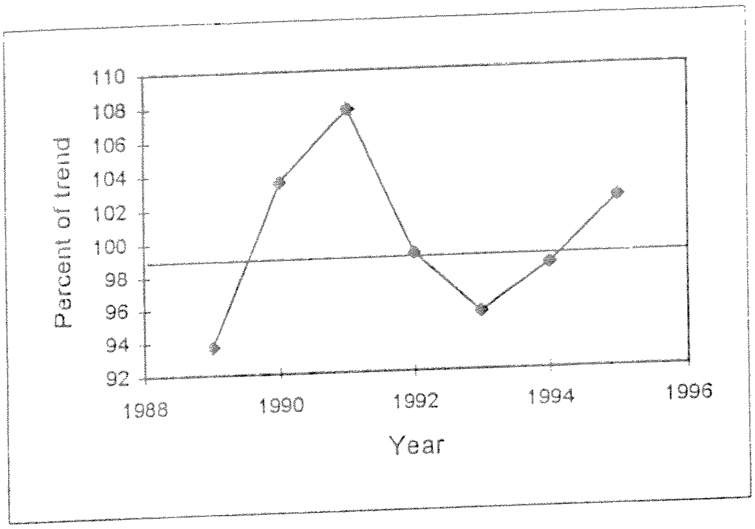
15-20 a)

Year	1989	1990	1991	1992	1993	1994	1995
Y	1.100	1.500	1.900	2.100	2.400	2.900	3.500
$\widehat{Y}$	1.174	1.449	1.764	2.119	2.514	2.949	3.424
$100\,Y/\widehat{Y}$	93.70	103.52	107.71	99.100	95.47	98.34	102.22

b)

$100(Y/\widehat{Y}-1)$	−6.30	3.52	7.71	−0.90	−4.53	1.66	2.22

c)



d) Largest fluctuation (by both methods) was in 1991.

15-21

Year	x	Y	xY	x^2	$\widehat{Y}$	$100(Y/\widehat{Y})$	$100(Y/\widehat{Y}-1)$
1990	−3	14.8	−44.4	9	13.832	107.00	7.00
1991	−2	20.7	−41.4	4	20.179	102.58	2.58
1992	−1	24.6	−24.6	1	26.525	92.74	−7.26
1993	0	32.9	0	0	32.871	100.09	0.09
1994	1	37.8	37.8	1	39.218	96.38	−3.62
1995	2	47.6	95.2	4	45.564	104.47	4.47
1996	3	51.7	155.1	9	51.911	99.59	−0.41
	0	230.1	177.7	28			

a) $a = \overline{Y} = \dfrac{230.1}{7} = 32.8714$ $\qquad b = \dfrac{\sum xY}{\sum x^2} = \dfrac{177.7}{28} = 6.3464$

$\widehat{Y} = 32.8714 + 6.4364x$ (where 1993 = 0 and x units = 1 year)

b) See the next-to-the-last column above.

c) See the last column above.

d) Largest fluctuation (by both methods) was in 1992.

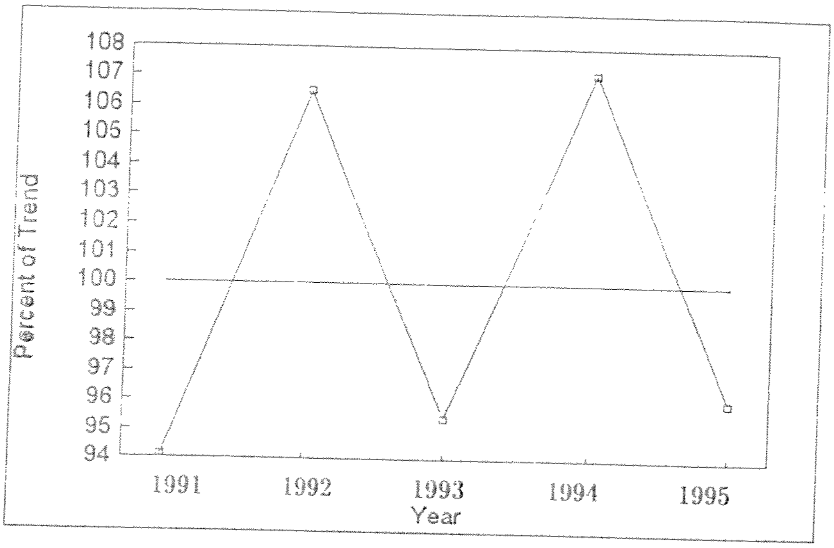
15-22 a)

Year	1991	1992	1993	1994	1995
Y	32.00	46.00	50.00	66.00	68.00
$\widehat{Y}$	34.00	43.20	52.40	61.60	70.80
$100\,Y/\widehat{Y}$	94.12	106.48	95.42	107.14	96.05

b)

$100(Y/\widehat{Y}-1)$	−5.88	6.48	−4.58	7.14	−3.95

c)



d) Largest fluctuation (by both methods) was in 1994.

15-23 a)

Year	1989	1990	1991	1992	1993	1994	1995
Y	2.2	2.1	2.4	2.6	2.7	2.9	2.8
$\widehat{Y} = 2.53 + .13x$	2.14	2.27	2.40	2.53	2.66	2.79	2.92
$100(Y/\widehat{Y})$	102.80	92.51	100.00	102.77	101.50	103.94	95.89

b) $100(Y/\widehat{Y}-1)$ 2.80 −7.49 0.00 2.77 1.50 3.94 −4.11

c) Largest fluctuation (by both methods) is in 1990.

d) It means that considerably less effort will be needed in 1990 to arrange for funds to finance the corporation's activities.

15-24

Year	x	Y	xY	x^2	$\widehat{Y}$	$100(Y/\widehat{Y})$	$100(Y/\widehat{Y}-1)$
1989	−3	21.0	−63.0	9	21.2643	98.76	−1.24
1990	−2	19.4	−38.8	4	22.3000	87.00	−13.00
1991	−1	22.6	−22.6	1	23.3357	96.85	−3.15
1992	0	28.2	0.0	0	24.3714	115.71	15.71
1993	1	30.4	30.4	1	25.4071	119.65	19.65
1994	2	24.0	48.0	4	26.4428	90.76	−9.24
1995	3	25.0	75.0	9	27.4785	90.98	−9.02
	0	170.6	29.0	28			

a) $a = \overline{Y} = \frac{170.6}{7} = 24.3714$ $\qquad b = \frac{\sum xY}{\sum x^2} = \frac{29}{28} = 1.0357$

$\widehat{Y} = 24.3714 + 1.0357x$ (where 1992 = 0 and x units = 1 year)

b) See the next-to-the-last column above.

c) See the last column above.

d) Largest fluctuation (by both methods) was in 1993.

15-25 a)

Year	x	Y	xY	x^2	$\hat{Y}$	$100(Y/\hat{Y})$	$100(Y/\hat{Y}-1)$
1991	−2	3.5	−7.0	4	3.84	91.15	−8.85
1992	−1	4.2	−4.2	1	3.82	109.95	9.95
1993	0	3.9	0.0	0	3.80	102.63	2.63
1994	1	3.8	3.8	1	3.78	100.53	0.53
1995	2	3.6	7.2	4	3.76	95.74	−4.26
	0	19.0	−0.2	10			

$a = \overline{Y} = \frac{19}{5} = 3.8$ $\qquad b = \frac{\sum xY}{\sum x^2} = \frac{-0.2}{10} = -0.02$

$\hat{Y} = 3.8 - 0.02x$ (where 1993 = 0 and x units = 1 year)

b) See the next-to-the-last column above.

c) See the last column above.

d) The data and both measures of cyclical variation show that Wombat Airlines appears to be carrying fewer and fewer passengers each year. We really don't have enough information to be able to decide whether Wombat is in the trough of a cycle or if their long run secular trend is decreasing.

15-26 a, b)

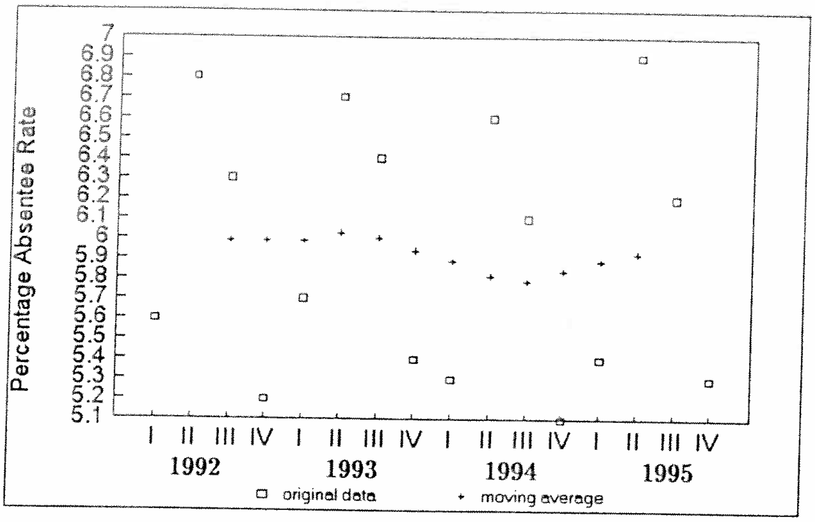
Year	Quarter	Actual Receivables	4–Quarter Moving Average	Centered Moving Average	Ratio of Actual to CMA
1991	Spring	102			
	Summer	120	97.50		
	Fall	90	99.50	98.500	0.9137
	Winter	78	101.00	100.250	0.7781
1992	Spring	110	102.25	101.625	1.0824
	Summer	126	103.50	102.875	1.2248
	Fall	95	103.75	103.625	0.9168
	Winter	83	104.25	104.000	0.7981
1993	Spring	111	104.75	104.500	1.0622
	Summer	128	105.50	105.125	1.2176
	Fall	97	106.50	106.000	0.9151
	Winter	86	108.25	107.375	0.8009
1994	Spring	115	109.75	109.000	1.0550
	Summer	135	111.00	110.375	1.2231
	Fall	103	112.75	111.875	0.9207
	Winter	91	115.00	113.875	0.7991
1995	Spring	122	116.75	115.875	1.0529
	Summer	144	118.50	117.625	1.2242
	Fall	110			
	Winter	98			

c)

Year	Spring		Summer		Fall		Winter	
1991					.9137		.7781	
1992	1.0824		1.2248		.9168		.7981	
1993	1.0622		1.2176		.9151		.8009	
1994	1.0550		1.2231		.9207		.7991	
1995	1.0529		1.2242					
Modified sum	2.1172		2.4473		1.8319		1.5972	
Modified mean	1.0586	+	1.2236	+	.9159	+	.7986	= 3.9967/4 = .999175
Seasonal index	1.0595		1.2246		.9167		.7993	

15-27 a)

Year	Quarter	Actual Absentee Rate	4–Quarter Moving Average	Centered Moving Average
1992	Spring	5.6		
	Summer	6.8		
			5.98	
	Fall	6.3		5.988
			6.00	
	Winter	5.2		5.988
			5.98	
1993	Spring	5.7		5.988
			6.00	
	Summer	6.7		6.025
			6.05	
	Fall	6.4		6.000
			5.95	
	Winter	5.4		5.938
			5.93	
1994	Spring	5.3		5.888
			5.85	
	Summer	6.6		5.813
			5.78	
	Fall	6.1		5.788
			5.80	
	Winter	5.1		5.838
			5.88	
1995	Spring	5.4		5.888
			5.90	
	Summer	6.9		5.925
			5.95	
	Fall	6.2		
	Winter	5.3		



b) In addition to the seasonal pattern (low in spring and winter, high in summer and fall), the moving average shows what looks like a cyclical pattern with a three-year period.

15-28

Year	Baseball		Football		Basketball		Hockey
1992	96		128		116		77
1993	92		131		126		69
1994	94		113		117		84
1995	97		118		126		89
1996	91		121		124		81
Modified sum	279		367		366		242
Modified mean	93.00	+	122.33	+	122.00	+	80.67 = 418.00/4 = 104.5
Seasonal index	89.00		117.06		116.75		77.20

15-29

Year	Spring	Summer	Fall	Winter
1990	108	128	94	70
1991	112	132	88	68
1992	109	134	84	73
1993	110	131	90	69
1994	108	135	89	68
1995	106	129	93	72
Modified sum	435	526	360	279
Modified mean	108.75 +	131.50 +	90.00 +	69.75 = 400

Thus, we need not adjust any further.

These indices show a pattern similar to that shown by those in problem 15-26, but the extreme values (in summer and winter) are even further away from 100 in this problem.

15-30 a, b)

Year	Quarter	Actual Enrollment	4-Quarter Moving Average	Centered Moving Average	Percentage of Actual to Moving Average
1991	Fall	220			
	Winter	203	175.00		
	Spring	193	178.75	176.875	109.117
	Summer	84	180.00	179.375	46.829
1992	Fall	235	183.25	181.625	129.387
	Winter	208	181.25	182.250	114.129
	Spring	206	181.50	181.375	113.577
	Summer	76	181.00	181.250	41.931
1993	Fall	236	181.75	181.375	130.117
	Winter	206	181.00	181.375	113.577
	Spring	209	182.25	181.625	115.072
	Summer	73	184.50	183.375	39.809
1994	Fall	241	183.75	184.125	130.889
	Winter	215	188.50	186.125	115.514
	Spring	206	188.00	188.250	109.429
	Summer	92	189.50	188.750	48.742
1995	Fall	239	191.25	190.375	125.542
	Winter	221	197.00	194.125	113.844
	Spring	213			
	Summer	115			

c)

Year	Fall	Winter	Spring	Summer
1991			109.117	46.829
1992	129.387	114.128	113.577	41.931
1993	130.117	113.577	115.072	39.809
1994	130.889	115.514	109.429	48.742
1995	125.542	113.844		
Modified sum	259.504	227.973	223.006	88.760
Modified mean	129.752 +	113.987 +	111.503 +	44.380 = 399.622/400 = .99905
Seasonal index	129.875	114.095	111.609	44.422

15-31

Year	Quarter	Actual Customers	4–Quarter Moving Average	Centered Moving Average	Ratio of Actual to CMA
1991	Spring	200			
	Summer	300	237.50		
	Fall	125	231.25	234.375	0.5333
	Winter	325	218.75	225.000	1.4444
1992	Spring	175	225.00	221.875	0.7887
	Summer	250	237.50	231.250	1.0811
	Fall	150	250.00	243.750	0.6154
	Winter	375	262.50	256.250	1.4634
1993	Spring	225	275.00	268.750	0.8372
	Summer	300	293.75	284.375	1.0549
	Fall	200	287.50	290.625	0.6882
	Winter	450	300.00	293.750	1.5319
1994	Spring	200	306.25	303.125	0.6598
	Summer	350	287.50	296.875	1.1789
	Fall	225	281.25	284.375	0.7912
	Winter	375	268.75	275.000	1.3636
1995	Spring	175	262.50	265.625	0.6588
	Summer	300	256.25	259.375	1.1566
	Fall	200			
	Winter	350			

Year	Spring	Summer	Fall	Winter
1991			0.5333	1.4444
1992	0.7887	1.0811	0.6154	1.4634
1993	0.8372	1.0549	0.6882	1.5319
1994	0.6598	1.1789	0.7912	1.3636
1995	0.6588	1.1566		
Modified sum	1.4485	2.2377	1.3036	2.9078
Modified mean	0.7243	+ 1.1189	+ 0.6518	+ 1.4539 = 3.9488/4 = 0.9872
Seasonal index	0.7336	1.1334	0.6603	1.4728

15-32 a)

Year	Quarter	Actual Starts	4–Quarter Moving Average	Centered Moving Average	Ratio of Actual to CMA
1991	Spring	8			
	Summer	10	7.50		
	Fall	7	7.75	7.625	0.9180
	Winter	5	7.75	7.750	0.6452
1992	Spring	9	7.75	7.750	1.1613
	Summer	10	8.00	7.875	1.2698
	Fall	7	8.25	8.125	0.8615
	Winter	6	8.50	8.375	0.7164
1993	Spring	10	8.50	8.500	1.1765
	Summer	11	8.50	8.500	1.2941
	Fall	7	8.50	8.500	0.8235
	Winter	6	8.75	8.625	0.6957
1994	Spring	10	9.00	8.875	1.1268
	Summer	12	9.25	9.125	1.3151
	Fall	8	9.50	9.375	0.8533
	Winter	7	9.75	9.625	0.7273

1995	Spring	11	10.00	9.875	1.1139
	Summer	13	10.25	10.125	1.2840
	Fall	9			
	Winter	8			

Year	Spring		Summer		Fall		Winter	
1991					0.9180		0.6452	
1992	1.1613		1.2698		0.8615		0.7164	
1993	1.1765		1.2941		0.8235		0.6957	
1994	1.1268		1.3151		0.8533		0.7273	
1995	1.1139		1.2840					
Modified sum	2.2881		2.5781		1.7148		1.4121	
Modified mean	1.1441	+	1.2891	+	0.8574	+	0.7061	= 3.9967/4 = 0.999175
Seasonal index	1.1450		1.2902		0.8581		0.7067	

b) $\frac{.7067}{1.2902} = .5477$, so his working capital need falls by 45.23% from summer to winter.

15-33 Since there can be no mathematical model for irregular variations, we have no way of predicting future irregular variations.

15-34 c and d

15-35 A stalled car at an intersection; an unannounced quiz; a snow storm in late April; a mail strike; a stopped clock.

15-36 The fact that these irregular variations even themselves out over time and the fact that they are frequently minor in magnitude enable management to live with them.

15-37 a) NOTE: Although numbers in this solution are displayed with at most 4 decimal places, all computations were done in an Excel spreadsheet, without rounding intermediate results.

Year	Quarter	Data (Contaminants)	4–Quarter Moving Average	Centered Moving Average	Ratio of Usage to CMA	Seasonal Index	Deseasonalized Data
1992	Winter	452				1.1764	384.2090
	Spring	385	388.00			0.9765	394.2584
	Summer	330	393.50	390.750	0.8445	0.8775	376.0651
	Fall	385	396.50	395.000	0.9747	0.9695	397.0986
1993	Winter	474	403.00	399.750	1.1857	1.1764	402.9094
	Spring	397	406.50	404.750	0.9809	0.9765	406.5470
	Summer	356	411.50	409.000	0.8704	0.8775	405.6945
	Fall	399	414.50	413.000	0.9661	0.9695	411.5386
1994	Winter	494	419.25	416.875	1.1850	1.1764	419.9098
	Spring	409	423.25	421.250	0.9709	0.9765	418.8356
	Summer	375	426.25	424.750	0.8829	0.8775	427.3467
	Fall	415	431.25	428.750	0.9679	0.9695	428.0414
1995	Winter	506	437.00	434.125	1.1656	1.1764	430.1101
	Spring	429	442.50	439.750	0.9756	0.9765	439.3165
	Summer	398	447.75	445.125	0.8941	0.8775	453.5573
	Fall	437	454.00	450.875	0.9692	0.9695	450.7327

1996	Winter	527	459.75	456.875	1.1535	1.1764	447.9605
	Spring	454	471.00	465.375	0.9756	0.9765	464.9177
	Summer	421				0.8775	479.7679
	Fall	482				0.9695	497.1468

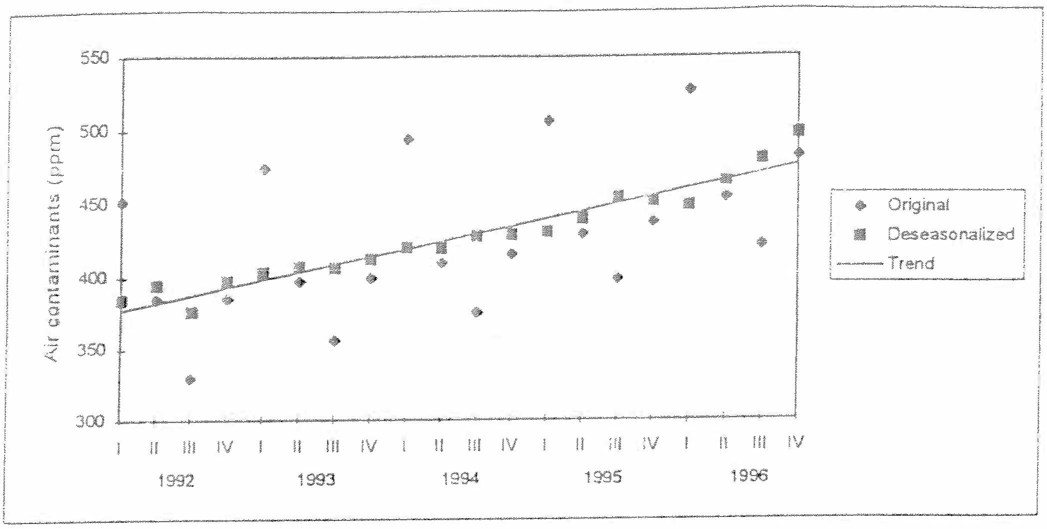
Year	Winter	Spring	Summer	Fall	
1992			0.8445	0.9747	
1993	1.1857	0.9808	0.8704	0.9661	
1994	1.1850	0.9709	0.8829	0.9679	
1995	1.1656	0.9756	0.8941	0.9692	
1996	1.1535	0.9756			
Modified sum	2.3506	1.9511	1.7533	1.9372	
Modified mean	1.1753 +	0.9756 +	0.8766 +	0.9686	= 3.9661/4 = 0.99902
Seasonal index	1.1764	0.9765	0.8775	0.9695	

b,c)

Year	Quarter	Deseason-alized Data (Y)	x	xY	x^2	Deseason-alized trend $\widehat{Y}$=426.7982 +2.5954x	Relative Cyclical Residual $100(Y/\widehat{Y}-1)$
1992	Winter	384.2090	−19	−7299.9709	361	377.4858	1.78
	Spring	394.2584	−17	−6702.3930	289	382.6765	3.03
	Summer	376.0651	−15	−4640.9768	225	387.8673	−3.04
	Fall	397.0986	−13	−5162.2820	169	393.0581	1.03
1993	Winter	402.9094	−11	−4432.0038	121	398.2489	1.17
	Spring	406.5470	−9	−3658.9229	81	403.4397	0.77
	Summer	405.6945	−7	−2839.8615	49	408.6304	−0.72
	Fall	411.5386	−5	−2057.6928	25	413.8212	−0.55
1994	Winter	419.9098	−3	−1259.7295	9	419.0120	0.21
	Spring	418.8356	−1	−418.8356	1	424.2028	−1.27
	Summer	427.3467	1	427.3467	1	429.3936	−0.48
	Fall	428.0414	3	1284.1241	9	434.5844	−1.51
1995	Winter	430.1101	5	2150.5503	25	439.7751	−2.20
	Spring	439.3165	7	3075.2156	49	444.9659	−1.27
	Summer	453.5573	9	4082.0160	81	450.1567	0.76
	Fall	450.7327	11	4958.0599	121	455.3475	−1.01
1996	Winter	447.9605	13	5823.4863	169	460.5383	−2.73
	Spring	464.9177	15	6973.7657	225	465.7291	−0.17
	Summer	479.7679	17	8156.0548	289	470.9198	1.88
	Fall	497.1468	19	9445.7899	361	476.1106	4.42
		8535.9637	0	6903.7406	2660		

$$a = \overline{Y} = \frac{8535.9637}{20} = 426.7982 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{6903.7406}{2660} = 2.5954$$

$\widehat{Y} = 426.7982 + 2.5954x$ (where 1994-II 1/2 = 0 and x units = 1/2 quarter) d)



15-38 a,b)

Year	Qtr	Actual Sales	4–Quarter Moving Average	Centered Moving Average	Percentage of Sales to CMA	Seasonal Index	Deseason-alized Sales
1991	I	19				75.886	25.038
	II	24	26.50			105.081	22.840
	III	38	27.00	26.750	142.056	142.050	26.751
	IV	25	28.00	27.500	90.909	76.984	32.474
1992	I	21	29.50	28.750	73.043	75.886	27.673
	II	28	29.00	29.250	95.726	105.081	26.646
	III	44	29.50	29.250	150.427	142.050	30.975
	IV	23	30.25	29.875	76.987	76.984	29.876
1993	I	23	29.50	29.875	76.987	75.886	30.309
	II	31	29.50	29.500	105.085	105.081	29.501
	III	41	29.75	29.625	138.397	142.050	28.863
	IV	23	30.75	30.250	76.033	76.984	29.876
1994	I	24	32.50	31.625	75.889	75.886	31.626
	II	35	32.00	32.250	108.527	105.081	33.308
	III	48				142.050	33.791
	IV	21				76.984	27.278

Year	I		II		III		IV	
1991					142.056		90.909	
1992	73.043		95.726		150.427		76.987	
1993	76.987		105.085		138.397		76.033	
1994	75.889		108.527					
Modified sum	75.889	+	105.085	+	142.056	+	76.987	= 400.017/4 = 100.004
Seasonal index	75.886		105.081		142.050		76.984	

15-39 a)

Year	Quarter	Deseasonalized Sales (Y)	x	xY	x^2
1991	I	25.038	−15	−375.570	225
	II	22.840	−13	−296.920	169
	III	26.751	−11	−294.261	121
	IV	32.474	−9	−292.266	81

1992	I	27.673	−7	−193.711	49
	II	26.646	−5	−133.230	25
	III	30.975	−3	−92.925	9
	IV	29.876	−1	−29.876	1
1993	I	30.309	1	30.309	1
	II	29.501	3	88.503	9
	III	28.863	5	144.315	25
	IV	29.876	7	209.132	49
1994	I	31.626	9	284.634	81
	II	33.308	11	366.388	121
	III	33.791	13	439.283	169
	IV	27.278	15	409.170	225
		466.825	0	262.975	1360

$a = \overline{Y} = \dfrac{466.825}{16} = 29.1766$ $\qquad b = \dfrac{\sum xY}{\sum x^2} = \dfrac{262.975}{1360} = .1934$

$\widehat{Y} = 29.1766 + .1934x$ (where 1992-IV 1/2 = 0 and x units = 1/2 quarter)

b)

Year	Quarter	Deseasonalized Sales (Y)	Deseasonalized Forecast ($\widehat{Y}$)	Percent of Trend $100(Y/\widehat{Y})$
1991	I	25.038	26.276	95.288
	II	22.840	26.662	85.665
	III	26.751	27.049	98.898
	IV	32.474	27.436	118.363
1992	I	27.673	27.823	99.461
	II	26.646	28.210	95.456
	III	30.975	28.596	108.319
	IV	29.876	28.983	103.081
1993	I	30.309	29.370	103.197
	II	29.501	29.757	99.140
	III	28.863	30.144	95.750
	IV	29.876	30.530	97.858
1994	I	31.626	30.917	102.293
	II	33.308	31.304	106.402
	III	33.791	31.691	106.626
	IV	27.278	32.078	85.036

15-40 A large irregular component; a change in weather produces a larger or smaller than expected seasonal index; a change in technology which affects the secular trend; an economic change which alters the time scale of the cyclical component.

15-41 Since we cannot expect the future to be exactly like the past, perhaps the only assurance we need about historical data is that they be accurate, reflective of what actually happened, and represent the best data available.

15-42 The decline in birth rates which has occurred will no doubt affect future college enrollments; we need be especially careful about the behavior in birth rates seventeen to eighteen years in the past when estimating college enrollments.

15-43 a) Changes in the federal tax laws are an irregular component and cannot really be forecast. There is some reason to believe that an astute political observer can anticipate the general direction of such changes for the near term.

b) Changing an inventory method, for example, from FIFO to LIFO, produces an effect on profits which can be calculated; the effects of this change on past profits can be calculated and adjustments in the time series made accordingly.

15-44 a,b)

Year	Quarter	Actual Admissions	4-Quarter Moving Average	Centered Moving Average	Ratio of Actual to CMA	Seasonal Index	Deseasonalized Admissions
1992	Spring	29				0.7851	36.9380
	Summer	30	35.75			0.8888	33.7534
	Fall	41	35.25	35.500	1.1549	1.1351	36.1202
	Winter	43	36.25	35.750	1.2028	1.1909	36.1071
1993	Spring	27	37.25	36.750	0.7347	0.7851	34.3905
	Summer	34	38.50	37.875	0.8977	0.8888	38.2538
	Fall	45	40.00	39.250	1.1465	1.1351	39.6441
	Winter	48	40.50	40.250	1.1925	1.1909	40.3057
1994	Spring	33	40.75	40.625	0.8123	0.7851	42.0329
	Summer	36	41.50	41.125	0.8754	0.8888	40.5041
	Fall	46	41.75	41.625	1.1051	1.1351	40.5251
	Winter	51	42.75	42.250	1.2071	1.1909	42.8248
1995	Spring	34	43.00	42.875	0.7930	0.7851	43.3066
	Summer	40	43.50	43.250	0.9249	0.8888	45.0045
	Fall	47				1.1351	41.4060
	Winter	53				1.1909	44.5042

Year	Spring	Summer	Fall	Winter
1992			1.1549	1.2028
1993	0.7347	0.8977	1.1465	1.1925
1994	0.8123	0.8754	1.1051	1.2071
1995	0.7930	0.9249		
Modified sum	0.7930 +	0.8977 +	1.1465 +	1.2028 = 4.0400/4 = 1.01
Seasonal index	0.7851	0.8888	1.1351	1.1909

c)

Year	Quarter	Deseasonalized Admissions (Y)	x	xY	x^2
1992	Spring	36.9380	−15	−554.0695	225
	Summer	33.7534	−13	−438.7939	169
	Fall	36.1202	−11	−397.3218	121
	Winter	36.1071	−9	−324.9643	81
1993	Spring	34.3905	−7	−240.7337	49
	Summer	38.2538	−5	−191.2691	25
	Fall	39.6441	−3	−118.9323	9
	Winter	40.3057	−1	−40.3057	1
1994	Spring	42.0329	1	42.0329	1
	Summer	40.5041	3	121.5122	9
	Fall	40.5251	5	202.6253	25
	Winter	42.8248	7	299.7733	49
1995	Spring	43.3066	9	389.7593	81
	Summer	45.0045	11	495.0495	121
	Fall	41.4060	13	538.2786	169
	Winter	44.5042	15	667.5623	225
		635.6207	0	450.2030	1360

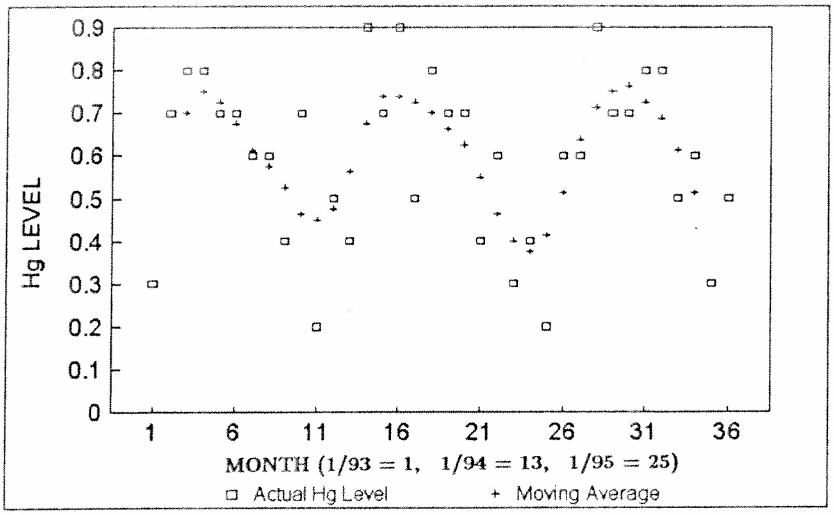
$$a = \overline{Y} = \frac{635.6207}{16} = 39.7263 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{450.2030}{1360} = 0.3310$$

$\widehat{Y} = 39.7263 + 0.3310x$ (where 1990-IV 1/2 = 0 and x units = 1/2 quarter)

15-45 $595{,}000 \times \frac{128}{100} = 761{,}600$ passengers

15-46

Year	Month	Actual Hg Level	4–Month Moving Average	Centered Moving Average
1993	Jan	0.3		
	Feb	0.7	0.650	
	March	0.8	0.750	0.7000
	April	0.8	0.750	0.7500
	May	0.7	0.700	0.7250
	June	0.7	0.650	0.6750
	July	0.6	0.575	0.6125
	Aug	0.6	0.575	0.5750
	Sept	0.4	0.475	0.5250
	Oct	0.7	0.450	0.4625
	Nov	0.2	0.450	0.4500
	Dec	0.5	0.500	0.4750
1994	Jan	0.4	0.625	0.5625
	Feb	0.9	0.725	0.6750
	March	0.7	0.750	0.7375
	April	0.9	0.725	0.7375
	May	0.5	0.725	0.7250
	June	0.8	0.675	0.7000
1994	July	0.7	0.650	0.6625
	Aug	0.7	0.600	0.6250
	Sept	0.4	0.500	0.5500
	Oct	0.6	0.425	0.4625
	Nov	0.3	0.375	0.4000
	Dec	0.4	0.375	0.3750
1995	Jan	0.2	0.450	0.4125
	Feb	0.6	0.575	0.5125
	March	0.6	0.700	0.6375
	April	0.9	0.725	0.7125
	May	0.7	0.775	0.7500
	June	0.7	0.750	0.7625
	July	0.8	0.700	0.7250
	Aug	0.8	0.675	0.6875
	Sept	0.5	0.550	0.6125
	Oct	0.6	0.475	0.5125
	Nov	0.3		
	Dec	0.5		



15-47 a,b)

Year	Quarter	Actual Pounds	4–Quarter Moving Average	Centered Moving Average	%-age of Actual to CMA	Seasonal Index	Deseason-alized Pounds
1992	Winter	3.1				74.7040	4.1497
	Spring	5.1	4.350			114.4814	4.4549
	Summer	5.6	4.400	4.3750	128.0000	128.9066	4.3442
	Fall	3.6	4.400	4.4000	81.8182	81.9080	4.3952
1993	Winter	3.3	4.450	4.4250	74.5763	74.7040	4.4174
	Spring	5.1	4.475	4.4625	114.2857	114.4814	4.4549
	Summer	5.8	4.500	4.4875	129.2479	128.9066	4.4994
	Fall	3.7		4.5250	81.7680	81.9080	4.5173

			4.550				
1994	Winter	3.4	4.600	4.5750	74.3169	74.7040	4.5513
	Spring	5.3	4.625	4.6125	114.9051	114.4814	4.6296
	Summer	6.0	4.700	4.6625	128.6863	128.9066	4.6545
	Fall	3.8	4.725	4.7125	80.6366	81.9080	4.6394
1995	Winter	3.7	4.750	4.7375	78.1003	74.7040	4.9529
	Spring	5.4	4.775	4.7625	113.3858	114.4814	4.7169
	Summer	6.1				128.9066	4.7321
	Fall	3.9				81.9080	4.7614

Year	Winter		Spring		Summer		Fall	
1992					128.0000		81.8182	
1993	74.5763		114.2857		129.2479		81.7680	
1994	74.3169		114.9051		128.6863		80.6366	
1995	78.1003		113.3858					
Modified sum	74.5763	+	114.2857	+	128.6863	+	81.7680	$= 399.3163/4 = 99.82907$
Seasonal index	74.7040		114.4814		128.9066		81.9080	

c)

Year	Quarter	Actual Pounds	Deseasonalized Pounds (Y)	x	xY	x^2
1992	Winter	3.1	4.1497	−15	−62.2455	225
	Spring	5.1	4.4549	−13	−57.9137	169
	Summer	5.6	4.3442	−11	−47.7862	121
	Fall	3.6	4.3952	−9	−39.5568	81
1993	Winter	3.3	4.4174	−7	−30.9218	49
	Spring	5.1	4.4549	−5	−22.2745	25
	Summer	5.8	4.4994	−3	−13.4982	9
	Fall	3.7	4.5173	−1	−4.5173	1
1994	Winter	3.4	4.5513	1	4.5513	1
	Spring	5.3	4.6296	3	13.8888	9
	Summer	6.0	4.6545	5	23.2725	25
	Fall	3.8	4.6394	7	32.4758	49
1995	Winter	3.7	4.9529	9	44.5761	81
	Spring	5.4	4.7169	11	51.8859	121
	Summer	6.1	4.7321	13	61.5173	169
	Fall	3.9	4.7614	15	71.4210	225
			72.8711	0	24.8747	1360

$$a = \overline{Y} = \frac{72.8711}{16} = 4.5544 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{24.8747}{1360} = .0183$$

$\widehat{Y} = 4.5544 + .0183x$ (where 1993-IV 1/2 = 0 and x units = 1/2 quarter)

d) In spring 1996, $x = 19$, so $\widehat{Y} = 4.5544 + .0183(19) = 4.9021$ (about 4.9 million pounds).

15-48 a) Gasoline mileage is affected by such things as government responses to the 1973 oil embargo and the resultant mandated fleet mileage standards.

b) This series is almost entirely irregular variation, because commercial aviation fatalities occur in random batches as the result of unpredictable airplane crashes.

c) Although total world demand has a long-run increasing trend, there are so many grain growers that each one's exports does not grow smoothly over time but depends instead on political and economic conditions in both importing and exporting nations.

d) In addition to seasonalities resulting from higher usages in the summer months, gasoline prices are also greatly affected by unpredictable geopolitical events.

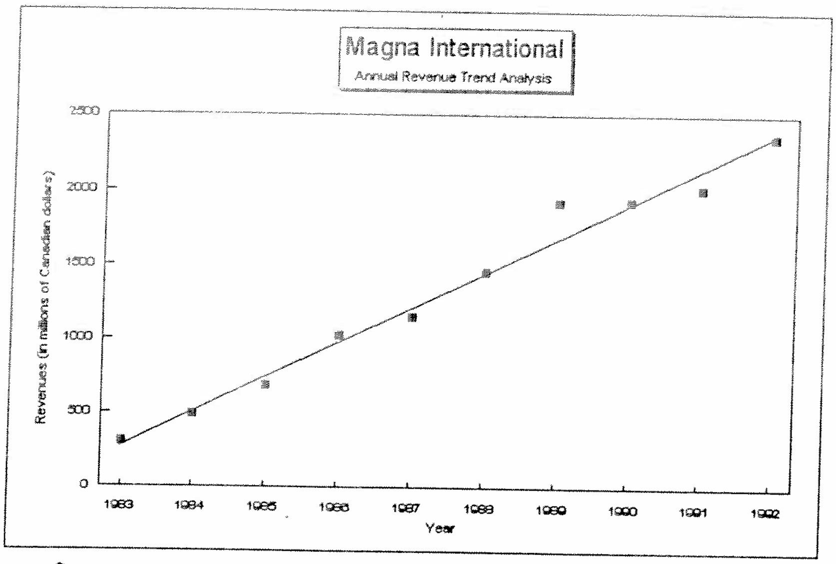
15-49 a)

Year	x	Y	xY	x^2
1983	−9	302.5	−2722.5	81
1984	−7	493.6	−3455.2	49
1985	−5	690.4	−3452.0	25
1986	−3	1027.8	−3083.4	9
1987	−1	1152.5	−1152.5	1
1988	1	1458.6	1458.6	1
1989	3	1923.7	5771.1	9
1990	5	1927.7	9636.0	25
1991	7	2017.2	14120.4	49
1992	9	2358.8	21229.2	81
	0	13352.3	38349.7	330

$$a = \overline{Y} = \frac{13352.3}{10} = 1335.23 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{38349.7}{330} = 116.2112$$

$\widehat{Y} = 1335.23 + 116.2112x$ (where 1987.5 = 0 and x units = .5 year)

b)



c) $\widehat{Y} = 1364.202 + 116.2112x - 0.8779x^2$

The coefficient c (-0.8779) is not significantly different from zero ($p = 0.4753$), and does not explain much variance in annual revenues (partial $R^2 = 0.0014$). In contrast, coefficient b is significant ($p = 0.0000$) and explains most of the variance in annual revenues (partial $R^2 = 0.9809$). Therefore, the least squares trend line (from 15-49a) is the better equation.

d) For 1993, $x = 11$.

$\widehat{Y} = 1335.23 + 116.2112(11)$

$\widehat{Y} = \$2613.5532$ (in millions of Canadian dollars)

15-50 a) Although sales of PC's have been growing at increasing rates, this growth cannot be sustained as even larger fractions of the population eventually come to own PC's. Because of this, a second-degree predicting equation will soon tend to overestimate the sales of PC's.

b) Here, too, a forecast based on a second-degree predicting equation will tend to be an overestimate, because of the saturation phenomenon mentioned in (a) and also because kids will tend to play with them less as the novelty of the games wears off.

c) As more states act to place caps on giving awards for damages in medical malpractice cases, the amounts paid for such claims will cease its current rapid growth. As insurance companies' liabilities stop growing so rapidly, so will the premiums they charge. Once again, second-degree forecasts will tend to be overestimated as a result.

d) Here is another instance of a growth rate that cannot be sustained which will lead to overestimates from a second-degree predicting equation.

15-51 a)

Year	x	Y	xY	x^2
1990	−5	596	−2980	25
1991	−3	688	−2064	9
1992	−1	740	−740	1
1993	1	812	812	1
1994	3	857	2571	9
1995	5	935	4675	25
	0	4628	2274	70

$$a = \overline{Y} = \frac{4628}{6} = 771.3333$$

$$b = \frac{\sum xY}{\sum x^2} = \frac{2274}{70} = 32.4857$$

$\widehat{Y} = 771.3333 + 32.4857x$ (where 1992.5 = 0 and x units = .5 year)

b) In 1997, $\widehat{Y} = 771.3333 + 32.4857(9) = 1064$ manuals.

15-52 a)

Year	Qtr	Actual Exports	4–Quarter Moving Average	Centered Moving Average	Percentage of Actual to CMA	Seasonal Index	Deseasonalized Exports
1992	I	1				43.343	2.307
	II	3	3.50			68.730	4.365
	III	6	3.75	3.625	165.517	173.374	3.461
	IV	4	3.50	3.625	110.345	114.551	3.492
1993	I	2	3.75	3.625	55.172	43.343	4.614
	II	2	4.00	3.875	51.613	68.730	2.910
	III	7	4.00	4.000	175.000	173.374	4.037
	IV	5	4.50	4.250	117.647	114.551	4.365
1994	I	2	4.75	4.625	43.243	43.343	4.614
	II	4	4.75	4.750	84.211	68.730	5.820
	III	8	4.50	4.625	172.973	173.374	4.614
	IV	5	4.25	4.375	114.286	114.551	4.365
1995	I	1	4.25	4.250	23.529	43.343	2.307
	II	3	4.50	4.375	68.571	68.730	4.365
	III	8				173.374	4.614
	IV	6				114.551	5.238

Year	I		II		III		IV	
1992					165.517		110.345	
1993	55.172		51.613		175.000		117.647	
1994	43.243		84.211		172.973		114.286	
1995	23.529		68.571					
Modified sum	43.243	+	68.571	+	172.973	+	114.286	= 399.073/4 = 99.76825
Seasonal index	43.343		68.730		173.375		114.551	

b,c)

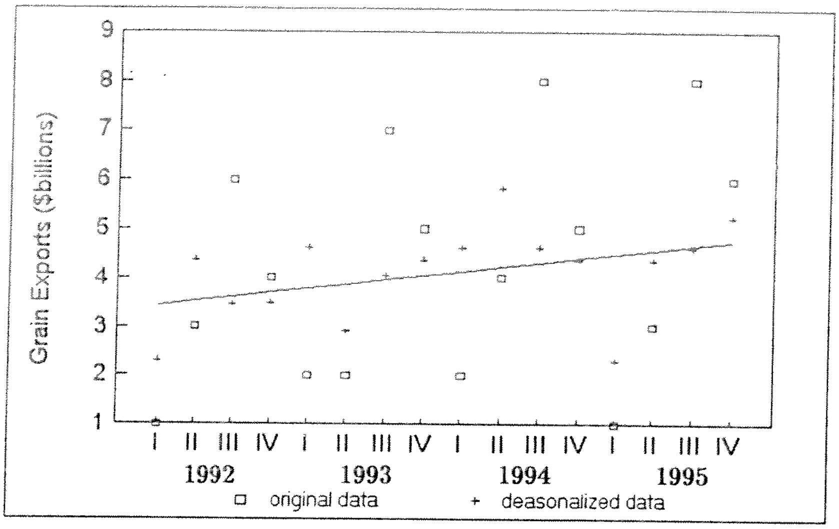
Year	Qtr	Deseasonalized Exports(Y)	x	xY	x^2	Deseasonalized trend $\widehat{Y}$=4.0930 +0.0433x	Relative Cyclical Residual 100($Y/\widehat{Y}$−1)
1992	I	2.307	−15	−34.605	225	3.444	−33.014
	II	4.365	−13	−56.745	169	3.530	23.654
	III	3.461	−11	−38.071	121	3.617	−4.313
	IV	3.492	−9	−31.428	81	3.703	−5.698
1993	I	4.614	−7	−32.298	49	3.790	21.741
	II	2.910	−5	−14.550	25	3.877	−24.942
	III	4.037	−3	−12.111	9	3.963	1.867
	IV	4.365	−1	−4.365	1	4.050	7.778

1994	I	4.614	1	4.614	1	4.136	11.557
	II	5.820	3	17.460	9	4.223	37.817
	III	4.614	5	23.070	25	4.310	7.053
	IV	4.365	7	30.555	49	4.396	−0.705
1995	I	2.307	9	20.763	81	4.483	−48.539
	II	4.365	11	48.015	121	4.569	−4.465
	III	4.614	13	59.982	169	4.656	−0.902
	IV	5.238	15	78.570	225	4.743	10.436
		65.488	0	58.856	1360		

$$a = \overline{Y} = \frac{65.488}{16} = 4.0930 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{58.856}{1360} = .0433$$

$\widehat{Y} = 4.0930 + .0433x$ (where 1993-IV 1/2 = 0 and x units = 1/2 quarter)

d)



15-53 $165 \times \frac{143}{100} = 235.95$, or 236 bicycles

15-54 Since such a major source of demand for heavy earth-moving equipment is going to be lost, historic trends in sales of such equipment will be poor predictors of future sales. The manufacturers would be better advised to abandon a times series forecasting model for an econometric model which includes such explanatory variables as the number of miles of roads currently under construction and scheduled for the next few years, the age of the current stocks of earth-moving equipment, etc.

15-55 a) NOTE: Although numbers in this solution are displayed with at most 4 decimal places, all computations were done in an Excel spreadsheet, without rounding intermediate results.

World Data:

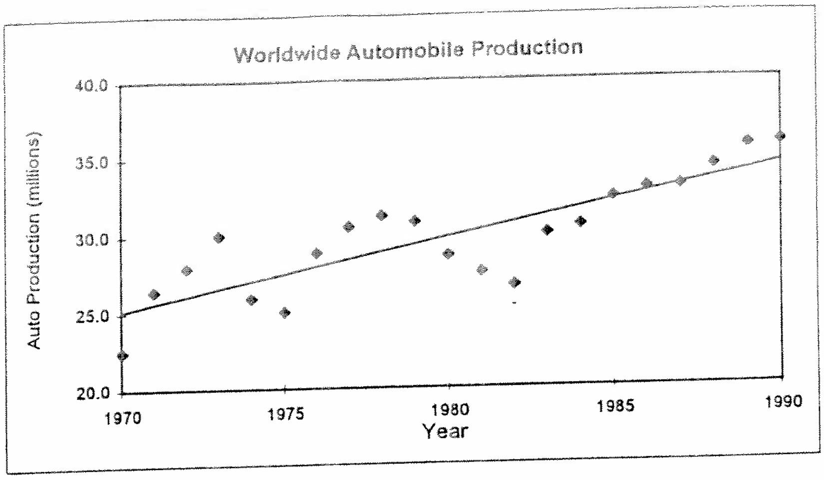
Year	x	Y	xY	x^2	$\widehat{Y}$	$100(Y/\widehat{Y}-1)$
1970	−10	22.5	−225.0	100	25.1481	−10.5298
1971	−9	26.4	237.6	81	25.6147	3.0659
1972	−8	27.9	−223.2	64	26.0813	6.9732
1973	−7	30.0	−210.0	49	26.5479	13.0032
1974	−6	25.9	−155.4	36	27.0145	−4.1257
1975	−5	25.0	−125.0	25	27.4812	−9.0286
1976	−4	28.8	−115.2	16	27.9478	3.0493
1977	−3	30.5	−91.5	9	28.4144	7.3399

1978	−2	31.2	−62.4	4	28.8810	8.0294
1979	−1	30.8	−30.8	1	29.3477	4.9487
1980	0	28.6	0.0	0	29.8143	−4.0728
1981	1	27.5	27.5	1	30.2809	−9.1837
1982	2	26.6	53.2	4	30.7475	−13.4890
1983	3	30.0	90.0	9	31.2142	−3.8898
1984	4	30.5	122.0	16	31.6808	−3.7271
1985	5	32.3	161.5	25	32.1474	0.4747
1986	6	32.9	197.4	36	32.6140	0.8768
1987	7	33.0	231.0	49	33.0806	−0.2438
1988	8	34.3	274.4	64	33.5473	2.2438
1989	9	35.6	320.4	81	34.0139	4.6631
1990	10	35.8	358.0	100	34.4805	3.8267
$\sum =$	0	626.1	359.3	770		

$a = \overline{Y} = 626.1/21 = 29.8143$ $\qquad b = \dfrac{\sum xY}{\sum x^2} = 359.3/770 = 0.4666$

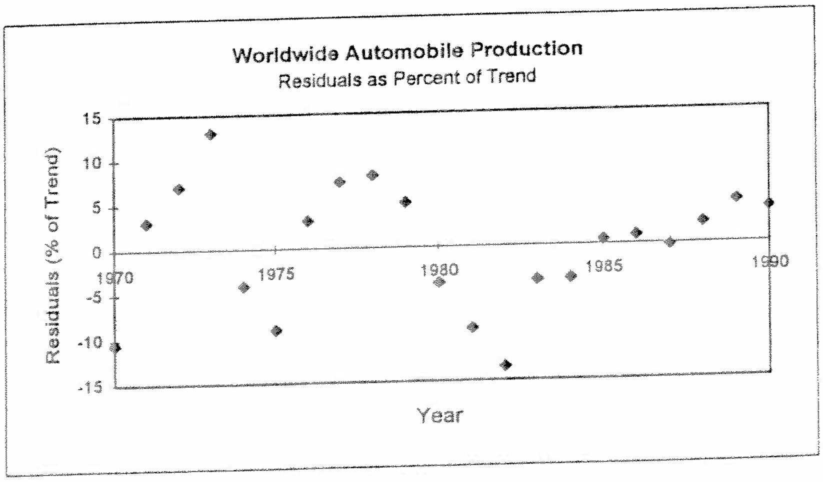
$\widehat{Y} = 29.8143 + 0.4666x$

b)



The variations from the trend appear to be cyclical.

c)



The length of the cycle is approximately 5 to 6 years, but seems to be increasing.

d) U.S.S.R. Data:

Year	x	Y	xY	x^2	$\widehat{Y}$	$100(Y/\widehat{Y}-1)$
1970	−10	3.4	−34.0	100	8.3446	−59.2550
1971	−9	5.3	−47.7	81	8.6682	−38.8571
1972	−8	7.3	−58.4	64	8.9919	−18.8155
1973	−7	9.2	−64.4	49	9.3155	−1.2398
1974	−6	11.2	−67.2	36	9.6391	16.1930
1975	−5	12.0	−60.0	25	9.9628	20.4484
1976	−4	12.4	−49.6	16	10.2864	20.5474
1977	−3	12.8	−38.4	9	10.6100	20.6404
1978	−2	13.1	−26.2	4	10.9337	19.8133
1979	−1	13.1	−13.1	1	11.2573	16.3688
1980	0	13.3	0.0	0	11.5810	14.8438
1981	1	13.2	13.2	1	11.9046	10.8816
1982	2	13.1	26.2	4	12.2282	7.1292
1983	3	13.2	39.6	9	12.5519	5.1637
1984	4	13.3	53.2	16	12.8755	3.2970
1985	5	13.3	66.5	25	13.1991	0.7642
1986	6	13.3	79.8	36	13.5228	−1.6474
1987	7	13.3	93.1	49	13.8464	−3.9462
1988	8	12.6	100.8	64	14.1700	−11.0800
1989	9	12.2	109.8	81	14.4937	−15.8254
1990	10	12.6	126.0	100	14.8173	−14.9644
$\sum =$	0	243.2	249.2	770		

$a = \overline{Y} = 243.2/21 = 11.5810$ $\qquad b = \dfrac{\sum xY}{\sum x^2} = 249.2/770 = 0.3236$

$\widehat{Y} = 11.5810 + 0.3236x$

Automobile production both worldwide and in the former U.S.S.R. were growing. However, production worldwide was marked by frequent cycles, whereas U.S.S.R. production was more steady. This difference is attributable to the variability of free-market economies subject to the laws of supply and demand. The steady growth to, and stabilization around, approximately 13 millions units per year in the U.S.S.R. reflects centralized control over production.

15-56 a)

Year	x	Y	xY	x^2
1988	−7	12	−84	49
1989	−5	11	−55	25
1990	−3	19	−57	9
1991	−1	17	−17	1
1992	1	19	19	1
1993	3	18	54	9
1994	5	20	100	25
1995	7	23	161	49
	0	139	121	168

$a = \overline{Y} = \dfrac{139}{8} = 17.3750$ $\qquad b = \dfrac{\sum xY}{\sum x^2} = \dfrac{121}{168} = 0.7202$

$\widehat{Y} = 17.3750 + 0.7202x$ (where $1991.5 = 0$ and x units $= .5$ year)

b) In 1999 $\widehat{Y} = 17.3750 + 0.7202(15) = 28.2$, or about 28 completions.

c) He should be very careful about predicting so far in advance because of the many things that can change in the home-building business in the meantime.

15-57 a)

	Spring	Summer	Fall	Winter
Raw data	31,000	52,000	39,000	29,000
Seasonal index	84	134	103	79
Deseasonalized data	36,905	38,806	37,864	36,709

b) In winter there are only about 79% as many homicides and assaults as there are in an average season.

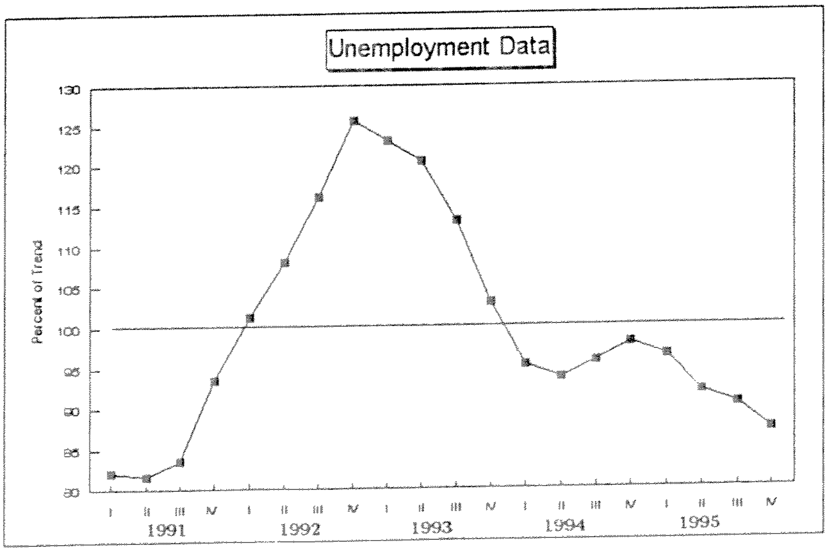
15-58 a,b)

Year	Qtr	Deseason-alized Unemploy-ment(Y)	x	xY	x^2	Deseason-alized trend $\widehat{Y}$=8.175 $-0.0379x$	Percent of trend $100(Y/\widehat{Y})$
1991	I	7.3	−19	−138.7	361	8.8951	82.1
	II	7.2	−17	−122.4	289	8.8193	81.6
	III	7.3	−15	−109.5	225	8.7435	83.5
	IV	8.1	−13	−105.3	169	8.6677	93.5
1992	I	8.7	−11	−95.7	121	8.5919	101.3
	II	9.2	−9	−82.8	81	8.5161	108.0
	III	9.8	−7	−68.6	49	8.4403	116.1
	IV	10.5	−5	−52.5	25	8.3645	125.5
1993	I	10.2	−3	−30.6	9	8.2887	123.1
	II	9.9	−1	−9.9	1	8.2129	120.5
	III	9.2	1	9.2	1	8.1371	113.1
	IV	8.3	3	24.9	9	8.0613	103.0
1994	I	7.6	5	38.0	25	7.9855	95.2
	II	7.4	7	51.8	49	7.9097	93.6
	III	7.5	9	67.5	81	7.8339	95.7
	IV	7.6	11	83.6	121	7.7581	98.0
1995	I	7.4	13	96.2	169	7.6823	96.3
	II	7.0	15	105.0	225	7.6065	92.0
	III	6.8	17	115.6	289	7.5307	90.3
	IV	6.5	19	123.5	361	7.4549	87.2
		163.5	0	−100.7	2660		

$$a = \overline{Y} = \frac{163.5}{20} = 8.175 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{-100.7}{2660} = -0.0379$$

$\widehat{Y} = 8.175 - 0.0379x$ (where 1993 - II 1/2 = 0 and x units = 1/2 quarter)

c)



15-59

Year	AIDS Cases (Y)	x	xY	x^2	x^2Y	x^4
1988	2	−2	−4	4	9	16
1989	4	−1	−4	1	4	1
1990	7	0	0	0	0	0
1991	13	1	13	1	13	1
1992	21	2	42	4	84	16
	47	0	47	10	109	34

a) $a = \overline{Y} = \frac{47}{5} = 9.4$ $\qquad$ $b = \frac{\sum xY}{\sum x^2} = \frac{47}{10} = 4.7$

$\widehat{Y} = 9.4 + 4.7x$ (where 1990 = 0 and x units = 1 year)

b) Equations 15.6 and 15.7 become:

$\sum Y = na + c\sum x^2$ $\qquad$ $47 = 5a + 10c$

$\sum x^2 Y = a\sum x^2 + c\sum x^4$ $\qquad$ $109 = 10a + 34c$

Solving these simultaneously, we get:

$a = 7.2571$ $\qquad$ $c = 1.0714$

$\widehat{Y} = 7.2571 + 4.7x + 1.0714x^2$

c)

Year	AIDS Cases (Y)	Linear $\widehat{Y}$	2nd degree $\widehat{Y}$
1988	2	0	2.1427
1989	4	4.7	3.6285
1990	7	9.4	7.2571
1991	13	14.1	13.0285
1992	21	18.8	20.9427

d) The second degree curve is the better estimator.

15-60 a)

Week	Sales (Y)	x	xY	x^2
1	41	−3	−123	9
2	52	−2	−104	4
3	79	−1	−79	1
4	76	0	0	0
5	72	1	72	1
6	59	2	118	4
7	41	3	123	9
	420	0	7	28

$a = \overline{Y} = \frac{420}{7} = 60$ $\qquad$ $b = \frac{\sum xY}{\sum x^2} = \frac{7}{28} = 0.25$

$\widehat{Y} = 60 + 0.25x$

b) For week 8, $\widehat{Y} = 60 + 0.25(4) = 61$

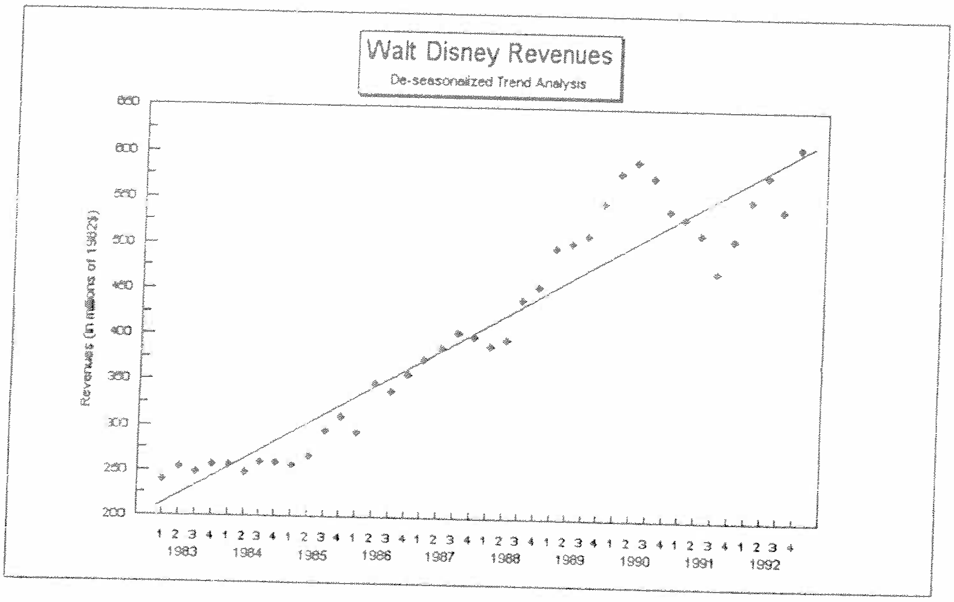
c) The data suggest that sales of the broiled whole chickens have peaked and are now declining, so a 2nd-degree curve would have have been better. The linear regression line predicts sales will continue to increase.

15-61 a)

Year	Quarter	1982 \$ Revenue (Y)	x	xY	x^2
1983	1	200.3	−20	−4006.0	400
	2	233.9	−19	−4444.1	361
	3	279.7	−18	−5034.6	324
	4	286.8	−17	−4875.6	289
1984	1	213.4	−16	−3414.4	256
	2	229.4	−15	−3441.0	225
	3	293.2	−14	−4104.8	196
	4	289.8	−13	−3767.4	169
1985	1	213.4	−12	−2560.8	144
	2	246.1	−11	−2707.1	121
	3	333.5	−10	−3335.0	100
	4	346.9	−9	−3122.1	81
1986	1	244.3	−8	−1954.4	64
	2	320.5	−7	−2243.5	49
	3	383.4	−6	−2300.4	36
	4	397.6	−5	−1988.0	25
1987	1	311.9	−4	−1247.6	16
	2	357.6	−3	−1072.8	9
	3	456.4	−2	−912.8	4
	4	446.1	−1	−446.1	1
1988	1	325.2	1	325.2	1
	2	367.4	2	734.8	4
	3	497.4	3	1492.2	9
	4	507.3	4	2029.2	16
1989	1	414.9	5	2074.5	25
	2	465.9	6	2795.4	36
	3	578.2	7	4047.4	49
	4	611.3	8	4890.4	64
1990	1	484.4	9	4359.6	81
	2	547.6	10	5476.0	100
	3	651.1	11	7162.1	121
	4	602.8	12	7233.6	144
1991	1	444.0	13	5772.0	169
	2	475.9	14	6662.6	196
	3	535.1	15	8026.5	225
	4	568.3	16	9092.8	256
1992	1	460.8	17	7833.6	289
	2	534.9	18	9628.2	324
	3	611.6	19	11620.4	361
	4	680.0	20	13600.0	400
		16448.3	0	57878.0	5740

$a = \overline{Y} = \dfrac{16448.3}{40} = 411.2075$ $\qquad b = \dfrac{\sum xY}{\sum x^2} = \dfrac{57878.0}{5740} = 10.0833$

$\widehat{Y} = 411.2075 + 10.0833x$ (where 1987-IV 1/2 = 0 and x units = 1/2 quarter)



$R^2 = 0.80$. Therefore, 80% of the variation in revenue is explained by the trend alone.

b)

Year	Quarter	Actual Revenue	4–Quarter Moving Average	Centered Moving Average	Ratio of Actual to CMA	Seasonal Index	Deseason-alized Revenue
1983	1	200.3				83.385	240.211
	2	233.9	250.175			92.377	253.201
	3	279.7	253.45	251.813	111.075	112.817	247.923
	4	286.8	252.325	252.888	113.410	111.420	257.404
1984	1	213.4	255.7	254.013	84.012	83.385	255.921
	2	229.4	256.45	256.075	89.583	92.377	248.330
	3	293.2	256.45	256.45	114.330	112.817	259.889
	4	289.8	260.625	258.538	112.092	111.420	260.096
1985	1	213.4	270.7	265.663	80.327	83.385	255.921
	2	246.1	284.975	277.838	88.577	92.377	266.408
	3	333.5	292.7	288.838	115.463	112.817	295.610
	4	346.9	311.3	302.0	114.868	111.420	311.343
1986	1	244.3	323.775	317.538	76.936	83.385	292.978
	2	320.5	336.45	330.113	97.088	92.377	346.947
	3	383.4	353.35	344.9	111.163	112.817	339.841
	4	397.6	362.625	357.988	111.065	111.420	356.847
1987	1	311.9	380.875	371.75	83.900	83.385	374.048
	2	357.6	393.0	386.938	92.418	92.377	387.108
	3	456.4	396.325	394.663	115.643	112.817	404.548
	4	446.1	398.775	397.55	112.212	111.420	400.376
1988	1	325.2	409.025	403.9	80.515	83.385	389.998
	2	367.4	424.325	416.675	88.174	92.377	397.717
	3	497.4	446.75	435.538	114.204	112.817	440.889
	4	507.3	471.375	459.063	110.508	111.420	455.303
1989	1	414.9	491.575	481.475	86.173	83.385	497.571
	2	465.9	517.575	504.575	92.335	92.377	504.345
	3	578.2	534.95	526.263	109.869	112.817	512.510
	4	611.3	555.375	545.163	112.132	111.420	548.643

1990	1	484.4	573.6	564.488	85.812	83.385	580.919
	2	547.6	571.475	572.538	95.644	92.377	592.787
	3	651.1	561.375	566.425	114.949	112.817	577.127
	4	602.8	543.45	552.413	109.121	111.420	541.014
1991	1	444.0	514.45	528.95	83.940	83.385	532.469
	2	475.9	505.825	510.138	93.289	92.377	515.170
	3	535.3	510.025	507.925	105.350	112.817	474.306
	4	568.2	524.775	517.4	109.838	111.420	510.050
1992	1	460.6	543.9	534.338	86.238	83.385	552.617
	2	535.0	571.825	557.863	95.884	92.377	579.039
	3	611.6				112.817	542.115
	4	680.0				111.420	610.301

Year	1	2	3	4
1983			111.075	113.410
1984	84.012	89.583	114.330	112.092
1985	80.327	88.577	115.463	114.868
1986	76.936	97.088	111.163	111.065
1987	83.900	92.418	115.643	112.212
1988	80.515	88.174	114.204	110.508
1989	86.173	92.335	109.869	112.132
1990	85.812	95.644	114.949	109.121
1991	83.940	93.289	105.350	109.121
1992	86.238	95.884		
Modified mean	83.526	92.533	113.008	111.608 = 400.674/4 = 100.1685
Seasonal Index	83.385	92.377	112.817	111.420

c)

Year	Quarter	Deseasonalized Revenue (Y)	x	xY	x^2
1983	1	240.211	−20	−4804.22	400
	2	253.201	−19	−4810.82	361
	3	247.923	−18	−4462.61	324
	4	257.404	−17	−4375.86	289
1984	1	255.921	−16	−4094.74	256
	2	248.330	−15	−3724.94	225
	3	259.889	−14	−3638.45	196
	4	260.096	−13	−3381.25	169
1985	1	255.921	−12	−3071.05	144
	2	266.408	−11	−2930.48	121
	3	295.610	−10	−2956.10	100
	4	311.343	−9	−2802.09	81
1986	1	292.978	−8	−2343.82	64
	2	346.947	−7	−2428.63	49
	3	339.841	−6	−2039.05	36
	4	356.847	−5	−1784.23	25
1987	1	374.048	−4	−1496.19	16
	2	387.108	−3	−1161.33	9
	3	404.548	−2	−809.095	4
	4	400.376	−1	−400.375	1
1988	1	389.998	1	389.998	1
	2	397.717	2	795.434	4
	3	440.889	3	1322.668	9
	4	455.303	4	1821.21	16

1989	1	497.571	5	2487.854	25
	2	504.345	6	3026.071	36
	3	512.510	7	3587.567	49
	4	548.643	8	4389.142	64
1990	1	580.919	9	5228.272	81
	2	592.787	10	5927.868	100
	3	577.127	11	6348.40	121
	4	541.014	12	6492.167	144
1991	1	532.469	13	6922.099	169
	2	515.170	14	7212.384	196
	3	474.306	15	7114.594	225
	4	510.050	16	8160.802	256
1992	1	552.617	17	9394.483	289
	2	579.039	18	10422.70	324
	3	542.115	19	10300.16	361
	4	610.301	20	12206.02	400
		16409.84	0	56034.6	5740

$a = \overline{Y} = \frac{16409.84}{40} = 410.2459$ $\qquad b = \frac{\sum xY}{\sum x^2} = \frac{56034.6}{5740} = 9.7621$

$\widehat{Y} = 410.246 + 9.7621x$ (where 1987-IV 1/2 = 0 and x units = 1/2 quarter)

d)

Year	Quarter	Actual Revenue (Y)	Deseasonalized trend: $\widehat{Y}=$ 410.246 + 9.7621x	Seasonal Index	Reseasonalized Estimates ($\widehat{Y}$*SI)/100	(Error) Reseason. Est. $-Y$	SSE
1983	1	200.3	215.003	83.385	179.281	− 21.019	441.806
	2	233.9	224.766	92.377	207.632	− 26.268	690.000
	3	279.7	234.528	112.817	264.588	− 15.112	228.373
	4	286.8	244.290	111.420	272.189	− 14.611	213.491
1984	1	213.4	254.052	83.385	211.842	− 1.559	2.429
	2	229.4	263.814	92.377	243.704	14.304	204.607
	3	293.2	273.576	112.817	308.642	15.442	238.440
	4	289.8	283.338	111.420	315.697	25.897	670.638
1985	1	213.4	293.100	83.385	244.402	31.002	961.130
	2	246.1	302.863	92.377	279.776	33.676	1134.073
	3	333.5	312.625	112.817	352.695	19.195	368.448
	4	346.9	322.387	111.420	359.205	12.305	151.405
1986	1	244.3	332.149	83.385	276.963	32.663	1066.854
	2	320.5	341.911	92.377	315.848	− 4.652	21.642
	3	383.4	351.673	112.817	396.749	13.349	178.183
	4	397.6	361.435	111.420	402.713	5.113	26.140
1987	1	311.9	371.197	83.385	309.523	− 2.377	5.648
	2	357.6	380.960	92.377	351.920	− 5.680	32.264
	3	456.4	390.722	112.817	440.802	− 15.598	243.297
	4	446.1	400.484	111.420	446.221	0.121	0.015
1988	1	325.2	420.008	83.385	350.224	25.024	626.208
	2	367.4	429.770	92.377	397.010	29.610	876.736
	3	497.4	439.532	112.817	495.869	− 1.531	2.344
	4	507.3	449.294	111.420	500.606	− 6.694	44.814
1989	1	414.9	459.075	83.385	382.785	− 32.115	1031.387
	2	465.9	468.819	92.377	433.082	− 32.818	1077.044
	3	578.2	478.581	112.817	539.922	− 38.278	1465.176
	4	611.3	488.343	111.420	544.114	− 67.186	4514.001

1990	1	484.4	498.105	83.385	415.345	– 69.055	4768.535
	2	547.6	507.867	92.377	469.154	– 78.446	6153.841
	3	651.1	517.629	112.817	583.976	– 67.124	4505.646
	4	602.8	527.391	111.420	587.622	– 15.178	230.381
1991	1	444.0	537.154	83.385	447.906	3.906	15.257
	2	475.9	546.916	92.377	505.226	29.326	859.985
	3	535.1	556.678	112.817	628.029	92.929	8635.872
	4	568.3	566.440	111.420	631.130	62.830	3947.570
1992	1	460.8	576.202	83.385	480.467	19.667	386.779
	2	534.9	585.964	92.377	541.297	6.397	40.927
	3	611.6	595.726	112.817	672.083	60.483	3658.181
	4	680.0	605.488	111.420	674.638	– 5.362	28.754
							$\sum = 49748.32$

From computer output: SST = 729718.4

$1 - \frac{SSE}{SST} = 1 - \frac{49748.32}{729718.4} = 0.9318 \Rightarrow$ Therefore, 93.18% of the variation in revenue is explained by the trend and seasonality, which accounts for 13.18% more of the variation than the trend alone.

e) Irregular variation

f) Based on the re-seasonalized forecasts and the actual revenues, it appears that Disney lost \$188.9907 (in millions of 1982 dollars) during fiscal year 1991, perhaps as a result of the Persian Gulf war and recession.

g) For the quarters in fiscal year 1993, $x = 21, 22, 23, 24$

$\widehat{Y} = 410.246 + 9.7621(21) = 615.2501 \quad 615.2501 \times 83.3851/100 = 513.0269$

$\widehat{Y} = 410.246 + 9.7621(22) = 625.0122 \quad 625.0122 \times 92.3772/100 = 577.3688$

$\widehat{Y} = 410.246 + 9.7621(23) = 634.7743 \quad 634.7743 \times 112.8174/100 = 716.1359$

$\widehat{Y} = 410.246 + 9.7621(24) = 644.5364 \quad 644.5364 \times 111.4204/100 = 718.1450$

Total TPR revenue for 1993 = \$2524.6766

This estimate may be in error due to the fact that the most recent theme park, Euro-Disney has not generated the numbers of visitors that they had expected. Since it has been losing money, revenues should be less than expected.

h) In order to convert the estimates in (g) into actual dollars, we would need the GNP deflator statistic for 1993.

15-62 a) Multiplying through by (SEASONAL INDICES)/100, we get the actual passenger counts:

	Spring	Summer	Fall	Winter
1994	652.30	397.85	689.30	598.00
1995	704.00	408.80	678.00	577.20

b) Summer saw the fewest passengers; spring saw the most.

c) For the 1996 fall season, $x = 13$ and

$\widehat{Y} = 584.75 - .45(13) = 578.9$ (deseasonalized)

so the actual predicted ridership is 578.9(113/100) = 654.157, or about 654,000 riders.

15-63 a,b)

Year	Quarter	Actual Attend.	Moving Average	Percentage of Actual to MA	Seasonal Index	Deseason-alized Attend.
1992	Spring	750			91.0433	823.7838
	Summer	1150	860.0000	133.7209	135.3110	849.8939
	Fall	680	870.0000	78.1609	73.6458	923.3385
1993	Spring	780	853.3333	91.4063	91.0433	856.7352
	Summer	1100	820.0000	134.1463	135.3110	812.9420
	Fall	580	826.6667	70.1613	73.6458	787.5534
1994	Spring	800	868.3333	92.1305	91.0433	878.7028
	Summer	1225	878.3333	139.4687	135.3110	905.3218
	Fall	610	825.0000	73.9394	73.6458	828.2889
1995	Spring	640	766.6667	83.4783	91.0433	702.9622
	Summer	1050	763.3333	137.5546	135.3110	775.9901
	Fall	600			73.6458	814.7104

Year	Spring		Summer		Fall		
1992			133.7209		78.1609		
1993	91.4063		134.1463		70.1613		
1994	92.1305		139.4687		73.9394		
1995	83.4783		137.5546				
Modified sum	91.4063		271.7009		73.9394		
Modified mean	91.4063	+	135.8505	+	73.9394	=	301.1962/3 = 100.39873
Seasonal index	91.0433		135.3110		73.6458		

15-64

Week	Day	Actual Customers	Moving Average	Percentage of Actual to MA
1	MON	345		
	TUE	310		
	WED	385		
	THU	416	487.4286	85.3458
	FRI	597	497.8571	119.9139
	SAT	706	501.1429	140.8780
	SUN	653	503.2857	129.7474
2	MON	418	517.4286	80.7841
	TUE	333	527.0000	63.1879
	WED	400	534.8571	74.7863
	THU	515	541.8571	95.0435
	FRI	664	538.2857	123.3546
	SAT	761	546.0000	139.3773
	SUN	702	533.2857	131.6368
3	MON	393	536.1429	73.3014
	TUE	387	530.5714	72.9402
	WED	311	523.4286	59.4159
	THU	535	508.5714	105.1966
	FRI	625	510.4286	122.4461
	SAT	711	514.0000	138.3268
	SUN	598	523.4286	114.2467

4	MON	406	510.4286	79.5410
	TUE	412	514.0000	80.1556
	WED	377	527.1429	71.5176
	THU	444	559.1429	79.4072
	FRI	650		
	SAT	803		
	SUN	822		

Week	MON	TUE	WED	THU	FRI	SAT	SUN
1				85.3458	119.9139	140.8780	129.7474
2	80.7841	63.1879	74.7863	95.0435	123.3546	139.3773	131.6368
3	73.3014	72.9402	59.4159	105.1966	122.4461	138.3268	114.2467
4	79.5410	80.1556	71.5176	79.4072			
Modified sum	79.5410	72.9402	71.5176	180.3893	122.4461	139.3773	129.7474
Modified mean	79.5410 +	72.9402 +	71.5176 +	90.1947 +	122.4461 +	139.3773 +	129.7474
Seasonal index	78.8914	72.3445	70.9335	89.4580	121.4460	138.2389	128.6877

(Sum of modified means = 705.7643/7 = 100.82347)

15-65 a)

Year	Sales (Y)	x	xY	x^2	x^2Y	x^4
1991	230	−2	−460	4	920	16
1992	250	−1	−250	1	250	1
1993	265	0	0	0	0	0
1994	300	1	300	1	300	1
1995	210	2	620	4	1240	16
	1355	0	210	10	2710	34

$$b = \frac{\sum xY}{\sum x^2} = \frac{210}{10} = 21$$

Equations 15.6 and 15.7 become:

$$\sum Y = na + c\sum x^2 \qquad 1355 = 5a + 10c$$

$$\sum x^2 Y = a\sum x^2 + c\sum x^4 \qquad 2710 = 10a + 34c$$

Solving these simultaneously, we get:

$a = 271$ $\qquad c = 0$

$\widehat{Y} = 271 + 21x + 0x^2$ (where 1993 = 0 and x units = 1 year)

b) The choice of a second-degree curve was unnecessary, since the best fitting second-degree equation has $c = 0$, and hence is a linear equation!

15-66

Year	1991	1992	1993	1994	1995
x	−2	−1	0	1	2
Actual sales (Y)	3.5	3.8	4.0	3.7	3.9
Trend($\widehat{Y}$=3.78+.07x)	3.64	3.71	3.78	3.85	3.92
% of trend ($100Y/\widehat{Y}$)	96.15	102.43	105.82	96.10	99.49

1993 had the largest percent of trend, 1994 the smallest.

15-67 a, b) NOTE: Although numbers in this solution are displayed with at most 4 decimal places, all computations were done in an Excel spreadsheet, without rounding intermediate results.

Year	Quarter	Data (Listings)	4–Quarter Moving Average	Centered Moving Average	Ratio of Data to CMA	Seasonal Index	Deseason-alized Data
1992	1	75				0.9854	76.1122
	2	77				1.0117	76.1062
			74.50				
	3	72		74.250	0.9697	1.0092	71.3469
			74.00				
	4	74		73.625	1.0051	0.9937	74.4680
			73.25				
1993	1	73		73.875	0.9882	0.9854	74.0825
			74.50				
	2	74		74.375	0.9950	1.0117	73.1411
			74.25				
	3	77		74.375	1.0353	1.0092	76.3015
			74.50				
	4	73		75.125	0.9717	0.9937	73.4617
			75.75				
1994	1	74		76.125	0.9721	0.9854	75.0974
			76.50				
	2	79		77.625	1.0177	1.0117	78.0830
			78.75				
	3	80		79.500	1.0063	1.0092	79.2743
			80.25				
	4	82				0.9937	82.5186
1995	1	80				0.9854	81.1863

Year	1		2		3		4	
1992					0.9697		1.0051	
1993	0.9882		0.9950		1.0353		0.9717	
1994	0.9721		1.0177		1.0063			
Modified sum	1.9602		2.0127		3.0113		1.9768	
Modified mean	0.9801	+	1.0063	+	1.0038	+	0.9884	= 3.9786/4 = 0.99465
Seasonal index	0.9854		1.0117		1.0092		0.9937	

c)

Year	Quarter	Deseason-alized Data (Y)	x	xY	x^2
1992	1	76.1122	−6	−456.6732	36
	2	76.1062	−5	−380.5312	25
	3	71.3469	−4	−285.3875	16
	4	74.4680	−3	−223.4041	9
1993	1	74.0825	−2	−148.1651	4
	2	73.1411	−1	−73.1411	1
	3	76.3015	0	0.0000	0
	4	73.4617	1	73.4617	1
1994	1	75.0974	2	150.1947	4
	2	78.0830	3	234.2491	9
	3	79.2743	4	317.0973	16
	4	82.5186	5	412.5931	25
1995	1	81.1863	6	487.1180	36
		991.1799	0	107.4118	182

$$a = \overline{Y} = \frac{991.1799}{13} = 76.2446 \qquad b = \frac{\sum xY}{\sum x^2} = \frac{107.4118}{182} = 0.5902$$

$\widehat{Y} = 76.2446 + 0.5902x$ (where 1993-3 = 0 and x units = 1 quarter)

CHAPTER 16

INDEX NUMBERS

16-1 Index for Base Year = 100

16-2 Price indices and quantity indices describe the change in a single variable, price and quantity (or number), respectively. A value index describes how the product of the two variables, price and quantity, changes over a period of time.

16-3 The CPI measures the overall price changes in a variety of consumer goods, i.e., a composite of variables.

16-4 An index may be used by itself or as a part of an intermediate computation to better understand some other information.

16-5 An index number measures the degree of change of a variable over a period of time or the degree of difference in a variable exhibited by several observations taken at the same time.

16-6 $$\text{Percentage relative} = \frac{\text{Current Value}}{\text{Base Value}} \times 100$$

16-7

	A	B	C	D	E	Total
1992	127	532	2290	60	221	3230
1995	152	651	2314	76	286	3479

$$1995\text{ Index} = \frac{\sum P_1}{\sum P_0} \times 100 = \frac{347900}{3230} = 107.7$$

16-8

	1992 P_0	1993 P_1	1994 P_2	1995 P_3
Class A	8.48	9.32	10.34	11.16
Class B	6.90	7.52	8.19	8.76
Class C	4.50	4.99	5.48	5.86
Class D	3.10	3.47	3.85	4.11
	22.98	25.30	27.86	29.89
$\text{Index} = \frac{\sum P_i}{\sum P_0} \times 100$:	$\frac{2298}{22.98} = 100.0$	$\frac{2530}{22.98} = 110.1$	$\frac{2786}{22.98} = 121.2$	$\frac{2989}{22.98} = 130.1$

16-9

	1993 P_0	1994 P_1	1995 P_2	1996 P_3
Eastern	3142	3564	4109	4372
State	2816	3474	3682	4019
Western	3582	3987	4406	4819
Central	4014	4197	4384	4671
	13,554	15,222	16,581	17,881
$\text{Index} = \frac{\sum P_i}{\sum P_0} \times 100$:	$\frac{1,355,400}{13,554} = 100.0$	$\frac{1,522,200}{13,554} = 112.3$	$\frac{1,658,100}{13,554} = 122.3$	$\frac{1,788,100}{13,554} = 131.9$

16-10

Commodity	1993 P_1	1994 P_0	1995 P_2
Dairy	\$2.34	\$2.38	\$2.60
Meat	3.19	3.41	3.36
Vegetables	.85	.89	.94
Fruit	1.11	1.19	1.18
	\$7.49	\$7.87	\$8.08

Index $= \dfrac{\sum P_i}{\sum P_0} \times 100$: $\dfrac{749}{7.87} = 95.2$ $\quad \dfrac{787}{7.87} = 100.0$ $\quad \dfrac{808}{7.87} = 102.7$

16-11

Material	1993 Q_0	1995 Q_1
A	86	95
B	395	380
C	1308	1466
D	430	469
E	113	108
	2332	2518

$$1995 \text{ Index} = \frac{\sum Q_1}{\sum Q_0} \times 100 = \frac{251800}{2332} = 108.0$$

16-12

Transaction	Q_0 1994	Q_1 1995
Savings withdrawal	169,000	158,000
Checking withdrawal	21,843,000	23,241,000
Savings deposit	293,000	303,000
Checking deposit	2,684,000	3,361,000
	24,989,000	27,063,000

$$1995 \text{ Index} = \frac{\sum Q_1}{\sum Q_0} \times 100 = \frac{2{,}706{,}300{,}000}{24{,}989{,}000} = 108.3$$

16-13

	1993	1994	1995
Total Quantity	187	203	231

Index $= \dfrac{\sum Q_1}{\sum Q_0} \times 100$: $\dfrac{18700}{187} = 100.0$ $\quad \dfrac{20300}{187} = 108.6$ $\quad \dfrac{23100}{187} = 123.5$

The total number of new texts increased 8.6% from 1993 to 1994. In 1995, the number of texts published was 23.5% higher than in 1993.

Data for Problems 16-14 to 16-16

	Selling Price				Number Sold			
	1993	1994	1995	1996	1993	1994	1995	1996
ED 107	1894	1906	1938	1957	84.6	86.9	96.4	107.5
ED Electra	2506	2560	2609	2680	38.4	42.5	55.6	67.5
ED Optima	1403	1440	1462	1499	87.4	99.4	109.7	134.6
ED 821	1639	1650	1674	1694	75.8	78.9	82.4	86.4

16-14 Laspeyres index

Base year: 1993	1993 P_0Q_0	1994 P_1Q_0	1995 P_2Q_0	1996 P_3Q_0
ED 107	160,232.4	161,247.6	163,954.8	165,562.2
ED Electra	96,230.4	98,304.0	100,185.6	102,912.0
ED Optima	122,622.2	125,856.0	127,778.8	131,012.6
ED 821	124,236.2	125,070.0	126,889.2	128,405.2
	503,321.2	510,477.6	518,808.4	527,892.0

Index $= \dfrac{\sum P_iQ_0}{\sum P_0Q_0} \times 100$:	$\dfrac{50,332,120}{503,321.2} = 100.0$	$\dfrac{51,047,760}{503,321.2} = 101.4$	$\dfrac{51,880,840}{503,321.2} = 103.1$	$\dfrac{52,789,200}{503,321.2} = 104.9$

16-15 Fixed weight index

Base prices: 1993

Fixed weights: 1996

	1993 P_0Q_3	1994 P_1Q_3	1995 P_2Q_3	1996 P_3Q_3
ED 107	203,605.0	204,895.0	208,335.0	210,377.5
ED Electra	169,155.0	172,800.0	176,107.5	180,900.0
ED Optima	188,843.8	193,824.0	196,785.2	201,765.4
ED 821	141,609.6	142,560.0	144,633.6	146,361.6
	703,213.4	714,079.0	725,861.3	739,404.5
Index $= \dfrac{\sum P_iQ_3}{\sum P_0Q_3} \times 100$:	$\dfrac{70,321,340}{703,213.4} = 100.0$	$\dfrac{71,407,900}{703,213.4} = 101.5$	$\dfrac{72,586,130}{703,213.4} = 103.2$	$\dfrac{73,940,550}{703,213.4} = 105.1$

16-16 Paasche index

Base period: 1994

	1993		1995		1996	
	P_1Q_1	P_0Q_1	P_2Q_2	P_0Q_2	P_3Q_3	P_0Q_3
ED 107	160,232.4	161,247.6	190,699.2	187,550.4	210,377.5	204,895.0
ED Electra	96,230.4	98,304.0	145,060.4	142,336.0	180,900.0	172,800.0
ED Optima	122,622.2	125,856.0	160,381.4	157,968.0	201,765.4	193,824.0
ED 821	124,236.2	125,070.0	137,937.6	135,960.0	146,361.6	142,560.0
	503,321.2	510,477.6	634,078.6	623,814.4	739,404.5	714,079.0

Index $= \dfrac{\sum P_iQ_i}{\sum P_0Q_i} \times 100$:	$\dfrac{50,332,120}{510,477.6} = 98.6$	$\dfrac{63,407,860}{623,814.4} = 101.6$	$\dfrac{73,940,450}{714,079.0} = 103.5$

16-17

Tape Length	1993 P_0	1994 P_1	1995 P_2	1993-5 Q	1993 P_0Q	1994 P_1Q	1995 P_2Q
30 min	\$2.20	\$2.60	\$2.85	32	\$ 70.40	\$ 83.20	\$ 91.20
60 min	2.60	2.90	3.15	119	309.40	345.10	374.85
90 min	3.10	3.20	3.25	75	232.50	240.00	243.75
120 min	3.30	3.35	3.40	16	52.80	53.60	54.40
					\$665.10	\$721.90	\$764.20

Index $= \dfrac{\sum P_iQ}{\sum P_0Q} \times 100$:	$\dfrac{66510}{665.10} = 100.0$	$\dfrac{72190}{665.10} = 108.5$	$\dfrac{76420}{665.10} = 114.9$

16-18

Fruit	June P_0	July P_1	August P_2	June Q_0	June P_0Q_0	July P_1Q_0	August P_2Q_0
Apples	.59	.64	.69	150	88.50	96.00	103.50
Oranges	.75	.65	.70	200	150.00	130.00	140.00
Peaches	.87	.90	.85	125	108.75	112.50	106.25
Watermelons	1.00	1.10	.95	350	350.00	385.00	332.50
Canteloupes	.95	.89	.90	150	142.50	133.50	135.00
					839.75	857.00	817.25

Index $= \dfrac{\sum P_iQ_0}{\sum P_0Q_0} \times 100$:	$\dfrac{83975}{839.75} = 100.00$	$\dfrac{85700}{839.75} = 102.1$	$\dfrac{81725}{839.75} = 97.30$

They are Laspeyres indices, since the weights are the base-month quantities.

16-19

Gadget	1993 P_0	1994 P_1	1995 P_2	1993-5 Q	1993 P_0Q	1994 P_1Q	1995 P_2Q
W	1.25	1.50	2.00	900	1125.00	1350.0	1800.0
X	6.50	7.00	6.25	50	325.00	350.0	312.5
Y	5.25	5.90	6.40	175	918.75	1032.5	1120.0
Z	.50	.80	1.00	200	100.00	160.0	200.0
					2468.75	2892.5	3432.5

$$\text{Index} = \frac{\sum P_i Q}{\sum P_0 Q} \times 100: \quad \frac{246875}{2468.75} = 100.0 \qquad \frac{289250}{2468.75} = 117.2 \qquad \frac{343250}{2468.75} = 139.0$$

16-20

Type of seats	1992 P_1	1993 P_0	1994 P_2	1995 P_3	1993 Q_0
Box seats	6.50	7.25	7.50	8.10	27
General admission	3.50	3.85	4.30	4.35	80

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_0Q_0)}{\sum P_0Q_0} = \left(\frac{\sum P_iQ_0}{\sum P_0Q_0}\right) \times 100$$

P_1Q_0	P_0Q_0	P_2Q_0	P_3Q_0
175.50	195.75	202.50	218.70
280.00	308.00	344.00	348.00
455.50	503.75	546.50	566.70

$$\text{Index:} \quad \frac{45550}{503.75} = 90.4 \qquad \frac{50375}{503.75} = 100.0 \qquad \frac{54650}{503.75} = 108.5 \qquad \frac{56670}{503.75} = 112.5$$

16-21

Material	1993 P_1	1994 P_2	1995 P_0	1995 P_0Q_0	1995 Q_0
Butadiene	17	15	11	50	4.5455
Styrene	85	89	95	210	2.2105
Rayon cord	348	358	331	1640	4.9547
Carbon black	62	58	67	630	9.4030
$Na_4P_2O_7$	49	56	67	90	1.3433

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_0Q_0)}{\sum P_0Q_0} = \left(\frac{\sum P_iQ_0}{\sum P_0Q_0}\right) \times 100$$

P_1Q_0	P_2Q_0	P_0Q_0
77.27	68.18	50
187.89	196.73	210
1724.24	1773.78	1640
582.99	545.37	630
65.82	75.22	90
2638.21	2659.28	2620

$$\text{Index:} \quad \frac{263821}{2620} = 100.7 \qquad \frac{265928}{2620} = 101.5 \qquad \frac{262000}{2620} = 100.0$$

16-22

Repair	1991 P_0	1993 P_1	1995 P_2	1991 P_0/P_0	1993 P_1/P_0	1995 P_2/P_0
Water pump	35	37	41	1.000	1.057	1.171
Engine valves	189	205	216	1.000	1.085	1.143
Wheel balancing	26	29	30	1.000	1.115	1.154
Tune-up	16	16	18	1.000	1.000	1.125
				4.000	4.257	4.593

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n}: \quad \frac{400.0}{4} = 100.0 \qquad \frac{425.7}{4} = 106.4 \qquad \frac{459.3}{4} = 114.8$$

16-23 (000's omitted in all figures)

	1993 P_0	1994 P_1	1995 P_2	1993 P_0/P_0	1994 P_1/P_0	1995 P_2/P_0
Savings	1845	2320	2089	1.000	1.257	1.132
Checking	385	447	491	1.000	1.161	1.275
				2.000	2.418	2.407

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n}: \quad \frac{200.0}{2} = 100.0 \qquad \frac{2.418}{2} = 120.9 \qquad \frac{2.407}{2} = 120.4$$

16-24

Product	1994 P_0	1996 P_1	$\frac{P_1}{P_0}$	P_0Q_0	$\left(\frac{P_1}{P_0}\right)(P_0Q_0)$
1-megabyte chips	\$ 42	\$ 65	1.5476	957	1481.07
4-megabyte chips	180	247	1.3722	487	668.27
16-megabyte chips	447	612	1.3691	349	477.83
			4.2889	1793	2627.17

a) $$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n} = \frac{428.89}{3} = 143.0$$

b) $$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_0Q_0)}{\sum P_0Q_0} = \frac{262{,}717}{1793} = 146.5$$

16-25

Destination	1991 P_1	1992 P_2	1993 P_3	1994 P_4	1995 P_0	1995 Q_0
Paris	690	714	732	777	783	2835
London	648	654	675	696	744	5175
Munich	702	723	753	768	798	2505
Rome	840	867	903	939	975	2145

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_0Q_0)}{\sum P_0Q_0} = \frac{\sum P_iQ_0}{\sum P_0Q_0} \times 100$$

P_1Q_0	P_2Q_0	P_3Q_0	P_4Q_0	P_0Q_0
1,956,150	2,024,190	2,075,220	2,202,795	2,219,805
3,353,400	3,384,450	3,493,125	3,601,800	3,850,200
1,758,510	1,811,115	1,886,265	1,923,840	1,998,990
1,801,800	1,859,715	1,936,935	2,014,155	2,091,375
8,869,860	9,079,470	9,391,545	9,742,590	10,160,370

Index: $\frac{886{,}986{,}000}{10{,}160{,}370} = 87.3$ $\qquad \frac{907{,}947{,}000}{10{,}160{,}370} = 89.4$ $\qquad \frac{939{,}154{,}500}{10{,}160{,}370} = 92.4$ $\qquad \frac{974{,}259{,}000}{10{,}160{,}370} = 95.9$ $\qquad \frac{1{,}016{,}037{,}000}{10{,}160{,}370} = 100.0$

16-26

	1992	1993	1994	1995	1992	1993	1994	1995
Group	P_1	P_2	P_0	P_3	P_1/P_0	P_2/P_0	P_0/P_0	P_3/P_0
Physicians	54	65	86	103	0.628	0.756	1.000	1.198
Student	39	41	55	76	0.709	0.745	1.000	1.382
Government	48	61	76	93	0.632	0.803	1.000	1.224
Teachers	46	58	75	96	0.613	0.773	1.000	1.280
					2.582	3.077	4.000	5.084

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n}: \quad \frac{258.2}{4} = 64.5 \quad \frac{307.7}{4} = 76.9 \quad \frac{400.0}{4} = 100.0 \quad \frac{508.4}{4} = 127.1$$

16-27

	1993	1994	1995	1993	1994	1995
Hotel	P_0	P_1	P_2	P_0/P_0	P_1/P_0	P_2/P_0
HH	35	37	42	1.000	1.057	1.200
RSI	25	26	28	1.000	1.040	1.120
EM	45	45	51	1.000	1.000	1.133
CI	37	38	44	1.000	1.027	1.189
FFM	26	30	31	1.000	1.154	1.192
				5.000	5.278	5.834

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n}: \quad \frac{500.0}{5} = 100.0 \quad \frac{527.8}{5} = 105.6 \quad \frac{583.4}{5} = 116.7$$

16-28

	1993	1994	1995	1993	1993	1994	1995
Maps	P_0	P_1	P_2	Q_0	P_0Q_0	P_1Q_0	P_2Q_0
City	.75	.90	1.10	1000	750	900	1100
County	.75	.90	1.00	400	300	360	400
State	1.00	1.50	1.50	1000	1000	1500	1500
U.S.	2.50	2.75	2.75	220	550	605	605
					2600	3365	3605

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_0Q_0)}{\sum P_0Q_0} = \frac{\sum P_iQ_0}{\sum P_0Q_0} \times 100: \quad \frac{336{,}500}{2600} = 129.4 \quad \frac{360{,}500}{2600} = 138.7$$

16-29 A value index depends on changes in both quantities and prices; as a result, a change in a value index cannot be identified with any single component.

16-30 The weighted aggregates index uses quantities for weights whereas the weighted average of relatives index uses values for weights.

16-31

	1991	1992	1993	1994	1995	1991	1992	1993	1994	1995
	V_0	V_1	V_2	V_3	V_4	V_0/V_0	V_1/V_0	V_2/V_0	V_3/V_0	V_4/V_0
Credit	5.66	6.32	6.53	6.98	7.62	1.000	1.117	1.154	1.233	1.346
Cash	2.18	2.51	2.48	2.41	2.33	1.000	1.151	1.138	1.106	1.069
						2.000	2.268	2.292	2.339	2.415

$$\text{Index} = \frac{\sum\left(\frac{V_i}{V_0} \times 100\right)}{n}: \quad \frac{200.0}{2} = 100.0 \quad \frac{226.8}{2} = 113.4 \quad \frac{229.2}{2} = 114.6 \quad \frac{233.9}{2} = 117.0 \quad \frac{241.5}{2} = 120.8$$

16-32

Product	1993 Q_1	1994 Q_2	1995 Q_0	1995 P_0	1993 Q_1P_0	1994 Q_2P_0	1995 Q_0P_0
Barges	92	118	85	33	3036	3894	2805
Cars	456	475	480	56	25536	26600	26880
Trucks	52	56	59	116	6032	6496	6844
					34604	36990	36529

$$\text{Index} = \frac{\sum Q_i P_0}{\sum Q_0 P_0} \times 100: \quad \frac{3{,}460{,}400}{36529} = 94.7 \qquad \frac{3{,}699{,}000}{36529} = 101.3 \qquad \frac{3{,}652{,}900}{36529} = 100.0$$

16-33

Model	1993 Q_0	1994 Q_1	1995 Q_2	1995 P_2	1995 P_2Q_2	1993 $\left(\frac{Q_0}{Q_0}\right)P_2Q_2$	1994 $\left(\frac{Q_1}{Q_0}\right)P_2Q_2$	1995 $\left(\frac{Q_2}{Q_0}\right)P_2Q_2$
Business	11.85	13.32	15.75	34	535.50	535.50	601.93	711.74
Scientific	10.32	11.09	10.18	69	702.42	702.42	754.83	692.89
Basic	7.12	7.48	7.89	13	102.57	102.57	107.76	113.66
						1340.49	1464.52	1518.29

$$\text{Index} = \frac{\sum\left(\frac{Q_i}{Q_0} \times 100\right)(P_2Q_2)}{\sum P_2Q_2}: \quad \frac{134049}{1340.49} = 100.0 \qquad \frac{146452}{1340.49} = 109.3 \qquad \frac{151829}{1340.49} = 113.3$$

16-34

Type of Crime	1992 Q_1	1993 Q_2	1994 Q_3	1995 Q_0	1992 Q_1/Q_0	1993 Q_2/Q_0	1994 Q_3/Q_0	1995 Q_0/Q_0
Assault & Rape	110	128	134	129	0.853	0.992	1.039	1.000
Murder	30	45	40	48	0.625	0.938	0.833	1.000
Robbery	610	720	770	830	0.735	0.867	0.928	1.000
Larceny	2450	2630	2910	2890	0.848	0.910	1.007	1.000
					3.061	3.707	3.808	4.000

$$\text{Index} = \frac{\sum\left(\frac{Q_i}{Q_0} \times 100\right)}{n}: \quad \frac{306.1}{4} = 76.5 \qquad \frac{370.7}{4} = 92.7 \qquad \frac{380.8}{4} = 95.2 \qquad \frac{400.0}{4} = 100.0$$

16-35

Type	1991 V_0	1992 V_1	1993 V_2	1994 V_3	1995 V_4	1996 V_5
Hard rock	642.4	721.5	842.6	895.3	905.6	951.2
Soft rock	325.8	347.8	398.5	406.3	418.7	426.4
Classical	118.3	123.6	174.3	176.2	174.9	185.3
Jazz	125.6	122.4	137.8	149.6	172.9	205.4
Alternative	208.7	252.7	405.9	608.9	942.7	987.4

V_1/V_0	V_2/V_0	V_3/V_0	V_4/V_0	P_5/V_0
1.123	1.312	1.394	1.410	1.481
1.068	1.223	1.247	1.285	1.309
1.045	1.473	1.489	1.478	1.566
0.975	1.097	1.191	1.377	1.635
1.211	1.945	2.918	4.517	4.731
5.421	7.050	8.239	10.067	10.722

$$\text{Index} = \frac{\sum\left(\frac{V_i}{V_0} \times 100\right)}{n}: \quad \frac{542.1}{5} = 108.4 \qquad \frac{705.0}{5} = 141.0 \qquad \frac{823.9}{5} = 164.8 \qquad \frac{1006.7}{5} = 201.3 \qquad \frac{1072.2}{5} = 214.4$$

16-36

Age (yrs.)	1973 "P_0"	1973 Q_0	1993 Q_1	1973 P_0Q_0	1993 P_0Q_1
< 4	5000	400	125	2,000,000	625,000
4 – 15	4000	295	200	1,180,000	800,000
16 – 25	24000	1230	1000	29,520,000	24,000,000
36 – 60	19000	700	450	13,300,000	8,550,000
> 60	7000	1100	935	7,700,000	6,545,000
				53,700,000	40,520,000

$$1993 \text{ index} = \frac{\sum P_0Q_1}{\sum P_0Q_0} \times 100: \quad \frac{4,052,000,000}{53,700,000} = 75.5$$

16-37

	Dec. Q_0	Jan. Q_1	Feb. Q_2	P	Dec. Q_0P	Jan. Q_1P	Feb. Q_2P
Cats	100	200	95	55	5500	11000	5225
Dogs	125	75	200	65	8125	4875	13000
Parrots	15	20	15	85	1275	1700	1275
Snakes	10	5	5	100	1000	500	500
					15900	18075	20000

$$\frac{\sum\left(\frac{Q_i}{Q_0} \times 100\right)(Q_0P)}{\sum Q_0P} = \frac{\sum Q_iP}{\sum Q_0P} \times 100: \quad \frac{1,590,000}{15900} = 100.0 \quad \frac{1,807,500}{15900} = 113.7 \quad \frac{2,000,000}{15900} = 125.8$$

16-38 Appropriate weighting for one period may become inappropriate in a short time. Unless the weights are changed the index becomes less informative.

16-39 1) A "normal" period--not at either the peak or the trough of a fluctuation.

2) A recent period

3) A period which coincides with the base period for one or more major indices, so that comparisons may be made.

16-40 The values from several adjoining periods are averaged.

16-41 No; the CPI does not measure all factors included in the "cost of living," and those factors it does measure are only measured for moderate-income, urban Americans rather than all Americans.

16-42 An index does not reflect changes in the quality of items and therefore understates or overstates the price level change if the quality changes.

16-43 a)

Model	1993 P_1	1994 P_2	1995 P_0	1995 Q_0	1993 P_1Q_0	1994 P_2Q_0	1995 P_0Q_0
I	139	155	149	7.6	1056.4	1178.0	1132.4
II	169	189	189	8.1	1368.9	1530.9	1530.9
III	199	205	219	3.4	676.6	697.0	744.6
					3101.9	3405.9	3407.9

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_0Q_0)}{\sum P_0Q_0} = \frac{\sum P_iQ_0}{\sum P_0Q_0} \times 100: \quad \frac{310190}{3407.9} = 91.0 \quad \frac{340590}{3407.9} = 99.9 \quad \frac{340790}{3407.9} = 100.0$$

b)

Model	1993 P_1	1993 Q_1	1994 P_2	1994 Q_2	1995 P_0	1993 P_1Q_1	1993 $\left(\frac{P_1}{P_0}\right)P_1Q_1$	1994 P_2Q_2	1994 $\left(\frac{P_2}{P_0}\right)P_2Q_2$
I	139	3.7	155	4.1	149	514.3	479.78	635.5	661.09
II	169	2.3	189	4.6	189	388.7	347.57	869.4	869.40
III	199	1.6	205	2.1	219	318.4	289.32	430.5	402.98
						1221.4	1116.67	1935.4	1933.47

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_iQ_i)}{\sum P_iQ_i}: \qquad \frac{111667}{1221.4} = 91.4 \qquad \frac{193347}{1935.4} = 99.9$$

16-44

Commodity	1991 V_0	1993 V_1	1995 V_2
Coffee	834	1436	1321
Sugar	96	118	122
Copper	241	258	269
Zinc	142	125	106
	1313	1937	1818

$$\text{Index} = \frac{\sum V_i}{\sum V_0} \times 100: \qquad \frac{131300}{1313} = 100.0 \qquad \frac{193700}{1313} = 147.5 \qquad \frac{181800}{1313} = 138.5$$

16-45

Type of coal	1989 Q_1	1990 Q_2	1991 Q_3	1992 Q_0	1992 Q_0P_0	1992 P_0	1989 Q_1P_0	1990 Q_2P_0	1991 Q_3P_0
Anthracite	7.4	6.8	7.1	7.2	90	12.50	92.5	85.0	88.75
Bituminous	595	580	601	625	5050	8.08	4807.6	4686.4	4856.08
					5140		4900.1	4771.4	4944.83

$$\text{Index} = \frac{\sum\left(\frac{Q_i}{Q_0} \times 100\right)(Q_0P_0)}{\sum Q_0P_0} = \left(\frac{\sum Q_iP_0}{\sum Q_0P_0}\right) \times 100: \qquad \frac{490010}{5140} = 95.3 \qquad \frac{477140}{5140} = 92.8 \qquad \frac{494483}{5140} = 96.2$$

(1992 index: 100.0)

16-46

Product	1991 P_0	1995 P_1	1991 Q_0	1991 P_0Q_0	1995 P_1Q_0
Cheese	1.45	1.49	2.6	3.77	3.874
Milk	1.60	1.61	47.6	76.16	76.636
Butter	.70	.80	3.1	2.17	2.480
				82.10	82.990

$$1995 \text{ index} = \frac{\sum P_1Q_0}{\sum P_0Q_0} \times 100 = \frac{8299}{82.10} = 101.1$$

16-47

Products	1992 P_0	1993 P_1	1994 P_2	1995 P_3	1992 P_0/P_0	1993 P_1/P_0	1994 P_2/P_0	1995 P_3/P_0
Jeans	\$13.00	\$13.00	\$15.00	\$15.00	1.000	1.000	1.154	1.154
Jackets	19.00	19.50	22.00	24.00	1.000	1.026	1.158	1.263
Shirts	12.00	11.00	12.00	13.00	1.000	0.917	1.000	1.083
					1.000	2.943	3.312	3.500

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n}: \qquad \frac{300.0}{3} = 100.0 \qquad \frac{2.943}{3} = 98.1 \qquad \frac{3.312}{3} = 110.4 \qquad \frac{3.500}{3} = 116.7$$

16-48 The problem of incomparability of indices would be present; there has been a basic change in what is being measured by the indices, because computer technology has changed significantly over the past few decades.

16-49

	1991	1992	1993	1991	1992	1993
Salesperson	V_1	V_2	V_0	V_1/V_0	V_2/V_0	V_0/V_0
A	704	985	1391	0.506	0.708	1.000
B	635	875	1306	0.486	0.670	1.000
C	752	1023	1523	0.494	0.672	1.000
D	503	696	1106	0.455	0.629	1.000
E	593	781	1215	0.488	0.643	1.000
				2.429	3.322	5.000

$$\text{Index} = \frac{\sum\left(\frac{V_i}{V_0} \times 100\right)}{n}: \quad \frac{242.9}{5} \quad \frac{3.322}{5} \quad \frac{5.000}{5}$$

$$= 48.6 \qquad = 66.4 \qquad = 100.0$$

16-50

	1993	1994	1995	1993	1993	1994	1995
Model	Q_0	Q_1	Q_2	P_0	Q_0P_0	Q_1P_0	Q_0P_0
Sport	45	48	56	89	4005	4272	4984
Touring	64	67	71	104	6656	6968	7384
Cross Country	28	35	27	138	3864	4830	3726
Sprint	21	16	28	245	5145	3920	6860
					19670	19990	22954

$$\text{Index} = \frac{\sum\left(\frac{Q_i}{Q_0} \times 100\right)(Q_0P_0)}{\sum Q_0P_0} = \left(\frac{\sum Q_iP_0}{\sum Q_0P_0}\right) \times 100: \quad \frac{1{,}967{,}000}{19670} \quad \frac{1{,}999{,}000}{19670} \quad \frac{2{,}295{,}400}{19670}$$

$$= 100.0 \qquad = 101.6 \qquad = 116.7$$

16-51 a) It ignores the fact that the various share prices are not equally important. A 10% rise in the price of Westinghouse stock does not have the same effect as a 10% rise in Coca Cola share prices.

b) Low point: DJIA = 1739 (Oct. 1987)

$$\text{Index 1992} = \frac{\sum P_i}{\sum P_0} \times 100: \qquad \frac{3301}{1739} \times 100 = 189.8217$$

$$\text{Index 1987} = \frac{\sum P_i}{\sum P_0} \times 100: \qquad \frac{3301}{2722} \times 100 = 121.2711$$

16-52

	1991	1993	1995
Costs	V_1	V_0	V_2
Wages	24,378	36,421	37,613
Lumber	1,816	2,019	2,136
Utilities	638	681	701
	26,832	39,121	40,450

$$\text{Index} = \frac{\sum V_i}{\sum V_0} \times 100: \quad \frac{2{,}683{,}200}{39121} \quad \frac{3{,}912{,}100}{39121} \quad \frac{4{,}045{,}000}{39121}$$

$$= 68.6 \qquad = 100.0 \qquad = 103.4$$

16-53

Products	1993 P_0	1994 P_1	1995 P_2	1993 P_0/P_0	1994 P_1/P_0	1995 P_2/P_0
Sirloin	$1.69	1.81	1.85	1.000	1.071	1.095
Chuck	.91	1.15	1.24	1.000	1.264	1.363
Bologna	1.45	1.58	1.53	1.000	1.090	1.055
Hot dogs	.99	1.03	1.01	1.000	1.040	1.020
Rib eye	2.39	2.61	2.56	1.000	1.092	1.071
				5.000	5.557	5.604

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n}: \quad \frac{500.0}{5} = 100.0 \qquad \frac{555.7}{5} = 111.1 \qquad \frac{560.4}{5} = 112.1$$

16-54 Depending on what is being measured, the choice of base periods can significantly distort the importance of a particular value.

16-55

Material	1992 P_0	1993 P_1	1994 P_2	1995 P_3
Aluminum	$ .96	$ .99	$1.33	$1.06
Steel	1.48	1.54	1.55	1.59
Brass tubing	.21	.25	.26	.31
Copper wire	.06	.08	.07	.09
	$2.71	$2.86	$2.91	$3.05

$$\text{Index} = \frac{\sum P_i}{\sum P_0} \times 100: \quad \frac{271}{2.71} = 100.0 \qquad \frac{286}{2.71} = 105.5 \qquad \frac{291}{2.71} = 107.4 \qquad \frac{305}{2.71} = 112.5$$

16-56

Product	1991 P_0	1991 Q_0	1992 Q_1	1993 Q_2	1994 Q_3	1995 Q_4
Wheat	4.40	610	620	640	630	650
Corn	3.60	390	390	410	440	440
Oats	1.20	100	90	120	130	150
Rye	24.00	10	20	10	10	20
Barley	2.10	160	150	120	190	180
Soybeans	5.60	130	140	160	120	130

1991 Q_0P_0	1992 Q_1P_0	1993 Q_2P_0	1994 Q_3P_0	1995 Q_4P_0
2684	2728	2816	2772	2860
1404	1404	1476	1584	1584
120	108	144	156	180
240	480	240	240	480
336	315	252	399	378
728	784	896	672	728
5512	5819	5824	5823	6210

$$\text{Index} = \frac{\sum Q_i P_0}{\sum Q_0 P_0} \times 100: \quad \frac{551200}{5512} = 100.0 \qquad \frac{581900}{5512} = 105.6 \qquad \frac{582400}{5512} = 105.7 \qquad \frac{582300}{5512} = 105.6 \qquad \frac{621000}{5512} = 112.7$$

16-57

Mineral	1994 P_0	1995 P_1	1995 Q_1	1994 P_0Q_1	1995 P_1Q_1
Copper	\$ .59	\$ .63	38.1	22.479	24.003
Lead	.17	.16	53.5	9.095	8.560
Zinc	.21	.23	86.4	18.144	19.872
				49.718	52.435

$$1995 \text{ Paasche index} = \frac{\sum P_1Q_1}{\sum P_0Q_1} \times 100 = \frac{5243.5}{49.718} = 105.5$$

16-58

Size	1991 P_1	1991 Q_1	1993 P_0	1993 Q_0	1995 P_2	1995 Q_2
Subcompact	62	32	68	65	70	86
Compact	76	45	78	68	80	73
Sedan	90	462	98	325	106	386

a)

1991 P_1Q_0	1993 P_0Q_0	1995 P_2Q_0
4030	4420	4550
5168	5304	5440
29250	31850	34450
38448	41574	44440

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_0Q_0)}{\sum P_0Q_0} = \frac{\sum P_iQ_0}{\sum P_0Q_0} \times 100: \quad \frac{3,844,800}{41574} \quad \frac{4,157,400}{41574} \quad \frac{4,444,000}{41574}$$

$= 92.5$ $\quad = 100.0$ $\quad = 106.9$

b)

1991 P_1Q_1	1991 $\left(\frac{P_1}{P_0}\right)P_1Q_1$	1993 $\left(\frac{P_0}{P_0}\right)P_0Q_0$	1995 P_2Q_2	1995 $\left(\frac{P_2}{P_0}\right)P_2Q_2$
1984	1808.94	4420	6020	6197.06
3420	3332.30	5304	5840	5989.74
41580	38185.72	31850	40916	44256.08
46984	43326.96	41574	52776	56442.88

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)(P_iQ_i)}{\sum P_iQ_i}: \quad \frac{4,332,696}{46984} \quad \frac{4,157,400}{41574} \quad \frac{5,644,288}{52776}$$

$= 92.2$ $\quad = 100.0$ $\quad = 106.9$

16-59

Product	1993 P_0	1994 P_1	1995 P_2	1996 P_3
Dishwasher	219	241	272	306
Washing machine	362	385	397	413
Dryer	229	241	261	275
Refrigerator	562	580	598	625
	1372	1447	1528	1619

$$\text{Index} = \frac{\sum P_i}{\sum P_0} \times 100: \quad \frac{137200}{1372} \quad \frac{144700}{1372} \quad \frac{152800}{1372} \quad \frac{161900}{1372}$$

$= 100.0$ $\quad = 105.5$ $\quad = 111.4$ $\quad = 118.0$

16-60

Department	1994 P_0	1995 P_1	1996 P_2	1994 P_0/P_0	1995 P_1/P_0	1996 P_2/P_0
Mechanical	3642	3891	4253	1.000	1.068	1.168
Chemical	3888	4052	4425	1.000	1.042	1.138
Biomedical	4251	4537	4724	1.000	1.067	1.111
Electrical	3764	4305	4297	1.000	1.144	1.142
				4.000	4.321	4.559

$$\text{Index} = \frac{\sum\left(\frac{P_i}{P_0} \times 100\right)}{n}: \quad \frac{400.0}{4} = 100.0 \qquad \frac{432.1}{4} = 108.0 \qquad \frac{455.9}{4} = 114.0$$

16-61

Indices (I)	1980 I_1	1981 I_2	1982 I_3	1983 I_4	1984 I_5	1985 I_6	1986 I_7	1987 I_8	1988 I_9	1989 I_{10}	1990 I_0	1991 I_{11}
Overall	81.7	85.6	88.0	89.6	91.7	93.5	94.1	94.2	94.9	97.0	100	103.3
Food	82.2	86.5	88.1	89.9	92.5	94.1	94.3	93.4	94.1	96.1	100	104.8
Housing	74.9	78.2	81.0	83.5	85.4	87.6	89.8	92.3	94.2	97.0	100	103.1

a) 1990 was the base year

b)

Indices (I)	1980 I_0	1981 I_1	1982 I_2	1983 I_3	1984 I_4	1985 I_5	1986 I_6	1987 I_7	1988 I_8	1989 I_9	1990 I_{10}	1991 I_{11}
Overall	100	104.8	107.7	109.7	112.2	114.4	115.2	115.3	116.2	118.7	122.4	126.4
Food	100	105.2	107.2	109.4	112.5	114.5	114.7	113.6	114.5	116.9	121.7	127.5
Housing	100	104.4	108.1	111.5	114.0	117.0	119.9	123.2	125.8	129.5	133.5	137.7

c) The price of housing increased the most from 1980 to 1991.

d) No. Consumers are more interested in the real value of money. Since the indices can vary greatly depending upon the base year used in the calculation, it doesn't have much meaning to consumers.

16-62 1993 real average wage = 521.35(100)/152 = \$342.99

16-63

Item	1992 P_0	1992 Q_0	1996 P_1	1996 Q_1	P_1Q_1	P_0Q_1
Cheese	1.19	2	2.09	1	2.09	1.19
Bread	.79	3	1.09	3	3.27	2.37
Eggs	.84	2	1.35	1	1.35	0.84
Milk	1.36	2	2.39	2	4.78	2.72
					11.49	7.12

$$\text{1996 Paasche index} = \frac{\sum P_1Q_1}{\sum P_0Q_1} \times 100 = \frac{1149}{7.12} = 161.4$$

16-64

Subject	1993 Q_1	1994 Q_2	1995 Q_3	1996 Q_0	1993 Q_1/Q_0	1994 Q_2/Q_0	1995 Q_3/Q_0	1996 Q_0/Q_0
Local	73	76	112	107	0.682	0.710	1.047	1.000
Snowboard	101	129	163	162	0.623	0.796	1.006	1.000
Handicapped	163	189	271	268	0.608	0.705	1.011	1.000
Regular	183	210	303	298	0.614	0.705	1.017	1.000
					2.527	2.916	4.081	4.000

$$\text{Index} = \frac{\sum\left(\frac{Q_i}{Q_0} \times 100\right)}{n}: \quad \frac{252.7}{4} = 63.2 \qquad \frac{291.6}{4} = 72.9 \qquad \frac{408.1}{4} = 102.0 \qquad \frac{400.0}{4} = 100.0$$

16-65

Product	1992 Q_1	1993 Q_2	1994 Q_0	1995 Q_3	1994 P_0	1992 Q_1P_0	1993 Q_2P_0	1994 Q_3P_0	1995 Q_0P_0
Wheat	4.6	6.7	4.0	5.2	2680	12328	17956	10720	13936
Feed grains	4.9	6.2	1.8	1.2	2270	11123	14074	4086	2724
Soybeans	4.7	5.7	1.2	1.8	3430	16121	19551	4116	6174
						39572	51581	18922	22834

$$\text{Index} = \frac{\sum Q_i P_0}{\sum Q_0 P_0} \times 100 : \quad \frac{3957200}{18922} \quad \frac{5158100}{18922} \quad \frac{1892200}{18922} \quad \frac{2283400}{18922}$$

$$= 209.1 \quad = 272.6 \quad = 100.0 \quad = 120.7$$

16-66

Time	1991 P_0	1991 Q_0	1996 P_1	P_1Q_1	P_0Q_1
Day	.17	5.2	.19	0.884	0.988
Evening	.13	8.7	.16	1.131	1.392
Night	.09	10.3	.12	0.927	1.236
				2.942	3.616

$$1996 \text{ Laspeyres index} = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100 = \frac{361.6}{2.942} = 122.9$$

16-67

Town	1992 P_1	1994 P_0	1996 P_2	1994 Q_0	1992 P_1Q_0	1994 P_0Q_0	1996 P_2Q_0
Greenville	21206	24210	26235	17	360502	411570	445995
Hampton	17129	19722	22109	14	239806	276108	309526
Middletown	25723	28657	32481	21	540183	601797	682101
					1140491	1289475	1437622

$$\text{Index} = \frac{\sum P_i Q_0}{\sum P_0 Q_0} \times 100 : \quad \frac{114049100}{1289475} \quad \frac{128947500}{1289475} \quad \frac{143762200}{1289475}$$

$$= 88.4 \quad = 100.0 \quad = 111.5$$

16-68

Item	1993 P_0	1994 P_1	1995 P_2	1996 P_3	1996 Q	1993 P_0Q	1994 P_1Q	1995 P_2Q	1996 P_3Q
Hamburger	0.58	0.62	0.69	0.79	1.8	1.044	1.116	1.242	1.422
Chicken	1.89	2.09	2.18	2.25	2.1	3.969	4.389	4.578	4.725
French fries	0.84	0.89	0.99	0.99	2.4	2.016	2.136	2.376	2.376
Onion rings	0.91	0.99	1.14	1.19	1.6	1.456	1.584	1.824	1.904
						8.485	9.225	10.020	10.427

$$\text{Index} = \frac{\sum P_i Q}{\sum P_0 Q} \times 100 : \quad \frac{848.5}{8.485} \quad \frac{922.5}{8.485} \quad \frac{1002.0}{8.485} \quad \frac{1042.7}{8.485}$$

$$= 100.0 \quad = 108.7 \quad = 118.1 \quad = 122.9$$

16-69

Item	1993 P_1	1994 P_2	1995 P_0	1996 P_3	1993 Q_1	1994 Q_2	1995 Q_0	1996 Q_3
Hamburger	0.58	0.62	0.69	0.79	2.1	2.5	2.0	1.8
Chicken	1.89	2.09	2.18	2.25	1.5	1.2	1.8	2.1
French fries	0.84	0.89	0.99	0.99	2.9	2.7	2.3	2.4
Onion rings	0.91	0.99	1.14	1.19	3.1	2.4	2.0	1.6

1993		1994		1996	
P_1Q_1	P_0Q_1	P_2Q_2	P_0Q_2	P_3Q_3	P_0Q_3
1.218	1.449	1.550	1.725	1.422	1.242
2.835	3.270	2.508	2.616	4.725	4.578
2.436	2.871	2.403	2.673	2.376	2.376
2.821	3.534	2.376	2.736	1.904	1.824
9.310	11.124	8.837	9.750	10.427	10.020

Index $= \frac{\sum P_iQ_i}{\sum P_0Q_i} \times 100$: $\frac{931.0}{11.124} = 83.7$ $\quad \frac{883.7}{9.750} = 90.6$ $\quad \frac{1042.7}{10.020} = 104.1$

16-70 Doubling a factor weight gives that factor extra impact in lieu of the missing factor; assigning low scores to a missing factor calls into question the entire rating process. Alternative responses to missing data include leaving out schools with missing information, or assigning average values to the missing factors. However, these alternatives still produce some distortions in the ratings.

16-71 a)

Company	1994 Q_0	1995 Q_1	Simple Index $Q_1/Q_0 \times 100$
GM	1,518,162	1,351,471	89.0
Ford	864,029	819,088	94.8
Chrysler	290,899	316,821	108.9
BMW	—	4,866	—
Honda	250,641	286,122	114.2
Nissan	171,804	184,284	107.3
Nummi	114,589	119,572	104.3
Toyota	140,090	199,840	142.7
Mazda	125,923	85,345	67.8
Mitsubishi	75,352	114,752	152.3
Subaru-Isuzu	14,098	39,579	280.7
	3,565,587	3,521,740	

b) The industry-wide index is given by: $\frac{\sum Q_1}{\sum Q_0} \times 100 = \frac{352,174,000}{3,565,587} = 98.8.$

The average of the individual indices is: $\frac{\sum\left(\frac{Q_1}{Q_0}\right) \times 100}{n} = \frac{1262.0}{10} = 126.2.$

The weighted average of relatives index is $\frac{\sum\left(\frac{Q_1}{Q_0}\right)(Q_0)(100)}{\sum Q_0} = \frac{\sum Q_1}{\sum Q_0} \times 100$, which is the same as the industry-wide average. It is lower than the average of the simple indices since it gives much more weight to the decreased production at GM and Ford.

CHAPTER 17

DECISION THEORY

17-1 No. If Wholesale had to make the decision today, the answer would be yes, since the manager would have to make the decision under uncertainty. If Wholesale waits a day, however, all uncertainty is removed by Leerie's firm order. The decision can then be made under certainty and the order placed in time for the lamp shop to receive its lamps for its sale.

17-2 Lisa is correct if she can obtain certain information; otherwise she isn't. She must determine her objective (presumably to maximize Adventures, Inc.'s profit), the available courses of action (which investments to make), the payoffs from these actions and the probability of the various payoffs being realized. The last two of these most likely will be difficult to determine.

17-3 All the elements needed to "solve" the problem by means of decision theory are present, and indeed the problem could be solved in that fashion. A closer inspection shows, however, that there is no risk of overstocking since Grambler absorbs all unsold copies. Thus the bookstore should order as many magazines as it can possibly sell since it faces no risk of loss. Decision theory would yield the same result after many more calculations.

17-4 We assume that the mechanics are paid for their vacations. Note that each mechanic works 2000 hours/year (50 weeks @ 40 hours/week).

a) The payoff table below gives both conditional and expected profits.

Mechanics needed Probability		5 .2	6 .3	7 .4	8 .1	Expected profit	
	5	66400	66400	66400	66400	66400	
Mechanics	6	47680	79680	79680	79680	73280	←
hired	7	28960	60960	92960	92960	70560	
	8	10240	42240	74240	106240	55040	

So the optimal decision is to hire six mechanics.

b) EVPI = .2(66400) + .3(79680) + .4(92960) + .1(106240) − 73280 = $11,712

17-5 Daily cost of car = 6750(.55)/312 = $11.90
Variable contribution per car per day = 24.95 − 2.25 = $22.70

The payoff table below gives both conditional and expected profits.

Cars rented Probability		10 .18	11 .19	12 .21	13 .15	14 .14	15 .13	Expected profit	
	10	108.00	108.00	108.00	108.00	108.00	108.00	108.00	
	11	96.10	118.80	118.80	118.80	118.80	118.80	114.71	
Cars	12	84.20	106.90	129.60	129.60	129.60	129.60	117.12	←
Bought	13	72.30	95.00	117.70	140.40	140.40	140.40	114.75	
	14	60.40	83.10	105.80	128.50	151.20	151.20	108.98	
	15	48.50	71.20	93.90	116.60	139.30	162.00	100.03	

They should buy 12 cars.

17-6 Labor cost = .375 × 72 boxes = \$27 per case sold

Purchase cost = 60 per case

\$87 total cost if sold

1.50 × 72 = \$108 = Total revenue per case sold

87 = Total cost

\$ 21 = Total profit per case

The payoff table below gives both conditional and expected profits.

Cases sold Probability		15 .05	16 .20	17 .30	18 .25	19 .10	20 .10	Expected profit
	15	315	315	315	315	315	315	315.00
	16	255	336	336	336	336	336	331.95
Cases	17	195	276	357	357	357	357	332.70 ←
Stocked	18	135	216	297	378	378	378	309.15
	19	75	156	237	318	399	399	265.35
	20	15	96	177	258	339	420	213.45

a) Order 17 cases

b) Expected profit \$332.70

17-7 a)

Jobs received Probability		24 .3	27 .2	30 .4	33 .1	Expected profit
	8	240	210	180	150	201
MBA's	9	180	270	240	210	225 ←
hired	10	120	210	300	270	225 ←
	11	60	150	240	330	177

Thus Emily should hire either 10 or 11 MBA's, and expect her company to make a profit of \$225,000 on their consulting work.

b) Expected profit under certainty $= .3(240) + .2(270) + .4(300) + .1(330)$
$= 279$

Thus EVPI = 279 − 225 = 54, i.e., \$54,000

17-8 The payoff table below gives both conditional and expected profits.

Demand (dozens) Probability		55 0.15	56 0.20	57 0.10	58 0.35	59 0.15	60 0.05	Expected profit
	55	646.80	646.80	646.80	646.80	646.80	646.80	646.80
	56	637.56	658.56	658.56	658.56	658.56	658.56	655.41
Pizzas	57	628.32	649.32	670.32	670.32	670.32	670.32	659.82
Ordered	58	619.08	640.08	661.08	682.08	682.08	682.08	662.13 ←
(dozens)	59	609.84	630.84	651.84	672.84	693.84	693.84	657.09
	60	600.60	621.60	642.60	663.60	684.60	705.60	648.90

They should order 58 dozen pizzas.

$$\text{EVPI} = 646.80(.15) + 658.56(.20) + 670.32(.10) + 682.08(.35) + 693.84(.15) + 705.60(.05) - 662.13 = \$11.718$$

Thus \$11.72 is the maximum they should pay for perfect information about the demand.

17-9 The payoff table below gives both conditional and expected profits.

1000 pairs demanded		45	50	55	60	65	Expected
Probability		.25	.30	.20	.15	.10	profit
	45	724.50	724.50	724.50	724.50	724.50	724.50
	50	648.25	805.00	805.00	805.00	805.00	765.81 ←
1000 pairs	55	572.00	728.75	885.50	885.50	885.50	760.10
produced	60	495.75	652.50	809.25	966.00	966.00	723.04
	65	419.50	576.25	733.00	889.75	1046.50	662.46

Manfred should recommend production of 50,000 pairs.

17-10 D.O.T. is \$8 better off on its budget for every sign it buys at \$21. This is similar to a marginal profit. \$21 is the cost D.O.T. has to absorb if it overstocks. This is the marginal loss.

$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{21}{29} = .7241,$ which corresponds to $-.595\sigma$, so D.O.T.should purchase $\mu - .595\sigma = 78 - .595(15) = 69.075$, or 69 signs.

17-11 $\text{MP} = \text{selling price} - \text{cost} = 3.25 - (.75 + 1.50) = \$1.00$
$\text{ML} = \$.75$ since there is a buyer for the shirts

$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{.75}{1.75} = .4286,$ which corresponds to $.18\sigma$, so they should buy $\mu + .18\sigma = 200 + .18(34) = 206.12$, or 206 patches.

17-12 a) $\text{MP} = 1.50 - 0.67 = 0.83$ $\text{ML} = 0.67$

$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{0.67}{1.50} = 0.4467$ which corresponds to $.13\sigma$, so he should order $\mu + .13\sigma = 375 + .13(20) = 377.6$, or 378 hot dogs.

b) Selling leftover hot dogs for \$0.50 leaves MP unchanged, but now $\text{ML} = 0.67 - 0.50 = 0.17$, so

$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{0.17}{1.00} = .0.17$ which corresponds to $.95\sigma$, so he should order $\mu + .95\sigma = 375 + .95(20) = 394$ hot dogs.

17-13 $\text{MP} = \$.80$ $\text{ML} = \$7.30$

$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{7.30}{8.10} = .9012,$ which corresponds to -1.29σ, so they should order $\mu - 1.29\sigma = 120 - 1.29(28) = 83.88$, or 84 rims.

17-14 $\text{MP} = 80$ $\text{ML} = 135$

$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{135}{215} = .6279,$ which corresponds to $-.33\sigma$, so they should prepare $\mu - .33\sigma = 190 - .33(32) = 179.44$ orders, or 90 chickens.

17-15 a) $\text{MP} = 35 - 30 = 5$ $\text{ML} = 30 - 5 = 25$

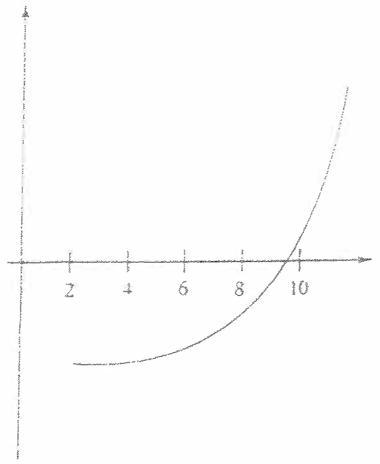
$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{25}{30} = .8333,$ which corresponds to $-.97\sigma$, so they should stock $\mu - .97\sigma = 300 - .97(50) = 251.5$, or 251 polyester-belted tires for next year.

b) $\text{MP} = 60 - 45 = 15$ $\text{ML} = 45 - 5 = 40$

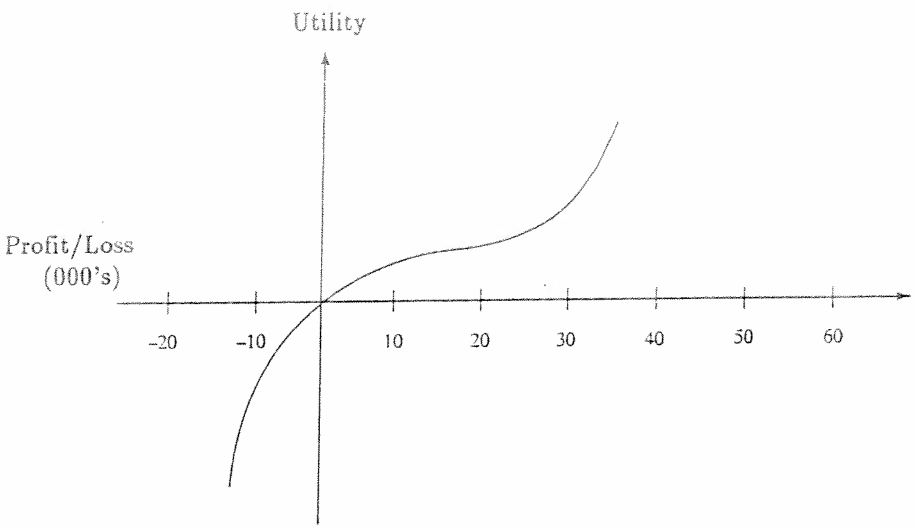
$p^* = \frac{40}{55} = .7273,$ which corresponds to $-.60\sigma$, so they should stock $\mu - .6\sigma = 200 - .6(20) = 188$ steel-belted tires for next year.

17-16 a) In a 50% tax bracket, to earn 9.43% after taxes, Bill must earn 18.86% before taxes.
18.86% of \$1,600,000 = \$301,760

b) Utility



17-17



17-18

Change	−1000	−500	0	500	1000	1500	Expected utility
Utility	0.0	0.1	0.7	0.8	0.9	1.0	
P(change | 2-month option)	.05	.15	.15	.25	.35	.05	0.685
P(change | 4-month option)	0	.05	.05	.20	.30	.40	0.870
P(change | no purchase)	0	0	1	0	0	0	0.700

She should purchase the 4-month option.

17-19 $7/20 = .35$, corresponding to 1.04σ, so $\sigma = 25/1.04 = 24.04$ hours.
The average number of hours needed per year to break even is

$2100/3(5) = 140$, corresponding to $z = \dfrac{140 - 185}{24.04/\sqrt{3}} = -3.24$.

$P(z > -3.24) = 1.000$, so they can be 98% sure the welder will pay for itself over a 3-year period.

17-20 $7/18 = .3889$, corresponding to 1.22σ, so $\sigma = 1/1.22 = .82$ month.
The probability that the upswing will last longer than six months is
$P(z > (6-8)/.82) = P(z > -2.44) = .9927$, which exceeds 95%. He should hire.

17-21 $6/20 = .3$, corresponding to 0.84σ, so $\sigma = 3/.84 = 3.571$ mpg. To save money by switching, the current mileage must be below 40, corresponding to $z = \dfrac{40-36}{3.571} = 1.12$. $P(z > 1.12) = 0.8686$. Since this is below the required 0.95, they shouldn't switch.

17-22 $4/14 = .2857$, corresponding to $.79\sigma$, so $\sigma = 12000/.79 = 15190$ miles.
Tax savings per mile $= .31 \times .34 = .1054$ so her breakeven mileage is
$12,250/.1054 = 116,224$ miles, corresponding to $z = \dfrac{116,224 - 120,000}{15,190} = -0.25$.
$P(z \geq -0.25) = .5987$ is the probability that the car will last long enough for Natalie to break even.

17-23 $9/20 = .45$, corresponding to 1.64σ, so $\sigma = 20/1.64 = 12.2$ tickets.
To pay for the radar unit, they will need to give at least $\dfrac{2000}{20} = 100$ additional tickets, corresponding to $z = \dfrac{100 - 115}{12.2} = -1.23$. $P(z > -1.23) = .8907$, so they cannot be 99% sure the unit will pay for itself in one year.

17-24 Broker A: $2/6 = .3333$, corresponding to $.97\sigma$, so $\sigma = 5/.97 = 5.155$.

Thus, $P(\text{price} \geq 60) = P\left(z \geq \dfrac{60-68}{5.155}\right) = P(z \geq -1.55) = .9394$

Broker B: $5/12 = .4167$, which corresponds to 1.38σ, so $\sigma = 5/1.38 = 3.623$.

Thus, $P(\text{price} \geq 60) = P\left(z \geq \dfrac{60-65}{3.623}\right) = Pz \geq -1.38) = .9162$

Since both of these are above .80, buy the stock.

17-25 With a three year decision horizon, we get the following tree:

```
PROFIT     PLANT TO   PROFIT      INITIAL      PROB  PROFIT     UPGRADE   PROFIT          FINAL       PROB   PROFIT
           BUILD                  ACCEPTANCE                    PILOT?                    ACCEPTANCE

                                                                                      +-- LOW  ----   50%   -18.00
              +-- NEW --------------------------------------------------------  12.00 --( )
              |                                                                        +-- HIGH ----   50%    42.00
              |
              |
              |                                                                        +-- LOW  ----   90%   -11.00
              |                                                      +-- YES ---   -7.00 --( )
13.35 --[ ]                                                          |                 +-- HIGH ----   10%    29.00
              |                         +-- LOW  ----    55%   3.00 --[ ]
              |                         |                            |
              |                         |                            +-- NO ------------------------------------  3.00
              |                         |
              +-- PILOT --   13.35 --( )
                                        |                                              +-- LOW  ----   10%   -10.00
                                        |                            +-- YES ---   26.00 --( )
                                        |                            |                 +-- HIGH ----   90%    30.00
                                        +-- HIGH ----    45%   26.00 --[ ]
                                                                     |
                                                                     +-- NO ------------------------------------  6.00
```

The pilot plant should be built.

17-26 a) Now the payoffs on the "operate by self, with snow-maker" branches become 95, 43, and −9, with an expected value of 43. Hence she should let the hotel operate the resort.

b) In this case, those three payoffs become 96, 48, and 0, with an expected value of 48. She should operate the resort by herself, using the snow-making equipment.

c) If we let the payoffs on those branches be $98 - 10x$, $58 - 50x$, and $18 - 90x$, then the expected value is $58 - 50x$. Setting this equal to the profit gained from letting the hotel chain operate the resort, we get

$$\begin{aligned} 58 - 50x &= 45 \\ 50x &= 13 \\ x &= 13/50 = .26 \end{aligned}$$

Hence if the operating cost increases by 26%, she will be indifferent to those two strategies, since either will earn a profit of \$45,000 for her.

17-27

```
                                                                                          SUCCESS
          PREVIEW             PREVIEW                       TYPE OF                       LEVEL
PROFIT    DECISION   PROFIT   RESULT      PROB   PROFIT     RELEASE    PROFIT                        PROB     PROFIT
                                                                                       +- SMASH      30.00%     22
                                                                                       |
                                                         +- LIMITED     7.200 -( )  MODEST     40.00%      9
                                                         |                             |
                                                         |                             +- BOMB       30.00%    -10
          +- NO --------------------------------- 7.200 -[ ]
          |                                              |                             +- SMASH      30.00%     12
          |                                              |                             |
          |                                              +- GENERAL     6.200 -( )  MODEST     40.00%      8
          |                                                                            |
          |                                                                            +- BOMB       30.00%     -2
          |
          |                                                                            +- SMASH       8.57%     22
          |                                                                            |
8.200 -[ ]                                               +- LIMITED     0.342 -( )  MODEST     40.00%      9
          |                                              |                             |
          |                                              |                             +- BOMB       51.43%    -10
          |                   +- GOOD      35%    3.200 -[ ]
          |                   |                          |                             +- SMASH       8.57%     12
          |                   |                          |                             |
          |                   |                          +- GENERAL     3.200 -( )  MODEST     40.00%      8
          |                   |                                                        |
          |                   |                                                        +- BOMB       51.43%     -2
          +- YES ----  8.200 -( )
                              |                                                        +- SMASH      41.54%     22
                              |                                                        |
                              |                          +- LIMITED    10.893 -( )  MODEST     40.00%      9
                              |                          |                             |
                              |                          |                             +- BOMB       18.46%    -10
                              +- EXCELLENT  65%  10.893 -[ ]
                                                         |                             +- SMASH      41.54%     12
                                                         |                             |
                                                         +- GENERAL     7.816 -( )  MODEST     40.00%      8
                                                                                       |
                                                                                       +- BOMB       18.46%     -2
```

a) See the first 6 branches of the tree above.

b) They should use a limited first-run release.

c) With perfect information, their expected profit would be

$$.3(22) + .4(9) + .3(-2) = 9.6.$$

Hence EVPI = 9.6 − 7.2 = 2.4, i.e., \$2,400,000.

d) Assuming "Claws" is previewed, we use Bayes' Theorem to find the revised probabilities in the tree:

Preview Rating	True Level	P(level)	P(rating | level)	P(rating & level)	P(level | rating)
	Smash	.3	.10	.03	.03/.35 = .0857
Good	Modest	.4	.35	.14	.14/.35 = .4000
	Bomb	.3	.60	.18	.18/.35 = .5143
			P(good) = .35		
	Smash	.3	.90	.27	.27/.65 = .4154
Excellent	Modest	.4	.65	.26	.26/.65 = .4000
	Bomb	.3	.40	.12	.12/.65 = .1846
			P(excellent) = .65		

Since 8.20 − 0.75 = 7.45 > 7.2, they should use the preview. If the preview ratings are "good," they should use general distribution. However, with "excellent" preview ratings, "Claws" should be given a limited first-run release. They will pay up to 8.2 − 7.2 = 1, i.e., $1,000,000 for the previews.

17-28

```
LATE                                    LATE                              LATE
TIME   UTILITY             DECISION     TIME    UTILITY           PROB    TIME    UTILITY
====   =======             ========     ====    =======           ====    ====    =======
                                                          +---     20%     10       95
                                                          |
                 +--- BUS               15.50   82.50 ---( )--     50%     15       85
                 |                                        |
                 |                                        +---     30%     20       70
                 |
                 |                                        +---     80%     15       85
                 +--- WALK              16.00   82.00 ---( )
                 |                                        +---     20%     20       70
                 |
14.00  86.00 ---[ ]                                       +---     50%     10       95
                 |                                        |
                 +--- BIKE              14.00   86.00 ---( )--     40%     15       85
                 |                                        |
                 |                                        +---     10%     30       45
                 |
                 |                                        +---     30%     10       95
                 |                                        |
                 |                                        +---     45%     15       85
                 +--- CAR               15.25   83.25 ---( )
                                                          +---     15%     20       70
                                                          |
                                                          +---     10%     25       60
```

a) To minimize expected late time, Sam should ride his bike.

b) To maximize expected utility (test score), Sam should also ride his bike.

17-29 a)

```
INCREASED                          INCREASED                         INCREASED
 REVENUE            OPTION          REVENUE                 PROB      REVENUE
=========          ========       ==========                ====     =========
                                                   +---     70%         2.8
                    TOTAL                          |
              +--  REDESIGN          0.6      ---( )--      20%        -3.2
              |                                    |
              |                                    +---     10%        -7.2
              |
              |      NEW                           +---     80%         1.3
  0.7      -[ ]-   RUNWAY            0.7      ---( )
              |                                    +---     20%        -1.7
              |
              |                                    +---     65%        -1.0
              +--  NOTHING          -2.05     ---( )
                                                   +---     35%        -4.0
```

b) Remodelling with the new runway is the best option.

c) With perfect information about the totally redesigned airport, they would choose that option if it would be successful and choose to remodel otherwise. Hence their expected return with perfect information is $0.7(2.8) + 0.3(0.7) = 2.17$, and so

$$\text{EVPI} = 2.17 - 0.7 = 1.47, \text{ i.e., } \$1,470,000.$$

e) With perfect information about the remodelled airport, they would choose that option if it would be successful and choose to totally redesign otherwise. Hence, their expected return with perfect information is $0.8(1.3) + 0.2(0.6) = 1.16$, and so

$$\text{EVPI} = 1.16 - 0.7 = 0.46, \text{ i.e., } \$460,000.$$

17-30 a) Let x be the number of print heads after which production is cheaper than purchase.

Then,
$$\begin{aligned} 24x + 28000 &\leq 35x \\ 28000 &\leq 11x \\ 2545.5 &\leq x \end{aligned}$$

Now 2545 print heads corresponds to $2545/1.15 = 2213$ units.

$$\text{P(demand} \geq 2213) = \text{P}\left(z \geq \frac{2213 - 3000}{700}\right) = \text{P}(z \geq -1.12) = .8686$$

b) Breakeven plus 1.5 standard deviations is $2213 + 1.5(700) = 3263 > \mu$, so the probability of being this far above breakeven is less than 50%. They should buy the modules.

17-31 On the surface it appears that all the necessary ingredients are present for analysis by decision theory. There are several possible states of nature and courses of action that Sarah might follow: the number of jars that might be bought by customers, and the number of jars she could stock. Sarah has the necessary data to compute payoffs, and the situation is uncertain. The factor that is not present is the objective; and without that, we simply cannot solve the problem. When Sarah is torn between satisfying her customers or adopting a more cautious purchasing policy, she makes it impossible to make the decision. Should she decide that customer service is her objective, she should choose to purchase eighteen jars--or more--if she so desires. If, however, she chooses some other decision criterion, then the outcome is not known until she makes that choice of her objective. For Sarah to solve her problem, then, is a two-step process--first choosing the objective and then making the decision in that light.

17-32 a) ML = 14.70 MP = 26.95 − 14.70 = 12.25

$$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{14.7}{26.95} = .545,$$ so he should order 44 tails (22 entrées).

b) Ordering 44 tails, his expected profit is

$$26.95[.07(18) + .09(19) + .11(20) + .16(21) + .57(22)] - 22(14.70) = \$244.44$$

Hence $\text{EVPI} = 12.25[.07(18) + .09(19) + .11(20) + .16(21) + .20(22) + .15(23) + .14(24) + .08(25)] - 244.14 = \$21.88$,

if requiring orders in advance doesn't change the demand distribution.

17-33 $7/24 = .2917$, corresponding to 0.81σ, so $\sigma = 6/.81 = 7.41$ hours. To pay for itself, the spreader will have to save $43.50/12.50 = 3.48$ hours or 208.8 minutes of labor. To do this, it will have to be used for

$208.8/8 = 26.1$ hours, corresponding to $z = \frac{26.1 - 48}{7.41} = -2.96$.

$P(z \geq -2.96) = .9985$. Hence, the probability that the spreader will pay for itself before it is scrapped is .9985.

17-34 ML = 26.00 MP = 42.75 − 26.00 = 16.75

$$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{26.00}{42.75} = .608,$$ so he should order 35 bags.

17-35 $2/10 = .2$, corresponding to 0.524σ, so $\sigma = 200/.524 = 381.68$ ties.

a) To break even, Archdale must sell $16500/3.50 = 4714$ ties, corresponding to

$z = \frac{4714 - 5000}{381.68} = -.75$. $P(z \geq -.75) = .7734$. Hence the probability that they at least break even is .7734.

b) To earn 10 percent, their revenues must be $16500(1.1) = \$18{,}150$, which they can get by selling $18150/3.50 = 5186$ ties, corresponding to

$z = \frac{5186 - 5000}{381.68} = .49$. $P(z \geq .49) = .3121$. Hence the probability of earning at least 10 percent on this investment is .3121.

17-36 a) The three numbers at some nodes are the costs in parts b, d.i, and d.ii.

```
           CASE #1                   TRIAL                      CASE #2                  TRIAL
COST       DECISION    COST          RESULT   PROB   COST       DECISION    COST         RESULT   PROB   COST
----       --------    ----          ------   ----   ----       --------    ----         ------   ----   ----
                                                                                                          95.0
                                                               +-- SETTLE ---------------------------   120.0
                                                       95.0    |                                         165.0
        +-- SETTLE ---------------------------------   120.0 --[ ]
        |                                              136.5    |   GO TO                +-- WIN ---  40%   82.5
        |                                                      +-- TRIAL -- 136.5 --( )
        |                                                                                +-- LOSE --  60%  172.5
        |
87.5    |                                                                                                  27.5
99.0 --[ ]                                                     +-- SETTLE ---------------------------    52.5
115.2   |                                              27.5    |                                          97.5
        |                      +-- WIN ---  60%        30.0 --[ ]
        |                      |                       30.0    |   GO TO                 +-- WIN ---  80%   12.0
        |                      |                               +-- TRIAL --  30.0 --( )
        |   GO TO     87.5     |                                                         +-- LOSE --  20%  102.0
        +-- TRIAL --  99.0 --[ ]
                     115.2     |                                                                          177.5
                               |                               +-- SETTLE ---------------------------   202.5
                               |                      177.5    |                                         247.5
                               +-- LOSE --  40%       202.5 --[ ]
                                                      243.0    |   GO TO                 +-- WIN ---  10%  162.0
                                                               +-- TRIAL -- 243.0 --( )
                                                                                         +-- LOSE --  90%  252.0
```

b) He should take #1 to trial. If he wins #1, he should take #2 to trial; but, if he loses #1, he should settle #2 out of court.

c)

```
                    MOCK TRIAL
COST                RESULT          PROB   ACTION          COST
====                ==========      ====   ============    ====
            +-- WIN #1      ----    60%    TRY BOTH         30.0
66.0 --( )
            +-- LOSE #1     ----    40%    SETTLE BOTH     120.0
```

He would pay up to 99 − 66 = 33, i.e., $33,000 for an absolutely reliable mock trial.

d) i. He would take #1 to trial, but settle #2 regardless of the outcome of the trial.
ii. He would take #1 to trial, and take #2 to trial regardless of the outcome of the first trial.

17-37 To earn 13% on an investment of $500,000, their profits must be at least $65,000. Necessary sales levels are:

1. At $130/set: $65000 = -125000 + (130 - 80)x$
$x = 190000/50 = 3{,}800$ units

2. At $140/set: $65000 = -125000 + (140 - 80)x$
$x = 190000/60 = 3{,}167$ units

Then the probabilities of earning 13% are:

1. At $130/set: $P(x \geq 3800) = P\left(z \geq \dfrac{3800 - 4000}{450}\right) = P(z \geq -.44) = .6700$
2. At $140/set: $P(x \geq 3167) = P\left(z \geq \dfrac{3167 - 3200}{300}\right) = P(z \geq -.11) = .5438$

They should proceed with a selling price of $130/set.

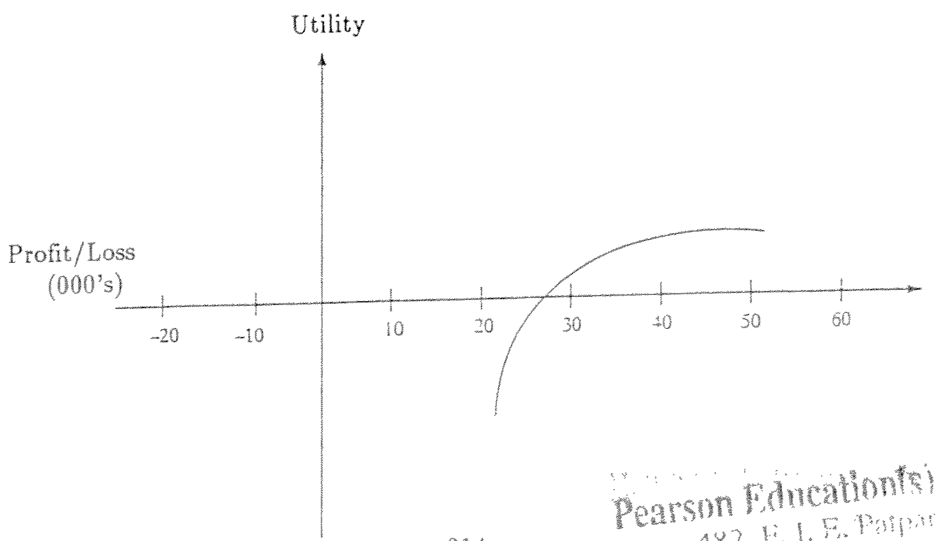
17-38 ML = 21.50 − 19.95 = 1.55 MP = 43.95 − 21.50 = 22.45

$$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{1.55}{24.00} = .065,$$ so she should stock 25 suits.

17-39 ML = 325 MP = 600 − 325 = 275

$$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{325}{600} = .542,$$ so they should order 8 − 2 = 6 more washer-dryer combinations.

17-40



17-41

```
                        EXTEND                              RISK
PROFIT                  CREDIT?   PROFIT                    CLASS   PROB   PROFIT
======                  ========  ======                    =====   ====   ======
                    +-- NO ------------------------------------------------  0.00
                    |
 10.60 --[ ]                                           +-- POOR -- 25%   -20.00
                    |                                  |
                    +-- YES ----   10.60 --( )--AVG.  -- 45%    18.00
                                                       |
                                                       +-- GOOD -- 30%    25.00
```

They should extend credit.

17-42 a) Assuming they take the analysis, we use Bayes' Theorem to find the revised probabilities:

Agency Rating	True Category	P(category)	P(rating | category)	P(rating & category)	P(category | rating)
	Poor	.25	.1	.025	.025/.25 = .10
A	Average	.45	.1	.045	.045/.25 = .18
	Good	.30	.6	.180	.180/.25 = .72
				P(A) = .250	
	Poor	.25	.2	.050	.05/.5 = .10
B	Average	.45	.8	.360	.36/.5 = .72
	Good	.30	.3	.090	.09/.5 = .18
				P(B) = .500	
	Poor	.25	.7	.175	.075/.25 = .70
C	Average	.45	.1	.045	.015/.25 = .18
	Good	.30	.1	.030	.030/.25 = .12
				P(C) = .250	

```
          PURCHASE             AGENCY                      EXTEND                  RISK
PROFIT    RATING?   PROFIT     RATING  PROB  PROFIT        CREDIT?  PROFIT         CLASS   PROB  PROFIT
======    ========  ======     ======  ====  ======        =======  ======         =====   ====  ======
                                                      +-- NO --------------------------------    0.00
                                                      |
                            +---- A --   25%  19.24 --[ ]                      +-- POOR -- 10%  -20.00
                            |                         |                        |
                            |                         +-- YES ----  19.24 --( )--AVG.  -- 18%   18.00
                            |                                                  |
                            |                                                  +-- GOOD -- 72%   25.00
                            |
                            |                         +-- NO --------------------------------    0.00
                            |                         |
       +-- YES ----  12.54 --( )--- B --   50%  15.46 --[ ]                    +-- POOR -- 10%  -20.00
       |                    |                         |                        |
       |                    |                         +-- YES ----  15.46 --( )--AVG.  -- 72%   18.00
       |                    |                                                  |
       |                    |                                                  +-- GOOD -- 18%   25.00
       |                    |
       |                    |                         +-- NO --------------------------------    0.00
       |                    |                         |
12.54 --[ ]                 +---- C --   25%   0.00 --[ ]                      +-- POOR -- 70%  -20.00
       |                                              |                        |
       |                                              +-- YES ----  -7.76 --( )--AVG.  -- 18%   18.00
       |                                                                       |
       |                                                                       +-- GOOD -- 12%   25.00
       |
       |                                              +-- NO --------------------------------    0.00
       |                                              |
       +-- NO -----------------------------  10.60 --[ ]                       +-- POOR -- 25%  -20.00
                                                      |                        |
                                                      +-- YES ----  10.60 --( )--AVG.  -- 45%   18.00
                                                                               |
                                                                               +-- GOOD -- 30%   25.00
```

Since $12.54 - 10.60 = 1.94 > 0.75$, they should purchase the credit rating.

b) Credit should be granted with an A or B rating, but not with a C rating.

c) $12.54 - 10.60 = 1.94$, so they will pay at most \$1940 for the rating.

d)

```
                     RISK            EXTEND
PROFIT               CLASS   PROB    CREDIT?   PROFIT
======               =====   ====    =======   ======
                 +-- POOR -- 25%   -- NO --      0.00
                 |
15.60 --( )-     AVG.  -- 45%   -- YES -     18.00
                 |
                 +-- GOOD -- 30%   -- YES -     25.00
```

$15.60 - 10.60 = 5$, so they would pay up to \$5000 for an absolutely reliable credit report.

17-43 a) John's contribution from sports fishing is
$150(500 - 135) = \$54{,}750$ if the weather is good, and
$105(500 - 135) = \$38{,}325$ if the weather is bad.

```
                  TYPE OF                        ACTUAL
PROFIT            FISHING      PROFIT            WEATHER     PROB      PROFIT
======            =======      ======            =======     ====      ======
                                              +-- GOOD --   70.00%    50.0000
               +-- TUNA ---   47.9000 --( )
               |                              +-- BAD ---   30.00%    43.0000
49.8225 --[ ]
               |                              +-- GOOD --   70.00%    54.7500
               +-- SPORT --   49.8225 --( )
                                              +-- BAD ---   30.00%    38.3250
```

John should rent the Jolly Roger out for sport fishing.

b) With perfect information, John's profit would be $.7(54.75) + .3(43) = 51.225$.
Hence EVPI $= 51.225 - 49.8225 = 1.4025$, i.e., \$1402.50.

c) Assuming John buys the forecast, we use Bayes' Theorem to find the revised probabilities:

Forecast Weather	Actual Weather	P(actual)	P(forecast | actual)	P(forecast & actual)	P(actual | forecast)
Good	Good	.7	.9	.63	$.63/.66 = .9545$
	Bad	.3	.1	.03	$.03/.66 = .0455$
			P(good weather) =	.66	
Bad	Good	.7	.1	.07	$.07/.34 = .2059$
	Bad	.3	.9	.27	$.27/.34 = .7941$
			P(bad weather) =	.34	

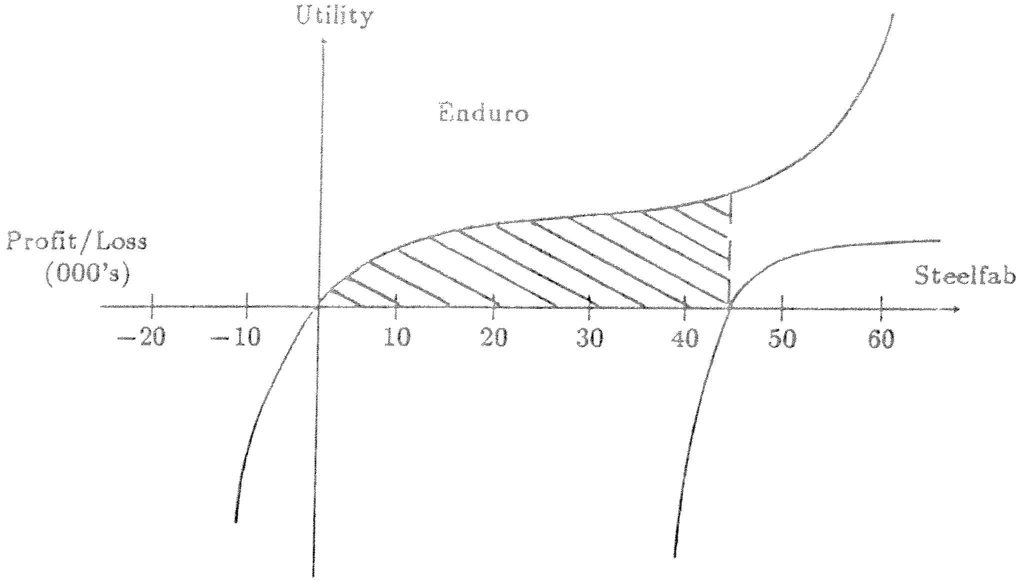
```
               FORECAST                             TYPE OF                       ACTUAL
PROFIT         WEATHER     PROB    PROFIT           FISHING      PROFIT           WEATHER     PROB     PROFIT
======         =======     ====    ======           =======      ======           =======     ====     ======
                                                                               +-- GOOD --   95.45%   50.0000
                                                 +-- TUNA ---   49.6815 --( )
                                                 |                             +-- BAD ---    4.55%   43.0000
            +-- GOOD ---   66%    54.0027 --[ ]
            |                                    |                             +-- GOOD --   95.45%   54.7500
            |                                    +-- SPORT --   54.0027 --( )
            |                                                                  +-- BAD ---    4.55%   38.3250
50.7517 --( )
            |                                                                  +-- GOOD --   20.59%   50.0000
            |                                    +-- TUNA ---   44.4413 --( )
            |                                    |                             +-- BAD ---   79.41%   43.0000
            +-- BAD ----   34%    44.4413 --[ ]
                                                 |                             +-- GOOD --   20.59%   54.7500
                                                 +-- SPORT --   41.7069 --( )
                                                                               +-- BAD ---   79.41%   38.3250
```

With the forecast, John is $50.7517 - 49.8225 = 0.9292$, i.e., \$929.20 better off. Since this exceeds the \$400 cost of the forecast, John should buy it. If the forecast is for good weather, John should rent the boat out for sports fishing; otherwise he should use if for commercial tuna fishing.

d) Since the forecast is worth only \$929.20, John would not buy it if he didn't get a discount for being Jim's friend.

17-44 a) Expected profit = \$80,000; therefore 1/2 of \$80,000 = \$40,000
\$40,000 ÷ 500,000 = 8% return.
Enduro would accept, but the 8% return is below Steelfab's 9% cutoff.

b)



c) 55/500 = 11% return, which is acceptable to both. Steelfab would bid up to \$611,111, where the \$55,000 equals a 9% return.

17-45

Decision	Expected payoff
Market the High Jump	$-42000(.5) + 26000(.3) + 71000(.2) = \$1000$
Do not market it	$0(.5) + 0(.3) + 0(.2) = \$0$

Sporty Sneaker Company should market the High Jump.

17-46 a)

Number of games	Expected payoff
20	$12600(.55) + 11000(.30) + 10600(.15) = \$11,820$
25	$18000(.55) + 16200(.30) + 8500(.15) = \$16,035$
35	$23000(.55) + 15000(.30) + 7100(.15) = \$18,215$

He should have 35 video games in the Cincinnati amusement center.

b) $\text{EVPI} = 23000(.55) + 16200(.30) + 10600(.15) - 18215 = \$885$.

17-47 a)

Type of text	Expected payoff
University	$8000(.7) + 16000(.3) = \$10,400$
Outside	$13000(.7) + 13000(.3) = \$13,000$

b) They should use outside textbooks.

17-48 MP = selling price − cost = 1.50 − 0.70 = 0.80
ML = cost = 0.70

$$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{0.70}{0.80 + 0.70} = 0.4667$$

Number of units	Probability of selling at least this many
500	.10 + .12 + .15 + .33 + .30 = 1.00
600	.12 + .15 + .33 + .30 = 0.90
700	.15 + .33 + .30 = 0.78
800	.33 + .30 = 0.63
900	.30 = 0.30

Since .63 > p^* > .30, Records and Tapes Unlimited should order 800 copies.

17-49 MP = selling price − cost = 0.75 − 0.35 = 0.40
ML = cost = 0.35

$$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{0.35}{0.40 + 0.35} = 0.4667,$$

which corresponds to 0.08σ, so they should order $\mu + 0.08\sigma = 960 + 0.08(140) = 971.2$, or $971(8) = 7768$ ounces (approximately 61 gallons)

17-50 a)

Number of beds	Expected profit
50	41000(.2) + 52000(.3) + 65000(.5) = \$56,300
75	− 12000(.2) + 68000(.3) + 80000(.5) = \$58,000
50	− 53000(.2) − 24000(.3) + 117000(.5) = \$40,700

He should build a 75-bed facility.

b) Expected profit with perfect information = 41000(.2) + 68000(.3) + 11700(.5) = \$87,100

c) EVPI = 87100 − 58000 = \$29,100

17-51

Order action	Expected payoff
Order more	6100(.65) + 1500(.35) = \$4490
Don't	0(.65) + 0(.35) = \$0

She should order more sweatshirts.

17-52 Since no salvage value is given, assume any unsold machines are worthless.
MP = selling price − cost = 89.95 − 75.50 = 14.45
ML = cost = 75.50

$$p^* = \frac{\text{ML}}{\text{MP} + \text{ML}} = \frac{75.50}{14.45 + 75.50} = 0.8394$$

Number of units	Probability of selling at least this many
15	.12 + .17 + .26 + .23 + .15 + .05 + .02 = 1.00
16	.17 + .26 + .23 + .15 + .05 + .02 = 0.88
17	.26 + .23 + .15 + .05 + .02 = 0.71

Since .88 > p^* > .71, Phones and More should order 16 phones.

17-53 The expected payoff for doubling the order is

$$0.15(240,000) + 0.85(20,000) = 53,000, \quad \text{and}$$

the expected payoff for not doubling the order is

$$0.15(100,000) + 0.85(80,000) = 83,000.$$

The owner should stick with the usual order.

17-54 a) $3150(1.08) = \$3402$

b.i) $100[0.25(25) + 0.50(35) + 0.25(50)] = \$3625$

ii) $525[0.25(0) + 0.50(35-30) + 0.25(50-30)] = \$3937.50$

c) They should sell now and use the proceeds to buy LEAPs, since that has the highest expected value.

STATISTICS AT WORK

DISCUSSION OF CONCEPTUAL CASE EXERCISES

ARRANGING DATA

The easy collection of large amounts of data by computers doesn't mean to say that the numbers are always presented in a manner which conveys *meaning*. The situation described in this fast-growing business is not unusual. Many firms have a great deal of data but very little information or analysis.

For some students, this will be the first exposure to a "case". Most should be able to follow the study questions and come to class armed with suggestions. But a few will balk at making suggestions because they haven't seen the data. Like most cases, LOVELAND COMPUTERS requires the student to make some assumptions.

It's important to get across the message that there are many possible answers -- and no single "right" answer -- here. Presumably, Lee would probably approach the problem by considering *operations* data, as well as the obvious financial data. This might include number of units shipped, some cross-tabulations by models and so on. Some information about the number of employees might also be meaningful.

To present the data, Lee will have to decide on a time frame. Collapsing 48 sheets of data into a table with six columns and 48 rows won't help either! Most likely, Lee would decide to organize some old data by year, and then give some detail on recent quarters. Although a student could make a case for presenting some monthly figures in this fast-moving industry, they may be subject to some irrelevant distortions (public holidays and four versus five-week months are examples). Certainly, we'd expect to see some information on the life-span of individual profit lines.

Some students -- not just male students -- will jump to the conclusion that "Lee" is male. Although it is useful to point out an instance where making an assumption could be making a mistake, the broader point is that the instructor has the opportunity to turn to any student in the class and say: "OK, you're Lee -- tell me what you're going to do."

Some students may want to know "what really happened", while others will recognize that the case is fictional. Many of the "facts" are "stolen" from the dramatic real-world successes -- and problems -- of the small computer companies of the 1980s. In class discussion, it may help to tie the fictional case to some examples from current business periodicals. But the most important point to communicate in the discussion of a case early in a course is that "what really happened" may not be as good as some of the suggestions that the students themselves have to offer. •

CASE 3

CENTRAL TENDENCY AND DISPERSION

This short case clearly relates to the concepts in the chapter and students should have no difficulty in coming up with the suggestion that some measure of central tendency of gross margins -- the mean probably makes sense -- and some measure of dispersion -- standard deviation would be fine -- may help. For a business presentation, two graphs -- or two lines plotted on the same axes if the means are sufficiently far apart -- may help to convey the point.

It's possible that the shape of the two curves might be different, with the established products conforming most closely to a normal distribution. The new products could have a "tail" in either direction depending on the number and severity of price cuts to move less-desired products on the one hand, or the ability to charge a true premium price for highly-desired new technology.

The information will help the new investors understand the nature of the industry, with some "old man river" predictable products and some greater variability -- and risks -- with new product introductions. With those risks, come the possibilities of some greater rewards. The investors will be interested to know what new product ideas the firm could proceed with if it had more capital.

Students who have had prior experience with the normal distribution may be uncomfortable assuming that "gross margin" (which is computed as a *percentage*) will necessarily behave in a "normal fashion." The range is limited at the top end at 100 percent (you get the computer for free and sell it -- all of the money is gross margin). And at the lower end, the range is constrained at *minus* 100 percent (you pay good money for the computer but give it away for free). If the average gross margin was close to either of these limits, the distribution would be very different from the familiar bell-shape. However, the indicated 28 percent overall gross margin indicates that these constraints should not affect the solution to this problem.

The "set up" of Walter leaving for the airport probably gives Lee a couple of hours to complete his analysis. Lee has gross margin by product by month already "clean" and in a spreadsheet. It's a simple matter to sort by the product lifetime (reporting month - month of introduction) and then break the data into two sets. A few keystrokes will produce the population statistics desired and the data can be graphed just as easily. Where the instructor has access to the right equipment, this might make an impressive demonstration. Even where students are not expected to have access to computers, instructors can make the point that this information can be generated with a simple business calculator costing less than $20.00. •

CASE 4

PROBABILITY

In class discussion, many students will volunteer that this problem should be approached with a probability tree -- but without experience, they may find it quite difficult to get started. As a conceptual case, it is more interesting to go through the idea of setting up the tree, than to grind through the calculations. However, once the tree(s) are set up, it may help to take a few guesses as to the probabilities involved as this will connect the exercise with decision-making in the "real world."

As a business decision, the probabilities under the two possible situations (Take/Don't Take the Holiday) should be compared. Nancy may be far from reaching her goal with or without the holiday. So, there should be two probability trees covering Take and Not Take. The remainder of the study questions can be approached as follows.

Lee will be computing conditional probabilities by multiplying "sick day" data and "snow day" data. If no student volunteers this point, it is worth noting in discussion that this analysis treats "snow days" and "sick days" as independent events. Although there are probably more of both in January, it is not unreasonable to treat them as independent phenomena.

Lee will need some weather information. Since the company's snow day policy has only been in effect for a short period of time, Lee should gather some snowfall information (such as number of storms with more than 6 inches accumulation) from a secondary source. This is a good time to encourage students to think about what information might be available to Lee, probably at the "expense" of a single phone call to the National Weather Service or the local library.

Finally, this is a good place to introduce a "first look" at the notion of *confidence*, which is treated in greater detail in later chapters. A glance at the probabilities suggests that, even without the King Day off, the production supervisor is unlikely to be "reasonably certain" of having the right number of production days if a very high standard -- 99 percent -- is used. In practice, since the business decision is to balance happy customers with a happy workforce, there are good arguments for setting a lower limit (75 percent) in which there is a "sporting chance" for this proposal. •

PROBABILITY DISTRIBUTIONS

This case sets up a reasonably plausible business situation in which the *normal approximation to the binomial* would give a workable solution. The number of "trials" is sufficiently large -- there are 200 computers in each lot -- and the probability that a modem will be needed is close to 0.5 ("about half the computers ... "), so the conditions for using this approximation are met.

The specific calculations that Lee will make are to determine the mean and standard deviation for the distribution. Looking at a table of normal probabilities will give Lee the "Z-score" for the acceptable probability. It's important to drive home the point that this is a *one-tail* test. So if a 95 percent probability of "success" is desired, the applicable Z-score is 1.64, not 1.96. Lee can then make a simple calculation -- on the back of an envelope -- to give Jeff the suggested order quantity.

Lee needs to know the "success rate," or the level of confidence, for this decision in order to be able to pick a point on the x-axis normal distribution. If Jeff wants to be "right" all the time, the only solution is to order a modem for every computer, and the case indicates that this is not a suitable solution. As a business decision, Jeff can tolerate occasional re-orders for small numbers of modems, but does not want this to be a frequent occurrence. This is similar to many "real world" business situations.

Because the distribution "number of modems per lot" is based on a yes/no decision (the customer does, or does not want a modem) the underlying distribution is the binomial distribution. But, as discussed above, with this large "n," Lee will be thinking about the normal curve as an approximation.

Assuming there are sufficient data to estimate "p" (the probability that a customer wants a modem) in the binder of data shown to Lee by Cohen, no additional data are needed, with the exception of a table of normal distribution probabilities by Z-score. In practice, the modem-rate "p" might trend up or down over time and Lee would want to check the data and re-work the calculation.

In addition to the statistics points made, it is worth emphasizing that, given a handy table of the normal curve, the managerial decision-making for an important business problem can be aided by a simple calculation made in a few minutes. •

SAMPLING AND SAMPLING DISTRIBUTIONS

The main thrust of this case scenario is to introduce students to a situation where sampling makes sense -- testing all of these components takes up too much time, but assuming "zero defects" for every shipment results in re-work if a component is found to be defective. Although modern concepts of manufacturing quality (covered in greater detail in chapter 10) would suggest that Loveland should develop consistent supplier relationships and aim for zero defects in component parts, there are some business situations where these quality ideals have not been achieved. In the meantime, some form of *acceptance sampling* seems prudent.

Most students will readily come up with the idea that some number of the components should be tested -- but what is the right number? More importantly, does it make sense to test the first n disk drives, or one from each box? Should the drives be selected at random, or according to some pre-determined scheme?

Lee will want some information about the usual distribution of *access times* from a "good lot," and a cut-off figure for an unacceptably slow access time (where the lot would be rejected.) This example will help students see that the sample mean (and distribution of possible sample means) reflect some underlying "true" population value for the lot being examined.

Whatever sampling scheme is adopted, the narrative has suggested that shipments are either fast or slow drives. Clearly a single shipment could contain both types of drives and the variance of measured access times would increase as well as a shift in the mean. This is a detail which may confuse some students at this point and it is best to focus discussion on the underlying mean access time for different lots.

An additional point which students may bring up is the option of taking a larger sample if the first sample mean is in a "gray area." In practice, some acceptance sampling is conducted in this way, with pre-determined rules for switching from an initial sample to a larger sample, and eventually examination of the entire shipment. In this case, this might get dangerously close to the error of "optional stopping," which would occur if, for example, a technician decided to stop sampling if the first few drives tested well within the established parameters. Clearly, this discussion goes well beyond the scope of an initial introduction to sampling. •

CASE 7

ESTIMATION

This situation presents an opportunity for students to think about how samples *estimate* true underlying population values. The key point is that we can calculate an expectation that an estimate will be accurate within a particular range. This decision is made because we can assume the shape of the underlying distribution -- in this case, most people would assume that total software expenditures would be roughly normally distributed.

The 500 telephone surveys are a *sample* of a population ("all people who bought LOVELAND CCOMPUTERS' machines in a certain time frame.") This number is *finite* but large, so treating the underlying population as *infinite* will not affect the result. The t-distribution is appropriate for this situation. Of course, some students may have planned for Lee to use the normal distribution table, but even though the underlying distribution is assumed normal, the distribution of sample means is the t-distribution.

Lee will help Margot to define "margin of error" as: the sample mean will fall in the following range "x" percent of the time. The compliment of this percentage is the confidence level for a two-tail test. For a business decision of this nature, the 90 percent confidence level might well be sufficient, although some may argue for a tougher standard such as "95 percent sure."

A sample of 500 customers may well be sufficient -- if the variance of the sample is small, a fact not given in the case. Some students may wonder whether the survey would yield a spuriously large variance because there is not one population, but two -- business customers and home users -- represented in the sample. This could mean two separate underlying variables, each with different means. This situation is addressed as the case continues. •

CASE 8

ONE-SAMPLE TESTS OF HYPOTHESES

Margot is not right in believing that Lee also needs to know the population standard deviation. Lee can estimate this from the sample standard deviation using the methods shown on pages 402-404.

Margot is exploring whether her customers are different from those mentioned in the Wall Street Journal article. She is probably doing this as a basis for some target marketing, i.e., identifying a segment within the market.

The idea would be stated in hypothesis testing terms thusly:

$$H_0:\ \mu = (\$)$$

(The average amount spent by people on software during the first year they own a machine is as reported in the Wall Street Journal article.)

$$H_1:\ \mu \neq (\$)$$

(The average amount spent by people on software during the first year they own a machine is different from that value reported in the Wall Street Journal article.)

TWO-SAMPLE TESTS OF HYPOTHESES

The key to this situation is that managers make informal hypotheses frequently in business. Here, inspection of the data - always a good thing to do as Margot now ruefully admits -- shows that there are probably two underlying distributions. On a blackboard or overhead, it might be useful to begin class discussion by drawing two very separate bell-shaped distributions on an x-axis. They clearly look as if data are being drawn from two different populations. But as the two curves are moved closer together, the resulting curve looks like the twin humps of a camel, as the two middle tails overlap. When data are plotted as a single graph, this summing of two distributions may mask the existence of two or more separate populations. In this case, Lee was "tipped off" because the variance was much higher than expected.

The graph that Lee made was a histogram showing the frequency with which respondents' total software purchases fell within certain ranges -- students who are not familiar with personal computing may be impressed with a demonstration that this can be achieved with a few keystrokes and that the worksheet program does "all the work."

Lee is going to test the hypothesis that there are two different distributions and that the mean expenditure for business customers is greater than that of home users. Because the hypothesis is "greater than" (and not "different from"), a one-tailed test can be used.

As a concluding comment, it may be worth reflecting on how business people have hunches -- hypotheses -- about their own target markets. Margot had already placed each customer into one of two different target markets. A full understanding of the differing needs of the two different markets will help her develop successful marketing strategies. •

QUALITY AND QUALITY CONTROL

This situation gives students an opportunity to apply the general principles of quality management, discussed in the chapter, to the specific situation of LOVELAND COMPUTERS.

A lively classroom discussion will probably bring out a range of opinions concerning Walter Azko's assertion that LOVELAND competes on price alone. The success of the mail-order computer companies probably relates more to *value* than to price. The most successful companies certainly offer discounts, but they are not the cheapest suppliers in their category. Customers estimate value as follows:

$$\text{Value} = \frac{\text{Quality}}{\text{Price}}$$

The total costs of replacing a defective machine include more than just the cost of goods sold ("COGS") for the computers that are thrown out. All of the packing, shipping, and handling costs are wasted, as are the financial and storage costs of carrying excess inventory. (The instructor can make this point by asking students to think about a firm with a very high defect rate. At 50 percent defects, half the costs of the firm go to receiving, storing and shipping goods that are eventually replaced.) And the customer incurs the costs of returning the defective machine (including the labor costs that go in to repacking the return) the costs of having money invested in a useless asset and the lost value of a working machine. More importantly, no matter how pleasant LOVELAND's staff are about handling the replacement, each defective machine "costs" LOVELAND's brand equity (image in the marketplace.) Students may offer additional suggestions.

This leads to the discussion of the interaction of quality and customer satisfaction. Firms that embrace the quality message do more than just produce to specification -- they are able to anticipate customer *requirements*. A good example of this is the difference between a perfectly made briefcase - size VHS Camcorder (it meets specifications) versus an 8 mm Handicam (which better meets consumers' needs for portability as well as for good pictures.)

When firms adopt continuous improvement, greater reliability of each part of a process should mean that end-of-the-line testing uncovers fewer and fewer rejects. In rejecting the quality message, Walter Azko has not even started to think that each defect has a *cause*. He's treating defects and random "acts of God."

It's important that Walter understand the main tenets of quality management. Efforts to instill quality into firms without top management leadership usually fall quickly by the wayside in the rush of business. ●

CASE 11

CHI-SQUARE AND ANOVA

The simple question -- whether a "difficult" question is more likely to come in on the early shift or the late shift -- is answered by accumulating calls in a two-by-two table for Difficulty and Shift. Chi-square is an appropriate test. The instructor can review the rules for Chi-square: that the expected values of the cells are clearly going to be more than 5, and that the rows and the columns require mutually exclusie sorting. (For example, a call cannot be both "easy" and "difficult.")

A feature of Chi-square that can be emphasized is that the test will accommodate quite different total call volumes for the two shifts, or quite different numbers of "difficult" questions.

As additional parameters get thrown into the mix, this clearly becomes a situation where analysis of variance will be useful. Once the cornerstone of many a PhD thesis, a useful ANOVA can now be accomplished on a small personal computer. In this situation, the data-gathering would probably be quite simple, since LOVELAND's telephone support staff undoubtedly log each call on a computer. Lee's value-added will be in organizing the three-dimensional data in a way which can be understood, and then choosing the appropriate tests of significance.

Of the many possible sources of bias in the data collection, the most obvious is that, at peak times, some customers will not be able to get through to the support staff. Some of these calls will be made later -- but some will be lost data. •

SIMPLE REGRESSION AND CORRELATION

This situation calls for the evaluation of some well-chosen weather statistics and the number of defects. Lee would probably want to get the daily highs and lows and run a correlation with the percent of bases that are defective. Percent is necessary to accommodate different production volumes on different days, and we'll assume that the problem is sufficiently severe that these data points are substantially above zero.

If a relationship is found, a simple regression can be done and Lee can generate a line showing how temperature affects defects, superimposed on a scatter plot of the data. Lee will be able to say whether this single factor accounts for a significant portion of the variance observed.

There are a few computational subtleties or difficulties in test selection here. But it may be useful to link the problem-solving shown in this situation with the discussion on quality management in Chapter 10. Nancy has identified "things gone right" and "things gone wrong" and has begun a search -- albeit somewhat unsystematically -- for the cause of the defects. •

CASE 13

MULTIPLE REGRESSION AND MODELING

For students who have even glanced at the chapter, it's clear that a multiple regression is required here. But how to set it up, what factors to consider and what to do with the results, are not so clear.

Class discussion can begin with eliciting a list of all the relevant data that LOVELAND might have available. Call volume and total sales have been mentioned, but sales of specific computer models would also seem to make sense. Someone is likely to note that LOVELAND is probably already capturing data about ads by asking callers, "Where did you here about us?"

As a business -- rather than statistics -- point, it can be emphasized that there's plenty of data already available to be fodder for a regression. Supplying 800 service is a competitive business and phone companies can readily supply call volume information broken down by time of day and place of origination. Magazines and newspapers can provide data on circulation and estimates of "pass through" readership (the number of people who read a magazine that they did not buy.) And, as a computer company, LOVELAND probably has a great deal of additional data already on disk.

This suggests that it may not be necessary to run an "experiment." But companies with less readily available data often do run tests of different advertising tactics. For example, the effectiveness of a full-color ad in one journal can be compared with the yield from a black-and-white ad in another.

The method for building a useful regression -- for say, call-volume response to advertising -- will depend on the instructor's preference. Those who had experience with punch cards and large main-frame computing environments may choose to build a model one factor at a time, using intuition to select the most likely factors. Only factors which make a significant contribution to the model are retained. This has the advantage of building a parsimonious model that is consistent with common sense. The downside to this approach is that a useful factor may be overlooked.

An alternative -- more likely to be used in environments of powerful personal computers -- is to begin with a massive regression that includes all factors that may contribute to the variance of the target parameter. By a process of "backwards elimination," the least significant variables are eliminated and the regression is run again. This technique permits unexpected variables to survive in the final model. But the downside is that it may be very difficult to make intuitive sense from some relationships. (IQ is correlated -- weakly but significantly -- with shoe size. It's not clear how university admissions officers should make use of this "spurious" relationship.) Further, when the original set of variables are not independent (for example, number of ad pages taken out by competitors might be strongly correlated with their price position) some of the variables which are discarded in backwards elimination may have been significant if chosen early in the forwards approach. Thorough statisticians sometimes handle this by running the following cross-check: After a model has been built, the "successful" factors are eliminated from the set of all possible variables and the regression is re-run. None of the remaining variables should have a significant contribution to variance. If they do, a choice between correlated variables must be made, based on a common sense approach to the business question under study.

CASE 14

NONPARAMETRIC METHODS

Although non-parametric statistics may be unfamiliar to some instructors, they can be quite useful in making sense of ordinal data, as this situation shows.

With the small number of respondents (by no means atypical for the expensive process of focus groups) inconclusive results are not unusual. But the firm still has to make a decision and operate! Manufacturers of consumer package goods can proceed to test markets to see how different options perform in the field. But that is not feasible with computers which are not being sold through retail stores. It's possible that the firm's sales records do contain enough information to be able to analyze customers preferences and to detect trends in these choices. But the customers' choices have been constrained because LOVELAND has not offered all three styles in all marketing seasons. So this is probably a decision which will be based on the "art" rather than "science" of management -- guessing the trend of customers' preferences. •

CASE 15

TIME SERIES

A time series analysis will once again make Lee look like a hero. There's enough information in the situation description to predict that Lee should be looking for (1) an underlying trend (exponential growth), (2) a monthly effect of seasonality, and (3) diurnal (daily) variation. Most students will readily accept that day-to-day variation can be factored out by a time series analysis but they may not have seen that Lee can handle both monthly and daily effects in the same analysis.

Lee will have asked for the daily call volume. Once again, LOVELAND is very likely to have captured this on disk. Lee can either include a quadratic function in a regression model to handle the underlying trend, or transform the data using a log function to turn the trend into a straight line.

Bert will use the time series analysis to make predictions about the future. He will probably use the trend to plan the equipment to be able to handle peak volume and then use the seasonality indices to optimally schedule staff.

In real-world situations, companies which are heavily dependent on phone "traffic" (airline reservation systems are a good example) will also analyze call volume by time of day. This analysis is important from an operations perspective because peak volume is often much greater than average volume. A good example of such high volume is the software maker, WordPerfect Corporation. By Spring, 1992 WordPerfect employed 858 operators who answered an average of 16,560 calls per day, with a peak volume of 20,070 calls. •

CASE 16

INDEX NUMBERS

Lee will begin by developing index numbers for three parameters that vary over time: the cost of living in the export country (as a measure of the agent's labor costs), the cost of jet fuel (a commodity that is traded on world markets), and the strength of the dollar. There are indexes of the dollar's strength relative to a "basket of foreign currencies published in the *Wall Street Journal*, and other financial periodicals. But Lee can obtain more specific information from the bank about the specific exchange rate to the Won (South Korea, 1991 about 720 won per $ US), the New Taiwan dollar (1991 about 27 per $ US), or whatever.

The instructor may wish to begin class discussion by asking for suggestions for the base period. Most prudent statisticians would look for a time of relative calm in international markets. Base periods taken immediately before or after a major market shift may yield index numbers that seem inconsistent with common sense. For jet fuel, for example, it would be advisable to choose a base period well after the initial shocks of the OPEC price increases of the early 70s. Fuel costs went up sharply, and then retreated to a new (higher) plateau. Stock brokers often show charts showing the rise of the Dow from the day after the 1929 "crash" to present times -- clearly this overstates the growth that would be expected in a "typical" year.

Index numbers for the three separate factors will give Uncle Walter some bargaining power. But some students will take the idea a step further and come up with a composite index. For example, the weighting might be: 20 percent of the cost of living index, 50 percent of the jet fuel index and 30 percent of the exchange rate index. The actual weights chosen would depend on how much each factor contributes to the agent's costs. The resultant index -- Azko's Shipping Index -- is as valid as the Dow-Jones Industrials, the Consumer Price Index (CPI), or any other index number. And it's much more useful in this particular situation.

Someone may mention that such a "market basket" index is adding "apples and oranges." But this is precisely how indexes such as the CPI are calculated, with weighted conponents for food, housing costs, transportation and even medical care. •

CASE 17

DECISION THEORY

This situation will allow students some in-class practice at building decision trees. Although it is clearly stated that the action will be to Take/Not Take the financing deal offered by the New Yorkers, students unfamiliar with the format of decision trees may see other places to start. For example, some may argue that the state of the economy is the most important factor and should be the first node.

After getting agreement that the *managerial decision* is to decide whether to take the financing offer, it may help to review the implicit consequences of that decision: Walter and Gratia will give up a 60 percent stake in the company; with the money in hand, they will expand capacity; and, at some point in the future, LOVELAND will become a public company.

Now to the uncertainties. Students will know that (1) the economy remained weak in the summer of 1992, and (2) Compaq started a "price war" on PCs, which IBM then joined. If students steer discussion too much towards what "obviously" was going to happen (based on their hindsight) the element of uncertainty can be recreated by asking students to predict these factors for *next* year.

For each branch of the decision trees (Accept/Reject Financing), the next logical step is to add nodes for the economy (Strong/Weak) yielding a total of four branches, and then, branching each again, for Price War/No Price War will give a total of eight end-points -- enough uncertainty to have Walter and Gratia perplexed!

The case doesn't hint at probabilities but it is interesting to see how students can make some intelligent estimates from their own general knowledge. If nothing else, the Price War "event" can be assigned at 50/50 likelihood (although a reasonable argument could be made that the partners would have read enough in the press to give 60/40 odds on a Price War.) Similarly, Walter and Gratia would have formed opinions from news accounts that a swift rebound in the economy was unlikely -- perhaps a 30/70 chance of a Strong economy.

At this point, the academic exercise should convey the "real-world" excitement of the business decision and students may be interested in completing the decision tree with some guesses at the monetary value of the different payoffs.

Students should leave this case with the understanding that, even when a business person says, "Who knows what'll happen to the economy?" most people can, in fact, make intelligent estimates of the probabilities of certain outcomes. And the formal exercise of drawing out the decision tree may clarify the arguments. •

SOLUTIONS TO

COMPUTER DATABASE

EXERCISES

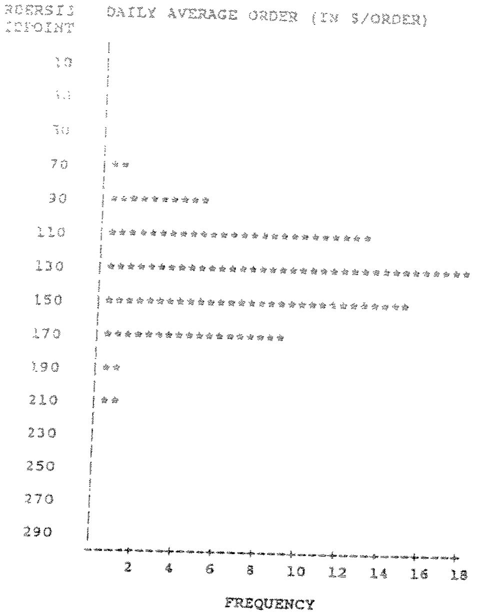
Pearson Education

ARRANGING DATA TO CONVEY MEANING: TABLES AND GRAPHS

HISTOGRAM & FREQUENCY DISTRIBUTION OF DAILY AVERAGE ORDER SIZE

QUARTER=3

FREQUENCY OF ORDERSIZ

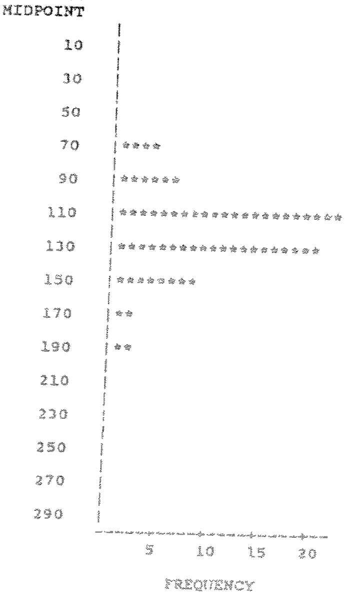


ORDERSIZ MIDPOINT	FREQ	CUM FREQ	PERCENT	CUM PERCENT
10	0	0	0.00	0.00
30	0	0	0.00	0.00
50	0	0	0.00	0.00
70	1	1	1.59	1.59
90	5	6	7.94	9.52
110	13	19	20.63	30.16
130	18	37	28.57	58.73
150	15	52	23.81	82.54
170	9	61	14.29	96.83
190	1	62	1.59	98.41
210	1	63	1.59	100.00
230	0	63	0.00	100.00
250	0	63	0.00	100.00
270	0	63	0.00	100.00
290	0	63	0.00	100.00

HISTOGRAM & FREQUENCY DISTRIBUTION OF DAILY AVERAGE ORDER SIZE

QUARTER=4

FREQUENCY OF ORDERSIZ



ORDERSIZ MIDPOINT	FREQ	CUM FREQ	PERCENT	CUM PERCENT
10	0	0	0.00	0.00
30	0	0	0.00	0.00
50	0	0	0.00	0.00
70	4	4	6.25	6.25
90	6	10	9.38	15.63
110	22	32	34.38	50.00
130	20	52	31.25	81.25
150	8	60	12.50	93.75
170	2	62	3.12	96.87
190	2	64	3.12	100.00
210	0	64	0.00	100.00
230	0	64	0.00	100.00
250	0	64	0.00	100.00
270	0	64	0.00	100.00
290	0	64	0.00	100.00

HISTOGRAM & FREQUENCY DISTRIBUTION OF DAILY AVERAGE ORDER SIZE

------------------------------ QUARTER=1 ------------------------------

FREQUENCY OF ORDERSIZ

ORDERSIZ MIDPOINT	DAILY AVERAGE ORDER (IN $/ORDER)	FREQ	CUM FREQ	PERCENT	CUM PERCENT
10		0	0	0.00	0.00
30		0	0	0.00	0.00
50		0	0	0.00	0.00
70	****	2	2	3.23	3.23
90	****************	8	10	12.90	16.13
110	**********************************	17	27	27.42	43.55
130	**************************************	19	46	30.65	74.19
150	********************	10	56	16.13	90.32
170	********	4	60	6.45	96.77
190	**	1	61	1.61	98.39
210	**	1	62	1.61	100.00
230		0	62	0.00	100.00
250		0	62	0.00	100.00
270		0	62	0.00	100.00
290		0	62	0.00	100.00
	2 4 6 8 10 12 14 16 18				

FREQUENCY

HISTOGRAM & FREQUENCY DISTRIBUTION OF DAILY AVERAGE ORDER SIZE

------------------------------ QUARTER=2 ------------------------------

FREQUENCY OF ORDERSIZ

ORDERSIZ MIDPOINT		FREQ	CUM FREQ	PERCENT	CUM PERCENT
10		0	0	0.00	0.00
30		0	0	0.00	0.00
50		0	0	0.00	0.00
70	*	1	1	1.56	1.56
90	**********	10	11	15.63	17.19
110	************************	24	35	37.50	54.69
130	*******************	19	54	29.69	84.37
150	*****	5	59	7.81	92.19
170	***	3	62	4.69	96.87
190	**	2	64	3.12	100.00
210		0	64	0.00	100.00
230		0	64	0.00	100.00
250		0	64	0.00	100.00
270		0	64	0.00	100.00
290		0	64	0.00	100.00
	5 10 15 20				

FREQUENCY

2.

HISTOGRAM & FREQUENCY DISTRIBUTION OF NUMBER OF ORDERS PER DAY

---------------------------------- QUARTER=3 ----------------------------------

FREQUENCY OF ORDERS

ORDERS MIDPOINT	NUMBER OF ORDERS PER DAY	FREQ	CUM FREQ	PERCENT	CUM PERCENT
105	**	1	1	1.59	1.59
115	**	1	2	1.59	3.17
125	********	4	6	6.35	9.52
135	****	2	8	3.17	12.70
145	****************************	14	22	22.22	34.92
155	********************************	16	38	25.40	60.32
165	**************************	13	51	20.63	80.95
175	********************	10	61	15.87	96.83
185	**	1	62	1.59	98.41
195	**	1	63	1.59	100.00
205		0	63	0.00	100.00
215		0	63	0.00	100.00
225		0	63	0.00	100.00
235		0	63	0.00	100.00
245		0	63	0.00	100.00
	2 4 6 8 10 12 14 16				

FREQUENCY

HISTOGRAM & FREQUENCY DISTRIBUTION OF NUMBER OF ORDERS PER DAY

---------------------------------- QUARTER=4 ----------------------------------

FREQUENCY OF ORDERS

ORDERS MIDPOINT	NUMBER OF ORDERS PER DAY	FREQ	CUM FREQ	PERCENT	CUM PERCENT
105		0	0	0.00	0.00
115		0	0	0.00	0.00
125	****	2	2	3.12	3.12
135	****	2	4	3.12	6.25
145	**********	5	9	7.81	14.06
155	********************	10	19	15.63	29.69
165	******************************	15	34	23.44	53.13
175	******************	9	43	14.06	67.19
185	****************	8	51	12.50	79.69
195	****************	8	59	12.50	92.19
205	******	3	62	4.69	96.87
215	**	1	63	1.56	98.44
225		0	63	0.00	98.44
235		0	63	0.00	98.44
245	**	1	64	1.56	100.00
	2 4 6 8 10 12 14				

FREQUENCY

HISTOGRAM & FREQUENCY DISTRIBUTION OF NUMBER OF ORDERS PER DAY

-------------------------------- QUARTER=1 --------------------------------

FREQUENCY OF ORDERS

ORDERS MIDPOINT	NUMBER OF ORDERS PER DAY	FREQ	CUM FREQ	PERCENT	CUM PERCENT
105	**	1	1	1.61	1.61
115	**	1	2	1.61	3.23
125	**	1	3	1.61	4.84
135	******	3	6	4.84	9.68
145	******	3	9	4.84	14.52
155	******************	9	18	14.52	29.03
165	************************	12	30	19.35	48.39
175	****************************	14	44	22.58	70.97
185	**************	7	51	11.29	82.26
195	********	4	55	6.45	88.71
205	**********	5	60	8.06	96.77
215	**	1	61	1.61	98.39
225	**	1	62	1.61	100.00
235		0	62	0.00	100.00
245		0	62	0.00	100.00

2 4 6 8 10 12 14

FREQUENCY

HISTOGRAM & FREQUENCY DISTRIBUTION OF NUMBER OF ORDERS PER DAY

-------------------------------- QUARTER=2 --------------------------------

FREQUENCY OF ORDERS

ORDERS MIDPOINT	NUMBER OF ORDERS PER DAY	FREQ	CUM FREQ	PERCENT	CUM PERCENT
105	**	1	1	1.56	1.56
115		0	1	0.00	1.56
125		0	1	0.00	1.56
135		0	1	0.00	1.56
145	****	2	3	3.12	4.69
155	**************	7	10	10.94	15.63
165	**************	7	17	10.94	26.56
175	**************************************	19	36	29.69	56.25
185	**************************	13	49	20.31	76.56
195	******************	9	58	14.06	90.62
205	********	4	62	6.25	96.87
215	**	1	63	1.56	98.44
225	**	1	64	1.56	100.00
235		0	64	0.00	100.00
245		0	64	0.00	100.00

2 4 6 8 10 12 14 16 18

FREQUENCY

3. The "average order size" histograms show a generally decreasing trend. For example, the following chart shows that the lower tail is getting heavier (both absolutely and relatively) while the upper tail is getting somewhat lighter.

	1991		1992	
Number of days with	III	IV	I	II
< \$100 per order	6(9.5%)	10(15.6%)	10(16.1%)	11(17.2%)
\$101–160 per order	46(73.0%)	50(78.0%)	46(74.2%)	48(75.0%)
> \$161 per order	11(17.5%)	4(6.3%)	6(9.7%)	5(7.8%)

The "number of orders per day" histograms, on the other hand, show a very strong increasing trend, as illustrated by the following chart.

	1991		1992	
Number of days with	III	IV	I	II
≤ 170 orders	51(81.0%)	34(53.1%)	30(48.4%)	17(26.6%)
≥ 171 orders	12(19.0%)	30(46.9%)	32(51.6%)	47(73.4%)

There are several possible explanations for these trends:

1) The number of customers buying from HH Industries is increasing, and their initial purchases are relatively small. With proper "nurturing," these customers should buy increasingly larger quantities as their confidence in the company's quality and service grows.
2) HH Industries' established customers are recognizing the expense of maintaining large inventories. Consequently, their orders will be more frequent and for smaller amounts than before.
3) Large construction or waste management firms which have traditionally maintained their own fleets of equipment may be tending more towards contracting for repairs with smaller service shops.
4) In addition, there might be a seasonal trend present, which would make sense since adverse winter weather might cause construction slowdowns. (This concept will be addressed in Chapter 15.)

CHAPTER 3

MEASURES OF CENTRAL TENDENCY AND DISPERSION IN FREQUENCY DISTRIBUTIONS

1. Using the raw data, we get the following measures of central tendency:

Average Order Size

	1991		1992	
	III	IV	I	II
Mean	133.9	120.7	126.0	119.9
Median	131.7	120.1	125.7	116.7

Daily Number of Orders

	1991		1992	
	III	IV	I	II
Mean	155.8	171.7	171.3	177.8
Median	156.0	168.5	171.5	177.0

Using the grouped data from the frequency distributions computed in Chapter 2, the measures of central tendency are:

Average Order Size

	1991		1992	
	III	IV	I	II
Mean	134.4	121.3	125.5	120.6
Median	133.3	119.6	123.7	117.1
Mode	132.5	117.8	123.6	114.7

Daily Number of Orders

	1991		1992	
	III	IV	I	II
Mean	155.2	171.3	170.2	177.8
Median	155.6	168.3	170.4	177.6
Mode	154.0	164.5	172.2	176.7

The raw data produces means and medians which are very close together, in which case either can be used as the primary measure of central tendency. The numbers calculated from the frequency distributions exhibit similar characteristics. Choosing the median as our best representative of central tendency, it is fairly obvious that the trends of the median generally support the conclusions drawn from the histograms.

The total sales figure has, in fact, increased steadily over the past year. However, when divided by the number of days in each quarter for an average daily sales rate, the results are less promising. While there is a net 2.75% increase over the entire 12-month period, the quarterly numbers show fluctuations.

	1991 III	1991 IV	1992 I	1992 II
Total sales	$1,310,788	$1,327,508	$1,346,084	$1,368,181
Number of days	63	64	62	64
Daily mean	$20806	$20742	$21711	$21378

2. Laurel is correct in wondering about the contribution of each profit center. The Pennsylvania warehouse, for instance, is actually seeing an increasing in the average number of orders per day, while the average dollar value per order, though fluctuating, is steadier than the corporate trend seems to show.

Profit Center 3 – Pennsylvania

	1991 III	1991 IV	1992 I	1992 II
Mean order size	98	102	101	95
Mean daily orders	23	31	31	35

3. Average Order Size

	1991 III	1991 IV	1992 I	1992 II
Interquartile range	38.4	27.4	33.9	20.3
Total range	146.5	120.8	125.6	120.2

4.

	1991 III	1991 IV	1992 I	1992 II
Number of Orders				
Variance	281.3	467.7	548.4	336.9
Std. deviation	16.8	21.6	23.4	18.4
Coeff. of Var.	10.8%	12.6%	13.7%	10.3%
Average Order Size				
Variance	774.1	582.4	702.7	597.1
Std. deviation	27.8	24.1	26.5	23.4
Coeff. of Var.	20.8%	20.0%	21.0%	20.4%

5. a) The Chebyshev ranges for 75% of the data are "mean $\pm$ 2 standard deviations," which are given by:

Number of Orders	Average Order Size
141.0 – 214.6	71.1 – 168.7

b)

	Number of Orders	Average Order Size
Total observations	64	64
$\times$ .75	48	48
Actual number in range	62 or 63	between 59 and 62

As usual, the Chebyshev ranges are very conservative and much wider than necessary to cover 75% of the data.

6. The coefficients of variation are:

Average order size	Florida	29.0%
	Arizona	37.7%
	Pennsylvania	32.6%
Number of orders	Florida	16.6%
	Arizona	21.9%
	Pennsylvania	21.8%

It appears that the Florida warehouse experiences the smallest relative dispersion, as might be expected since it is both the oldest and exists in the climate least likely to experience significant seasonal variations. It is interesting to note that the average order size numbers are definitely more widely dispersed than the daily number of orders.

7. Histograms are quite understandable, even to people who are unfamiliar with statistics. While the average order values don't show a trend as clearly as those for the number of orders, Laurel can nevertheless make her point that the "daily figure" shouldn't be the only number of interest to the staff in keeping track of the health of the business. In particular, each profit center has unique characteristics and trends which should be carefully followed. The ideas proposed by Laurel to Stan as potentially explaning the trends should be investigated, and promotions designed to take advantage of the information gathered.

PROBABILITY I: INTRODUCTORY IDEAS

1.

New England	.1652
Northeast	.4503
Southeast	.0516
Midwest	.2877
North Central	.0168
South Central	.0116
West	.0168

2. P(New England + Northeast + Midwest) =
P(New England) + P(Northeast) + P(Midwest) = .9032

3. P(Next Day Air) = .2297
P(heavy) = .0413
P(Next Day Air or heavy) =
P(Next Day Air) + P(heavy) − P(Next Day Air and heavy) =
.2297 + .0413 − .0039 = .2671

4. P(heavy and in targeted area) = .0413
P(heavy and outside targeted area) = .0000

Note that the sum of these two is the total probability of a package being heavy (.0413).

5. P(targeted area | Next Day Air) = .8652

6. P(Next Day Air | outside targeted area) = .32
P(Next Day Air | inside targeted area) = .22

7. Since roughly 90% of the packages being sent from the Pennsylvania warehouse are to destinations within the targeted region, it appears that the warehouse is being used effectively. However, a greater percentage of those packages being sent outside of the targeted area are sent by Next Day Air than those being sent within the targeted area (see question 6). If a reason can be established for this phenomenon, such as insufficient inventory levels of certain parts at the other warehouses, some money could be saved here.

8. P(New England + Northeast + Midwest | Next Day Air) = .196

9. P(Southeast + South Central | Next Day Air) = .564

10. It seems that an inordinate amount of Next Day Air shipping is being done from the Florida warehouse to regions which should be covered by the satellite warehouses (almost 20% to Pennsylvania's territory, and 24% to Arizona's). An examination of the particular orders involved may result in adjusted inventories (both quantities and the choise of items) being stocked at the satellite warehouses.

PROBABILITY II: DISTRIBUTIONS

1. Probability $= \dfrac{54}{250(2)} = .108$

2. Using the binomial distribution with $n = 2$ and $p = .108$, we get P(one machine down) $= 2(.096336) = .192672$, so the expected number of days/year with one machine down is $.192672(250) = 48.17$.

   Similarly, P(two machines down) $= .011664$, so the expected number of days/year with two machines down is $.011664(250) = 2.92$.

3. Expected cost/year $= 48.17(68 + 7.50) + 2.92(100 + 15) = \$3972.64$

4. <u>LEASE OPTION</u>:

   Expected number of days/year with one machine down $= 2(.0475)(250)$ $= 23.75$

   Expected number of days/year with two machines down $= .0025(250)$ $= 0.625$

   Expected cost/year $= 12(350) + 23.75(7.50) + 0.625(15) = \$4387.50$
   Expected cost for three years $= 3(4387.5) = \$13,162.50$

   <u>NEW MACHINE OPTION</u>:

   Expected number of days/year with the machine down $= .017(250)$ $= 4.25$

   Expected first-year cost $= 8750 + 4.25(15) = \$8813.75$
   Expected yearly cost for subsequent years $= 4.25(175 + 15) = \$807.50$
   Expected cost for three years $= 8813.75 + 2(807.50) = \$10,428.75$

   <u>DO NEITHER OPTION</u>:

   Expected cost for three years $= 3(3972.64) = \$11,917.92$

   Hence the best choice for HH Industries is to purchase the new, state-of-the-art machine.

5. The average number of calls received per hour is 27.49.

6. Computing Poisson probabilities with $\lambda = 27.49$, we find

   $$P(x \leq 32) = .8314 \quad \text{and} \quad P(x \leq 40) = .9905$$

   Hence, if they want to be at least 98% certain that each one handles a maximum of eight calls per hour, HH Industries should use five sales reps.

7. No. With only four reps, $P(x \leq 32 + 2) = .9059$. Since this is still below the desired 98%, five reps will still be needed, even though Stan handles two calls per hour.

8. From the histogram below, it appears that the purchases are approximately normally distributed.

CUSTOMER PURCHASES

FREQUENCY OF DOLLARS

DOLLARS MIDPOINT	FREQ	CUM FREQ	PERCENT	CUM PERCENT
0	3	3	0.26	0.26
1500	6	9	0.52	0.78
3000	16	25	1.39	2.17
4500	30	55	2.61	4.78
6000	49	104	4.26	9.04
7500	80	184	6.96	16.00
9000	112	296	9.74	25.74
10500	130	426	11.30	37.04
12000	171	597	14.87	51.91
13500	146	743	12.70	64.61
15000	147	890	12.78	77.39
16500	97	987	8.43	85.83
18000	75	1062	6.52	92.35
19500	46	1108	4.00	96.35
21000	24	1132	2.09	98.43
22500	12	1144	1.04	99.48
24000	6	1150	0.52	100.00

40 80 120 160

FREQUENCY

9. mean = 12817.49 median = 12793 standard deviation = 4248.45

10. $P(x > 20000) = P\left(z > \dfrac{20000 - 12817.49}{4248.45}\right) = P(z > 1.69) = .0455$

$P(x < 10000) = P\left(z < \dfrac{10000 - 12817.49}{4248.45}\right) = P(z < -0.66) = .2546$

The actual proportion above 20,000 is 45/1150 = .0383.
The actual proportion below 10,000 is 289/1150 = .2513.

CHAPTER 6

SAMPLING AND SAMPLING DISTRIBUTIONS

1. Sampling randomly from the drawers of purchase orders will be Bob's best option for collecting a random sample in this case. We can't be sure that there isn't a yearly, or even monthly, trend when it comes to competitive procurements, so cluster sampling is not guaranteed to generate a truly random sample. Systematic sampling is a possibility but, as Bob mentioned, counting through the files and then recording the data from each nth purchase order is quite cumbersome.

2.

n	mean	standard deviation
25	14154.52	4508.91
50	14041.54	4600.49
100	13203.22	4383.00
250	13102.13	4383.26
500	12913.19	4270.37

ESTIMATION

1. $1.96\sqrt{\frac{pq}{n}} = .05$, so $n = \left(\frac{1.96}{.05}\right)^2 pq = 1536.64pq$

p	.2	.3	.4	.5
n	246	323	369	384

2. $\overline{p} = \frac{214}{384} = .557$

Standard error $= \sqrt{\frac{.557(.443)}{384}} = .025$

Confidence interval $= .557 \pm (1.96)(.025) = (.508, .606)$

3. 90% Confidence interval: $.557 \pm (1.64)(.025) = (.516, .598)$

4. $1.96\sqrt{\frac{.557(.443)}{n}} = .03$

$n = \left(\frac{1.96}{.03}\right)^2(.557)(.443) = 1053$ total purchase orders, or 669 more

5. Confidence interval: $.58 \pm (1.96)(.015) = (.55, .61)$. Bob's arithmetic is correct.

Laurel is concerned that the rest of the staff understand exactly what interval estimates tell us, i.e., that if we took a large number of samples of this size, found the sample proportions and respective standard errors, then the true population proportion would be contained in approximately 95% of these intervals. Bob should be prepared to explain clearly what confidence intervals <u>are</u> and what they <u>aren't</u>.

6. $\overline{x} = 5945.20 \qquad s = 2284.85$

7. $s/\sqrt{n} = 2284.85/\sqrt{50} = 323.13$

8. $\overline{x} \pm 1.96\, s/\sqrt{n} = 5945.20 \pm 1.96(323.13) = 5945.20 \pm 633.33$, or $(5311.87, 6578.53)$

9. To sell \$300,000 of metric parts in 50 weeks requires average weekly sales of \$6000. Our best estimate of the distribution of $\overline{x}$ is normal with mean 5945.20 and standard deviation 323.13. For our target mean sales of \$6000, we get a z-score of $(6000 - 5945.20)/323.13 = 0.17$. The corresponding probability is approximately $.5 - .0675 = .4325$ that total sales will be above \$300,000. Hence, this data seems to indicate that, unless Hal can expect an increase in sales, he should consider dropping the metric line.

10. $\overline{x} = 6681.60 \qquad s = 2098.87$

Now to predict next year's sales, our best estimate of the distribution of $\overline{x}$ is normal with mean 6681.60 and standard deviation $2098.87/\sqrt{50} = 296.83$. Our z-score is now $(6000 - 6681.60)/296.83 = -2.30$. The corresponding probability that sales will exceed \$300,000 is .9893, which exceeds our goal of 95% certainty. Given that the metric market is still increasing, it is a fairly safe bet for HH Industries to continue to sell metric parts.

TESTING HYPOTHESES

1. $H_0: p = .6$ $\qquad$ $H_1: p < .6$

$\overline{p} = .58$ $\qquad$ $\overline{q} = .42$ $\qquad$ $\sigma_{\overline{p}} = .015$

We can use the normal approximation to the binomial, so the limit of the acceptance region for our left-tailed test, at the .01 significance level, is

$$.6 - 2.33(.015) = .565$$

Since .58 is greater than .565, we accept the hypothesis that the true proportion of competitively bid purchase orders is .6.

2. $H_0: \mu = .140$ $\qquad$ $H_1: \mu \neq .140$

At the .05 significance level, the limits of the acceptance region are

$$\mu \pm 1.96\sigma/\sqrt{n} = .140 \pm 1.96(.003)/\sqrt{310} = .140 \pm .0003$$

Since $\overline{x} = .142 > .1403$, we reject H_0 and conclude that the O-ring dimension is not within tolerance limits. Note that if the test were based on the s value of .0037 instead of σ, the acceptance region limits of (.1396, .1404) would generate the same conclusion.

TESTING HYPOTHESES: TWO-SAMPLE TESTS

1. Sample 1 (Pre-CONEXPO UCCs): $\bar{x}_1 = 28.6825$ $s_1 = 8.9205$ $n = 63$
Sample 2 (Post-CONEXPO UCCs): $\bar{x}_2 = 34.9683$ $s_2 = 8.8772$ $n = 63$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$ $\alpha = .05$

$z_{CRIT} = z_{.05} = \pm\ 1.96$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1{}^2 + s_2{}^2}{n_1 + n_2}} = \sqrt{\frac{8.9205^2 + 8.8772^2}{63 + 63}} = 1.1212$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{28.6825 - 34.9683 - 0}{1.1212} = -5.6065$$

Since $-5.6065 < -1.96$, we reject the H_0. The pre-CONEXPO and post-CONEXPO call populations, do not have the same number of unique customers per day.

2. Sample 1 (Picked-up): $\bar{p}_p = 0.03698$ $s_p = 8.9205$ $n = 63$
Sample 2 (Mailed): $\bar{p}_m = 0.03183$ $s_m = 8.8772$ $n = 63$

$H_0: p_p = p_m$ $H_1: p_p > p_m$

$$\hat{p} = \frac{n_p\bar{p}_p + n_m\bar{p}_m}{n_p + n_m} = \frac{63(.03698) + 63(.03183)}{126} = 0.0344$$

$$\hat{\sigma}_{\bar{p}_p - \bar{p}_m} = \sqrt{\frac{\hat{p}\hat{q}}{n_p} + \frac{\hat{p}\hat{q}}{n_m}} = \sqrt{\frac{.0344(.9656)}{63} + \frac{.0344(.9656)}{63}} = 0.0325$$

$$z = \frac{(\bar{p}_p - \bar{p}_m) - (p_p - p_m)_{H_0}}{\hat{\sigma}_{\bar{p}_p - \bar{p}_m}} = \frac{(.03698 - .03183) - 0}{.0325} = 0.1585$$

For this small z value (0.1585), we would accept the H_0 for most all values of alpha. In fact, this difference would not be considered significant unless one set the alpha value at .94.

3. Sample 1 (Pre-CONEXPO sales): $\bar{x}_1 = 860.25$ $s_1 = 487.9892$ $n = 40$
Sample 2 (Post-CONEXP0 sales): $\bar{x}_2 = 958.125$ $s_2 = 513.9713$ $n = 40$

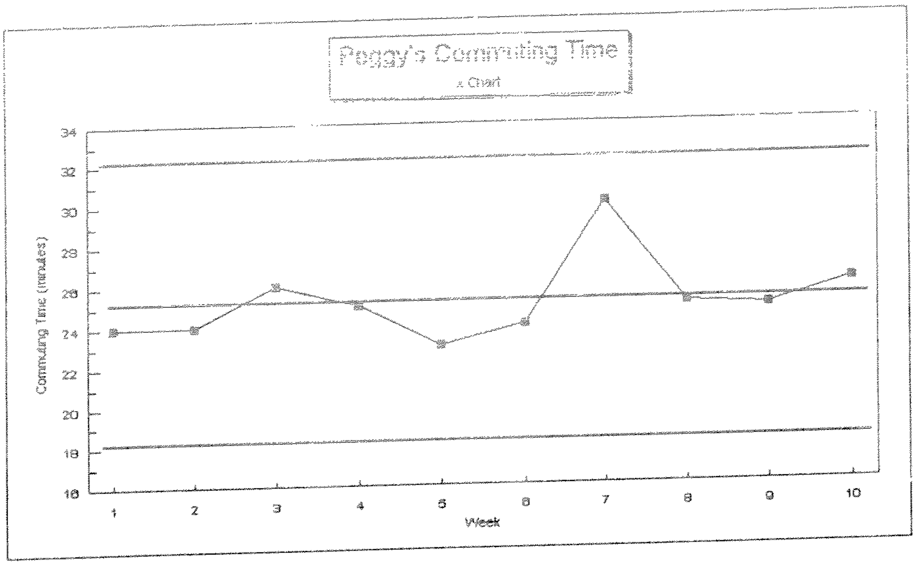
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$ $\alpha = .15$
$z_L = z_{.15} = -1.04$

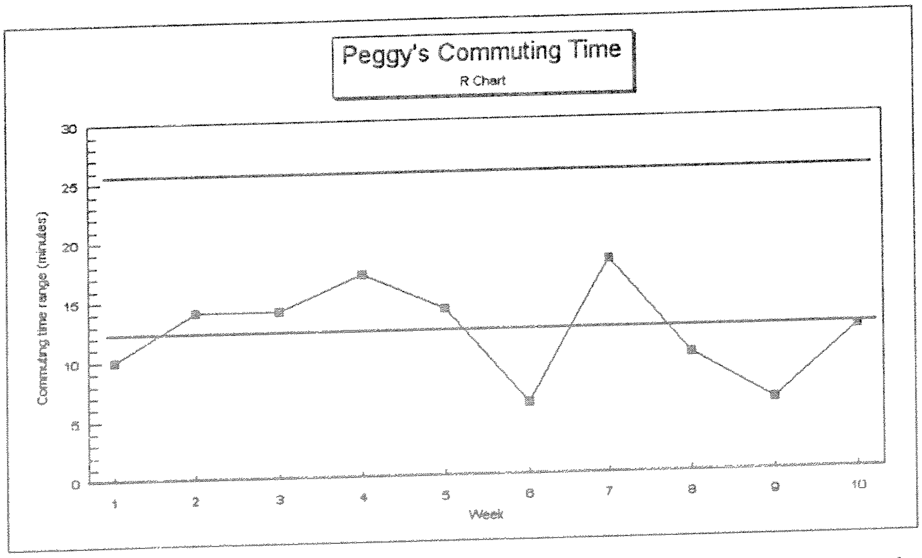
$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1{}^2 + s_2{}^2}{n_1 + n_2}} = \sqrt{\frac{487.9892^2 + 513.9713^2}{40 + 40}} = 79.2345$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{860.25 - 958.125 - 0}{79.2345} = -1.2353$$

Since $-1.2353 < -1.04$, we reject the H_0. Post-CONEXPO sales were significantly greater than pre-CONEXPO sales. However, the difference cannot be attributed solely to the CONEXPO meeting. There are other factors which could have also affected sales, such as the hiring of exceptional sales representatives, or seasonal trends.

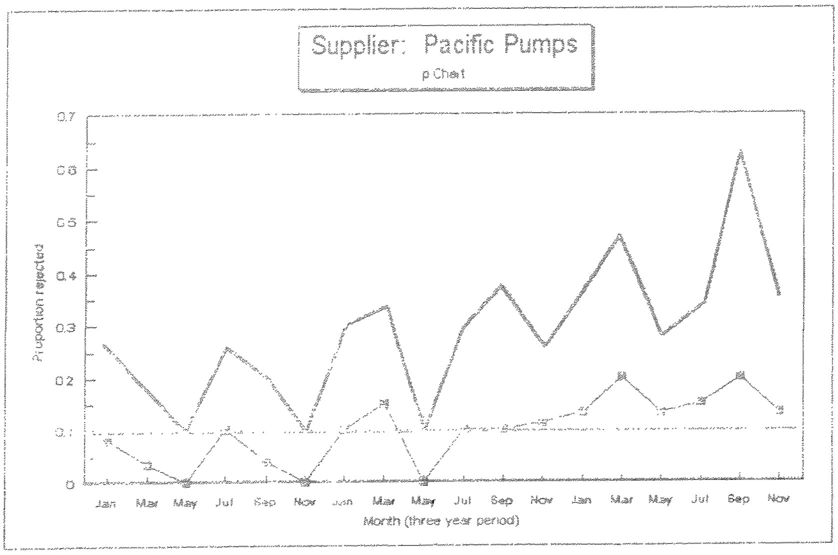
QUALITY AND QUALITY CONTROL





The process is under-control. Peggy's commuting time is relatively consistent with some fluctutations (such as in week 7). There is no pattern in the data to suggest that her commuting time is increasing, or decreasing. Rather, the variability that does exist is most likely due to unusual events or events out of her control, such as heavy traffic, traffic accidents, or mishaps at home prior to her commute.

2.



The data is well within the control limits. However, the proportion of pumps rejected has been greater than average over the last eight months. If the trend continues, it may signify a problem with Pacific Pumps.

CHI-SQUARE AND ANALYSIS OF VARIANCE

1. An analysis of variance of the daily sales data by salesperson yields an observed F of 19.46 (with 3/452 *df*), with a prob value of .0001. Since this prob value is smaller than the 1% significance level of the test, we conclude that the four salespeople have <u>different</u> mean daily sales.

2.

salesperson	mean	standard deviation
Barry	2508.44	1608.30
Debbie	2783.04	1794.49
Jeff	2138.15	1433.72
Mike	3790.57	1971.43

From these values, it seems clear that Stan is correct in his observation that "... Mike's in a class by himself." Running a new ANOVA for the remaining three salespeople yields an observed F of 4.56 (with 2/339 *df*), with a prob value of .0112. Since this prob value is larger than the 1% significance level of the test, we conclude that Barry, Debbie, and Jeff have mean daily sales that <u>are not significantly different</u>.

3.

Interval	Poisson probability with $\lambda = 27.49$	f_e	f_o
0 - 20	.08641	17.1	49
21 - 25	.27603	54.7	34
26 - 30	.36178	71.6	39
31 - 35	.20791	41.2	32
$\geq$ 36	.06787	13.4	44

$$\chi^2 = \frac{\sum (f_e - f_o)^2}{f_e} = 154.1$$

With 3 *df* (5 − 1 − 1, because $\lambda = 27.49$ was estimated from the data) $\chi^2_{.05} = 7.815$, so we reject the hypothesis that the data come from a Poisson distribution.

4. Estimating μ by $\bar{x} = 27.49$ and σ by $s = 9.34$, we'll see if the data come from a normal distribution. We use z-scores of −0.84, −0.25, 0.25, and 0.84 to divide the distribution into five equally likely intervals.

Interval	f_e	f_o
$(-\infty, 19.64)$	39.6	41
$[19.64, 25.16)$	39.6	42
$[25.16, 29.83)$	39.6	35
$[29.83, 35.34)$	39.6	36
$[35.34, +\infty)$	39.6	44

$$\chi^2 = \frac{\sum (f_e - f_o)^2}{f_e} = 1.545$$

With 2 *df* (5 − 1 − 2, because $\mu = 27.49$ and $\sigma = 9.34$ were estimated from the data) $\chi^2_{.05} = 5.991$, so we cannot reject the hypothesis that the data come from a normal distribution.

5. Recalculating using the normal distribution ($\mu = 27.49$, $\sigma = 9.34$):

4 sales reps, $P(x \leq 32) = P(z \leq 0.48) = .6844$
5 sales reps, $P(x \leq 40) = P(z \leq 1.34) = .9099$
6 sales reps, $P(x \leq 48) = P(z \leq 2.20) = .9861$

Hence six sales reps are required to meet the goal of 98%.

6. $P(z \leq 2.05) = .98$, so examining the data on an hourly basis, we get the estimates for μ and σ, with resulting numbers of sales reps required.

hour	μ	σ	$\frac{\mu + 2.05\sigma}{8}$	number of reps needed
8	14.50	4.82	3.05	4
9	26.32	6.88	5.05	6
10	27.91	5.17	4.81	5
11	31.55	7.04	5.75	6
12	23.77	5.60	4.41	5
1	30.73	5.09	5.15	6
2	38.86	7.24	6.71	7
3	34.27	8.08	6.35	7
4	19.50	5.04	3.73	4

Hence, a combination of six full-time reps and one part-time rep (or perhaps five full-time reps and two part-time reps) would cover the entire day, while meeting the goal of 98%.

CHAPTER 12

SIMPLE REGRESSION AND CORRELATION

1. The regression equation is SERVICE = −3.8228 + 0.7839AGE. The standard error of estimate is 9.07, so the approximate confidence intervals are

   For a 25-year-old:

   $$(-3.8228 + 0.7839(25)) \pm 2(9.07) = 15.77 \pm 18.14 = (0, 33.91) \text{ months}$$

   For a 65-year-old:

   $$(-3.8228 + 0.7839(65)) \pm 2(9.07) = 47.13 \pm 18.14 = (28.99, 65.27) \text{ months}$$

   Although there is some overlap of these intervals, it looks like personnel turnover will be less if HH Industries continues its retiree-hiring program.

2. $r^2 = 0.6555, \quad r = 0.8096$

3. H_0: $B = 1.0$ $\qquad$ H_1: $B \neq 1.0$

   Since $s_B = 0.0922$, the acceptance region for H_0 is $B \pm t_{38,.05} s_B = 1.0 \pm 1.645(0.0922) = 1.0 \pm 0.15 = (0.85, 1.15)$. Since the regression coefficient of 0.7839 falls outside this interval, Gary's null hypothesis must be rejected.

CHAPTER 13

MULTIPLE REGRESSION AND MODELING TECHNIQUES

1. The regression equation on gender is

$$\text{SERVICE} = 39.5714 - 19.3022\,\text{GENDER}$$

with $r^2 = 0.3738$, and $r = -0.6114$.

The regression equation on years of education is

$$\text{SERVICE} = 98.3246 - 3.3874\,\text{EDUCATION}$$

with $r^2 = 0.6110$, and $r = -0.7817$.

2. Using all three explanatory variables, the regression equation is

$$\text{SERVICE} = 48.9872 + 0.4741\,\text{AGE} - 3.3737\,\text{GENDER} - 3.0436\,\text{EDUCATN}$$

with $r^2 = 0.7888$, and $r = 0.8881$. This equation has a higher r^2 than the simple regressions, as well as a smaller standard error of estimate (7.30, versus 9.07 for the regression on AGE, 12.23 for the regression on GENDER, and 9.64 for the regression on EDUCATN). This multiple regression is better than any of the simple regressions.

3.

Variable	AGE	GENDER	EDUCATN
Prob > |T|	0.0001	0.2794	0.0003

On the basis of the Prob > |T| values, GENDER is the most appropriate variable to drop from the regression. The new regression equation is

$$\text{SERVICE} = 48.5950 + 0.5108\,\text{AGE} - 3.3005\,\text{EDUCATN}$$

with $r^2 = 0.7817$, and a standard error of estimate of 7.32. This r^2 and s_e are very close to those of the three-variable regression, confirming again that GENDER added essentially nothing to the explanatory capability of this two-variable regression.

CHAPTER 14

NONPARAMETRIC METHODS

1. For the four salespeople, we get the following rank sums from their daily sales data:

Barry	23193.0
Debbie	25990.5
Jeff	19944.0
Mike	35063.5

The Kruskal-Wallis K statistic is 64.04, with 3 *df*. The associated prob value is 0.0001. Since this is smaller than the 5% significance level of the test, we conclude that the four salespeople have different mean daily sales. (Note that even if we used a 1% significance level, the same conclusion would have been reached. This is in agreement with the result reached by the ANOVA we did in Chapter 9.)

Excluding Mike from the analysis, we get the following rank sums:

Barry	19729.5
Debbie	21852.0
Jeff	17071.5

Now the Kruskal-Wallis statistic is 10.30, with 2 *df* and a prob value of 0.0058. So even after excluding Mike, we conclude at $\alpha = .05$ (or even at $\alpha = .01$), that the remaining three salespeople have different daily mean sales. This is in contrast to the result in Chapter 9, where a standard ANOVA led to the conclusion that Barry, Debbie, and Jeff had mean daily sales that were not significantly different.

2. The rank correlation between states "shipped to" and states "received from" is .7367. An upper-tailed test is appropriate, for which the z-value for the .01 level of significance is 2.33. The upper limit of our acceptance region is:

$$\rho_{s_{H_0}} + 2.33\sigma_{r_s} = 0 + 2.33/\sqrt{49} = 0.3329$$

Since our calculated rank correlation coefficient is outside of the acceptance region, we reject the null hypothesis of no correlation, and conclude that there is a relationship between the states HH Industries ships to and the states it receives shipments from.

TIMES SERIES

1. The monthly seasonal indices are:

Jan	0.9786
Feb	0.9396
Mar	1.1180
Apr	0.9909
May	1.1419
Jun	0.9537
Jul	1.0933
Aug	1.2006
Sep	0.9325
Oct	1.0205
Nov	0.8673
Dec	0.7629

When the data are deseasonalized and regressed on time (with 0 between December 1988 and January 1989, and x units $= \frac{1}{2}$ month) the regression equation is:

$$\text{Deseasonalized sales} = 2153.498 + 8.696x$$

For this regression, $r^2 = 0.7246$.

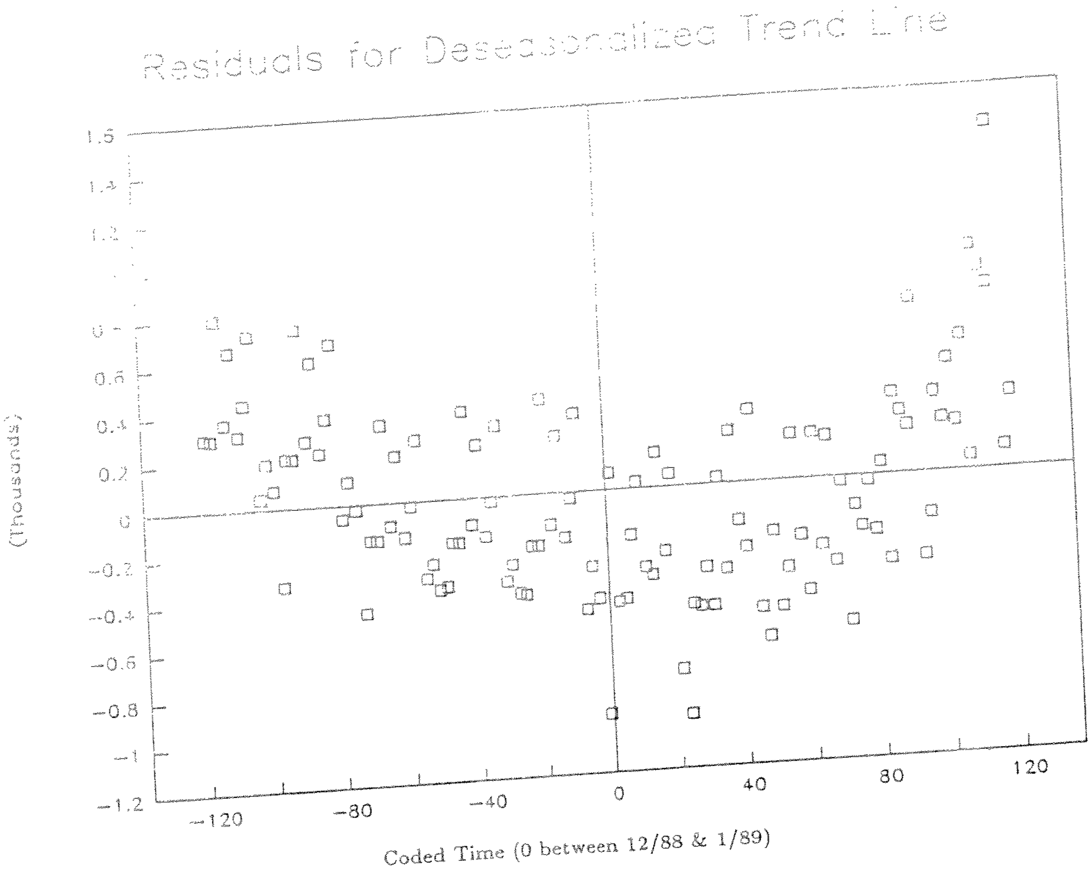
2. The 1994 forecast numbers are:

	Deseasonalized	Seasonalized
Jan	3205.73	3117.17
Feb	3223.12	3028.47
Mar	3240.51	3623.14
Apr	3257.91	3228.20
May	3275.30	3740.15
Jun	3292.69	3140.20
Jul	3310.08	3619.15
Aug	3327.47	3995.25
Sep	3344.87	3118.98
Oct	3362.26	3431.25
Nov	3379.65	2931.01
Dec	3397.04	2591.65

3. As can be seen from the graph below, 19 of the first twenty residuals are positive. Then the values are predominantly negative until the 102nd observation. Finally, all but two of the last 18 residuals are positive. This indicates that a second-degree regression might be more appropriate to estimate the trend. In fact, the regression on time and (time)^2 is

$$\text{Deseasonalized sales} = 1847.353 + 8.696x + 0.0637x^2$$

with an r^2 of 0.8743 and an observed t-value of 11.79 for the coefficient of x^2. Since prob > |T| is .0001, this confirms the suspicion that a second-degree regression for the trend might be more appropriate.



CHAPTER 16

INDEX NUMBERS

1.

Year	Employee Contribution	Index	Total Co. Insurance Cost	Index	Ratio of Indices
1973	$3,690	1.00	$15,000	1.00	1.00
1974	4,320	1.17	17,600	1.17	1.00
1975	4,500	1.22	19,500	1.30	1.07
1976	4,725	1.28	22,000	1.47	1.15
1977	5,775	1.57	24,300	1.62	1.03
1978	6,450	1.75	27,300	1.82	1.04
1979	7,425	2.01	31,000	2.07	1.03
1980	8,250	2.24	34,500	2.30	1.03
1981	9,300	2.52	38,800	2.59	1.03
1982	10,950	2.97	44,100	2.94	0.99
1983	11,850	3.21	49,000	3.27	1.02
1984	13,120	3.56	55,200	3.68	1.03
1985	14,450	3.92	61,300	4.09	1.04
1986	16,200	4.39	68,100	4.54	1.03
1987	20,100	5.45	79,400	5.29	0.97
1988	22,365	6.06	92,800	6.19	1.02
1989	24,035	6.51	108,000	7.20	1.11
1990	27,625	7.49	126,400	8.43	1.13
1991	31,185	8.45	150,000	10.00	1.18
1992	35,100	9.51	180,000	12.00	1.26
1993	40,160	10.88	219,000	14.60	1.34

Note: Ratio = Total Insurance Cost/Employee Contribution

2. The insurance cost index for $270,000 is 18.00. If we want the ratio to be no greater than 1.20, then the employee contribution index must be at least 15.00, which translates to a contribution of $55,350 or more. With 20 employees having individual coverage and 45 having family coverage, the following equation gives us the divisor we need:

$$20x + 45(2.5x) = 132.5x$$

Hence, dividing 55,350 by 132.5, we see that the individual employee deduction for 1994 must be at least $418 (almost a $100 increase from the 1991 deduction). The family coverage deduction, therefore, must be at least $1045 for the year.