

APPLYING THE TOOLS

Example One

1. Problem Statement
2. Analyst Statement
3. SOLUTION

EXAMPLE ONE

The Spring Mills Company produces and distributes a wide variety of manufactured goods. Due to its variety, it has a large number of customers. It classifies these customers as small, medium and large, depending on the volume of business each does with Spring Mills. Recently, Spring Mills has noticed a problem with its accounts receivable. It is not getting paid back by its customers in as timely manner as it would be. This obviously costs Spring Mills money. If a customer delays a payment of \$300 for 20 days, say, then the company loses potential interest on this amount. The company has gathered data on 280 customer accounts. For each of these accounts, the data set lists three variables: Size, the size of customer (coded 1 for small, 2 for medium, 3 for large); Days, the number of days since the customer was billed; and Amount, the amount the customer owes. What information can we obtain from these data?

STATISTICAL ANALYST WRITER

I am working for the Spring Mills Company, and my boss, Sharon Sanders, has asked me to report on the account receivable problem our company currently experiencing. My task is to describe data on our customers, analyze the magnitude of interest lost because of late payments from our customers and suggest a solution for remedying the problem. Mr. Sanders knows basic statistics but she might need a refresh on the meaning of box plots.

SOLUTION

It is always a good idea to get a rough sense of the data first. We do this by calculating several summary measures for Days and Amount, a histogram of Amount and a scatter plot of Amount versus Days in Figures 3.16, 3.17, and 3.18. Figure 3.16 indicate positive skewness in the Amount variable- the mean is considerably larger than the median, probably because of some large amounts due. Also, the standard deviation of Amount is quite large. The positive skewness is confirmed by the histogram. The scatterplot suggests some suspicious behavior, with two distinct groups of points.

The next logical step is to see whether the different customer sizes have any effect on either Days, Amount, or the relationship between Days and Amount. To do this, it is useful to “unstack” the Days and Amount variables-that is, to create a new Days and Amount variable for each group customer sizes. For example, the Days and Amount variable for customer size1 are named Days1 and Amount1. (We used StatPro’s Unstack procedure, which is quite straightforward, to accomplish this, but copying and pasting also works.) Summary measures and

a variety of charts based on these unstuck variable appear in Figure 3.19 through 3.27

Figure 3.16 Summary Measures for the Combined Data

	A	B	C	D
1	Summary measures for selected variable			
2			Days	Amount
3		Count	280.000	280.000
4		Sum	4102.000	130000.000
5		Mean	14.650	464.286
6		Median	13.000	320.000
7		Standard deviation	7.221	378.055
8		Minimum	2.000	140.000
9		Maximum	39.000	2220.000

Figure 3.17 Histogram of All Amounts Owed

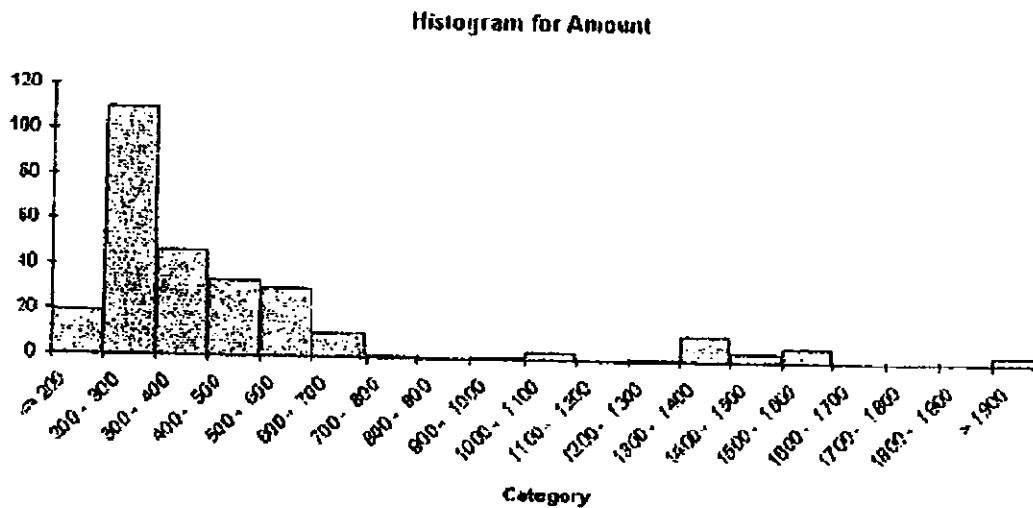


Figure 3.18 Scatterplot of Amount versus Days for All Customers

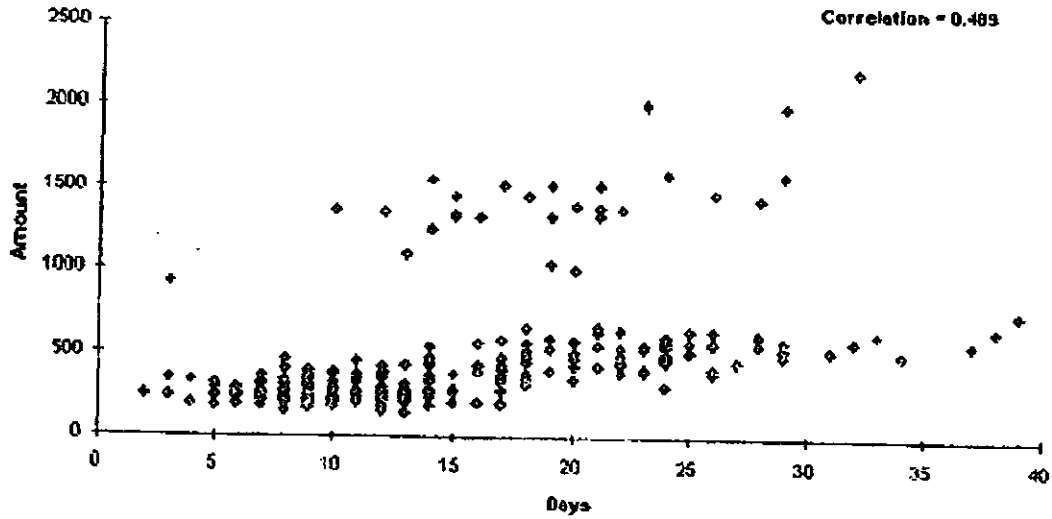


Figure 3.19 Summary Measures Broken Down by Size

	A	B	C	D	E	F	G	H
1	Summary measures for selected variable							
2			Days1	Amount1	Days2	Amount2	Days3	Amount3
3		Count	150.000	150.000	100.000	100.000	30.000	30.000
4		Sum	1470.000	38180.000	2055.000	48190.000	577.000	43630.000
5		Mean	9.800	254.533	20.550	481.000	19.233	1454.000
6		Median	10.000	250.000	20.000	470.000	19.000	1395.000
7		Standard deviation	3.128	49.285	6.622	99.155	6.191	293.000
8		Minimum	2.000	140.000	8.000	280.000	3.000	930.000
9		Maximum	17.000	410.000	39.000	750.000	32.000	2220.000

Figure 3.20 Histogram of Amount for Small Customers

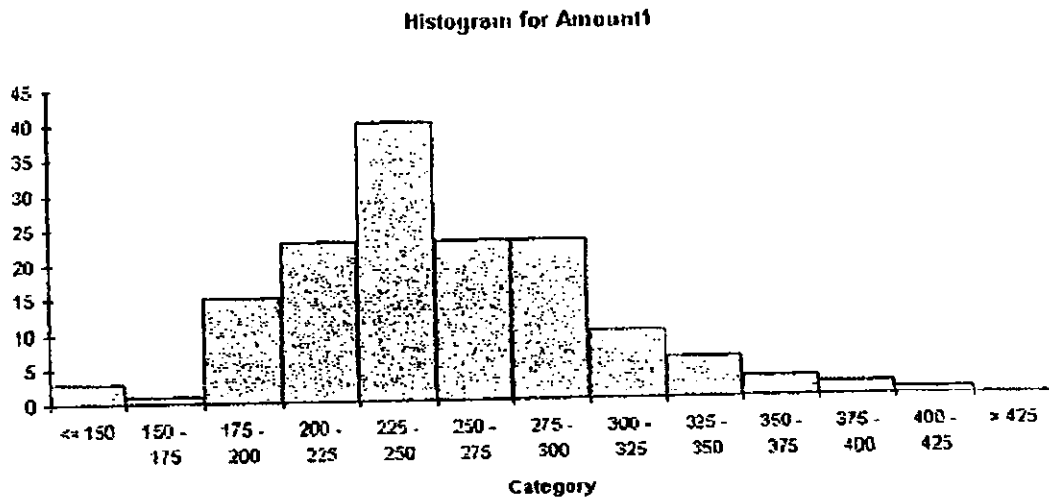


Figure 3.21 Histogram of Amount for Medium Customers

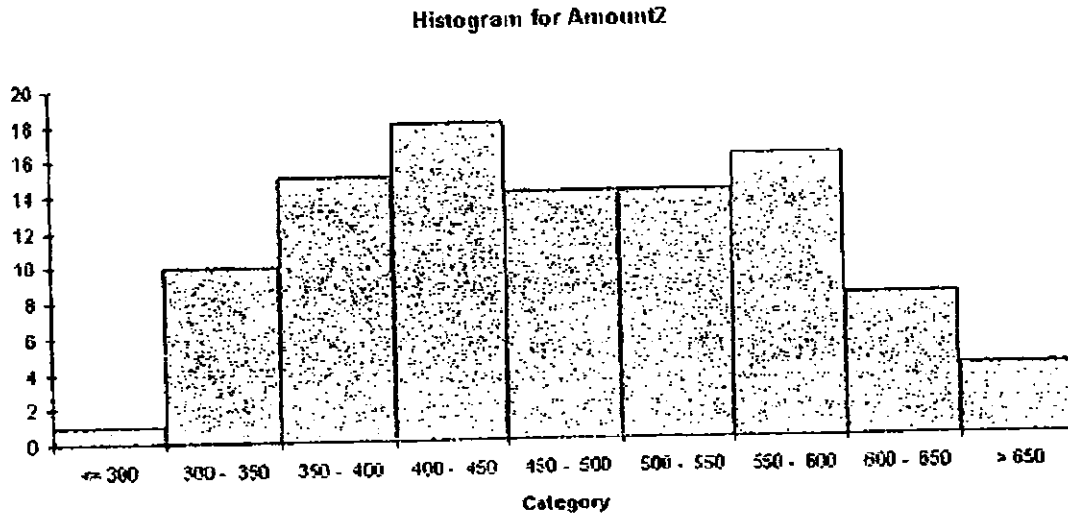


Figure 3.22 Histogram of Amount for Large Customers

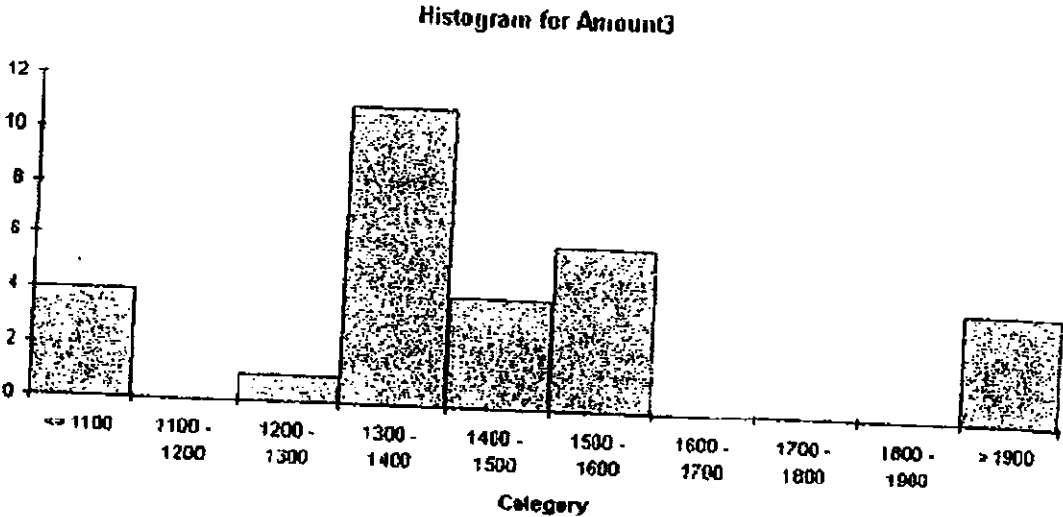


Figure 3.23 Boxplots of Days Owed by Different Size Customers

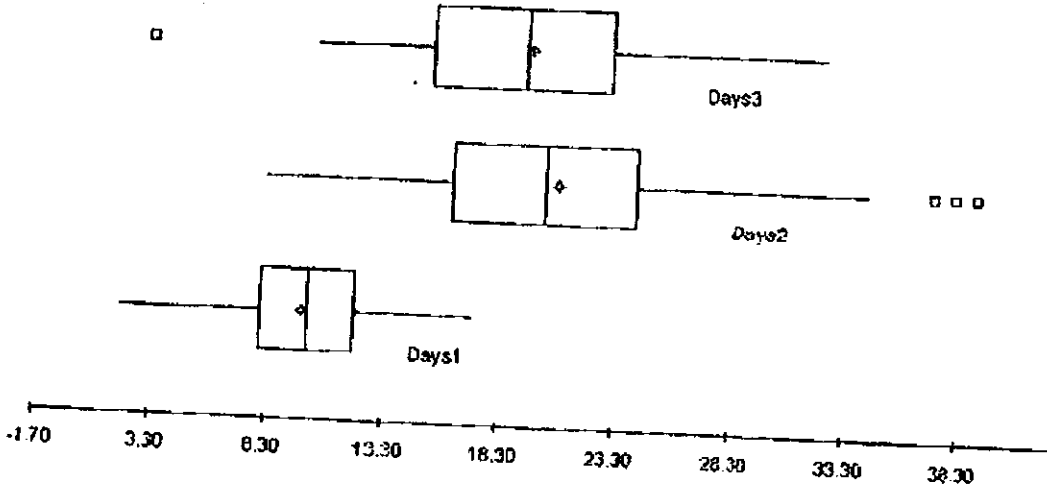


Figure 3.24 Boxplots of Amounts Owed by Different Size Customers

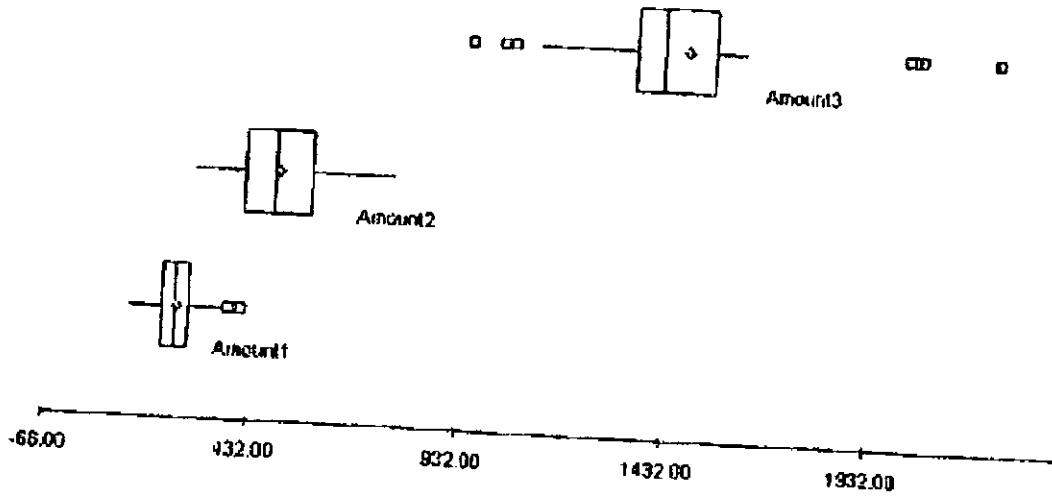


Figure 3.25 Scatterplot of Amount versus Days for Small Customers

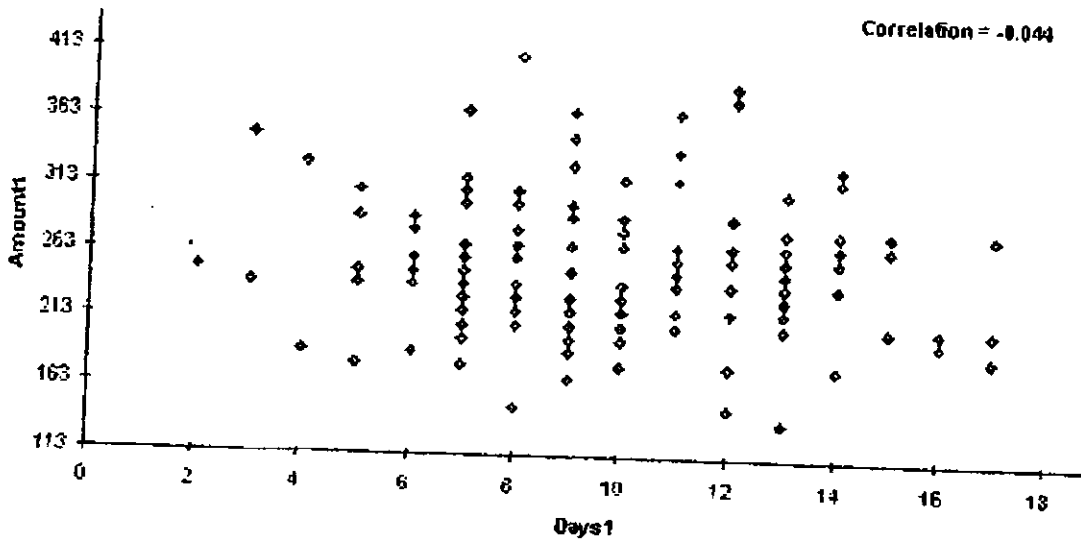


Figure 3.26 Scatterplot of Amount versus Days for Medium Customers

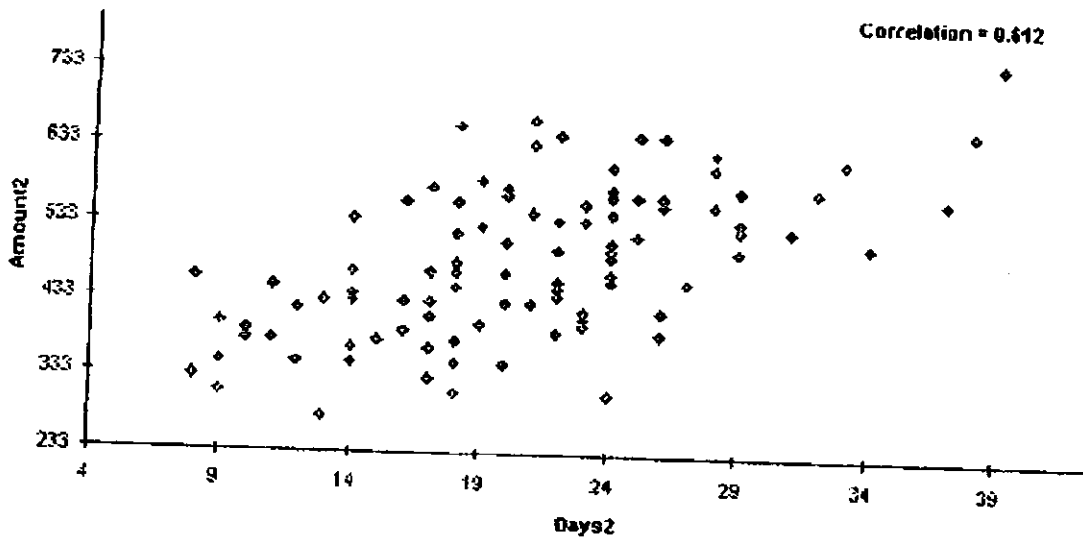
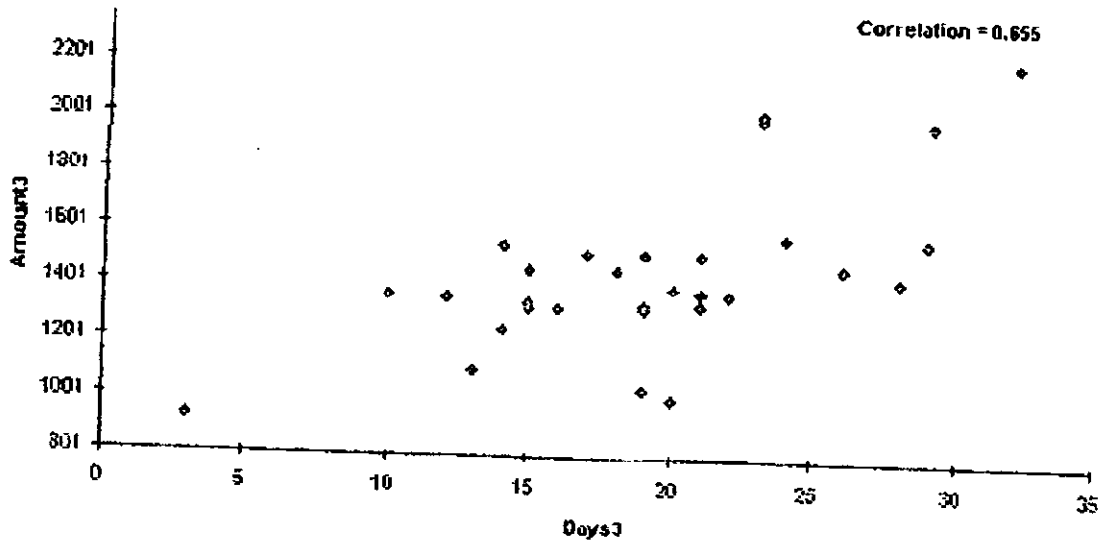


Figure 3.27 Scatterplot of Amount versus Days for Large Customers



There is obviously a lot going on here and most of it is clear from the figures. We point out the following: (1) there are many fewer large customers than small or medium customers; (2) the large customers tend to owe considerably more than small or medium customers; (3) the small customers do not tend to be as long overdue as the medium or large customers; and (4) there is no relationship between Days and Amount for the small customers, but there is a definite positive relationship between these variables for the medium and large customers.

We've now done the "obvious" analysis. There is still much more we can do, however. For example, suppose Spring Mills wants a breakdown of customers who owe at least \$500. We first create a new variable called "Large?" next to the original variable that equals 1 for all amounts greater than \$500 and equals 0 otherwise. (See figure 3.28 for some of the data.) We do this by entering the formula in cell D6 and copying down.

=IF(C6>=\$B\$3,1,0)

We then use a pivot table to create a *count* of the number of 1's in this new variable for each value of Size variable.

Figure 3.28 Checking for Amounts greater than \$500

	A	B	C	D
1	Distribution of receivables - checking for large values			
2				
3	Value to check for	\$500		
4				
5	Size	Days	Amount	Large?
6	1	7	\$180	0
7	1	8	\$210	0
8	1	10	\$210	0
9	1	8	\$150	0
10	1	9	\$300	0
11	1	5	\$240	0
12	1	4	\$330	0
13	1	10	\$290	0
14	1	5	\$240	0
15	1	13	\$270	0
16	1	12	\$220	0
17	1	11	\$260	0

Figure 3.29 shows the result. Actually, we created this pivot table twice, once (on top) showing counts as percentages of each column, and once showing them as percentages of each row. The top table shows, for example, that about 73% of all customers with amounts less than \$500 are small customers. The bottom table shows, for example, that about 45% of all medium-sized customers owe more than \$500. When you hear the expression "slicing and dicing the data" this is what it means. These two pivot tables are based on the *same* counts, but they

portray them slightly different ways. Neither is better than the other; each provides useful information.

Finally, we investigate the amount of interest Spring Mills is losing by the delays in its customers' payment. We assume that the company can make 12% annual interest on excess cash. Then we create a Lost variable for each customer size that indicates the amount of interest Spring Mills loses on each customer group. (See Figure 3.30.) The typical formula for lost interest in cell C10 is

$$=B10+A10*\$C\$7/365$$

Figure 3.29 Pivot Tables for Counts of Customers Who Owe More than \$500

	A	B	C	D
1				
2	Count of Large	Large?		
3	Size	0	1	Grand Total
4	1	73.17%	0.00%	53.57%
5	2	26.83%	60.00%	35.71%
6	3	0.00%	40.00%	10.71%
7	Grand Total	100.00%	100.00%	100.00%
8				
9	Count of Large	Large?		
10	Size	0	1	Grand Total
11	1	100.00%	0.00%	100.00%
12	2	55.00%	45.00%	100.00%
13	3	0.00%	100.00%	100.00%
14	Grand Total	73.21%	26.79%	100.00%

Figure 3.30 Summary Measures of Lost Interest

	A	B	C	D	E	F	G	H	I
1	Interest lost								
2									
3	Summary measures for selected variable								
4		Lost1	Lost2	Lost3					
5	Sum	\$ 122.68	\$ 338.65	\$ 287.25					
6									
7	Annual Interest Rate		12%						
8									
9	Days1	Amount1	Lost1	Days2	Amount2	Lost2	Days3	Amount3	Lost3
10	7	\$180	\$0.41	17	\$470	\$2.63	19	\$1,330	\$8.31
11	8	\$210	\$0.55	22	\$540	\$3.91	20	\$1,400	\$9.21
12	10	\$210	\$0.69	28	\$56	\$5.16	14	\$1,550	\$7.13
13	8	\$150	\$0.39	24	\$470	\$3.71	15	\$1,460	\$7.20
14	9	\$300	\$0.89	26	\$650	\$5.56	23	\$2,030	\$15.35
15	5	\$240	\$0.39	29	\$530	\$5.05	19	\$1,520	\$9.49
16	4	\$330	\$0.43	21	\$550	\$3.80	15	\$1,330	\$6.56
17	10	\$290	\$0.95	33	\$620	\$6.73	17	\$1,520	\$8.50
18	5	\$240	\$0.39	16	\$430	\$2.26	21	\$1,390	\$9.60
19	13	\$270	\$1.15	27	\$460	\$4.08	24	\$1,590	\$12.55

This is the amount owed multiplied by the number of days owed multiplied by the interest rate, divided by the n number of days in a year. Then we calculate sums of these amounts in row 5. Although Spring Mills is losing more per customer from the large customers, it is losing more in *total* from the medium-sized customers-because there are more of them. This is shown graphically in Figure 3.31 (page 102) by a pie chart of the sums in row 5. This pie chart shows, for example, that 46% of the lost interest is due to the medium-sized customers.

If Spring Mills really wants to decrease its receivables, it might want to target the medium-sized customer group, from which it is losing the most interest. Or it could target depends on the cost effectiveness of targeting any particular customer group. However, the analysis presented here gives the company a much better picture of what's currently going on.

Figure 3.31 Pie chart of Lost interest by Customer Size

Interest Lost by Customer Type

