# Chapter 3. Displaying and Describing Categorical Data

## DATA DESK

To make a bar chart or pie chart, select the variable. In the **Plot** menu, choose **Bar Chart or Pie Chart**.

To make a frequency table, in the **Calc** menu choose **Frequency Table**.

**COMMENTS**

These commands treat the data as categorical even if they are numerals. If you select a quantitative variable by mistake, you’ll see an error message warning of too many categories.

## EXCEL

First make a pivot table (Excel’s name for a frequency table). From the **Data** menu, choose **Pivot Table** and **Pivot Chart Report**.

When you reach the Layout window, drag your variable to the row area and drag your variable again to the data area. This tells Excel to count the occurrences of each category. Once you have an Excel pivot table, you can construct bar charts and pie charts.

Click inside the Pivot Table.

Click the Pivot Table Chart Wizard button. Excel creates a bar chart.

A longer path leads to a pie chart; see your Excel documentation.

**COMMENTS**

Excel uses the pivot table to specify the category names and find counts within each category. If you already have that information, you can proceed directly to the Chart Wizard.

## EXCEL 2007

To make a bar chart:

- Select the variable in Excel you want to work with.
- Choose the **Column** command from the Insert tab in the Ribbon.
- Select the appropriate chart from the drop-down dialog.

To change the bar chart into a pie chart:

- Right-click the chart and select **Change Chart Type**...from the menu. The Chart type dialog opens.
- Select a pie chart type.
- Click the **OK** button. Excel changes your bar chart into a pie chart.

## JMP

JMP makes a bar chart and frequency table together. From the **Analyze** menu, choose **Distribution**.

In the Distribution dialog, drag the name of the variable into the empty variable window beside the label “Y, Columns”; click **OK**.

To make a pie chart, choose **Chart** from the **Graph** menu.

In the Chart dialog, select the variable name from the Columns list, click on the button labeled “Statistics,” and select “N” from the drop-down menu.

Click the **Categories, X, Levels** button to assign the same variable name to the X-axis.

Under Options, click on the second button—labeled “Bar Chart”—and select “Pie” from the drop-down menu.

## MINITAB

To make a bar chart, choose **Bar Chart** from the **Graph** menu.

Select “Counts of unique values” in the first menu, and select “Simple” for the type of graph. Click **OK**.

In the Chart dialog, enter the name of the variable that you wish to display in the box labeled “Categorical variables.” Click **OK**.
To make a bar chart, open the Chart Builder from the Graphs menu. Click the Gallery tab. Choose Bar Chart from the list of chart types. Drag the appropriate bar chart onto the canvas. Drag a categorical variable onto the x-axis drop zone. Click OK.

COMMENTS
A similar path makes a pie chart by choosing Pie chart from the list of chart types.

The TI-Nspire Handheld does not display plots for categorical variables.

The TI-89 won’t do displays for categorical variables.

**Chapter 4. Displaying and Summarizing Quantitative Data**

**DATA DESK**
To make a histogram:
- Select the variable to display.
- In the Plot menu, choose Histogram.

To calculate summaries:
- In the Calc menu, open the summaries submenu. Options offer separate tables, a single unified table, and other formats.

**COMMENTS**
Excel’s Data Analysis add-in does offer something called a histogram, but it just makes a crude frequency table, and the Chart Wizard cannot then create a statistically appropriate histogram. The DDXL add-in provided on our DVD adds these and other capabilities to Excel. Excel’s STDEV function should not be used for data values larger in magnitude than 100,000 or for lists of more than a few thousand values. It is programmed with an unstable formula that can generate rounding errors when these limits are exceeded.

**EXCEL**
Excel cannot make histograms or dotplots without a third-party add-in. To calculate summaries, click on an empty cell. Type an equal sign and choose “Average” from the popup list of functions that appears to the left of the text-editing box. Enter the data range in the box that says “Number 1.” Click the OK button. To compute the standard deviation of a column of data directly, use the STDEV from the popup list of functions in the same way.

**EXCEL 2007**
In Excel 2007 there is another way to find some of the standard summary statistics. For example, to compute the mean:
- Click on an empty cell.
- Go to the Formulas tab in the Ribbon. Click on the drop down arrow next to “AutoSum” and choose “Average”.
- Enter the data range in the formula displayed in the empty box you selected earlier.
- Press Enter. This computes the mean for the values in that range.

To compute the standard deviation:
- Click on an empty cell.
- Go to the Formulas tab in the Ribbon and click the drop down arrow next to “AutoSum” and select “More functions…”.
- In the dialog window that opens, select “STDEV” from the list of functions and click OK. A new dialog window opens. Enter a range of fields into the text fields and click OK.
Excel 2007 computes the standard deviation for the values in that range and places it in the specified cell of the spreadsheet.

**JMP**
To make a histogram and find summary statistics:
- Choose Distribution from the Analyze menu.
- In the Distribution dialog, drag the name of the variable that you wish to analyze into the empty window beside the label “Y, Columns.”

- Click OK. JMP computes standard summary statistics along with displays of the variables.
**Minitab**

To make a histogram:
- Choose **Histogram** from the **Graph** menu.
- Select “Simple” for the type of graph and click **OK**.
- Enter the name of the quantitative variable you wish to display in the box labeled “Graph variables.” Click **OK**.

To calculate summary statistics:
- Choose **Basic statistics** from the **Stat** menu. From the **Basic Statistics** submenu, choose **Display Descriptive Statistics**.
- Assign variables from the variable list box to the Variables box. MINITAB makes a Descriptive Statistics table.

**SPSS**

To make a histogram in SPSS open the Chart Builder from the **Graphs** menu.
- Click the **Gallery** tab.
- Choose **Histogram** from the list of chart types.
- Drag the histogram onto the canvas.
- Drag a scale variable to the y-axis drop zone.
- Click **OK**.

To calculate summary statistics:
- Choose **Explore** from the **Descriptive Statistics** submenu of the **Analyze** menu. In the Explore dialog, assign one or more variables from the source list to the Dependent List and click the **OK** button.

**TI-Nspire**

To plot a histogram using a named list, press ` several times so that the entire list is highlighted. Press ` for Data, and ` for Quick Graph. Then press ` for Plot Type, and ` for Histogram.

To create the plot on a full page, press ` for Data & Statistics. Move the cursor to “Click to add variable,” and then press ` and select the list name. Then press ` for Plot Type, and ` for Histogram.

**TI-89**

To make a histogram:
- Select ` (Plots), then 1: **Plot Setup**. Select a plot and press ` to define it.
- Select plot type 4: **Histogram**. Use VAR-LINK to select the data list.
- Enter a number for the histogram bucket (bar) width.
- Press ` to complete the plot definition. Press ` to display the histogram.
- Press ` to adjust the window appropriately, then press ` (Graph).

To calculate summary statistics:
- To compute summary statistics, press ` (Calc). Input the name of the list using VAR-LINK. Press `.
- Use the down arrow to scroll through the output.
- To create a boxplot, press ` (Plots) then ` (Graph). Select a plot to define and press ` (Graph). Select either 3: **Box Plot** or 4: **Mod Box**.

**Plot** (to identify outliers). Select the mark type of your choice (for outliers). Press ` (Graph) to display the graph.

**COMMENTS**

If the data are stored as a frequency table (say, with data values in list1 and frequencies in list2), change Use Freq and Categories to YES and use VAR-LINK to select list2 as the frequency variable on the plot definition screen.

If the data are stored as a frequency table (say, with data values in list1 and frequencies in list2), use VAR-LINK to select list2 as the frequency variable in 1-Var Stats.

For the plot, change Use Freq and Categories to YES and use VAR-LINK to select list2 as the frequency variable on the plot definition screen.

---

**Chapter 5. Understanding and Comparing Distributions**

There are two ways to organize data when we want to compare groups. Each group can be in its own variable (or list, on a calculator). In this form, the experiment comparing coffee cups would have four lists, one for each type of cup:

<table>
<thead>
<tr>
<th>CUPPS</th>
<th>SIGG</th>
<th>Nissan</th>
<th>Starbucks</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>1.5</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>17.5</td>
<td>7</td>
<td>20.5</td>
<td>15.5</td>
</tr>
<tr>
<td>11</td>
<td>0.5</td>
<td>12.5</td>
<td>6</td>
</tr>
<tr>
<td>6.5</td>
<td>6</td>
<td>24.5</td>
<td>6</td>
</tr>
</tbody>
</table>
But there's another way to think about and organize the data. What is the variable of interest (the What) in this experiment? It's the number of degrees lost by the water in each cup. And the Who is each time she tested a cup. We could gather all the temperature values into one variable and put the names of the cups in a second variable listing the individual results, one on each row. Now the Who is clearer—it's an experimental run, one row of the table. Most statistics packages prefer data on groups organized in this way.

That's actually the way we've thought about the wind speed data in this chapter, treating wind speeds as one variable and the groups (whether seasons, months, or days) as a second variable.

<table>
<thead>
<tr>
<th>Container</th>
<th>Temperature Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUPPS</td>
<td>6</td>
</tr>
<tr>
<td>CUPPS</td>
<td>6</td>
</tr>
<tr>
<td>CUPPS</td>
<td>6</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Nissan</td>
<td>2</td>
</tr>
<tr>
<td>Nissan</td>
<td>1.5</td>
</tr>
<tr>
<td>Nissan</td>
<td>2</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Container</th>
<th>Temperature Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIGG</td>
<td>12</td>
</tr>
<tr>
<td>SIGG</td>
<td>16</td>
</tr>
<tr>
<td>SIGG</td>
<td>9</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Starbucks</td>
<td>13</td>
</tr>
<tr>
<td>Starbucks</td>
<td>7</td>
</tr>
<tr>
<td>Starbucks</td>
<td>7</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

**DATA DESK**

If the data are in separate variables, select the variables and choose **Boxplot side by side** from the Plot menu. The boxes will appear in the order in which the variables were selected.

If the data are a single quantitative variable and a second variable holding group names, select the quantitative variable as Y and the group variable as X. Then choose **Boxplot y by x** from the Plot menu. The boxes will appear in alphabetical order by group name.

Data Desk offers options for assessing whether any pair of medians differ.

**EXCEL**

Excel cannot make boxplots.

**COMMENT**

The DDXL add-on provided on the DVD adds the ability to make boxplots to Excel.

**JMP**

Choose **Fit y by x**. Assign a continuous response variable to Y, **Response** and a nominal group variable holding the group names to X, **Factor**, and click **OK**. JMP will offer (among other things) dotplots of the data. Click the red triangle and, under **Display Options**, select Boxplots. **Note:** If the variables are of the wrong type, the display options might not offer boxplots.

**MINITAB**

Choose **Boxplot...** from the Graph menu. If your data are in the form of one quantitative variable and one group variable, choose **One Y and with Groups**. If your data are in separate columns of the worksheet, choose **Multiple Y’s**.

**SPSS**

To make a boxplot in SPSS, open the **Chart Builder** from the Graphs menu. Click the **Gallery** tab. Choose **Boxplot** from the list of chart types. Drag a single or 2-D (side-by-side) boxplot onto the canvas.

Drag a scale variable to the y-axis drop zone. To make side-by-side boxplots, drag a categorical variable to the x-axis drop zone. Click **OK**.

**TI-NSPRIRE**

To compute summary statistics using a named list, press **C**, **1** for Calculator, **C**, **2** for Statistics, **C** for Stat Calculations, and **C** for One-Variable Statistics. Complete the dialog boxes. To create a box plot using a named list, press **C** several times so that the entire list is highlighted. Press **C**, **1** for Data, and **C** for Quick Graph. Then press **C**, **1** for Plot Type, and **C** for Box Plot.

To create the plot on a full page, press **C**, and then press **C** for Data & Statistics. Move the cursor to “Click to add variable,” and then press **C** and select the list name. Then press **C**, **1** for Plot Type, and **C** for Box Plot.
Chapter 6. The Standard Deviation as a Ruler and the Normal Model

### DATA DESK

To make a “Normal Probability Plot” in Data Desk,
- Select the Variable.
- Choose Normal Prob Plot from the Plot menu.

**COMMENTS**
Data Desk places the ordered data values on the vertical axis and the Normal scores on the horizontal axis.

### EXCEL

Excel offers a “Normal probability plot” as part of the Regression command in the Data Analysis extension, but (as of this writing) it is not a correct Normal probability plot and should not be used.

### JMP

To make a “Normal Quantile Plot” in JMP,
- Make a histogram using Distributions from the Analyze menu.
- Click on the drop-down menu next to the variable name.
- Choose Normal Quantile Plot from the drop-down menu.
- JMP opens the plot next to the histogram.

**COMMENTS**
JMP places the ordered data on the vertical axis and the Normal scores on the horizontal axis. The vertical axis aligns with the histogram’s axis, a useful feature.

### MINITAB

To make a “Normal Probability Plot” in MINITAB,
- Choose Probability Plot from the Graph menu.
- Select “Single” for the type of plot. Click OK.
- Enter the name of the variable in the “Graph variables” box. Click OK.

**COMMENTS**
MINITAB places the ordered data on the horizontal axis and the Normal scores on the vertical axis.

### SPSS

To make a Normal “P-P plot” in SPSS,
- Choose P-P from the Graphs menu.
- Select the variable to be displayed in the source list.
- Click the arrow button to move the variable into the target list.
- Click the OK button.

**COMMENTS**
SPSS places the ordered data on the horizontal axis and the Normal scores on the vertical axis. You may safely ignore the options in the P-P dialog.

### TI-NSPIRE

To create a normal probability plot using a named list, press several times so that the entire list is highlighted. Press for Data, and for Quick Graph. Then press for Plot Type, and for Normal Probability Plot.

To create the plot on a full page, press, and then for Plot Type, and for Normal Probability Plot.

To compute the area under a normal curve, press for Calculator, for Probability, for Distributions, and for Normal Cdf. Complete the dialog box.

To compute the value for a given percentile, press for Calculator, for Probability, for Distributions, for Inverse Normal. Complete the dialog box.
Chapter 7. Scatterplots, Association, and Correlation

DATA DESK

To make a scatterplot of two variables, select one variable as Y and the other as X and choose Scatterplot from the Plot menu. Then find the correlation by choosing Correlation from the scatterplot’s HyperView menu. Alternatively, select the two variables and choose Pearson Product-Moment from the Correlations submenu of the Calc menu.

EXCEL

To make a Scatterplot with the Excel Chart Wizard:
- Click on the Chart Wizard Button in the menu bar. Excel opens the Chart Wizard’s Chart Type Dialog window.
- Make sure the Standard Types tab is selected, and select XY (Scatter) from the choices offered.
- Specify the scatterplot without lines from the choices offered in the Chart subtype selections. The Next button takes you to the Chart Source Data dialog.
- If it is not already frontmost, click on the Data Range tab, and enter the data range in the space provided.
- By convention, we always represent variables in columns. The Chart Wizard refers to variables as Series. Be sure the Column option is selected.
- Excel places the leftmost column of those you select on the x-axis of the scatterplot. If the column you wish to see on the x-axis is not the leftmost column in your spreadsheet, click on the Series tab and edit the specification of the individual axis series.
- Click the Next button. The Chart Options dialog appears.
- Select the Titles tab. Here you specify the title of the chart and names of the variables displayed on each axis.
- Type the chart title in the Chart title: edit box.
- Type the x-axis variable name in the Value (X) Axis: edit box. Note that you must name the columns correctly here. Naming another variable will not alter the plot, only mislabel it.
- Type the y-axis variable name in the Value (Y) Axis: edit box.
- Click the Next button to open the chart location dialog.
- Select the As new sheet: option button.
- Click the Finish button.
- Often, the resulting scatterplot will not be useful. By default, Excel includes the origin in the plot even when the data are far from zero. You can adjust the axis scales. To change the scale of a plot axis in Excel:
  - Double-click on the axis. The Format Axis Dialog appears.
  - If the scale tab is not the frontmost, select it.
  - Enter new minimum or new maximum values in the spaces provided. You can drag the dialog box over the scatterplot as a straightedge to help you read the maximum and minimum values on the axes.
- Click the OK button to view the rescaled scatterplot.
- Follow the same steps for the x-axis scale.

Compute a correlation in Excel with the CORREL function from the drop-down menu of functions. If CORREL is not on the menu, choose More Functions and find it among the statistical functions in the browser. In the dialog that pops up, enter the range of cells holding one of the variables in the space provided. Enter the range of cells for the other variable in the space provided.

COMMENTS

Normal models strictly go to infinity on either end, which is 1EE99 on the calculator. In practice, any “large” number will work. For example, the percentage of the Normal model over two standard deviations above the mean can use Lower Value 2 and Upper Value 99. To find area more than 2 standard deviations below the mean, use Lower Value –99, and Upper value –2.
EXCEL 2007

To make a scatterplot in Excel 2007:
- Select the columns of data to use in the scatterplot. You can select more than one column by holding down the control key while clicking.
- In the Insert tab, click on the Scatter button and select the Scatter with only Markers chart from the menu.
- Unfortunately, the plot this creates is often statistically useless.
- To make the plot useful, we need to change the display:
  - With the chart selected click on the Gridlines button in the Layout tab to cause the Chart Tools tab to appear.
  - Within Primary Horizontal Gridlines, select None. This will remove the gridlines from the scatterplot.
  - To change the axis scaling, click on the numbers of each axis of the chart, and click on the Format Selection button in the Layout tab.
  - Select the Fixed option instead of the Auto option, and type a value more suited for the scatterplot. You can use the popup dialog window as a straightedge to approximate the appropriate values.

Excel 2007 automatically places the leftmost of the two columns you select on the x-axis, and the rightmost one on the y-axis. If that's not what you'd prefer for your plot, you'll want to switch them.
- To switch the X and Y-variables:
  - Click the chart to access the Chart Tools tabs.
  - Click on the Select Data button in the Design tab.
  - In the popup window’s Legend Entries box, click on Edit.
  - Highlight and delete everything in the Series X Values line, and select new data from the spreadsheet. (Note that selecting the column would inadvertently select the title of the column, which would not work well here.)
  - Do the same with the Series Y Values line.
  - Press OK, then press OK again.

JMP

To make a scatterplot and compute correlation, choose Fit Y by X from the Analyze menu.
- In the Fit Y by X dialog, drag the Y variable into the “Y, Response” box, and drag the X variable into the “X, Factor” box. Click the OK button.

Once JMP has made the scatterplot, click on the red triangle next to the plot title to reveal a menu of options. Select Density Ellipse and select .95. JMP draws an ellipse around the data and reveals the Correlation tab. Click the blue triangle next to Correlation to reveal a table containing the correlation coefficient.

MINITAB

To make a scatterplot, choose Scatterplot from the Graph menu. Choose “Simple” for the type of graph. Click OK. Enter variable names for the Y-variable and X-variable into the table. Click OK.

To compute a correlation coefficient, choose Basic Statistics from the Stat menu. From the Basic Statistics submenu, choose Correlation. Specify the names of at least two quantitative variables in the “Variables” box. Click OK to compute the correlation table.

SPSS

To make a scatterplot in SPSS, open the Chart Builder from the Graphs menu. Then:
- Click the Gallery tab.
- Choose Scatterplot from the list of chart types.
- Drag the scatterplot onto the canvas.
- Drag a scale variable you want as the response variable to the y-axis drop zone.
- Drag a scale variable you want as the factor or predictor to the x-axis drop zone.
- Click OK.

To compute a correlation coefficient, choose Correlate from the Analyze menu. From the Correlate submenu, choose Bivariate. In the Bivariate Correlations dialog, use the arrow button to move variables between the source and target lists. Make sure the Pearson option is selected in the Correlation Coefficients field.

TI-Nspire

To create a scatterplot using named lists, press ⇧ several times so that the first list is highlighted. Then press ▼ so that the second list is highlighted. Press  أعلن,  for Data, and  for Quick Graph.

To create the plot on a full page, press  then  for Data & Statistics. Move the cursor to “Click to add variable,” and then press  and select the list name. Repeat for the other axis.

To find the correlation, press  for Calculator,  for Statistics,  for Stat Calculations, and  for Linear Regression. Complete the dialog boxes.
Appendix B

Guide to Statistical Software

To create a scatterplot, press \( \text{Fr} \) (Plots). Select choice 1: Plot Setup. Select a plot to define and press \( \text{Fn} \). Select Plot Type 1: Scatter. Select a mark type. Specify the lists where the data are stored as Xlist and Ylist, using VAR-LINK. Press \( \text{Enter} \) to finish. Press \( \text{Fr} \) to display the plot. To find the correlation, press \( \text{Fr} \) (CALC), then arrow to 3: Regressions, press the right arrow, and select 1:LinReg(a+bx).

**COMMENTS**

Notice that if you TRACE (press \( \text{Fr} \)) the scatterplot, the calculator will tell you the x- and y-value at each point.

**TI-89**

**Chapter 8. Linear Regression**

**DATA DESK**

Select the y-variable and the x-variable. In the Plot menu choose Scatterplot, from the scatterplot HyperView menu, choose Add Regression Line to display the line. from the HyperView menu, choose Regression to compute the regression.

**COMMENTS**

Alternatively, find the regression first with the Regression command in the Calc menu. Click on the x-variable’s name to open a menu that offers the scatterplot.

**EXCEL**

Make a scatterplot of the data. With the scatterplot front-most, select Add Trendline... from the Chart menu. Click the Options tab and select Display Equation on Chart. Click OK.

**COMMENTS**

The computer section for Chapter 7 shows how to make a scatterplot. We don’t repeat those steps here.

**EXCEL 2007**

- Click on a blank cell in the spreadsheet.
- Go to the Formulas tab in the Ribbon and click More Functions ➞ Statistical.
- Choose the CORREL function from the drop-down menu of functions.
- In the dialog that pops up, enter the range of one of the variables in the space provided.
- Enter the range of the other variable in the space provided.
- Click OK.

**COMMENTS**

The correlation is computed in the selected cell. Correlations computed this way will update if any of the data values are changed. Before you interpret a correlation coefficient, always make a scatterplot to check for nonlinearity and outliers. If the variables are not linearly related, the correlation coefficient cannot be interpreted.

**JMP**

Choose Regression from the Stat menu. From the Regression submenu, choose Fitted Line Plot. In the Fitted Line Plot dialog, click in the Response Y box, and assign the y-variable from the Scatterplot window, click on the red triangle beside the heading labeled “Bivariate Fit...” and choose “Fit Line.” JMP draws the least squares regression line on the scatterplot and displays the results of the regression in tables below the plot.

**MINITAB**

Choose Interactive from the Graphs menu. From the interactive Graphs submenu, choose Scatterplot. In the Create Scatterplot dialog, drag the y-variable into the y-axis target, and the x-variable into the x-axis target. Click on the Fit tab. Choose Regression from the Method popup menu. Click the OK button.

**SPSS**

Variable list. Click in the Predictor X box, and assign the x-variable from the Variable list. Make sure that the Type of Regression Model is set to Linear. Click the OK button.
Chapter 9. Regression Wisdom

**DATA DESK**

Click on the HyperView menu on the Regression output table. A menu drops down to offer scatterplots of residuals against predicted values, Normal probability plots of residuals, or just the ability to save the residuals and predicted values.

Click on the name of a predictor in the regression table to be offered a scatterplot of the residuals against that predictor.

**COMMENTS**

If you change any of the variables in the regression analysis, Data Desk will offer to update the plots of residuals.

**EXCEL**

The Data Analysis add-in for Excel includes a Regression command. The dialog box it shows offers to make plots of residuals.

**COMMENTS**

Do not use the Normal probability plot offered in the regression dialog. It is not what it claims to be and is wrong.

**JMP**

From the Analyze menu, choose Fit Y by X. Select Fit Line. Under Linear Fit, Select Plot Residuals. You can also choose to Save Residuals. Subsequently, from the Distribution menu, choose Normal quantile plot or histogram for the residuals.

**MINITAB**

From the Stat menu, choose Regression. From the Regression submenu, select Regression again. In the Regression dialog, enter the response variable name in the “Response” box and the predictor variable name in the “Predictor” box. To specify saved results, in the Regression dialog, click Storage. Check “Residuals” and “Fits.” Click OK. To specify displays, in the Regression dialog, click Graphs. Under “Residual Plots,” select “Individual plots” and check “Residuals versus fits.” Click OK. Now back in the Regression dialog, click OK. Minitab computes the regression and the requested saved values and graphs.

**SPSS**

From the Analyze menu, choose Regression. From the Regression submenu, choose Linear. After assigning variables to their roles in the regression, click the “Plots...” button. In the Plots dialog, you can specify a Normal probability plot of residuals and scatterplots of various versions of standardized residuals and predicted values. A plot of *ZRESID against *PRED will look most like the residual plots we’ve discussed. SPSS standardizes the residuals by dividing by their standard deviation. (There's no need to subtract their mean; it must be zero.) The standardization doesn't affect the scatterplot.

**TI-NSPRIRE**

To plot and find the equation of the regression line, first create a scatterplot. Using named lists, press several times so that the first list is highlighted. Then press so that the second list is highlighted. Press for Data, and for Quick Graph. Then press for Actions, for Regression, and for Show Linear. To find the equation of the regression line on a full page, press for Calculator, for Statistics, for Stat Calculations, and for Linear Regression. Complete the dialog boxes.

To see the plot on a full page, press, and then press for Data & Statistics. Move the cursor to “Click to add variable,” and then press and select the list name. Repeat for the other axis. Then press for Actions, for Regression, and for Show Linear.

**COMMENTS**

Each time you execute a LinReg command, the calculator automatically computes the residuals and stores them in a data list named RESID. If you don’t want to see this (or any other calculator-generated list) anymore, press [F1] (Tools) and select choice 3: Setup Editor. Leaving the box for lists to display blank will reset the calculator to show only lists 1 through 6.

**TI-89**

To find the equation of the regression line (and add the line to a scatterplot), choose LinReg (a+bx) from the Calc Regressions menu and tell it the list names and a function to store the equation. To make a residuals plot, define a PLOT as a scatterplot. Specify your explanatory datalist as Xlist. For Ylist, find the list name resid from VAR-LINK by arrowing to the STATVARS portion. then press to locate the list. press ENTER to finish the plot definition and to display the plot.

Data Desk will offer to update the plots of residuals. SPSS standardizes the residuals by dividing by their standard deviation. (There’s no need to subtract their mean; it must be zero.) The standardization doesn’t affect the scatterplot.

**COMMENTS**

If you change any of the variables in the regression analysis, Data Desk will offer to update the plots of residuals.
To make a residuals plot, define a Plot as a scatterplot. Specify your explanatory datalist as Xlist. For Ylist, find the list name resid from VAR-LINK by arrowing to the STATVARS portion. Then press $r$ and locate the list. Press $\pi$ to finish the plot definition and $\alpha$ to display the plot.

### Chapter 10. Re-expressing Data: Get It Straight!

#### DATA DESK

To re-express a variable in Data Desk, select the variable and Choose the function to re-express it from the Manip > Transform menu. Square root, log, reciprocal, and reciprocal root are immediately available. For others, make a derived variable and type the function. Data Desk makes a new derived variable that holds the re-expressed values. Any value changed in the original variable will immediately be re-expressed in the derived variable.

#### EXCEL

To re-express a variable in Excel, use Excel’s built-in functions as you would for any calculation. Changing a value in the original column will change the re-expressed value.

#### JMP

To re-express a variable in JMP, double-click to the right of the last column of data to create a new column. Name the new column and select it. Choose Formula from the Cols menu. In the Formula dialog, choose the transformation and variable that you wish to assign to the new column. Click the OK button. JMP places the re-expressed data in the new column.

#### MINITAB

To re-express a variable in MINITAB, choose Calculator from the Calc menu. In the Calculator dialog, specify a name for the new re-expressed variable. Use the Functions List, the calculator buttons, and the Variables list box to build the expression. Click OK.

#### SPSS

To re-express a variable in SPSS, Choose Compute from the Transform menu. Enter a name in the Target Variable field. Use the calculator and Function List to build the expression. Move a variable to be re-expressed from the source list to the Numeric Expression field. Click the OK button.

#### TI-NSPIRE

To re-express data, create a new list and enter the formula in the cell in the second row. For example, if one column has a list named time, another list can be created using the formula log(time).

#### TI-89

To re-express data stored in a list, perform the re-expression on the whole list and store it in another list. For example, to use the common (base 10) logarithms of the data in list1, on the home screen, enter the command `log(list1)` `STO` list2.

### COMMENTS

Each time you execute a LinReg command, the calculator automatically computes the residuals and stores them in a data list named RESID. If you don’t want to see this (or any other calculator-generated list) anymore, press $\Omega$ (Tools) and select choice 3: Setup Editor. Leaving the box for lists to display blank will reset the calculator to show only lists 1 through 6.

Or choose Manip > Transform > Dynamic > Box-Cox to generate a continuously changeable variable and a slider that specifies the power. Set plots to Automatic Update in their HyperView menus and watch them change dynamically as you drag the slider.

The log and square root re-expressions are found in the Transcendental menu of functions in the formula dialog.

To find the log command, press $\Omega$ then $y$ (L) arrow to log, and press $\pi$.

Natural logs are LN (press $2\pi$).

For square roots, press $2p$.
## Chapter 11. Understanding Randomness

### DATA DESK
Generate random numbers in Data Desk with the **Generate Random Numbers** command in the **Manip** menu. A dialog guides you in specifying the number of variables to fill, the number of cases, and details about the values. For most simulations, generate random uniform values.

### EXCEL
The **RAND** function generates a random value between 0 and 1. You can multiply to scale it up to any range you like and use the **INT** function to turn the result into an integer.

### JMP
In a new column, in the **Cols** menu choose **Column Info**…
In the dialog, click the **New Property** button, and choose **Formula** from the drop-down menu.

### MINITAB
In the **Calc** menu, choose **Random Data**…
In the Random Data submenu, choose **Uniform**…

### SPSS
The **RV.UNIFORM(min, max)** function returns a random value that is equally likely between the min and max limits.

### TI-NSPIRE
To generate random numbers, move the cursor to highlight the name of a blank list. Use **5:RandInt** from the **(Calc) Probability** menu. This command will produce any number of random integers in the specified range.

### TI-89
To generate random integers, press **c**, **1** for Calculator, **b**, **5** for Probability, **4** for Random, and **2** for Integer. Then type the range for the random integers, such as `randInt(1,6)`.

### Comments
- **Bernoulli Trials** generate random values that are 0 or 1, with a specified chance of a 1.
- **Binomial Experiments** automatically generate a specified number of Bernoulli trials and count the number of 1’s.

## Chapter 16. Random Variables

### TI-NSPIRE
To compute the mean and standard deviation for a discrete random variable, enter the values in one named list and the probabilities in another. Then press **g**, **1** for Calculator, **e**, **1** for Stat Calculations, and **1** for One-Variable Statistics. Enter 2 for the prompt for the number of lists, **e** to OK, ·, and complete the dialog box.

### TI-89
To calculate the mean and standard deviation of a discrete random variable, enter the probability model in two lists:
- In one list (say, list1) enter the x-values of the variable.
- In a second list (say, list2) enter the associated probabilities $P(X = x)$.
- From the **STAT CALC** (**f**3) menu select 1-VarStats. Use **VAR-LINK** to enter the list name list1 in the List box and list2 in the Freq box.

### Comments
- Statistics, **1** for Stat Calculations, and **1** for One-Variable Statistics. Enter 2 for the prompt for the number of lists, **e** to OK, **2**, and complete the dialog box.
- You can enter the probabilities as fractions; the calculator will change them to decimals for you.
- Notice that the calculator knows enough to compute only the standard deviation $\sigma$, but mistakenly uses $\mu$ when it should say $\mu$. Make sure you don’t make that mistake!
Chapter 17. Probability Models

The only important differences among these functions are in what they are named and the order of their arguments. In these functions, pdf stands for “probability density function”—what we’ve been calling a probability model. The letters cdf stand for “cumulative distribution function,” the technical term when we want to accumulate probabilities over a range of values. These technical terms show up in many of the function names. The term “cumulative” in a function name says that it corresponds to a cdf.

Generically, the four functions are as follows:

- **Geometric pdf** \( (\text{prob}, x) \): Finds the individual geometric probability of getting the first success on trial \( x \) when the probability of success is \( \text{prob} \).
- **Geometric cdf** \( (\text{prob}, x) \): Finds the cumulative probability of getting the first success on or before trial \( x \), when the probability of success is \( \text{prob} \).
- **Binomial pdf** \( (n, \text{prob}, x) \): Finds the probability of getting \( x \) successes in \( n \) trials when the probability of success is \( \text{prob} \).
- **Binomial cdf** \( (n, \text{prob}, x) \): Finds the probability of getting \( x \) or fewer successes in \( n \) trials when the probability of success is \( \text{prob} \).

### DATA DESK

- **BinomDist** \( (x, n, \text{prob}) \) (pdf)
- **CumBinomDist** \( (x, n, \text{prob}) \) (cdf)

**COMMENTS**

Data Desk does not compute Geometric probabilities. These functions work in derived variables or in scratchpads.

### EXCEL

- **Binomdist** \( (x, n, \text{prob}, \text{cumulative}) \)

**COMMENTS**

Set cumulative = true or for cdf, false for pdf.
Excel’s function fails when \( x \) or \( n \) is large.
Possibly, it does not use the Normal approximation.
Excel does not compute Geometric probabilities.

### JMP

- **Binomial Probability** \( (\text{prob}, n, x) \) (pdf)
- **Binomial Distribution** \( (\text{prob}, n, x) \) (cdf)

**COMMENTS**

JMP does not compute Geometric probabilities.

### MINITAB

Choose **Probability Distributions** from the Calc menu.
Choose **Binomial** from the Probability Distributions submenu.
To calculate the probability of getting \( x \) successes in \( n \) trials, choose **Probability**.

### SPSS

- **PDF.GEOM** \( (x, \text{prob}) \)
- **CDF.GEOM** \( (x, \text{prob}) \)

### TI-NSPIRE

To compute geometric and binomial probabilities, press \( \text{\textasciitilde} \) for Probability, and \( \text{\textasciitilde} \) for Distributions. Select the menu item.

**COMMENTS**

Pdf is for the probability distribution function; Cdf will display cumulative probabilities. Complete the dialog box.

### TI-89

Find the commands under the **\textasciitilde** (Distributions) menu.
- **F**: **Geometric Pdf** will ask for \( p \) and \( x \). It returns the probability of the first success occurring on the \( x \)th trial.
- **G**: **Geometric Cdf** will ask for \( p \) and the upper and lower values of interest, say \( a \) and \( b \). It returns \( P(a \leq X \leq b) \), the probability the first success occurs between the \( a \)th and \( b \)th trials, inclusive.
- **A**: **Binomial Pdf** asks for \( n, p, \) and \( x \).
- **B**: **Binomial Cdf** asks for \( n, p, \) and the lower and upper values of interest.

**COMMENTS**

For Geometric variables, when finding \( P(X \geq a) \) specify an upper value of infinity, \( 10^99 \), or a very large number.
For Binomial variables, when finding \( P(X \geq a) \), the upper value is \( n \).
## Chapter 19. Confidence Intervals for Proportions

### DATA DESK

Data Desk does not offer built-in methods for inference with proportions.

**COMMENTS**
For summarized data, open a Scratchpad to compute the standard deviation and margin of error by typing the calculation. Then use $z$-interval for individual $\mu$s.

### EXCEL

Inference methods for proportions are not part of the standard Excel tool set.

**COMMENTS**
For summarized data, type the calculation into any cell and evaluate it.

### JMP

For a categorical variable that holds category labels, the Distribution platform includes tests and intervals for proportions. For summarized data, put the category names in one variable and the frequencies in an adjacent variable. Designate the frequency column to have the role of frequency. Then use the Distribution platform.

**COMMENTS**
JMP uses slightly different methods for proportion inferences than those discussed in this text. Your answers are likely to be slightly different, especially for small samples.

### MINITAB

Choose Basic Statistics from the Stat menu.

- Choose 1Proportion from the Basic Statistics submenu.
- If the data are category names in a variable, assign the variable from the variable list box to the Samples in columns box. If you have summarized data, click the Summarized Data button and fill in the number of trials and the number of successes.
- Click the Options button and specify the remaining details.

**COMMENTS**
If you have a large sample, check Use test and interval based on normal distribution. Click the OK button.

### SPSS

SPSS does not find confidence intervals for proportions.

### TI-Nspire

To compute a confidence interval for a population proportion, press $\text{2nd}$, $\text{0}$ for Calculator, $\text{5}$ for Statistics, $\text{4}$ for Confidence Intervals, and $\text{5}$ for 1-Prop $z$-interval. Complete the dialog box. Be sure to enter the number of successes, $x$, as a whole number, and the C level as a decimal, such as .99.

**COMMENTS**
Beware: When you enter the value of $x$, you need the count, not the percentage. The count must be a whole number. If the number of successes are given as a percentage, you must first multiply $np$ and round the result.

### TI-89

To calculate a confidence interval for a population proportion:

- Go to the Ints menu (2nd) and select 5:1-PropZInt.
- Enter the number of successes observed and the sample size.
- Specify a confidence level.
- Calculate the interval.

**COMMENTS**
Beware: When you enter the value of $x$, you need the count, not the percentage. The count must be a whole number. If the number of successes are given as a percentage, you must first multiply $np$ and round the result.

## Chapter 20. Testing Hypotheses About Proportions

### DATA DESK

Data Desk does not offer built-in methods for inference with proportions. The Replicate Y by X command in the Manip menu will "reconstruct" summarized count data so that you can display it.

**COMMENTS**
For summarized data, open a Scratchpad to compute the standard deviation and margin of error by typing the calculation. Then perform the test with the $z$-test for individual $\mu$s found in the Test command.

### EXCEL

Inference methods for proportions are not part of the standard Excel tool set.

**COMMENTS**
For summarized data, type the calculation into any cell and evaluate it.
### Appendix B: Guide to Statistical Software

#### JMP

For a categorical variable that holds category labels, the Distribution platform includes tests and intervals of proportions. For summarized data, put the category names in one variable and the frequencies in an adjacent variable. Designate the frequency column to have the role of frequency. Then use the Distribution platform.

**COMMENTS**

JMP uses slightly different methods for proportion inferences than those discussed in this text. Your answers are likely to be slightly different.

---

#### MINITAB

Choose Basic Statistics from the Stat menu.
- Choose 1Proportion from the Basic Statistics submenu.
- If the data are category names in a variable, assign the variable from the variable list box to the Samples in columns box.
- If you have summarized data, click the Summarized Data button and fill in the number of trials and the number of successes.
- Click the Options button and specify the remaining details.

**COMMENTS**

If you have a large sample, check Use test and interval based on Normal distribution.

Click the OK button.

When working from a variable that names categories, MINITAB treats the last category as the “success” category. You can specify how the categories should be ordered.

#### SPSS

SPSS does not find hypothesis tests for proportions.

#### TI-Nspire

To compute a hypothesis test for a population proportion, press 2, 1 for Calculator, 6, 7 for Statistics, and 5 for 1-Prop z-test. Complete the dialog box. Be sure to enter the number of successes, \( x \), as a whole number.

**COMMENTS**

Specify whether to calculate the result or draw the result (a normal curve with \( p \)-value area shaded.)

#### TI-89

To do the mechanics of a hypothesis test for a proportion,
- Select 5:1-PropZTest from the STAT TESTS menu.
- Specify the hypothesized proportion.
- Enter the observed value of \( x \).
- Specify the sample size.
- Indicate what kind of test you want: one-tail lower tail, two-tail, or one-tail upper tail.

**COMMENTS**

Beware: When you enter the value of \( x \), you need the count, not the percentage. The count must be a whole number. If the number of successes is given as a percent, you must first multiply \( np \) and round the result to obtain \( x \).

---

**Chapter 22. Comparing Two Proportions**

#### DATA DESK

Data Desk does not offer built-in methods for inference with proportions. Use Replicate Y by X to construct data corresponding to given proportions and totals.

**COMMENTS**

For summarized data, open a Scratchpad to compute the standard deviations and margin of error by typing the calculation.

#### EXCEL

Inference methods for proportions are not part of the standard Excel tool set.

**COMMENTS**

For summarized data, type the calculation into any cell and evaluate it.

#### JMP

For a categorical variable that holds category labels, the Distribution platform includes tests and intervals of proportions. For summarized data, put the category names in one variable and the frequencies in an adjacent variable. Designate the frequency column to have the role of frequency. Then use the Distribution platform.

**COMMENTS**

JMP uses slightly different methods for proportion inferences than those discussed in this text. Your answers are likely to be slightly different.
## MINITAB

To find a hypothesis test for a proportion, Choose Basic Statistics from the Stat menu.
Choose 2Proportions ... from the Basic Statistics submenu. If the data are organized as category names in one column and case IDs in another, assign the variables from the variable list box to the Samples in one column box. If the data are organized as two separate columns of responses, click on Samples in different columns: and assign the variables from the variable list box. If you have summarized data, click the Summarized Data button and fill in the number of trials and the number of successes for each group.

Click the Options button and specify the remaining details. Remember to click the Use pooled estimate of p for test box when testing the null hypothesis of no difference between proportions. Click the OK button.

**COMMENTS**

When working from a variable that names categories, MINITAB treats the last category as the “success” category. You can specify how the categories should be ordered.

---

## SPSS

SPSS does not find hypothesis tests for proportions.

---

## TI-Nspire

To compute a confidence interval for the difference between two population proportions, press $c$, for Calculator, $b$, for Statistics, $6$ for Confidence Intervals, and $6$ for 2-Prop Z-interval. Complete the dialog box. Be sure to enter each number of successes as a whole number, and the C level as a decimal, such as .99.

To compute a hypothesis test for the difference between two population proportions, press $c$, for Calculator, $b$, for Statistics, $7$ for Stat Tests, and $6$ for 2-Prop Z-test. Complete the dialog box. Be sure to enter each number of successes as a whole number.

• Indicate what kind of test you want: one-tail upper tail, lower tail, or two-tail.
• Specify whether results should simply be calculated or displayed with the area corresponding to the P-value of the test shaded.

**COMMENTS**

_Beware:_ When you enter the value of $x$, you need the count, not the percentage. The count must be a whole number. If the number of successes is given as a percent, you must first multiply $np$ and round the result to obtain $x$.

---

## TI-89

To calculate a confidence interval for the difference between two population proportions,
• Select 6:2-PropZInt from the STAT Ints menu.
• Enter the observed counts and the sample sizes for both samples.
• Specify a confidence level.
• Calculate the interval.

To do the mechanics of a hypothesis test for equality of population proportions,
• Select 6:2-PropZTest from the STAT Tests menu.
• Enter the observed counts and sample sizes.
• Indicate what kind of test you want: one-tail upper tail, lower tail, or two-tail.
• Specify whether results should simply be calculated or displayed with the area corresponding to the P-value of the test shaded.

**COMMENTS**

_Beware:_ When you enter the value of $x$, you need the count, not the percentage. The count must be a whole number. If the number of successes is given as a percent, you must first multiply $np$ and round the result to obtain $x$.

---

## Chapter 23. Inferences About Means

### DATA DESK

Select variables.
From the Calc menu, choose Estimate for confidence intervals or Test for hypothesis tests. Select the interval or test from the drop-down menu and make other choices in the dialog.

### EXCEL

Specify formulas. Find $t^*$ with the TINV(alpha, df) function.

**COMMENTS**

Not really automatic. There’s no easy way to find P-values in Excel.

### JMP

From the Analyze menu, select Distribution. For a confidence interval, scroll down to the “Moments” section to find the interval limits. For a hypothesis test, click the red triangle next to the variable’s name and choose Test Mean from the menu. Then fill in the resulting dialog.

**COMMENTS**

“Moment” is a fancy statistical term for means, standard deviations, and other related statistics.

### MINITAB

From the Stat menu, choose the Basic Statistics submenu. From that menu, choose 1-sample t . . . Then fill in the dialog.

**COMMENTS**

The dialog offers a clear choice between confidence interval and test.
Chapter 24. Comparing Means

There are two ways to organize data when we want to compare two independent groups. The data can be in two lists, as in the table at the start of this chapter. Each list can be thought of as a variable. In this method, the variables in the batteries example would be Brand Name and Generic. Graphing calculators usually prefer this form, and some computer programs can use it as well.

There’s another way to think about the data. What is the response variable for the battery life experiment? It’s the Time until the music stopped. But the values of this variable are in both columns, and actually there’s an experiment factor here, too—namely, the Brand of the battery. So, we could put the data into two different columns, one with the Times in it and one with the Brand: Then the data would look as shown in the table to the right.

This way of organizing the data makes sense as well. Now the factor and the response variables are clearly visible. You’ll have to see which method your program requires. Some packages even allow you to structure the data either way.

The commands to do inference for two independent groups on common statistics technology are not always found in obvious places. Here are some starting guidelines.

<table>
<thead>
<tr>
<th>Brand</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand name</td>
<td>194.0</td>
</tr>
<tr>
<td>Brand name</td>
<td>205.5</td>
</tr>
<tr>
<td>Brand name</td>
<td>199.2</td>
</tr>
<tr>
<td>Brand name</td>
<td>172.4</td>
</tr>
<tr>
<td>Brand name</td>
<td>184.0</td>
</tr>
<tr>
<td>Brand name</td>
<td>169.5</td>
</tr>
<tr>
<td>Generic</td>
<td>190.7</td>
</tr>
<tr>
<td>Generic</td>
<td>203.5</td>
</tr>
<tr>
<td>Generic</td>
<td>203.5</td>
</tr>
<tr>
<td>Generic</td>
<td>206.5</td>
</tr>
<tr>
<td>Generic</td>
<td>222.5</td>
</tr>
<tr>
<td>Generic</td>
<td>209.4</td>
</tr>
</tbody>
</table>

**DATA DESK**

Select variables.
From the Calc menu, choose Estimate for confidence intervals or Test for hypothesis tests. Select the interval or test from the drop-down menu and make other choices in the dialog.

**COMMENTS**

Data Desk expects the two groups to be in separate variables.

**EXCEL**

From the Data Tab, Analysis Group, choose Data Analysis. Alternatively (if the Data Analysis Tool Pack is not installed), in the Formulas Tab, choose More functions > Statistical > TTEST, and specify Type=3 in the resulting dialog.
Fill in the cell ranges for the two groups, the hypothesized difference, and the alpha level.

**COMMENTS**

Excel expects the two groups to be in separate cell ranges. Notice that, contrary to Excel’s wording, we do not need to assume that the variances are not equal; we simply choose not to assume that they are equal.
### JMP

From the **Analyze** menu, select **Fit y by x**. Select variables: a **Y**, **Response** variable that holds the data and an **X**, **Factor** variable that holds the group names. JMP will make a dotplot. Click the **red triangle** in the dotplot title, and choose **Unequal variances**. The **t**-test is at the bottom of the resulting table. Find the P-value from the **Prob>F** section of the table (they are the same).

**COMMENTS**

JMP expects data in one variable and category names in the other. Don’t be misled: There is no need for the variances to be unequal to use two-sample **t** methods.

### MINITAB

From the **Stat** menu, choose the **Basic Statistics** submenu. From that menu, choose **2-sample t**... Then fill in the dialog.

**COMMENTS**

The dialog offers a choice of data in two variables, or data in one variable and category names in the other.

### SPSS

From the **Analyze** menu, choose the **Compare Means** submenu. From that, choose the **Independent-Samples t-test** command. Specify the data variable and “group variable.” Then type in the labels used in the group variable. SPSS offers both the two-sample and pooled-**t** results in the same table.

**COMMENTS**

SPSS expects the data in one variable and group names in the other. If there are more than two group names in the group variable, only the two that are named in the dialog box will be compared.

### TI-Nspire

To compute a confidence interval for the difference between two population means, press **4**, **2** for Calculator, **4** for Statistics, **4** for Confidence Intervals, and **4** for 2-Sample **t**-interval. Select between Data and Stats, **2** to OK, and **2**. Complete the dialog box. Be sure to enter the C level as a decimal, such as .99.

To compute a hypothesis test for the difference between two population means, press **4**, **2** for Calculator, **4** for Statistics, **7** for Stat Tests, and **4** for 2-Sample **t**-test. Select between Data and Stats, **2** to OK, and **2**. Complete the dialog box.

### TI-89

For a confidence interval:

In the **STAT Ints** menu, choose **4:2-SampTInt**. You must specify if you are using data stored in two lists or if you will enter the means, standard deviations, and sizes of both samples. You must also indicate whether to pool the variances (when in doubt, say no) and specify the desired level of confidence.

To test a hypothesis:

In the **STAT TESTS** menu, choose **4:2-SampTTest**. You must specify if you are using data stored in two lists or if you will enter the means, standard deviations, and sizes of both samples. You must also indicate whether to pool the variances (when in doubt, say no) and specify whether the test is to be two-tail, lower-tail, or upper-tail.

### Chapter 25. Paired Samples and Blocks

#### DATA DESK

Select variables.

From the **Calc** menu, choose **Estimate** for confidence intervals or **Test** for hypothesis tests. Select the interval or test from the drop-down menu, and make other choices in the dialog.

**COMMENTS**

Data Desk expects the two groups to be in separate variables and in the same “Relation”—that is, about the same cases.

#### EXCEL

In Excel 2003 and earlier, select **Data Analysis** from the **Tools** menu.

In Excel 2007, select **Data Analysis** from the **Analysis** Group on the **Data** Tab.

From the **Data Analysis** menu, choose **t-test: paired two-sample for Means**. Fill in the cell ranges for the two groups, the hypothesized difference, and the alpha level.

**COMMENTS**

Excel expects the two groups to be in separate cell ranges.

**Warning:** Do not compute this test in Excel without checking for missing values. If there are any missing values (empty cells), Excel will usually give a wrong answer. Excel compacts each list, pushing values up to cover the missing cells, and then checks only that it has the same number of values in each list. The result is mismatched pairs and an entirely wrong analysis.

#### JMP

From the **Analyze** menu, select **Matched Pairs**. Specify the columns holding the two groups in the **Y Paired Response** Dialog. Click **OK**.
<table>
<thead>
<tr>
<th>MINITAB</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>From the Stat menu, choose the Basic Statistics submenu. From that menu, choose Paired t... Then fill in the dialog.</td>
<td>Minitab takes “First sample” minus “Second sample.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPSS</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>From the Analyze menu, choose the Compare Means submenu. From that, choose the Paired-Samples t-test command. Select pairs of variables to compare, and click the arrow to add them to the selection box.</td>
<td>You can compare several pairs of variables at once. Options include the choice to exclude cases missing in any pair from all tests.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TI-NSPRIE</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>For inference on a matched pair design, compute a third list of differences such as diff = time2-time1. Then construct the confidence interval or conduct the hypothesis test in the same way as 1-sample procedures, using the list of differences.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TI-89</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>If the data are stored in two lists, say, list1 and list2, create a list of the differences. Move the cursor to the name of an empty list, and then use VAR-LINK to enter the command list1-list2. Press ENTER to perform the subtraction.</td>
<td>Since inference for paired differences uses one-sample t-procedures, select 2:T-Test or 2:TInterval from the STAT Tests or Ints menu. Specify as your data the list of differences you just created, and apply the procedure.</td>
</tr>
</tbody>
</table>

Chapter 26. Comparing Counts

<table>
<thead>
<tr>
<th>DATA DESK</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select variables. From the Calc menu, choose Contingency Table. From the table’s HyperView menu, choose Table Options. (Or Choose Calc &gt; Calculation Options &gt; Table Options.) In the dialog, check the boxes for Chi Square and for Standardized Residuals. Data Desk will display the chi-square and its P-value below the table, and the standardized residuals within the table.</td>
<td>Data Desk automatically treats variables selected for this command as categorical variables even if their elements are numerals. The Compute Counts command in the table’s HyperView menu will make variables that hold the table contents (as selected in the Table Options dialog), including the standardized residuals.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXCEL</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excel offers the function CHITEST(actual_range, expected_range), which computes a chi-square value for homogeneity. Both ranges are of the form UpperLeftCell:LowerRightCell, specifying two rectangular tables that must hold counts (although Excel will not check for integer values). The two tables must be of the same size and shape.</td>
<td>Excel’s documentation claims this is a test for independence and labels the input ranges accordingly, but Excel offers no way to find expected counts, so the function is not particularly useful for testing independence. You can use this function only if you already know both tables of counts or are willing to program additional calculations.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>JMP</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>From the Analyze menu, select Fit Y by X. Select variables: a Y, Response variable that holds responses for one variable, and an X, Factor variable that holds responses for the other. Both selected variables must be Nominal or Ordinal. JMP will make a plot and a contingency table. Below the contingency table, JMP offers a Tests panel. In that panel, the Chi Square for independence is called a Pearson ChiSquare. The table also offers the P-value. Click on the Contingency Table title bar to drop down a menu that offers to include a Deviation and Cell Chi square in each cell of the table.</td>
<td>JMP will choose a chi-square analysis for a Fit Y by X if both variables are nominal or ordinal (marked with an N or O), but not otherwise. Be sure the variables have the right type. Deviations are the observed—expected differences in counts. Cell chi-squares are the squares of the standardized residuals. Refer to the deviations for the sign of the difference. Look under Distributions in the Analyze menu to find a chi-square test for goodness-of-fit.</td>
</tr>
</tbody>
</table>
**Minitab**

From the Stat menu, choose the Tables submenu. From that menu, choose Chi Square Test . . . . In the dialog, identify the columns that make up the table. Minitab will display the table and print the chi-square value and its P-value.

**Comments**

Alternatively, select the Cross Tabulation . . . command to see more options for the table, including expected counts and standardized residuals.

**SPSS**

From the Analyze menu, choose the Descriptive Statistics submenu. From that submenu, choose Crosstabs . . . . In the Crosstabs dialog, assign the row and column variables from the variable list. Both variables must be categorical. Click the Cells button to specify that standardized residuals should be displayed. Click the Statistics button to specify a chi-square test.

**Comments**

SPSS offers only variables that it knows to be categorical in the variable list for the Crosstabs dialog. If the variables you want are missing, check that they have the right type.

**TI-Nspire**

To conduct a $\chi^2$ goodness of fit test, enter the observed and the expected values into two named lists. Then press $\mathrm{C}$, $\mathrm{1}$ for Calculator, $\mathrm{C}$, $\mathrm{1}$ for Statistics, $\mathrm{C}$ for Stat Tests, and $\mathrm{C}$ for $\chi^2$ GOF. Complete the dialog box.

To conduct a $\chi^2$ test of independence or homogeneity, first enter the data into a matrix. Press $\mathrm{C}$, $\mathrm{1}$ for Calculator, $\mathrm{C}$, $\mathrm{6}$ for Statistics, $\mathrm{C}$ for Stat Tests, and $\mathrm{C}$ for $\chi^2$ 2-way Test. Complete the dialog box.

**TI-89**

To test a hypothesis of homogeneity or independence, you need to enter the data as a matrix. From the home screen, press $\mathrm{O}$ and select $\mathrm{6}$:Data/Matrix Editor, then select $\mathrm{3}$:New. Specify type as Matrix and name the matrix in the Variable box. Specify the number of rows and columns. Type the entries, pressing $\mathrm{P}$ after each. Press $\mathrm{2}$N to leave the editor.

To do the test, choose $\mathrm{8}$:Chi2 2-way from the STAT Tests menu.

**Comments**

You can change the regression by dragging the icon of another variable over either the Y- or X-variable name in the table and dropping it there. The regression will recompute automatically.

**Excel**

In Excel 2003 and earlier, select Data Analysis from the Tools menu. In Excel 2007, select Data Analysis from the Analysis Group on the Data Tab.

Select Regression from the Analysis Tools list.

Click the OK button.

Enter the data range holding the Y-variable in the box labeled "Y-range".

Enter the range of cells holding the X-variable in the box labeled "X-range".

Select the New Worksheet Ply option.

Select Residuals options. Click the OK button.

**Comments**

The Y and X ranges do not need to be in the same rows of the spreadsheet, although they must cover the same number of cells. But it is a good idea to arrange your data in parallel columns as in a data table.

Although the dialog offers a Normal probability plot of the residuals, the data analysis add-in does not make a correct probability plot, so don’t use this option.
### JMP

- From the **Analyze** menu, select **Fit Y by X**.
- Select variables: a **Y**, Response variable, and an **X**, Factor variable. Both must be continuous (quantitative).
- JMP makes a scatterplot.
- Click on the red triangle beside the heading labeled **Bivariate Fit**, and choose **Fit Line**. JMP draws the least squares regression line on the scatterplot and displays the results of the regression in tables below the plot.
- The portion of the table labeled **"Parameter Estimates"** gives the coefficients and their standard errors, \( t \)-ratios, and P-values.

**COMMENTS**

JMP chooses a regression analysis when both variables are “Continuous.” If you get a different analysis, check the variable types. The Parameter table does not include the residual standard deviation \( s_e \). You can find that as Root Mean Square Error in the Summary of Fit panel of the output.

### MINITAB

- Choose **Regression** from the **Stat** menu.
- Choose **Regression...** from the **Regression** submenu.
- In the Regression dialog, assign the Y-variable to the **Response** box and assign the X-variable to the **Predictors** box.
- Click the **Graphs** button.
- In the Regression-Graphs dialog, select **Standardized residuals**, and check **Normal plot of residuals** and **Residuals versus fits**.

**COMMENTS**

You can also start by choosing a **Fitted Line plot** from the **Regression** submenu to see the scatterplot first—usually good practice.

### SPSS

- Choose **Regression** from the **Analyze** menu.
- Choose **Linear** from the **Regression** submenu.
- In the Linear Regression dialog that appears, select the Y-variable and move it to the dependent target. Then move the X-variable to the independent target.
- Click the **Plots** button.

**COMMENTS**

To compute a hypothesis test for a population slope, first enter the data into two named lists. Then press **c**, **6** for Calculations, **6** for Statistics, and **7** for Linear Reg t-tests. Complete the dialog box.

### TI-Nspire

To compute a confidence interval for a population slope, first enter the data into two named lists. Then press **c**, **1** for Calculator, **c**, **6** for Statistics, **c**, **6** for Confidence Intervals, and **c**, **7** for Linear Reg t-intervals. Select slope, **c**, **2** to OK, and **c**, **3**. Complete the dialog box. Be sure to enter the C level as a decimal, such as .99.

To compute a hypothesis test for a population slope, first enter the data into two named lists. Then press **c**, **6** for Calculations, **c**, **7** for Statistics, and **c**, **7** for Linear Reg t-test. Complete the dialog box.

**COMMENTS**

Under **STAT Tests** choose **A:LinRegTTest**. Specify the two lists where the data are stored and (usually) choose the two-tail option. Select an equation name to store the resulting line. In addition to reporting the calculated value of \( t \) and the P-value, the calculator will tell you the coefficients of the regression equation \( (a \text{ and } b) \), the values of \( r^2 \) and \( r \), the value of \( s \) used in prediction and confidence intervals, and the standard error of the slope. For 95% prediction and confidence intervals, choose **7:LinRegTint** from the **STAT Ints** menu. Specify the two lists where the data are stored, and select an equation name to store the resulting line. Select for an interval for the slope or for a response. If for a response, enter the \( x \)-value.

### TI-89

- Select the response variable as **Y** and the factor variable as **X**.
- From the **Calc** menu, choose **ANOVA**.
- Data Desk displays the ANOVA table.
- Select plots of residuals from the ANOVA table’s HyperView menu.

**COMMENTS**

Data Desk expects data in “stacked” format. You can change the ANOVA by dragging the icon of another variable over either the **Y** or **X** variable name in the table and dropping it there. The analysis will recompute automatically.

### Chapter 28. Analysis of Variance

- **DATA DESK**
  - Select the response variable as **Y** and the factor variable as **X**.
  - From the **Calc** menu, choose **ANOVA**.
  - Data Desk displays the ANOVA table.
  - Select plots of residuals from the ANOVA table’s HyperView menu.

**COMMENTS**

Data Desk expects data in “stacked” format. You can change the ANOVA by dragging the icon of another variable over either the **Y** or **X** variable name in the table and dropping it there. The analysis will recompute automatically.
**EXCEL**

- In Excel 2003 and earlier, select **Data Analysis** from the Tools menu.
- In Excel 2007, select **Data Analysis** from the Analysis Group on the Data Tab.
- Select **Anova Single Factor** from the list of analysis tools.
- Click the **OK** button.
- Enter the data range in the box provided.
- Check the **Labels in First Row** box, if applicable.
- Enter an alpha level for the F-test in the box provided.
- Click the **OK** button.

**COMMENTS**
The data range should include two or more columns of data to compare. Unlike all other statistics packages, Excel expects each column of the data to represent a different level of the factor. However, it offers no way to label these levels. The columns need not have the same number of data values, but the selected cells must make up a rectangle large enough to hold the column with the most data values.

**JMP**

- From the **Analyze** menu select **Fit Y by X**.
- Select variables: a quantitative **Y**, Response variable, and a categorical **X**, Factor variable.
- JMP opens the **One way** window.
- Click on the red triangle beside the heading, select **Display Options**, and choose **Boxplots**.
- From the same menu choose the **Means/ANOVA.t-test** command.
- JMP opens the one way ANOVA output.

**COMMENTS**
JMP expects data in “stacked” format with one response and one factor variable.

**MINITAB**

- Choose **ANOVA** from the Stat menu.
- Choose **One-way...** from the **ANOVA** submenu.
- In the One-way **Anova** dialog, assign a quantitative **Y** variable to the Response box and assign a categorical **X** variable to the Factor box.
- Check the **Store Residuals** check box.
- In the ANOVA-Graphs dialog, select **Standardized residuals**, and check **Normal plot of residuals** and **Residuals versus fits**.
- Click the **OK** button to return to the Regression dialog.
- Click the **OK** button to compute the regression.

**COMMENTS**
If your data are in unstacked format, with separate columns for each treatment level, choose **One-way (unstacked)** from the **ANOVA** submenu.

**SPSS**

- Choose **Compare Means** from the **Analyze** menu.
- Choose **One-way ANOVA** from the **Compare Means** submenu.
- In the One-Way **ANOVA** dialog, select the **Y**-variable and move it to the dependent target. Then move the **X**-variable to the independent target.
- Click the **OK** button.

**COMMENTS**
SPSS expects data in stacked format. The **Contrasts** and **Post Hoc** buttons offer ways to test contrasts and perform multiple comparisons. See your SPSS manual for details.

**TI-89**

Under **STAT Tests**, choose **C:ANOVA**

- Specify the input method (Data or Stats) according to whether you have data entered as one list for each group or summary statistics for each group, and specify the number of groups. Press ÷.
- If Data, you will then be asked to supply the name of each list.
- If Stats, you will be asked for the stats for each group. Enter n, \( \bar{x} \), and s for each group separated by commas and within curly braces (\{and\}).
- Press \( \div \) to perform the calculations.

**COMMENTS**
In addition to the ANOVA table output, the calculator creates three new lists—the means for each group (in the order specified) and individual 95% confidence interval upper and lower bounds.
## Chapter 29. Multiple Regression

### DATA DESK
- Select Y- and X-variable icons.
- From the Calc menu, choose Regression.
- Data Desk displays the regression table.
- Select plots of residuals from the Regression table’s HyperView menu.

**COMMENTS**
You can change the regression by dragging the icon of another variable over either the Y- or an X-variable name in the table and dropping it there. You can add a predictor by dragging its icon into that part of the table. The regression will recompute automatically.

### EXCEL
- In Excel 2003 and earlier, select Data Analysis from the Tools menu.
- In Excel 2007, select Data Analysis from the Analysis Group on the Data Tab.
- Select Regression from the Analysis Tools list.
- Click the OK button.
- Enter the data range holding the Y-variable in the box labeled “Y-range.”
- Enter the range of cells holding the X-variables in the box labeled “X-range.”
- Select the New Worksheet Ply option.
- Select Residuals options. Click the OK button.

**COMMENTS**
The Y and X ranges do not need to be in the same rows of the spreadsheet, although they must cover the same number of cells. But it is a good idea to arrange your data in parallel columns as in a data table. The X-variables must be in adjacent columns. No cells in the data range may hold non-numeric values. Although the dialog offers a Normal probability plot of the residuals, the data analysis add-in does not make a correct probability plot, so don’t use this option.

### JMP
- From the Analyze menu select Fit Model.
- Specify the response, Y. Assign the predictors, X, in the Construct Model Effects dialog box.
- Click on Run Model.

**COMMENTS**
JMP chooses a regression analysis when the response variable is “Continuous.” The predictors can be any combination of quantitative or categorical. If you get a different analysis, check the variable types.

### MINITAB
- Choose Regression from the Stat menu.
- Choose Regression . . . from the Regression submenu.
- In the Regression dialog, assign the Y-variable to the Response box and assign the X-variables to the Predictors box.
- Click the Graphs button.

**COMMENTS**
In the Regression-Graphs dialog, select Standardized residuals, and check Normal plot of residuals and Residuals versus fits.
- Click the OK button to return to the Regression dialog.
- Click the OK button to compute the regression.

### SPSS
- Choose Regression from the Analyze menu.
- Choose Linear from the Regression submenu.
- When the Linear Regression dialog appears, select the Y-variable and move it to the dependent target. Then move the X-variables to the independent target.
- Click the Plots button.

**COMMENTS**
In the Linear Regression Plots dialog, choose to plot the *SRESID against the *ZPRED values.
- Click the Continue button to return to the Linear Regression dialog.
- Click the OK button to compute the regression.

### TI-89
Under STAT Tests choose B:MultReg Tests
- Specify the number of predictor variables, and which lists contain the response variable and predictor variables.
- Press □ to perform the calculations.

**COMMENTS**
The first portion of the output gives the F-statistic and its P-value as well as the values of $R^2$, $\text{Adj}^2R^2$, the standard deviation of the residuals (s), and the Durbin-Watson statistic, which measures correlation among the residuals.
- The rest of the main output gives the components of the $F$-test, as well as values of the coefficients, their standard errors, and associated $t$-statistics along with P-values. You can use the right arrow to scroll through these lists (if desired).
- The calculator creates several new lists that can be used for assessing the model and its conditions: Yhatlist, resid, sresid (standardized residuals), leverage, and cookd, as well as lists of the coefficients, standard errors, $t$’s, and P-values.