This module will cover:

- Advanced graphing
  - Changing data types
  - Value Order
  - Making similar graphs
  - Zooming worksheet
  - Brushing
- Multi-graphs: By variables
- Interactively upgrading graphs / time axes
- Interactive Graph Layouts
  - Including: Subsetting while graphing
- Secondary Axes
- Line / Bar combo graph

Cont’d
MTB – 02 Intermediate Minitab®

➢ Cont’d

• Data Manipulation
  • Calculations using IF statements
  • Splitting vs Subsetting worksheets
  • Creating tables of summary statistics
  • “Lookup” tables
  • Forward lags

• Working Environment
  • Graphs into PowerPoint®
  • Which worksheet generated my graph?
  • Options > Graph > Annotation
MTB – 02 Advanced Graphing

• We will work with this sample dataset which has a strength value for a product made in a 12 hour shift process:

   Open in Minitab: *Strength.CSV*

   D = Days; N = Nights
MTB – 02  Change Data Type

- We notice that the first column is text when we need it to be date/time so it can be used for graphing, etc.
- We fix this with Data > Change Data Type > Text to Date/Time…

We change the data “in place”. A safer option would be to create a new column, but we haven’t invested much in this worksheet so far.

The data is in International date format – by far the best!
Let’s start with a simple graph of Strength vs Time
This is good start

We notice a production gap. This was an overhaul.

The gap in production would not have been obvious is a simple Time Series graph.
MTB – 02  Scatterplot vs Time

• We make edits to the graph:
  • Add a Lowess Smoother with a Degree of Smoothing of 0.1, a line colour of black and a width of 7
  • We get rid of the X axis label
  • We add minor ticks to the X-axis
  • We add gridlines
  • We add a grouping variable of crew to the symbols
  • Change connect line colour
  • Edit Legend
  • Add and edit a reference line at the lower spec. limit.

Yogi Berra: “You can observe a lot just by watchin.”

Is there a seasonal effect? Are C crew strength values low? Did the process start with low strength after the overhaul?
• Let’s systematically look at this dataset. We start with *Graph > Boxplot > With Groups* and plot Strength by Crew:

C crew strength values look low, but is it statistically significant?
While our boxplot has focus, we click on Editor > Make Similar Graph and create new graphs of Strength vs Shift and Speed.

It would make sense to list the speeds as Slow, Medium, Fast but Minitab will default to alphabetical. We fix on next slide.
We choose the User Specified Order, and edit in the window provided.

As soon as we make the change, the graph shows the Yellow status meaning the data has changed. We right click on it and choose Update Graph.

Our updated graph!
We click on one of the graphs to give it focus, and then click on **Editor > Layout Tool**.

- We change to 1 Row, 3 Columns.
- We select our BoxPlots to be included by clicking & using the arrow or by double clicking.
- You can move graphs from one panel to another by clicking and dragging.
- When you are done you click **Finish**.
Our completed graph layout.

All three factors look like they might be statistically significant.
Let’s do a formal test.
One-way ANOVA: Strength versus Crew

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crew</td>
<td>3</td>
<td>15537.8</td>
<td>5179.3</td>
<td>61.72</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>1748</td>
<td>146682.1</td>
<td>83.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1751</td>
<td>162219.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S = 9.160  R-Sq = 9.58%  R-Sq(adj) = 9.42%

Individual 95% CIs For Mean Based on Pooled StDev

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>469</td>
<td>61.833</td>
<td>9.125</td>
</tr>
<tr>
<td>B</td>
<td>407</td>
<td>63.545</td>
<td>9.507</td>
</tr>
<tr>
<td>C</td>
<td>407</td>
<td>55.318</td>
<td>8.838</td>
</tr>
<tr>
<td>D</td>
<td>469</td>
<td>59.937</td>
<td>9.164</td>
</tr>
</tbody>
</table>

Pooled StDev = 9.160

The factor is significant (P value of 0.000) and explains about 9% of the variability in strength.
A similar analysis for Speed and Shift shows they are also statistically significant, explaining about 22% and 6% respectively of the variation in strength. What about our possible seasonal trend? One tool we could use is:

We use 2 since there are 2 shifts in a day.
The data shows all sorts of non random behaviour.

(OK this is a “manufactured” data set!) :
Some manufacturing processes exhibit differences based on day of the week. Could this be happening here?

We then use an alternative tool, Stat > ANOVA > Interval Plot. ANOVA shows that this factor explains about 10% of the variability in strength.

If only typical data had such obvious sources of variation!
Here is how we can assign a formula to a column:

1. We right click on the new column.
2. You can also assign formulae through the Calc > Calculator checkbox.
3. The symbol shows that this column has a formula assigned. Any change to the input columns will update this column.
4. We type in the column name manually.

MTB – 02 Assigning Formulae
ANOVA shows that weekend is statistically significant. However, this might be because Thursday is unusual. We calculate a new value of strength which is missing for Thursdays:

A repeat ANOVA analysis using $StrengthNotThurs$ shows no statistical significance. It looks like the $Weekend$ variable was confounded with the $DayOfWk$ variable.
We click on the Brush Icon (or click on the graph and then Editor > Brush)

We click single points or click and drag to highlight multiple points of interest. If we hold down <Shift> we can highlight more than one range or one point.

The highlighted points have their row shown in the Brushing window.

In the worksheet, highlighted rows are shown via a dot beside the row number.
We can also change what appears in the Brushing window. With the points still highlighted, we click on Editor > Set ID Variables. Here we have selected all of the variables to appear in the Brushing window.
If we choose *Data > Subset* we can copy these brushed rows into a new worksheet. If we also click on *Exclude* we can copy only the non brushed rows into a new spreadsheet. Given the highlighted points, this means we could end up with a worksheet of out of spec. or in spec. rows respectively.

**Note:** Another way to subset from the current worksheet is to use the "*Rows that match* " *Condition* option to use a calculator style criteria to subset rows of interest e.g. ‘Speed’ = “Slow”
MTB – 02 Splitting

• You can also split a worksheet on the basis of unique values of a categorical variable.
  – Be careful not to use a continuous variable or you may get hundreds of worksheets!

You can see the new sheets in the Project Manager window.
Holding down the `<Ctrl>` key and using the scroll wheel let you zoom the font size of the worksheet.
MTB – 02 Descriptions

- You can add column descriptions via right clicking on the column:

After you’ve entered a column description, you see the red triangle. If you hover over it, you will see the description.

You can add worksheet descriptions as well!
• We start with Graph > Scatterplot

Use of a By variable to split a graph into Panels
• Same graph after further edits:
  • We get rid of the X axis label
  • We add minor ticks to the X-axis
  • We add gridlines
  • We change panels to not alternate, to have a margin of 0.009, Don’t Alternate Panels, Both Variable Names and Levels changing font colour & to bold
  • Change symbol type & colour
  • Add and edit a reference line at the lower spec. limit.
MTB – 02 Multi-graphs

• We start with Graph > Scatterplot

Use of a By variable to split a graph into Pages
• We will work with this sample dataset which has the output and max. output capability of the wind power generation farms in Ontario:
  
  Open in Minitab: WindByStation.CSV

• An obvious graph would be to look at the output of the various stations vs date:
MTB – 02 Interactive Upgrading of Graphs

- The resulting graph is a little too busy to be very useful.
We can put each generating station in a separate panel of the graph:
And we can subset the data to be the latest data
MTB – 02 Interactive Upgrading of Graphs

• This graph is much more readable

Scatterplot of Output vs Date

Panel variable: Generator

- Generator
  - AMARANTH
  - COMBER
  - DILLON
  - GOSFIELDWGS
  - GREENWICH
  - GREENWICH-T1
  - GREENWICH-T2
  - KINGSBRIDGE
  - PAROCHES
  - PORTALMA-T1
  - PORTALMA-T3
  - PRINCEFARM
  - RIPLEY SOUTH
  - SPENCE
  - UNDERWOOD
  - WOLFE ISLAND
MTB – 02 Interactive Upgrading of Graphs

• After some edits:

We edit the graph:
• X axis label removed
• X axis ticks to 90 degrees, Font size 7, 5 ticks
• Title
• Legend deleted
• Panel footnote deleted
• Size of Symbols (1 to 0.75)
• Add gridlines
• Panel Options:
  • Don’t alternate
  • Margin 0.009
Why the picket fence appearance?
- This is because all 24 values for each day are plotted at the start of that day
- We need a date/time variable that includes the hours & minutes portion

The dataset includes the time portion of the date/time
We create a date + time variable

Minitab® uses the same date/time convention as Excel®. Dates start at 1900 and go up by 1 for each day. Time values go from 0 to 1 representing fractions of a day.

We select (or type in) yyyy-mm-dd hh:mm

The we format the column by right clicking on the grey column header:
Now instead of starting over to create a graph of output vs DateTime, we use the make similar graph dialogue:
Can you see what additional edits have been made to the graph??
• What if we wanted to show the output of the major farms (only) overlaid on a graph.

• How would we define a “major farm”?  
  – By average power output  
    • We would like to add a column with average output by station.
Here is the dialogue:

This creates a new column “Mean1” which has, for every row, the average output for the station appearing in that row.

This option puts the average value on each line instead of creating a table of average values.
Now we can create a graph showing just recent data for the major farms:
After some additional edits
After some additional edits

This may not be a very useful view here, but sometimes this capability is “just what the Doctor ordered”!
Open in Minitab: *NaturalGas.CSV*

Data to be plotted

<table>
<thead>
<tr>
<th>Date</th>
<th>MetersCubed</th>
<th>DegreeDaysC</th>
<th>Cost</th>
<th>Year</th>
<th>DayIntoYear</th>
<th>SumConsumption</th>
<th>SumCost</th>
<th>MMDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/01/01</td>
<td>207.800</td>
<td>564.0</td>
<td>0.00</td>
<td>2001</td>
<td>1</td>
<td>207.80</td>
<td>0.00</td>
<td>01-01</td>
</tr>
<tr>
<td>2/01/01</td>
<td>378.600</td>
<td>926.0</td>
<td>0.00</td>
<td>2001</td>
<td>32</td>
<td>378.60</td>
<td>0.00</td>
<td>02-01</td>
</tr>
<tr>
<td>3/01/01</td>
<td>284.600</td>
<td>676.0</td>
<td>0.00</td>
<td>2001</td>
<td>60</td>
<td>663.20</td>
<td>0.00</td>
<td>03-01</td>
</tr>
<tr>
<td>4/01/01</td>
<td>287.500</td>
<td>653.0</td>
<td>0.00</td>
<td>2001</td>
<td>91</td>
<td>860.70</td>
<td>0.00</td>
<td>04-01</td>
</tr>
</tbody>
</table>

Day into year. MMDD is that date +1 converted into MM-DD format

This is the cumulative consumption so far in this year.

This is the cumulative cost so far in this year. (Costs not avail. for first years)
MTB – 02 Graph Layouts

• We generate a scatter plot with connect and groups:
• The graph is pretty close to what we want right out of the box:
We edit the graph:
- Y axis label
- X axis label
- Title
- Size of Symbols (1 to 0.75)
- Connect lines (to solid)
- Add gridlines
Following a similar approach, we create a graph for Cumulative Cost by Year.

You notice that there are no costs until March of 2008. We need to change the graph so that we only plot data from 2009 and on.
Here is how we do that:
• After some simple modification, here is what our graph looks like:

Much Better!

Cummulative Cost by Year

SumCost

Year
- 2009
- 2010
- 2011
- 2012
MTB – 02 Graph Layouts

- While one of the graphs has focus, we then use Editor > Layout Tool
• We change to 1 Row, move the cost to the right panel and bring the consumption into the left panel.
After clicking *Finish*, we get this.

A thing of beauty!
Open in Minitab: *NaturalGas.CSV*

- First we plot consumption vs date and degree days vs date, overlaid on the same graph:
MTB-02 Secondary Axes

- The resulting graph is a good starting point:

We double click on the Y axis, select the “Secondary” tab and change the second line to plot on a secondary Y axis.
We take this graph and make customizations:

- Get rid of the label “Date” on the X axis
- Change the font colour of the Y1 axis to a blue
- Change the colour of the Y1 axis ticks to blue
- Do the same for the Y2 axis to make it red
- Get rid of the legend
- Add gridlines
- Edit the title
We could stop here or continue to make this a bar / line plot.

We right click on the graph > Add > Data Display > Project Lines.
We click twice on one of the red project lines (where it does not overlap the black project line) and then choose type of none to get rid of this project line.
MTB-02 Line / Bar Combo

Natural Gas Consumption, Degree Days

We make further edits:
• Get rid of Y1 symbols, line
• Change range of time axis
• Change colour and width of Y1 project lines
• Change size of Y2 symbols

Note: consumption correlates very well with degree days except during the summer (when the main consumption is water heating).
MTB-02 Forward Lags

- Consider this data from a batch process:

<table>
<thead>
<tr>
<th></th>
<th>C1-D</th>
<th>C2</th>
<th>C3-T</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DateTime</td>
<td>PropertyX</td>
<td>Product</td>
<td>CleanOut</td>
</tr>
<tr>
<td>1</td>
<td>2012-01-07 13:10</td>
<td>41.7</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2012-01-07 17:40</td>
<td>35.5</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2012-01-07 22:12</td>
<td>37.8</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2012-01-08 02:08</td>
<td>37.7</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2012-01-08 06:41</td>
<td>36.1</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2012-01-08 11:13</td>
<td>37.3</td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2012-01-08 15:43</td>
<td>39.4</td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2012-01-08 20:10</td>
<td>40.4</td>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2012-01-08 23:58</td>
<td>37.2</td>
<td>A</td>
<td>0</td>
</tr>
</tbody>
</table>

Open in Minitab: BatchData.CSV

- Standard practice says that a “Clean Out” is needed on the last batch prior to a product change.
- You want to highlight these batches and flag batches that missed the clean out. (Note: there is a column which records if a clean out was performed.
- We need to look at the next row to see if there is a product change.
MTB-02  Forward Lags

• One approach to this would be to use a DO loop and look at the next row however this approach is complex and slow. (Looping code in Minitab® macros is slow to execute).

• The simplest approach is to sort the data into reverse order and then use the standard lag function. In this case the batches have a time stamp that can be used for the sort. (If not you could create a row number variable first).

• Once the data is sorted in reverse you can use a calculator function to see if a product change has happened and if the clean out was done and create appropriate Status text.
MTB-02  Forward Lags

• Here is the code which goes into the calculator with added comments: *(my text editor adds the handy colouring)*

    ```
    Name C5 'Status'
    Let 'Status' = if('Product'=lag('Product'), &
    "", &
    if(CleanOut=0, &
    "0 Missed C/O", &
    "0")
    # Same product
    # - status is blank
    # Different product
    # - no clean out - set status to "0 Missed...
    # - with clean out - set to "0" which
    # will be used to circle plot symbol
    ```

• Here is the worksheet after the sort & calculation:

  The latest batch shows a missed clean out – but until there is data for the next batch that may or may not be true.

  Next batch in time for the batch on row 5 is the batch on row 4. Thus the row 4 batch missed a clean out.
MTB-02  Forward Lags

- Here is a graph that can be generated with the data now available:

The latest batch shows a missed clean out – but until there is data for the next batch that may or may not be true.

Once the graph has been generated, the data labels are edited to increase the Font Size, and so the Text Angle = 90 and then the horizontal & vertical offset is adjusted so the plot symbols are circled.

The Scatterplot with groups > Labels sub-dialogue is used to select our Status text column to be used for data labels.
Another approach is possible, but you likely would only use it for large, complex datasets. In this approach for our example:

- You create a row number and a row number plus 1 variables
- You create a table of product types vs row number
- You then use the Data > Code > Use Conversion Table (Minitab®’s version of the lookup function), and add the product corresponding to row number plus 1 to the current row in a new column
- Now you can this new column for the value in the next row.
- The advantage is that you do not need to sort the dataset
• Using the same example we use Minitab®’s equivalent of a Lookup Table to add information to each row.

• We create a table of statistics to get a list of all the Products. Note: this example is so simple you would not need to do this but for more complex examples you would not want to create the list by hand. You might in fact read in information from a database or copy in a table of information from Excel®, etc.
• In this example we add in the batch weight for each product type by hand in column 8:

• Then we use Data > Code > Use Conversion Table… to add the batch weight to each row.

Existing column whose values are to be define information needed from conversion table

New column to be created with the looked up values.
Now the worksheet contains a batch weight for each row.
If you right click on a graph one option is to “Send Graph to Microsoft PowerPoint”. This will create a new slide in PowerPoint® containing the Minitab graph. PowerPoint will be opened if it is not already.

My usual approach is to “Copy Graph”. Then in PowerPoint® I paste it exactly where I want it in the document.
MTB 02 Which Worksheet?

• You have multiple graphs open and you do not know which of the worksheets was used to generate the graph. How can you tell?
MTB 02  Which Worksheet?

- Method 1: Right click on graph

This option on the pop up menu shows you the corresponding worksheet.
Method 2: Expand the Graph Selection window:

You can not only see the highlighted graph’s worksheet but you can double click on any of the graphs to switch to it.
Minitab has many customizable options. One for example allows you to add standard text to each graph you create:

As new graphs are created they get the custom added footnote.
• We learned how:
  - To Change data types (e.g. from text into numeric)
  - To override Minitab’s default alphabetic Value Ordering of a column
  - To make similar graphs to avoid having to “recustomize”
  - Zooming a worksheet allows you to move around quickly
  - To brushing point(s) of interest on a graph
  - To use By variables to split graphs across panels or pages
  - To upgrade graphs / time axes
  - To use the Graph Layout tool
  - To Subset while graphing
  - To create a Secondary Axis for a scatterplot
  - To create a Line / Bar combo graph using “Project Lines”
• (Cont’d) We learned how:
  – To do calculations using IF statements
  – To Split or Subset a worksheet
  – To create tables of summary statistics
  – To use “Lookup” tables
  – To create a “Forward lag”
  – To copy graphs into PowerPoint®
  – To find out which worksheet generated my graph?
  – To add a standard footnote to graphs