**Chapter 3**

**Descriptive Statistics: Numerical Measures**

**Case Problem 1: Pelican Stores**

1. Descriptive statistics for all customers are shown followed by the same descriptive statistics for 4 subgroups of customers.

Net Sales (All Customers)

|  |  |
| --- | --- |
| Mean | $77.60 |
| Median | $59.71 |
| Std. Dev. | $55.66 |
| Range | $274.36 |
| Skewness | 1.715 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **NET SALES BY CUSTOMER TYPE** | | | |
|  | **Married** | **Single** | **Regular** | **Promotion** |
| Mean | $78.03 | $77.04 | $61.99 | $85.25 |
| Median | 59.00 | 69.00 | 51.00 | 63.64 |
| Std. Deviation | 57.67 | 46.21 | 35.07 | 61.38 |
| Range | 274.36 | 163.30 | 137.25 | 274.36 |
| Skewness | 1.732 | 1.254 | 1.351 | 1.520 |

A few observations can be made:

a. Customers taking advantage of the promotional coupons spent more money on average. The mean amount spent by all customers is $77.60; the average amount spent by promotional customers was $85.25.

b. The standard deviation of sales is $55.66. This indicates a fairly wide variability in purchase amounts across customers. This variability is quite a bit smaller for the regular customers.

c. The distribution of the sales data is skewed to the right. The mean ($77.60) is larger than the median ($59.71) and the skewness measure (1.715) is positive. Positive skewness is typical for this kind of data. There are no negative sales amounts and there are a few large purchases.

There are many other descriptive statistics students may generate using the other variables. These will lead to other observations concerning the demographics of the Pelican customers and their buying behavior. For example, the following crosstabulation shows data for the 70 female customers classified by type of customer and marital status.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Gender | Marital Status |  |  |
|  |  | Female |  | Female Total | Grand Total |
| Type of Customer | Data | Married | Single |  |  |
| Promotional | Average of Age | 44 | 33 | 43 | 43 |
|  | Average of Net Sales | 86.48 | 75.96 | 85.20 | 85.20 |
|  | Count of Customer | 58 | 8 | 66 | 66 |
| Regular | Average of Age | 44 | 42 | 44 | 44 |
|  | Average of Net Sales | 58.81 | 89.50 | 64.49 | 64.49 |
|  | Count of Customer | 22 | 5 | 27 | 27 |
| Total Average of Age | | 44 | 36 | 43 | 43 |
| Total Average of Net Sales | | 79 | 81 | 79 | 79 |
| Total Count of Customer | | 80 | 13 | 93 | 93 |

We see that for the 58 female-married promotional customers the average net sales was $86.48, and that for the 8 female-single promotional customers the average net sales was $75.96. Thus, for the promotional customers the average net sales are greater for the married female customers. Note, however, that this effect is just the opposite for the regular customers. For the female-married promotional customers the average net sales is also much greater than the average net sales for the female-married regular customers.

2. The correlation coefficient for the association of sales with age is *r* = .01. There does not appear to be any relationship between net sales and age.

**Case Problem 2: Motion Picture Industry**

This case provides the student with the opportunity to use numerical measures to continue the analysis of the motion picture industry data first presented in Chapter 2. Developing and interpreting descriptive statistics such as the mean, median, standard deviation and range are emphasized. Five-number summaries and the identification of outliers are also of interest. Interpretations and insights can vary. We illustrate some below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Opening Gross** | **Total Gross** | **Number of Theaters** | **Weeks in Top 60** |
| Mean | 9.37 | 33.04 | 1278 | 8.68 |
| Standard Error | 1.89 | 6.32 | 137.87 | 0.64 |
| Median | 0.39 | 5.85 | 410 | 7 |
| Mode | 0.04 | #N/A | 202 | 1 |
| Standard Deviation | 18.87 | 63.16 | 1378.69 | 6.39 |
| Sample Variance | 356.25 | 3989.78 | 1900784.58 | 40.83 |
| Kurtosis | 13.81 | 12.32 | -1.35 | -0.42 |
| Skewness | 3.43 | 3.28 | 0.56 | 0.67 |
| Range | 108.43 | 380.15 | 3905 | 26 |
| Minimum | 0.01 | 0.03 | 5 | 1 |
| Maximum | 108.44 | 380.18 | 3910 | 27 |
| Sum | 937.43 | 3303.84 | 127794 | 868 |
| Count | 100 | 100 | 100 | 100 |

**Five-Number Summary**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Opening Gross** | **Total Gross** | **Number of Theaters** | **Weeks in Top 60** |
| Minimum | 0.01 | 0.03 | 5 | 1 |
| First Quartile | 0.06 | 0.40 | 46 | 3 |
| Median | 0.39 | 5.85 | 410.00 | 7 |
| Third Quartile | 12.43 | 47.43 | 2627 | 13 |
| Maximum | 108.44 | 380.18 | 3910 | 27 |

**Interpretation**

**Opening Weekend Gross Sales.** The mean opening weekend gross sales is $9.37 million. The five-number summary is .01, .06, .39, 12.43 and 108.44. Thus the opening weekend gross sales is highly variable and ranges from a low of $10,000 to a high of $108.44 million. 50% of the motion pictures had an opening weekend gross sales of $390,000 or less, and 25% had a relatively low opening weekend gross sales of $60,000 or less. The top 25% of the motion pictures had an opening weekend gross sales of $12.43 million or more.

**Total Gross Sales.** The mean total gross sales is $33.04 million. The five-number summary is .03, .40, 5.85, 47.43 and 380.18. Thus the total gross sales is also highly variable and ranges from a low of $30,000 to a high of $380.18 million. 50% of the motion pictures had a total gross sales of $5.85 million or less, and 25% had a relatively low total gross sales of $400,000 or less. The top 25% of the motion pictures had total gross sales of $47.43 million or more.

**Number of Theaters.** The mean number of theaters for motion pictures is 1278 theaters. The five-number summary is 5, 46, 410, 2627 and 3910. Thus the number of theaters for a motion picture is also highly variable and ranges from a low of 5 theaters to a high of 3910 theaters. 50% of the motion pictures were shown in 410 or fewer theaters. 25% of the motion pictures were shown in 46 or fewer theaters. The top 25% of the motion pictures were shown in 2627 or more theaters.

**Number of Weeks in Top 60.** The mean number of weeks in the top 60 for motion pictures is 8.68 weeks. The five-number summary is 1, 3, 7, 13 and 27. Thus the number of weeks in the top 60 is also highly variable and ranges from a low of 1 week to a high of 27 weeks. 50% of the motion pictures were on the top 60 list for 7 or fewer weeks. 25% of the motion pictures were on the top 60 list for 3 or fewer weeks. The top 25% of the motion pictures were on the top 60 list for 13 or more weeks.

**General Observations**. The data show that there is a wide variation in the performance of motion pictures for the four variables being studied. Motion pictures range from the low gross sales movies shown in relatively few theaters to the highly successful motion pictures with hundreds of millions in gross sales and playing in almost 4000 theaters. The profiles of motion pictures using the means and medians are shown below.

|  |  |  |
| --- | --- | --- |
| **Profile** | **Mean** | **Median** |
| Opening Weekend Gross Sales | $ 9.37 million | $ .39 million |
| Total Gross Sales | $33.04 million | $ 5.85 million |
| Number of Theaters | 1278 | 410 |
| Number of Weeks in Top 60 | 8.68 | 7 |

The relatively few extremely high performance blockbuster motion pictures tend to inflate the mean in the above profile calculations. The profile based the median gives a better picture of the middle or more typical performance characteristics in the motion picture industry.

**Outliers**

We will use outliers to identify the highly successful blockbuster motion pictures in the data set. Using Q3 + 1.5(IQR) to identify the levels required to qualify as a high performance outlier, we have the following.

Opening Weekend Gross Sales

Q3 + 1.5(IQR) = 12.43+ 1.5(12.43 - .06) = $31 million

Total Gross Sales

Q3 + 1.5(IQR) = 47.43 + 1.5(47.43 - .40) = $118 million

Number of Theaters

Q3 + 1.5(IQR) = 2627 + 1.5(2627 - 46) = 6499 theaters

Number of Weeks on the Top 60 List

Q3 + 1.5(IQR) = 13 + 1.5(13 - 3) = 28 weeks

There are no outliers in terms of the number of theaters or the number of weeks on the top 60 list. There were motion pictures that were high on these two variables, but not high enough to be considered outliers.

However, there were six motion pictures that outperformed the other motion pictures in terms of reaching outlier levels in opening weekend gross sales and total gross sales. These motions pictures are considered the “blockbuster” motion pictures in the data set. To be in this category the motion picture had an opening weekend gross sales greater than $31 million and a total gross sales greater than $118 million. The six blockbuster motion pictures in this category ranked by total gross sales are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Motion Picture** | **Opening Gross Sales**  **($ millions)** | **Total**  **Gross Sales**  **($ millions)** | **Number of Theaters** | **Weeks in Top 60** |
| Star Wars: Episode III | 108.44 | 380.18 | 3,663 | 19 |
| Harry Potter and the Goblet of Fire | 102.69 | 287.18 | 3,858 | 13 |
| War of the Worlds | 77.06 | 234.21 | 3,910 | 19 |
| Wedding Crashers | 33.90 | 209.22 | 3,131 | 23 |
| Batman Begins | 48.75 | 205.28 | 3,858 | 18 |
| Mr. and Mrs. Smith | 50.34 | 186.22 | 3,451 | 21 |

**Correlation**

We also computed the sample correlation coefficient between total gross sales and each of the other three variables. Positive correlations were shown for all three relationships.

Total gross sales and opening weekend gross sales + .96

Total gross sales and number of theaters + .71

Total gross sales and number of weeks in top 60 + .53

The fact that the sample correlation coefficients are positive is to be expected. The motion pictures with the highest total gross sales generally have higher opening weekend gross sales, are shown in more theaters, and have a higher number of weeks in the top 60. The best predictor of total gross sales is the opening weekend gross sales with a sample correlation coefficient of + .96.

**Case Problem 3: Heavenly Chocolates Website Transactions**

1. Descriptive statistics for the time spent on the website, number of pages viewed, and amount spent are shown below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Time (min)** | **Pages Viewed** | **Amount Spent ($)** |
| Mean | 12.8 | 4.8 | 68.13 |
| Median | 11.4 | 4.5 | 62.15 |
| Standard Deviation | 6.06 | 2.04 | 32.34 |
| Skewness | 1.45 | .65 | 1.05 |
| Range | 28.6 | 8 | 140.67 |
| Minimum | 4.3 | 2 | 17.84 |
| Maximum | 32.9 | 10 | 158.51 |
| Sum | 640.5 | 241 | 3406.41 |

The mean time a shopper is on the Heavenly Chocolates website is 12.8 minutes, with a minimum time of 4.3 minutes and a maximum time of 32.9 minutes. The fact that the mean time on the website (12.8 minutes) is greater than the median time (11.4 minutes) and the value of skewness is 1.45 indicates that the time on the website is skewed to the right. The following histogram provides further evidence of the skewness in the data.



The mean number of pages viewed during a visit is 4.8 pages with a minimun of 2 pages and a maximum of 10 pages. The fact that the mean number of pages viewed (4.8) is greater than the median (4.5) and the value of skewness is .65 indicates the number of pages viewed is slightly skewed to the right. A histogram of the number of pages viewed provides additonal evidence that the data are slightly skewed to the right.



The mean amount spent for an on-line shopper is $68.13 with a minimum amount spent of $17.84 and a maximum amount spent of $158.51. The fact that the median amount spent ($68.13) is greater than the median amount spent ($62.15) and the value of skewess is 1.05 indicates that the amount spent is skewed to the right. The following histogram provides further evidence of the skewness in the data.



2. Summary by Day of Week

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Day of Week** | **Frequency** | **Total Amount Spent ($)** | **Average Amount Spent ($)** |
|  | Sunday | 5 | 218.15 | 43.63 |
|  | Monday | 9 | 813.38 | 90.38 |
|  | Tuesday | 7 | 414.86 | 59.27 |
|  | Wednesday | 6 | 341.82 | 56.97 |
|  | Thursday | 5 | 294.03 | 58.81 |
|  | Friday | 11 | 945.43 | 85.95 |
|  | Saturday | 7 | 378.74 | 54.11 |
|  | Total | 50 | 3406.41 | 68.13 |

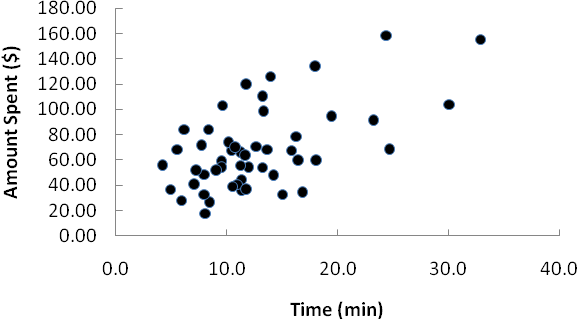
The above summary shows that Monday and Friday are the best days in terms of both the total amount spent and the averge amount spent per transaction. Friday had the most purchases (11) and the highest value for total amount spent ($945.43). Monday, with nine transactions, had the highest average amount spent per transaction ($90.38). Sunday was the worst sales day of the week in terms of number of transactions (5), total amount spent ($218.15), and average amount spent per transaction ($43.63). However, the sample size for each day of the week are very small, with only Friday having more than ten transactions. We would suggest a larger sample size be taken before recommending any specific stratgegy based on the day of week statistics.

3. Summary by Type of Browser

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Browser** | **Frequency** | **Total Amount Spent ($)** | **Average Amount Spent ($)** |
|  | Firefox | 16 | 1228.21 | 76.76 |
|  | Internet Explorer | 27 | 1656.81 | 61.36 |
|  | Other | 7 | 521.39 | 74.48 |

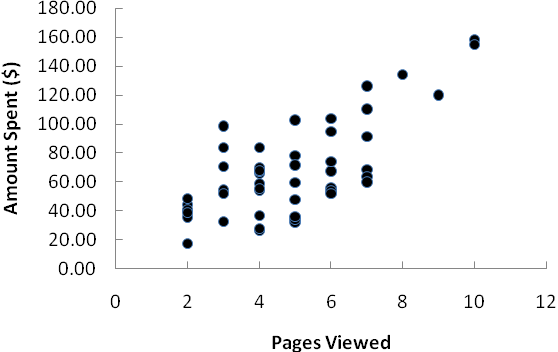
Internet Explorer was used by 27 of the 50 shoppers (54%). But, the average amount spent spent by customers who used Internet Explorer ($61.36) is less than the average amount spent by customers who used Firefox ($76.76) or some other type of browser ($74.48). This result would suggest targeting special promotion offers to Firefox users or users of other types of browsers. But, before recommending any specific strategies based upon the type of browser, we would suggest taking a larger smaple size.

4. A scatter diagram showing the relationship between time spent on the website and the amount spent follows:



The sample correlation coefficient between these two variables is .580. The scatter diagram and the sample correlation coefficient indicate a postive relationship between time spent on the website and the total amount spent. Thus, the sample data support the conclusion that customers who spend more time on the website spend more.

5. A scatter diagram showing the relationship between the number of pages viewed and the amount spent follows:



The sample correlation coefficient between these two variables is .724. The scatter diagram and the sample correlation coefficient indicate a postive relationship between time spent on the website and the number of pages viewed. Thus, the sample data support the conclusion that customers who view more website pages spend more.

5. A scatter diagram showing the relationship between the number of pages viewed and the time spent on the website follows:



The sample correlation coefficient between these two variables is .596. The scatter diagram and the sample correlation coefficient indicate a postive relationship between the number of pages viewed and the time spent on the website.

**Summary**: The analysis indicates that on-line shoppers who spend more time on the company’s website and/or view more website pages spend more money during their visit to the website. If Heavenly Chocolates can develop an attractive website such that on-line shoppers are willing to spend more time on the website and/or view more pages, there is a good possiblity that the company will experience greater sales. And, consideration should also be given to developing marketing strategies based upon possible differences in sales associated with the day of the week as well as differences in sales associated with the type of browser used by the customer.