modified 2/16/2010

EXCERPTS FROM:

Solutions Manual to Accompany

Statistics for Business and Economics

Eleventh Edition

David R. Anderson
University of Cincinnati

Dennis J. Sweeney
University of Cincinnati

Thomas A. Williams
Rochester Institute of Technology

Contents

1. Data and Statistics

12. a. The population is all visitors coming to the state of Hawaii.
    b. Since airline flights carry the vast majority of visitors to the state, the use of questionnaires for passengers during incoming flights is a good way to reach this population. The questionnaire actually appears on the back of a mandatory plants and animals declaration form that passengers must complete during the incoming flight. A large percentage of passengers complete the visitor information questionnaire.
    c. Questions 1 and 4 provide quantitative data indicating the number of visits and the number of days in Hawaii. Questions 2 and 3 provide qualitative data indicating the categories of reason for the trip and where the visitor plans to stay.

21. a. The two populations are the population of women whose mothers took the drug DES during pregnancy and the population of women whose mothers did not take the drug DES during pregnancy.
    b. It was a survey.
    c. 63 / 3.980 = 15.8 women out of each 1000 developed tissue abnormalities.
    d. The article reported "twice" as many abnormalities in the women whose mothers had taken DES during pregnancy. Thus, a rough estimate would be 15.8/2 = 7.9 abnormalities per 1000 women whose mothers had *not* taken DES during pregnancy.
    e. In many situations, disease occurrences are rare and affect only a small portion of the population. Large samples are needed to collect data on a reasonable number of cases where the disease exists.

2. Descriptive Statistics: Tabular and Graphical Methods

15. a/b.

Waiting Time	Frequency	Relative Frequency
0 - 4	4	0.20
5 - 9	8	0.40
10 - 14	5	0.25
15 - 19	2	0.10
20 - 24	1	0.05
Totals	20	1.00

c/d.

Waiting Time	Cumulative Frequency	Cumulative Relative Frequency
Less than or equal to 4	4	0.20
Less than or equal to 9	12	0.60
Less than or equal to 14	17	0.85
Less than or equal to 19	19	0.95
Less than or equal to 24	20	1.00

e. 12/20 = 0.60

29. a.

x \ ***y***	1	2	Total
A	5	0	5
B	11	2	13
C	2	10	12
Total	18	12	30

b.

x \ ***y***	1	2	Total
A	100.0	0.0	100.0
B	84.6	15.4	100.0
C	16.7	83.3	100.0

c.

		y: 1	y: 2
x	A	27.8	0.0
	B	61.1	16.7
	C	11.1	83.3
	Total	100.0	100.0

d. Category A values for *x* are always associated with category 1 values for *y*. Category B values for *x* are usually associated with category 1 values for y. Category C values for *x* are usually associated with category 2 values for *y*.

50. a.

	Fuel Type					
Year Constructed	Elec	Nat. Gas	Oil	Propane	Other	Total
1973 or before	40	183	12	5	7	247
1974-1979	24	26	2	2	0	54
1980-1986	37	38	1	0	6	82
1987-1991	48	70	2	0	1	121
Total	149	317	17	7	14	504

b.

Year Constructed	Frequency	Fuel Type	Frequency
1973 or before	247	Electricity	149
1974-1979	54	Nat. Gas	317
1980-1986	82	Oil	17
1987-1991	121	Propane	7
Total	504	Other	14
		Total	504

c. Crosstabulation of Column Percentages

	Fuel Type				
Year Constructed	Elec	Nat. Gas	Oil	Propane	Other
1973 or before	26.9	57.7	70.5	71.4	50.0
1974-1979	16.1	8.2	11.8	28.6	0.0
1980-1986	24.8	12.0	5.9	0.0	42.9
1987-1991	32.2	22.1	11.8	0.0	7.1
Total	100.0	100.0	100.0	100.0	100.0

d. Crosstabulation of row percentages.

	Fuel Type					
Year Constructed	Elec	Nat. Gas	Oil	Propane	Other	Total
1973 or before	16.2	74.1	4.9	2.0	2.8	100.0
1974-1979	44.5	48.1	3.7	3.7	0.0	100.0
1980-1986	45.1	46.4	1.2	0.0	7.3	100.0
1987-1991	39.7	57.8	1.7	0.0	0.8	100.0

e. Observations from the column percentages crosstabulation

For those buildings using electricity, the percentage has not changed greatly over the years. For the buildings using natural gas, the majority were constructed in 1973 or before; the second largest percentage was constructed in 1987-1991. Most of the buildings using oil were constructed in 1973 or before. All of the buildings using propane are older.

Observations from the row percentages crosstabulation

Most of the buildings in the CG&E service area use electricity or natural gas. In the period 1973 or before most used natural gas. From 1974-1986, it is fairly evenly divided between electricity and natural gas. Since 1987 almost all new buildings are using electricity or natural gas with natural gas being the clear leader.

3. Descriptive Statistics: Numerical Methods

5. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{3181}{20} = \$159$

b. Median 10th $160 Los Angeles
11th $162 Seattle

$$\text{Median} = \frac{160+162}{2} = \$161$$

c. Mode = $167 San Francisco and New Orleans

d. $i = \left(\frac{25}{100}\right)20 = 5$

5th $134
6th $139

$$Q_1 = \frac{134+139}{2} = \$136.50$$

e. $i = \left(\frac{75}{100}\right)20 = 15$

15th $167
16th $173

$$Q_3 = \frac{167+173}{2} = \$170$$

19. a. Range = 60 - 28 = 32
IQR = Q_3 - Q_1 = 55 - 45 = 10

b. $\bar{x} = \frac{435}{9} = 48.33$

$\Sigma(x_i - \bar{x})^2 = 742$

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{742}{8} = 92.75 \qquad s = \sqrt{92.75} = 9.63$$

c. The average air quality is about the same. But, the variability is greater in Anaheim.

34. a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{765}{10} = 76.5$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{442.5}{10-1}} = 7$$

b. $z = \frac{x - \bar{x}}{s} = \frac{84 - 76.5}{7} = 1.07$

Approximately one standard deviation above the mean. Approximately 68% of the scores are within one standard deviation. Thus, half of (100-68), or 16%, of the games should have a winning score of 84 or more points.

$$z = \frac{x - \bar{x}}{s} = \frac{90 - 76.5}{7} = 1.93$$

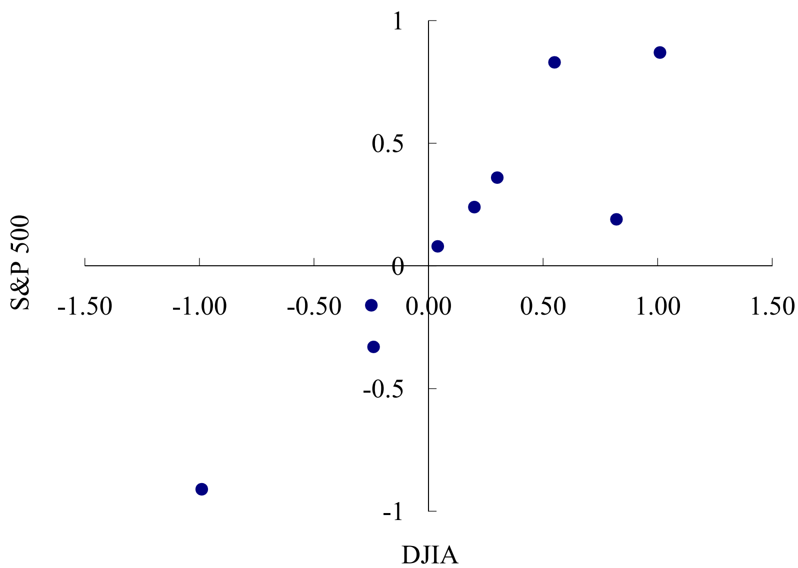
Approximately two standard deviations above the mean. Approximately 95% of the scores are within two standard deviations. Thus, half of (100-95), or 2.5%, of the games should have a winning score of more than 90 points.

c. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{122}{10} = 12.2$

$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{559.6}{10-1}} = 7.89$

Largest margin 24: $z = \frac{x - \bar{x}}{s} = \frac{24 - 12.2}{7.89} = 1.50$. No outliers.

50. a.



b. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{1.44}{9} = .16$ $\bar{y} = \frac{\Sigma x_i}{n} = \frac{1.17}{9} = .13$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
0.20	0.24	0.04	0.11	0.0016	0.0121	0.0044
0.82	0.19	0.66	0.06	0.4356	0.0036	0.0396
-0.99	-0.91	-1.15	-1.04	1.3225	1.0816	1.1960
0.04	0.08	-0.12	-0.05	0.0144	0.0025	0.0060
-0.24	-0.33	-0.40	-0.46	0.1600	0.2166	0.1840
1.01	0.87	0.85	0.74	0.7225	0.5476	0.6290
0.30	0.36	0.14	0.23	0.0196	0.0529	0.0322
0.55	0.83	0.39	0.70	0.1521	0.4900	0.2730
-0.25	-0.16	-0.41	-0.29	0.1681	0.0841	0.1189
			Total	2.9964	2.4860	2.4831

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{2.4831}{8} = .3104$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{2.9964}{8}} = .6120$$

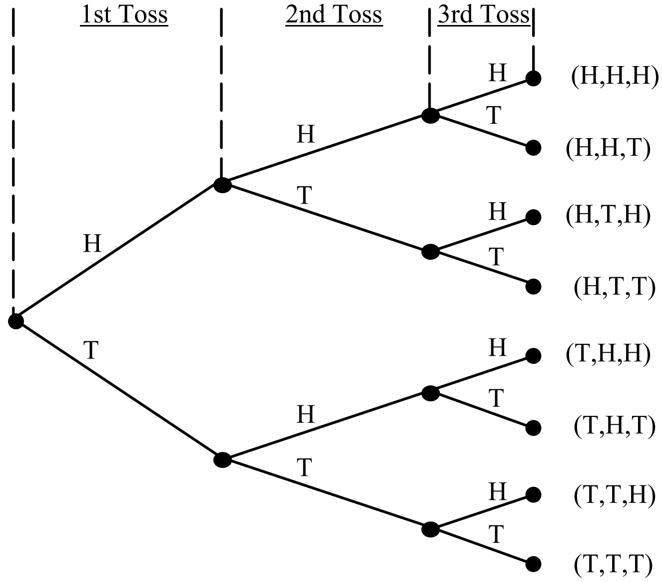
$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{2.4860}{8}} = .5574$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{.3104}{(.6120)(.5574)} = .9098$$

c. There is a strong positive linear association between DJIA and S&P 500. If you know the change in either, you will have a good idea of the stock market performance for the day.

4. Introduction to Probability

4. a.



b. Let: H be head and T be tail
(H,H,H) (T,H,H)
(H,H,T) (T,H,T)
(H,T,H) (T,T,H)
(H,T,T) (T,T,T)

c. The outcomes are equally likely, so the probability of each outcomes is 1/8.

7. No. Requirement (4.4) is not satisfied; the probabilities do not sum to 1. $P(E_1) + P(E_2) + P(E_3) + P(E_4) = .10 + .15 + .40 + .20 = .85$

21. a. Use the relative frequency method. Divide by the total adult population of 227.6 million.

Age	Number	Probability
18 to 24	29.8	0.1309
25 to 34	40.0	0.1757
35 to 44	43.4	0.1907
45 to 54	43.9	0.1929
55 to 64	32.7	0.1437
65 and over	37.8	0.1661
Total	227.6	1.0000

b. P(18 to 24) = .1309
c. P(18 to 34) = .1309 + .1757 = .3066
d. P(45 or older) = .1929 + .1437 + .1661 = .5027

26. a. Let D = Domestic Equity Fund
$P(D) = 16/25 = .64$
b. Let A = 4- or 5-star rating
13 funds were rated 3-star of less; thus, 25 – 13 = 12 funds must be 4-star or 5-star.
$P(A) = 12/25 = .48$
c. 7 Domestic Equity funds were rated 4-star and 2 were rated 5-star. Thus, 9 funds were Domestic Equity funds and were rated 4-star or 5-star
$P(D \cap A) = 9/25 = .36$

d. $P(D \cup A) = P(D) + P(A) - P(D \cap A)$
$= .64 + .48 - .36 = .76$

28. Let: B = rented a car for business reasons
P = rented a car for personal reasons

a. $P(\text{B} \cup \text{P}) = P(\text{B}) + P(\text{P}) - P(\text{B} \cap \text{P})$
$= .54 + .458 - .30 = .698$

b. $P(\text{Neither}) = 1 - .698 = .302$

31. a. $P(\text{A} \cap \text{B}) = 0$

b. $P(\text{A} | \text{B}) = \dfrac{P(\text{A} \cap \text{B})}{P(\text{B})} = \dfrac{0}{.4} = 0$

c. No. $P(\text{A} \mid \text{B}) \neq P(\text{A})$; $\therefore$ the events, although mutually exclusive, are not independent.

d. Mutually exclusive events are dependent.

34. a. Let O = flight arrives on time
O^c = flight arrives late
S = Southwest flight
U = US Airways flight
J = JetBlue flight

Given: $P(\text{O} \mid \text{S}) = .834$ $P(\text{O} \mid \text{U}) = .751$ $P(\text{O} \mid \text{J}) = .701$
$P(\text{S}) = .40$ $P(\text{U}) = .35$ $P(\text{J}) = .25$

$$P(\text{O} \mid \text{S}) = \frac{P(\text{O} \cap \text{S})}{P(\text{S})}$$

$\therefore$ $P(\text{O} \cap \text{S}) = P(\text{O} \mid \text{S})P(\text{S}) = (.834)(.4) = .3336$

Similarly
$P(\text{O} \cap \text{U}) = P(\text{O} \mid \text{U})P(\text{U}) = (.751)(.35) = .2629$
$P(\text{O} \cap \text{J}) = P(\text{O} \mid \text{J})P(\text{J}) = (.701)(.25) = .1753$

Joint probability table

	On time	Late	Total
Southwest	.3336	.0664	.40
US Airways	.2629	.0871	.35
JetBlue	.1753	.0747	.25
Total:	.7718	.2282	1.00

b. Southwest Airlines; $P(\text{S}) = .40$

c. $P(\text{O}) = P(\text{S} \cap \text{O}) + P(\text{U} \cap \text{O}) + P(\text{J} \cap \text{O}) = .3336 + .2629 + .1753 = .7718$

d. $P(\text{S} | \text{O}^c) = \dfrac{P(\text{S} \cap \text{O}^c)}{P(\text{O}^c)} = \dfrac{.0664}{.2282} = .2910$

Similarly, $P(\text{U} | \text{O}^c) = \dfrac{.0871}{.2282} = .3817$

$P(\text{J} | \text{O}^c) = \dfrac{.0747}{.2282} = .3273$

Most likely airline is US Airways; least likely is Southwest

42. M = missed payment
D_1 = customer defaults
D_2 = customer does not default
$\text{P}(\text{D}_1) = .05$ $\text{P}(\text{D}_2) = .95$ $\text{P}(\text{M} \mid \text{D}_2) = .2$ $\text{P}(\text{M} \mid \text{D}_1) = 1$

a. $\text{P}(\text{D}_1 | \text{M}) = \dfrac{\text{P}(\text{D}_1)\text{P}(\text{M}|\text{D}_1)}{\text{P}(\text{D}_1)\text{P}(\text{M}|\text{D}_1) + \text{P}(\text{D}_2)\text{P}(\text{M}|\text{D}_2)} = \dfrac{(.05)(1)}{(.05)(1) + (.95)(.2)} = \dfrac{.05}{.24} = .21$

b. Yes, the probability of default is greater than .20.

43. Let: S = small car

S^c = other type of vehicle

F = accident leads to fatality for vehicle occupant

We have P(S) = .18, so P(S^c) = .82. Also P(F | S) = .128 and P(F | S^c) = .05. Using the tabular form of Bayes Theorem provides:

Events	Prior Probabilities	Conditional Probabilities	Joint Probabilities	Posterior Probabilities
S	.18	.128	.023	.36
S^c	.82	.050	.041	.64
	1.00		.064	1.00

From the posterior probability column, we have P(S | F) = .36. So, if an accident leads to a fatality, the probability a small car was involved is .36.

56. a. P(A) = 200/800 = .25

b. P(B) = 100/800 = .125

c. $P(A \cap B)$ = 10/800 = .0125

d. $P(A \mid B) = P(A \cap B) / P(B)$ = .0125 / .125 = .10

e. No, $P(A \mid B) \neq P(A) = .25$

59. a. P(Oil) = .50 + .20 = .70

b. Let S = Soil test results

Events	$P(A_i)$	$P(S \mid A_i)$	$P(A_i \cap S)$	$P(A_i \mid S)$
High Quality (A_1)	.50	.20	.10	.31
Medium Quality (A_2)	.20	.80	.16	.50
No Oil (A_3)	.30	.20	.06	.19
	1.00		P(S) = .32	1.00

P(Oil) = .81 which is good; however, probabilities now favor medium quality rather than high quality oil.

60. a. Let F = female. Using past history as a guide, P(F) = .40.

b. Let D = Dillard's

$$P(F|D) = \frac{.40\left(\frac{3}{4}\right)}{.40\left(\frac{3}{4}\right)+.60\left(\frac{1}{4}\right)} = \frac{.30}{.30+.15} = .67$$

The revised (posterior) probability that the visitor is female is .67.

We should display the offer that appeals to female visitors.

5. Discrete Probability Distributions

2. a. Let x = time (in minutes) to assemble the product.
   b. It may assume any positive value: $x > 0$.
   c. Continuous

14. a. $f(200) = 1 - f(-100) - f(0) - f(50) - f(100) - f(150)$
   $= 1 - .95 = .05$
   This is the probability MRA will have a $200,000 profit.
   b. P(Profit) $= f(50) + f(100) + f(150) + f(200)$
   $= .30 + .25 + .10 + .05 = .70$
   c. P(at least 100) $= f(100) + f(150) + f(200)$
   $= .25 + .10 + .05 = .40$

19. a. $E(x) = \Sigma x f(x) = 0\ (.56) + 2\ (.44) = .88$
   b. $E(x) = \Sigma x f(x) = 0\ (.66) + 3\ (.34) = 1.02$
   c. The expected value of a 3 - point shot is higher. So, if these probabilities hold up, the team will make more points in the long run with the 3 - point shot.

24. a. Medium $E(x) = \Sigma x f(x) = 50\ (.20) + 150\ (.50) + 200\ (.30) = 145$
   Large: $E(x) = \Sigma x f(x) = 0\ (.20) + 100\ (.50) + 300\ (.30) = 140$
   Medium preferred.
   b. Medium

x	$f(x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2 f(x)$
50	.20	-95	9025	1805.0
150	.50	5	25	12.5
200	.30	55	3025	907.5
				σ^2 = 2725.0

Large

y	$f(y)$	$y - \mu$	$(y - \mu)^2$	$(y - \mu)^2 f(y)$
0	.20	-140	19600	3920
100	.50	-40	1600	800
300	.30	160	25600	7680
				σ^2 = 12,400

Medium preferred due to less variance.

26. a. $f(0) = .3487$
   b. $f(2) = .1937$
   c. $P(x \leq 2) = f(0) + f(1) + f(2) = .3487 + .3874 + .1937 = .9298$
   d. $P(x \geq 1) = 1 - f(0) = 1 - .3487 = .6513$
   e. $E(x) = n\,p = 10\ (.1) = 1$
   f. $\text{Var}(x) = n\,p\ (1 - p) = 10\ (.1)\ (.9) = .9, \qquad \sigma = \sqrt{.9} = .9487$

29. a. $f(x) = \binom{n}{x} (p)^x (1-p)^{n-x}$

$$f(3) = \frac{10!}{3!(10-3)!}(.30)^3(1-.30)^{10-3}$$

$$f(3) = \frac{10(9)(8)}{3(2)(1)}(.30)^3(1-.30)^7 = .2668$$

   b. $P(x \geq 3) = 1 - f(0) - f(1) - f(2)$

$$f(0) = \frac{10!}{0!(10)!}(.30)^0(1-.30)^{10} = .0282$$

$$f(1) = \frac{10!}{1!(9)!}(.30)^1(1-.30)^9 = .1211$$

$$f(2) = \frac{10!}{2!(8)!}(.30)^2(1-.30)^8 = .2335$$

$P(x \geq 3) = 1 - .0282 - .1211 - .2335 = .6172$

39. a. $f(x) = \dfrac{2^x e^{-2}}{x!}$

b. $\mu = 6$ for 3 time periods

c. $f(x) = \dfrac{6^x e^{-6}}{x!}$

d. $f(2) = \dfrac{2^2 e^{-2}}{2!} = \dfrac{4(.1353)}{2} = .2706$

e. $f(6) = \dfrac{6^6 e^{-6}}{6!} = .1606$

f. $f(5) = \dfrac{4^5 e^{-4}}{5!} = .1563$

58. Since the shipment is large we can assume that the probabilities do not change from trial to trial and use the binomial probability distribution.

a. $n = 5$

$$f(0) = \binom{5}{0}(0.01)^0(0.99)^5 = 0.9510$$

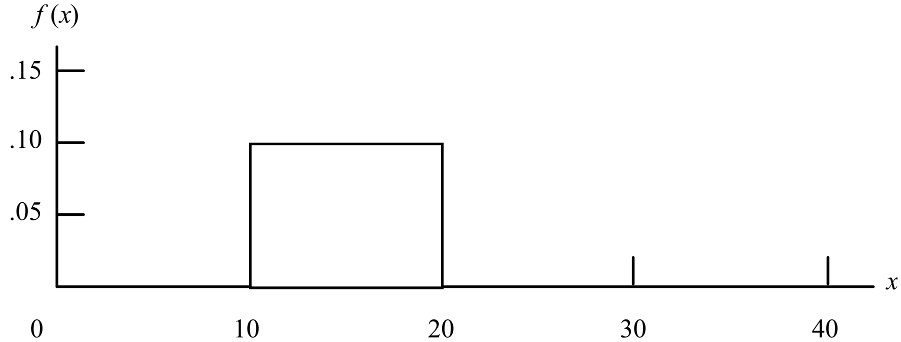
b. $f(1) = \binom{5}{1}(0.01)^1(0.99)^4 = 0.0480$

c. $1 - f(0) = 1 - .9510 = .0490$

d. No, the probability of finding one or more items in the sample defective when only 1% of the items in the population are defective is small (only .0490). I would consider it likely that more than 1% of the items are defective.

6. Continuous Probability Distributions

2. a.



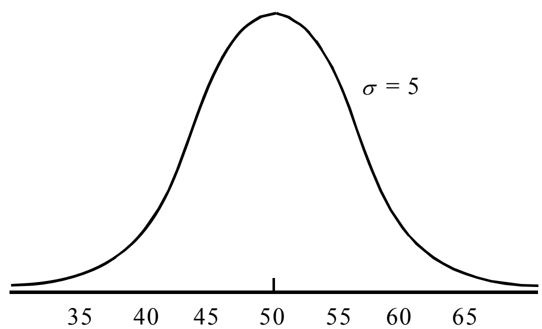
b. $P(x < 15) = .10(5) = .50$

c. $P(12 \leq x \leq 18) = .10(6) = .60$

d. $E(x) = \frac{10+20}{2} = 15$

e. $\text{Var}(x) = \frac{(20-10)^2}{12} = 8.33$

9. a.



b. .683 since 45 and 55 are within plus or minus 1 standard deviation from the mean of 50 (Use the table or see characteristic 7a of the normal distribution).

c. .954 since 40 and 60 are within plus or minus 2 standard deviations from the mean of 50 (Use the table or see characteristic 7b of the normal distribution).

13. a. $P(-1.98 \leq z \leq .49) = P(z \leq .49) - P(z < -1.98) = .6879 - .0239 = .6640$

b. $P(.52 \leq z \leq 1.22) = P(z \leq 1.22) - P(z < .52) = .8888 - .6985 = .1903$

c. $P(-1.75 \leq z \leq -1.04) = P(z \leq -1.04) - P(z < -1.75) = .1492 - .0401 = .1091$

15. a. The z value corresponding to a cumulative probability of .2119 is $z = -.80$.

b. Compute .9030/2 = .4515;
z corresponds to a cumulative probability of .5000 + .4515 = .9515. So $z = 1.66$.

c. Compute .2052/2 = .1026;
z corresponds to a cumulative probability of .5000 + .1026 = .6026. So $z = .26$.

d. The z value corresponding to a cumulative probability of .9948 is z = 2.56.

e. The area to the left of z is 1 - .6915 = .3085. So $z = -.50$.

41. a. $P(defect) = 1 - P(9.85 \leq x \leq 10.15) = 1 - P(-1 \leq z \leq 1) = 1 - .6826 = .3174$
Expected number of defects = 1000(.3174) = 317.4

b. $P(defect) = 1 - P(9.85 \leq x \leq 10.15) = 1 - P(-3 \leq z \leq 3) = 1 - .9974 = .0026$

Expected number of defects = 1000(.0026) = 2.6

c. Reducing the process standard deviation causes a substantial reduction in the number of defects.

7. Sampling and Sampling Distributions

3. 459, 147, 385, 113, 340, 401, 215, 2, 33, 348

19. a. The sampling distribution is normal with

$E(\bar{x}) = \mu = 200$ and $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 50/\sqrt{100} = 5$

For ± 5, $195 \le \bar{x} \le 205$. Using Standard Normal Probability Table:

At $\bar{x} = 205$, $z = \dfrac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \dfrac{5}{5} = 1$ $\quad P(z \le 1) = .8413$

At $\bar{x} = 195$, $z = \dfrac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \dfrac{-5}{5} = -1$ $\quad P(z < -1) = .1587$

$P(195 \le \bar{x} \le 205) = .8413 - .1587 = .6826$

b. For ± 10, $190 \le \bar{x} \le 210$. Using Standard Normal Probability Table:

At $\bar{x} = 210$, $z = \dfrac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \dfrac{10}{5} = 2$ $\quad P(z \le 2) = .9772$

At $\bar{x} = 190$, $z = \dfrac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \dfrac{-10}{5} = -2$ $\quad P(z < -2) = .0228$

$P(190 \le \bar{x} \le 210) = .9772 - .0228 = .9544$

37. a. Normal distribution: $E(\bar{p}) = .12$, $\sigma_{\bar{p}} = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{(.12)(1-.12)}{540}} = .0140$

b. $z = \dfrac{\bar{p}-p}{\sigma_{\bar{p}}} = \dfrac{.03}{.0140} = 2.14$ $\quad P(z \le 2.14) = .9838$ $\quad P(z < -2.14) = .0162$

$P(.09 \le \bar{p} \le .15) = .9838 - .0162 = .9676$

c. $z = \dfrac{\bar{p}-p}{\sigma_{\bar{p}}} = \dfrac{.015}{.0140} = 1.07$ $\quad P(z \le 1.07) = .8577$ $\quad P(z < -1.07) = .1423$

$P(.105 \le \bar{p} \le .135) = .8577 - .1423 = .7154$

44. a. Normal distribution because of central limit theorem ($n > 30$)

$E(\bar{x}) = 115.50$, $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{35}{\sqrt{40}} = 5.53$

b. $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{10}{35/\sqrt{40}} = 1.81$ $\quad P(z \le 1.81) = .9649$, $P(z < -1.81) = .0351$

$P(105.50 \le \bar{x} \le 125.50) = P(-1.81 \le z \le 1.81) = .9649 - .0351 = .9298$

c. At $\bar{x} = 100$, $z = \dfrac{100-115.50}{35/\sqrt{40}} = -2.80$ $\quad P(\bar{x} \le 100) = P(z \le -2.80) = .0026$

Yes, this is an unusually low spending group of 40 alums. The probability of spending this much or less is only .0026.

53. a. Normal distribution with $E(\bar{p}) = .15$ and $\sigma_{\bar{p}} = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{(0.15)(0.85)}{150}} = 0.0292$

b. $P(.12 \le \bar{p} \le .18) = ?$

$z = \frac{.18 - .15}{.0292} = 1.03$ $P(z \leq 1.03) = .8485$, $P(z < -1.03) = .1515$

$P(.12 \leq \bar{p} \leq .18) = P(-1.03 \leq z \leq 1.03) = .8485 - .1515 = .6970$

8. Interval Estimation

7. Margin of error = $z_{.025}(\sigma/\sqrt{n})$ = 1.96(600/ $\sqrt{50}$) = 166.31

A larger sample size would be needed to reduce the margin of error to $150 or less. Section 8.3 can be used to show that the sample size would need to be increased to n = 62.

$1.96(600/\sqrt{n}) = 150$ Solving for n yields n = 62

14. $\bar{x} \pm t_{\alpha/2}(s/\sqrt{n})$ df = 53

a. $22.5 \pm 1.674\,(4.4/\sqrt{54})$

22.5 ± 1 or 21.5 to 23.5

b. $22.5 \pm 2.006\,(4.4/\sqrt{54})$

22.5 ± 1.2 or 21.3 to 23.7

c. $22.5 \pm 2.672\,(4.4/\sqrt{54})$

22.5 ± 1.6 or 20.9 to 24.1

d. As the confidence level increases, there is a larger margin of error and a wider confidence interval.

18. For the JobSearch data set, $\bar{x} = 22$ and $s = 11.8862$

a. $\bar{x}$ = 22 weeks

b. margin of error = $t_{.025}s/\sqrt{n} = 2.023(11.8862)/\sqrt{40} = 3.8020$

c. The 95% confidence interval is $\bar{x}$ ± margin of error = 22 ± 3.8020 or 18.20 to 25.80

d. Skewness = 1.0062, data are skewed to the right.

This modest positive skewness in the data set can be expected to exist in the population.

Regardless of skewness, this is a pretty small data set. Consider using a larger sample next time.

29. a. $n = \dfrac{(1.96)^2(6.25)^2}{2^2} = 37.52$ Use $n = 38$

b. $n = \dfrac{(1.96)^2(6.25)^2}{1^2} = 150.06$ Use $n = 151$

34. Use planning value p^* = .50

$n = \dfrac{(1.96)^2(0.50)(0.50)}{(0.03)^2} = 1067.11$ Use $n = 1068$

36. a. $\bar{p}$ = 46/200 = .23

b. $\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}} = \sqrt{\dfrac{.23(1-.23)}{200}} = .0298$, $\bar{p} \pm z_{.025}\sqrt{\dfrac{\bar{p}(1-\bar{p})}{n}}$ = .23 ± 1.96(.0298)

= .23 ± .0584 or .1716 to .2884

39. a. $n = \dfrac{z_{.025}^2 p^*(1-p^*)}{E^2} = \dfrac{(1.96)^2(.156)(1-.156)}{(.03)^2} = 562$

b. $n = \dfrac{z_{.005}^2 p^*(1-p^*)}{E^2} = \dfrac{(2.576)^2(.156)(1-.156)}{(.03)^2} = 970.77$ Use 971

9. Hypothesis Testing

1. a. H_0: $\mu \leq 600$ H_a: $\mu > 600$ assuming that you give benefit of doubt to the manager.
   b. We are not able to conclude that the manager's claim is wrong.
   c. The manager's claim can be rejected. We can conclude that $\mu > 600$.

2. a. H_0: $\mu \leq 14$ H_a: $\mu > 14$ Research hypothesis
   b. There is no statistical evidence that the new bonus plan increases sales volume.
   c. The research hypothesis that $\mu > 14$ is supported. We can conclude that the new bonus plan increases the mean sales volume.

7. a. H_0: $\mu \leq 8000$
   H_a: $\mu > 8000$ Research hypothesis to see if the plan increases average sales.
   b. Claiming $\mu > 8000$ when the plan does not increase sales. A mistake could be implementing the plan when it does not help.
   c. Concluding $\mu \leq 8000$ when the plan really would increase sales. This could lead to not implementing a plan that would increase sales.

10. a. $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{26.4 - 25}{6 / \sqrt{40}} = 1.48$
   b. Upper tail p-value is the area to the right of the test statistic
   Using normal table with $z = 1.48$: p-value = 1.0000 - .9306 = .0694
   c. p-value > .01, do not reject H_0
   d. Reject H_0 if $z \geq 2.33$
   $1.48 < 2.33$, do not reject H_0

24. a. $t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} = \dfrac{17 - 18}{4.5 / \sqrt{48}} = -1.54$
   b. Degrees of freedom = $n - 1 = 47$
   Because $t < 0$, p-value is two times the lower tail area
   Using t table: area in lower tail is between .05 and .10; therefore, p-value is between .10 and .20.
   Exact p-value corresponding to $t = -1.54$ is .1303
   c. p-value > .05, do not reject H_0.
   d. With $df = 47$, $t_{.025} = 2.012$
   Reject H_0 if $t \leq -2.012$ or $t \geq 2.012$
   $t = -1.54$; do not reject H_0

30. a. H_0: $\mu = 600$, H_a: $\mu \neq 600$
   b. $t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}} = \dfrac{612 - 600}{65 / \sqrt{40}} = 1.17$ $df = n - 1 = 39$
   Because $t > 0$, p-value is two times the upper tail area
   Using t table: area in upper tail is between .10 and .20; therefore, p-value is between .20 and .40.
   Exact p-value corresponding to $t = 1.17$ is .2491
   c. With $\alpha = .10$ or less, we cannot reject H_0. We are unable to conclude there has been a change in the mean CNN viewing audience.
   d. The sample mean of 612 thousand viewers is encouraging but not conclusive for the sample of 40 days. Recommend additional viewer audience data. A larger sample should help clarify the situation for CNN.

34. a. H_0: $\mu = 2$ H_a: $\mu \neq 2$
   b. $\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{22}{10} = 2.2$

c. $s = \sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n-1}} = .516$

d. $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{2.2 - 2}{.516/\sqrt{10}} = 1.22$

Degrees of freedom = n - 1 = 9
Because $t > 0$, p-value is two times the upper tail area
Using t table: area in upper tail is between .10 and .20; therefore, p-value is between .20 and .40.
Exact p-value corresponding to t = 1.22 is .2535

e. p-value > .05; do not reject H_0. No reason to change from the 2 hours for cost estimating purposes.

36. a. $z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{.68 - .75}{\sqrt{\dfrac{.75(1-.75)}{300}}} = -2.80$

Lower tail p-value is the area to the left of the test statistic
Using normal table with z = -2.80: p-value =.0026
p-value ≤ .05; Reject H_0

b. $z = \dfrac{.72 - .75}{\sqrt{\dfrac{.75(1-.75)}{300}}} = -1.20$

Lower tail p-value is the area to the left of the test statistic
Using normal table with z = -1.20: p-value =.1151
p-value > .05; Do not reject H_0

c. $z = \dfrac{.70 - .75}{\sqrt{\dfrac{.75(1-.75)}{300}}} = -2.00$

Lower tail p-value is the area to the left of the test statistic
Using normal table with z = -2.00: p-value =.0228
p-value ≤ .05; Reject H_0

d. $z = \dfrac{.77 - .75}{\sqrt{\dfrac{.75(1-.75)}{300}}} = .80$

Lower tail p-value is the area to the left of the test statistic
Using normal table with z = .80: p-value =.7881
p-value > .05; Do not reject H_0

40. a. $\bar{p} = \dfrac{414}{1532} = .2702$ (27%)

b. H_0: $p \le .22$, H_a: $p > .22$

$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{.2702 - .22}{\sqrt{\dfrac{.22(1-.22)}{1532}}} = 4.75$

Upper tail p-value is the area to the right of the test statistic
Using normal table with z = 4.75: p-value ≈ 0 so Reject H_0.
Conclude that there has been a significant increase in the intent to watch the TV programs.

c. These studies help companies and advertising firms evaluate the impact and benefit of commercials.

45. a. H_0: $p = .30$ H_a: $p \ne .30$

b. $\bar{p} = \dfrac{24}{50} = .48$

c. $z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.48 - .30}{\sqrt{\frac{.30(1-.30)}{50}}} = 2.78$

Because $z > 0$, p-value is two times the upper tail area
Using normal table with $z = 2.78$: p-value = 2(.0027) = .0054
p-value $\leq$ .01; reject H_0.
We would conclude that the proportion of stocks going up on the NYSE is not 30%. This would suggest not using the proportion of DJIA stocks going up on a daily basis as a predictor of the proportion of NYSE stocks going up on that day.

58. At $\mu_0 = 28$, $\alpha = .05$. Note however for this two - tailed test, $z_{\alpha/2} = z_{.025} = 1.96$
At $\mu_a = 29$, $\beta = .15$. $z_{.15} = 1.04$
$\sigma = 6$

$$n = \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(1.96 + 1.04)^2 (6)^2}{(28-29)^2} = 324$$

59. At $\mu_0 = 25$, $\alpha = .02$. $z_{.02} = 2.05$
At $\mu_a = 24$, $\beta = .20$. $z_{.20} = .84$
$\sigma = 3$

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_a)^2} = \frac{(2.05 + .84)^2 (3)^2}{(25-24)^2} = 75.2 \quad \text{Use 76}$$

65. a. H_0: $\mu \geq 6883$ H_a: $\mu < 6883$

b. $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{5980 - 6883}{2518/\sqrt{40}} = -2.268$

Degrees of freedom = $n - 1 = 39$
Lower tail p-value is the area to the left of the test statistic
Using t table: p-value is between .025 and .01
Exact p-value corresponding to t = -2.268 is 0.0145 (one tail)

c. We should conclude that Medicare spending per enrollee in Indianapolis is less than the national average.

d. Using the critical value approach we would:
Reject H_0 if $t \leq -t_{.05}$ = -1.685
Since $t = -2.268 \leq -1.685$, we reject H_0.

67. H_0: $\mu = 2.357$ H_a: $\mu \neq 2.357$

$\bar{x} = \frac{\Sigma x_i}{n} = 2.3496$ $s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}} = .0444$

$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2.3496 - 2.3570}{.0444/\sqrt{50}} = -1.18$

Degrees of freedom = 50 - 1 = 49
Because $t < 0$, p-value is two times the lower tail area
Using t table: area in lower tail is between .10 and .20; therefore, p-value is between .20 and .40.
Exact p-value corresponding to t = -1.18 is .2437
p-value > .05; do not reject H_0.
There is not a statistically significant difference between the National mean price per gallon and the mean price per gallon in the Lower Atlantic states.

73. a. H_0: $p \geq .24$ H_a: $p < .24$

b. $\bar{p} = \frac{81}{400} = .2025$

c. $z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{.2025 - .24}{\sqrt{\dfrac{.24(1-.24)}{400}}} = -1.76$

Lower tail *p*-value is the area to the left of the test statistic
Using normal table with z = -1.76: *p*-value =.0392
p-value ≤ .05; reject H_0.
The proportion of workers not required to contribute to their company sponsored health care plan has declined. There seems to be a trend toward companies requiring employees to share the cost of health care benefits.

10. Statistical Inference about Means and Proportions with Two populations

7. a. μ_1 = Population mean 2002

μ_2 = Population mean 2003

H_0: $\mu_1 - \mu_2 \leq 0$ H_a: $\mu_1 - \mu_2 > 0$

b. With time in minutes, $\bar{x}_1 - \bar{x}_2$ = 172 - 166 = 6 minutes

c. $$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(172-166)-0}{\sqrt{\frac{12^2}{60} + \frac{12^2}{50}}} = 2.61$$ p-value = 1.0000 - .9955 = .0045

p-value $\leq$ .05; reject H_0. The population mean duration of games in 2003 is less than the population mean in 2002.

d. $$\bar{x}_1 - \bar{x}_2 \pm z_{.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = (172-166) \pm 1.96\sqrt{\frac{12^2}{60} + \frac{12^2}{50}} = 6 \pm 4.5 = (1.5 \text{ to } 10.5)$$

e. Percentage reduction: 6/172 = 3.5%. Management should be encouraged by the fact that steps taken in 2003 reduced the population mean duration of baseball games. However, the statistical analysis shows that the reduction in the mean duration is only 3.5%. The interval estimate shows the reduction in the population mean is 1.5 minutes (.9%) to 10.5 minutes (6.1%). Additional data collected by the end of the 2003 season would provide a more precise estimate. In any case, most likely the issue will continue in future years. It is expected that major league baseball would prefer that additional steps be taken to further reduce the mean duration of games.

20. a. 3, -1, 3, 5, 3, 0, 1

b. $\bar{d} = \sum d_i / n = 14/7 = 2$

c. $$s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{26}{7-1}} = 2.08$$

d. $\bar{d} = 2$

e. With 6 degrees of freedom $t_{.025}$ = 2.447, $2 \pm 2.447\left(2.082/\sqrt{7}\right)$ = 2 ± 1.93 = (.07 to 3.93)

23. a. μ_1 = population mean grocery expenditures, μ_2 = population mean dining-out expenditures

H_0: $\mu_d = 0$ H_a: $\mu_d \neq 0$

b. $$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{850-0}{1123/\sqrt{42}} = 4.91$$ $df = n - 1 = 41$ p-value ≈ 0

Conclude that there is a difference between the annual population mean expenditures for groceries and for dining-out.

c. Groceries has the higher mean annual expenditure by an estimated \$850.

$$\bar{d} \pm t_{.025}\frac{s_d}{\sqrt{n}} = 850 \pm 2.020\frac{1123}{\sqrt{42}} = 850 \pm 350 = (500 \text{ to } 1200)$$

25. a. $H_0: \mu_d = 0$ $H_a: \mu_d \neq 0$

Use difference data: -3, -2, -4, 3, -1, -2, -1, -2, 0, 0, -1, -4, -3, 1, 1

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-18}{15} = -1.2 \qquad s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{54.4}{15-1}} = 1.97$$

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{-1.2 - 0}{1.97/\sqrt{15}} = -2.36 \qquad df = n - 1 = 14$$

Using t table, the 1-tail area is between .01 and .025, so the Two-tail p-value is between .02 and .05.
The exact p-value corresponding to t = -2.36 is .0333
Since the p-value $\leq$ .05, reject H_0. Conclude that there is a difference between the population mean weekly usage for the two media.

b. $\bar{x}_{TV} = \frac{\sum x_i}{n} = \frac{282}{15} = 18.8$ hours per week for cable television, $\bar{x}_R = \frac{\sum x_i}{n} = \frac{300}{15} = 20$ for radio.
Radio has greater usage.

31. a. Professional Golfers: $\bar{p}_1$ = 688/1075 = .64, Amateur Golfers: $\bar{p}_2$ = 696/1200 = .58
Professional golfers have the better putting accuracy.

b. $\bar{p}_1 - \bar{p}_2 = .64 - .58 = .06$
Professional golfers make 6% more 6-foot putts than the very best amateur golfers.

c. $$\bar{p}_1 - \bar{p}_2 \pm z_{.025}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} = .64 - .58 \pm 1.96\sqrt{\frac{.64(1-.64)}{1075} + \frac{.58(1-.58)}{1200}} = .06 \pm .04 \ (.02 \text{ to } .10)$$

The confidence interval shows that professional golfers make from 2% to 10% more 6-foot putts than the best amateur golfers.

38. $H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 \neq 0$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(4.1 - 3.4) - 0}{\sqrt{\frac{(2.2)^2}{120} + \frac{(1.5)^2}{100}}} = 2.79$$

p-value = 2(1.0000 - .9974) = .0052
p-value $\leq$ .05, reject H_0. A difference exists with system B having the lower mean checkout time.

41. a. $n_1 = 10$ $n_2 = 8$
$\bar{x}_1 = 21.2$ $\bar{x}_2 = 22.8$
$s_1 = 2.70$ $s_2 = 3.55$
$\bar{x}_1 - \bar{x}_2$ = 21.2 - 22.8 = -1.6 so Kitchens are less expensive by \$1600.

b. $$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{2.70^2}{10} + \frac{3.55^2}{8}\right)^2}{\frac{1}{9}\left(\frac{2.70^2}{10}\right)^2 + \frac{1}{7}\left(\frac{3.55^2}{8}\right)^2} = 12.9$$. Use df = 12, $t_{.05}$ = 1.782

$$-1.6 \pm 1.782\sqrt{\frac{2.70^2}{10}+\frac{3.55^2}{8}}$$ = -1.6 ± 2.7 = (-4.3 to 1.1)

47. $\bar{p}_1 = .276$ Most recent week, $\bar{p}_2 = .487$ One Week Ago, $\bar{p}_3 = .397$ One Month Ago

a. Point estimate = $\bar{p}_1 - \bar{p}_2 = .276 - .487 = -.211$

Margin of error: $z_{.025}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1}+\frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} = 1.96\sqrt{\frac{.276(1-.276)}{240}+\frac{.487(1-.487)}{240}} = .085$

95% confidence interval: -.211 ± .085 (-.296, -.126)

b. H_0: $p_1 - p_3 \geq 0$ H_a: $p_1 - p_3 < 0$

c. $\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_3}{n_1 + n_3} = \frac{(240)(.276)+(240)(.397)}{240+240} = .3365$

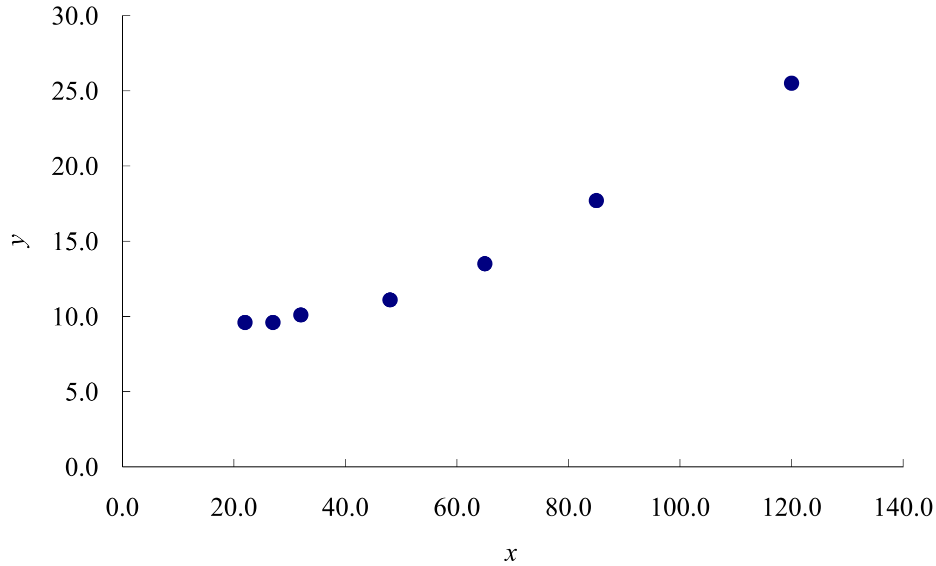
$s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{(.3365)(.6635)\left(\frac{2}{240}\right)} = .0431$

$z = \frac{.276 - .397}{.0431} = -2.81$ p-value = .0025

With p-value ≤ .01, we reject H_0 and conclude that bullish sentiment has declined over the past month.

14. Simple Linear regression

13. a.



b. The summations needed to compute the slope and the y-intercept are:

$\Sigma x_i = 399 \quad \Sigma y_i = 97.1 \quad \Sigma(x_i - \bar{x})(y_i - \bar{y}) = 1233.7 \quad \Sigma(x_i - \bar{x})^2 = 7648$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{1233.7}{7648} = 0.16131$$

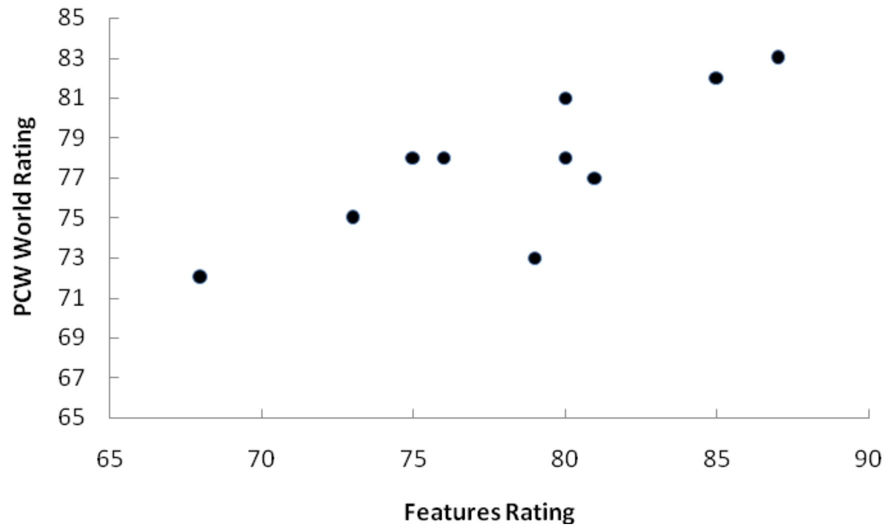
$$b_0 = \bar{y} - b_1\bar{x} = 13.87143 - (0.16131)(57) = 4.67675$$

$$\hat{y} = 4.68 + 0.16x$$

c. $\hat{y} = 4.68 + 0.16x = 4.68 + 0.16(52.5) = 13.08$ or approximately \$13,080.

The agent's request for an audit appears to be justified.

14. a.



b. There appears to be a positive linear relationship between x = features rating and y = PCW World Rating.

c. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{784}{10} = 78.4 \quad \bar{y} = \frac{\Sigma y_i}{n} = \frac{777}{10} = 77.7$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 147.20 \quad \Sigma(x_i - \bar{x})^2 = 284.40$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{147.20}{284.40} = .51758$$

$b_0 = \bar{y} - b_1\bar{x} = 77.7 - (.51758)(78.4) = 37.1217$

$\hat{y} = 37.1217 + .51758x$

d. $\hat{y} = 37.1217 + .51758(70) = 73.35$ or 73

18. a. The estimated regression equation and the mean for the dependent variable are:

$\hat{y} = 1790.5 + 581.1x \quad \bar{y} = 3650$

The sum of squares due to error and the total sum of squares are

SSE $= \sum(y_i - \hat{y}_i)^2 = 85{,}135.14$ SST $= \sum(y_i - \bar{y})^2 = 335{,}000$

Thus, SSR = SST - SSE = 335,000 - 85,135.14 = 249,864.86

b. r^2 = SSR/SST = 249,864.86/335,000 = .746

We see that 74.6% of the variability in y has been explained by the least squares line.

c. $r = \sqrt{.746} = +.8637$

21. a. The summations needed in this problem are:

$\Sigma x_i = 3450 \quad \Sigma y_i = 33{,}700 \quad \Sigma(x_i - \bar{x})(y_i - \bar{y}) = 712{,}500 \quad \Sigma(x_i - \bar{x})^2 = 93{,}750$

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{712{,}500}{93{,}750} = 7.6$$

$b_0 = \bar{y} - b_1\bar{x} = 5616.67 - (7.6)(575) = 1246.67$

$\hat{y} = 1246.67 + 7.6x$

b. $7.60

c. The sum of squares due to error and the total sum of squares are:

SSE $= \sum(y_i - \hat{y}_i)^2 = 233{,}333.33$ SST $= \sum(y_i - \bar{y})^2 = 5{,}648{,}333.33$

Thus, SSR = SST - SSE = 5,648,333.33 - 233,333.33 = 5,415,000

r^2 = SSR/SST = 5,415,000/5,648,333.33 = .9587

We see that 95.87% of the variability in y has been explained by the estimated regression equation.

d. $\hat{y} = 1246.67 + 7.6x = 1246.67 + 7.6(500) = \$5046.67$

35. a. s = 145.89

$\bar{x} = 3.2 \quad \Sigma(x_i - \bar{x})^2 = 0.74$

$$s_{\hat{y}_p} = s\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 145.89\sqrt{\frac{1}{6} + \frac{(3-3.2)^2}{0.74}} = 68.54$$

$\boldsymbol{\hat{y}_p} = 1790.54 + 581.08\boldsymbol{x} = 1790.54 + 581.08(3) = 3533.78$

$\hat{y}_p \pm t_{\alpha/2}s_{\hat{y}_p}$ = 3533.78 ± 2.776 (68.54) = 3533.78 ± 190.27 or $3343.51 to $3724.05

b. $$s_{ind} = s\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} = 145.89\sqrt{1 + \frac{1}{6} + \frac{(3-3.2)^2}{0.74}} = 161.19$$

$\hat{y}_p \pm t_{\alpha/2}s_{ind}$ = 3533.78 ± 2.776 (161.19) = 3533.78 ± 447.46 or $3086.32 to $3981.24

44. a/b. The scatter diagram shows a linear relationship between the two variables.

c. The Minitab output is shown below:

```
The regression equation is
Rental$ = 37.1 - 0.779 Vacancy%

Predictor        Coef     SE Coef          T        P
Constant       37.066       3.530      10.50    0.000
Vacancy%      -0.7791      0.2226      -3.50    0.003

S = 4.889       R-Sq = 43.4%     R-Sq(adj) = 39.8%

Analysis of Variance

Source            DF          SS          MS          F        P
Regression         1      292.89      292.89      12.26    0.003
Residual Error    16      382.37       23.90
Total             17      675.26

Predicted Values for New Observations

New Obs     Fit     SE Fit       95.0% CI              95.0% PI
1         17.59       2.51   ( 12.27,   22.90)  (     5.94,   29.23)
2         28.26       1.42   ( 25.26,   31.26)  (    17.47,   39.05)

Values of Predictors for New Observations

New Obs  Vacancy%
1            25.0
2            11.3
```

d. Since the p-value = 0.003 is less than $\alpha = .05$, the relationship is significant.

e. $r^2 = .434$. The least squares line does not provide a very good fit.

f. The 95% confidence interval is 12.27 to 22.90 or \$12.27 to \$22.90.

g. The 95% prediction interval is 17.47 to 39.05 or \$17.47 to \$39.05.

47. a. Let x = advertising expenditures and y = revenue

$\hat{y} = 29.4 + 1.55x$

b. SST = 1002   SSE = 310.28   SSR = 691.72

MSR = SSR / 1 = 691.72

MSE = SSE / (n - 2) = 310.28/ 5 = 62.0554

F = MSR / MSE = 691.72/ 62.0554= 11.15

$F_{.05}$ = 6.61 (1 degree of freedom numerator and 5 denominator)

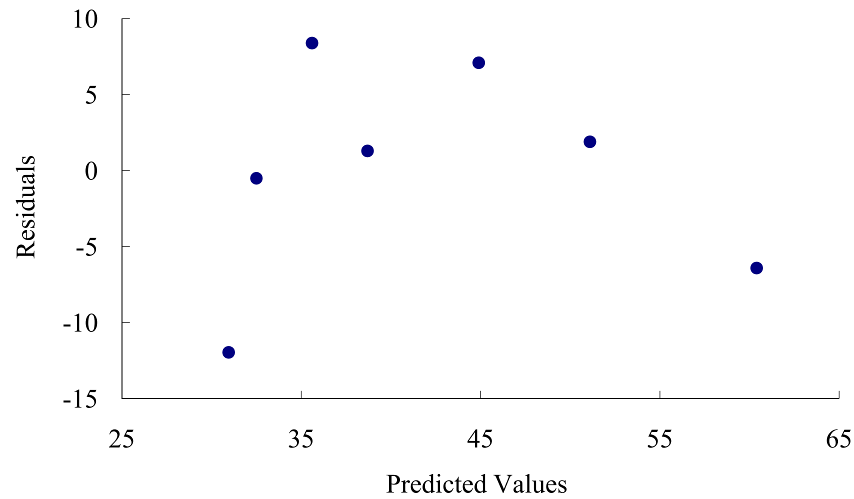
Since $F = 11.15 > F_{.05} = 6.61$ we conclude that the two variables are related.

Or: Using F table (1 degree of freedom numerator and 5 denominator), p-value is between .01 and .025

Using Excel or Minitab, the p-value corresponding to F = 11.15 is .0206.

Because p-value $\leq \alpha = .05$, we conclude that the two variables are related.

c.

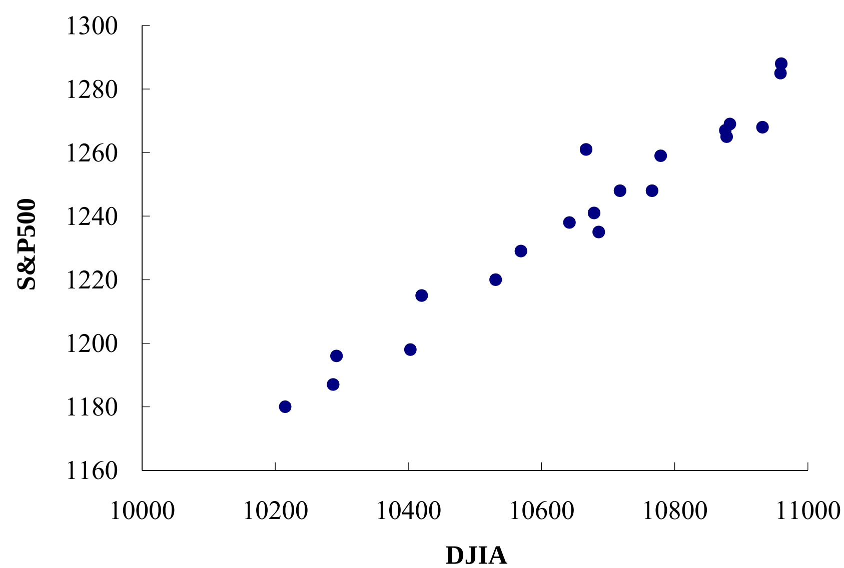


d. The residual plot leads us to question the assumption of a linear relationship between x and y. Even though the relationship is significant at the .05 level of significance, it would be extremely dangerous to extrapolate beyond the range of the data.

55. No. Regression or correlation analysis can never prove that two variables are casually related.

57. The purpose of testing whether $\beta_1 = 0$ is to determine whether or not there is a significant relationship between x and y. However, rejecting $\beta_1 = 0$ does not necessarily imply a good fit. For example, if $\beta_1 = 0$ is rejected and r^2 is low, there is a statistically significant relationship between x and y but the fit is not very good.

60. a.



b. The Minitab output is shown below:

```
The regression equation is
S&P500 = - 182 + 0.133 DJIA

Predictor       Coef   SE Coef       T      P
Constant     -182.11     71.83   -2.54  0.021
DJIA        0.133428  0.006739   19.80  0.000


S = 6.89993   R-Sq = 95.6%   R-Sq(adj) = 95.4%

Analysis of Variance

Source          DF     SS     MS       F      P
Regression       1  18666  18666  392.06  0.000
Residual Error  18    857     48
Total           19  19523
```

c. Using the F test, the p-value corresponding to F = 392.06 is .000. Because the p-value $\leq \alpha$ =.05, we reject $H_0: \beta_1 = 0$; there is a significant relationship.

d. With R-Sq = 95.6%, the estimated regression equation provided an excellent fit.

e. $\hat{y} = -182.11 + .133428\text{DJIA} = -182.11 + .133428(11,000) = 1285.60$ or 1286.

f. The DJIA is not that far beyond the range of the data. With the excellent fit provided by the estimated regression equation, we should not be too concerned about using the estimated regression equation to predict the S&P500.

15. Multiple Regression

5. a. The Minitab output is shown below:

```
The regression equation is
Revenue = 88.6 + 1.60 TVAdv

Predictor        Coef     SE Coef          T        P
Constant       88.638       1.582      56.02    0.000
TVAdv          1.6039      0.4778       3.36    0.015

S = 1.215       R-Sq = 65.3%     R-Sq(adj) = 59.5%

Analysis of Variance

Source            DF          SS          MS          F        P
Regression         1      16.640      16.640      11.27    0.015
Residual Error     6       8.860       1.477
Total              7      25.500
```

b. The Minitab output is shown below:

```
The regression equation is
Revenue = 83.2 + 2.29 TVAdv + 1.30 NewsAdv

Predictor        Coef     SE Coef          T        P
Constant       83.230       1.574      52.88    0.000
TVAdv          2.2902      0.3041       7.53    0.001
NewsAdv        1.3010      0.3207       4.06    0.010

S = 0.6426      R-Sq = 91.9%     R-Sq(adj) = 88.7%

Analysis of Variance

Source            DF          SS          MS          F        P
Regression         2      23.435      11.718      28.38    0.002
Residual Error     5       2.065       0.413
Total              7      25.500
```

c. No, it is 1.60 in part (a) and 2.29 above. In part (b) it represents the marginal change in revenue due to an increase in television advertising with newspaper advertising held constant.

d. Revenue = 83.2 + 2.29(3.5) + 1.30(1.8) = $93.56 or $93,560

7. a. The Minitab output is shown below:

```
The regression equation is
PCW Rating = 66.1 + 0.170 Performance

Predictor       Coef   SE Coef      T      P
Constant      66.062     3.793  17.42  0.000
Performance  0.16989   0.05407   3.14  0.014

S = 2.59221   R-Sq = 55.2%   R-Sq(adj) = 49.6%

Analysis of Variance
```

```
Source          DF       SS      MS     F      P
Regression       1   66.343  66.343  9.87  0.014
Residual Error   8   53.757   6.720
Total            9  120.100
```

b. The Minitab output is shown below:

```
The regression equation is
PCW Rating = 40.0 + 0.113 Performance + 0.382 Features

Predictor       Coef  SE Coef     T      P
Constant      39.982    7.855  5.09  0.001
Performance  0.11338  0.03846  2.95  0.021
Features      0.3820   0.1093  3.49  0.010

S = 1.67285   R-Sq = 83.7%   R-Sq(adj) = 79.0%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       2  100.511  50.255  17.96  0.002
Residual Error   7   19.589   2.798
Total            9  120.100
```

Note that the coefficient of Performance changed slightly when Features is included in the model. But there is a huge increase in the Adjusted R-Squared, and both variables have low p-values in part b. Hence we can expect better predictions from the 2-variable model.

c. $\hat{y} = 40.0 + .113(80) + .382(70) = 75.78$ or 76

7. a. The Minitab output is shown below:

```
The regression equation is
Price = 356 - 0.0987 Capacity + 123 Comfort

Predictor        Coef     SE Coef          T        P
Constant        356.1       197.2       1.81    0.114
Capacity     -0.09874     0.04588      -2.15    0.068
Comfort        122.87       21.80       5.64    0.001

S = 51.14       R-Sq = 83.2%     R-Sq(adj) = 78.4%

Analysis of Variance

Source            DF          SS           MS         F         P
Regression         2       90548        45274     17.31     0.002
Residual Error     7       18304         2615
Total              9      108852
```

b. $b_1 = -.0987$ is an estimate of the change in the price with respect to a 1 cubic inch change in capacity with the comfort rating held constant. $b_2 = 123$ is an estimate of the change in the price with respect to a 1 unit change in the comfort rating with the capacity held constant.

c. $\hat{y} = 356 - .0987(4500) + 123 (4) = 404$

23. Note: The Minitab output is shown in Exercise 5

a. $F = 28.38$

Using F table (2 degrees of freedom numerator and 5 denominator), p-value is less than .01

Actual p-value = .002

Because p-value $\leq \alpha$, there is a significant relationship.

b. $t = 7.53$
Using t table (5 degrees of freedom), area in tail is less than .005; p-value is less than .01
Actual p-value = .001
Because p-value $\leq \alpha$, β_1 is significant and x_1 should not be dropped from the model.

c. $t = 4.06$
Actual p-value = .010
Because p-value $\leq \alpha$, β_2 is significant and x_2 should not be dropped from the model.

NOTE: These answers seem to imply that a variable whose p-value is above alpha should be dropped.
THAT IS NOT NECESSARILY TRUE!

29. a. $\hat{y} = 83.2 + 2.29(3.5) + 1.30(1.8) = 93.555$ or \$93,555

More accurate answer: In Exercise 5b, the Minitab output shows that $b_0 = 83.230$, $b_1 = 2.2902$, and $b_2 = 1.3010$; hence, $\hat{y} = 83.230 + 2.2902x_1 + 1.3010x_2$. Using this estimated regression equation, we obtain
$\hat{y} = 83.230 + 2.2902(3.5) + 1.3010(1.8) = 93.588$ or \$93,588

The difference between these two estimates (\$93,588 - \$93,555 = \$33) is simply due to the fact that additional significant digits are used in Minitab's computations.

The Minitab output is shown below:

```
     Fit   Stdev.Fit          95% C.I.          95% P.I.
  93.588       0.291   ( 92.840, 94.335)  ( 91.774, 95.401)
```

Note that the value of FIT ($\hat{y}$) is 93.588.

b. Confidence interval estimate: 92.840 to 94.335 or \$92,840 to \$94,335
c. Prediction interval estimate: 91.774 to 95.401 or \$91,774 to \$95,401

34. a. \$15,300
b. Estimate of sales = 10.1 - 4.2(2) + 6.8(8) + 15.3(0) = 56.1 or \$56,100
c. Estimate of sales = 10.1 - 4.2(1) + 6.8(3) + 15.3(1) = 41.6 or \$41,600

35. a. Let Type = 0 if a mechanical repair
Type = 1 if an electrical repair
The Minitab output is shown below:

```
The regression equation is
Time = 3.45 + 0.617 Type

Predictor         Coef      SE Coef          T        P
Constant        3.4500       0.5467       6.31    0.000
Type            0.6167       0.7058       0.87    0.408

S = 1.093        R-Sq = 8.7%      R-Sq(adj) = 0.0%

Analysis of Variance

Source            DF          SS          MS          F        P
Regression         1       0.913       0.913       0.76    0.408
Residual Error     8       9.563       1.195
Total              9      10.476
```

b. The estimated regression equation did not provide a good fit. In fact, the p-value of .408 shows that the relationship is not significant for any reasonable value of α.
c. Person = 0 if Bob Jones performed the service and Person = 1 if Dave Newton performed the service. The Minitab output is shown below:

```
The regression equation is
Time = 4.62 - 1.60 Person

Predictor        Coef     SE Coef          T        P
Constant       4.6200      0.3192      14.47    0.000
Person        -1.6000      0.4514      -3.54    0.008

S = 0.7138      R-Sq = 61.1%     R-Sq(adj) = 56.2%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         1      6.4000      6.4000     12.56    0.008
Residual Error     8      4.0760      0.5095
Total              9     10.4760
```

d. We see that 61.1% of the variability in repair time has been explained by the repair person that performed the service; an acceptable, but not good, fit.

36. a. The Minitab output is shown below:

```
The regression equation is
Time = 1.86 + 0.291 Months + 1.10 Type - 0.609 Person

Predictor        Coef     SE Coef          T        P
Constant       1.8602      0.7286       2.55    0.043
Months        0.29144     0.08360       3.49    0.013
Type           1.1024      0.3033       3.63    0.011
Person        -0.6091      0.3879      -1.57    0.167

S = 0.4174      R-Sq = 90.0%     R-Sq(adj) = 85.0%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      9.4305      3.1435     18.04    0.002
Residual Error     6      1.0455      0.1743
Total              9     10.4760
```

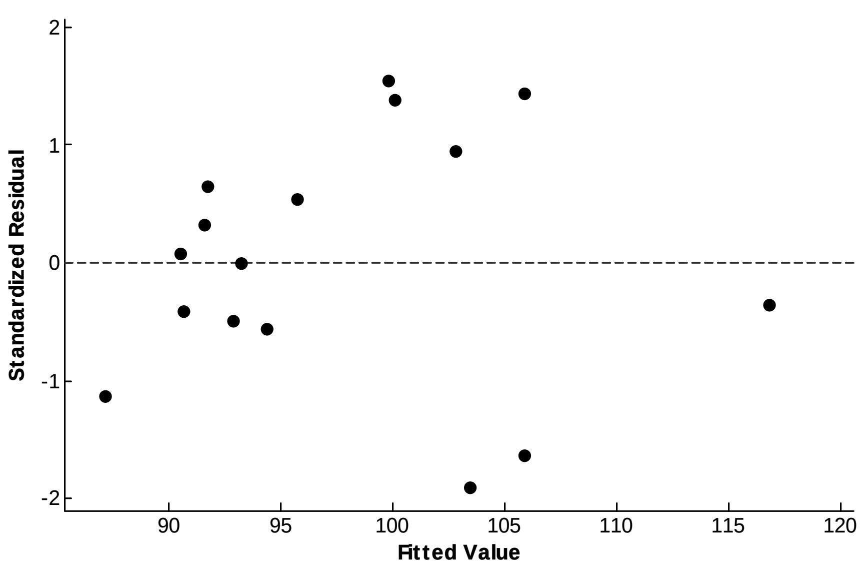
b. Since the p-value corresponding to $F = 18.04$ is $.002 < \alpha = .05$, the overall model is statistically significant.

c. The p-value corresponding to $t = -1.57$ is $.167 > \alpha = .05$; thus, the addition of Person is not statistically significant. Person is highly correlated with Months (the sample correlation coefficient is -.691); thus, once the effect of Months has been accounted for, Person will not add much to the model.

42. a. The Minitab output is shown below:

```
The regression equation is
Speed = 71.3 + 0.107 Price + 0.0845 Horsepwr

Predictor        Coef     SE Coef          T        P
Constant       71.328       2.248      31.73    0.000
Price         0.10719     0.03918       2.74    0.017
Horsepwr     0.084496    0.009306       9.08    0.000

S = 2.485       R-Sq = 91.9%     R-Sq(adj) = 90.7%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         2      915.66      457.83     74.12     0.000
Residual Error    13       80.30        6.18
Total             15      995.95

Source       DF      Seq SS
Price         1      406.39
Horsepwr      1      509.27

Unusual Observations
Obs      Price     Speed      Fit     SE Fit   Residual   St Resid
  2       93.8   108.000  105.882      2.007      2.118     1.45 X

X denotes an observation whose X value gives it large influence.
```

b. The standardized residual plot is shown below. There appears to be a very unusual trend in the standardized residuals.



c. The Minitab output shown in part (a) did not identify any observations with a large standardized residual; thus, there does not appear to be any outliers in the data.

d. The Minitab output shown in part (a) identifies observation 2 as an influential observation.

16. Regression Analysis: Model Building

4. a. The Minitab output is shown below:

```
The regression equation is
Y = 943 + 8.71 X

Predictor       Coef       Stdev    t-ratio        p
Constant      943.05       59.38      15.88    0.000
X              8.714       1.544       5.64    0.005

s = 32.29       R-sq = 88.8%     R-sq(adj) = 86.1%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    1       33223       33223     31.86    0.005
Error         4        4172        1043
Total         5       37395
```

b. p-value = .005 < α = .01; reject H_0

5. The Minitab output is shown below:

```
The regression equation is
Y = 433 + 37.4 X - 0.383 XSQ

Predictor       Coef       Stdev    t-ratio        p
Constant       432.6       141.2       3.06    0.055
X             37.429       7.807       4.79    0.017
XSQ          -0.3829      0.1036      -3.70    0.034

s = 15.83       R-sq = 98.0%     R-sq(adj) = 96.7%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    2       36643       18322     73.15    0.003
Error         3         751         250
Total         5       37395
```

b. Since the linear relationship was significant (Exercise 4), this relationship must be significant. Note also that since the p-value of .003 < α = .05, we can reject H_0.

c. The fitted value is 1302.01, with a standard deviation of 9.93. The 95% confidence interval is 1270.41 to 1333.61; the 95% prediction interval is 1242.55 to 1361.47.

12. a. A portion of the Minitab output follows:

```
The regression equation is
Scoring Avg. = 46.3 + 14.1 Putting Avg.


Predictor       Coef  SE Coef     T      P
Constant      46.277    6.026  7.68  0.000
Putting Avg.  14.103    3.356  4.20  0.000


S = 0.510596   R-Sq = 38.7%   R-Sq(adj) = 36.5%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   4.6036  4.6036  17.66  0.000
Residual Error  28   7.2998  0.2607
Total           29  11.9035
```

b. A portion of the Minitab output follows:

```
The regression equation is
Scoring Avg. = 59.0 - 10.3 Greens in Reg. + 11.4 Putting Avg. - 1.81 Sand
  Saves

Predictor          Coef  SE Coef      T      P
Constant         59.022    5.774  10.22  0.000
Greens in Reg.  -10.281    2.877  -3.57  0.001
Putting Avg.     11.413    2.760   4.14  0.000
Sand Saves      -1.8130   0.9210  -1.97  0.060


S = 0.407808   R-Sq = 63.7%   R-Sq(adj) = 59.5%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       3   7.5795  2.5265  15.19  0.000
Residual Error  26   4.3240  0.1663
Total           29  11.9035
```

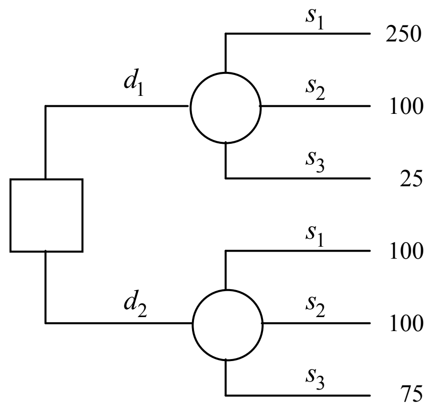
c. SSE(reduced) = 7.2998   SSE(full) = 4.3240   MSE(full) = .1663

$$F = \frac{\dfrac{\text{SSE(reduced) - SSE(full)}}{\text{number of extra terms}}}{\text{MSE(full)}} = \frac{\dfrac{7.2998 - 4.3240}{2}}{.1663} = 8.95$$

The *p*-value associated with $F = 8.95$ (2 degrees of freedom numerator and 26 denominator) is .001. With a *p*-value < α =.05, the addition of the two independent variables is statistically significant.

21. Decision Analysis

1. a.



b. EV(d_1) = .65(250) + .15(100) + .20(25) = 182.5
EV(d_2) = .65(100) + .15(100) + .20(75) = 95
The optimal decision is d_1

4. a. The decision to be made is to choose the type of service to provide. The chance event is the level of demand for the Myrtle Air service. The consequence is the amount of quarterly profit. There are two decision alternatives (full price and discount service). There are two outcomes for the chance event (strong demand and weak demand).

b. EV(Full) = 0.7(960) + 0.3(-490) = 525
EV(Discount) = 0.7(670) + 0.3(320) = 565
Optimal Decision: Discount service

c. EV(Full) = 0.8(960) + 0.2(-490) = 670
EV(Discount) = 0.8(670) + 0.2(320) = 600
Optimal Decision: Full price service

7. a. EV(Small) = 0.1(400) + 0.6(500) + 0.3(660) = 538
EV(Medium) = 0.1(-250) + 0.6(650) + 0.3(800) = 605
EV(Large) = 0.1(-400) + 0.6(580) + 0.3(990) = 605

Best decision: Build a medium or large-size community center.

Note that using the expected value approach, the Town Council would be indifferent between building a medium-size community center and a large-size center.

b. The Town's optimal decision strategy based on perfect information is as follows:

If the worst-case scenario, build a small-size center
If the base-case scenario, build a medium-size center
If the best-case scenario, build a large-size center

Using the consultant's original probability assessments for each scenario, 0.10, 0.60 and 0.30, the expected value of a decision strategy that uses perfect information is:

EVwPI = 0.1(400) + 0.6(650) + 0.3(990) = 727

In part (a), the expected value approach showed that EV(Medium) = EV(Large) = 605.

Therefore, EVwoPI = 605 and EVPI = 727 - 605 = 122

The town should seriously consider additional information about the likelihood of the three scenarios. Since perfect information would be worth $122,000, a good market research study could possibly make a significant contribution.

c. EV(Small) = 0.2(400) + 0.5(500) + 0.3(660) = 528
EV(Medium) = 0.2(-250) + 0.5(650) + 0.3(800) = 515
EV(Large) = 0.2(-400) + 0.5(580) + 0.3(990) = 507

Best decision: Build a small-size community center.

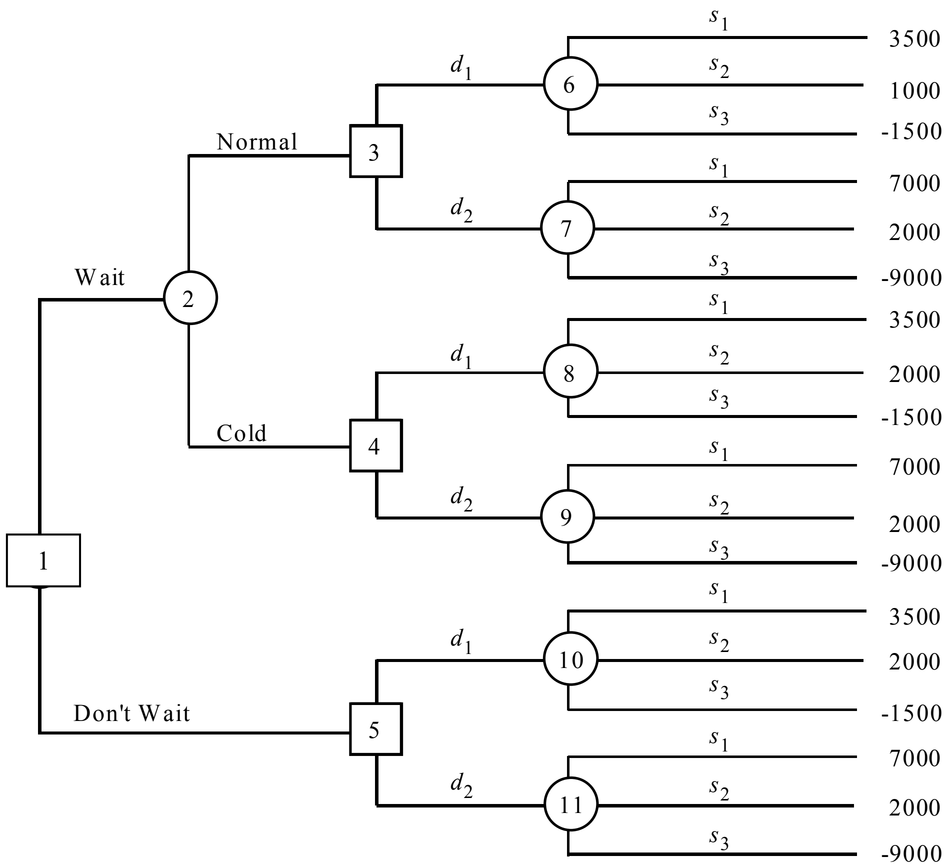
d. If the promotional campaign is conducted, the probabilities will change to 0.0, 0.6 and 0.4 for the worst case, base case and best case scenarios respectively.

EV(Small) = 0.0(400) + 0.6(500) + 0.4(660) = 564
EV(Medium) = 0.0(-250) + 0.6(650) + 0.4(800) = 710
EV(Large) = 0.0(-400) + 0.6(580) + 0.4(990) = 744

In this case, the recommended decision is to build a large-size community center. Compared to the analysis in Part (a), the promotional campaign has increased the best expected value by $744,000 - 605,000 = $139,000. Compared to the analysis in part (c), the promotional campaign has increased the best expected value by $744,000 - 528,000 = $216,000.

Even though the promotional campaign does not increase the expected value by more than its cost ($150,000) when compared to the analysis in part (c), it appears to be a good investment. That is, it eliminates the risk of a loss, which appears to be a significant factor in the mayor's decision-making process.

12. a.



b. Using Node 5,

EV (node 10) = 0.4(3500) + 0.3(1000) + 0.3(-1500) = 1250

EV (node 11) = 0.4(7000) + 0.3(2000) + 0.3(-9000) = 700

Decision: d_1 Blade attachment Expected Value \$1250 (at Node 5)

c. EVwPI = 0.4(7000) + 0.3(2000) + 0.3(-1500) = \$2950

EVPI = \$2950 - \$1250 = \$1700

d. EV (node 6) = 0.35(3500) + 0.30(1000) + 0.35(-1500) = 1000

EV (node 7) = 0.35(7000) + 0.30(2000) + 0.35(-9000) = -100

EV (node 8) = 0.62(3500) + 0.31(1000) + 0.07(-1500) = 2375

EV (node 9) = 0.62(7000) + 0.31(2000) + 0.07(-9000) = 4330

EV (node 3) = Max(1000,-100) = 1000 d_1 Blade attachment

EV (node 4) = Max(2375,4330) = 4330 d_2 New snowplow

The expected value of node 2 is

EV (node 2) = 0.8 EV(node 3) + 0.2 EV(node 4)

= 0.8(1000) + 0.2(4330) = 1666

EV (node 1) = Max(node 2, node 5) = Max(1666,1250) = \$1666 Wait

The optimal strategy is

"Wait until September and then,

If normal weather, choose the blade attachment,

but if unseasonably cold, choose snowplow"