

Chapter 1

Looking at Data—Distributions

Excel provides more than 70 functions related to statistics and data analysis as well as tools in the **Analysis ToolPak**. Additionally, the **Chart Wizard** gives a step-by-step approach to creating informative graphs.

We first discuss the **Chart Wizard**. The figures are from Excel 2001 for Macintosh. They differ only cosmetically from Excel 2000 for Windows, as illustrated in Figure 1.1.

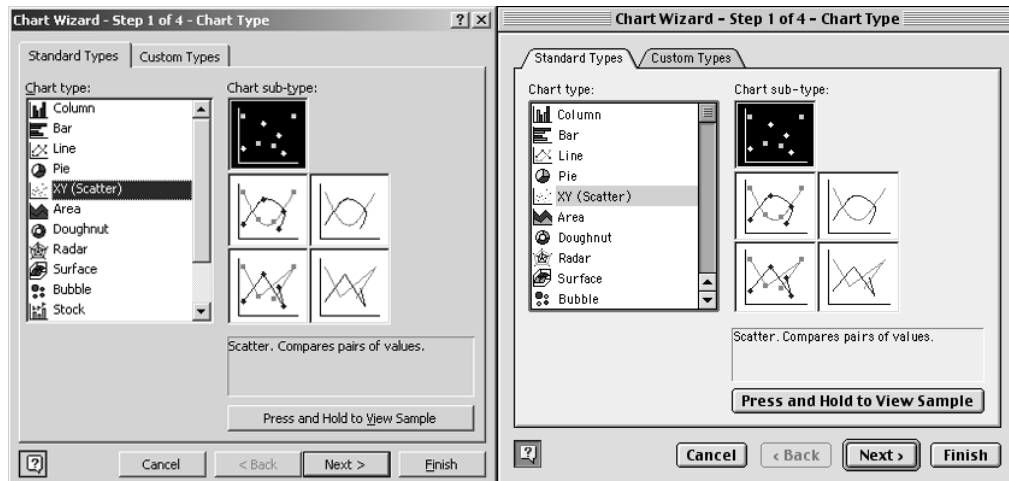


Figure 1.1: Windows 2000 (left) and Macintosh 2001 (right) Chart Wizard Opening

1.1 Displaying Distributions with Graphs

The Chart Wizard

The **Chart Wizard** is a step-by-step approach to creating informative graphs. Its interface provides a sequence of four steps in **Excel 97/2000 (Windows)** and **Excel 98/2001 (Macintosh)** that guide the user through the creation of a customized graph (called a Chart by Excel). The user supplies details about the chart type, formatting, titles, legends, and so on, in dialog boxes. The Chart Wizard can be activated either from the button on the **Standard Toolbar** or by choosing **Insert – Chart** from the Menu Bar. The chart can be inserted in the current sheet or in a new sheet. The following applies to Excel 97/98/2000/2001. **Users of Excel 5/95 will have five steps paralleling the ones presented here.**

Example 1.1. (Page 6 in the text.) Figure 1.2 shows the education level of young adults aged 24 to 35 given both as a count and as a percent. Create a bar chart of the percents.

	A	B	C
1	Education	Count (millions)	Percent
2	Less than high school	4.7	12.3
3	High school graduate	11.8	30.7
4	Some college	10.9	28.3
5	Bachelor's degree	8.5	22.1
6	Advanced degree	2.5	6.6

Figure 1.2: Educational Level

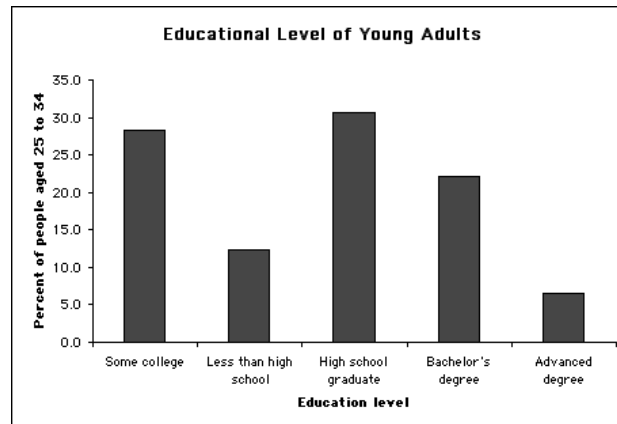


Figure 1.3: Excel Bar Chart

Solution. Figure 1.3 is a bar graph produced by Excel that displays the same information as in Figure 1.2. For other types of graphical displays, make appropriate choices from the same sequence of dialog boxes. The following steps describe

how it is obtained. First enter the data and labels in cells A1:C6 and format the display as in Figure 1.2.

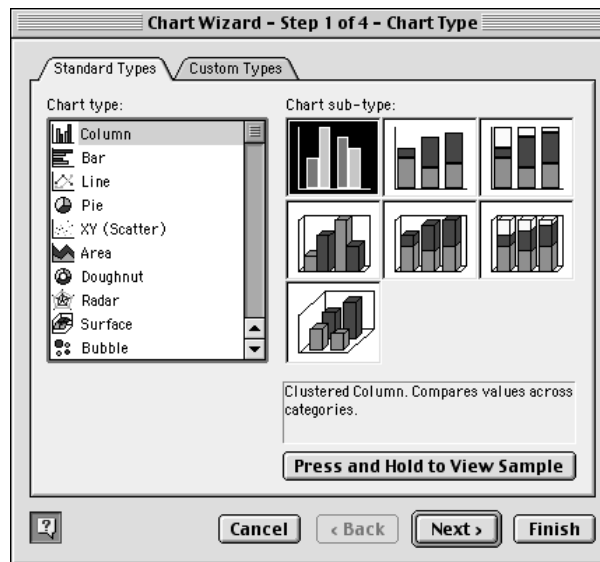


Figure 1.4: ChartWizard—Step 1

Step 1. Select cells A1:C6 and click on the **ChartWizard**. The ChartWizard (Figure 1.4) displays the types of graphs that are available. In the left field select **Column** for Chart type.

In the right field select **Clustered Column** for Chart sub-type, which is the first choice in the top row. When you select a sub-type, an explanation of the chart appears in the box below all the choices, and you can preview your chart’s appearance using the “Press and Hold to View Sample” button. Click Next.

Step 2. The next dialog box (Figure 1.5) with title **Chart Source Data** previews your chart and allows you to select the data range for your chart. Since you had already selected cells A1:C6 prior to invoking the ChartWizard, this block appears in the text area **Data range**. Had you not selected the data range, then you would input the range now or you can make corrections to the data range. The preview chart shows bar charts for both the counts and the percents. To remove the counts click on the **Series** tab at the top of the dialog box, then highlight the series “Count (millions)” and click the **Remove** button (right side of Figure 1.5). The bar graph of counts vanishes. Click Next.

Step 3. A dialog box **Chart Options** (Figure 1.6) appears with the default



Figure 1.5: ChartWizard—Step 2

chart. Rarely is the default satisfactory; you will generally need to make cosmetic changes to its appearance.

- Click the **Titles** tab. Enter “Educational Level” for Category (X) axis and “Percent of people aged 25 to 34” for Value (Y) axis.
- Click the **Legend** tab. We don’t require a legend since only one variable is plotted, so make sure the check box **Show Legend** is cleared.
- Additional tabs are available to customize other types of charts. They are not required here. Click Next.

Step 4. The final step lets you decide if you want the chart placed on the same worksheet as the data or in another worksheet. With each choice there is a field for entering the worksheet name. We will embed the chart on the same worksheet, so we select the radio button **As object in:** (Figure 1.7). As our current workbook only has one sheet, Excel has used the default name Sheet1. We could also embed the chart on another sheet in the same workbook. Click the Finish button.

The chart appears with eight handles indicating that it is selected. The chart can be resized by selecting a handle and then dragging the handle to the desired size. The chart can also be moved. Click the interior of the chart and drag it to another location (holding the mouse button down). Then click outside the chart to deselect.

You will also find the **Chart Toolbar** (Figure 1.8) embedded on the worksheet. This is used for embellishments of the chart. Use of this toolbar is described

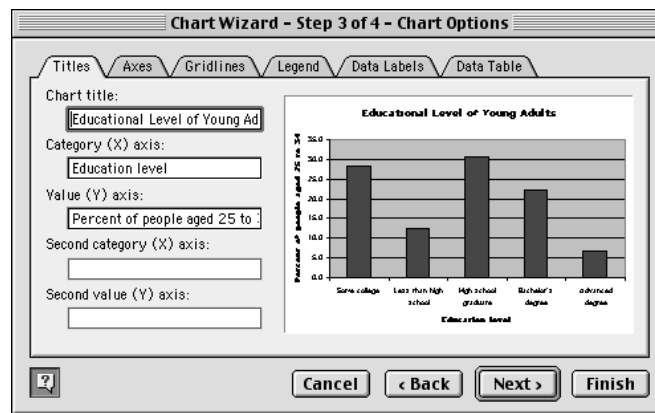


Figure 1.6: ChartWizard—Step 3



Figure 1.7: ChartWizard—Step 4

in the next section on creating histograms. Note that the Chart Toolbar may also be called from the Menu Bar by **View – Toolbars – Chart**. Also, if you select the Chart by clicking once within its area, the Menu Bar will change. In place of the word **Data** there will now appear **Chart** from which a pull-down menu will provide the same tools as are displayed with icons on the Chart Toolbar.

Note: You might get an error message “Cannot add chart to shared workbook” even if you are not sharing your workbook. This is a bug introduced when Shared Workbooks were implemented in Excel 97/98. It occurs under certain conditions if you try to create a chart using the data analysis tools.

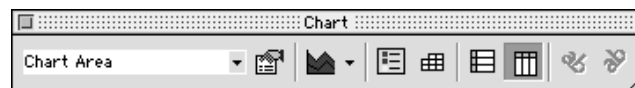


Figure 1.8: Chart Toolbar

You can still output your chart to a new workbook, and then, if desired, copy the chart to the existing workbook. This is one workaround. Fortunately, this problem can be fixed by installing an updated file ProcDBRes to replace the existing one of the same name in the folder/directory “Microsoft Office 98:Office:Excel Add-Ins:Analysis Tools” (the folder for a default installation—your location may differ). This file is available for download at the Microsoft Software Library. There was a similar problems with Excel 97 (fixed in Excel 2000). Details may be found at the URLs <http://support.microsoft.com/default.aspx?scid=kb;EN-US;q183188> and <http://support.microsoft.com/default.aspx?scid=kb;en-us;Q178243>.

Pie Charts

It is easy as pie to produce a chart as in Figure 1.9 using Excel. Select Pie in place of Column in Step 1 of the ChartWizard and follow the remaining steps.

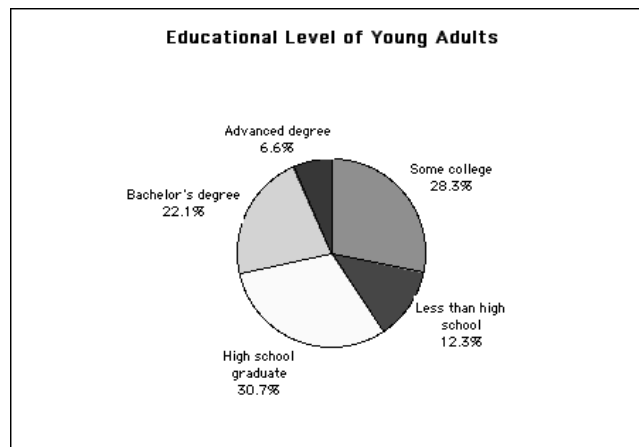


Figure 1.9: Pie Chart

Alternatively, since we have already created a bar graph, it is instructive to use the **Chart Toolbar** as an illustration of how easily modifications may be made. This interface is a vast improvement over the previous version of Excel.

1. Select the completed bar graph by clicking once within its border.
2. From the Menu Bar select **Chart – Chart Type. . .** You will be presented with a box that is identical to Figure 1.4 but for the title, which contains only the words **Chart Type** without mention of Step 1 of the ChartWizard. In the left field, referring to Figure 1.4, select **Pie** for Chart type and in the right field select **Pie** for Chart sub-type, which is the first choice in the top row.

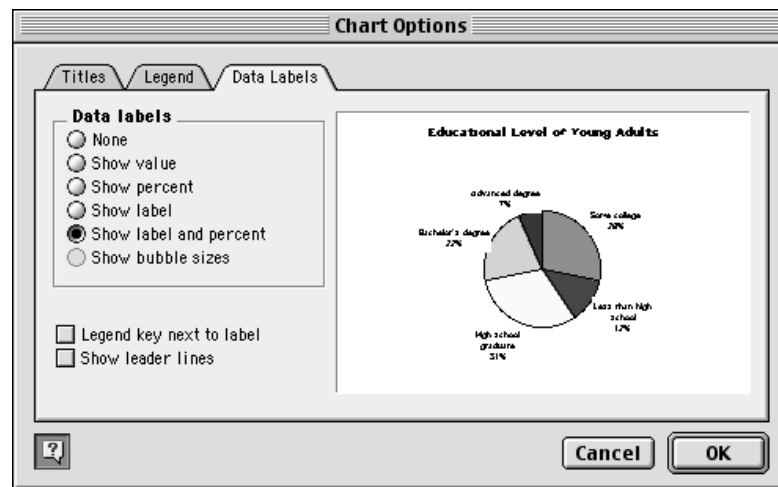


Figure 1.10: Chart Options

3. From the Menu Bar select **Chart – Chart Options. . .** Figure 1.10 appears with three tabs: Titles, Legend, and Data Labels. Click the tab **Data Labels** and then select the radio button **Show label and percent** and check the box **Show leader lines**. Click OK. A pie chart now replaces the bar graph.
4. Double-clicking on any label will bring up the **Format Data Label** dialog box which allows you to change background colors, number of decimals, and other formatting features.

Histograms

The ChartWizard is designed for use with data that are already grouped, for instance, categorical variables, and can therefore be used to construct a histogram of quantitative variables that have been grouped into categories or intervals. However, for raw numerical data, Excel provides additional commands within the **Analysis ToolPak** for constructing histograms.

To determine whether this toolpak is installed, choose **Tools – Add-Ins** from the Menu Bar. The **Add-Ins** dialog box appears. Depending on whether other Add-Ins have been loaded, your box might appear slightly different. If the Analysis ToolPak box is not checked, then select it and click OK. The **Analysis ToolPak** will now be an option in the pull-down menu when you choose **Tools – Data Analysis**. Note that you can also use the Select button to add customized add-ins to complement Excel.

Histogram from Raw Data

Example 1.2. (Example 1.9, page 14 in the text.) Make a histogram of the percent of adult residents in each of the 50 states who identified themselves in the 2000 census as “Spanish/Hispanic/Latino.” (Figure 1.11).

	A	B	C	D	E
1	Percent of Hispanics in the adult population, by state (2000)				
2	1.5	5.7	5.6	38.7	1.2
3	3.6	6.4	2.7	13.8	2.0
4	21.3	10.7	2.4	4.3	28.6
5	2.8	3.1	1.3	1.0	8.1
6	28.1	2.3	1.8	1.6	0.8
7	14.9	5.8	1.6	4.3	4.2
8	8.0	1.3	4.5	6.5	6.0
9	4.0	2.4	16.7	2.6	0.6
10	16.1	0.6	1.4	7.0	2.9
11	5.0	4.0	12.3	2.2	5.5

Figure 1.11: Percent of Hispanics in the Adult Population

Solution. The data appears in Table 1.2 of the text.

1. Excel requires a contiguous block of data for the histogram tool and it is convenient to enter these 50 data in a block of ten rows by 5 columns. So enter the data in a block (cells A2:E11) with the label “Percent of hispanics in the adult population, by state (2000)”.
2. From the Menu Bar choose **Tools – Data Analysis** and scroll to the choice Histogram (Figure 1.12). Click OK.

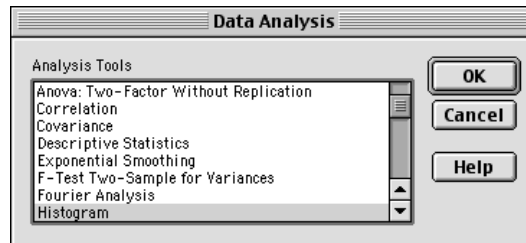


Figure 1.12: Data Analysis Tools

3. In the dialog box (Figure 1.13) type the reference for the range A2:E11 in the **Input range** area, which is the location on the workbook for the data. As with the Bar Chart, you may instead click and drag from cell A2 to E11. The choice depends on whether your preference is for strokes (keyboard) or clicks (mouse). Leave the **Bin range** blank to allow Excel to select the bins, check the **Labels** box blank, type a cell location, say A14,

for **Output range** to denote the upper left cell of the output range, and check the box **Chart output** to obtain a histogram on the same sheet of the workbook as the data. The option Pareto (sorted histogram) constructs a histogram with the vertical bars sorted from left to right in decreasing height. If Cumulative Percentage is checked, the output will include a column of cumulative percentages.

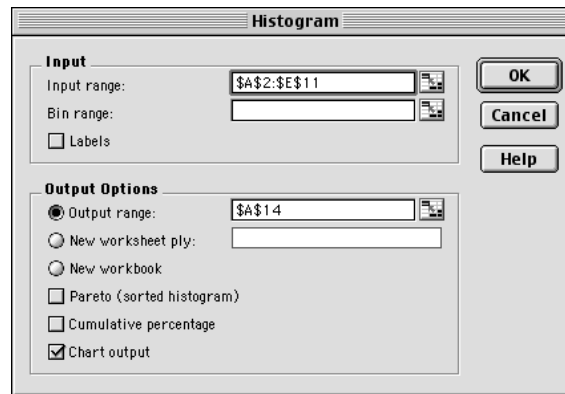


Figure 1.13: Histogram Tool

- The output appears in Figure 1.14. The entries under Bin in A13:A18 are not the midpoints of the bin intervals, as you might expect. Rather they are the **upper limits** of the boundaries for each interval. The corresponding frequencies appear in cells B13:B18 with the histogram to the right. We shall shortly modify the histogram by changing the labels and allowing adjacent bars to touch. But first, we explain how to customize the selection of bins.

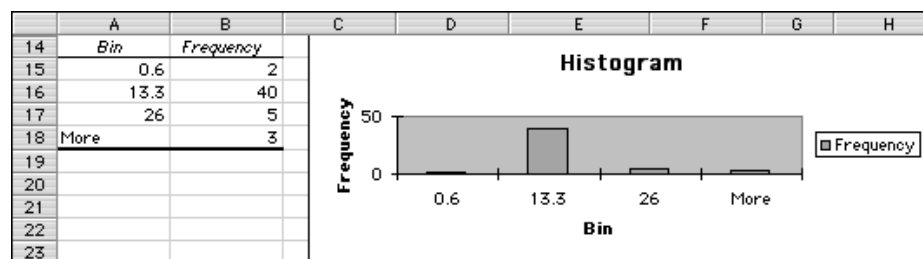


Figure 1.14: Output Table and Default Histogram

Changing the Bin Intervals

If the bin intervals are not specified, then Excel creates them automatically, choosing the number of bins roughly equal to the square root of the number of observations beginning and ending at the minimum and maximum, respectively, of the data set. In creating a histogram from raw data, we let Excel choose the default bins. Here we select our own bin intervals.

1. Type “Bin” (or another appropriate label) in an empty location, say G1. Then enter the values 5.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0 directly below in cells G2:G9. An easy way to accomplish this is to type 5.0 and 10.0 in cells G2 and G3 respectively, then select G2 and G3, click the fill handle in the lower right corner of G3, then drag the fill handle down to cell G9 and release the mouse button.

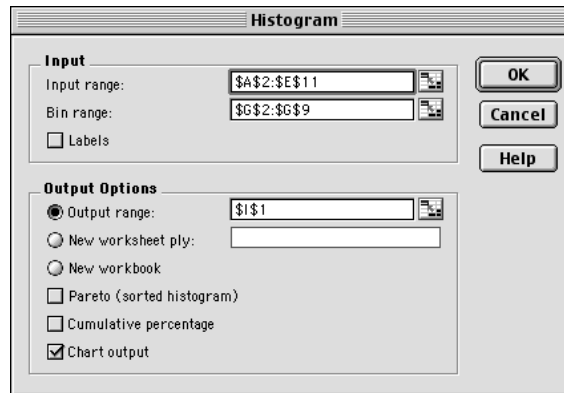


Figure 1.15: Histogram Tool—Specified Bin Intervals

2. Repeat the earlier procedure for creating a histogram, but this time type G2:G9 in the text area for **Bin range** and select a location (cell I1) in Figure 1.15) to mark where the output with your selected bin intervals will appear (see Figure 1.16).

Enhancing the Histogram

While the default histogram captures the overall features of the data set, it is inadequate for presentation. Excel provides a set of tools for enhancing the histogram. These are too numerous for all to be mentioned here, but a few will be discussed with reference to the example. The other options may be invoked analogously.

Legend. To remove the legend (which is not needed here) select **Chart – Chart Options...** from the Menu Bar, click the **Legend** tab, and clear the box **Show legend**.

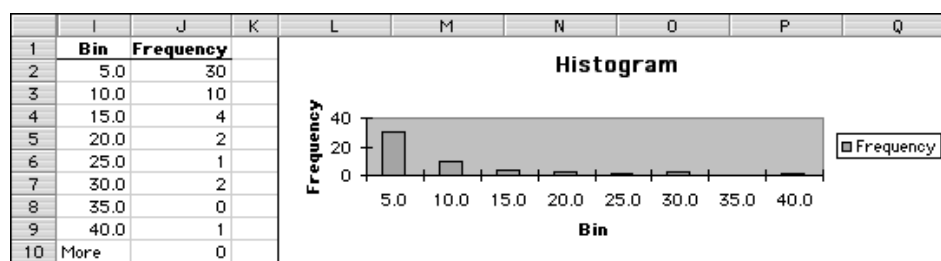


Figure 1.16: Output Table and New Bin Histogram

Resize. Both the histogram (called the **Plot Area**) and the box (called the **Chart Area**) that contains it can be resized and moved. Select the Chart Area by clicking once within its boundary, resize using any of the eight handles that appear, or move it by dragging or cutting and pasting from **Edit** on the Menu Bar to a new location. Likewise, select the Plot Area by clicking once within its boundary and then resize or move as with the Chart Area. The X axis labels may appear placed by default horizontally, vertically or diagonally to accommodate the selected size. This can also be changed. After removing the Legend and resizing click **outside** the Chart Area to deselect.

Bar Width. Adjacent bars do not touch in the default histogram, which looks more like a bar chart for categorical data. To adjust the bar width, click and select any one of the bars, and then from menu **Bar** select **Format – Selected Data Series...** to bring up the **Format Data Series** dialog box. Select the Options tab (Figure 1.17) and change the Gap width from 150% to 0%.

Chart Title. Click on the title word Histogram. A rectangular grey border with handles will surround the word, indicating that it is selected for editing. Begin typing “Percent of Hispanics in an Adult Population”, hold down the **Alt** key (**Windows**) or the **Command** key (**Macintosh**), and press Enter. You may now type a second line of text in the **Formula Bar** entry area. Continue typing “by state (2000)”, and then press Enter. If you want to move the title within the Chart Area, use the handles. To change the font of your title, select the title, then from the Menu Bar choose **Format – Selected Chart Title...** The dialog box has three tabs: Patterns, Font, and Alignment. Select the **Font** tab and pick a font face, style, and size.

X axis Title. Click on the word Bin at the bottom of the chart and type “Percent of Hispanics.” Change the font by selecting the X axis title, then from the Menu Bar choose **Format – Selected Axis Title...** and complete the dialog box as desired in the same fashion as for the Chart Title.

Y axis Title. Click on the word Frequency on the left side, and then from the

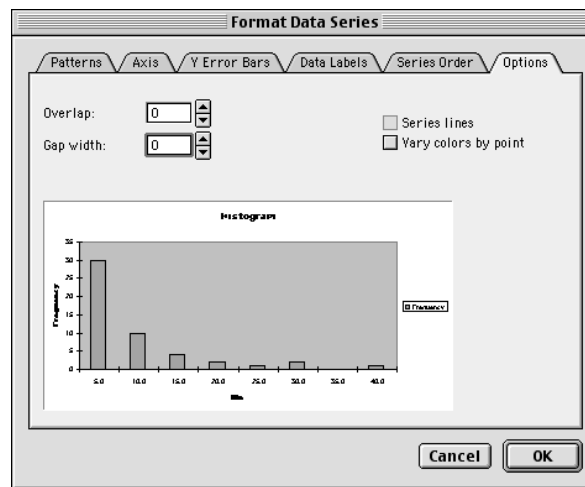


Figure 1.17: Format Data Series

Menu Bar choose **Format – Selected Axis Title...** for any desired formatting.

X axis Format. Double-click the X axis, and in the **Format Axis** dialog box you can click on various tabs to change the appearance of the X axis. If you click on the **Alignment** tab you can change the orientation of the X axis labels.

Y axis Format. Double-click the Y axis, and in the **Format Axis** dialog box you can click on various tabs to change the appearance of the Y axis. Click on the **Scale** tab and change the **Maximum** to 30. Sometimes if you resize the chart you will need to experiment with the **Major** or **Minor** units to achieve a pleasing result. Click OK.

More Interval. The “More” interval with 0 counts is unattractive, especially if it appears on your graph (which it may, depending on your default settings, prior to enhancement). In the workbook in cell I10 (refer to Figure 1.16), change the label More to 45.0. The histogram is dynamically linked to the data in columns I and J and the label More on the X axis becomes 45.0. Of course, knowing that the count is zero in the bin labeled 45.0, we could redo the histogram ignoring this bin, if we wished to exclude it.

Plot Area Pattern. The default histogram has a border around the Plot Area and the Plot Area is shaded grey. Both defaults can be changed by double-clicking the **Plot Area** to bring up the **Format Plot Area** dialog box. If desired, select the radio button **None** for Border and also select the radio button **None** for Area.

At the conclusion of the formatting, the histogram will look like Figure 1.18.

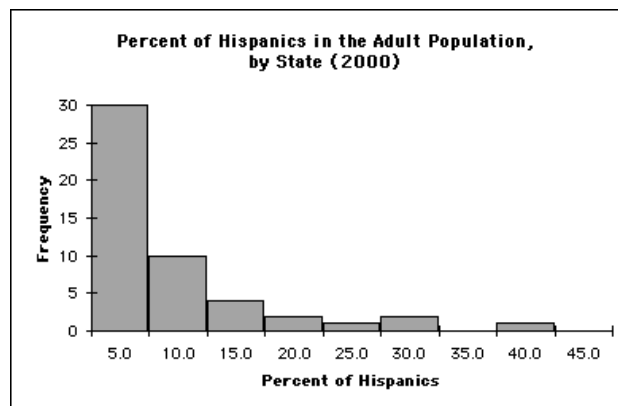


Figure 1.18: Final Histogram after Editing

Histogram from Grouped Data

The **Histogram** tool requires the raw data as input. When numerical data have already been grouped into a frequency table, it is the **ChartWizard** that is the appropriate tool. First use it to obtain a bar chart, and then modify it exactly as you would enhance a histogram.

Example 1.3. (See Exercise 1.17, page 25 in the text.) Figure 1.19 gives the frequencies of vocabulary scores of all 947 seventh graders in Gary, Indiana, on the vocabulary part of the Iowa Test of Basic Skills. Column A is the interval for the score and column B is the score. Construct a histogram using the **ChartWizard**. The final histogram should be similar to that shown in Figure 1.20.

	A	B	C
1	Score	Bin	Number of Students
2	2.0 - 2.9	3	9
3	3.0 - 3.9	4	28
4	4.0 - 4.9	5	59
5	5.0 - 5.9	6	165
6	6.0 - 6.9	7	244
7	7.0 - 7.9	8	206
8	8.0 - 8.9	9	146
9	9.0 - 9.9	10	60
10	10.0 - 10.9	11	24
11	11.0 - 11.9	12	5
12	10.0 - 12.9	13	1
13	Total		947

Figure 1.19: Vocabulary Scores—Grouped Data

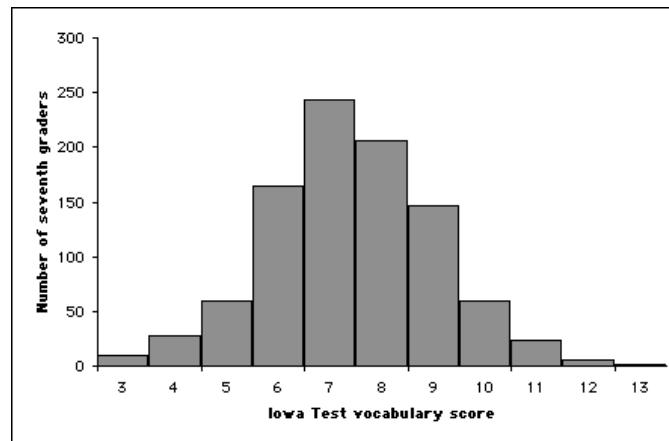


Figure 1.20: Histogram from Grouped Data

1.2 Describing Distributions with Numbers

The most direct way to obtain the common summary statistics is through the **Descriptive Statistics Tool**, which provides preformatted output very quickly. It is explained in this section. An alternative is the **Formula Palette** which provides greater flexibility of output and many more functions and formulas over its predecessor. We first describe the Descriptive Statistics Tool and then the Formula Palette.

The Descriptive Statistics Tool

Example 1.4. (Based on problems 1.55 - 1.57, pages 59 - 60 in the text.) Figure 1.21 shows the calories and sodium levels measured in three types of hot dogs: beef, meat (mainly pork and beef), and poultry. Data is from *Consumer Reports*, June 1986, pp. 366-367. Describe the data using the Descriptive Statistics Tool.

Solution. For illustration purposes we consider only the beef calories data.

1. From the Menu Bar choose **Tools – Data Analysis** and double-click **Descriptive Statistics** (or, equivalently, select **Descriptive Statistics** and click OK) in the **Data Analysis Dialog** box. A dialog box **Descriptive Statistics** appears which prompts for user input.
2. Complete the input as follows. The **Input range:** is A3:A23, corresponding to the beef calories, including labels. (If you selected this range prior to invoking Descriptive Statistics, it will already be inserted by Excel.) Check the box **Labels in first row**. The **Confidence level for mean:** is not

	A	B	C	D	E	F
1	Beef hot dogs		Meat hotdogs		Poultry hotdogs	
2						
3	Calories	Sodium	Calories	Sodium	Calories	Sodium
4	186	495	173	458	129	430
5	181	477	191	506	132	375
6	176	425	182	473	102	396
7	149	322	190	545	106	383
8	184	482	172	496	94	387
9	190	587	147	360	102	542
10	158	370	146	387	87	359
11	139	322	139	386	99	357
12	175	479	175	507	170	528
13	148	375	136	393	113	513
14	152	330	179	405	135	426
15	111	300	153	372	142	513
16	141	386	107	144	86	358
17	153	401	195	511	143	581
18	190	645	135	405	152	588
19	157	440	140	428	146	522
20	131	317	138	339	144	545
21	149	319				
22	135	298				
23	132	253				

Figure 1.21: Hot Dog Data

needed at this time (it gives the half-width). Check the **Kth largest:** or **Kth smallest:** boxes if needed. We have selected $K = 5$ for illustration.

3. The **Output Options** tell Excel where to place the output. Select cell A26. Finally check the box Summary Statistics and click OK. The output appears in Figure 1.22. We have formatted the output by reducing the number of decimal points using the **Decimal** button in the **Formatting Toolbar**. We can read off the summary statistics:

mean = 156.85
 standard deviation = 5.0629
 median = 152.50
 minimum = 111 maximum = 190
 5th smallest = 139 5th largest = 181

Formula Palette

The **Function Wizard** was replaced by the **Formula Palette** in **Excel 97/98**. This is a tool that assists in entering formulas and functions included in Excel, particularly complex ones. The functions can perform decision-making, action-taking, or value-returning operations. The Formula Palette simplifies this process by guiding you step by step.

It can be fired up in one of two ways. When you select a cell and press the

	A	B
26	<i>Calories</i>	
27		
28	Mean	156.85
29	Standard Error	5.0629
30	Median	152.50
31	Mode	149
32	Standard Deviation	22.6420
33	Sample Variance	512.6605
34	Kurtosis	-0.813
35	Skewness	-0.031
36	Range	79
37	Minimum	111
38	Maximum	190
39	Sum	3137
40	Count	20
41	Largest(5)	181
42	Smallest(5)	139

Figure 1.22: Descriptive Statistics Output

Paste Function button f_x next to the autosum button Σ on the **Standard Toolbar** (or, equivalently, choose **Insert – Function...** from the Menu Bar), an equal sign (=) appears both in the cell and in the **Formula Bar**. The **Paste Function** dialog box (Figure 1.23) appears showing all available functions grouped by category on the left and the function name on the right. Both lists have scroll bars for choices not directly visible on the screen. At the bottom of the box, the selected function is shown with the arguments it takes and a brief description. (In previous versions of Excel, a similar dialog box called **Function Wizard – Step 1 of 2** appeared.) When you click OK in the **Paste Function** box, the **Formula Palette** box appears below the **Formula Bar**, requesting parameters and an input range for the function you selected. In addition, the Formula Bar is now activated showing the Formula Palette’s drop-down list control with the 10 most recently used functions, and an equal (=) sign appears in the Formula Bar showing the selected function partially constructed and awaiting completion of its arguments. You may enter these either directly into the Formula Bar or in the Formula Palette box.

The **Formula Palette** is usually invoked in a second, more direct way. Select a cell and press (=) on the **Formula Bar** to open the **Formula Palette** dialog box (Figure 1.24). On the far left side of the **Formula Toolbar** is a button with the most recently used function, in this case **SUM**. If this is the function you need then click on the word **SUM** and the **Formula Palette** dialog box will expand requesting the required parameter or the data range for the function (which can be typed directly or *referenced* by using the mouse to point to the data by clicking and dragging over cells in the data range). As you input this information, Excel will correspondingly build the function both in the **Formula Bar** and in the cell you had selected in the workbook. When you have completed entering the requested input, click OK to complete the function. If you want some other function than the default, click the small arrow in the box to the right of the function name

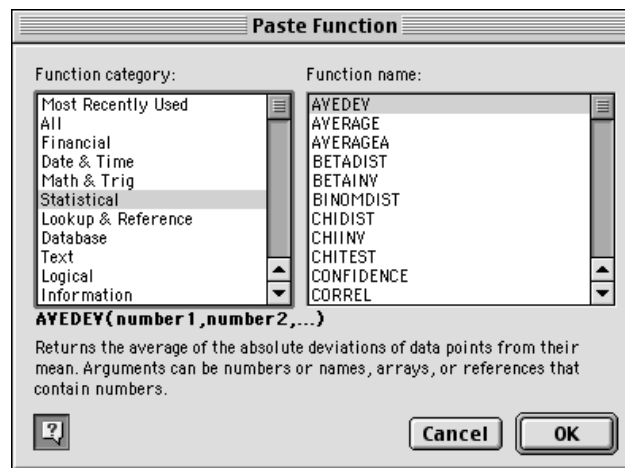


Figure 1.23: Paste Function



Figure 1.24: Formula Palette—Default

to generate a drop-down list of your 10 most recently used functions or you may select **More functions...**. If you select the latter, then the **Paste Function** dialog box appears. An OK (checkmark symbol) and a cancel (X) button appear to the right of this arrow. Click the checkmark and the formula is entered into the active cell. Click the cancel to discard the formula without making changes.

Recommendation. The **Paste Function** button on the Standard Toolbar duplicates the actions of the **Formula Palette**. Since Excel formulas start with an equal (=) sign, we recommend that you begin your formulas by pressing the (=) symbol on the Formula Toolbar instead of using the Paste Function. This activates the **Formula Palette**, and you can either type the formula by hand into the Formula Bar or order up a function from the **Paste Function** box, if required. Experienced users of Excel often **customize** the **Standard Toolbar** and replace the Paste Function button with some other one.

The Five-Number Summary

Example 1.4 continued. Find the five-number summary {minimum, first quartile, median, third quartile, maximum} for the calorie distribution of the beef hot dogs data shown in Table 1.9 page 59 of the text.

	A	B	C	D
1	Beef hot dogs		Five-number summary	
2				
3	Calories	Min	=MIN(A4:A23)	111
4	186	Q1	=QUARTILE(A4:A23,1)	140.5
5	181	Med	=MEDIAN(A4:A23)	152.5
6	176	Q3	=QUARTILE(A4:A23,3)	177.25
7	149	Max	=MAX(A4:A23)	190
8	184			
9	190			
10	158			
11	139			
12	175			
13	148			
14	152			
15	111			
16	141			
17	153			
18	190			
19	157			
20	131			
21	149			
22	135			
23	132			

Figure 1.25: Five-Number Summary

Solution

- Copy the beef calories data from Figure 1.21 onto a new worksheet and enter the labels “Min,” “Q1,” “Med,” “Q3,” and “Max” as shown in cells B3:B7 of your worksheet (Figure 1.25).
- Click the equal (=) symbol on the **Formula Bar** to start the **Formula Palette** and use the drop-down list to select **More functions...** In the **Paste Function** dialog box, select **Statistical** from the left and scroll down and select **QUARTILE** on the right. Click OK.
- The **Formula Palette** dialog box appears. Move it out of the way and enter the data **Array** by selecting cells A4:A23 with your mouse (or more mundanely by typing A4:A23 into the dialog box). Click in the text area for **Quart** and type “1” to indicate the first quartile. The completed formula appears in the **Formula Toolbar** and value of the formula 140.5 shows in the dialog box (Figure 1.26). Click OK and the value 140.5 is printed in C5.
- Continue in this fashion using the Formula Palette to complete the five number summary. Of course you can still enter the formulas by hand in the Formula Bar once you are familiar with them. Cells C4:C7 show the syntax while the values are in D4:D7.

The five-number summary is {111, 140.5, 152.5, 177.25, 190}. Note that Excel uses a slightly different definition of quartiles for a finite data set than the text.

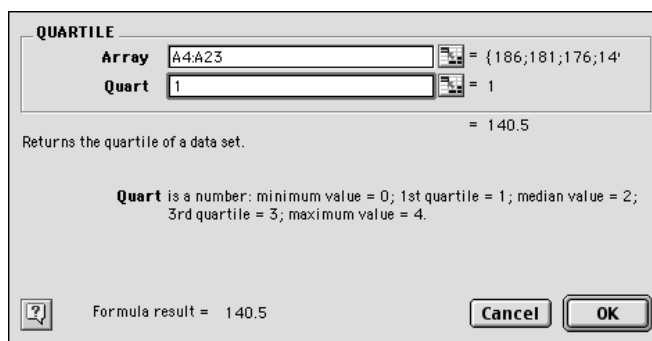


Figure 1.26: Quartile Formula

1.3 The Normal Distribution

Areas under a normal curve can be found using the `NORMDIST` function. The syntax is `= NORMDIST(x, μ, σ , cumulative)`, where μ is the mean and σ is the standard deviation. The parameter `cumulative` indicates whether the density (set `cumulative = "false"` or `"0"`) or whether the cumulative distribution (set `cumulative = "true"` or `"1"`) is wanted. The formula `= NORMDIST($x, \mu, \sigma, 1$)` returns $F(x)$, which is the area to the left of x under an $N(\mu, \sigma)$ density and can be used to produce a table of normal areas as found in many statistics texts. Another formula, `= NORMINV(p, μ, σ)`, returns the inverse $F^{-1}(p)$ of the cumulative, that is a value x such that the area to the left of x is the specified p . For $N(0, 1)$, use `NORMSDIST` and `NORMSINV` instead.

Normal Distribution Calculations

Example 1.5. (Example 1.26, page 75 in the text.) The NCAA requires Division I athletes to score at least 820 on the combined mathematics and verbal parts of the SAT exam in order to compete in their first year college year. In 2000, the scores of the more than one million students taking the SATs were approximately normal with mean 1019 and standard deviation 209. What percent of all students had SAT scores of at least 820?

Solution

Click on a cell (activate it) where you want to locate the answer, say A1. Since we are looking for the area to the right of the point 820, the syntax is

$$\Phi(x) = 1 - \text{NORMDIST}(x, \mu, \sigma, 1)$$

so enter the formula `= 1 - NORMDIST(820, 1019, 209, 1)` since the upper tail area is wanted. The answer 0.829 appears in cell A1. (Users of **Excel 5/95** may also

use the **Function Wizard** while users of **Excel 97/98** may use the **Formula Palette** instead to enter the formula.)

Example 1.6. (Example 1.27, page 76 in the text.) The NCAA considers a student a “partial qualifier” eligible to practice and receive an athletic scholarship, but not to compete, if the combined SAT score is at least 720. What percent of all students who take the SAT would be partial qualifiers?

Solution

Because we want the area between 720 and 820 the syntax is

$$\text{NORMDIST}(820, 1019, 209, 1) - \text{NORMDIST}(720, 1019, 209, 1) = 0.094.$$

Example 1.7. (Example 1.28, page 77 in the text.) Scores on the SAT verbal test in recent years follow approximately the $N(505, 110)$ distribution. How high must a student score in order to place in the top 10% of all students taking the SAT?

Solution

We are looking for the SAT score with area 0.1 to its right under the normal curve, equivalently with area 0.9 to its left. Therefore the Excel formula is the inverse normal

$$\Phi^{-1}(x) = \text{NORMINV}(p, \mu, \sigma)$$

for the inverse of the cumulative normal distribution. Enter the formula = **NORMINV**(0.90, 505, 110) and read off 645.97, the 90th percentile of the SAT scores.

Graphing the Normal Curve

By combining the **Chart Wizard** and the **NORMDIST** function, we can create a graph of any normal curve. *In fact, the procedure described here can be used to plot the graph of any function Excel can evaluate.*

Constructing a Graph of the Standard Normal Curve

1. Enter the labels z and $f(z)$ in cells A2, B2. (See Figure 1.27.) $f(z)$ will represent the density or height of the normal curve at the point z .
2. Enter -3.5 and -3.4 in cells A3 and A4, respectively. These two points mark the ends of the range of values over which the normal density will be graphed. Next we create a column of z values at which the standard normal density will be calculated. Select A3:A4, check the fill handle in the lower right corner of A4, and drag to cell A73 to fill the column with decreasing values of z decremented by 0.1. Format the values with two decimal places.

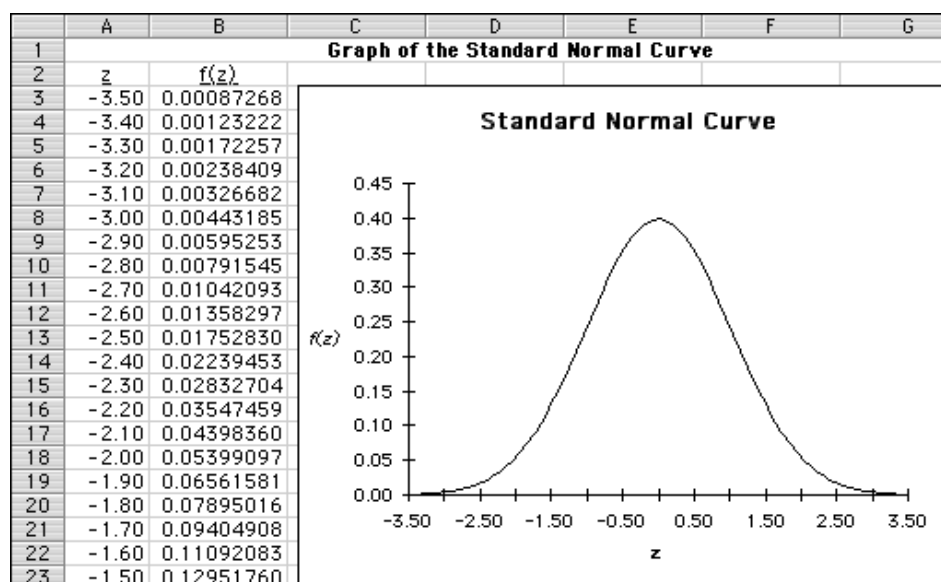


Figure 1.27: Graphing a Normal Density

3. Select cell B3 and enter = `NORMDIST(A3, 0, 1, 0)` in the **Formula Bar**. Cell B3 now contains the value 0.00087268, the standard normal density evaluated at $z = -3.50$.
4. Select cell B3, click the fill handle, and drag down to B73. This copies the formula you just entered in B3 into cells B4:B73 relative to the corresponding cell references in column A. Column B is filled with values $f(z)$ of the standard normal density corresponding to each value of z in column A.
5. **Users of Excel 5/95.** Select cells A2:B73, click the **Chart Wizard** button, and then click in cell C2 and drag to I26 to locate the graph. A dialog box **Chart Wizard – Step 1 of 5** appears.
 - In Step 1 you are given an opportunity to correct or confirm your range.
 - In Step 2 select **XY (Scatter)** chart.
 - In Step 3 select format **6**.
 - In Step 4 select the radio button **Columns** for Data Series, enter “1” for Column for Category (X) Axis Labels, and enter “1” for Row for Legend Text.
 - In Step 5 select the radio button **No** for Add a Legend?, type “Standard Normal Curve” as the Chart Title, and type z and $f(z)$ for Category (X) and Value (Y) titles, respectively. Finally, click Finish.

Users of Excel 97/98/2000/2001. Select cells A2:B73 and click the **ChartWizard**. A dialog box **ChartWizard – Step 1 of 4 – Chart Type** appears.

- In Step 1 select **XY (Scatter)** for Chart Type and the lower right Chart sub-type **Scatter without markers**.
 - In Step 2 under the **Data Range** tab, the range will already be indicated and the Series radio button for **Columns** will be selected. You may edit the range if it is incorrect. Under the **Series** tab no changes are necessary.
 - In Step 3 under the **Titles** tab, type “Standard Normal Curve” as the Chart Title, z for Value (X) Axis, and $f(z)$ for Value (Y) Axis. Under the **Axes** tab, both check boxes should be selected. Under the **Gridlines** tab, clear all check boxes. Under the **Legend** tab, clear the Show legend. Finally, under the **Data Labels** tab, select the radio button **None**.
 - In Step 4 embed the graph in the current workbook by selecting the radio button **As object in**. Finally, click Finish.
6. Activate the graph for editing and format the display as you wish to present it using the editing features discussed previously.

Constructing a Normal Table

It is very easy in Excel to produce a table of normal areas. This method described below can be adapted to produce tables of other continuous distributions.

Figure 1.28 gives areas under a standard normal curve for values of $z \geq 0$. Here is the procedure for producing this table.

1. Enter the label and values in column A and row 3. Column A is the first decimal while row 1 is the second decimal of z .
2. Enter the formula `=NORMDIST($A4 + B$3)` in cell B4. (Remember that the \$ sign prefix makes the corresponding row or column label absolute.) Select cell B4, click the fill handle in the lower right corner of B4, and drag to K4.
3. Select cells B4:K4, click the fill handle in the lower right corner of K4, and drag to K43 to fill the block B4:K43.

Normal Quantile Plots

Excel does not provide a normal quantile (probability) plot, but it is very easy to construct such a graph. We defer this to the next chapter since it is an application of a scatterplots.

	A	B	C	D	E	F	G	H	I	J	K
1	Normal Table Constructed in Excel										
2											
3	z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
4	0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
5	0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
6	0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
7	0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
8	0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
9	0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
10	0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
11	0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
12	0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
13	0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
14	1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
15	1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
16	1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
17	1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
18	1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
19	1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
20	1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
21	1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
22	1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
23	1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
24	2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
25	2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
26	2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
27	2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
28	2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
29	2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
30	2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
31	2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
32	2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
33	2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
34	3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
35	3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
36	3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
37	3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
38	3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
39	3.50	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
40	3.60	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
41	3.70	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
42	3.80	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
43	3.90	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Figure 1.28: Normal Table

1.4 Boxplots

A boxplot is one of the most important exploratory tools available to the data analyst. Unfortunately this tool is not part of the Analysis ToolPak. The Microsoft Personal Support Center does have a Web page “XL: How to Create a BoxPlot – Box and Whisker Chart (Q155130)” located at

<http://support.microsoft.com/support/kb/articles/q155/1/30.asp>

with instructions for creating a reasonable boxplot using a **Volume-Open-High-Low-Close** Chart but the sequence of steps is complicated enough to discourage a student from using this tool regularly. We have provided a macro *boxplot.xls* which will produce a boxplot of a single data set or side-by-side boxplots of multiple data sets. This boxplot can be downloaded from the Freeman web site.

Example 1.8. (See Exercise 1.49, page 57 in the text.) Figure 1.29 shows the scores on the Survey of Study Habits and Attitudes for 18 first-year college women. Find the five-number summaries for both sets of SSHA scores and make side-by-side modified boxplots for the two distributions.

	A	B	C	D	E
1				Women	Men
2				154	108
3				109	140
4				137	114
5				115	91
6				152	180
7				140	115
8				154	126
9				178	92
10				101	169
11				103	146
12				126	109
13				126	132
14				137	75
15				165	88
16				165	113
17				129	151
18				200	70
19				148	115
20					187
21					104

Figure 1.29: Boxplot Input

Solution. Prepare the data in columns on a worksheet as in Figure 1.29. The data need not be in adjacent columns nor need the number of observations in each data set be the same. If the data are on the same worksheet as the boxplot macro then click the button **Run Boxplot** to bring up the Boxplot dialog box Figure 1.30 where you select the data. If the data sets have the same column lengths



Figure 1.30: Boxplot Dialog Box

then an entire block can be selected. Otherwise each column needs to be selected separately. In this example, first select cells D1:D19, then hold down the **Control (Windows)** or **Command (Macintosh)** keys and select E1:E21. Check the box “First Row Contains Labels”. You can also label the data axis (Y-axis) if desired. The boxplot, Figure 1.30, will appear in the next worksheet, named “Box Plot 1”, in the same workbook together with the five number summaries for the data sets. Outliers are also plotted.

The data need not be on the same sheet as the boxplot button as long as the boxplot.xls file is open. If the data is in another workbook then the macro is run as follows. From the Menu Bar select **Tools – Macro...** to open the **Macro** dialog box and select **Run** which opens the Boxplot dialog box. Complete this box as before.

Figure 1.31 shows the resulting boxplot together with the five number summary. Outlier values are also printed indicating that the value 200 among the women’s scores is an outlier according to the $1.5 \times IQR$ rule.

Note that Excel uses a slightly different definition for the quartiles than is given in the text. For the women’s scores the quartiles calculated are the same but not for the men’s scores. This may affect judgment on which observations would be considered outliers.

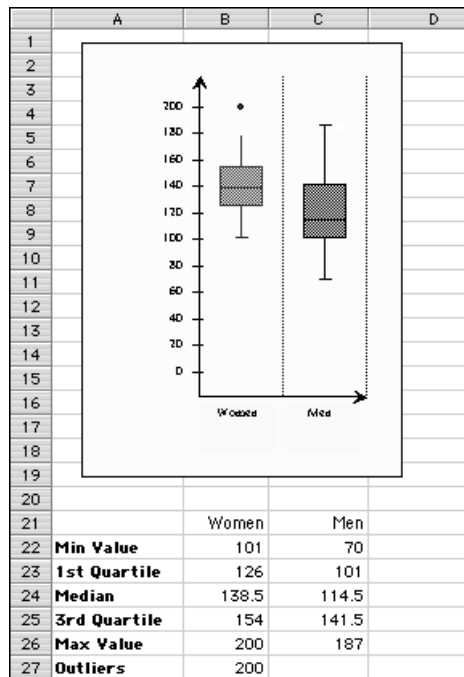


Figure 1.31: Boxplot Output