

## Lab Activity: Descriptive Statistics

In this lab activity, you will create graphs and compute descriptive statistics for a built-in dataset in Statcato.

### Student Learning Outcomes

By the end of this chapter, you should be able to do the following:

- Create graphs for data in Statcato (stem-and-leaf plot, histogram, box plot, pie chart)
- Compute descriptive statistics in Statcato (quartiles, interquartile range, mean, median, mode, skewness, standard deviation, variance)
- Interpret data using graphs and descriptive statistics

### Preliminary

**Read** Chapter 2 Descriptive Statistics in:

Illowsky, Barbara, and Susan Dean. Collaborative Statistics. Connexions. 2 Mar. 2010 <<http://cnx.org/content/col10522/1.37/>>.

Make sure you understand the following **key terms** (LR:Key Terms):


outlier, median, quartile, percentile, interquartile range, mean, mode, skewness, standard deviation

### Loading Data

This lab uses a dataset built into Statcato (*Rainfall Totals for Seeded and Unseeded Clouds*, data from Miller, A.J., Shaw, D.E., Veitch, L.G. & Smith, E.J. “Analyzing the results of a cloud-seeding experiment in Tasmania”, *Communications in Statistics – Theory & Methods*, vol.A8(10), pp. 1017-1047).



Here you will load the built-in dataset.

- Go to **File > Load Dataset**. Or click the  icon in the toolbar.
- Under Built-in Datasets, select the **Choose a sample dataset** radio button. Select “Rainfall Totals for Seeded and Unseeded Clouds” in the drop-down menu.
- Click **Load Dataset**.

The rainfall data should now be in column C1.

### Creating Graphs


You will now construct a stem-and-leaf plot, histogram, box plot, and pie chart for the data.



#### Stem-and-Leaf Plot

Go to **Graph > Stem-and-Leaf Plot**.

- For **Graph Variables**, choose **C1** in list box.
- Leave the default options in the **Graph Options** panel.
- Click **OK**.

The stem-and-leaf plot should now be displayed as a table in the Log window. Highlight the Stem-and-Leaf Plot results in the Log, and then copy (**Edit > Copy**, **Ctrl+C**, or click ). Paste the results to the provided space in **LR: Graphs**.

### Histogram

Go to **Graph > Histogram**.

- For **Graph Variables**, choose **C1** in list box. Choose **Frequency** under **Heights of bars represent**.
- For **X-axis**
  - Choose the option **Provide the class width and the minimum**.
    - Class width: 1
    - Minimum: 0
  - For **Label**, enter the word *rainfall*.
- In **Y-axis** panel, enter the word *frequency* for **Label**.
- In **Other Options** panel, enter *Histogram for Rainfall* in the **Title** text box.
- Click **OK**.

The generated histogram should be shown in a separate window. You can copy the graph by going to **Graph > Copy Graph to Clipboard** and then paste it the provided space to **LR: Graphs**.

### Box Plot

Go to **Graph > Box Plot**.

- For **Graph Variables**, choose **C1** in list box.
- In **Graph Options** panel, enter *Box Plot for Rainfall* in the **Plot Title** text box.
- Click **OK**.

The generated box plot should be shown in a separate window. Hover your mouse over the dot in the center of the box plot to see the five-number summary. You can also hover your mouse over the dot/triangle outside of the box and whiskers to any outliers. Copy and paste the box plot to **LR: Graphs**.

### Pie Chart

Go to **Calculate > Frequency Table**.

- For **Source Data**, select **C1** in the drop-down menu.
- In the **Computing Frequencies** panel, select **Source data contains numerical data**.
  - First upper class limit: 1
  - Class width: 1
- For **Results**, under **Store Frequency Table in**
  - Category column: c5
  - Frequency column: c6
- Click **OK**.

Now you should see the upper class limits of classes of width 1 and the corresponding class frequencies in C5 and C6, respectively.

Go to **Graph > Pie Chart**.

- In **Graph Variables** panel, choose **Summary Data from Dataheet**.
  - Categories: C5
  - Frequencies: C6
- In **Graph Options** panel, enter *Pie Chart for Rainfall* in the **Plot Title** text box. Leave the default options.
- Click **OK**.

The generated pie chart should be shown in a separate window. Copy and paste the pie chart to **LR: Graphs**.

## Computing Descriptive Statistics



Go to **Statistics > Basic Statistics > Descriptive Statistics**.

- In the **Input Variable(s):** text box, enter **C1**.
- In the **Statistics** panel, select the following statistics: Mean, Standard deviation, Variance, Minimum, Maximum, First quartile, Median, Third quartile, Interquartile range, Mode, Percentile, Skewness (Biased). Enter **10** in the text box for Percentile.
- Click **OK**.

The selected descriptive statistics should now be displayed in the Log. Copy them to **LR: Descriptive Statistics**.

## Discussion

Answer the following questions in **LR: Discussion**.

1. Are the data discrete or continuous? How do you know?
2. In general, when should the following graphs be used?

- a. stem-and-leaf graph
  - b. histogram
  - c. box plot
  - d. pie chart
3. What is the shape of histogram? Is it symmetrical or skewed?
  4. A positive skewness indicates that the data is skewed to the right, and a negative skewness indicates that the data is skewed to the left. Based on the skewness, is the data used in this lab skewed to the right or left? Does it match what you see in the histogram?
  5. Looking at your box plot, does it appear that the data are concentrated in some areas or spread out evenly? Explain.
  6. Suppose you want to show that a particular class of values in your data has the highest proportion of rainfall. Which of the four graphs would you use and why?
  7. Are there any potential outliers in the data? How do you determine the potential outliers in the data using some of the descriptive statistics computed in the lab?
  8. Which is a better measure of the center of this data, mean or median?