**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: The 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.

**Descriptive Statistics**

Descriptive Statistics provided by Minitab follow:

Variable N Mean SE Mean StDev Minimum Q1

Opening Gross Sales ($mi 100 27.51 2.65 26.52 0.07 12.97

Total Gross Sales $milli 100 90.47 6.81 68.12 29.14 39.35

Number of Theaters 100 3114.3 61.1 610.8 1038.0 2849.3

Weeks in Release 100 14.58 0.505 5.050 6.000 11.250

Variable Median Q3 Maximum

Opening Gross Sales ($mi 19.08 32.06 169.19

Total Gross Sales $milli 72.40 107.07 381.01

Number of Theaters 3102.5 3553.3 4375.0

Weeks in Release 14.500 17.000 43.000

**Interpretation**

**Opening Weekend Gross Sales.** The mean opening weekend gross sales is $27.51 million. The five-number summary is .07, 12.97, 19.08, 32.06, and 169.19. Thus the opening weekend gross sales is highly variable and ranges from a low of $70,000 to a high of $169.19 million. 50% of the motion pictures had an opening weekend gross sales of $19.08 million or less, and 25% had a relatively low opening weekend gross sales of $12.97 million or less. The top 25% of the motion pictures had an opening weekend gross sales of $32.06 million or more.

**Total Gross Sales.** The mean total gross sales is $90.47 million. The five-number summary is 29.14, 39.35, 72.40, 107.07, and 381.01. Thus the total gross sales is also highly variable and ranges from a low of $29.14 million to a high of $381.01 million. 50% of the motion pictures had a total gross sales of $72.40 million or less, and 25% had a relatively low total gross sales of $39.35 million or less. The top 25% of the motion pictures had total gross sales of $107.07 million or more.

**Number of Theaters.** The mean number of theaters is 3114.3. The five-number summary is 1038, 2849.3, 3102.5, 3553.3, and 4375. Thus the number of theaters for a motion picture is also highly variable and ranges from a low of 1038 theaters to a high of 4375 theaters. 50% of the motion pictures were shown in 3103 (3102.5) or fewer theaters. 25% of the motion pictures were shown in 2850 (2849.3) or fewer theaters. The top 25% of the motion pictures were shown in 3554 (3553.3) or more theaters.

**Number of Weeks in Release.** The mean number of weeks in release for motion pictures is 14.58 weeks. The five-number summary is 6, 11.25, 14.5, 17, and 43. Thus the number of weeks in release is also highly variable and ranges from a low of 6 weeks to a high of 43 weeks. 50% of the motion pictures were in release for 15 (14.5) or fewer weeks. 25% of the motion pictures were in release for 12 (11.25) or fewer weeks. The top 25% of the motion pictures were in release for 17 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 over 4000 theaters. The profiles of motion pictures using the means and medians are shown below.

|  |  |  |
| --- | --- | --- |
| **Profile** | **Mean** | **Median** |
| Opening Weekend Gross Sales | $27.51 million | $19.08 million |
| Total Gross Sales | $90.47 million | $72.4 million |
| Number of Theaters | 3114.3 | 3102.5 |
| Number of Weeks in Release | 14.58 | 14.5 |

The relatively few extremely high performance blockbuster motion pictures tend to inflate the mean in the above financial profile calculations. The profile based the median gives a better picture of the middle or more typical financial 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) = 32.6 + 1.5(32.6 – 12.97) = $62.045 million

Total Gross Sales

Q3 + 1.5(IQR) = 107.07 + 1.5(107.07 - 39.35) = $208.65 million

Number of Theaters

Q3 + 1.5(IQR) = 3553.3 + 1.5(3553.3- 2849.3) = 4609.3 theaters

Number of Weeks in Release

Q3 + 1.5(IQR) = 17 + 1.5(17 – 11.25) = 25.625 weeks

There are two no outliers in terms of the number of theaters. There were motion pictures that were high on this variable, but not high enough to be considered outliers.

There are two outliers in terms of the number of weeks in release. They are *Midnight in Paris* and *The Help*.

However, there were six motion pictures that outperformed the other motion pictures in terms of reaching outlier levels in both 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 $62.045 million and a total gross sales greater than $208.65 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 Release** |
| *The Hangover Part II* | 85.95 | 254.46 | 3,675 | 16 |
| *Fast Five* | 86.2 | 209.84 | 3,793 | 15 |
| *Pirates of the Caribbean: On Stranger Tides* | 90.15 | 241.07 | 4,164 | 19 |
| *Transformers: Dark of the Moon* | 97.85 | 352.39 | 4,088 | 15 |
| *The Twilight Saga: Breaking Dawn Part 1* | 138.12 | 281.29 | 4,066 | 14 |
| *Harry Potter and the Deathly Hallows Part 2* | 169.19 | 381.01 | 4,375 | 19 |

*Harry Potter and the Deathly Hallows Part 2* was the top motion picture in terms of both opening weekend gross sales and total gross sales for 2011. It was also shown in the most theaters (4375). It is interesting to note that the motion picture in the data with the longest run, and *Midnight in Paris* at 23 weeks, was not among our list of blockbuster motion pictures.

**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 + .887

Total gross sales and number of theaters + .641

Total gross sales and number of weeks in Release + .333

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 release. The best predictor of total gross sales is the opening weekend gross sales with a sample correlation coefficient of + .887.

**Case Problem 3: Business Schools of Asia-Pacific**

The Minitab output is shown below.

Variable N Mean Median Tr Mean StDev SE Mean

FT Enrol 25 165.2 126.0 158.9 140.8 28.2

Std/Facu 25 8.48 7.00 8.30 5.06 1.01

Local$ 25 12375 11513 11970 7778 1556

Foreign$ 25 16582 17765 16543 9135 1827

Age 25 28.360 29.000 28.261 3.785 0.757

%Foreign 25 28.08 27.00 26.61 25.01 5.00

GMAT 25 0.560 1.000 0.565 0.507 0.101

English 25 0.3200 0.0000 0.3043 0.4761 0.0952

Work 25 0.7600 1.0000 0.7826 0.4359 0.0872

Salary 25 37292 41400 36448 23459 4692

Variable Min Max Q1 Q3

FT Enrol 12.0 463.0 43.0 270.0

Std/Facu 2.00 19.00 5.00 13.50

Local$ 1000 33060 5223 17263

Foreign$ 1000 33060 8750 22650

Age 22.000 37.000 25.000 31.000

%Foreign 0.00 90.00 4.75 45.00

GMAT 0.000 1.000 0.000 1.000

English 0.0000 1.0000 0.0000 1.0000

Work 0.0000 1.0000 0.5000 1.0000

Salary 7000 87000 14550 53750

Many graphical and numerical summaries and interpretations are possible with this data set. The students should be encouraged to investigate the data set with minimum guidance and direction. A comparison of student reports usually shows interesting and unique insights provided by the descriptive statistics.

Some of the results are summarized here.

a. Full-time enrollment varies considerably. The Indian Institute of Management (463) and the Macquaire Graduate School of Management in Sydney (12) provide the extremes. The mean full-time enrollment is 165 students.

b. The mean number of students per faculty member is 8.5. This should be surprisingly low to most students. The schools appear to provide a good potential for student access to faculty.

c. The mean local tuition cost is $12,375

d. The mean foreign tuition cost is $16,582

e. The mean age is 28.36 years and median age is 29. This indicates a substantial number of older students are in the programs. Apparently, a good number of the students have given up ongoing careers to return to school.

f. The mean percentage of foreign students is 28%, with four schools having 1% or 0% foreign. Schools with a high percentage of foreign students include Australian National University (80%) and Asian Institute of Management in Bangkok (90%).

g. About half of the schools (56%) require the GMAT test for admission to the school.

h. English tests are only required in 32% of the schools.

i. Work experience appears to be an important factor in gaining admission to Asia-Pacific business schools, with 76% requiring work experience. This is consistent with item e where the mean age was noted to be 28.36 years.

j. The mean starting salary is $37,292 and the median starting salary is $41,400. Starting salary ranges from only $7,000 for four schools to a high of $87,000 at the International University of Japan.

k. As probably expected, the mean tuition is greater for foreign students. The mean difference is $16,582 - $12,375 = $4,207 per year.

l. The mean starting salary is greater for schools that require work experience. This appears reasonable with means of $41,305 for the 19 schools requiring work experience and $24,583 for the 6 schools not requiring work experience.

m. The mean starting salary is greater for schools that require English tests. The mean is $45,088 for the 8 schools requiring English tests and the mean is $33,624 for schools not requiring English Tests.

**Case Problem 4: Heavenly Chocolates Website Traffic**

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 |

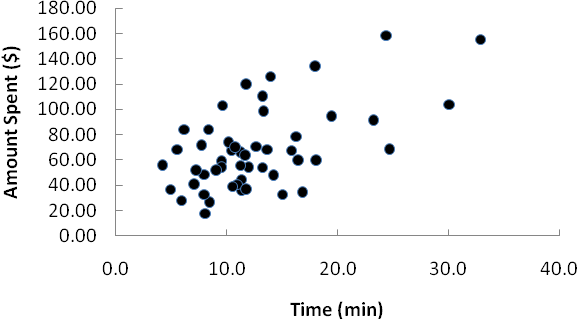
The above summary shows that Monday and Friday are the best days in terms of both the total amount spent and the averge amount spent per transaction. Friday had the most purchases (11) and the highest value for total amount spent ($945.43). Monday, with nine transactions, had the highest average amount 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 |

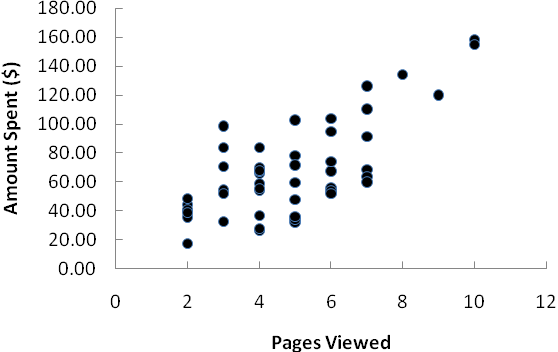
Internet Explorer was used by 27 of the 50 shoppers (54%). But, the average amount spent spent by customers who used Internet Explorer ($61.36) is less than the average amount spent by customers who used Firefox ($76.76) or some other type of browser ($74.48). This result would suggest targeting special 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:



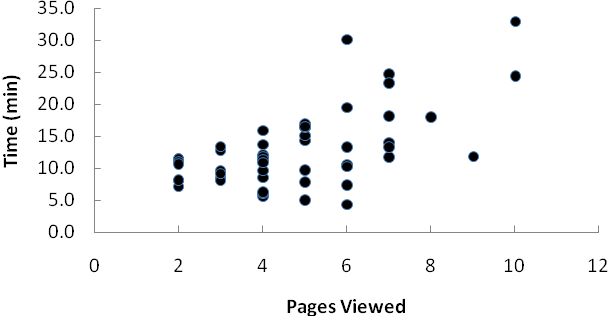
The sample correlation coefficient between these two variables is .580. The scatter diagram and the sample correlation coefficient indicate a postive relationship between time spent on the website and the total amount spent. Thus, the sample data support the conclusion that customers who spend more time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

6. 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.

**Case Problem 5 African Elephant Populations**

This case provides the student with the opportunity to use the geometric mean in conjunction with a graph (such as the boxplot) to analyze changes over time in the populations of elephants in several African nations.

1. Let’s calculate the proportional change for each country over the ten year period 1979-1989. We’ll begin by considering the Central African Republic. We have:

19000=63000, so =0.301587 and



So the mean annual change in the elephant population for the Central African Republic during this period is (0.887036 – 1)100 = -11.3%. During the period of 1979-1989, the elephant population in the Central African Republic declined at an annual rate of 11.3%.

Repeating these calculations for each nation yields the values in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| Country |  |  | Mean Annual Change |
| Angola | 1.0000 | 1.0000 | 0.0000 |
| Botswana | 2.5500 | 1.0981 | 0.0981 |
| Cameroon | 1.3086 | 1.0273 | 0.0273 |
| Cen African Rep | 0.3016 | 0.8870 | -0.1130 |
| Chad | 0.2067 | 0.8541 | -0.1459 |
| Congo | 6.4815 | 1.2055 | 0.2055 |
| Dem Rep of Congo | 0.2250 | 0.8614 | -0.1386 |
| Gabon | 5.6716 | 1.1895 | 0.1895 |
| Kenya | 0.2923 | 0.8843 | -0.1157 |
| Mozambique | 0.3394 | 0.8976 | -0.1024 |
| Somalia | 0.2469 | 0.8695 | -0.1305 |
| Sudan | 0.0299 | 0.7039 | -0.2961 |
| Tanzania | 0.2529 | 0.8716 | -0.1284 |
| Zambia | 0.2733 | 0.8784 | -0.1216 |
| Zimbabwe | 1.4333 | 1.0367 | 0.0367 |

The elephant populations in several nations (Central African Republic, Chad, Democratic Republic of the Congo, Kenya, Mozambique, Somalia, Sudan, Tanzania, and Zambia) declined at an annual rate of 10% or more from 1979-1989, with Sudan losing almost 30% of its elephant population annually. Sudan’s elephant population has almost entirely disappeared. During the same period a few nations (Botswana, Congo, and Gabon) experienced growth in their elephant populations.

2. Now let’s calculate the proportional change for each country over the ten year period 1989-2007. We’ll again begin by considering the Central African Republic. We have:

3334=19000, so =0.175474 and



So the mean annual change in the elephant population for the Central African Republic during this period is (0.907845 – 1)100 = -9.2155%. During the period of 1979-1989, the elephant population in the Central African Republic declined at an annual rate of 9.2%.

Repeating these calculations for each nation yields the values in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| Country |  |  | Mean Annual Change |
| Angola | 0.2040 | 0.9155 | -0.0845 |
| Botswana | 3.4409 | 1.0711 | 0.0711 |
| Cameroon | 0.7258 | 0.9824 | -0.0176 |
| Cen African Rep | 0.1755 | 0.9078 | -0.0922 |
| Chad | 2.0758 | 1.0414 | 0.0414 |
| Congo | 0.3157 | 0.9380 | -0.0620 |
| Dem Rep of Congo | 0.2790 | 0.9315 | -0.0685 |
| Gabon | 0.9294 | 0.9959 | -0.0041 |
| Kenya | 1.6651 | 1.0287 | 0.0287 |
| Mozambique | 1.4026 | 1.0190 | 0.0190 |
| Somalia | 0.0117 | 0.7809 | -0.2191 |
| Sudan | 0.0750 | 0.8660 | -0.1340 |
| Tanzania | 2.0875 | 1.0417 | 0.0417 |
| Zambia | 0.7130 | 0.9814 | -0.0186 |
| Zimbabwe | 2.3048 | 1.0475 | 0.0475 |

Only two countries (Somalia and Sudan) continue to experience average annual declines in their elephant populations of 10% or more from 1989-2007, while the elephant populations in most other nations had relatively small mean annual changes in their elephant populations.

3. Now we compare the results of our two analyses and draw conclusions.

|  |  |  |
| --- | --- | --- |
| **Country** | **Mean Annual Change 1979-1989** | **Mean Annual Change**  **1989-2007** |
| Dem Rep of Congo | -0.1386 | -0.0685 |
| Tanzania | -0.1284 | 0.0417 |
| Zambia | -0.1216 | -0.0186 |
| Sudan | -0.2961 | -0.1340 |
| Kenya | -0.1157 | 0.0287 |
| Cen African Rep | -0.1130 | -0.0922 |
| Mozambique | -0.1024 | 0.0190 |
| Zimbabwe | 0.0367 | 0.0475 |
| Somalia | -0.1305 | -0.2191 |
| Botswana | 0.0981 | 0.0711 |
| Cameroon | 0.0273 | -0.0176 |
| Chad | -0.1459 | 0.0414 |
| Gabon | 0.1895 | -0.0041 |
| Angola | 0.0000 | -0.0845 |
| Congo | 0.2055 | -0.0620 |

We can use a set of boxplots to support this analysis. We can see from these boxplots that the population of elephants declined dramatically from 1979 to 1989, and have generally started to come back between 1989 and 2007. We can also see that the declining trend that was established between 1979 and 1989 continues for the elephant populations in some African nations.



Several nations appear to have reversed the declines in elephant populations they experienced from 1979-1989, but the growth rates are still generally low (and in some countries still negative). At 1989-2007 rates of change, it will take many decades for the elephant populations to recover to their 1979 levels.