

The Managing Ashland MultiComm Services Case

Chapter 1

1. The company's own historical data would be a primary data source. The research team might use secondary data sources published by some private market research companies, e.g., GoodData[®], Market Analyst[™], M/A/R/C Research, etc. or secondary data sources published by governmental agencies, e.g., the Federal Communications Commission, Broadband.com, etc.
2. The team might conduct a survey using any of the probability samples.
3. Operational definitions like "Internet-based methods of program delivery", "marketplace" or "subscription renewal rates" will need to be properly defined so that there is no potential confusion on what constitute "Internet-based methods of program delivery", which marketplace they should be concentrating on or what constitutes renewal of subscriptions.

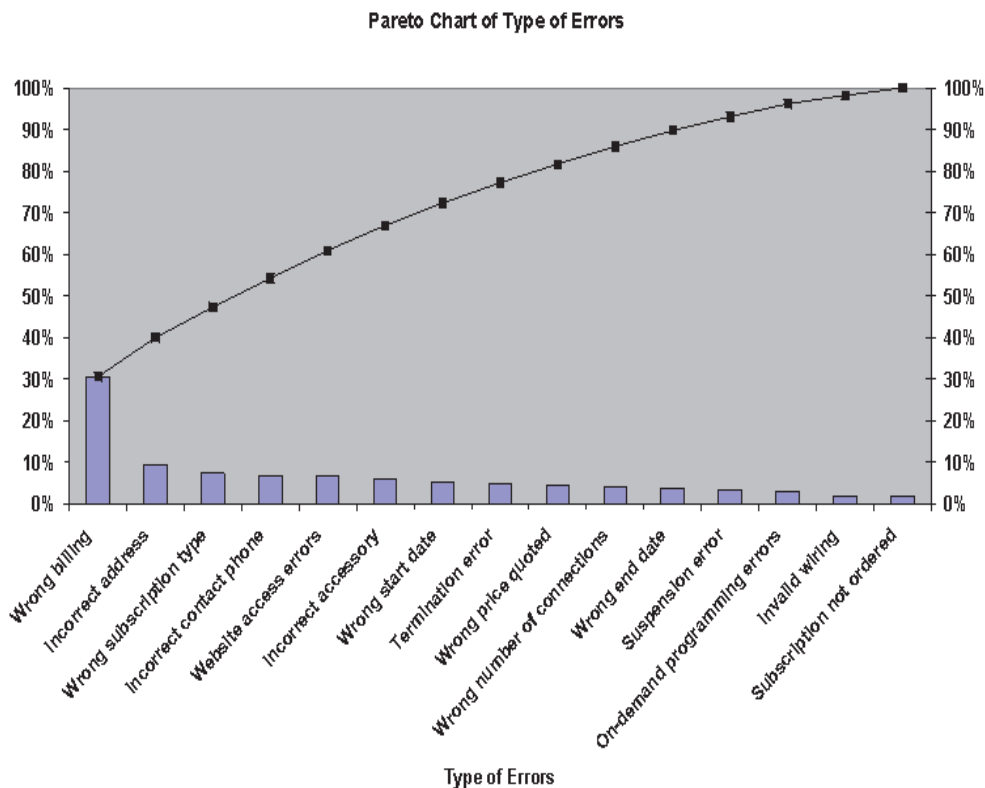
Chapter 2

1. You need to examine the type of customer service errors, the cost of customer service errors, and then within the large category of wrong billing errors, look at the type of wrong billing errors.

Since there are many categories of customer service errors, you should construct a Pareto chart for the type and cost of errors to help you focus on the vital few categories that account for most of the errors.

<u>Type of Errors</u>	<u>Frequency</u>	<u>Percentage</u>	<u>Cumulative Pct.</u>
Wrong billing	137	30.58%	30.58%
Incorrect address	42	9.38%	39.96%
Wrong subscription type	33	7.37%	47.32%
Incorrect contact phone	31	6.92%	54.24%
Website access errors	30	6.70%	60.94%
Incorrect accessory	27	6.03%	66.96%
Wrong start date	24	5.36%	72.32%
Termination error	22	4.91%	77.23%
Wrong price quoted	20	4.46%	81.70%
Wrong number of connections	19	4.24%	85.94%
Wrong end date	17	3.79%	89.73%
Suspension error	15	3.35%	93.08%
On-demand programming errors	14	3.13%	96.21%
Invalid wiring	9	2.01%	98.21%
Subscription not ordered	8	1.79%	100.00%

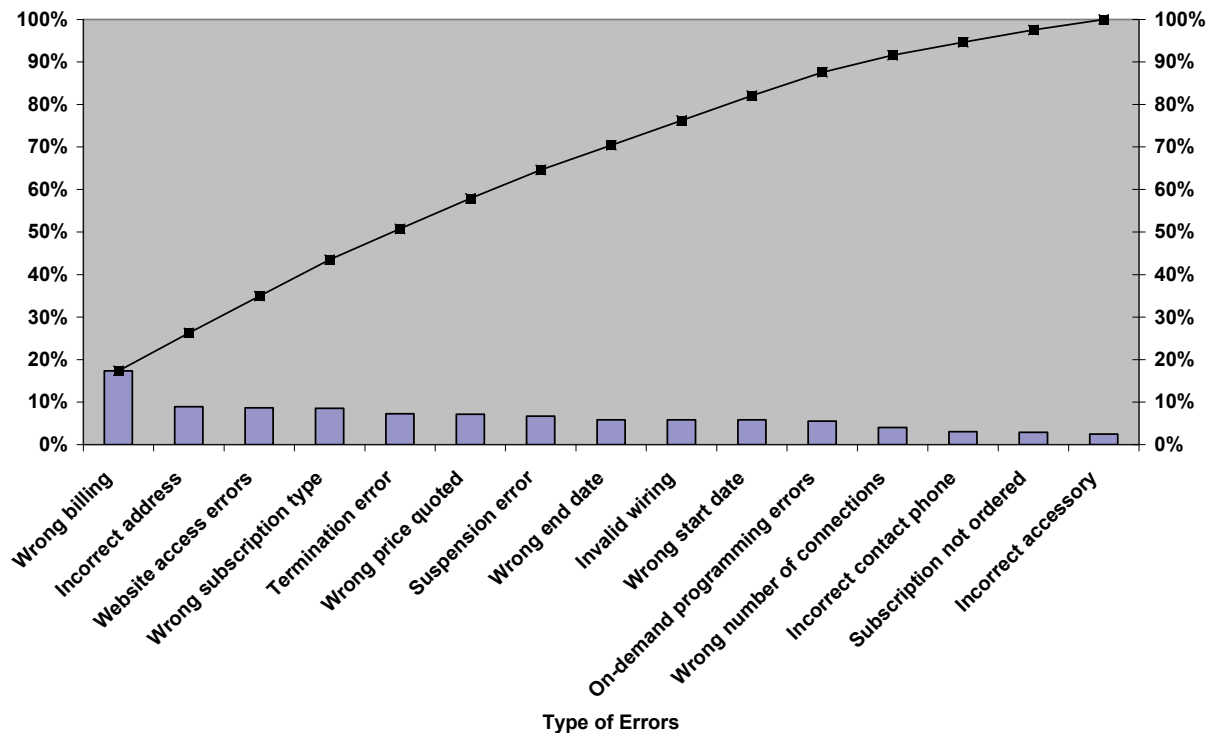
596 Solutions to Managing Ashland MultiComm Services Case



Observe that the first six categories of wrong billing, incorrect addresses, wrong subscription type, incorrect contact phone, website access errors, and incorrect accessory account for about two-thirds of the billing errors. Wrong billing alone accounts for more than 30% of the errors. However, before proceeding you need to examine the cost of the billing errors in addition to the frequency.

<u>Type of Errors</u>	<u>Cost(\$000)</u>	<u>Percentage</u>	<u>Cumulative Pct.</u>
Wrong billing	121.7	17.36%	17.36%
Incorrect address	62.4	8.90%	26.25%
Website access errors	60.7	8.66%	34.91%
Wrong subscription type	60.1	8.57%	43.48%
Termination error	50.9	7.26%	50.74%
Wrong price quoted	50.3	7.17%	57.92%
Suspension error	46.8	6.67%	64.59%
Wrong end date	40.9	5.83%	70.42%
Invalid wiring	40.8	5.82%	76.24%
Wrong start date	40.8	5.82%	82.06%
On-demand programming errors	38.8	5.53%	87.59%
Wrong number of connections	28.1	4.01%	91.60%
Incorrect contact phone	21.3	3.04%	94.64%
Subscription not ordered	20.3	2.90%	97.53%
Incorrect accessory	17.3	2.47%	100.00%

Pareto Chart of the Cost of Errors

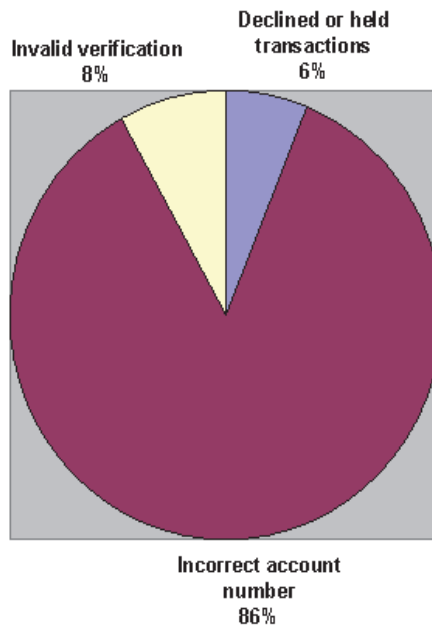


Observe that the first six categories of wrong billing, incorrect addresses, website access errors, wrong subscription type, termination errors, and wrong price quoted account for just less than 60% of the cost of billing errors. Wrong billing alone accounts for more than 17% of the errors.

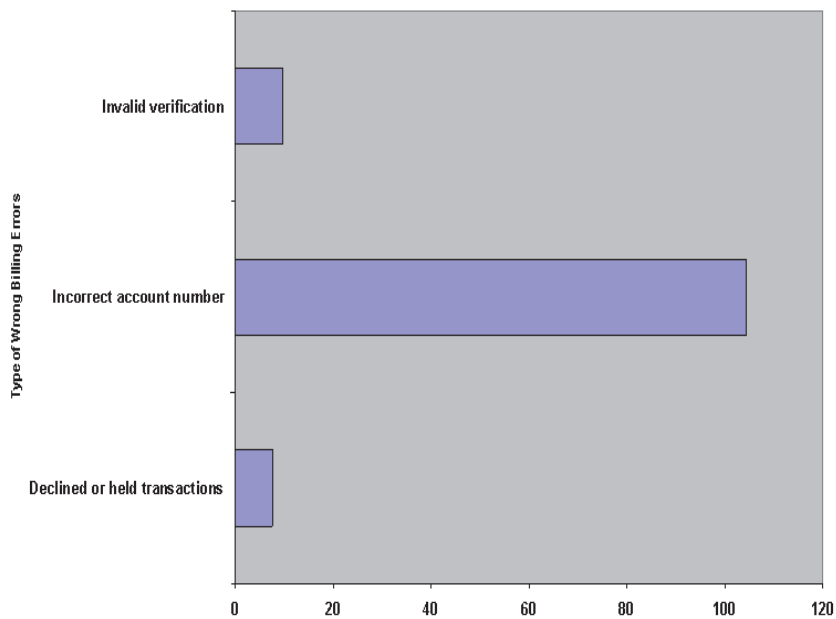
Certainly wrong billing errors should be investigated first, followed by incorrect addresses, website access errors, and wrong subscription type .

Examining wrong billing error costs, you can construct bar and pie charts since there are only three categories.

Type of Wrong Billing Errors(\$)

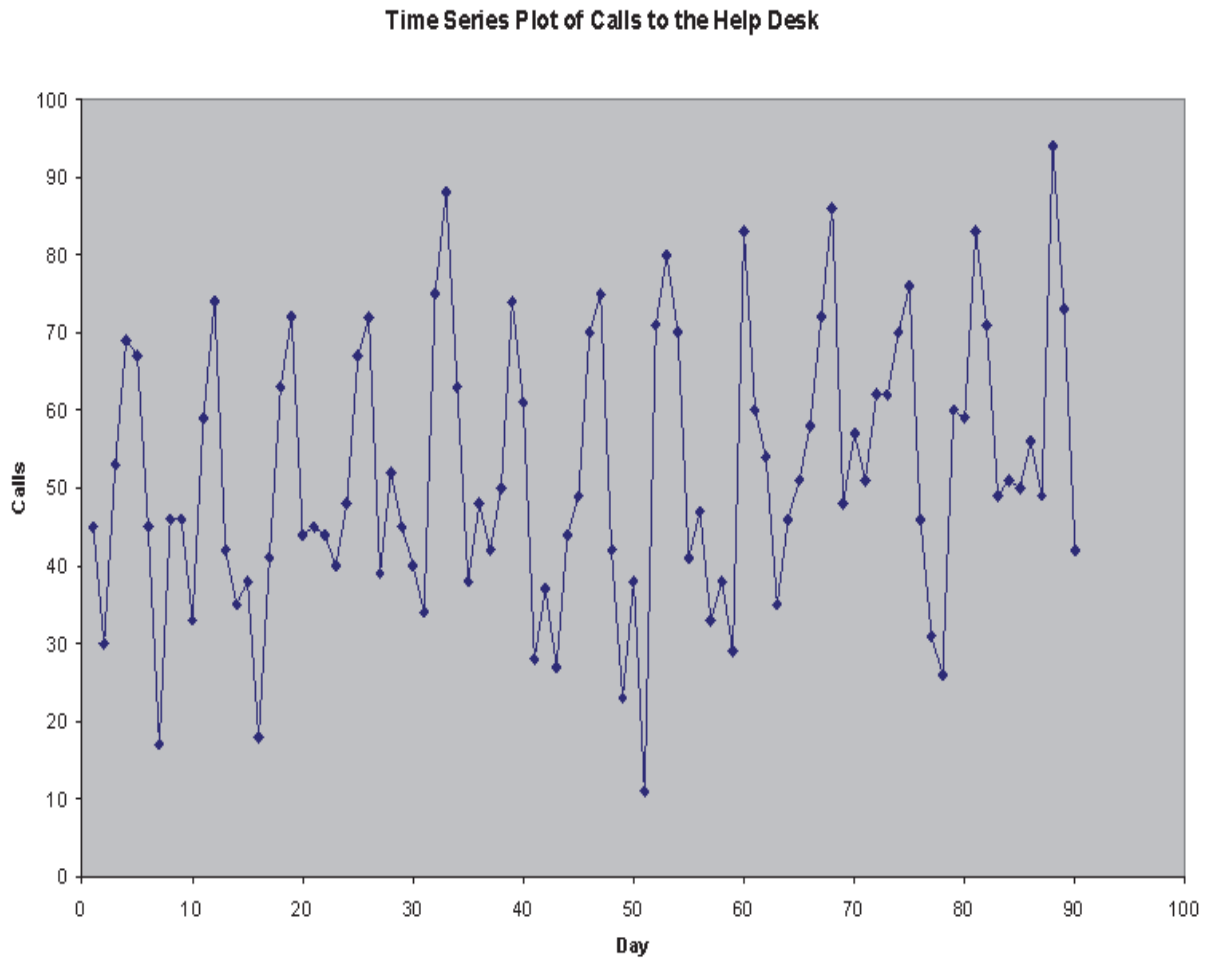


Type of Wrong Billing Errors(\$)



You can see that almost 90% of the cost of wrong billing errors is due to incorrect account numbers, so the first action for the task force or management to take is to reduce these errors.

2. Since the data for calls to the help desk are collected over a period of time, before constructing charts, you need to plot the data over time to determine if there is a pattern.



There does not appear to be any trend in the number of daily calls over the time period studied, so you can now construct other charts, such as a stem-and-leaf display, histogram, and polygons.

Stem-and-Leaf Display for Calls

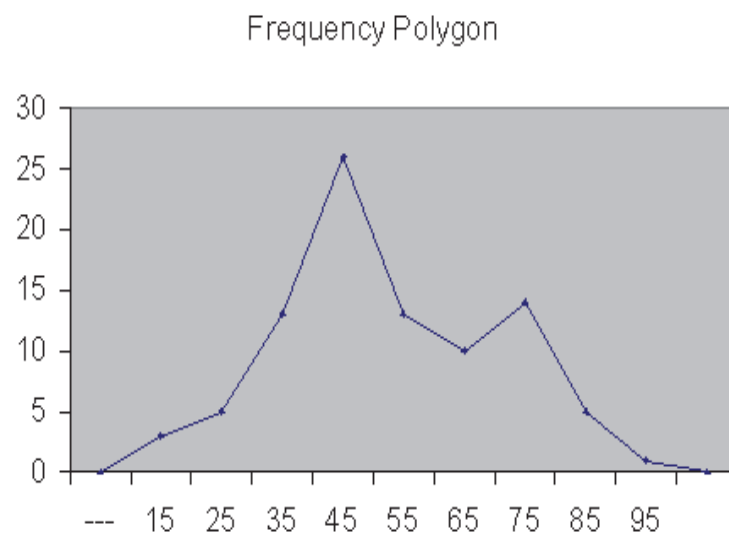
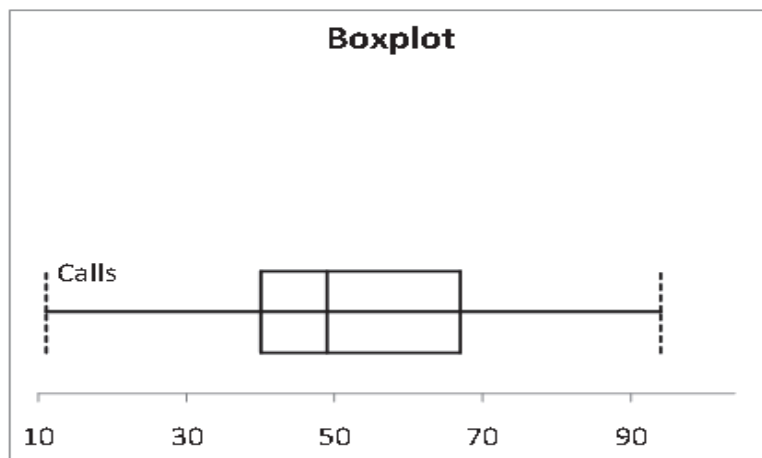
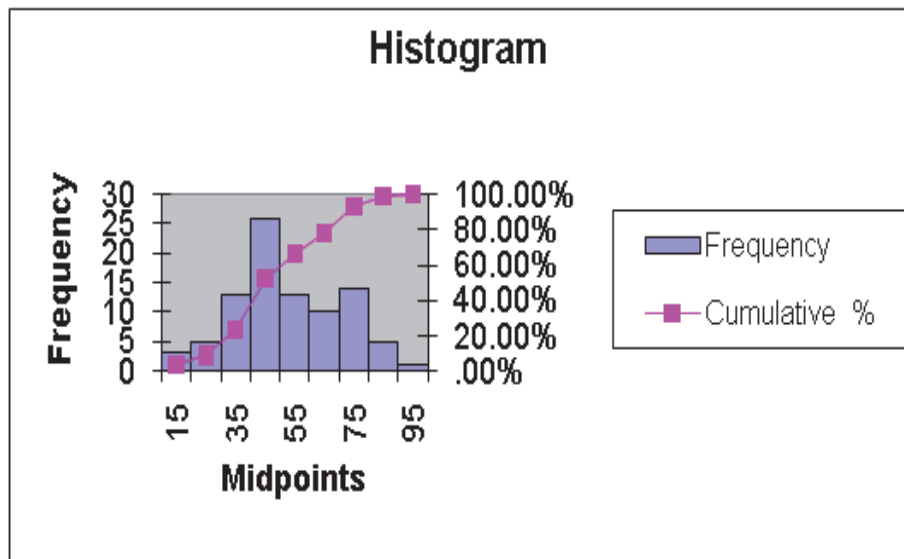
Stem 10
unit:

Statistics	
Sample Size	90
Mean	51.84444
Median	49
Std. Deviation	17.42909
Minimum	11
Maximum	94

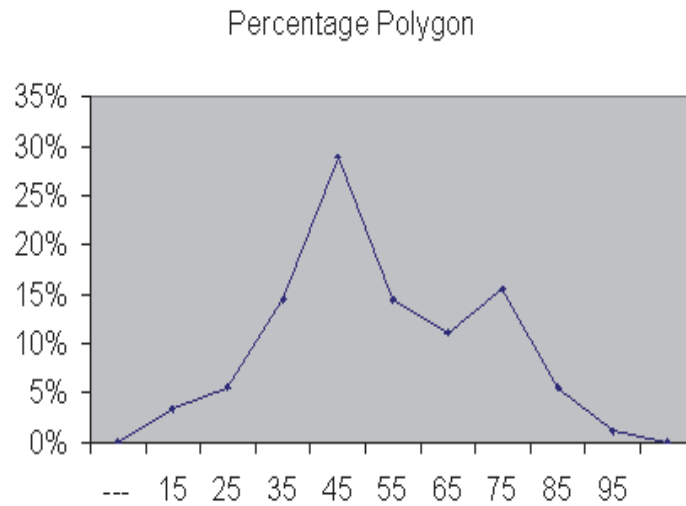
1	1 7 8
2	3 6 7 8 9
3	0 1 3 3 4 5 5 7 8 8 8 9
4	0 0 1 1 2 2 2 2 4 4 4 5 5 5 5 6 6 6 6 7 8 8 8 9 9 9
5	0 0 1 1 1 2 3 4 6 7 8 9 9
6	0 0 1 2 2 3 3 7 7 9
7	0 0 0 1 1 2 2 2 3 4 4 5 5 6
8	0 3 3 6 8
9	4

Frequencies (Calls)

<i>Class</i>	<i>Frequency</i>	<i>Percentage</i>	<i>Cumulative %</i>	<i>Midpoints</i>
10 but less than 19.99	3	3.33%	3.33%	15
20 but less than 29.99	5	5.56%	8.89%	25
30 but less than 39.99	13	14.44%	23.33%	35
40 but less than 49.99	26	28.89%	52.22%	45
50 but less than 59.99	13	14.44%	66.67%	55
60 but less than 69.99	10	11.11%	77.78%	65
70 but less than 79.99	14	15.56%	93.33%	75
80 but less than 89.99	5	5.56%	98.89%	85
90 but less than 99.99	1	1.11%	100.00%	95



602 Solutions to Managing Ashland MultiComm Services Case



Help desk calls appear to be skewed to the right. 28.89% of the days have between 40 and 50 calls while 57.77% of the days have between 30 and 60 calls. Only about 9% of the days had fewer than 30 calls. Fewer than 7% of the days had 80 or more calls.

Chapter 3

1. For the daily number of help desk calls, the following descriptive statistics are computed:

Excel Results

<i>Calls</i>	
Mean	51.8444
Standard Error	1.8372
Median	49
Mode	45
Standard Deviation	17.4291
Sample Variance	303.7733
Kurtosis	-0.4121
Skewness	0.1790
Range	83
Minimum	11
Maximum	94
Sum	4666
Count	90

Refer to Chapter 2 for the stem-and-leaf display and the box plot.

2. Another useful graphical display is the histogram. Refer to Chapter 2 for the histogram.
3. The data are slightly right skewed due to the large number of calls on certain days (The mean is slightly greater than the median). Twenty-five percent of the calls/day are below 40, 50% are less than 49 while 75% are less than 67.

The data have a range of 83 calls/day, with the middle fifty percent of the days varying by 27 calls. The standard deviation is about 33.62% of the mean. You can see the right skewness from the boxplot, which shows a larger distance from the median to the third quartile as compared to the smaller distance from the first quartile to the median.

Chapter 5

1. Assume that the assumptions for using a binomial distribution are satisfied and let π denote the probability that a customer will subscribe to the *3-For-All* service,

(a) $\pi = 0.02, P(X < 3) = 0.921572$

(b) $\pi = 0.02, P(X = 0) + P(X = 1) = 0.36417 + 0.371602 = 0.735772$

(c) $\pi = 0.02, P(X > 4) = 1 - P(X \leq 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)] = 1 - (0.36417 + 0.371602 + 0.1858 + 0.06067 + 0.014548) = 1 - 0.99679 = 0.00321$

(d) $P(X \geq 4) = 1 - P(X < 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)] = 1 - (0.36417 + 0.371602 + 0.1858 + 0.06067) = 1 - 0.982242 = 0.017758$

The likelihood that you would get 4 or more subscribers in a sample of 50 if the probability of a subscription is 0.02 is only 0.017758. Thus, you can conclude that it is more likely than 0.02 that you will get new subscribers when no free premium channels are included.

2. (a) $\pi = 0.06, P(X < 3) = 0.416246$

(b) $\pi = 0.06, P(X = 0) + P(X = 1) = 0.045331 + 0.144673 = 0.190003$

(c) $\pi = 0.06, P(X > 4) = 1 - P(X \leq 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)] = 1 - (0.045331 + 0.144673 + 0.226243 + 0.231067 + 0.173293) = 1 - 0.820596 = 0.1794$

(d) The likelihood that you would get 4 or more subscribers in a sample of 50 if the probability of a subscription is 0.06 is 0.1794. This is much higher than when no free premium channels are offered and the probability of a new subscription is 0.02.

(e) If no premium channels were offered, the probability of getting six or more subscriptions is very small (0.000478) assuming that the probability of a new subscription is 0.02. Thus, you can conclude that it is more likely than 0.02 that you will get new subscribers when no free premium channels are included.

(f) Offering free premium channels increases the likelihood of subscriptions assuming that the estimate of the probability of new subscriptions is accurate.

3. 1 free premium channel, 5 subscriptions: $\pi = 0.04$, $P(X \geq 5) = 0.048971$
- 3 free channels, 6 subscriptions: $\pi = 0.07$, $P(X \geq 6) = 0.135046$
- 4 free channels, 6 subscriptions: $\pi = 0.08$, $P(X \geq 6) = 0.208126$
- 5 free channels, 7 subscriptions: $\pi = 0.085$, $P(X \geq 7) = 0.129051$

One of the questions relates to the accuracy of the estimates of the probability of getting new subscriptions. The ultimate question is to determine the number of premium channels to offer free. On one hand you want to maximize the chance of a new subscription, on the other had, you don't want to give away more premium channels than necessary.

There is no single answer here, but a lot can be discussed. The likelihood of getting five or more new subscriptions with one free premium channel is about 0.05 if the probability of a new subscription is 0.04. The other results are not that unexpected given the probability of new subscriptions in the past. The number of free premium channels offered did not greatly affect the results since the number of new subscriptions ranged from 5 to 7. If, in fact, without any free premium channels, there were either four or six new subscriptions, it would lead you to believe that the probability of a new subscription with that offer might be greater than 0.02. A reasonable conclusion might be to offer one premium channel as an incentive so that you can advertise that there is something being given away for free. Additional free premium channels may not produce sufficiently more new subscriptions.

Chapter 6

1 $n = 1$ $\mu = 1.005$ $\sigma = 0.10$

(a) $P(< 1) = 0.4801$

(b) $P(\text{between } 0.95 \text{ and } 1.0) = 0.1889$

(c) $P(\text{ between } 1.0 \text{ and } 1.05) = 0.1936$

(d) $P(< 0.95 \text{ or } > 1.05) = 0.6175$

- 2 The probability that the upload speed is less than 1.0 if the mean upload speed is 1.05 is 0.3085. The probability that the upload speed is less than 1.0 if the standard deviation is reduced to 0.075 is 0.4734. It will be better to focus on increasing the mean. Once that was done, reducing the standard deviation could result in additional increases in the probability that the upload speed is less than 1.0.

Chapter 7

1 $n = 25$ $\mu = 1.005$ $\sigma = 0.10$

(a) $P(< 1) = 0.4013$

(b) $P(\text{between } 0.95 \text{ and } 1.0) = 0.3983$

(c) $P(\text{ between } 1.0 \text{ and } 1.05) = 0.5865$

(d) $P(< 0.95 \text{ or } > 1.05) = 1 - 0.9848 = 0.0152$

(e) $P(\bar{X} < 0.952) = P(Z < -2.65) = 0.004$

If the upload speed is normally distributed with an mean of 1.005 and a standard deviation of 0.10, the probability is 0.004 of obtaining a sample that will yield a sample mean upload speed of 0.952 or less, a rather unlikely event. The fact that today's sample of 25 yields a sample mean upload speed of 0.952 indicates that the distribution of the upload speed is most likely not normally distributed with a mean of 1.005 and a standard deviation of 0.10.

- 2 The results here are based on a sample of 25. The standard error of the mean is 0.02 and thus there are more means than individual upload speeds close to the population mean.

Chapter 8

95% confidence intervals constructed:

$n = 418$

Question 1: Yes $p = 0.1986$; 0.1603 to 0.2368

Question 2: Yes $p = 0.6268$; 0.5804 to 0.6732

Question 3: Basic $p = 0.3923$; 0.3455 to 0.4392

Enhanced $p = 0.6077$; 0.5609 to 0.6545

$n = 254$

Question 4: Every day $p = 0.1969$; 0.1480 to 0.2457

Question 4: Most days $p = 0.5669$; 0.5060 to 0.6279

Question 4: Occasionally or never $p = 0.2362$; 0.1840 to 0.2885

$n = 418$

Question 5: Almost every day $p = 0.0335$; 0.0162 to 0.0507

Question 5: Several times a week $p = 0.0837$; 0.0572 to 0.1103

Question 5: Almost never (rarely) $p = 0.7488$; 0.7072 to 0.7904

Question 5: Never $p = 0.1340$; 0.1013 to 0.1666

Question 6: Toll-free phone number $p = 0.5502$; 0.5025 to 0.5979

Question 6: AMS website $p = 0.2536$; 0.2119 to 0.2953

Question 6: Direct mail $p = 0.11$; 0.08 to 0.14

Question 6: Good Tunes & More $p = 0.0239$; 0.0093 to 0.0386

Question 6: Other $p = 0.0622$; 0.0390 to 0.0854

Question 7: Yes $p = 0.0957$; 0.0675 to 0.1239

Question 9: Yes $p = 0.8469$; 0.8124 to 0.8814

Question 10: Yes $p = 0.0909$; 0.0633 to 0.1185

Question 8: Trial Weekly rate (\$) Willing to Pay

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	1.9667
Sample Mean	20.73
Sample Size	40
Confidence Level	95%

Intermediate Calculations	
Standard Error of the Mean	0.3110
Degrees of Freedom	39
<i>t</i> Value	2.0227
Interval Half Width	0.6290

Confidence Interval	
Interval Lower Limit	20.10
Interval Upper Limit	21.35

A small proportion of homes subscribe to telephone service from Ashland (between 16.03% and 23.68%). A large proportion of homes subscribe to Internet service from Ashland (between 58.04% and 67.32%). fewer than half the homes have Basic cable service (between 34.55% and 43.92%). Of the 56.1% to 65.5% who have Enhanced service, most (76.38%) watch cable stations available only with enhanced service every day or most days. Very few homes watch premium on-demand services requiring an extra fee (11.72%). Most homes obtained their current service from a toll-free phone number or the AMS website (80.38%). A small percentage of homes (between 6.75% and 12.39%) would consider the *3-For-All* service for a trial period. An overwhelming proportion of homes use another phone service provider (between 81.24% and 88.14%). A small proportion of homes (between 6.33% and 11.85%) would sign up for the *3-For-All* service if offered the Ashland Gold card. For those willing to sign up for the *3-For-All* service for a trial period, they would be willing to pay between \$20.10 and \$21.35.

Chapter 9

Since the population standard deviation is unknown and the sample size is large enough at 50, the t test for the mean can be used.

$$H_0 : \mu \geq 0.97 \quad H_1 : \mu < 0.97$$

PHStat output:

t Test for Hypothesis of the Mean	
Data	
Null Hypothesis $\mu =$	0.97
Level of Significance	0.05
Sample Size	50
Sample Mean	0.95872
Sample Standard Deviation	0.15596782
Intermediate Calculations	
Standard Error of the Mean	0.022057181
Degrees of Freedom	49
t Test Statistic	-0.51139809
Lower-Tail Test	
Lower Critical Value	-1.67655116
p -Value	0.30568461
Do not reject the null hypothesis	

Since $t_{STAT} = -0.511 < t_{CRITICAL} = -1.676$ or the p -value = 0.31 > 0.05, do not reject the null hypothesis. There is insufficient evidence to conclude that the mean upload speed is less than 0.97.

Chapter 10

- 1 You need to test whether the variances are equal since in order to conduct the t test for the difference between two independent means, you first need to determine whether the population variances of the two groups are equal.

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

PHStat output:

F Test for Differences in Two Variances

Data	
Level of Significance	0.05
Larger-Variance Sample	
Sample Size	15
Sample Variance	0.1272
Smaller-Variance Sample	
Sample Size	15
Sample Variance	0.0890

Intermediate Calculations	
F Test Statistic	1.4302
Population 1 Sample Degrees of Freedom	14
Population 2 Sample Degrees of Freedom	14

Two-Tail Test	
Upper Critical Value	2.9786
p-Value	0.5119
Do not reject the null hypothesis	

F-Test (Normal Distribution)

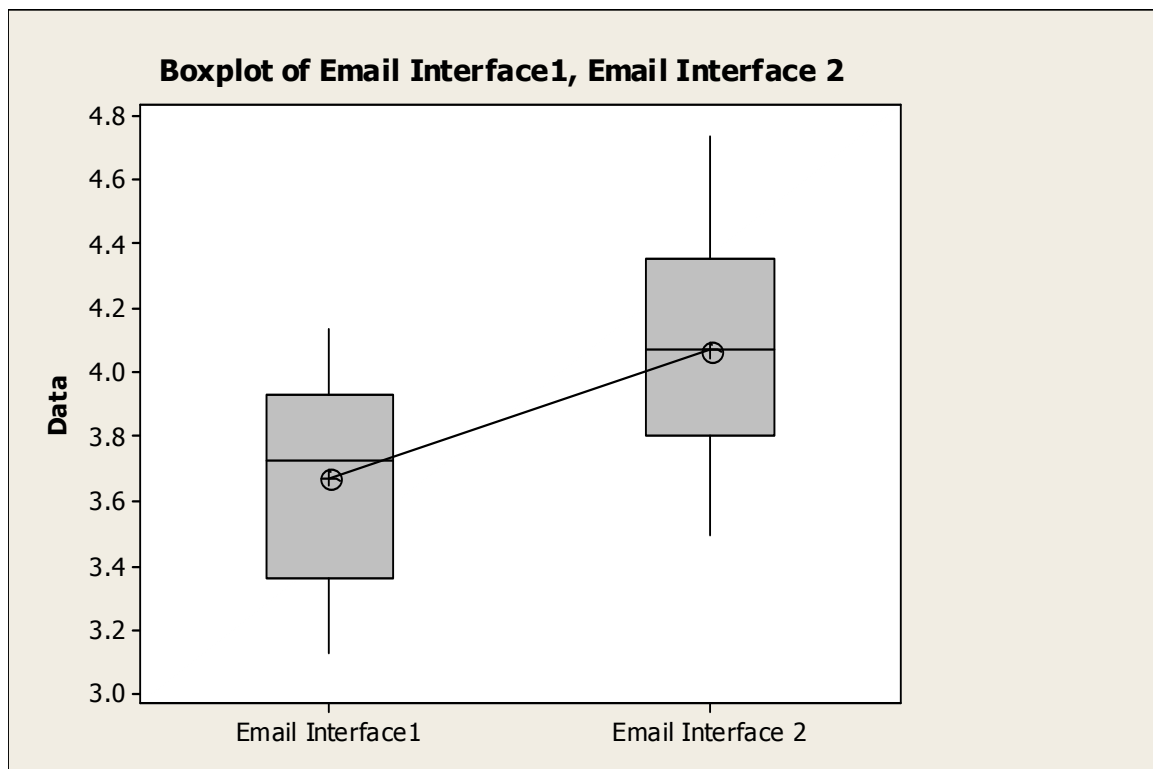
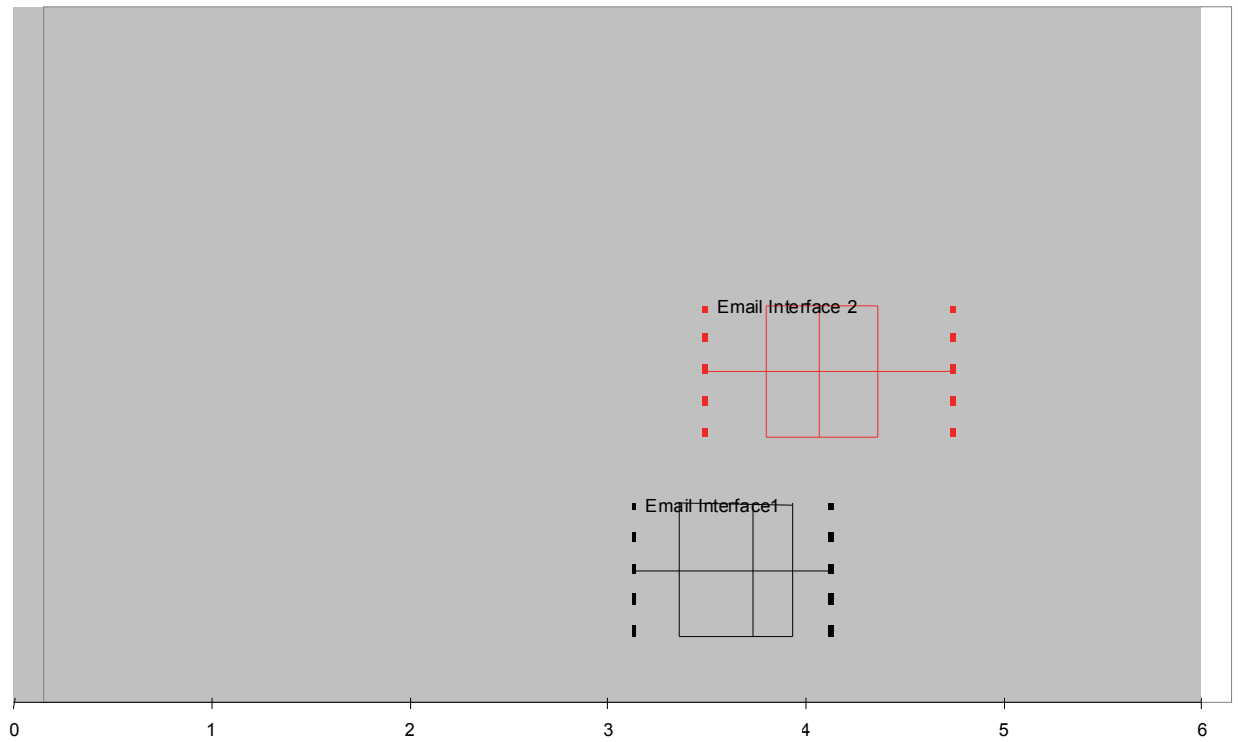
Test statistic = 1.4302, p -value = 0.512

Since $F_{STAT} = 1.4302 < 2.9786$, you do not reject H_0 .

There is insufficient evidence of a difference in the variation of the download (update) times between interface one and interface two.

612 Solutions to Managing Ashland MultiComm Services Case

Download time of Email Interfaces



The boxplots do not show serious departure from the normality assumption.

Because the sample sizes are each 15, you assume that the two populations are normally distributed with roughly equal variances. Since the boxplots do not show serious departure from normality and the F test shows insufficient evidence of a difference in the variances between the two interfaces, you use the pooled-variance t test for the difference in means of the two independent samples.

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$$

PHStat output:

Pooled-Variance t Test for the Difference Between Two Means

(assumes equal population variances)

Data	
Hypothesized Difference	0
Level of Significance	0.05
Population 1 Sample	
Sample Size	15
Sample Mean	3.674
Sample Standard Deviation	0.298275999
Population 2 Sample	
Sample Size	15
Sample Mean	4.072
Sample Standard Deviation	0.356715172

Intermediate Calculations	
Population 1 Sample Degrees of Freedom	14
Population 2 Sample Degrees of Freedom	14
Total Degrees of Freedom	28
Pooled Variance	0.108107143
Standard Error	0.1201
Difference in Sample Means	-0.398
t Test Statistic	-3.3150

Two-Tail Test	
Lower Critical Value	-2.0484
Upper Critical Value	2.0484
p-Value	0.0025
Reject the null hypothesis	

Since $t_{STAT} = -3.315 < -2.0481$ or the p -value of $0.0025 < 0.05$, you reject the null hypothesis at the 5% level of significance. There is enough evidence to conclude that the two population mean download (update) times are different. A t test statistic of -3.315 indicates that the mean download (update) time is lower for Email interface 1 than for Email interface 2.

614 Solutions to Managing Ashland MultiComm Services Case

- 2 Since the two email interfaces were used for the same 15 subscribers, the appropriate test to use is the paired-sample t test. Because of the small sample size of 15, you assume that both populations are normally distributed.

$$H_0 : \mu_D = 0 \qquad H_1 : \mu_D \neq 0$$

PHStat output:

Paired t Test

Data	
Hypothesized Mean Diff.	0
Level of significance	0.05

Intermediate Calculations	
Sample Size	15
DBar	-0.3980
degrees of freedom	14
S_D	0.4286
Standard Error	0.1107
t Test Statistic	-3.5963

Two-Tail Test	
Lower Critical Value	-2.1448
Upper Critical Value	2.1448
p-Value	0.0029
Reject the null hypothesis	

Since $t_{STAT} = -3.5963 < -2.1448$ or a p -value of $0.0029 < 0.05$, you reject the null hypothesis at the 5% level of significance. There is enough evidence to conclude that the mean difference between the two email interfaces is not zero.

- 3 First you need to determine whether there is a difference in the variance of the three email systems.

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2$$

PHStat output:

Levene Test on Email Systems

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Email1	8	26.7	3.3375	5.9655
Email2	8	19.4	2.425	7.9507
Email3	8	26.2	3.275	5.1079

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	4.1575	2	2.0788	0.3278	0.7241	3.4668
Within Groups	133.1688	21	6.3414			
Total	137.3263	23				
<i>Level of significance</i>						0.05

Since $F_{STAT} = 0.3278 < 3.4668$ or $p\text{-value} = 0.7241 > 0.05$, you do not reject H_0 . There is insufficient evidence of a difference in the variation of the update times between the email systems.

Now that you can assume that the email systems do not differ in their variances, you can test to determine whether there is a difference in their update times.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_1: \text{At least one of the means differs}$$

PHStat output:

**ANOVA: Single
Factor
SUMMARY**

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Email1	8	332	41.4625	18.6741
Email2	8	288	36	14.6000
Email3	8	284	35.55	17.3543

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	173.3308	2	86.6654	5.1354	0.0153	3.4668
Within Groups	354.3988	21	16.8761			
Total	527.7296	23				
<i>Level of significance</i>						0.05

Since $F_{STAT} = 5.1354 > 3.4668$ or $p\text{-value} = 0.0153 < 0.05$, you reject H_0 . There is evidence of a difference in the mean update times between the email systems.

Now, you can use the Tukey-Kramer multiple comparisons to determine which email systems differ in their mean update times.

PHstat output:

Tukey Kramer Multiple Comparisons

Group	Sample Mean	Sample Size
1: Email1	41.4625	8
2: Email2	36	8
3: Email3	35.55	8

Other Data	
Level of significance	0.05
Numerator d.f.	3
Denominator d.f.	21
MSW	16.87613
Q Statistic	3.57

Comparison	Absolute Difference	Std. Error of Difference	Critical Range	Results
Group 1 to Group 2	5.4625	1.452417422	5.1851	Means are different
Group 1 to Group 3	5.9125	1.452417422	5.1851	Means are different
Group 2 to Group 3	0.45	1.452417422	5.1851	Means are not different

Email system 1 has a higher update time than Email systems 2 and 3.

Chapter 11

1 PHStat output:

Observed Frequencies					
	Column variable				
Continue subscription after trial period	No discount	Moderate	Substantial	Rest.Card	Total
Yes	24	30	38	51	143
No	76	70	62	49	257
Total	100	100	100	100	400

Expected Frequencies					
	Column variable				
Row variable	No discount	Moderate	Substantial	Rest.Card	Total
Yes	35.75	35.75	35.75	35.75	143
No	64.25	64.25	64.25	64.25	257
Total	100	100	100	100	400

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	4
Degrees of Freedom	3

Results	
Critical Value	7.8147
Chi-Square Test Statistic	17.7954
p-Value	0.00048471
Reject the null hypothesis	

**Expected frequency assumption
is met.**

$\chi^2 = 17.7954 > 7.8147$. Reject H_0 . There is evidence of a difference in the proportion of subscribers who continue subscriptions among the four trial plans.

2 PHStat output:

Observed Frequencies			
	Internet		
Phone	Yes	No	Total
yes	55	28	83
No	207	128	335
Total	262	156	418

Calculations

fo-fe	
2.9761	-2.9761
-2.9761	2.9761

Expected Frequencies			
	Internet		
Phone	Yes	No	Total
yes	52.0239	30.9761	83
No	209.9761	125.0239	335
Total	262	156	418

(fo-fe) ² /fe	
0.1702	0.2859
0.0422	0.0708

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	3.8415
Chi-Square Test Statistic	0.5692
p-Value	0.4506
Do not reject the null hypothesis	

Expected frequency assumption

$\chi^2 = 0.5692 < 3.8415$. Do not reject H_0 . There is no evidence of a relationship between having AMS telephone service and AMS Internet service.

PHStat output:

Observed Frequencies			
	Discount		
Type	Yes	No	Total
Basic	8	156	164
Enhanced	32	222	254
Total	40	378	418

Calculations

fo-fe	
-7.6938	7.6938
7.6938	-7.6938

Expected Frequencies			
	Discount		
Type	Yes	No	Total
Basic	15.6938	148.3062	164
Enhanced	24.3062	229.6938	254
Total	40	378	418

(fo-fe) ² /fe	
3.7718	0.3991
2.4354	0.2577

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	2
Degrees of Freedom	1

Results	
Critical Value	3.8415
Chi-Square Test Statistic	6.8640
p-Value	0.0088
Reject the null hypothesis	

**Expected frequency assumption
is met.**

$\chi^2 = 6.864 > 3.8415$. Reject H_0 . There is evidence of a difference in willingness to sign up for a discount trial between those who have Basic service and those who have Enhanced service.

PHStat output:

	Watches Premium or On-Demand Services				
Type of cable Service	Almost Every day	Several Times a Week	Almost Never	Never	Total
Basic	2	5	127	30	164
Enhanced	12	30	186	26	254
Total	14	35	313	56	418

Expected Frequencies					
	Watches Premium or On-Demand Services				
Type of cable Service	Almost Every day	Several Times a Week	Almost Never	Never	Total
Basic	5.4928	13.7321	122.8038	21.9713	164
Enhanced	8.5072	21.2679	190.1962	34.0287	254
Total	14	35	313	56	418

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	4
Degrees of Freedom	3

Results	
Critical Value	7.8147
Chi-Square Test Statistic	17.8570
p-Value	0.00047076
Reject the null hypothesis	

Expected frequency assumption is met.

$\chi^2 = 17.857 > 7.8147$. Reject H_0 . There is evidence of a difference in how often premium or on demand services are watched based on type of service .

PHStat output:

	Watches Premium or On-Demand Services				
Discount	Almost Every day	Several Times a Week	Almost Never	Never	Total
Yes	4	5	27	4	40
No	10	30	286	52	378
Total	14	35	313	56	418

Expected Frequencies					
	Watches Premium or On-Demand Services				
Discount	Almost Every day	Several Times a Week	Almost Never	Never	Total
Yes	1.3397	3.3493	29.9522	5.3589	40
No	12.6603	31.6507	283.0478	50.6411	378
Total	14	35	313	56	418

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	4
Degrees of Freedom	3

Results	
Critical Value	7.8147
Chi-Square Test Statistic	7.4440
p-Value	0.0590
Do not reject the null hypothesis	

**Expected frequency assumption
is met.**

$\chi^2 = 7.444 < 7.8147$. Do not reject H_0 . There is insufficient evidence of a difference in willingness to sign up for a discount trial based on how often premium or on demand services are watched.

PHStat output:

Observed Frequencies						
	Method for Current Subscription					
Discount	Toll-free phone	AMS website	Direct mail reply card	Good Tunes & More	Other	Total
Yes	11	21	5	1	2	40
No	219	85	41	9	24	378
Total	230	106	46	10	26	418

Expected Frequencies						
	Method for Current Subscription					
Discount	Toll-free phone	AMS website	Direct mail reply card	Good Tunes & More	Other	Total
Yes	22.0096	10.1435	4.4019	0.9569	2.4880	40
No	207.9904	95.8565	41.5981	9.0431	23.5120	378
Total	230	106	46	10	26	418

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	5
Degrees of Freedom	4

Results	
Critical Value	9.4877
Chi-Square Test Statistic	19.1369
p-Value	0.0007
Reject the null hypothesis	

Expected frequency assumption is violated.

$\chi^2 = 19.1369 > 9.4877$. Reject H_0 . There is evidence of a difference in willingness to sign up for a discount trial based on the method of current subscription.

PHStat output:

Observed Frequencies						
	Method for Current Subscription					
Gold Card	Toll-free phone	AMS website	Direct mail reply card	Good Tunes & More	Other	Total
Yes	10	20	5	1	2	38
No	220	86	41	9	24	380
Total	230	106	46	10	26	418

Expected Frequencies						
	Method for Current Subscription					
Gold Card	Toll-free phone	AMS website	Direct mail reply card	Good Tunes & More	Other	Total
Yes	20.9091	9.6364	4.1818	0.9091	2.3636	38
No	209.0909	96.3636	41.8182	9.0909	23.6364	380
Total	230	106	46	10	26	418

Data	
Level of Significance	0.05
Number of Rows	2
Number of Columns	5
Degrees of Freedom	4

Results	
Critical Value	9.4877
Chi-Square Test Statistic	18.7689
p-Value	0.0009
Reject the null hypothesis	

Expected frequency assumption is violated.

$\chi^2 = 18.7689 > 9.4877$. Reject H_0 . There is evidence of a difference in willingness to sign up for a gold card based on the method of current subscription.

Chapter 12

- 1 The method was placing too much emphasis on the last three time periods instead of considering a large amount of data that was available. This overemphasis would tend to lead to large fluctuations in the forecast accuracy when the trend differed from what had occurred in the last three months.
- 2 There are many factors that might be considered other than the number of telemarketing hours. Among them are:
 1. The amount of newspaper and radio and television advertising for new subscriptions.
 2. The email advertising for new subscriptions.
 3. Whether or not special promotional campaigns were being used in a particular month.
 4. The average experience of the telemarketers used in a given month.
 5. The training programs provided to the telemarketers in a given month.
- 3 (a) Excel output:

<i>Regression Statistics</i>	
Multiple R	0.9283
R Square	0.8617
Adjusted R Square	0.8554
Standard Error	409.7766
Observations	24

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	23022485.1539	23022485.1539	137.1065	0.0000
Residual	22	3694170.8044	167916.8547		
Total	23	26716655.9583			

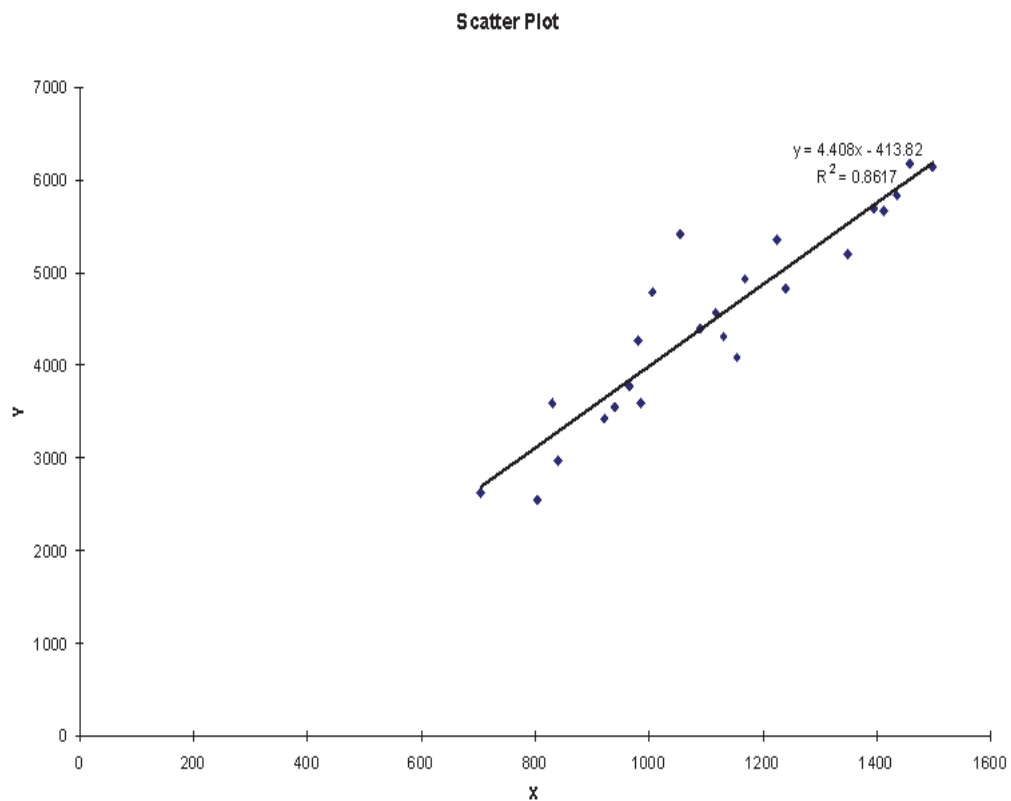
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-413.8204	427.0878	-0.9689	0.3431	-1299.5473	471.9065
Hours	4.4080	0.3765	11.7092	0.0000	3.6272	5.1887

626 Solutions to Managing Ashland MultiComm Services Case

Durbin-Watson Calculations

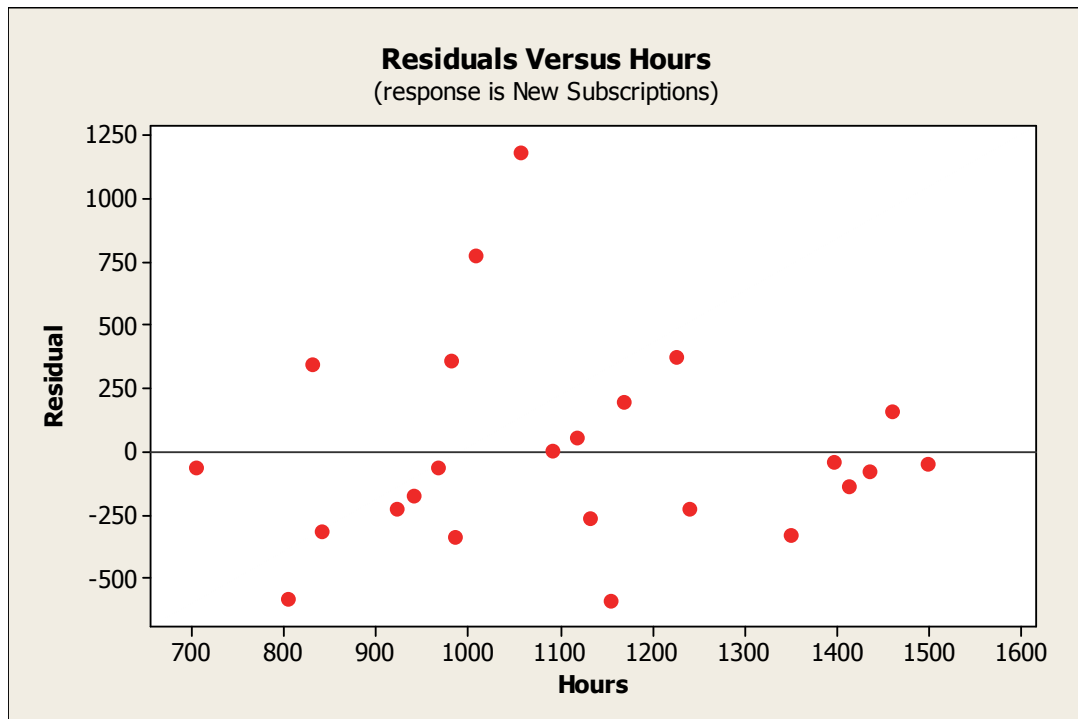
Sum of Squared Difference of Residuals 6283406.9317
Sum of Squared Residuals 3694170.8044

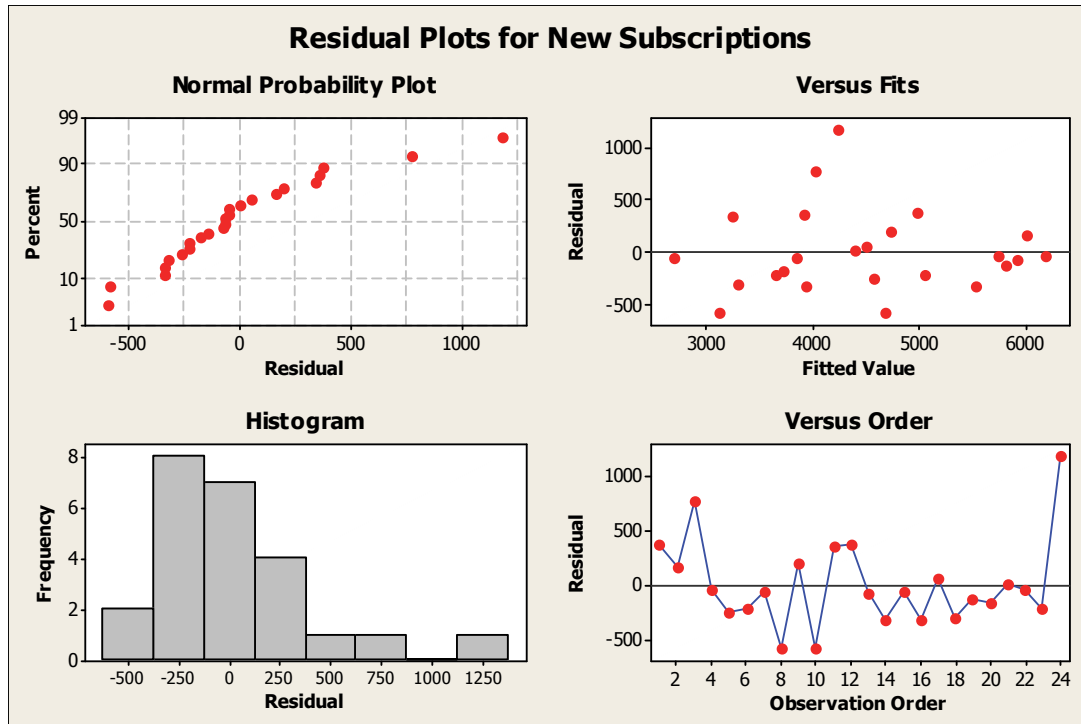
Durbin-Watson Statistic	1.7009
--------------------------------	---------------



Confidence Interval Estimate (b)

Data	
X Value	1200
Confidence Level	95%
Intermediate Calculations	
Sample Size	24
Degrees of Freedom	22
t Value	2.073875
XBar, Sample Mean of X	1112.542
Sum of Squared Differences from XBar	1184892
Standard Error of the Estimate	409.7766
h Statistic	0.048122
Predicted Y (YHat)	4875.72
For Average Y	
Interval Half Width	186.4241
Confidence Interval Lower Limit	4689.296
Confidence Interval Upper Limit	5062.144
For Individual Response Y	
Interval Half Width	870.033
Prediction Interval Lower Limit	4005.687
Prediction Interval Upper Limit	5745.753





The regression model initially fit to the data is a simple linear model that predicts

$$\text{New Subscriptions} = -413.8204 + 4.4080 \text{ Telemarketing Hours}$$

The Y intercept (b_0), equal to -413.8204 , has no direct interpretation since sales of below zero new subscriptions is not feasible.

The slope (b_1), equal to 4.4080 , can be interpreted to mean that for each increase of one telemarketing hour, the mean new subscriptions are predicted to increase by 4.4080 per month.

The r^2 value of 0.8617 can be interpreted to mean that 86.17% of the variation in new subscriptions can be explained by variation in the number of telemarketing hours from month to month.

Since the data were collected for 24 consecutive months, you need to determine whether there is any autocorrelation among the residuals. The Durbin-Watson D statistic of 1.7009 is > 1.45 , the upper critical value for $n = 24$ and $\alpha = 0.05$. Thus, there is no evidence of any positive autocorrelation among the residuals. You could also examine the plot of the residuals over time to see whether a pattern existed. In this case, the plot does not seem to indicate any evidence of positive association among consecutive residuals.

Before making any predictions based on the model, you need to determine whether the simple linear model was appropriate for these data. The plot of the residuals versus the fitted Y values (or the X values) indicates no evidence of any pattern. Thus, we may conclude that the simple linear model was appropriate for these data.

You can examine the validity of the assumption of normality among the residuals by studying the normal probability plot. If the points in this plot fall on an approximate

straight line, there would be no reason to suspect serious departure from normality. Although there is some skewness in the normal probability plot and the histogram, the normality assumption does not appear to have been seriously violated.

- (b) For $X = 1,200$, the predicted value of

$$\text{New Subscriptions} = -413.82 + 4.40795(1200) = 4,875.72$$

Confidence interval estimate is 4,689.296 to 5,062.144. The prediction interval is 4,005.687 to 5,745.753

- (c) The value of 2,000 for X is beyond the range of our X values. Thus, you would be assuming that the regression model fit for a range of 704 to 1,498 telemarketing hours would be valid for a month in which the telemarketing hours was 2,000. This represents a situation which is well beyond the range of the number of telemarketing hours that has been used in the past 24 months. Many things could change with this increase of telemarketing hours, and the type of increase in new subscriptions with expanding telemarketing hours may not continue when you further expand by another 33.5% above the maximum amount used in the past.

Chapter 13

Let Y = number of new subscriptions, X_1 = number of hours spent on telemarketing,

$X_2 = 1$ if presentation type is personal formal; 0 if presentation type is personal informal, $X_3 = X_1X_2$

Excel Output:

Regression Statistics	
Multiple R	0.9608
R Square	0.9231
Adjusted R Square	0.9115
Standard Error	75.8181
Observations	24

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	1379445.1974	459815.0658	79.9903	0.0000
Residual	20	114967.7609	5748.3880		
Total	23	1494412.9583			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-341.0273	132.1268	-2.5811	0.0178	-616.6388	-65.4158
Hours	5.0972	0.4727	10.7842	0.0000	4.1113	6.0832
Presentation	206.2773	169.3167	1.2183	0.2373	-146.9109	559.4655
Interaction	-0.4210	0.6358	-0.6621	0.5155	-1.7473	0.9054

Testing the significance of the interaction:

$$H_0 : \beta_3 = 0 \quad \text{vs.} \quad H_1 : \beta_3 \neq 0$$

Since the p -value of the t -test statistic for the significance of X_3 is $0.5155 > 0.05$, do not reject the null hypothesis. There is insufficient evidence of an interaction between number of hours spent on telemarketing and the presentation type.

The Excel results of the multiple regression without the interaction term are as follows:

Regression Statistics	
Multiple R	0.9599
R Square	0.9214
Adjusted R Square	0.9139
Standard Error	74.7973
Observations	24

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	1376925.4887	688462.7444	123.0575	0.0000
Residual	21	117487.4696	5594.6414		
Total	23	1494412.9583			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-277.0791	88.9423	-3.1153	0.0052	-462.0448	-92.1134
Hours	4.8646	0.3119	15.5970	0.0000	4.2160	5.5132
Presentation	96.3074	32.4011	2.9724	0.0073	28.9257	163.6891

Durbin-Watson Calculations

Sum of Squared Difference of Residuals	191570.393
Sum of Squared Residuals	117487.4696

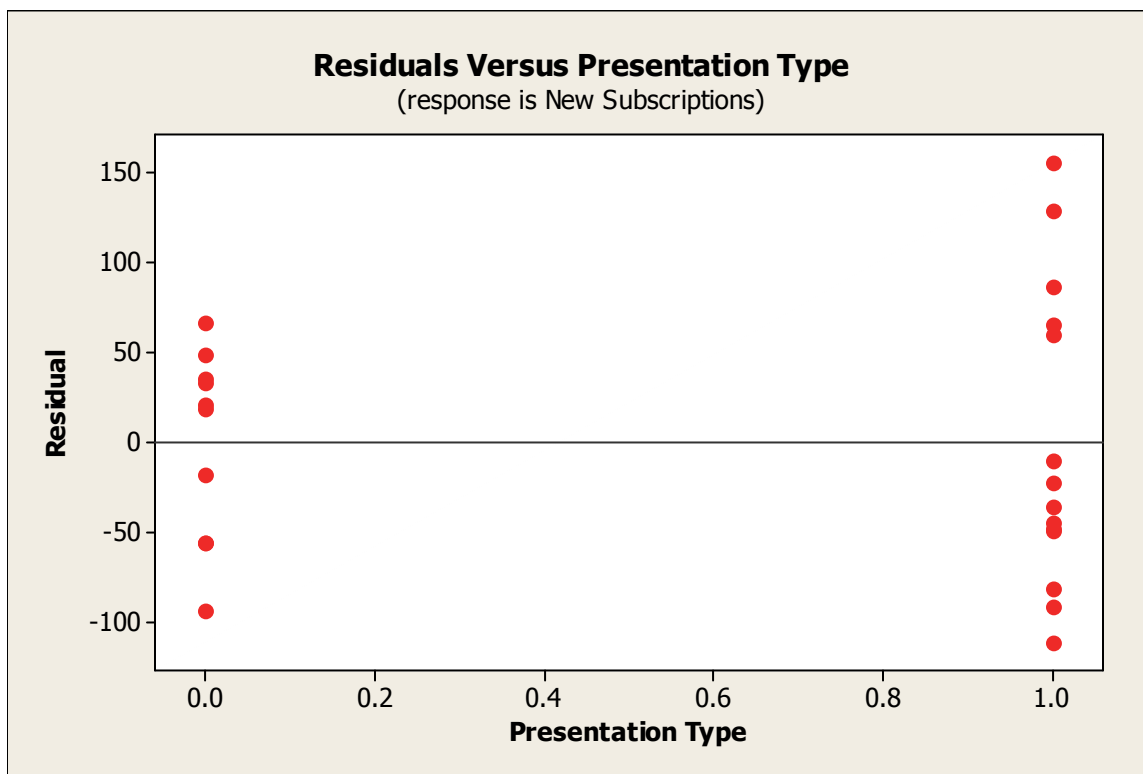
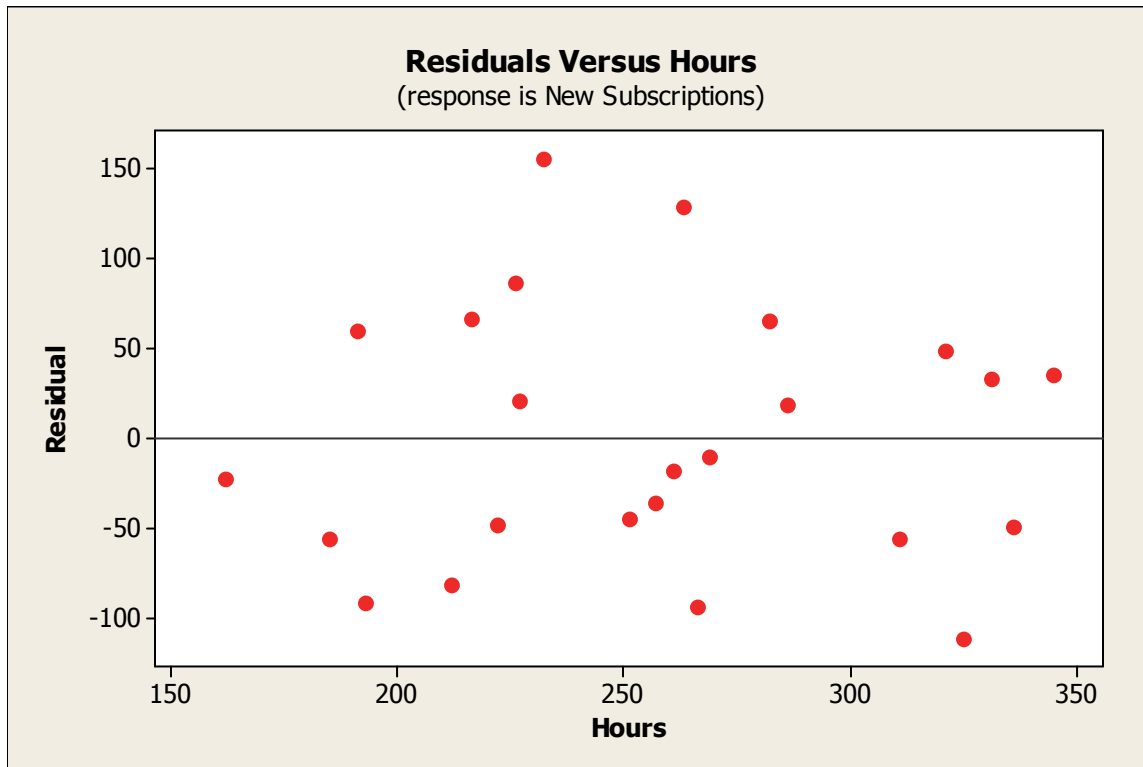
Durbin-Watson Statistic	1.630560209
--------------------------------	--------------------

Regression Analysis

Coefficients of Partial Determination

Intermediate Calculations			
SSR(X1,X2)	1376925.489		
SST	1494412.958		
SSR(X2)	15938.42976	SSR(X1 X2)	1360987.059
SSR(X1)	1327497.441	SSR(X2 X1)	49428.04753

Coefficients	
r² Y1.2	0.920534668
r² Y2.1	0.296126138



The 5% critical values of the Durbin-Watson statistics are $d_L = 1.19$ and $d_U = 1.55$. The Durbin-Watson test statistic = 1.63 > 1.55. There is no evidence of autocorrelation in the data.

Testing for the overall significance of the multiple regression:

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{vs.} \quad H_1 : \text{not all } \beta_j = 0.$$

Since the p -value of the overall F test statistic is essentially zero, reject the null hypothesis and conclude that there is evidence that the number of new subscriptions depend on the number of hours spent on telemarketing and/or the sales presentation type.

Testing for the effect of the individual independent variable on the number of new subscriptions:

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0$$

Since the p -value of the t -test statistic for the significance of X_1 is essentially zero, reject the null hypothesis and conclude that there is evidence that the number of hours spent on telemarketing alone has significant effect on the number of new subscriptions.

$$H_0 : \beta_2 = 0 \quad \text{vs.} \quad H_1 : \beta_2 \neq 0$$

Since the p -value of the t -test statistic for the significance of X_2 is $0.007 < 0.05$, reject the null hypothesis and conclude that there is evidence that the sales presentation type alone has significant effect on the number of new subscriptions.

There is no pattern in the residuals versus hours or presentation type. The best model to predict the number of new subscriptions is

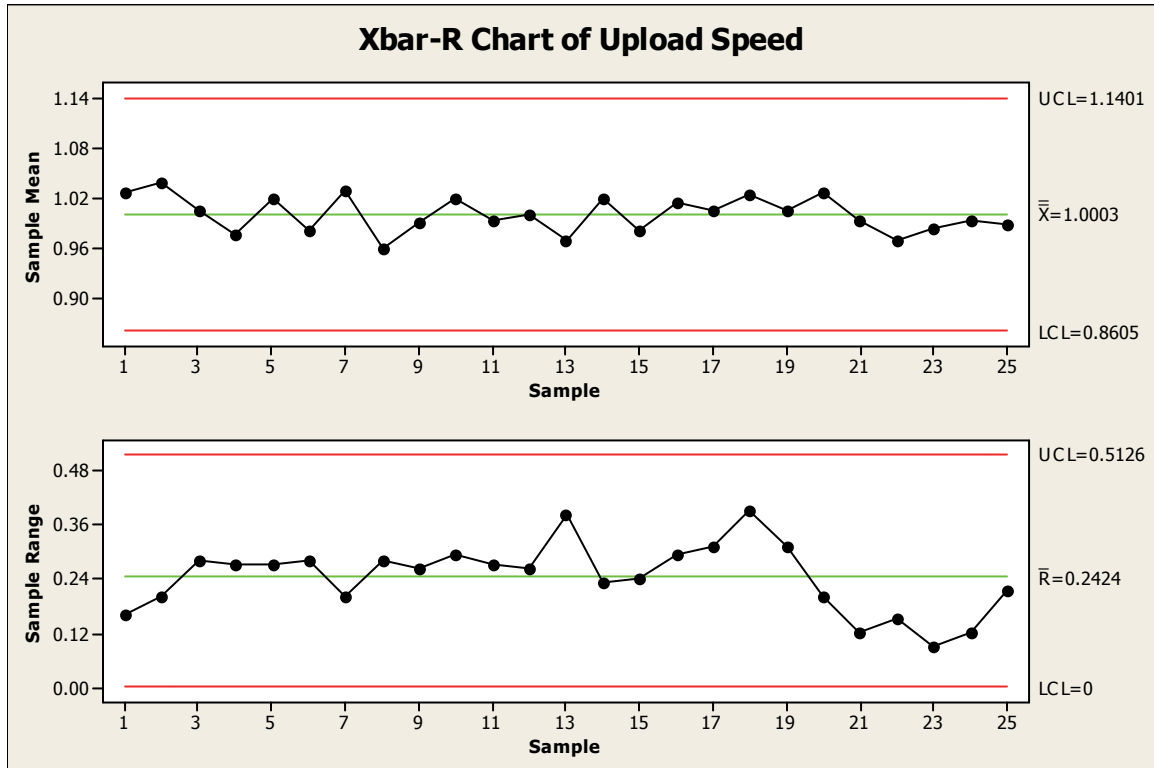
$$\hat{Y} = -277.0791 + 4.8646 X_1 + 96.3074 X_2$$

92.14% of the variation in the number of new subscriptions can be explained by variation in the hours and the presentation type. Holding constant the presentation style, 92.05% of the variation in new subscriptions can be explained by variation in telemarketing hours. Holding constant the number of telemarketing hours, 29.61% of the variation in new subscriptions can be explained by variation in the presentation style.

Since the regression coefficient for presentation type is positive, this means that holding constant the number of telemarketing hours, using a formal presentation style is predicted to increase the number of new subscriptions by a mean of 96.3074. Holding constant the presentation style, for each increase of one telemarketing hour, the mean number of new subscriptions is predicted to increase by 4.8646.

Chapter 14

1 (a)



Since there are five observations for each day, you should use the \bar{X} chart in conjunction with the Range chart.

Mean and Range Charts:

$$\bar{\bar{X}} = 1.0003 \quad \bar{R} = 0.2424$$

Range Chart:

$$\bar{R} = .02424 \quad UCL = 0.5126 \quad LCL \text{ does not exist}$$

- (b) There are no points outside the control limits and no violations of the rules 1 - 5.

\bar{X} Chart:

$$\bar{\bar{X}} = 1.0003 \quad UCL = 1.1401 \quad LCL = 0.8605$$

There are no points outside the control limits and no violations of the rules 1 - 5.

- (c) The process is stable, so any attempt to reduce the common cause variation in the upload speed or to improve the upload speed must be undertaken by management by changing the process.