CHAPTER 1

Appendix

Getting Started with Statistical Computing

Most statistical analyses rely heavily on statistical software. In this appendix, we discuss the use of Excel 365, JMP 14, Minitab 19, SPSS 26, CrunchIt, R, and a TI-83/-84 calculator for conducting statistical analysis. As specialized statistical packages, JMP, Minitab, and SPSS are the most popular software choices both in industry and in colleges and schools of business. R is an extremely powerful statistical environment that is free to anyone; it relies heavily on members of the academic and general statistical communities for support. As an all-purpose spreadsheet program, Excel provides a limited set of statistical analysis options in comparison. However, given its pervasiveness and wide acceptance in industry and the computer world at large, we believe it is important to give Excel proper attention. It should be noted that for users who want more statistical capabilities but want to work in an Excel environment, there are a number of commercially available add-on packages (if you have JMP, for instance, it can be invoked from within Excel). Finally, instructions are provided for the TI-83/-84 calculators. While generally sufficient for an introductory course, most statistical analysis is beyond the capabilities of even the best calculator, so those seeking to continue their learning of statistics should consider learning one of the specialized statistical packages.

Even though basic guidance is provided in this and subsequent appendices, it should be emphasized that PSBE is not bound to any of these programs. Computer output from statistical packages is very similar, so you can feel quite comfortable using any one these packages.

File Naming Conventions

Each program has its own file extensions for saving data worksheets and output. All use the typical interface to open and save (or "save as" to change a name) files from the File menu. The extensions are shown below; to access data files from the CD or website, the naming convention is TTxx-yyyaaaaaa.ext, where "TT" is the type of data set (ca for cases, eg for examples, ex for exercises, or ta for tables); is the chapter number; "zzz" is the number of the exercise, example, or table; and "aaaaa" is a short description of the topic. File extensions depend on the software:

	Data file extension	Output file extension
EXCEL	.xls or .xlsx	.xls or .xlsx Excel embeds output, including graphics, into the worksheet

(*Continued*)

JMP	.jmp	.jmpprj Projects contain all data, reports, and output
MINITAB	.mtw	.mpj Projects contain both data and output
SPSS	.sav	.spv
R	.csv R can read many formats; comma separated is typical	.Rdata Saves the entire workspace

Getting Help

If you encounter a question not answered in this material, most software platforms offer help (both general and contextual) in dialog boxes. To access help topics, click Help in the menu bar at the top of the screen; for contextual help, click Help in a dialog box. Several of these packages (Minitab, JMP, SPSS, and R) also have tutorials available that will help you get started. Click on the "Tutorial" option from the Help pull-down menu.

Convention in These Appendices

In this material, options that are "clicked" appear in bold type. A series of bold options with arrows between them indicates a series of clicks.

Commands that are actually typed (applies mostly to R) are shown in Courier font.

Getting Started

EXCEL

We assume that the reader is familiar with the basic layout and usage of Excel. As noted earlier, Excel provides a number of standard statistical analysis procedures but is not as comprehensive as a stand-alone statistical package. Therefore, for a few topics covered in this book, software support will be found only in a statistical package or in an enhanced add-on version of Excel rather than in standard Excel. Excel is the only software platform with a dynamic worksheet (meaning it updates as data are changed that impact formulas); all the other programs have the capability to compute new columns, but once computed, the data residing there are static.

Built-In Statistical Functions and Charts

Excel has a variety of built-in statistical functions that can be used to compute common descriptive statistics for a given set of data or to compute probabilities for well-known statistical distributions. To find these functions, select the **Formulas** tab found in the main menu. Click **More Functions**, which allows you to select the category Statistical to reveal all the statistical functions.

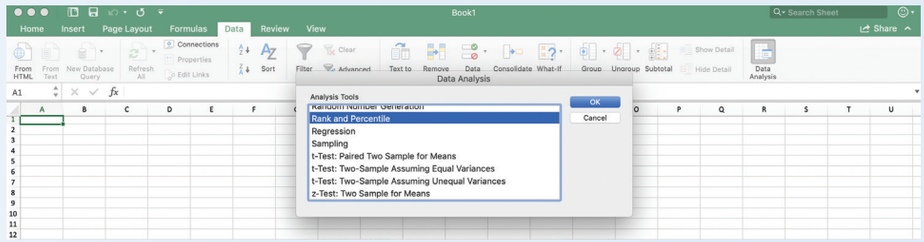
In addition to the built-in statistical functions, a number of graphing options are available that may prove useful for data analysis. The available charts are found by selecting the Insert tab found in the main menu. One then finds a variety of graphing options in the Charts group. A few statistical options (for example, regression fitting) can be implemented within the charts.

Installing Data Analysis ToolPak Add-In

Excel's built-in statistical functions can be useful for isolated computations. However, attempting to do a more complete statistical analysis with a collection of "raw" functions can be a laborious and clumsy process. Excel provides an add-in known as Analysis ToolPak, which enables you to perform a more integrative statistical analysis. This add-in is not loaded with the standard installation of Excel. To install this add-in, click **File ➔ Options ➔ Add-ins** and then, in the Manage box, choose "Excel Add-ins" and click **Go**. Select **Analysis ToolPak** and finally click **OK**.

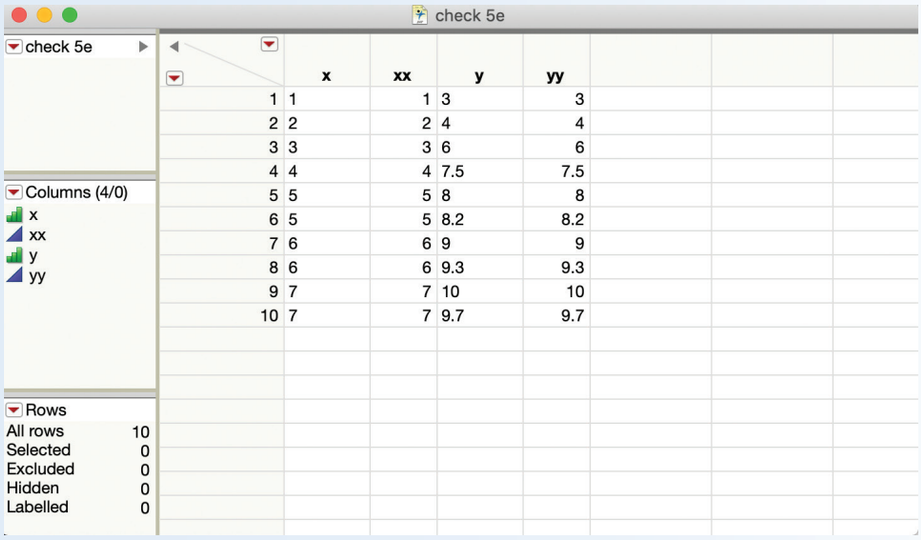
Invoking Data Analysis ToolPak Procedures

Once the Data Analysis ToolPak is installed, the statistical analysis routines are found by first selecting the Data tab found on the main toolbar. You will then see the Data Analysis command in the Analysis group. The following figure shows a blank Excel spreadsheet with the Data Analysis command invoked, resulting in the appearance of the Data Analysis menu box.



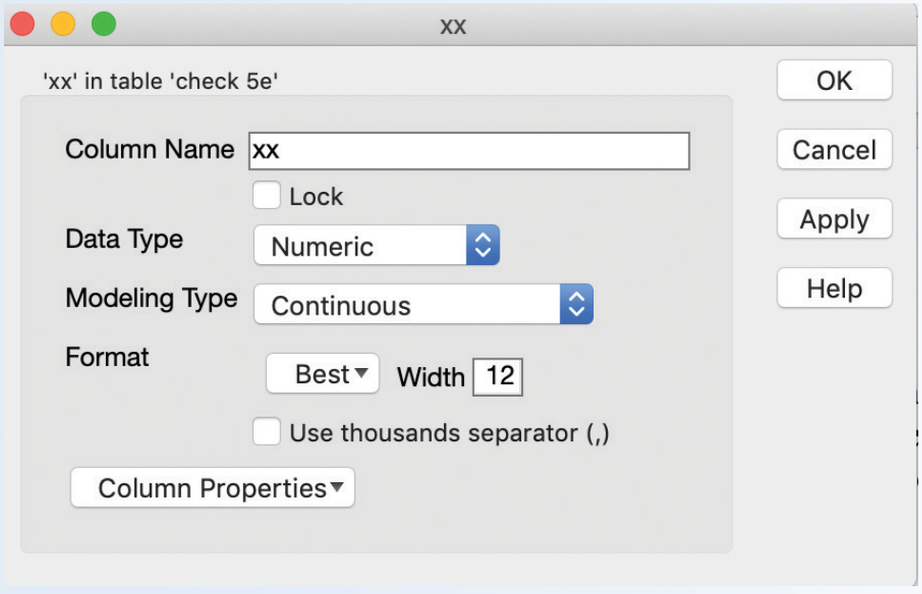
Within the Data Analysis menu box, there are 19 menu choices. When you select one of these, a box specific to the statistical routine will appear that asks for you to indicate where the data reside and where you want the output to be displayed. To indicate where the data for analysis reside, you specify the range of cells for the data in the Input Range box. This can be accomplished by first clicking the cursor in the Input Range box and then typing in the cell range, or, more easily, you can highlight the data by clicking and dragging the mouse over the cell range. The statistical output can be placed either in the current worksheet (placement indicated with Output Range box), in a new worksheet tabbed with the current workbook (New Worksheet Ply option), or in an entirely new workbook (New Workbook option).

Upon entering JMP on either Mac or Windows, you will find the JMP home window, which is partitioned into two sections: recent files and a list of open windows (if you selected to remember them the last time you used the software). Upon opening a data set (illustrated below), a data table will be shown in a separate window.



Modeling Types

Variables in JMP take on a property called "modeling type," which is just a classification for what measurements in a variable mean. For example, the chromosomal sex of an individual (male vs. female) is a different type of measurement than the age of an individual—one is a category, whereas the other is a numeric quantity. In JMP, variables are designated as being Nominal (categories), Ordinal (ordered categories), or Continuous (numeric measurements on a scale, like age). This designation is important because JMP will help you produce analyses and graphical output that is appropriate for the type of variable you have. To change or set the modeling type of a variable, simply double-click on the variable name and select the data and modeling types appropriate for that variable (see figure below).



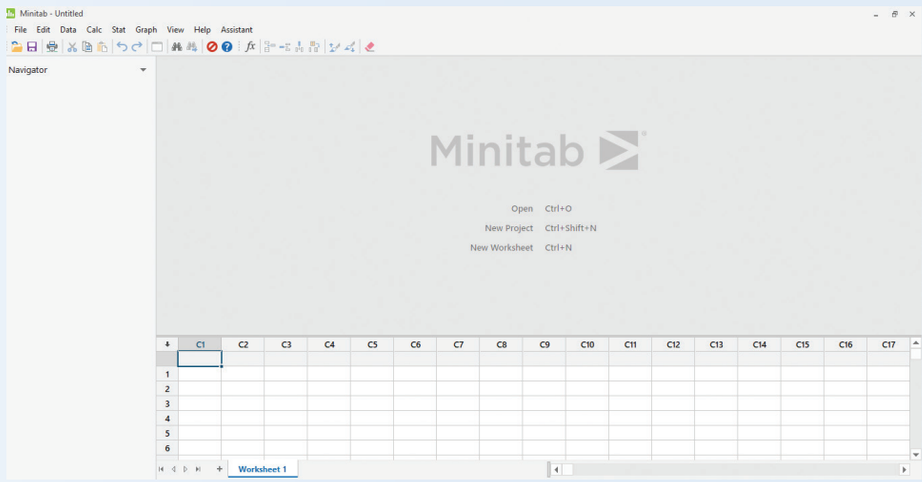
Invoking Statistical Procedures

To produce an analysis or create a graph, users can make a sequence of selections from a series of menus that all begin in the menu bar. In JMP, analyses and graphics are grouped by their context within "platforms." For example, the Fit Y by X platform under the Analyze menu allows users to test hypotheses when there is one Y variable and one X variable (for instance, a two-group *t* test or a simple regression). Which type of analysis is returned depends on the modeling types of the variables specified (as described previously).

Once a platform is launched, additional options are available under the "Red Triangles" in the output window. These Red Triangles are special menus that show contextualized options—that is, analyses and options that make sense for the types of variables specified. In this regard, JMP is said to have a "progressive" interface: launching a platform is the first step, and once in a platform you can produce any number of analyses. If you are looking for a specific analysis, the "Statistics Index," available under the Help menu, provides a list of all available procedures and can even launch an example for a given analysis. If you need additional help, select the question mark tool in the menu and click on any object in JMP to see the documentation for that object.

MINITAB

Minitab version 19 is the first to have both PC and Mac versions. Older versions ran only on PCs. Upon entering Minitab, you will find the display partitioned into three windows, as seen in the following figure. The Session window is the area where all statistical output and Minitab commands generating statistical output (graphical and nongraphical) are displayed. The Data Window displays a spreadsheet environment (known as a worksheet) where the data can be directly entered and edited. Each column represents a variable to be analyzed. The third window the Navigator keeps track of all the analyses that have been done in a project. **You may also see a command line entry window and a command history window if those have been activated on previous uses. You can use the View menu option to hide or see any window. If you have an older version, there will only be two windows: the worksheet and the session.**



Invoking Statistical Procedures

There are two ways to invoke procedures:

1. You can type commands in the Command Line box at upper right. To do so, you must first enable the command language:
   - Click **View → Command Line/History**

   This will open the command line window. Here, you can then type desired commands. If your version is 18 or earlier, the command line is invoked by clicking **Editor → Show Command line**. You will see the MTB> prompt appear either in a separate window or in the session window.
2. Make a sequence of selections beginning in the toolbar menu. For example, in this chapter, we produce a graph known as a boxplot. To create this graph, you would click **Graph** and then select **Boxplot**. In this appendix, such a sequence of selections will be presented as **Graph → Boxplot**. Once the sequence of selections has been made, dialog and/or option boxes will be encountered that allow you to indicate which variable(s) will be part of the analysis, along with other information. If further help is needed, you can click the **Help** button that appears with every dialog box. Once all appropriate information is provided, click the **OK** button to get the desired output.

SPSS

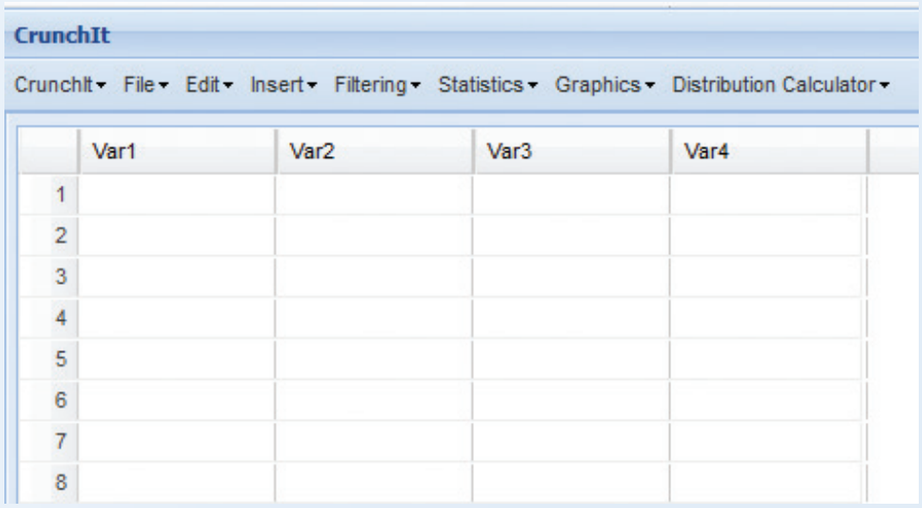
When SPSS first starts, you will see a spreadsheet-type data window. You can use this to type in data of your own or use the File menu (or icon) to open a data file. SPSS can read many types of data files. If your file is not an SPSS .sav file, use the drop-down in the open file dialog to change the "Files of type" to the appropriate one. If not an SPSS file, you will see an intermediate screen asking you to confirm what SPSS thinks columns should be.

SPSS also has the Variable View accessible by a separate tab at the bottom of the spreadsheet window. This screen indicates the variable type (scale, nominal, or ordinal), number of decimal places (if appropriate), and other characteristics of your data. Check this to ensure that the variables are understood to be of the appropriate type. If you have typed in data, SPSS will default to naming the columns VAR0001, VAR0002, etc. Use the Variable View screen to give them actual descriptive names by clicking in the name column, deleting the placeholder name, and typing in a new one.

After doing any analysis or graph, you will have another output window. You can access the menu of commands from any window.

CRUNCHIT!

Upon entering CrunchIt!, you will be shown a blank data set with rows and columns (see the figure below). To enter data, click in a cell and enter a value. To change a column name, double-click the column and enter a new column name.



Invoking Statistical Procedures

Users can make a sequence of selections from a series of menus that all begin in the menu toolbar. Once the sequence of selections has been made, you will encounter dialog and/or option boxes that allow you to indicate which variable(s) will be part of the analysis, along with other information. If further help is needed, you can click the Help button that appears in dialog boxes. Once all appropriate information is provided, click "Calculate" to get the desired output.

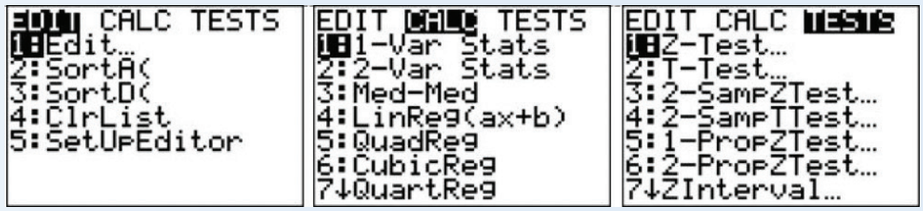
CrunchIt! Files

CrunchIt! provides file options from the File menu, including creating a new data set, importing data from a file and URL, and exporting datasets to a file. CrunchIt also provides direct access to datasets from this book by selecting "Load from The Practice of Statistics in Business and Economics."

TI-83/-84

In this section we provide a very basic overview, but for more instruction, Texas Instruments provides getting started tutorials at education.ti.com ➔ **Products ➔ Graphing calculators**. Here, select your calculator and Support Resources.

After pressing the [STAT] button, you have three options: EDIT, CALC, and TESTS (shown below). Selecting "EDIT" enters the data-table editor, allowing you to enter numeric data; "CALC" includes options for descriptive statistics as well as regression procedures; and "TESTS" includes hypothesis-testing procedures.



Invoking Statistical Procedures

After entering data, statistical procedures can be selected from the "CALC" and "TEST" sections. After making the necessary selections, your calculator will return the results of tests and procedures.

R is command-line software, although some "menu" interfaces (like R Commander) can help—especially beginners. To load R Commander, after installing the package, click **Packages → Load Package** and select Rcmdr. This also allows for an easier means of inputting data.

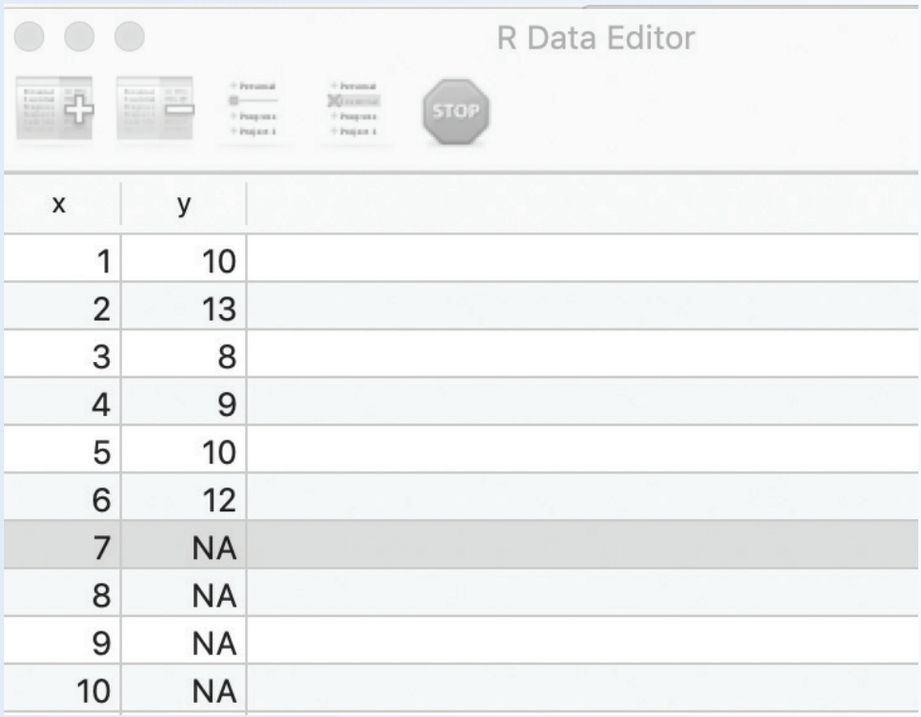
R works from data frames (a collection of variables). There are several methods of inputting data. For a small data set, you may want to directly enter the data from the command line, as in the following example, which creates a data frame called mydat with variables x and y:

```
> x=c(1,2,3,4,5,6,7,8)
> y=c(10,13,8,7,9,8,4,10)
> mydat <- data.frame(x,y)
```

Another method evokes a spreadsheet-like input frame (this example creates variables x and y with 0 decimal places):

```
> mydata <- data.frame(x=numeric(0),y=numeric(0))
> mydata <- edit(mydata)
```

To add rows, click the icon with the blue dot. To replace the "NA" placeholder, use the Tab key to move across rows and the down arrow to move down the spreadsheet.



You can also input data by reading from a file. R can read many types, including .csv (comma separated variable) format, .xls and .xlsx (Excel), as well as others. An example command to read a .csv file that indicates that the first row has variable names is given below.

```
>mydata <- read.csv("file.txt",head=T)
```

R commands have many possible parameters to give graphs titles and so on. For full documentation on any command, click Help, select R functions(text), and enter the name of the command in the box.

The R Video Technology Manuals: Introduction Video would be helpful here.

Displaying Distributions with Graphs

Bar Graphs

1. With pretabulated frequencies, the spreadsheet should have two columns with a column name in the top row: one column should have the category names, and the other column should have the total counts of each category.
2. Select all cells, including the column names.
3. Click **Insert → Chart → Bar**.
4. A bar graph will be created automatically. If you want to change the orientation of the bars, click **Change Chart Type** and select a different orientation.

When you have only one column that requires counting, you will need to create a Pivot Table.

1. Click Data **→ Summarize with Pivot Table**. Select a location for the table (typically a new worksheet).
2. Find that the field name appears in a section titled Field Name. Drag this to the Rows area. You want to also have this field name in the section titled Σ Values. Drag the field from the field section into the Σ Values section. Excel will then automatically make the counts
3. Create a corresponding bar graph as above.

Pie Charts

1. Follow the steps for making a bar graph.
2. Change the created bar graph into a pie chart by clicking **Change Chart Type** in the Chart Design tab, then select the Pie chart type. Add labels for the category names and percents using the Add Chart Element option.

Note: Alternatively, select **Insert → Chart → Pie**.

Pareto Chart

1. Highlight both columns of the data table.
2. Click the **Data** tab and then click **Sort** in the Sort & Filter group.
3. Select the column with the frequencies and choose the **Descending** (largest to smallest) option and.
4. To convert the counts into percent, click the field name found in the Σ Values section, select the Value Field Setting option, click the **Show Data As** tab, then finally select "% of Grand Total."
5. Click **OK**.

Histograms

1. In the Data Analysis menu box, select **Histogram** and click **OK**.
2. Enter the cell range of the data into the Input Range box. If you want Excel to automatically select the classes, leave the Bin Range box empty.
3. Place a check mark next to the Chart Output option. Click **OK**.

Note: To remove gaps between bars, right-click on any one of the bars, select the Format Data Series option, then set the gap width to 0%. With no gap, it is best to border the bars with line edges. Before closing the Format Data Series box, click the Border Color option and select the Solid line option and then click Close.

If you wish to change the automatically selected classes, enter upper values for each class into the spreadsheet and input their cell range in the Bin Range box.

Stemplots

Stemplots are not available in standard Excel or with the Data Analysis ToolPak.

Time Plots

1. Click and drag the mouse to highlight the cell range of the data you wish to time plot (include the column name if you wish it to appear as a chart label).
2. With the cell range highlighted, click **Insert → Chart → Line**.
3. Within Change Chart Type choices, you can choose whether to have data symbols at the data values.

For videos to help with these topics, see the Excel Video Technology Manuals on Bar Chart, Pie Chart, Histogram, and Time Plot.

Bar Graphs

Using the Distribution Platform (does not separate bars) for a single column of data:

1. Click **Analyze → Distribution**.
2. Select variables of interest and click "Y" to cast the variable into that role.
3. Click **OK**.

If you have one column of categories and another with the frequencies, in the Analyze distribution menu,

1. Select the variable containing the categories and click **Y** to cast the variable into that role.
2. Select the frequency variable. Click **Freq** to cast it into that role.
3. Click **OK**.

Note: Frequency Bar Graphs are produced for nominal and ordinal variables, and histograms are produced for continuous variables. Remember, you can change the modeling types of variables by clicking the icon next to the variable name in columns list in the dataset.

Using Graph Builder (properly separates bars):

1. Click **Graph → Graph Builder**.
2. Drag a nominal or ordinal variable of interest to the X axis. If you have a frequency variable, drag it to the Y axis.
3. Click the bar chart icon in the toolbar.

Pareto Chart

1. Click **Graph → Pareto Plot**. (Note: This does not separate bars.)
2. Select the variable containing the categories and click **Y** to cast the variable into that role.
3. Select the frequency variable. Click **Freq** to cast it into that role.
4. Click **OK**.
5. If you want to remove the cumulative percent curve, click the red triangle next to Pareto Plot in the output and uncheck that option.

If using Graph Builder (and you have a frequency variable), sort the data by descending frequency first.

1. Click **Tables → Sort**.
2. Click to enter the frequency variable into the By box. Click that name in the box to highlight. Click the down arrow option. Proceed with Graph Builder as described above.

Pie Chart

Using the Pareto Plot Platform:

1. Produce the Pareto plot as described above.
2. Click the red triangle and select "Pie Chart."

Using Graph Builder:

1. Click **Graph → Graph Builder**.
2. Drag the nominal or ordinal variable of interest to the X axis. If you have a frequency variable, drag it to the Y axis.
3. Click the pie chart in the toolbar.

Histograms

1. Click **Analyze → Distribution**.
2. Select the continuous variable(s) of interest and click **Y** to cast variables into that role.
3. Click **OK**.
4. If you wish to change the automatically selected classes, use the **Graph → Graph Builder** option; drag the variable of interest to the X axis, then select the histogram icon. Click on a value on the axis, then change the minimum, maximum, and/or the bin width (Increment) and click **OK**.

Stemplots

1. Click **Analyze → Distribution**.
2. Select the continuous variable(s) of interest and click **Y** to cast variables into that role.
3. Click **OK**.
4. Click the red triangle next to a variable's name and select **Stem and Leaf**.

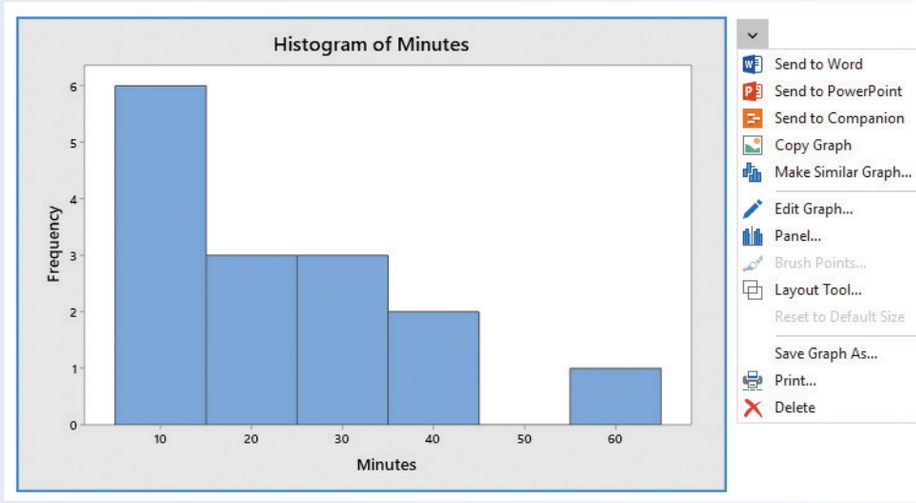
Time Plots

1. Click **Graph ➔ Graph Builder**.
2. Drag the time variable to the X axis.
3. Drag a continuous outcome variable to the Y axis.
4. Click the line chart (next to the bar chart icon) in the toolbar.

For videos to help with these topics, see the JMP Video Technology Manuals on Bar Chart, Pie Chart, Histogram, Stemplot, and Time Plot.

MINITAB

Where Minitab graphical output will appear depends on your version. Minitab 19 includes all output in the Session (output) window. Previous versions have all output in a separate graph window. In either case, clicking anywhere in the graph will invoke the Graph editor. Clicking on a specific item in the graph (for example, the default title) will open a dialog box specific to that item. Click **OK** when your editing is complete. To copy a graph to Word or some other application, in Version 19 click the down arrow at the top of the graph. A menu (shown below) will appear. To access these options in previous versions, right click in the graph.



Bar Graphs

1. Click **Graph ➔ Bar Chart**.
2. If the frequencies have been pretabulated, select "Values from a table" in the "Bars represent" box. If the frequencies have not been tabulated, select "Counts of unique values." Click **Simple** (the default) for the type of bar graph, then click **OK**.
3. For pretabulated frequencies, click to select the frequencies column into the Graph variables box and click to select the column that has the names of the categories into the Categorical variables box. If the frequencies have not been pretabulated, click the column that has data on the categorical names that need to be counted into the Categorical variables box.
4. Click **OK**.

Pareto Chart

1. Follow the instructions for a bar chart.
2. Before clicking **OK**, click **Chart Options**.
3. Check the box next to "Decreasing Y".
4. Click **OK** to return to the main dialog and **OK** to generate the graph.

Pie Charts

1. Click **Graph ➔ Pie Chart**.
2. If the frequencies have been pretabulated, select the "Chart values from a table" option; if the frequencies have not been pretabulated, select "Chart counts of unique values" option by clicking the appropriate radio button.
   If the frequencies have been pretabulated, click to select the data column into the Summary variables box and click the column that has the names of the categories into the Categorical variables box. If the frequencies have not been pretabulated, click in the column that has data on the categorical names that need to be counted into the Categorical variables box.
3. If you wish to have the pie slices labeled by categorical names and have percents reported (as in Figure 1.1(b) in your text), click **Label**, then click the **Slice Labels** tab and place check marks next to the desired labels. Otherwise, there will only be a color code key to the right of the pie.
4. Click **OK**.

Histograms

1. Click **Graph ➔ Histogram**.
2. Select **Simple** (the default) for the type of histogram, then click **OK**.
3. Click to select the data column into the Graph Variables box and then click **OK**.
4. If you wish to change the automatically selected classes, double-click on the horizontal axis to make the Edit Scale box appear. Now, click the **Binning** tab and then choose the Midpoint/Cutpoint positions option found in the Interval Definition section. Depending on whether you choose the Interval type as "Midpoint" or "Cutpoint," you then give the desired values of the midpoints (that is, the middle values of the classes) or the cutpoints (that is, lower and upper values of the classes). You need only give the first few; Minitab will complete the pattern to hold all the data.
5. Click **OK**.

Stemplots

1. Click **Graph ➔ Stem-and-Leaf**.
2. Click to select the data column into the Graph Variables box.
3. Depending on the values you have, you may want to split the stems. To do this, enter a value into the Increment box. We recommend doing an initial plot to check the scaling. If increments were (for example), 10, enter 5 to split those 10s into two lines each.
4. Click **OK**.

Time Plots

1. Click **Graph ➔ Time Series Plot**.
2. Select **Simple** (the default) for the type of time series plot, then click **OK**.
3. Click-in the data column into the Series box.
   Note: By default, Minitab will label the time periods as "1," "2," "3," and so on. If you wish to label the time periods by year, as in Figure 1.7 in your text, click the Time/Scale button, select the Calendar option, and select the desired time periods (for example, "Year") from the adjacent menu and a starting value. Click **OK** to return to the main dialog.
4. Click **OK**.

For videos to help with these topics, see the Minitab Video Technology Manuals on Bar Chart, Pie Chart, Histogram, Stemplot, and Time Plot.

SPSS

If this is the first time the Chart Builder has been used, you will see an initial screen reminding you to ensure that variables are the appropriate type for the graph you are building. Click to place a check in the box at the lower left if you do not want to see this warning again.

Graphs here (and with other software like Minitab) have too many possible options to detail here. We give only basic directions.

If creating several graphs with Chart Builder, click **Reset** at the bottom of the dialog box to start a new graph.

Bar Graphs

1. Click **Graphs ➔ Chart Builder**.
2. Drag the bar graph icon  into the graph area.
3. Drag the categorical variable to the X axis. If you have a table of categories and counts (or percents), drag the frequency variable to the Y axis.
4. If you want to give the graph a meaningful title, click **Edit Title 1** in the "Edit properties of" box at the upper right. Move the radio button to **Custom** and enter the desired title.
5. Click **OK**.

Pie Charts

1. Click **Graphs ➔ Chart Builder**.
2. Drag the pie chart icon into the graph area.
3. Drag the categorical variable to the "Slice by?" box. If you have a table of categories and counts (or percents), drag the frequency variable to the "Angle Variable?" box.
4. If you want to give the graph a meaningful title, click **Edit Title 1** in the "Edit properties of" box at the upper right. Move the radio button to **Custom** and enter the desired title.
5. Click **OK**.

Pareto Plot

You could use **Data ➔ Sort** to sort a table of data (categories and frequencies) in decreasing order by the frequency variable and then create a bar chart as described above or use the following. Here, we assume you have a table of categories and frequencies.

1. Click **Analyze ➔ Quality Control ➔ Pareto Chart**.
2. Select "Simple" (the default) and click **Define**.
3. Click the radio button next to "Sums of Variable." Highlight the frequency variable and click the arrow to move it into that box.
4. Select the categorical variable to plot, then click the arrow next to "Category Axis" to move the variable to the category axis section.
5. To add a title to your plot, click the **Titles . . .** button. A new box will appear to enter titles or subtitles. Click **Continue** to close this dialog box.
6. If desired, click to uncheck the box next to "Display cumulative line."
7. Click **OK**.

Histograms

There are at least two ways to create a histogram.

1. Click **Graphs → Chart Builder**.
2. Select the Histogram graph type and simply drag the variable of interest to the X axis.
3. To control binning, the highlight should be on "Bar 1" in the "Edit Properties of" box. Click **Set Parameters**. Move the radio buttons from "Automatic" to "Custom" and enter values for the starting value, the "Custom value for anchor" and "Interval width." Click **Continue** to return to the main dialog.
4. If you want to give the graph a meaningful title, click **Edit Title 1** in the "Edit properties of" box at the upper right. Move the radio button to **Custom** and enter the desired title.
5. Click **Continue**.

Or use the following, as you likely will want summary statistics as well for the continuous (scale) variable of interest. *This also is used to create stemplots and boxplots*.

1. Click **Analyze → Descriptive Statistics → Explore**.
2. Select the variable of interest on the left, then click the right arrow to move the variable to the Dependent List box.
3. Click **Charts** and check the box(es) to select Histogram (or Stem-and-Leaf or both). Click **Continue** and **OK**.
4. To change the binning (bar scaling), double-click in the graph for the Chart Editor, then click in a bar of the graph for Properties. Select the **Binning** tab. Move the radio button to **Custom** and enter either a number of bars (bins) or a bin width. To change the minimum X value in a bar, click to place a check next to "Custom value for anchor" and enter the value.
5. Click **Apply → Close**.
6. Close the Chart Editor window.

Stemplots

See the above on using Explore to create a histogram.

Note: This procedure also produces a boxplot and descriptive statistics by default.

Time Plots

With Sequence Charts:

1. Click **Analyze → Forecasting → Sequence Chart**.
2. Select the variable of interest on the left, then click the right arrow next to "Variables" to move the variable into the box.
3. If you have a variable identifying time, select it and click the right arrow next to "Time Axis Labels."
4. Click **OK**.

With Scatter/Dot (requires a time variable):

1. Click **Graph → Chart Builder**.
2. Select "Simple Scatter"  and drag the icon to the graph area.
3. Select the variable of interest and drag it to the "Y-Axis?" box.
4. Select the time variable and drag it to the "X-Axis?" box.
5. If you want to give the graph a meaningful title, click **Edit Title 1** in the "Edit properties of" box at the upper right. Move the radio button to **Custom** and enter the desired title.

6. Click **Continue** for an initial scatterplot.
7. Double-click the scatterplot in the output window to open the editor.
8. In the toolbar, select the "Interpolation Line"  at the upper right.
9. Close the editor to finalize the graph.

For videos to help with these topics, see the SPSS Video Technology Manuals on Bar Chart, Pie Chart, Histogram, Stemplot, and Time Plot.

CRUNCHIT! Bar Graphs

With summarized data:

1. Click **Graphics → Bar Chart With Summarized Data**.
2. For "Labels," select the column identifying groups.
3. For "Heights," select the frequency variable.
4. Add a title and X and Y axis labels, if desired.
5. Click **Calculate**.

With raw data:

1. Click **Graphics → Bar Chart With Raw Data**.
2. For "Sample," select the column of interest; to avoid many "short" bars, you can enter a value in Cutoff that will gather all categories with frequencies less than the specified value into an "Other" category.
3. Add a title and X and Y axis labels, if desired.
4. Click **Calculate**.

Pie Charts

With summarized data:

1. Click **Graphics → Pie Chart With Summarized Data**.
2. For "Labels," select the column identifying groups.
3. For "Sizes," select the frequency variable.
4. Add a title, if desired.
5. Click **Calculate**.

With raw data:

1. Click **Graphics → Pie Chart With Raw Data**.
2. For "Sample," select the column of interest; to avoid many small slices, you can enter a value in Cutoff that will gather all categories with frequencies less than the specified value into an "Other" category.
3. Add a title and X and Y axis labels, if desired.
4. Click **Calculate**.

Pareto Plot

Pareto plots are not available in CrunchIt. For a Pareto chart, you would sort the data first by decreasing order of the frequency, which is not a function available in CrunchIt!

Histograms

1. Click **Graphics → Histogram**.
2. For "Sample," select the column of interest.
3. If desired, specify number of bins, bin width, start point, title, and axis labels.
4. Click **Calculate**.

Stemplots

1. Click **Graphics → Stem and Leaf**.
2. For "Sample," select the column of interest.
3. If desired, enter a title.
4. Click **Calculate**.

Time Plots

1. Click **Graphics → Scatter Plot**.
2. For "X," enter the time variable which must be numeric, such as day, year, or an index (1, 2, …, n).
3. For "Y," select the variable of interest.
4. In the Parameters section, change "Points" to "Lines" or "Both."
5. Click **Calculate**.

For videos to help with these topics, see the CrunchIt! Help Videos on Pictures for Categorical Data and Pictures for Quantitative Data.

TI-83/-84

TI Calculators try to graph everything they can at the same time. For that reason, before creating any statistical graph/plot, you should check to see that no functions are entered on the Y= screen; if so, use CLEAR to erase those functions. Also, make sure only one STAT PLOT is "On" at a time; use STAT PLOTS option 4: PlotsOff to turn them all off.

Bar Graphs

TI calculators can handle only numeric data. You can create the look of a bar graph using the following, but actual descriptive category names will not be shown.

1. Press STAT ENTER to select "Edit."
2. In L1 enter sequential values (1, 2, 3 . .) up to as many categories you have.
3. Enter the values associated with each category in L2.
4. Press WINDOW, then set the Xmin and Xmax to match the values in L1 and adjust Ymin and Ymax to be an appropriate range for your Y variable.
5. Press 2nd Y= (STAT PLOT) ENTER and turn the plot "On" if needed by using the ▶ and pressing ENTER to move the highlight. Select the histogram.
6. Select L1 for Xlist and L2 for Freq.
7. Press GRAPH.

Pie Charts

Pie Charts are not available on a TI-83.

Pareto Plot

Pareto Plots are not explicitly available on a TI-83, but a bar chart with descending frequencies can be made by following the steps for Bar Graphs while entering categories in order of descending frequencies.

Histograms

1. Press 2nd Y= = STAT PLOT and select the histogram.
2. Select a plot (press ENTER to select Plot 1).
3. Enter the name of the list that contains the data by pressing 2nd 1 = [L1], 2nd 2 = [L2], etc.
4. Get an initial histogram by pressing ZOOM 9.
5. Adjust the windowing (if needed) using WINDOW. Reset Xmin, Xmax, Xscl (the bar width), and Ymax as needed.
6. Press GRAPH. Do NOT press ZOOM 9 after adjusting the window!

Stemplots

Stemplots are not available on a TI-83 or T-84.

Time Plots

1. Press STAT ENTER to enter the list editor.
2. In L1, enter time values (or an index from 1 to n).
3. In L2, enter data for outcome variable.
4. Press 2nd Y= = STAT PLOT and select the connected scatterplot.
5. Select L1 for Xlist, and L2 for Ylist.
6. Press ZOOM 9.

For videos to help with these topics, see the TI-83/-84 Video Technology Manuals on Histogram and Time Plot.

Bar Graphs

A basic bar graph with raw data in a variable named cat in the mydat data frame can be created using the command

```
> barplot(table(mydat$cat))
```

If data are already summarized into a variable named c1, use the command

```
> barplot(mydat$Freq,names.arg=mydat$c1)
```

If more space is desired between bars (to show all the category names), add space=n as an additional parameter. Increase n so all labels are shown.

Pie Charts

A pie chart with raw data in a variable named cat in data frame mydat can be created with the command

```
> pie(table(mydat$cat))
```

If data are already summarized, modify the command to the one below with frequencies in Frq and slice labels in c2.

```
> pie(mydat$Frq,names=c2)
```

Pareto Plot

Pareto plots are not available in R. To create one, sort the data by decreasing frequency (here, the frequency variable is freq) before creating a barplot. An example is that creates data frame sortd is below. Note the use of square brackets.

```
> Sortd <- mydat[order-mydat$freq),]
```

Histograms

A basic histogram for data stored in variable y can be created using the command

```
> hist(mydat$y)
```

To set your own bins (the example below uses bins < 5, 5 < x ≤10, … to >25), modify the command to

```
> hist(mydat$y,breaks=c(5,10,15,20,25))
```

Stemplots

A basic stemplot can be created using the command

```
> stem(mydat$y)
```

Time Plots

A time series plot using an index for time can be done as a connected scatterplot. Use type="b" to have both points and lines or type="l" to simply have connected lines.

```
> plot(mydat$x,mydat$y,type="b")
```

For videos to help with these topics, see the R Video Technology Manuals on Bar Chart, Pie Chart, Histogram, Stemplot, and Time Plot.

Describing Distributions with Numbers

Now that we have provided a general overview, we discuss more specifically how these software programs can be used to create the graphs presented in this chapter.

Numeric Summaries of Distributions

1. Click **Data → Data Analysis** and locate "Descriptive Statistics" in the menu box. Click **OK**.
2. Enter the cell range of the data into the Input Range box.
3. Place a check mark next to the Summary Statistics option. Click **OK**.
4. If you wish to compute the first and third quartiles, click an empty cell in the spreadsheet and then proceed to the statistical function menu as described in the Excel overview section of this Appendix. Scroll down the list of functions and double-click on the **QUARTILE.EXC** function choice. In the Array box, input the cell range of the data. In the Quartile box, input the value "1" to get the first quartile or the value "3" to get the third quartile and then click **OK**.

Boxplots

Boxplots are not available in standard Excel (including in the Data Analysis ToolPak), but they are available in the enhanced add-on version of Excel.

For videos to help with these topics, see the Excel Video Technology Manuals on Boxplots, Sorting Data, and Summary Statistics.

Numeric Summaries of Distributions

1. Click **Analyze ➔ Distribution**.
2. Select continuous variables of interest and click **Y** to cast variables into that role.
3. Click **OK**. This gives the mean, standard deviation, and five-number summary. For other statistics, click the red triangle next to "Summary Statistics" and check the ones you want.

A boxplot that identifies potential outliers is included in the histogram by default.

Boxplots Using Graph Builder

1. Click **Graph ➔ Graph Builder**.
2. Drag a continuous variable of interest to the X or Y axis.
3. Click the boxplot icon in the toolbar.

For videos to help with these topics, see the Excel Video Technology Manuals on Boxplots and Summary Statistics.

Numeric Summaries of Distributions

1. Click **Stat ➔ Basic Statistics ➔ Display Descriptive Statistics**.
2. Click in the data column(s) for which you want to get numerical summaries into the Variables box.
3. To select additional options, click **Statistics**. Click the boxes to select or deselect the statistic(s) of choice. Click **OK** to close the pop-up.
4. Click **OK**.

Boxplots

1. Click **Graph ➔ Boxplot**.
2. For a single variable, select **One Y Simple** (the default), then click **OK**. If you have multiple boxplots that you want to display together, select "Multiple Y's Simple" (if the variables are in separate columns) or "One Y With Groups" (if there is a column with the group identifier). Click **OK**.
3. Click in the data column(s) for which you want to construct boxplots into the Graph variables box.
4. Click **OK**.

For videos to help with these topics, see the Minitab Video Technology Manuals on Boxplots, Sorting Data, and Summary Statistics.

Numeric Summaries of Distributions and Boxplots

1. Click **Analyze ➔ Descriptive Statistics ➔ Explore**.
2. Select the variable of interest on the left, then click the right arrow next to "Dependent List" to move the variable to that section.
3. For a boxplot, click the **Plots…** button and place a check mark in the box next to "Stem-and-Leaf." Click **Continue**.
4. Click **OK**.

Boxplots can also be created using the Chart Builder.

1. Click **Graphs → Chart Builder**.
2. Click **Boxplot** and drag the appropriate icon into the graph area (do you want a single boxplot or one with groups?).
3. Drag the continuous variable to the "Y-axis?" box. For a plot with a grouping variable, drag that variable to the "X-axis?" box. The grouping variable MUST be categorical.
4. If you want to give the graph a meaningful title, click **Edit Title 1** in the "Edit properties of" box at the upper right. Move the radio button to **Custom** and enter the desired title.
5. Click **OK**.

For videos to help with these topics, see the SPSS Video Technology Manuals on Boxplots, Sorting Data, and Summary Statistics.

CRUNCHIT!

Numeric Summaries of Distributions

1. Click **Statistics → Descriptive Statistics**.
2. Place a check box next to each variable of interest.
3. If you want percentiles other than the standard five-number summary, enter those in the box.
4. Click **Calculate**.

Boxplots

1. Click **Graphics → Box Plot**.
2. Check the box next to each variable you want to visualize.
3. Enter a title and X and Y axis labels, if desired.
4. Click Calculate.

For videos to help with these topics, see the CrunchIt! Help Videos on Pictures for Quantitative Data and Descriptive Statistics.

TI-83/-84

Numeric Summaries of Distributions

1. Enter the data in a list, using STAT ENTER to access the list editor.
2. Press STAT and ▶ to CALC. Select option 1:1-Var Stats.
3. Enter L1 (as 2nd 1 or the list with your data). Press the down arrow to Calculate.
4. Within the results, press the down arrow to reveal the five-number summary.

Boxplots

1. Press 2nd Y= = STAT PLOT and use the down arrow and right arrow to select the box plot ⊡ or ⊡. The first does not identify potential outliers using the 1.5IQR criteria; the second does.
2. Enter L1 (as 2nd 1) or the list with your data as Xlist.
3. You can change the mark for potential outliers on the third line of the screen. We do NOT recommend the single pixel, as it can be hard to see.
4. Press ZOOM 9.

For videos to help with these topics, see the TI-83/-84 Video Technology Manuals on Boxplots, Sorting Data, and Summary Statistics.

Numeric Summaries of Distributions

Summary statistics for variable x can be found with a command like

```
> summary(x,statistics=c("mean","sd","IQR,
+ "quantiles"),quantiles=c(0,.25,.5,.75,1))
```

The command "summary(mydat)" will find the mean and five-number summary for all variables in a data frame (or simply specify a single variable).

Boxplots

The command is simply

```
> boxplot(x)
```

For videos to help with these topics, see the R Video Technology Manuals on Boxplots, Sorting Data, and Summary Statistics.

Density Curves and the Normal Distributions

Normal Distribution Calculations

EXCEL

Excel does not provide a means to visualize areas under the Normal curve, but it can compute areas under the Normal curve or work backward.

1. Click an empty cell in the spreadsheet, then **Formulas → More Functions → Statistical**. Scroll down to find **NORM.DIST**.
2. To find the area to the left of the point of interest for any Normal distribution, enter the value, the mean (μ), the standard deviation (σ), and "1" for Cumulative. Click **OK**.
3. For a standard Normal (Z) distribution, use **NORM.S. DIST**. Enter the z-score and "1" for Cumulative. Click **OK**.
4. To do backward Normal calculations, use **NORM.S.INV** for standard Normal (enter the probability to the left of the point, i.e., $P(Z \leq z)$; for any Normal distribution, use **NORM.INV** and enter the probability to the left of the point, the mean (μ), and the standard deviation (σ)). Click **OK**.

Normal Quantile Plots

Neither Normal quantile plots nor Normal probability plots are available as a basic function in Excel. To create these, this author recommends watching the video below.

For videos to help with these topics, see the Excel Video Technology Manuals on Normal Calculations and Normal Quantile Plots.

Normal Distribution Calculations

1. Click **Applets → Distribution Calculator**.
2. Enter a mean and standard deviation if not standard Normal ($\mu = 0, \sigma = 1$) and click Reset Distribution.
3. Select the type of calculation desired (input values to obtain a probability or enter a probability to obtain a value—the inverse calculation).
4. Enter the value(s) type to the right, enter the value(s) of interest, and press the Enter key to see shading on distribution.

Normal Quantile Plot

1. Click **Analyze ➔ Distribution**.
2. Select the continuous variables of interest and click **Y** to cast variables into that role.
3. Click **OK**.
4. Click the red triangle next to the variable name, then select **Normal Quantile Plot**.

For videos to help with these topics, see the JMP Video Technology Manuals on Normal Calculations and Normal Quantile Plots.

Normal Distribution Calculations

1. Click **Graph ➔ Probability Distribution Plot**.
2. Select "View Probability" and click **OK**.
3. Enter values for the mean and/or standard deviation.
4. Click the **Shaded Area** tab. If you want to find the area under the curve associated with a specified value, select the "X Value" option. You can choose to find the area to the left or right of that specified value or even between two values by clicking the appropriate picture.
5. Enter the specified value(s) in the X value box.
6. Click **OK**.

For more precise values:

1. Click **Calc ➔ Probability Distributions ➔Normal**.
2. Choose the "Cumulative probability" option if you wish to find the area to the left of a specified value. Choose the "Inverse cumulative probability" option if you wish to find the value associated with a specified area to the left of that value.
3. Select the "Input constant" option.
4. Enter the specified value of X (or z) or enter the specified area.
5. Click **OK**.

Normal Quantile Plots

1. Click **Graph ➔ Probability Plot**.
2. Select "single" (for only one variable, the default) and click **OK**.
3. Click in the data column of interest into the Variable box.
4. Click **OK**.

For videos to help with these topics, see the Minitab Video Technology Manuals on Normal Calculations and Normal Quantile Plots.

Normal Distribution Calculations

SPSS does not have a probability look up graphical interface. Cumulative probabilities can be found and inverse calculations performed using Compute Variable.

1. Click **Transform ➔ Compute Variable**.
2. In "Target Variable," enter a name for the new variable.
3. To find a cumulative probability, select **CDF and Noncentral**, then **CDF.normal** (this will enter the formula in the editor). Enter the desired parameters into the formula (quantity to look up, mean, and standard deviation). Click **OK**.

4. To find a variable value given a cumulative probability, select **Inverse DF** and then **IDF.Normal**. Enter the parameters and click **OK**.
5. A new column will be added to the data set containing the result.

Normal Quantile Plots

1. Click **Analyze → Descriptive Statistics → Q-Q Plots**.
2. Select the variable of interest on the left, then click the right arrow to move it to the Variables box.
3. Click **OK**. The Q-Q plot appears in the output window after some initial processing details.

For videos to help with these topics, see the SPSS Video Technology Manuals on Normal Calculations and Normal Quantile Plots.

CRUNCHIT!

Normal Distribution Calculations

1. Click **Distribution Calculator → Normal**.
2. Enter the mean and standard deviation (or accept defaults for standard normal),
3. To look up proportions, use the probability tab; to look up a variable value given a cumulative proportion, $P(X \leq x)$, use the quantile tab.
4. After entering a value, click **Calculate**.

Normal Quantile Plots

1. Click **Graphics → Q-Q Plot**.
2. For "Sample," select the column of interest.
3. Click **Calculate**.

For videos to help with these topics, see the CrunchIt! Help Video on Distribution Calculators.

TI-83/-84

Normal Distribution Calculations

There are two ways to find Normal distribution probabilities: using normalcdf from the 2nd VARS = [DISTR] menu or using ShadeNorm from the DRAW menu (2nd PRGM). If using ShadeNorm, be sure to use 2nd PRGM = [DRAW], 1:ClrDraw between calculations. Otherwise, the Normal curve will become increasingly filled in, and you won't be able to discern the area of interest.

For probabilities such as $P(X > 2)$ or $P(X < 1)$, the upper (lower) end of the area is $+$ ($-$) infinity. For the calculator, $+\infty$ is 1 2nd , 99 = 1e99, but practically speaking, any "very large" value will work. The same applies to $-\infty$, and large, negative value will work.

To look up proportions of the normal distribution:

1. Press 2nd VARS = [DISTR]. ▼ to 2:normalcdf(and press ENTER (or just press 2).
2. To look up proportions from a normal distribution with a mean other than 0 and standard deviation other than 1, enter the mean and standard deviation after entering the range. For example, normalcdf(90,110,100,15) will look up the proportion between 90 and 110 for a normal distribution with a mean of 100 and a standard deviation of 15. If you are using standard Normal, there is no need to enter the mean and standard deviation.

3. To use ShadeNorm, you need to size the window. An appropriate size is $\mu \pm 3\sigma$ for the X axis. Y sizing will depend on the distribution. For standard Normal, Xmin = [-][3], Xmax = [3], Ymin = [-][.][1], and Ymax = [.][4] will work well. The parameters for the command (low value, high value, μ, σ) are the same.

To find value in a standard normal distribution with a particular proportion to the left:

1. Press [2nd][VARS] = [DISTR]➔3:invNorm(and press [ENTER] (or just press [3]).
2. Enter the proportion desired, a comma, the mean, and the standard deviation and press [ENTER]. If you are using standard Normal, there is no need to enter the mean and standard deviation. For example, invNorm(.50) will return 0 (at $z = 0$, half of the distribution is to the left).

To look up the value of a normal variable that is not standard, add the mean and standard deviation after the proportion.

Normal Quantile Plots

1. Enter the data in a list.
2. Press [2nd][Y=] = STAT PLOT, select one of the three plots, and select the Quantile plot, ⌞.
3. Enter L1 (as [2nd][1], or the list with your data) as Xlist.
4. Select a data axis (your text uses the Y axis) and a mark type for each data point.
5. Press [ZOOM][9].

For videos to help with these topics, see the TI-83/-84 Video Technology Manuals on Normal Calculations and Normal Quantile Plots.

Normal Distribution Calculations

To find the cumulative probability, i.e., P(X ≤ x), use the command

```
> pnorm(x,mu,sd)
```

To find value in a standard Normal distribution with a particular proportion to the left, use

```
> qnorm(Prob,mu,sd)
```

Normal Quantile Plots

The command is

```
> qqnorm(mydat$x)
```

For videos to help with these topics, see the R Video Technology Manuals on Normal Calculations and Normal Quantile Plots.