Instructional Tips and Solutions for Digital Cases

Chapter 2—Organizing and Visualizing Data

Instructional Tips

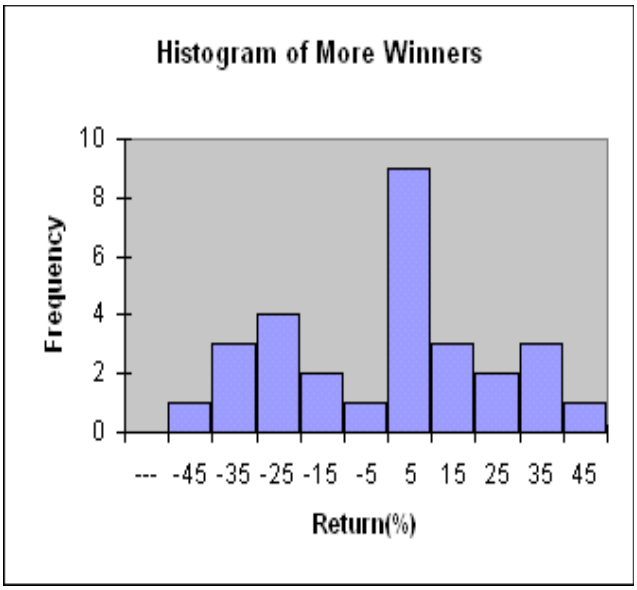
1. Students should develop a frequency distribution of the More Winners data along with at least one graph such as a histogram, polygon, or cumulative percentage polygon.
2. One objective is to have students look beyond the actual statistical results generated to evaluate the claims presented. For the More Winners data, this might include a comparison with tables and charts developed for the entire Mutual Funds data set. Such a comparison would lead to the realization that all eight funds in the "Big Eight" are high-risk funds that may have a great deal of variation in their return.
3. The presentation of information can lead to different perceptions of a business. This can be seen in the aggressive approach taken in the home page.

Solutions

1. Yes. There is a breathless, exaggerated style to the writing and the illustrations are very busy and colorful without conveying much information. There is also a certain aggressiveness in exclamations to "show me the data." Claims are made, but supporting evidence is scant. The style is reminiscent of a misleading infommercial. The graphs on pages 5 and 6 have poor design that obscures their meaning, if any. Also, nowhere in the document does EndRun disclose its principals and the address of its operations, something that a reputable business would surely do. And a testimonial page at the end is more suitable for an infommercial selling a consumer product and not something one would expect to see from a reputable financial services firm.
2.

Frequencies (Return(%))

Bins	*Frequency*	*Percentage*	*Cumulative %*	*Midpts*
-50	0	0.00%	.00%	---
-40	1	3.45%	3.45%	-45
-30	3	10.34%	13.79%	-35
-20	4	13.79%	27.59%	-25
-10	2	6.90%	34.48%	-15
-0.01	1	3.45%	37.93%	-5
9.99	9	31.03%	68.97%	5
20	3	10.34%	79.31%	15
30	2	6.90%	86.21%	25
40	3	10.34%	96.55%	35
50	1	3.45%	100.00%	45



Although the claim is literally true, the data show a wide range of returns for the 29 mutual funds selected by EndRun investors. Although 18 funds had positive returns, 11 had negative returns for the five year period. Of the funds having negative returns, many had large losses, with 27.59% having annualized losses of 20% or more. Many of the positive returns were small, with 31.03% having an annualized return between 0 and 10%. All of this raises questions about the effectiveness of the EndRun investment service.

3. Since mutual funds are rated by risk, it would be important to know the "risk" of the funds EndRun chooses. "High" risk funds, as all eight turn out to be, are not a wise choice for certain types of investors. An in-depth analysis would also see if the eight funds were representative of the performance of that group (no, the eight are among the weakest performers, as it turns out). In addition, examining summary measures (discussed in Chapter 3) would also be helpful in evaluating the "Big Eight" funds.

4. You would hope that one's investment "grew" over time. Whether this is reason to be truly proud would again be based on a comparison to a similar group of funds. You would also like to know such things as whether the gain in value is greater than any inflation that might have occurred during that period. Even more sophisticated reasoning would look at financial planning analysis to see if an investment in the "big eight" was a worthy one or one that showed a real gain after tax considerations. A warning flag, however, is that the business feels the need to state that it is "proud" even as it does not state a comparative (such as "we are proud to have *outperformed* all of the leading national investment services.") Such an emotional claim suggests a lack of rational data that could otherwise be used to make a more persuasive case for using EndRun's service.

Chapter 3—Numerical Descriptive Measures

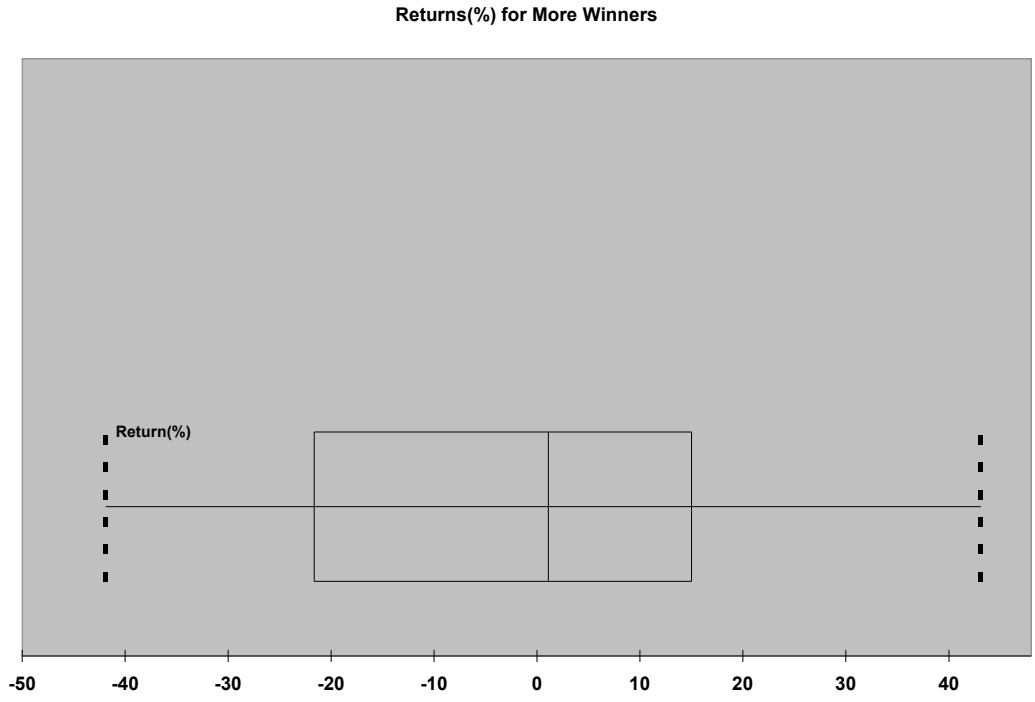
Instructional Tips

1. Students should compute descriptive statistics and develop a boxplot for the More Winners sample. They should compare the measures of central tendency and take note of the measures of variation. The boxplot can be used to evaluate the symmetry of the data.
2. All too often means and standard deviations are computed on data from a scale (usually 5 or 7 points) that is ordinal at best. They should be cautioned that such statistics are of questionable value.

Solutions

1.

Return(%)	
Mean	**-0.61724**
Standard Error	**4.533863**
Median	**1.1**
Mode	**1.1**
Standard Deviation	**24.4156**
Sample Variance	**596.1215**
Range	**85**
Minimum	**-41.9**
Maximum	**43.1**
Sum	**-17.9**
Count	**29**
Largest(1)	**43.1**
Smallest(1)	**-41.9**



For the sample of 29 investors, the average annualized rate of return is -0.62% and the median annualized rate of return is only 1.1%, Thus, half the investors are either losing money or have a very small return. In addition, there is a very large amount of variability with a standard deviation of over 24% in the annualized return. The data appear fairly symmetric since the distance between the minimum return and the median is about the same as the distance between the median and the largest return. However, the first quartile is more distant from the median than is the third quartile.

2. Calculating mean responses for a categorical variable is a naïve error at best. No methodology for collecting this survey is offered. For several questions, the neutral response dominates, surely not an enthusiastic endorsement of EndRun! Strangely, for the question "How satisfied do you expect to be when using EndRun's services in the coming year?" only 19 responses appear, compared with 26 or 27 responses for the other questions (see the next question). Eliminating the means and considering the questions as categorical variables and then developing a bar chart for each question would be more appropriate.
3. As proposed, the question expects that the person being surveyed will be using EndRun. Most likely, the missing responses reflect persons who had already planned not to use EndRun and therefore could not answer the question as posed. Survey questions that would uncover reasons for planning to use or not use would be more insightful.

Chapter 4—Basic Probability

Instructional Tips

The main goal of the Digital case for this chapter is to have students be able to distinguish between what is a simple probability, a joint probability, and a conditional probability.

Solutions

1.

	Return not less than 15%	Return less than 15%
Best 10 Customers	**8**	**2**
Other Customers	**0**	**19**

The claim "four-out-of-five chance of getting annualized rates of return of no less than 15%," is literally accurate, but it applies only to EndRun's best 10 customers. A more accurate probability would consider all customers (8/29, or about 28%). In fact, none of EndRun's other customers achieved a return of not less than 15%. Another issue is that you do not know the actual return rates for each customer, so you cannot calculate any meaningful descriptive statistics.

2.

		Invested at EndRun	
		Yes	No
Made money?	**Yes**	**18**	**94**
	No	**11**	**45**

The 7% probability calculated (11/168 = 6.55%) is actually the joint probability of investing at EndRun and making money. The probability of being an EndRun investor who lost money is the conditional probability of losing money given an investment in EndRun which is equal to 11/29 = 37.93%.

3. Since the patterns of security markets are somewhat unpredictable by their nature, any probabilities based on past performance are not necessarily indicative of future events. Even if EndRun had the "best" probability for "success", that would be no guarantee that their investment strategy would work in tomorrow's market.

Chapter 5—Discrete Probability Distributions

Instructional Tips

This digital case involves computing expected values and standard deviations of probability distributions and then using portfolio risk to obtain a good expected return with a lower risk than what would be involved if an entire investment was made in one fund.

1. Students need to realize that a very good return may occur only under certain circumstances.
2. Students need to realize that how the probabilities of the various events are obtained is of crucial importance to the results.
3. Using PHStat2, students can determine the expected portfolio return and portfolio risk of different combinations of two different funds.

Solutions

1. Yes! “With EndRun's Worried Bear Fund, you can get a four hundred percent rate of return in times of recession!” However, EndRun itself estimates the probability of recession at only 20% in its own calculations. “With EndRun's Happy Bull Fund, you can make twelve times your initial investment (that's a 1,200 percent rate of return!) in a fast expanding, booming economy.” In this case, EndRun itself estimates the probability of a fast expanding economy at only 10%.
2. Estimating the probabilities of the outcomes is very subjective. It is never made clear how the value of the outcomes were determined.
3. There are several factors to consider. Most obviously, if an investor believed in a different set of probabilities, then the Worried Bear fund would not necessarily have the better expected return. An investor more concerned about risk would want to examine other measures (such as the standard deviation of each investment, the expected portfolio return, and the portfolio risk of different combination of investments). Investors who hedge might also invest in a lower expected return fund if the pattern of outcomes is radically different (as it is in the case of the two EndRun funds).

EndRun Portfolio Analysis

Outcomes	P	Happy Bull	Worried Bear
fast expanding economy	0.1	1200	-300
expanding economy	0.2	600	-200
weak economy	0.5	-100	100
recession	0.2	-900	400

Weight Assigned to X	0.5

Statistics	
E(X)	10
E(Y)	60
Variance(X)	382900
Standard Deviation(X)	618.7891
Variance(Y)	50400
Standard Deviation(Y)	224.4994
Covariance(XY)	-137600
Variance(X+Y)	158100
Standard Deviation(X+Y)	397.6179
Portfolio Management	
Weight Assigned to X	0.5
Weight Assigned to Y	0.5
Portfolio Expected Return	35
Portfolio Risk	198.809
Portfolio Management	
Weight Assigned to X	0.3
Weight Assigned to Y	0.7
Portfolio Expected Return	45
Portfolio Risk	36.94591
Portfolio Management	
Weight Assigned to X	0.2
Weight Assigned to Y	0.8
Portfolio Expected Return	50
Portfolio Risk	59.4979
Portfolio Management	
Weight Assigned to X	0.1
Weight Assigned to Y	0.9
Portfolio Expected Return	55
Portfolio Risk	141.0142
Portfolio Management	
Weight Assigned to X	0.7
Weight Assigned to Y	0.3
Portfolio Expected Return	25
Portfolio Risk	366.5583

Portfolio Management	
Weight Assigned to X	0.9
Weight Assigned to Y	0.1
Portfolio Expected Return	15
Portfolio Risk	534.6821

Note that of the two funds, Worried Bear has both a higher expected return and a lower standard deviation. From the results above, it appears that a good approach is to invest more in the Worried Bear fund than the Happy Bull fund to achieve a higher expected portfolio return while minimizing the risk. A reasonable choice is to invest 30% in the Happy Bull fund and 70% in the Worried Bear fund to achieve an expected portfolio return of 45 with a portfolio risk of 36.94. This risk is substantially below the standard deviation of 618 for the Happy Bull fund and 224 for the Worried Bear fund. The expected portfolio return of 45 is much higher than the expected return for investing in only the Happy Bull fund and is somewhat below the expected return for investing completely in the Worried Bear fund. Of course, with the knowledge about EndRun accumulated through Digital cases in Chapters 2 - 5, a reasonable course of action would be *not to invest any money with EndRun*!

Chapter 6—The Normal Distribution and Other Continuous Distributions

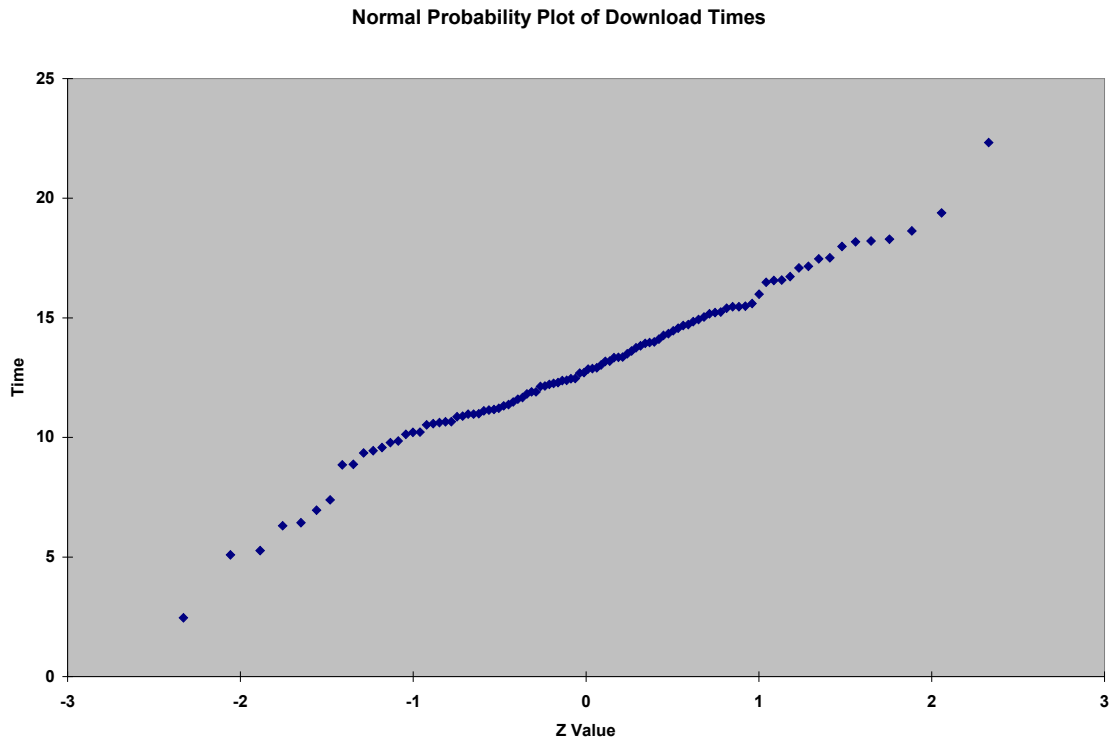
Instructional Tips

This digital case consists of two parts – determining whether the download times are approximately normally distributed and then evaluating the validity of various statements made concerning the download times that relate to understanding the meaning of probabilities from the normal distribution.

Solution

1.

Statistics	
Sample Size	100
Mean	12.8596
Median	12.785
Std. Deviation	3.279278
Minimum	2.46
Maximum	22.33



From the normal probability plot, the data appear to be approximately normally distributed. In addition, the distance from the minimum value to the median is approximately the same as the distance from the median to the maximum value.

2. "• A 15-second download is less likely than a 14 or 13-second download."

This is false since the probability of an exact download time is zero. Statements should be made concerning the likelihood that the download time is less than a specific value. For example, the probability of a download time less than 15 seconds is 0.743 or 74.3%.

Normal Probabilities

Download Time	
Mean	12.8596
Standard Deviation	3.279278

Probability for X <=	
X Value	15
Z Value	0.6527047
P(X<=15)	0.7430267

"• If we can strive to eliminate times greater than 22.7 seconds, then more times will fall within 3 standard deviations."

This is false since eliminating those times will reduce the mean and the standard deviation. There will still be 99.7% of the values within ± 3 standard deviations. All that can be said is the probability of obtaining a download time less than 22.7 seconds will increase.

"• One time out of every 10 times, an individual user will experience a download time that is greater than 17.06 seconds."

The probability of a download time above 17.06 seconds is 10%. However, this does not mean that one of every ten downloads will take more than 17.06 seconds. It means that if the data is normally distributed with $\mu = 12.8596$ seconds and the standard deviation equal to 3.279278 seconds, 10% of all downloads will take more than 17.06 seconds.

"• Since over 99 percent of download times fall within plus or minus 3 standard deviations, our home page download process meets the Six Sigma benchmark for industrial quality. (Recall that senior management held a meeting last month on the importance of the Six Sigma methodology.)"

Note: Six Sigma is discussed in Chapter 17 of the text. This statement is "double talk". In a normal distribution, 99.7% percent of all measurements fall within plus or minus 3 standard deviations. Six Sigma is a managerial approach designed to create processes that results in no more than 3.4 defects per million. The QRT needs to determine the requirements of the customers and then determine the capability of the current process (see Section 17.6) before embarking on quality improvement efforts.

3. If the standard deviation was assumed to be the same as it was previously, the probability of obtaining a download time below a specific number of seconds would increase. For example the probability of having a download time below 15 seconds with a mean of 7.8596 seconds instead of a mean of 12.8596 seconds is 98.53% instead of 74.30%.

Normal Probabilities

Download Time	
Mean	7.8596
Standard Deviation	3.279278

Probability for X <=	
X Value	15
Z Value	2.1774305
P(X<=15)	0.9852758

Chapter 7—Sampling and Sampling Distribution

Instructional Tips

This digital case focuses on two concepts – the need for random sampling and the application of the sampling distribution of the mean.

Solutions

1. "For our investigation, members of our group went to their favorite stores …One member thought her box of Oxford's Pennsylvania Dutch-Style Chocolate Brownie Morning Squares was short, but her son opened the box and starting eating that cereal before we could weigh the box..." These comments suggest that a non-random, informal collection procedure was used. When the data are examined, you discover that the sample size is only 5 for each of the two cereals. Drawing a random sample, and using a larger sample size would add rigor by reducing the variability in the sample means.

2. (a)

Oxford O's	Alpine Frosted Flakes
360.4	366.1
361.8	367.2
362.3	365.6
364.2	367.8
371.4	373.5
364.02	368.04

(b) If □ = 15, then $\sigma_{\bar{X}} = \frac{15}{\sqrt{5}} = 6.7082$, and with an expected population mean of 368 grams,

Normal Probabilities

Cereal Weight for Oxford O's	
Mean	368
Standard Deviation	6.7082
Probability for X <=	
X Value	364.02
Z Value	-0.593304
P(X<=364.02)	0.2764889

The likelihood of obtaining a sample average weight of no more than 364.02 grams if the population weight is 368 grams is 27.65%.

Normal Probabilities

Cereal Weights for Alpine Frosted Flakes	
Mean	368
Standard Deviation	6.7082
Probability for X <=	
X Value	368.04
Z Value	0.005962851
P(X<=368.04)	0.502378833

The likelihood of obtaining a sample average weight of no more than 368.04 grams if the population weight is 368 grams is 50.24%.

(c)

Normal Probabilities

Cereal Weight for Oxford O's	
Mean	368
Standard Deviation	15
Probability for X <=	
X Value	364.02
Z Value	-0.265333
P(X<=364.02)	0.3953764

The likelihood of obtaining an individual weight of no more than 364.02 grams if the population weight is 368 grams is 39.54%.

Normal Probabilities

Cereal Weights for Alpine Frosted Flakes	
Mean	368
Standard Deviation	15
Probability for X <=	
X Value	368.04
Z Value	0.002666667
P(X<=368.04)	0.501063851

The likelihood of obtaining an individual weight of no more than 368.04 grams if the population weight is 368 grams is 50.11%.

3. There is a fairly high chance that an individual box of Oxford O's or the mean of a sample of five boxes will have a weight below 364.02 grams. There is more than a 50% chance that an individual box of Alpine Frosted Flakes or the average of a sample of five boxes will have a weight below 368.04 grams. This is true even though four of the five boxes in each sample contain less than 368 grams.
4. Arguments for being reasonable:
   - Statistical procedure used is invalid.
   - The mean of the one group actually exceeds 368.
   - Confusion over conclusions that can be drawn from a sample.
   - Possibility of investigator bias.

   Arguments against:
   - Data are available for independent review.
   - Oxford is producing some boxes of cereal that had less cereal than claimed on their boxes.
   - Right of individuals to freely express non-libelous opinions.
5. Even for the Oxford O's sample, you cannot prove cheating without using statistical inference. When the techniques of the next two chapters are applied, it will turn out that with these samples, there is insufficient evidence that the population mean is less than 368 grams.

Chapter 8—Confidence Interval Estimation

Instructional Tips

This digital case focuses on two concepts – the need to develop confidence interval estimates rather than point estimates, and using statistical methods to determine sample size.

Solutions

Pay A Friend

Data	
Sample Size	**50**
Number of Successes	**22**
Confidence Level	**95%**
Intermediate Calculations	
Sample Proportion	0.44
Z Value	-1.95996279
Standard Error of the Proportion	0.070199715
Interval Half Width	0.137588829
Confidence Interval	
Interval Lower Limit	**0.302411171**
Interval Upper Limit	**0.577588829**

Conbanco

Data	
Sample Size	**50**
Number of Successes	**28**
Confidence Level	**95%**
Intermediate Calculations	
Sample Proportion	0.56
Z Value	-1.95996279
Standard Error of the Proportion	0.070199715
Interval Half Width	0.137588829
Confidence Interval	
Interval Lower Limit	**0.422411171**
Interval Upper Limit	**0.697588829**

The confidence interval estimate for each of the two groups includes 0.50 or 50%. The proportion in the population using Pay A Friend is estimated to be between 30.2% and 57.8%, while the proportion using Conbanco is estimated to be between 42.2% and 69.8%.

Pay a Friend

Data	
Sample Standard Deviation	**12.00647704**
Sample Mean	**30.36636364**
Sample Size	**22**
Confidence Level	**95%**
Intermediate Calculations	
Standard Error of the Mean	2.559789506
Degrees of Freedom	21
t Value	2.079614205
Interval Half Width	5.323374619
Confidence Interval	
Interval Lower Limit	**25.04**
Interval Upper Limit	**35.69**

Conbanco

Data	
Sample Standard Deviation	**7.098347673**
Sample Mean	**23.17285714**
Sample Size	**28**
Confidence Level	**95%**
Intermediate Calculations	
Standard Error of the Mean	1.341461619
Degrees of Freedom	27
t Value	2.051829142
Interval Half Width	2.752450042
Confidence Interval	
Interval Lower Limit	**20.42**
Interval Upper Limit	**25.93**

The 95% confidence interval estimate for the mean payment amount is $25.04 to $35.69 for Pay a Friend and $20.42 to $25.93 for Conbanco.

Since the confidence intervals for both Pay a Friend and Conbanco include 0.50 or 50%, there is no evidence that customers use the two forms of payment in unequal numbers. Since there is some overlap in the two confidence intervals for the mean, it is hard to conclude that there is a difference in the mean purchases for the two forms of payment. However, these data are useful in pointing out the fact that when comparing differences between the means of two groups, confidence interval estimates for each group should not be compared. In fact, the correct procedure is to use the *t*-test for the difference between the means and the confidence interval estimate for the difference between two means (to be covered in Chapter 10). The results of this test indicate a significant difference in the mean purchase amount between the two forms of payment.

t-Test: Two-Sample Assuming Equal Variances

	PAF	*Conbanco*
Mean	**30.36636**	**23.17286**
Variance	**144.1555**	**50.38654**
Observations	**22**	**28**
Pooled Variance	**91.41046**	
Hypothesized Mean Difference	**0**	
df	**48**	
t Stat	**2.640876**	
P(T<=t) one-tail	**0.005563**	
t Critical one-tail	**1.677224**	
P(T<=t) two-tail	**0.011126**	
t Critical two-tail	**2.010634**	

Using the range of the data divided by 6 as an estimate of the population standard deviation [(72.12 – 12.84)/6] equal to 9.88, the sample size necessary for 95% confidence with a sampling error of ± $3 is 42. Thus, a sample size of 50 is appropriate.

Chapter 9—Fundamentals of Hypothesis Testing: One-Sample Tests

Instructional Tips

There are several objectives involved in this digital case.

1. Have students question the validity of data collected.
2. Have students looking for hidden issues that could invalidate a set of conclusions.
3. Have students use hypothesis testing to draw conclusions about a claimed value.
4. Increase students' understanding of the effect of sampling on a conclusion.

Solutions

1. Issues that could be raised about the testing process – the size of the sample, how the sample was selected, the selection of only two brands of cereals, the identity of the independent testers (not disclosed), whether, as discussed in a subsequent chapters, there is a single sample or in fact, samples of two different cereals.. Also, if you read all of the materials related to the television station, you could raise issues about the independence of the consumer reporter and wonder why only one out of four plants was chosen for this analysis.

2.

t Test for Hypothesis of the Mean

Data	
Null Hypothesis μ=	**368**
Level of Significance	**0.05**
Sample Size	**80**
Sample Mean	**370.433375**
Sample Standard Deviation	**14.70776355**
Intermediate Calculations	
Standard Error of the Mean	1.644377955
Degrees of Freedom	79
t Test Statistic	1.479814901
Lower-Tail Test	
Lower Critical Value	**-1.664370757**
***p*-Value**	**0.928550208**
Do not reject the null hypothesis	

The mean weight is actually above the hypothesized weight of 368 grams by 1.48 standard deviation units. Clearly, with a *p*-value of 0.929, there is no reason to believe that the mean weight is *below* 368 grams.

However, as noted in the press release, samples of two different cereals were selected, so the question can be raised as to whether separate analyses should have been done on each cereal.

3. The claim is true since 42 boxes contain more than 368 grams. However, if the mean were equal to 368, you would expect that approximately half of the boxes would contain more than 368 grams, so the result is certainly not surprising. Of course, the Oxford CEO does not mention that 38 boxes contained less than 368 grams.
4. Sample statistics will vary from sample to sample. It is possible that a sample with a mean below 368 grams and a sample with a mean above 368 grams will both lead to the conclusion that there is insufficient evidence that the population mean is below 368 grams. In fact, if you use the CCACC sample of 10 cereal boxes discussed in Chapter 7, the results of the test for whether the population mean is below 368 are not significant.

t Test for Hypothesis of the Mean

Data	
Null Hypothesis μ=	**368**
Level of Significance	**0.05**
Sample Size	**10**
Sample Mean	**366.03**
Sample Standard Deviation	**4.165746565**
Intermediate Calculations	
Standard Error of the Mean	1.31732473
Degrees of Freedom	9
t Test Statistic	-1.495455111
Lower-Tail Test	
Lower Critical Value	**-1.833113856**
***p*-Value**	**0.08450497**
Do not reject the null hypothesis	

Chapter 10—Two-Sample Tests

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Understand that just having sample statistics does not mean that claims can be made about differences between groups without using hypothesis testing.
2. Use two-sample tests of hypothesis to determine whether there are significant differences between two groups.

Solutions

1. Although the means of the two samples are different, without the necessary tests of hypothesis, you cannot infer that the two processes are statistically different. This, of course, assumes that CCACC has drawn random samples, something that is unclear in their posting.
2.

t-Test: Two-Sample Assuming Equal Variances

	Plant 1	*Plant 2*
Mean	372.441	365.637
Variance	180.8843	101.1672
Observations	10	10
Pooled Variance	141.0257	
Hypothesized Mean Difference	0	
df	18	
t Stat	1.28115	
P(T<=t) one-tail	0.108201	
t Critical one-tail	1.734063	
P(T<=t) two-tail	0.216401	
t Critical two-tail	2.100924	

F-Test Two-Sample for Variances

	Plant 1	*Plant 2*
Mean	372.441	365.637
Variance	180.8843	101.1672
Observations	10	10
df	9	9
F	1.787974	
P(F<=f) one-tail	0.199863	
F Critical one-tail	3.178897	

Wilcoxon Rank Sum Test for Plant 1 and Plant 2

Data	
Level of Significance	**0.05**

Population 1 Sample	
Sample Size	10
Sum of Ranks	119
Population 2 Sample	
Sample Size	10
Sum of Ranks	91

Intermediate Calculations	
Total Sample Size n	20
*T*1 Test Statistic	119
*T*1 Mean	105
Standard Error of *T*1	13.22876
Z Test Statistic	**1.058301**

Upper-Tail Test	
Upper Critical Value	**1.644853**
***p*-value**	**0.144959**
Do not reject the null hypothesis	

The *t*-test for the difference between the means indicates a test statistic of $t_{STAT} = 1.28$ and a one-tail *p*-value of 0.108. The *F*-test for the equality of variances indicates a test statistic $F_{STAT} = 1.788$ and a two-tailed *p*-value of 0.40. The Wilcoxon rank sum test (covered in Section 12.6) indicates a test statistic of $Z_{STAT} = 1.058$ and a one-tail *p*-value of 0.145. Thus, there is insufficient statistical evidence to indicate any difference in the mean, median, or variability between Plant 1 and Plant 2.

Chapter 11—Analysis of Variance

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Understand that just having sample statistics does not mean that claims can be made about differences between groups without using hypothesis testing.
2. Use the one-factor Analysis of Variance to determine whether there are significant differences between two groups.
3. See that there can be anomalies that can occur when analyzing data in which one analysis can lead to a certain conclusion, and a different analysis might lead to another conclusion.

Solutions

1. Yes, because Oxford Cereals operates four plants, a careful examination would explore if there are differences among the four plants. A proper sample of the population of cereal boxes would include boxes from all four plants. In addition, as in an earlier case, it is unclear if the CCACC sample is randomly drawn from all cereal boxes available. From their posting, it seems as if their members actively excluded boxes from plants other than #1 and #2.
2. In order to determine whether there is a difference in the weights among the four plants, a one-factor analysis of variance needs to be done.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Plant 1	20	7448	372.4	132.1037
Plant 2	20	7324.07	366.2035	218.1177
Plant 3	20	7393.12	369.656	222.0002
Plant 4	20	7531.72	376.586	131.1284

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1155.949	3	385.3162	2.19132	0.095938	2.724946
Within Groups	13363.65	76	175.8375			
Total	14519.6	79				

Kruskal-Wallis Test of Cereal Weights

Data	
Level of Significance	**0.05**
Intermediate Calculations	
Sum of Squared Ranks/Sample Size	134728.4
Sum of Sample Sizes	80
Number of groups	4
H Test Statistic	6.496991
Test Result	
Critical Value	**7.814725**
***p*-Value**	**0.089781**
Do not reject the null hypothesis	

The ANOVA results with an F_{STAT} test statistic equal to 2.19 < 2.72 or a *p*-value = 0.0959 > 0.05, indicates that there is insufficient evidence to conclude that there is a difference in the means of the four plants. The nonparametric Kruskal-Wallis test (covered in Section 12.7) provides similar results with a χ^2_{STAT} test statistic = 6.497 < 7 815 or a *p*-value = 0.0898 > 0.05.

Interestingly, had CCACC argued that something was amiss only in Plant 2, but not in Plants 1, 3, and 4, there is some evidence that this is the case. Using an a priori research hypothesis that focused on testing differences between plants 1, 3, and 4 as compared to plant 2, the following results are obtained.

t-Test: Two-Sample Assuming Equal Variances

	Plant 1, 3, 4	***Plant 2***
Mean	**372.880667**	**366.2035**
Variance	**164.518528**	**218.1177**
Observations	**60**	**20**
Pooled Variance	**177.574747**	
Hypothesized Mean Difference	**0**	
df	**78**	
t Stat	**1.94065012**	
P(T<=t) one-tail	**0.02795698**	
t Critical one-tail	**1.66462542**	
P(T<=t) two-tail	**0.05591397**	
t Critical two-tail	**1.99084752**	

Since t_{STAT} = 1.94 > 1.664 or the *p*-value = 0.028 < 0.05, there is evidence that the mean weight of cereal boxes in plants 1, 3, and 4 is greater than the mean weight in plant 2.

3. The one-way ANOVA shows that the null hypothesis cannot be rejected, so you cannot claim a statistical difference among the four plants. The mean weight of the 80 boxes in the sample is 371.2 grams, consistent with a claim that boxes average 368 grams.

   Interestingly, an analysis that pits Plant #2 against the other plants indicates that a statistically significant difference does occur. There may be something different happening in Plant #2, after all. That said, if the source of cereal boxes for sale were randomly distributed, consumers would, over time, be unlikely to be "cheated."

   Quantifiable claims must be substantiated by the proper statistical analysis. While the CCACC may, in fact, have at least one valid point, the group cannot offer any legitimate evidence to support their claims. So, at least at this point, you should not testify on the group's behalf.

Chapter 12—Chi-Square Tests and Nonparametric Tests

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Understand the difference between the results from a one-way table and a two-way contingency table.
2. Be able to use the chi-square test to determine whether a relationship exists between two categorical variables.
3. Be able to see the importance of examining differences between groups in their response to a categorical variable.

Solutions

1. They are literally true since 181 of the respondents prefer the Sun Low Concierge Class program as compared to 119 who prefer the T. C. Resorts TCPass Plus. However, since the program is described as aimed at business travelers, other interpretations of the data can be made.
2. By examining the preferences of business travelers, the target for the program, especially those business travelers who use travel programs, or by examining the resort last visited by type of traveler.
3.

Program Preference by Travel Program

Observed Frequencies			
	Program Preference		
Uses Travel Program	**TC Pass Plus**	**Concierge Class**	**Total**
Yes	**55**	**20**	**75**
No	**64**	**161**	**225**
Total	**119**	**181**	**300**

Expected Frequencies			
	Program Preference		
Uses Travel Program	TC Pass Plus	Concierge Class	Total
Yes	29.75	45.25	75
No	89.25	135.75	225
Total	119	181	300

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

Results	
Critical Value	**3.841455338**
Chi-Square Test Statistic	**47.3606017**
***p*-Value**	**5.90578E-12**
Reject the null hypothesis	

Expected frequency assumption is met.

There is a significant difference in preference for TCPass Plus versus Concierge Class based on whether the respondent uses a travel rewards program ($\chi^2_{STAT} = 47.361 > 3.841$, p-value = 0.000 < 0.05. Those who use travel rewards programs clearly prefer TCPass Plus (73.3%) over Concierge Class, while those who do not use travel rewards programs prefer Concierge Class (71.6%).

Program Preference by Travel Program

Observed Frequencies			
	Program Preference		
Customer Type	**TC Pass Plus**	**Concierge Class**	**Total**
Business	**34**	**16**	**50**
Leisure	**85**	**165**	**250**
Total	**119**	**181**	**300**

Expected Frequencies			
	Program Preference		
Customer Type	TC Pass Plus	Concierge Class	Total
Business	19.83333333	30.16666667	50
Leisure	99.16666667	150.8333333	250
Total	119	181	300

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

Results	
Critical Value	**3.841455338**
Chi-Square Test Statistic	**20.12628256**
***p*-Value**	**7.24936E-06**
Reject the null hypothesis	

Expected frequency assumption is met.

There is a significant difference in preference for TCPass Plus versus Concierge Class based on whether the respondent is a business or leisure traveler ($\chi^2_{STAT} = 20.126 > 3.841$, p-value = 0.000 < 0.05. Business travelers clearly prefer TCPass Plus (68%) over Concierge Class, while leisure travelers prefer Concierge Class (66%).

4. Further analysis indicates that of 41 business travelers who use travel reward programs, 31 prefer TCPass Plus. Of 34 leisure travelers who use travel reward programs, 24 prefer TCPass Plus. Thus, it is reasonable to conclude that TCPass Plus is preferred by the target audience of business travelers and also by those who use travel reward programs.

   Among other factors that might be included in future surveys are whether the travel program influences the choice of accommodation, what attributes of a resort chain are desirable for business travelers, and the reasons for the attractiveness of Concierge Class for leisure travelers.

Chapter 13—Simple Linear Regression

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Perform a simple linear regression analysis to determine the usefulness of an independent variable in predicting a dependent variable.
2. Understand the danger in making predictions that extrapolate beyond the range of the independent variable.

Solutions

1.

Regression Analysis

Regression Statistics	
Multiple R	0.698234618
R Square	0.487531581
Adjusted R Square	0.44482588
Standard Error	2.234863491
Observations	14

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	57.01890785	57.01890785	11.41607709	0.005480622
Residual	12	59.93537787	4.994614822		
Total	13	116.9542857			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-1.941218839	2.379988792	-0.815642009	0.430597414
Average Disposable Income($000)	0.192948059	0.05710603	3.378768576	0.005480622

Yes, there is a correlation between the variables, but not a very strong one, given the r^2 value of only 0.49. The sales projection claim should be discarded as Triangle is attempting to extrapolate sales outside the range of the X values. This raises a related point: Sunflowers clearly has not done business in areas of "exceptional affluence," so there is no track record on which to base a decision to accept or reject Triangle's proposal.

2. No, because the r^2 value of mean disposable income with sales is only 0.49 as compared to an r^2 value of 0.904 for store size. In fact, a multiple regression analysis reveals that given that store size is included in the regression model, adding mean disposable income does not significantly improve the model.
3. Yes, given the r^2 value of only 0.49, it is less significant than other single factors such as store size. However, opening a new retail location would be based on a number of factors (some of these factors such as competitive retail analysis, demographic and geographic profiles, regional economic analysis, and sales potential forecast analysis, are actually mentioned by Triangle in its proposal).
4. The Sunflowers brand perception and merchandise mix would be important as well. For example, a store selling hip junior swimsuit fashions would not do well in a community of senior citizens in wintry Minnesota. The financial health of the Sunflowers chain would be another factor—many retail chains have gone out of business due to unwise overexpansion.

Chapter 14—Introduction to Multiple Regression

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Evaluate the contribution of dummy variables to a multiple regression model.
2. Determine whether an interaction term needs to be included in a regression model that has a dummy variable.

Solutions

1.

Regression Analysis

Regression Statistics	
Multiple R	0.931595697
R Square	0.867870543
Adjusted R Square	0.849645791
Standard Error	487.1843555
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	45210568.15	11302642.04	47.62042926	2.44077E-12
Residual	29	6883109.291	237348.5962		
Total	33	52093677.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5432.871895	516.4342024	10.51996918	2.06405E-11
Price	-53.0446975	5.235748581	-10.1312537	4.90559E-11
Promotion	3.564958066	0.565215244	6.307257458	6.8805E-07
Shelf Location	815.3759143	169.3097032	4.815884139	4.23097E-05
Dispensers	100.3258952	180.4650079	0.555929908	0.582522982

The presence of dispensers does not make a significant contribution to the multiple regression model since the *p*-value = 0.5825 > 0.05. Therefore it should be eliminated from consideration in the model.

Regression Analysis

Regression Statistics	
Multiple R	0.93083963
R Square	0.866462417
Adjusted R Square	0.853108658
Standard Error	481.5414066
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	45137213.65	15045737.88	64.88528515	3.20721E-13
Residual	30	6956463.788	231882.1263		
Total	33	52093677.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5533.560441	478.0310999	11.57573313	1.36426E-12	4557.291698	6509.829184
Price	-53.1560416	5.171316284	-10.2790157	2.40096E-11	-63.7172675	-42.5948157
Promotion	3.4475624	0.51821306	6.652789493	2.28496E-07	2.389231231	4.505893568
Shelf location	823.8004992	166.6769534	4.942497943	2.74026E-05	483.4010988	1164.1999

Regression with Interaction Terms

Regression Statistics	
Multiple R	0.942343966
R Square	0.88801215
Adjusted R Square	0.86801432
Standard Error	456.4560263
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	46259818.53	9251963.706	44.40542491	1.85186E-12
Residual	28	5833858.91	208352.1039		
Total	33	52093677.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5392.816086	687.0076402	7.849717778	1.50017E-08	3985.54315	6800.089023
Price	-55.6660643	7.170933758	-7.76273582	1.86365E-08	-70.3550727	-40.9770559
Promotion	4.335268841	0.691008467	6.273828826	8.78762E-07	2.919800572	5.75073711
Shelf location	963.5819136	911.118414	1.057581428	0.299285645	-902.7616484	2829.925476
Price*Shelf	8.628987422	9.975146401	0.8650487	0.394363297	-11.8041966	29.0621715
Promotion*Shelf	-2.07339807	1.001589281	-2.070108089	0.047779128	-4.12506301	-0.02173313

$SSR(X_1,X_2,X_3,X_4,X_5)$ = 46,259,818.53 with 5 degrees of freedom
$SSR(X_1,X_2,X_3)$ = 45,137,213.65 with 3 degrees of freedom
Thus, $SSR(X_1,X_2,X_3,X_4,X_5)$ - $SSR(X_1,X_2,X_3)$ = 46,259,818.53 - 45,137,213.65 = 1,122,604.88

To test a null hypothesis of no interaction effect,
F_{STAT} = 1,122,604.88/2 divided by $MSE(X_1,X_2,X_3,X_4,X_5)$ = 208,352.1039
F_{STAT} = 561,302.44/208,352.1039 = 2.694 < 3.34, there is no evidence that the interaction terms together significantly improve the regression model. Testing each interaction term separately, from the previous output since the *t* statistic for the interaction of promotion and shelf location is -2.07 with a *p*-value of 0.0478, it is a candidate for inclusion in the regression model.

Regression Analysis

Regression Statistics	
Multiple R	0.940754611
R Square	0.885019238
Adjusted R Square	0.869159823
Standard Error	454.4709208
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	46103906.72	11525976.68	55.80402648	3.31112E-13
Residual	29	5989770.717	206543.8178		
Total	33	52093677.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5000.710782	514.0114463	9.728792652	1.22694E-10	3949.438762	6051.982801
Price	-51.2067026	4.963082378	-10.3175201	3.23215E-11	-61.3573513	-41.0560539
Promotion	4.452725243	0.674590482	6.600634552	3.11043E-07	3.073032041	5.832418446
Shelf location	1662.849646	418.5247884	3.973121048	0.000430336	806.8698756	2518.829416
Promotion*Shelf	-2.14915821	0.993413794	-2.16340686	0.038889122	-4.18091866	-0.11739777

Thus, there is a significant effect of shelf location on sales with end aisle location having a positive effect on sales. However, the effect of the end aisle location is not the same across different levels of promotion with a slight decrease in its effect with increasing levels of promotion expenses. In addition, there is no evidence of any patterns in the residual plots.

2. You would recommend using the end aisle location but not use in-store coupon dispensers.

3. Actual sales by linear display feet (the linear size of the product stock area), the number of OmniPower coupons dispensed per store, the number of coupon dispensers per store, and the amount or existence of special in-store signage or advertising panels.

Chapter 15—Multiple Regression Model Building

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Be able to determine which one of a set of competing claims concerning regression results are correct.
2. Use model building approach to determine the best fitting model.
3. Evaluate the contribution of dummy variables to a multiple regression model.
4. Determine whether an interaction term needs to be included in a regression model that has a dummy variable.
5. Use the coefficient of partial determination to evaluate the importance of each independent variable.

Solutions

1.

Best Subsets Regression for Predicting OmniPower Bars sold

Intermediate Calculations	
*R*2T	0.902879
1 - *R*2T	0.097121
n	34
T	5
n - T	29

Model	Cp	k	R Square	Adj. R Square	Std. Error
X1	107.2231	2	0.54044	0.526078336	864.9457
X1X2	44.34219	3	0.757726	0.742095357	638.0653
X1X2X3	13.87386	4	0.866462	0.853108658	481.5414
X1X2X3X4	5	5	0.902879	0.889482975	417.6862
X1X2X4	31.07428	4	0.808858	0.789744002	576.1157
X1X3	70.70057	3	0.669452	0.648126018	745.2966
X1X3X4	69.62005	4	0.679768	0.647745201	745.6997
X1X4	103.7214	3	0.558865	0.530404528	860.9888
X2	183.1004	2	0.286327	0.264024476	1077.872
X2X3	152.3072	3	0.396151	0.35719323	1007.339
X2X3X4	124.6271	4	0.49555	0.445104867	935.9249
X2X4	148.6162	3	0.408512	0.370351656	996.9757
X3	228.6949	2	0.13363	0.106556372	1187.597
X3X4	216.0288	3	0.182748	0.130021543	1171.898
X4	249.0457	2	0.065476	0.036271692	1233.425

The only model that has a C_p close to or less than the number of terms in the model is the model with all four independent variables ($C_p = 5$).

Regression Analysis

Regression Statistics	
Multiple R	0.950199441
R Square	0.902878978
Adjusted R Square	0.889482975
Standard Error	417.6862048
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	47034286.24	11758571.56	67.39913192	2.91747E-14
Residual	29	5059391.205	174461.7657		
Total	33	52093677.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	5079.497556	436.907215	11.62603267	1.94744E-12	4185.921481	5973.073631
Price	-50.35076989	4.565528109	-11.02846564	6.8443E-12	-59.6883284	-41.01321137
Promotion	3.739140134	0.45810945	8.162110899	5.32561E-09	2.802200597	4.676079672
Shelf Location	770.7578086	145.4667039	5.298517033	1.10652E-05	473.2448314	1068.270786
Number of Dispensers	107.8739853	32.71333956	3.297553437	0.002583173	40.96765705	174.7803136

Now the number of dispensers makes a significant contribution to the regression model (p-value = 0.00258). However, the need for interaction terms must be determined prior to the selection of a final model.

Regression Analysis

Regression Statistics	
Multiple R	0.96009987
R Square	0.92179176
Adjusted R Square	0.9007357
Standard Error	395.851313
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	48019522.62	6859931.804	43.77796989	8.5067E-13
Residual	26	4074154.816	156698.2622		
Total	33	52093677.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4633.96393	660.8118467	7.012531565	1.90645E-07
Price	-49.4903825	6.639664883	-7.453747045	6.4894E-08
Promotion	4.72233798	0.616742898	7.656898832	3.98385E-08
Shelf Location	1458.09724	850.1444634	1.71511702	0.098220374
Number of Dispensers	112.959469	42.54881228	2.654820729	0.013364798
Price*Shelf	2.36399254	8.95813567	0.263893362	0.79394251
Promotion*Shelf	-2.1693067	0.892191533	-2.431436093	0.022235822
Shelf*Dispensers	-16.2036751	63.63227257	-0.254645551	0.801000223

$SSR(X_1,X_2,X_3,X_4,X_5,X_6,X_7)$ = 48,019,522.62 with 7 degrees of freedom

$SSR(X_1,X_2,X_3,X_4)$ = 47,034,286.24 with 4 degrees of freedom

Thus, $SSR(X_1,X_2,X_3,X_4,X_5)$ - $SSR(X_1,X_2,X_3)$ = 48,019,522.62 - 47,034,286.24 = 985,236.38

To test a null hypothesis of no interaction effect,

F_{STAT} = 985,236.38 /3 divided by $MSE(X_1,X_2,X_3,X_4,X_5,X_6,X_7)$ = 156,698.2622

F_{STAT} = 328,412.1267/156,698.2622= 2.096 < 2.98, there is no evidence that the interaction terms together significantly improve the regression model. Testing each interaction term separately, from the previous output since the *t* statistic for the interaction of promotion and shelf location is -2.43 with a *p*-value of 0.022, it is a candidate for inclusion in the regression model.

Regression Analysis

Regression Statistics	
Multiple R	0.959846036
R Square	0.921304414
Adjusted R Square	0.90725163
Standard Error	382.6385132
Observations	34

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	47994134.95	9598826.99	65.56028053	1.39552E-14
Residual	28	4099542.49	146412.2318		
Total	33	52093677.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4549.354259	450.6308677	10.09552293	7.79946E-11
Price	-48.41339383	4.250332669	-11.3904951	5.04359E-12
Promotion	4.740215982	0.573574734	8.264338889	5.40529E-09
Shelf Location	1606.929657	352.7174462	4.555855328	9.33165E-05
Number of Dispensers	107.6795585	29.96848747	3.593092864	0.001236438
Promotion*Shelf	-2.141543368	0.836400262	-2.56042885	0.016135997

In addition, there is no evidence of any patterns in the residual plots.

It appears that the predicted increase in sales from the number of dispensers is approximately 108 bars for each dispenser available. As was the case with the regression analysis in Chapter 14, there is a negative interaction effect of promotion amount with shelf location, with the effect of end aisle placement decreasing with increasing promotion.

Mari has overstated her claim when she writes “I'm sure if we look at the number of dispensers we will find the reason for the seeming lack of success,” We cannot be totally sure, but only believe that there is statistical evidence that indicates an effect due to the number of dispensers.

Ted offers an unsubstantiated opinion and the question that would arise eventually is how best to measure the effectiveness of the methods he mentions.

2. All the independent variables of price, promotion expenses, shelf location, and number of dispensers can be used to predict sales.

3. If only one independent variable could be used, the coefficient of partial determination would be helpful in determining which independent variable explained the most variation in sales holding constant the effect of the other independent variables.

Coefficients	
r2 Y1.2345	0.822496514
r2 Y2.1345	0.70923983
r2 Y3.1245	0.425709561
r2 Y4.1235	0.315576057
r2 Y5.1234	0.189716248

Price has the highest coefficient of partial determination followed by promotion expenses, shelf location, number of dispensers, and the interaction of promotion expenses and shelf location. Another approach would be to perform a cost-benefit analysis on each variable and use the results as a basis for selection.

Chapter 16—Time-Series Forecasting

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Be able to develop a time-series forecasting model using quarterly data.
2. Be able to compare the results of two forecasts and plot the raw time-series data on a graph.
3. Interpret the results of the time-series forecasting model including the compound growth rate and the seasonal multiplier.

Solutions

1. *Ashland Herald*

Ashland Herald Regression

Regression Statistics	
Multiple R	0.996366173
R Square	0.99274555
Adjusted R Square	0.990107568
Standard Error	0.001609736
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	0.003900635	0.000975159	376.327666	1.10551E-11
Residual	11	2.85037E-05	2.59125E-06		
Total	15	0.003929139			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5.044105529	0.001141807	4417.653248	1.00447E-35
Coded Q	0.003017445	8.9987E-05	33.53202179	1.98391E-12
Q1	-0.01261371	0.001169831	-10.7825107	3.46505E-07
Q2	-0.008951114	0.001152395	-7.76739835	8.63802E-06
Q3	-0.010059338	0.001141807	-8.81001951	2.58147E-06

The regression model for the *Ashland Herald* is

Log(Circulation) = 5.044 + 0.003 Coded Quarter – 0.0126 Quarter 1 – 0.00895 Quarter 2 – 0.010059 Quarter 3

The interpretation of the slopes is as follows:

- The estimated quarterly compound growth rate in sales is 0.697%
- 0.9714 is the seasonal multiplier for the first quarter as compared to the fourth quarter. Sales are 2.86% lower for the first quarter as compared to the fourth quarter.
- 0.9796 is the seasonal multiplier for the second quarter as compared to the fourth quarter. Sales are 2.04% lower for the second quarter as compared to the fourth quarter.
- 0.9771 is the seasonal multiplier for the third quarter as compared to the fourth quarter. Sales are 2.29% lower for the third quarter as compared to the fourth quarter.

Oxford Glen Journal

Oxford Glen Journal Regression

Regression Statistics	
Multiple R	**0.972433653**
R Square	**0.94562721**
Adjusted R Square	**0.925855286**
Standard Error	**0.017825405**
Observations	**16**

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	**4**	**0.060786879**	**0.01519672**	**47.82676826**	**6.87159E-07**
Residual	**11**	**0.003495196**	**0.000317745**		
Total	**15**	**0.064282075**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	**4.800407531**	**0.012643792**	**379.6651645**	**5.31459E-24**
Coded Q	**-0.008459265**	**0.00099647**	**-8.48922860**	**3.69877E-06**
Q1	**-0.160736195**	**0.012954116**	**-12.4081179**	**8.2427E-08**
Q2	**-0.10346711**	**0.012761048**	**-8.10804175**	**5.74866E-06**
Q3	**-0.092339127**	**0.012643792**	**-7.30311952**	**1.53677E-05**

The regression model for the *Oxford Glen Journal* is

Log(Circulation) = 4.8004 - 0.00846 Coded Quarter – 0.1607 Quarter 1

– 0.1035 Quarter 2 – 0.09234 Quarter 3

The interpretation of the slopes is as follows:

- The estimated quarterly compound growth rate in sales is -1.93%
- 0.6907 is the seasonal multiplier for the first quarter as compared to the fourth quarter. Sales are 30.93% lower for the first quarter as compared to the fourth quarter.
- 0.7880 is the seasonal multiplier for the second quarter as compared to the fourth quarter. Sales are 21.2% lower for the second quarter as compared to the fourth quarter.
- 0.8085 is the seasonal multiplier for the third quarter as compared to the fourth quarter. Sales are 19.15% lower for the third quarter as compared to the fourth quarter.

These results refute the claims of the *Oxford Glen Journal*. First, it is more appropriate to examine the data from four years than just the last year. Second, examining the data from the four years, the quarterly growth rate for the *Ashland Herald* is +0.7% as compared to a negative growth rate of almost 2% for the *Oxford Glen Journal*. Finally, the *Ashland Herald* has small seasonal effects of 2 – 3 % as compared to the fourth quarter, while the *Oxford Glen Journal* has large seasonal effects of between 19 and 31% as compared to the fourth quarter.

2. *Ashland Herald*: Its steady continuous growth with little variability from season to season.

   Oxford Glen Journal: The decline that occurred in year 3 did not continue. Sales in year 4 stabilized at about the same level as year 3.

3. Among other variables might be the actual number of readers per circulated copy, the demographics of the readers, rates of renewal or the so-called "churn rate" for subscribers.

Chapter 19 (Online Chapter)—Decision Making

Instructional Tips

The objectives for the digital case in this chapter are to have students:

1. Be able to use several criteria to determine a chosen course of action.
2. Revise probabilities in light of new information and determine if a previous course of action selected has changed.
3. Realize that better is a subjective word in making decisions.

Solutions

1.

Probabilities & Payoffs Table:

	P	StraightDeal	Happy Bull	Worried Bear
fast expanding	0.1	150	1200	-300
expanding	0.2	100	600	-200
stable	0.5	95	-100	100
recession	0.2	80	-900	400

Statistics for:	StraightDeal	Happy Bull	Worried Bear
Expected Monetary Value	98.5	10	60
Variance	340.25	382900	50400
Standard Deviation	18.44586675	618.7891402	224.4994432
Coefficient of Variation	0.187267683	61.87891402	3.741657387
Return to Risk Ratio	5.339949668	0.016160594	0.267261242

	StraightDeal	Happy Bull	Worried Bear
Expected Opportunity Loss	271.5	360	310
	EVPI		

Better is a subjective term that cannot be solely determined by a statistical analysis. If you accept the probabilities of the various events, StraightDeal should be selected since it has the highest expected monetary value ($98.50), the highest return-to-risk ratio (5.34), and the lowest expected value of perfect information ($271.50).

2.

Bayes' Theorem Calculations

		Probabilities		
Event	Prior	Conditional	Joint	Revised
fast expanding	0.1	0.9	0.09	0.1765
expanding	0.2	0.75	0.15	0.2941
stable	0.5	0.5	0.25	0.4902
recession	0.2	0.1	0.02	0.0392
		Total:	0.51	

Probabilities & Payoffs Table:

	P	StraightDeal	Happy Bull	Worried Bear
fast expanding	0.1765	150	1200	-300
expanding	0.2941	100	600	-200
stable	0.4902	95	-100	100
recession	0.0392	80	-900	400

Statistics for:	StraightDeal	Happy Bull	Worried Bear
Expected Monetary Value	105.59	303.96	-47.07
Variance	437.9369	304298.3184	36607.4151
Standard Deviation	20.92694196	551.6324124	191.3306434
Coefficient of Variation	0.198190567	1.814819096	-4.06481077
Return to Risk Ratio	5.045648819	0.551019108	-0.24601391

	StraightDeal	Happy Bull	Worried Bear
Expected Opportunity Loss	347.37	149	500.03
		EVPI	

Now the choice of which fund to invest in is much more difficult. Although Happy Bull has a higher expected monetary value than StraightDeal and a lower expected value of perfect information, it also has a much lower return-to-risk ratio. Perhaps a better approach would be to use the portfolio management approach covered in Section 5.2 to invest a proportion of assets in StraightDeal and a proportion in Happy Bull. For example, investing 70% in StraightDeal and 30% in Happy Bull would provide a portfolio expected return of $165.10 and a portfolio risk of $178.14, substantially below the standard deviation of Happy Bull of $551.63.