Introduction

These easy step-by-step tutorials introduce you to the Minitab environment and provide a quick overview of some of Minitab's most important features. Each of the tutorials are designed to explain the fundamentals of using Minitab—how to use the menus and dialog boxes, how to manage and manipulate data and files, how to produce graphs, and more. There are five successive tutorials:

Session One: Minitab Basics

Session Two: Doing a Simple Analysis

Session Three: Advanced Minitab

Session Four: Quality Control and Improvement

Session Five: Designing an Experiment

Session One: Minitab Basics

Overview of Session One

The story

Clones are genetically identical cells descended from the same individual. Researchers have identified a single poplar clone that yields fast-growing, hardy trees. These trees may one day be an alternative energy resource to conventional fuel.
Researchers at The Pennsylvania State University planted Poplar Clone 252 on two different sites: one site was by a creek with rich, well-drained soil, and the other site was on a ridge with dry, sandy soil. They measured the diameter in centimeters, height in meters, and dry weight of the wood in kilograms of a sample of three-year-old trees. These researchers want to see if they can predict how much a tree weighs from its diameter and height measurements.

Congratulations! You have been hired as data analyst for the project, and you will be performing the statistical analysis.

What you will learn

In Session One you will learn how to:

· open a worksheet· enter and edit data· save data· compute some basic statistics· do arithmetic · plot the data· compute a correlation coefficient· edit and add comments to the output· print and save your results

Time required

About 30 minutes.

Proceed to Step 1: Start Minitab.

Step 1: Start Minitab

· To start Minitab:
From the Taskbar, choose Start > Programs > Minitab 13 for Windows > Minitab.

Step 2: Open a Worksheet

When you start Minitab, you begin with a new, empty project. You can add data to your project in many ways, but the most common way is to open a worksheet. Note that you are only copying the data from the worksheet to the project; any changes that you make to the data added to your project will not affect the original file.
In this session, you will use the file POPLAR1.MTW. This file is one of the dozens of worksheets that are shipped with Minitab. Most of these worksheets are in the Data subdirectory or folder.

1 Activate the Project Manager by typing [Ctrl]+[I] or by clicking the Project Manager button on the toolbar.
2 Right-click on the Worksheets folder in the Project Manager and choose Open Worksheet.
3 Make sure the file type is Minitab (*.mtw) and the current subdirectory is Data.
4 Click on POPLAR1.MTW and click Open. Verify dialog box.
5 If the Data window is not already visible, open it to view the columns in your worksheet: choose Window > POPLAR1.MTW or press [Ctrl]+[D].

This worksheet contains three variables, labeled Diameter, Height, and Weight. Each variable contains 15 observations
—all the data collected so far. Verify Data window.

Step 3: Enter Data from the Keyboard

The worksheet POPLAR1 contained the data collected so far, but you just received new observations from the field, and there are five new rows to enter.

1 Press until you reach the first blank cell in row 16 or, with your mouse, click on the first blank cell in row 16.
2 Make sure the data entry arrow points to the right. If it does not, click on it to change its direction. Verify data window.
3 Type the following from left to right across each row:

1.52	[Enter]	2.9	[Enter]	.07	[Ctrl]+[Enter]
4.51	[Enter]	5.27	[Enter]	.79	[Ctrl]+[Enter]
1.18	[Enter]	2.2	[Enter]	.03	[Ctrl]+[Enter]
3.17	[Enter]	4.93	[Enter]	.44	[Ctrl]+[Enter]
3.33	[Enter]	4.89	[Enter]	.52	[Ctrl]+[Enter]

Tip If you make a mistake: click on or move to a cell (the contents will be automatically selected, type the correct value, and press [Enter].

Step 4: Enter Patterned Data

You now want to create a new variable that will indicate whether an observation was taken from the site with rich, well-drained soil (1), or from the site with dry, sandy soil (2). This new variable, called Site, will contain ten 1's followed by ten 2's.

You can always type data in the Data window, but if your data follow a pattern, you may find it easier to use the Autofill feature to create patterned data.

1 Go to column C4 and select the Label cell, directly under the column number.
2 Type Site and press [Enter].
3 In row 1, type 1.
4 Place your mouse over the square handle in the lower right corner of the selected cell. The crosshair cursor should change from white to solid black, which indicates that you are in Autofill mode.
5 Right-click and drag down to row 10. Autofill will automatically fill the selected cells with 1's.
6 In row 11, type 2

7 Click with the Autofill cursor and drag down to row 20.

The new Site column will appear in the Columns folder in the Project Manager.

Note To view the column info, double-click on the folder POPLAR1.MTW in the Project Manager and click on the Columns folder. The column info will appear in the right pane of the Project Manager.

Step 5: Save Your Project

It is a good idea to save your work frequently. Now is probably a good time to save, since you have just entered new data.

1 In the Project Manager, right-click on the top-most folder (labeled Untitled) and choose Save Project As….
2 In File name, enter POPLAR1 for the name of your project. If you omit the extension .MPJ, Minitab will automatically add it once you save the document.
3 Click Save. Verify dialog box.
4 If you see a message box asking if you want to replace an existing file, click Yes.
Once you have saved your project, the project folder in the Project Manager will be labeled Poplar1.mpj.

Step 6: Compute Descriptive Statistics

Minitab offers a wide array of basic statistics to help you analyze your data, such as descriptive statistics, t-tests, z-tests, and correlations. You decide to produce summary tables and boxplots describing the variables Diameter, Height, and Weight for the trees at each site.

1 Choose Stat > Basic Statistics > Display Descriptive Statistics.
2 In the variable list box, click Diameter and drag the mouse so that you highlight Diameter, Height, and Weight. Then click Select.
3 Check By variable, and enter Site. Verify dialog box.

Checking By variable tells Minitab to generate separate statistics for Diameter, Height, and Weight for each level of the variable Site.

Note When you select a series of columns, Minitab uses a dash to abbreviate the series. In this example, Diameter
—Weight means the variables Diameter, Height, and Weight.

4 Click Graphs.
5 Check Boxplot of data and click OK in each dialog box. Verify dialog box.

Minitab displays text output in the Session window and each graph (three, in this case) in its own Graph window.

Now you can tile the graphs to view all of them at one time on your screen.

1 Activate the Project Manager by typing [Ctrl]+[I] or by clicking the Project Manager button on the toolbar.
2 Click on the Graphs folder.
3 In the right pane of the Project Manager, select the graphs you want. You can select multiple graphs by clicking on the graph titles while holding down the [Ctrl] key,

or by clicking below or to the side of the graph titles and dragging across the desired selections.
4 Right-click on the selected graphs and choose Tile.

Judging from the boxplots , poplars grown at Site 2 are larger than those grown at Site 1. The Session window data confirm that the median values for Diameter, Height, and Weight of the poplars are larger for Site 2 than for Site 1. Also, the variable Weight has a very large standard deviation relative to its mean. At Site 2, the minimum weight is only 0.03 kg while the maximum is 1.11 kg. It appears that some of our poplars are doing very well, while others are barely alive.

Step 7: Perform Arithmetic

Now on to the task of predicting how much the trees weigh. Based on previous work, the researchers have found that the weight of a tree is closely related to the square of its diameter multiplied by its height. Since you have diameter and height data, you can calculate this new variable using Minitab's calculator. The calculator performs the equation you enter and puts the result in the variable you specify.

1 Choose Calc > Calculator.
2 You decide to call the new variable "D2H" for diameter squared times height. In Store result in variable, type D2H.
3 In Expression, type C1**2*C2. Click OK. Verify dialog box.

This expression tells Minitab to square the variable Diameter (C1), multiply by the variable Height (C2), and put the result in a new variable called D2H.

Tip You could also use the mouse to create the equation: (1) select Diameter from the variable list, (2) click the **, 2, and * buttons on the calculator, and (3) select Height from the variable list.

The Data window shows the new variable D2H that you just created.

Now save the project changes.

4 Choose File > Save Project, or press [Ctrl] +[S].

Step 8: Create a Scatter Plot

The researchers have determined that there is a relationship between weight and D2H. You want to see if your poplars' data exhibit this relationship as well by plotting Weight by D2H on a scatter plot:

1 Choose Graph > Plot.
2 In Y (the vertical axis), enter Weight.
3 In X (the horizontal axis), enter D2H and click OK. Verify dialog box.

Looking at the scatter plot, you see a positive linear relationship between Weight and D2H. That is, as D2H increases, so does Weight. You also notice an unusual data point—a tree that has a very low weight for a relatively high D2H value. For now, you decide to ignore it, but it is something you may want to check on later. Next, you will compute the correlation between these two variables to quantify the relationship.

Step 9: Compute a Correlation Coefficient

From the scatter plot, you have seen that as D2H increases, so does Weight. Now you want to measure the association between these two variables by computing a correlation coefficient. The correlation coefficient measures the linear relationship between two variables and assumes a value between -1 and +1.

1 Choose Stat > Basic Statistics > Correlation.

2 In Variables, enter Weight and D2H. Click OK. Verify dialog box.

Verify Session window output.

Step 10: Edit the Session Window Output

It is time to create a report of your results: the text results, such as the summary descriptive statistics you computed; and the graphs, such as the scatter plot.
First you will edit the text output in the Session window to make it more appropriate for a report. You can edit text in Minitab's Session window the same way you can edit with a word processor, even finding and replacing text and changing fonts.
By default, the Session window is read-only, so that you cannot accidentally delete results. To begin editing, you will have to make the Session window editable:

1 Press [Ctrl]+[M] to make the Session window active.
2 Pull down the Editor menu. Verify Editor menu.

· If there is no check mark next to Output Editable, then select it to enable Session window editing.
· If there is already a check mark next to Output Editable, then Session window editing is already enabled. Press [Esc] twice to close the menu.

Now you can edit your output.

3 Delete all the text above the Descriptive Statistics output and all the text between the Descriptive Statistics output and the Correlation output. Select the text by dragging over it with your mouse, then delete it by choosing Edit > Cut or pressing [Delete].
4 Scroll to the top of the Session window and type four comment lines as shown below:

Original data set was POPLAR1.MTW
Five lines of data were added.

Two variables were added. Site and D2H = (Diameter**2)* Height.
The resulting data set was stored in the project POPLAR.MPJ.

5 Save your work. Choose File > Save Project.

The Session window is ready to print.

Step 11: Print Your Work

You will first print your output from the Session window, and then your graphs from the Graph windows.

1 With the Session window active, choose File > Print Session Window, then click OK.

You could go to each Graph window and print them separately, but if you have more than one graph there is a faster way.

2 Activate the Project Manager by typing [Ctrl]+[I] or by clicking the Project Manager button on the toolbar.
3 Left-click on the Graphs folder in the Project Manager.
4 In the left pane of the Project Manager, select the four graphs you have created: click below Plot Weight * D2H and drag up. Verify Project Manager.
5 Right-click on the selected graphs and choose Print. Click OK.

Step 12: Save Your Work

When you save your project, you save all your work at once: all the data, all the output in the Session window, and all the open Graph windows. When you reopen the project, all that information will be waiting for you, right where you left it.

1 Choose File > Save Project.

More If you want to use output or data in another application or another Minitab project, you can save your Session window output, data, and graphs as separate files. These separate files are copies of what is currently in your project—the contents of your project are not changed in any way.

Step 13: Exit Minitab

If you want to take a break before continuing to another session, you can exit Minitab.

1 Choose File > Exit.

2 Minitab may ask if you want to save changes to your project. Since you already saved your project above, there is no need to do it again here. Click No.

Otherwise you may continue with Session 2: Doing A Simple Analysis.

Overview of Session Two

The story

Researchers at The Pennsylvania State University planted hundreds of poplar trees and grew them under a variety of controlled conditions. After three years, they measured the diameter in centimeters, height in meters, and dry weight of the wood in kilograms of a sample of trees.
You believe there is a close relationship between the dry weight of wood from young poplar trees and a variable that is a function of the diameter and the height of the trees. But what is that relationship?

These fast-growing, hardy trees may one day serve as an alternative source of fuel or chemicals. As data analyst for the project, you will determine if diameter and height measurements can be used to reliably predict the yield of wood.

What you will learn

In this session, you learn will how to:

· use simple regression to find the relationship between the trees' diameter and height
· find and correct errors in your data, then quickly re-run your analysis
· generate graphs to visualize the relationship between variables
· customize the appearance of those graphs to make them more informative
· brush the graphs to identify key data points

Time required

About 30 minutes.

Proceed with Step 1: Start a New Project.

Step 1: Start a New Project

· If you are not already running Minitab, start the program.
· If you have just completed Session One, start a new project: choose File > New, click Minitab Project, and click OK.

If you have not saved your changes to the previous project, Minitab will give you the chance to do so.

Step 2: Open a Worksheet

You will get data from a Minitab saved worksheet named POPLAR2.MTW that is located in the Data subdirectory or folder.

1 Choose File > Open Worksheet.

2 Move to the Data subdirectory and select the worksheet POPLAR2.MTW.

3 Click Open, then OK (if necessary).

Step 3: Perform a Simple Regression

Towards the end of Session One, you saw that as D2H (the diameter squared multiplied by the height) increases, so does Weight. One way to find out how well D2H predicts weight is to use a simple regression command:

1 Choose Stat > Regression > Regression.
2 In Response, enter Weight.
3 In Predictors, enter D2H. Verify dialog box.

You decide you might as well do a series of plots for residual analysis to check for any potential problems.

4 Click Graphs.
5 Under Residuals for Plots, click Standardized.
6 Under Residual Plots, check Histogram of residuals and Normal plot of residuals.
7 In Residuals versus the variables, enter D2H. Verify dialog box.
8 Click OK in each dialog box.

In the Session window output, Minitab displays the regression equation, the table of coefficients, the analysis of variance table, and—in the table of unusual observations—the identity of the outlier and influential observations (rows 12 and 15). Before proceeding with further analysis, you want to examine rows 12 and 15 to make sure they contain valid data.

Minitab also displays three graph plots. A quick glance at the Residuals Versus D2H plot shows you that the data contains an outlier.

Step 4: Edit the Data

1 To view the worksheet, click on the Data window, choose Window > POPLAR2.MTW, or press [Ctrl]+[D].
2 Now go to the first unusual observation, in row 12 of the column named Weight:

· Choose Editor > Go To. (Select the first Go To.)
· In Enter column number or name, type Weight.
· In Enter row number, type 12 and click OK. Verify dialog box.

The Data window now shows the 12th observation of Weight as the highlighted cell.

Both Weight and D2H seem rather large, so you double-check the researchers' log sheets. It turns out that poplar #12 is a very healthy tree—the values are correct.

3 Click on the Weight value in row 15 to highlight it, or press three times.

Double-checking the log sheet shows that this value is actually an error. The correct value should be 0.70, not 0.07.

4 Type .7 and press [Enter]. Verify data window.

Step 5: Run the Regression Again

Now you are ready to run the regression again. Simply repeat the menu selection you made earlier. The Regression dialog box and Graphs subdialog box will contain the same settings as before. You are ready to go!

1 First, close all the graphs that you created before correcting the data. Choose Window > Close All Graphs and click OK.
2 Choose Stat > Regression > Regression and click OK.

Tip To set a dialog box back to its defaults, press [F3].

As before, Minitab displays the text output in the Session window, and displays each of the three graphs in its own Graph window. First, look at the Session window output.

If you have a good model and have satisfied all the statistical assumptions, then you can measure the diameter and height of any poplar in this population and be able to predict its weight without cutting it down, drying it, and weighing it on a scale.

From the regression output, you see a high t-ratio and a low p-value for D2H in the table of coefficients, indicating strong evidence of a relationship between D2H and Weight. The large F-statistic and low p-value in the analysis of variance table quantify this relationship in a different way. The R2 and adjusted R2 values of greater than 98%

further reinforce the assertion that there is a strong linear relationship between D2H and Weight.

Before making a final conclusion, however, you decide to look at the plots.

You notice from the Residuals Versus D2H plot that the variance does not appear to be constant—an important assumption for a regression model to meet. It is larger at bigger values of D2H. In the interest of time we will continue with our session, but this is something you would want to examine more closely.

Step 6: Draw a Fitted Regression Line

Next, you want to display a scatter plot with the regression line drawn on it to see how closely the measured data lie to the least-squares regression line.
To plot Weight versus D2H:

1 Choose Stat > Regression > Fitted Line Plot.

2 In Response (Y), enter Weight.

3 In Predictor (X), enter D2H. Verify dialog box. Click OK.

Verify Graph window output.

Step 7: Change the Graph Title

You would like to change the title of your graph. You could redo the dialog box and add a title in the Options subdialog box. You can also edit the graph directly, after it has been created.
Minitab's Graph Editor is very similar to most drawing packages. If you know how to use a drawing package, you should be able to edit Minitab graphs very easily.
In this step, you will learn how to:

· enter graph editing mode
· change the text of the title
· resize the text box so that the title fits on one line

Put the graph in edit mode

1 Make the Graph window active by clicking on it or choosing its name from the Window menu.
2 Maximize the Graph window.
3 Choose Editor > Edit.

More A graph can be in one of three modes: View mode allows you to view your graph but nothing else, Edit mode allows you to edit your graph, and Brush mode allows you to identify data points. You choose the mode from the Editor menu.

4 Two palettes should appear. If they do not, open them:

· Choose Editor > Show Tool Palette.
· Choose Editor > Show Attribute Palette.

Your Graph window should look like this.

Your palettes may be in different positions. You can move a palette around by dragging the title bar at the top of the palette.

Change the text of the title

1 On the Tool palette, click the selection tool if it is not already selected.
2 Click anywhere on the title and press [Enter], or double-click on the title.

A text box containing the current title will appear on the screen.

3 In the box, edit the title to say Regression Plot for Poplar2 Data. Click OK.

Resize the text box of the title

Because the title is longer, it is now on three lines. You need to resize the box that surrounds the title.

1 If necessary, click on the title to make it active. Handles will appear.
2 Place the cursor over the second handle up on the right edge. The cursor changes to crossed diagonal lines.
3 Click and hold the mouse button down and drag the right edge so the title is on two lines.
4 Place the cursor over the second handle up on the left edge. Click and hold the mouse button down and drag the left edge so the title is on one line.

Now you can position the title where you want it.

1 Put the cursor in the middle of the title.
2 Click and hold the mouse button down and drag the title to the position you want. Release the mouse button.

Step 8: Make the Regression Line Red

You decide to make the regression line red, so it is easier to see.

1 Click anywhere on the regression line. Handles will appear.
2 On the Attribute palette, click the line color tool . A Color palette appears.
3 Click the red square. The line will become red.
3 Click in a blank space on the graph to remove the handles.

Step 9: Brush the Graph to Identify Points

One point has a very large value for D2H. You want to know what point this is. Brushing allows you to identify points on a plot. You will switch to Brush mode, and open the Brushing palette if necessary.

1 Choose Editor > Brush. Your cursor will change to a hand.
2 If the Brushing palette does not appear (see below), choose Editor > Show Brushing Palette.
3 Click on the point you want to identify. Its row number appears in the Brushing palette.

Suppose you would like more information on points you select. You can include data for up to ten worksheet columns in the Brushing palette.

1 Choose Editor > Set ID Variables. Verify dialog box.
2 Check Use columns, then enter C1–C4.
3 Click OK.

The Brushing palette widens to display the additional information. If you like, you can move it so both the palette and the plot can be easily seen.

You want to identify the two points on the extreme right of the plot. You can select a block of points by enclosing them in a rectangle. To draw the rectangle, you begin at the upper-left corner and drag down to the lower-right corner of the rectangle.

4 Move the cursor to the location where you want to begin drawing the rectangle. This location will be the upper-left corner.
5 Hold the mouse button down and drag down to the lower-right until the rectangle encloses the two points.

The points are enclosed in a rectangle and identified in the Brushing palette.

These are the same points identified previously as being an outlier and influential observation. Here, brushing lets you quickly see the diameter, height, weight, and site information for these points.

The brushed points are also marked in the Data window.

Step 10: Save and Exit

1 Choose File > Save Project.
2 In File name, type POPLAR2 for the name of your project. If you omit the extension .MPJ, Minitab will automatically add it once you save the project.
3 Click Save.
4 If you see a message box asking if you want to replace an existing file, click Yes.
4 If you want to take a break at this point, you can exit Minitab by choosing File > Exit, or you can continue on with Session Three: Advanced Minitab.

Overview of Session Three

The story

How feasible are energy plantations? How much wood for energy can you realistically expect from these plantations, and how can you maximize yield?
In an effort to maximize yield, researchers designed an experiment to determine how two factors, Site and Treatment, influence the weight of four-year-old poplar clones. They planted trees on two sites: Site 1—a moist site with rich soil, and Site 2—a dry, sandy site. They applied four different treatments to the trees: Treatment 1 was the control (no treatment), Treatment 2 was fertilizer, Treatment 3 was irrigation, and Treatment 4 was both fertilizer and irrigation. To account for a variety of weather conditions, the researchers replicated the data by planting half the trees in Year 1, and the other half in Year 2.

As data analyst for the project, you will perform the statistical analysis on the sample data stored in the Minitab file called POPLAR3.MTW.

What you will learn

In this session, you will learn how to:

· quickly generate basic statistics to describe the variables you are interested in
· change the codes the field researchers were using for missing values into missing value codes that Minitab will recognize
· subset the data to focus on just the group of trees that you want to examine further
· create boxplots to see at a glance the differences between categories of trees
· use analysis of variance to determine which variables are contributing to the differences between trees

Time required

About 40 minutes.

Proceed to Step 1: Start a New Project.

Step 1: Start a New Project

- If you are not already running Minitab, start the program.
- If you have just completed Session Two, start a new project: choose File > New, click Minitab Project, then click OK.

If you have not saved your changes to the previous project, Minitab will give you the chance to do so.

Step 2: Open a Worksheet

You will get data from a Minitab saved worksheet named POPLAR3.MTW that is located in the Data subdirectory or folder.

1 Choose File > Open Worksheet.
2 Move to the Data subdirectory and select the file POPLAR3.MTW. Click Open.

Two windows can show you information about this worksheet.

3 Choose Window > Project Manager or press [Ctrl]+[I].

If the left page of the Project Manager is too narrow, simply click the divider between the panes and drag to the right.

4 In the left pane of the Project Manager, click on the Columns folder. Verify.

The right pane displays information about the data in the worksheet.

5 If the worksheet is not visible, open the Data window by pressing [Ctrl]+[D].

The Data window shows you the columns of data in detail.

This worksheet contains seven variables: Site, Year, Treatment (experimental treatment), Diameter (cm), Height (m), Weight (kg), and Age (years).

Tip If you want to adjust the column widths to fit the data, point with your mouse to the top of a line dividing the two columns until the mouse cursor turns into a two-sided arrow. Then, press the mouse button down and drag the column border to make it wider or narrower.

Step 3: Generate Descriptive Statistics

You want to maximize yield, so you will focus on what factors influence the weight of trees. Begin by looking at the descriptive statistics for the variable Weight.

1 Choose Stat > Basic Statistics > Display Descriptive Statistics. Verify dialog box.
2 In Variables, enter Weight. Click OK.

When you examine the session window output, notice the minimum value for Weight. It is certainly impossible to have a weight of -99 kilograms! The real story here is that our data gatherers in the field recorded the value -99 to represent a dead tree.

Leaving values of -99 in the worksheet is going to considerably throw off any analyses you do. In fact, it has already affected the output of the descriptive statistics just computed. The means and medians are artificially low, while the standard deviation is artificially high. You need to convert all -99's to missing values.

Missing values do not affect the results of any statistical analyses. Minitab represents a missing value for numerical data with an asterisk (*).

Step 4: Recode the Data

Minitab provides many data manipulation tools. One of the most useful is the Code command, which allows you change all the occurrences of one value into another value. In this case, you want to change all the -99's to *, the missing value symbol.

1 Choose Manip > Code > Numeric to Numeric.
2 In Code data from columns, enter Weight.
3 In Into columns, enter Weight. This will replace the old values in Weight with the new, coded values.
4 In Original values, type -99. This is the value you want to change.
5 In New, type *. This is the missing value symbol. Click OK. Verify dialog box.

In the Data window, you will see that all occurrences of the value -99 in the variable Weight have been replaced with *, the code for a missing data value.

Step 5: Tally the Data

How many trees of each age are you dealing with? Use the Tally command to find out:

1 Choose Stat > Tables > Tally.
2 In Variables, enter Age. Click OK. Verify dialog box.

The session window output shows that you have 147 three-year-old trees and 151 four-year-old trees.

Step 6: Split the Data by Age

Suppose you want to analyze the data for just the four-year-old trees. Here is a technique you can use to create a new data set with just the four-year-old trees.

Make a separate worksheet for the four-year-old trees

1 Choose Manip > Split Worksheet. Verify dialog box.
2 In By variables, enter Age. Click OK.

Minitab will split the POPLAR3 worksheet using the values of Age. Since there are two unique values in the age column (3 and 4), Minitab will create two new worksheets. The worksheet that contains the data for the three-year-old trees will be named POPLAR3.MTW(Age = 3); the worksheet that contains the data for the four-year-old trees will be named POPLAR3.MTW(Age = 4).

Rename the worksheet that contains the four-year-old trees' data

1 Press [Ctrl]+[I]. Verify dialog box.
2 In the left pane of the Project Manager, right-click on POPLAR3.MTW(Age = 4).
3 Click Rename. Verify dialog box.
4 Type 4YROLDS.MTW.
5 Press [Enter].

You will now perform the analysis on the four-year-old poplar data.

Step 7: Check for Normality with a Histogram

You will now create a histogram of the variable Weight.

1 Make sure the Data window named 4YROLDS.MTW is active. To make a Data window active, click on it, or choose its name from the Window menu. Notice that the active Data window has asterisks after its name.

2 Choose Graph > Histogram.
3 In X, enter Weight. Click OK. Verify dialog box.

The histogram appears in its own window.

You see from the histogram that the weights of the poplars are approximately normally distributed (in a bell-shaped curve).

Step 8: Compare Weight by Treatment with Boxplots

Now you will want to look at the weight for each treatment. Boxplots are good for graphically comparing different levels of a variable.

1 Choose Graph > Boxplot.
2 In Y, enter Weight.
3 In X, enter Treatment. Verify dialog box.

This tells Minitab to produce a separate boxplot of weight for each treatment.

4 Click OK.

Examine the boxplot of Weight by Treatment. The line drawn across each box indicates the median, or middle, of the data. The bottom and top edges of the box mark the first (25th percentile) and third (75th percentile) quartiles, respectively.

The boxplots suggest that Treatments 2 and 4 (fertilizer and fertilizer/irrigation) have produced the heaviest trees, while Treatments 1 and 3 (control and irrigation) have yielded lighter trees.

You might also expect the site to have an impact on weight. The Site 1 trees planted in the rich, well-drained soil would be expected to weigh more than the Site 2 trees planted in the dry, sandy soil. You can determine if this assumption is true by looking at a boxplot of weight for each site.

Rather than repeat your previous menu selection from the beginning, recall the last Boxplot dialog box and change the X, or category, variable.

5 Choose Edit > Edit Last Dialog, or press [Ctrl]+[E].
6 In X, enter Site. Verify dialog box.
7 Click OK.

The new boxplot will appear. Surprisingly, the Site 1 tree weights do not seem very different from the Site 2 tree weights. The spreads are different for each site, but the medians are almost the same.

Step 9: Perform an Analysis of Variance

You have seen from the boxplots that poplar weights differ noticeably among the four treatments, but not as noticeably between the two sites. Now you decide to use analysis of variance to see if there are statistically significant differences in weight due to the different levels of the factors site and treatment.
When you have two or more factors, Minitab gives you a choice between balanced ANOVA and the general linear model (GLM). Balanced ANOVA requires a balanced design, that is, you must have the same number of observations for each site/treatment combination. Because your design is not balanced, you will use the general linear model.

1 Choose Stat > ANOVA > General Linear Model.
2 In Responses, enter Weight.

Next, you will enter the model you want GLM to fit. You decide to look at a model with Site, Treatment, and the Site*
Treatment interaction.

3 In Model, type Site | Treatment.

The vertical bar tells Minitab that you want to include all possible interactions in the model. To make a vertical bar on most keyboards, press [Shift]+[\], or you can use the symbol ! instead.

4 Click OK. Verify dialog box.

The session window output for the GLM lists each factor in the model and the number of levels in each factor. Next, GLM lists the analysis of variance table, and finally, it lists unusual observations.

Suppose you want to perform an F-test for each effect in the model. For example, to test the null hypothesis that the treatment effect is the same for both sites (the Site*Treatment interaction), compare Minitab's p-value with the commonly used alpha level of 0.05. Because the p-value is 0.091 (a value larger than 0.05) you cannot reject the null hypothesis. That is, you cannot conclude that the treatment effect differs for the two sites.

Now you can look at the main effects, Site and Treatment. The Site p-value of 0.219 is also larger than 0.05, so you cannot conclude that poplar weights differ significantly between the two sites. The p-value for Treatment is small (0.000) thereby supporting the conclusion that mean weights do differ significantly for different treatments.

This agrees with what you saw earlier in the boxplots—that poplar weights were different for different treatments, but only varied slightly between the two sites. Before you decide that Treatment is the only important factor influencing poplar weight, take a look at the Year effect—remember that the researchers planted half the trees in Year 1 and half in Year 2.

Step 10: Compare Weight by Year with Boxplots

You decide to look at a boxplot to compare the weight of poplars planted in Year 1 with those planted in Year 2.

1 Choose Graph > Boxplot.
2 In Y, enter Weight.
3 In X, enter Year.

This says to draw a separate boxplot of weight for each year.

Notice the first two rows of the Data display table. IQRange Box instructs Minitab to display a box showing the interquartile range, from the 25th to the 75th percentile. Outlier Symbol instructs Minitab to display an asterisk (*) for all outlier values. You decide to also display a confidence interval box within the IQ Range Box.

4 In the Data display table, in the Display column, click in the row for item 3.
5 Click the down arrow beside Display, and choose CI Box.
6 Click in the next cell to the right.
7 Click the down arrow beside For each and choose Graph. Verify dialog box.

This row tells Minitab to include a confidence interval on each boxplot.

By default, Minitab draws boxplots vertically, but you also can draw them horizontally.

8 Click Options.
9 Check Transpose X and Y. Verify dialog box. Click OK in each dialog box.

In the boxplot, the inner boxes show a 95% confidence interval for the median. The boxplot suggests that poplars planted in Year 2 are heavier than those planted in Year 1. But why is year important? Trees were planted in two different years simply to replicate the data.

You interview the field researchers and learn that they did not apply herbicides to control weeds during the first year planting. As a result, many young trees either died or were severely stunted. To improve the trees' ability to survive, researchers did apply herbicides when they planted poplars the second year.

You draw three preliminary conclusions from your analysis. One, fertilization appears to be an effective way to maximize the weight of poplar clones. Two, it is important to control weeds while the trees are very young. Three, given proper planting and nutrient conditions, poplar clones may not require a high-quality site in order to yield a substantial amount of biomass.

Not only were the Year 2 trees heavier, their weights were more consistent. But before you recommend the use of herbicides and fertilizers, you want to look more closely at the Year 2 trees.

Specifically, you want to know if you still see Site and Treatment effects, when you look at the Year 2 trees alone.

Step 11: Quickly Repeat the Entire Analysis

You decide to repeat the analysis on Year 2 trees only. First, you need to create this subset by splitting the data you just used for the four-year-old trees. Then, rather than redoing all the dialog boxes to repeat the analysis, you will use Minitab's Command Line Editor.

Split the 4YROLDS worksheet using the values of Year

1 Make sure the Data window named 4YROLDS.MTW is active. To make a Data window active, click on it, or choose its name from the Window menu.
2 Choose Manip > Split Worksheet.
3 In By variables, enter Year. Click OK.

Rename the worksheet that contains the data for the Year 2 trees

1 Press [Ctrl]+[I].
2 In the left pane of the Project Manager, right-click on 4YROLDS.MTW(Year = 2).
3 Choose Rename.
4 Type YEAR2.MTW.
5 Press [Enter].

Repeat the analysis on the Year 2 trees

1 Make sure the Data window named YEAR2.MTW is active. To make a Data window active, click on it, or choose its name from the Window menu.

2 Press [Ctrl]+[I].

3 In the left pane of the Project Manager, click the History folder.

The right pane of the Project Manager displays all the commands you have executed in the current project.

4 Scroll through the commands until you find "Histogram."

This was the command you used to check for normality.

5 Click on the "Histogram" command to select it.

6 Scroll to the bottom of the commands, hold down [Shift], and click the last command.

All the commands from “Histogram” to the end are selected. Verify history folder contents.

7 Choose Edit > Command Line Editor, or as a shortcut, press [Ctrl]+[L].

A dialog box appears containing the Minitab commands from the section you highlighted. Verify dialog box.

8 Click Submit Commands.

The entire analysis, a histogram of Weight, boxplots of Weight by Treatment and Weight by Site, an analysis of variance, and a boxplot of Weight by Year are all done, with no further work.

Tip You can also select consecutive commands in the History folder by clicking and dragging through them. You can select nonconsecutive commands by holding down [Ctrl] while you click them.

Step 12: Save and Exit

1 Choose File > Save Project.
2 In File name, enter POPLAR3 for the name of your project. If you omit the extension .MPJ, Minitab will automatically add it once you save the project.
3 Click OK.
4 If you see a message box asking if you want to replace an existing file, click Yes.
5 If you want to take a break at this point, you can exit Minitab by choosing File > Exit, or you can go on to Session Four: Quality Control and Improvement.
6

Overview of Session Four

The story

You work for an automobile manufacturer in a department that assembles engines. One of the parts, a camshaft, must be 600 mm ±2 mm long to meet engineering specifications. There has been a chronic problem with camshaft length being out of specification—a problem that has caused poor-fitting assemblies down the production line and high scrap and rework rates.
Your supervisor wants to run Xbar and R charts to monitor this characteristic. For a month, data are collected on the length of five camshafts per shift (1 sample of size 5 per shift). You have been asked to lead a problem-solving team and recommend a solution.

What you will learn

In this session you will learn how to:

- produce Xbar and R charts
- produce histograms with normal curves
- perform a process capability analysis

Time required

About 30 minutes.

Proceed with Step 1: Start a New Project.

Step 1: Start a New Project

- If you are not already running Minitab, start the program.
- If you have just completed Session Three, start a new project: choose File > New, click Minitab Project, then click OK.

If you have not saved your changes to the previous project, Minitab will give you the chance to do so.

Step 2: Open a Worksheet

You will get data from a Minitab saved worksheet named CAMSHAFT.MTW that is located in the Data subdirectory or folder.

1. Choose File > Open Worksheet.
2. Move to the Data subdirectory and select the worksheet CAMSHAFT.MTW. Click OK.
3. If it is not visible, open the Data window by pressing [Ctrl]+[D].

The Data window shows you the columns of data in detail.

This worksheet contains the results of the sampling plan from the last month. For now, you are concerned with the first column, Length, which contains 100 observations (20 samples of 5 camshafts each). Recall that the camshaft lengths are measured in millimeters.

Step 3: Examine Ranges with an R Chart

First, you want to produce a control chart to look at the range of camshaft lengths within the sample subgroups. You hope that the plotted points fall inside the control limits in a random manner.

1. Choose Stat > Control Charts > R.
2. Choose Single column and enter Length.

3 In Subgroup size, type 5. Verify dialog box.
4 Click OK.

The R chart for Length does not show any points out of control.

Step 4: Test for Special Causes with an Xbar Chart

You will create an Xbar chart to see if there is a problem with camshaft lengths being outside acceptable limits. In addition, you will instruct Minitab to use eight common tests that point out special causes for variation.

1 Choose Stat > Control Charts > Xbar.
2 Choose Single column and enter Length.
3 In Subgroup size, type 5. Verify dialog box.
4 Click Tests.
5 Check Perform all eight tests. Verify dialog box.
6 Click OK in each dialog box.

The Xbar chart shows that the process is out of control. Specifically, one point has failed test 1, and two points have failed test 6. To find out what these tests mean, look in the Session window.

7 Choose Window > Session.

By looking at the Session window output, you see that the process produced one point more than 3 sigmas from the center line, and four of five points more than 1 sigma from the center line (on one side of the center line).

Now that you have confirmed that a problem does exist, it is time to look for causes and solutions. Unfortunately, the sampling plan did not allow for detailed inspection of precisely where and when the problems occurred because only one sample was taken per shift. A better plan would have been to take multiple samples per shift for the troubleshooting phase, and to switch to this monitoring plan after special causes were found and eliminated. Nonetheless, you are determined to get what you can out of the data that you have.

Step 5: Create a Histogram with Normal Curve

The histogram with normal curve is useful for examining a variable's distribution. You decide to examine the variable Length.

1 Choose Stat > Basic Statistics > Display Descriptive Statistics.
2 In Variables, enter Length. Verify dialog box.
3 Click Graphs.

The Graphs subdialog box appears.

4 Check Histogram of data, with normal curve.

5 Click OK in each dialog box.

Examine the histogram. In general, we expect a variable such as Length to follow the normal distribution. In this case, the histogram would be approximately bell-shaped. The histogram you just created is certainly not bell-shaped. In fact, it would appear from the spikes at 598, 599, and 601 that we may be dealing with more than one separate and distinct distributions.

An examination of the inventory records indicates that there are two suppliers for the camshafts. Now you are starting to understand the odd histogram. You decide to obtain measurements from both suppliers and run Xbar and R charts separately on each set of data with a subgroup size of 5 for each. The data for each supplier are stored in the columns Supp1 and Supp2 of your worksheet.

Step 6: Display Combined Xbar and R Charts

Your worksheet contains variables named Supp1 and Supp2 with data for Suppliers 1 and 2. You could repeat the same procedures to produce control charts for Supplier 1 as you did to produce the charts for Length. However, there is another command, Xbar-R, that you can use to display both charts together.

More The Xbar-R command is actually a macro—a collection of Minitab commands that are executed as a single command. You can use macros to automate repetitive tasks as well as to create you own custom Minitab feature.

Evaluate Supplier 1

1 Choose Stat > Control Charts > Xbar-R.
2 Choose Single column and enter Supp1. Verify dialog box.
3 In Subgroup size, type 5. Click OK.

According to the Xbar-R Chart, both the means and ranges for Supplier 1 appear to be in control, although you notice that the mean is 599.5 mm, not 600. The average range for Supplier 1 is 1.36 mm.

Evaluate Supplier 2

You can produce the same control charts for Supplier 2, using the variable named Supp2, as you did for Supplier 1.

1 Press [Ctrl]+[E].

This keyboard shortcut, for Edit > Edit Last Dialog, brings up the Xbar-R Chart dialog box again.

2 In Single column, enter Supp2. Click OK.

You do not need to enter a subgroup size because it was still set to 5 from the last time you used this dialog box. Minitab “remembers” the dialog box settings from the last time a dialog box was used in a session.

Supplier 2’s Xbar-R chart reveals problems. From the chart, you can see that two points are above the upper control limit.

The R chart does not indicate that the process is out of control. However, you notice that the center line is at 3.72, which is almost three times larger than Supplier 1’s Rbar of 1.36.

As team leader, you recommend that longer production runs be accepted from Supplier 1 until Supplier 2 can demonstrate that camshaft production is in control. You will work with Supplier 2 to reduce process variability to an acceptable level. Because of the statistical evidence to support your position, your recommendation is implemented.

Step 7: Preparing for a Process Capability Analysis

So now you have managed to reduce variability by using only Supplier 1. The number of poor-quality assemblies being produced down the line has dropped significantly, but problems have not completely disappeared. You decide to run a capability study to see whether Supplier 1 alone is capable of meeting your engineering specifications of 600 mm ±2 mm.
Before you can proceed with capability analysis, the process must be in control. The control charts have shown that, thanks to your recommendation, the process is now in control. You also expect the camshaft lengths to be normally distributed. Now you want to view a histogram to check normality.

First you want to look at the distribution of camshaft lengths for Supplier 1:

1 Choose Stat > Basic Statistics > Display Descriptive Statistics.
2 In Variables, enter Supp1. Click OK.

Minitab produces a histogram with normal curve again, based on your earlier selections.

You are satisfied by the bell shape of the distribution, and you do not see multiple modes or peaks as you did earlier.

You are ready to proceed.

Step 8: Perform a Process Capability Analysis

Now you are ready to run a process capability analysis to see if Supplier 1 is capable of meeting your engineering specifications of 600 mm ±2 mm.

1 Choose Stat > Quality Tools > Capability Analysis (Normal).
2 Choose Single column and enter Supp1.
3 In Subgroup size, type 5.

Next, enter the specification limits.

4 In Lower spec, type 598.
5 In Upper spec, type 602. Verify dialog box.
6 Click Options.
7 In Target (adds Cpm to table), type 600. Verify dialog box.
8 Click OK in each dialog box.

From the graph you can see that the process mean falls short of the target and the process distribution mean lies to the left of the target. Also, the left tail of the distribution falls outside the lower specification limit. Therefore, some camshafts will not meet the lower specification of 598 mm.

The Capability Analysis command also produces a table of statistics. The Cpk index indicates whether the process will produce units within the tolerance limits. A Cpk index of 1 means that a process is exactly capable of meeting specifications, while less than 1 means that the process is not meeting specification limits. Ideally, you would like to see a Cpk much larger than 1, because the larger the index, the more capable the process. The Cpk index for Supplier 1 is only 0.90, indicating that they need to improve by reducing variability and by centering the process around the target.

Since Supplier 1 is currently your best supplier of camshafts, you will work with them to improve their process and, therefore, your own. Minitab offers analysis of variance (ANOVA), regression, design of experiments (DOE), and many other statistical tools that you will use to continuously improve your processes.

Step 9: Save and Exit

1 Choose File > Save Project.
2 In File name, enter CAMSHFT1 for the name of your project. If you omit the extension .MPJ, Minitab will automatically add it once you save the project.
3 Click OK.
4 If you see a message box asking if you want to replace an existing file, click Yes.
7 If you want to take a break at this point, you can exit Minitab by choosing File > Exit, or you can go on to Session Five: Designing an Experiment.

Overview of Session Five

The story

For this lesson, assume that you work at a chemical plant. You are studying one of the reactions that produces a chemical product. You would like some way to increase the yield of a product that is produced from the reaction. From past experience, you have seen that varying the temperature, the pressure, and the type of catalyst seems to change the yield of the reaction. A problem is that everyone you work with has their own theory about how each of these factors affects the reaction. You want to make real improvements, so you decide to run an experiment to determine the actual effects of the three factors.

What you will learn

In Session Five, you will learn how to:

- design a factorial experiment to tell which factors are important to the reaction
- fit a full model to the data
- use several simple graphical methods to help determine which effects are active (important) or inactive
- fit a reduced model to the data, and then assess the adequacy of the model

Time required

About 30 minutes.

Step 1: Start a New Project

If you are not already running Minitab, start the program.

- If you have just completed Session Four, start a new project: choose File > New, click Minitab Project, then click OK.

If you have not saved your changes to the previous project, Minitab will give you the chance to do so.

Step 2: Select a Design

Suppose you want to design an experiment to test three factors: time, temperature, and type of catalyst.

1 Choose Stat > DOE > Factorial > Create Factorial Design. Verify dialog box.
2 Click Display Available Designs. Verify dialog box.

Since you have three factors that are of interest, you can see in the table that you have two options. You can choose

- a fractional factorial design of resolution III with 4 runs, or
- a full factorial design with 8 runs

A two-level design with three factors has 2**3 (or eight) possible factor combinations. By choosing a design with all possible combinations, called a full factorial design, you will get results that show effects free from confounding, that is, all effects are distinguishable from other effects. However, you may also be able to obtain meaningful results by doing fewer runs or combinations. Designs that use less than all possible combinations are called fractional factorial designs.

You decide that the full factorial design with 3 factors and 8 runs is more appropriate than the fractional factorial design. At your chemical plant, runs that manipulate the factors of interest—time, temperature, and type of catalyst
—are not expensive or time-consuming. Also, the experiment can be performed at a non-peak period without disturbing the workflow at the plant. If the runs of the experiment were costly or difficult to perform, you may have made a different decision.

3 Click OK. You are now back in the main dialog box.
4 Choose 2-level factorial (default generators).
5 In Number of factors, choose 3.
6 Click Designs.

The box at the top shows all available designs for the design type and number of factors you selected.

7 In the Designs box, select Full factorial.
8 In Number of replicates, choose 2. Verify dialog box.
9 Click OK. This selects the design and brings you back to the main dialog box. Notice that the remaining buttons are now enabled.

Step 3: Name Factors and Set Factor Levels

You can enter factor levels (settings) as numeric or text. If your factors could be continuous, use numeric levels; if your factors are categorical, use text levels. Continuous variables can take on any value on the measurement scale being used (for example, length of reaction time). In contrast, categorical variables can only assume a limited number of possible values (for example, type of catalyst).
You now need to choose settings for your factors. In a two-level factorial design, you set factors at two levels. Many experimenters advocate choosing limits as far apart as possible (within limits of safety if you know them). After some deliberation, you choose the following settings:

Factor	Low Setting	High Setting
Temperature	20° C	40° C
Pressure	1 atmosphere	4 atmospheres
Catalyst	A	B

1 Click Factors.
2 Click on the first row of the Name column to change the name of the first factor. Then, use the arrow keys to navigate within the table, moving across rows or down columns. In the row for:

· Factor A, type Temp in Name, 20 in Low, and 40 in High.
· Factor B, type Pressure in Name, 1 in Low, and 4 in High.
· Factor C, type Catalyst in Name, A in Low, and B in High.

Verify dialog box.

3 Click OK. This brings you back to the main dialog box.

More If you have a design that includes center points and you have both numeric and text factors, you need to be aware that there really is no true center to the design. In this case, center points are called pseudo-center points See Help or Chapter 19, Factorial Designs, in Minitab User's Guide 2 for a discussion of pseudo-center points.

Step 4: Randomize and Store the Design

1 Click Options.
2 In Base for random data generator, type 9.

Entering a base for the random data generator allows you to control the randomization so that you obtain the same pattern every time. This way you will get the same design order that is used in this sample session.

3 Make sure Store design in worksheet is checked. Verify dialog box. Click OK.
4 You are now back in the main dialog box. Click OK. This will generate the design and store the design in the worksheet.

Tip It is usually a good idea to randomize the run order. Randomizing the order of the runs lessens the effects of other factors that are not included in the study, particularly effects that are time-dependent.

Step 5: View the Design

Open the Data window so you can see what the structure of a typical design looks like.

1 Choose Window > Worksheet 1, or as a shortcut, press [Ctrl]+[D].

The Data window should now look like this:

Notice the columns named StdOrder (C1) and RunOrder (C2). Every time you create a design, Minitab reserves C1 and C2 to store the standard order and run order.

· StdOrder shows what the order of the runs in the experiment would be if the experiment were done in standard, or Yates', order.
· RunOrder shows what the order of the runs in the experiment would be if the experiment was run in random order.

If you do not randomize a design, the standard order and run order are the same.

In addition, Minitab stores the center point indicators in C3 and the block numbers in C4. Since you did not add center points or block the design, Minitab sets all the values in C3 and C4 to one.

Next in the worksheet are the factor columns, beginning with C5. In this example, the factors are in C5 through C7. Since you entered the factor levels in the Factors subdialog box, you see the actual levels in the worksheet.

More   You can use Stat > DOE > Display design to switch back and forth between a random and standard order display, and between a coded and uncoded display in the worksheet.

There are two ways to change the factor settings or names: use Stat > DOE > Modify Design, or type new factor names directly in the Data window.

Step 6: Collect and Enter Data in the Worksheet

At this point, you may want to create a data collection form for your experiment. Print the Data window with its grid lines:

1 In the Data window, click on the name field of C8 and type Yield.
2 Choose File > Print Worksheet, and make sure Print Grid Lines is checked. Click OK.

Now you would perform all sixteen runs of the experiment, and record the observed yields. Suppose you came up with the following product yields (in grams):

66 66 102 98 65 54 107 68 53 66 55 85 108 89 52 63

3 Type the observed yields into the Yield column of the Data window:

Step 7: Screen the Design — Fit a Model

When you screen a design, the object is to select factors that have large effects. Now that you have created a factorial design and collected the response data, you can fit a model to the results and generate some graphs to evaluate the effects. You will use the output from fitting a mathematical model, and you will also use two graphical methods to help see which factors are important for improving the yield in the reaction.

Fit a model

Since you have created and stored a factorial design, you will notice that Minitab has enabled the DOE > Factorial menu commands Analyze Factorial Design and Factorial Plots. If you plot the responses rather than the fitted values (least-squares means), you can generate main effects plots, interaction plots, and cube plots either before or after you actually fit a model. In this sample session, you will fit the model first.

1 Choose Stat > DOE > Factorial > Analyze Factorial Design.
2 In Responses, enter Yield. Verify dialog box.
3 Click Graphs.
4 To generate two effects plots that will help you determine which effects are active, check Normal and Pareto. Use the default a level (0.10). Verify dialog box.
5 Click OK. This brings you back to the main dialog box.

Now you have selected the model you want to fit, the graphs you want to display, and you have set all other options.

8 To display the requested output in the Session window, and each graph in a separate Graph window, click OK in the main dialog box.

Step 8: Identify Important Effects

You can use both the Session window output and the two effects plots to help you decide which effects are important to your process. First, you will look at the Session window output.
You fit the full model, which includes the three main effects, three two-way interactions, and one three-way interaction. Use the values in the P column of the Estimated Effects and Coefficients table to determine which of the effects are significant. Using a = 0.05, the main effects for Pressure and Catalyst, and the Pressure*Catalyst interaction are significant; that is, their p-values are less than 0.05.

Step 9: Screen the Design — Effects Plots

Now you can use the normal probability plot and the Pareto chart of the effects to see which effects influence the response, Yield.
Active effects are effects that are significant or important. In the normal plot of the effects, points that do not fit the line well usually signal active effects. Active effects are larger and further from the fitted line than inactive effects. Inactive effects tend to be smaller and centered around zero, the mean of all the effects.

1 To make the normal probability plot appear in the active window, choose Window > Effects Plot for Yield. Verify effects plot.

The normal probability plot labels effects that are lower than the a level you chose in the Analyze Factorial Design–
Graphs subdialog box. Here, the effects of Pressure, Catalyst, and the Pressure*Catalyst interaction are significant using a = 0.10.

A Pareto chart of the effects is another useful tool that you can use to help determine which effects are active.

2 To make the Pareto chart appear in the active window, choose Window > Effects Pareto for Yield. Verify Pareto chart.

Minitab displays the absolute value of the effects on the Pareto chart.

The Pareto chart uses the same a as the normal plot to determine the significance of effects. So again, you see that Pressure, Catalyst, and Pressure*Catalyst are significant (a = 0.10).

Later, you will fit a model without the terms Temp, Temp*Pressure, and Temp*Catalyst, which seem to be inactive. You will check to see how good the model is after you fit the reduced model.

Step 10: Fit a Reduced Model

Next, you want to fit a new model using only the terms you identified as important by looking at the results of fitting the full model—in other words, screening out the unimportant effects. After you fit the model, you will generate several plots to visualize the effects, evaluate the fit of the reduced model, and do a residual analysis.
You will fit a model that includes Pressure, Catalyst, and the Pressure*Catalyst interaction.

1 Choose Stat > DOE > Factorial > Analyze Factorial Design.
2 Click Terms.
3 Set up the model you want to fit.

· From Include terms in the model up through order, choose 2. Notice this moves ABC to the Available Terms box.
· Click on A:Temp in the Selected Terms box, then click . This will move the A:Temp variable to the Available Terms list box.
· Repeat these actions to move the AB and AC interactions to the Available Terms box.

4 Verify dialog box. Click OK. You are now back in the main dialog box.
5 Click Graphs. Uncheck Normal and Pareto.
6 Check Histogram, Normal plot, Residuals versus fits, and Residuals versus order. Verify dialog box. Click OK and return to the main dialog box.
7 Click OK in the Analyze Factorial Design dialog box.

The output will display in the Session window and the residual plots will display in Graph windows.

Was your choice of active effects a good one? Is your model valid? You will try to answer these questions next when you evaluate the reduced model.

Step 11: Evaluate the Reduced Model

The Session window output provides information as to how good the model is. Examine the P column, which contains p-values for each of the terms in the model. A good standard by which to evaluate the model is to look at p-values. If all terms have p-values less than the a level appropriate for your experiment, you can be confident that you have a good model. Here, you choose to use a = 0.05.
The p-value for each term in the model is less than 0.05, indicating a model that is a good candidate for further exploration and validation. This model is considerably simpler and fits the data almost as well as the model with all terms. The residual error only increased by a small amount.

You can further check the model by using the residual plots. The fitted values are the results predicted by your model. The residuals are the actual yields minus the predicted yields. The following graphs display:
Normal Probability Plot of the Residuals

Residuals Versus the Order of the Data
Histogram of the Residuals
Residuals Versus the Fitted Values
The residuals plots were satisfactory, and showed no cause for concern.

Step 12: Draw Conclusions — Display Factorial Plots

Now generate two graphs that will allow you to visualize the effects—a main effects plot and an interactions plot. When the plots are based on the means of the response data, you can generate them either before or after you actually fit a model to the data. When you are plotting the fitted values (least-squares means), you need to fit the model first.

1 Choose Stat > DOE > Factorial > Factorial Plots. Verify dialog box.
2 Check Main effects and click Setup.
3 In Responses, type Yield.
4 Next, select the terms you want to plot:

· Click on B:Pressure in the Available box, then click on the single arrow that points to the right. This will move the B:Pressure variable to the Selected box.
· Repeat these actions to move C:Catalyst to the Selected box. Verify dialog box. Click OK.

5 Check Interaction and click Setup.
6 Repeat steps 3 and 4. Verify Interactions dialog box.
9 Click OK in the main Factorial Plots dialog box to display each plot in a separate Graph window.

Step 13: Draw Conclusions — Evaluate the Plots

First, take a look at a plot that shows the basic effect of changing pressure, or using catalyst A versus catalyst B. These one-factor effects are called main effects. The numerical values for all the effects are shown in the Session window.

1 Choose Window > Main Effects for Yield to make the main effects plot the active window:

Two main effects plots display in this graph—one for pressure and one for catalyst. Verify main effects graph. The main effect of:

· pressure is the difference between the low setting and the high setting on the graph
· catalyst is the difference between the two categories

You can see that the type of catalyst has a bigger main effect than pressure. That is, the line connecting the mean responses for catalyst A and catalyst B has a steeper slope than the line connecting the mean responses at the low and high settings of pressure. Although the type of catalyst appears to affect the Yield more than pressure, it is very important to look at the interaction. An interaction can magnify or cancel out a main effect.

To calculate main effects, Minitab subtracts the mean response at the low or first level of the factor from the mean response at the high or second level of the factor. The table below summarizes the findings:

Factor Size of Effect Interpretation
Pressure +14.13 runs at 4 atmospheres of pressure had higher yields than runs at 1 atmosphere of pressure
Catalyst –30.37 runs that used catalyst A had higher yields than runs that used catalyst B
If you have no interactions between the factors, this graph will adequately describe where you can get the biggest payoff for changes in your factors.
The next step, then, is to look at the significant interaction. Although you have already verified a significant interaction with the Session window output, you can look at the interaction plot to see how big this effect is.

2 To make the interaction plot the active window, choose Window > Interaction Plot for Yield. Verify Interactions Plot.

An interaction plot shows the impact that changing the settings of one factor has on another factor. Because an interaction can magnify or diminish main effects, evaluating interactions is extremely important. The significant interaction between pressure and catalyst shows up as two lines with sharply differing slopes.

The yields for catalyst A are greater than yields for catalyst B at both 1 and 4 atmospheres of pressure. However, you can see that the difference in yields between runs using catalyst A and runs using catalyst B at 4 atmospheres is much greater than the difference in yields between runs using catalyst A and runs using catalyst B at 1 atmosphere.

In order to get the highest yield for your experiment, your results suggest that you should set pressure to 4 atmospheres and use catalyst A.

Step 14: Save and Exit

1 Choose File > Save Project.
2 In File name, enter SS5DOE for the name of your project. If you omit the extension .MPJ, Minitab will automatically add it once you save the project.
3 Click Save.
4 If you see a message box asking if you want to replace an existing file, click Yes.
5 To close Minitab, choose File > Exit.

Step 15: What You Learned

So, that is the end of your analysis! Let’s summarize what you did:

1 You decided on a design for the experiment, then generated and saved settings using the Create Factorial Design command.
2 You ran the experiment and entered the responses.
3 You fit the full model to look at some numerical values and generated two effects plots to see which terms seemed to be active.
4 You screened out unimportant effects, then fit a reduced model.
5 You generated main effects and interactions plots with the Factorial Plots command to visualize the effects.

6 You evaluated the reduced model with the p-values in the Analyze Factorial Design output and the various residuals plot.

You could have used additional analysis techniques in Minitab as well.

Let’s summarize what you have learned:

· From looking at the effects plots, you determined that pressure, type of catalyst, and the interaction between pressure and catalyst were active. Evaluating interactions is extremely important, because an interaction can magnify or cancel out main effects.
· You can eliminate (screen out) the other terms without significantly affecting predictions.
· Now that you have a model to predict the yield, you can apply this model to help obtain higher yields in future experiments.

In order to get the highest yield for your experiment, your results suggest that you should:

· set pressure to 4 atmospheres
· use Catalyst A
· evaluate higher levels of pressure with future experiments

All that for 16 runs. Quite a payoff!