Basic Statistics

Descriptive Statistics

Statistics

- Statistics is a tool that can be overused, misused and misinterpreted.
  - Requires knowledge about the application area for its most accurate use.
- ***When using statistical analysis on your data:***
  - Be objective.
  - Have a healthy skepticism – treat all data as suspect.
  - Choose a representative sample.
  - Check that your displays of data accurately and effectively represent the facts.
  - Make sure you have an adequate sample size to support your conclusions.

Descriptive Statistics

- Descriptive statistics are numerical measures used to summarize or describe data.
- Common measures of central tendency:
  - ***Mean***
  - ***Median***
- Common measures of dispersion (or variability):
  - ***Range***
  - ***Variance***
  - ***Standard Deviation***
  - ***Coefficient of Variation***

Descriptive Statistics

- Close Minitab, save if you want, and re-open Minitab.
- Open the Excel spreadsheet called ***Exercise 7a.xls***
- Lets calculate some descriptive statistics.

Display Descriptive Statistics

- Select **Stat** → **Basic Statistics** → **Display Descriptive Statistics…**
- Select ***Moisture*** as the **Variable**.
- Click on the **Statistics…** button:
  - Default stats: **Mean**, **SE of Mean**, **Standard Deviation**, **Minimum**, **Maximum, First quartile**,and **Median, Third quartile, N nonmissing**, **N missing**,
- Click **OK** twice.

Display Descriptive Statistics

Descriptive Statistics: Moisture

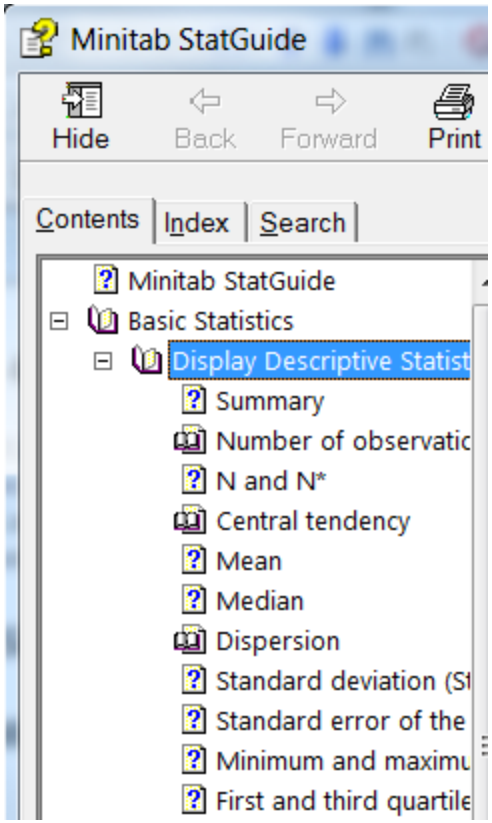
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Moisture	150	0	7.6053	0.0155	0.1903	7.2400	7.3700	7.7150	7.7500	7.8100

Youre output should look like the above.

What do some of these statistics mean?

Minitab StatGuide and Help

- The Minitab *StatGuide* and **Help** are ***excellent***.
- If you've forgotten what any statistic means, click **Help** → **StatGuide**.
- Lets take a look at the ***Mean***.



Minitab StatGuide and Help

- Minitab StatGuide
  - Basic Statistics
    - Display Descriptive Statist
      - Summary
      - Number of observatic
      - N and N*
      - Central tendency
      - Mean
      - Median
      - Dispersion
      - Standard deviation (St
      - Standard error of the
      - Minimum and maximu
      - First and third quartile
      - Graphs
      - Histogram of data

Central Tendency – Mean Topics

The mean (also called the average) is a measure of where the center of your distribution lies. It is simply the sum of all observations divided by the number of observations.

The mean is strongly influenced by extreme values.

Example Output Data

Variable	N	N*	Mean	SE Mean	StDev	Minimum
Precipitation	11	1	3.636	0.717	2.378	1.000

Variable	Q1	Median	Q3	Maximum
Precipitation	2.000	3.000	4.000	10.000

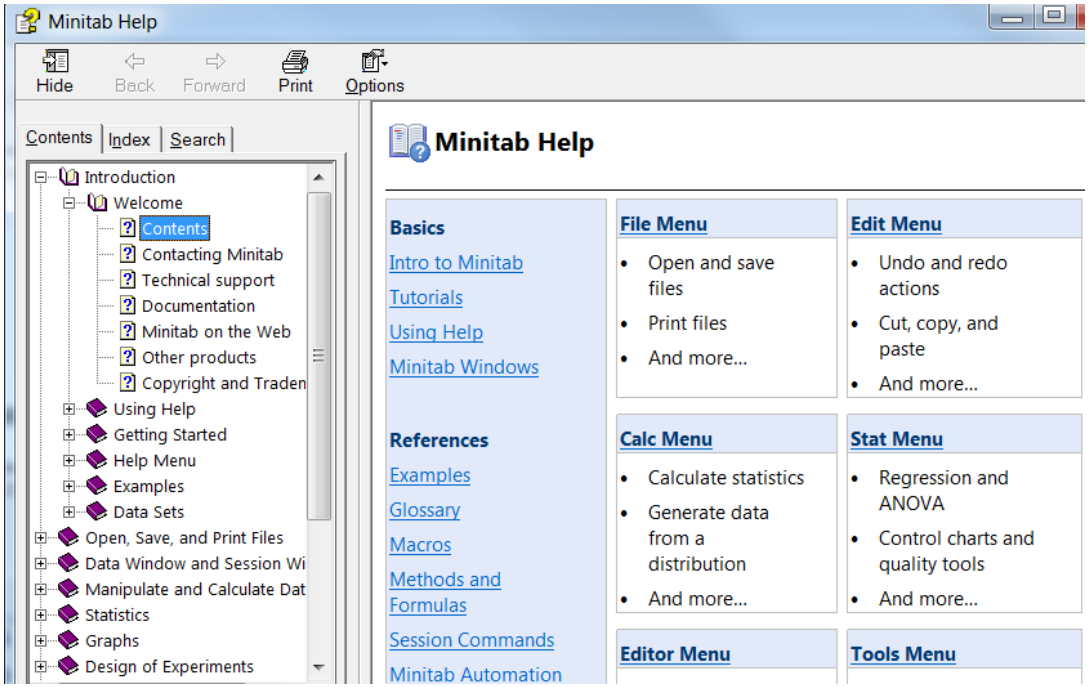
Interpretation More

For the precipitation data, the mean is:

(2 + 3 + 10 + 5 + 4 + 4 + 3 + 3 + 1 + 2 + 3) / 11 = **3.636**.

Minitab StatGuide and Help

- Close the StatGuide
- Select **Help → Help**



Minitab StatGuide and Help

- There's all the usual help items. I use the **Search** and the **Methods and Formulas** link the most.
- Select **Methods and Formulas → Basic Statistics → Descriptive Statistics → Mean**

Mean

A commonly used measure of the center of a batch of numbers, which is also called the average. It is the sum of all observations divided by the number of (nonmissing) observations.

$$\bar{x} = \frac{\Sigma x_i}{N}$$

where:

x_i is the ith observation

N is the number of nonmissing observations

Store Descriptive Statistics

- Sometimes it becomes necessary to do analysis on a **summary**.
  - For example, averages by some grouping variable.
- Minitab makes this very easy with the ***Store Descriptive Statistics*** module.
- The variable(s) being summarized and the grouping variable(s) are stored in the current worksheet.

Store Descriptive Statistics

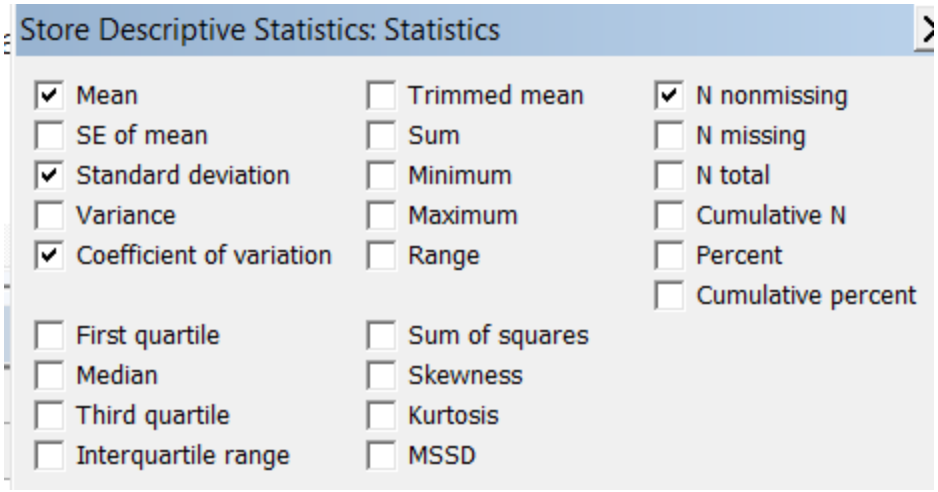
- In the this dataset, we want to summarize ***Moisture*** by the variable ***Roll Position***.
- Use **Store Descriptive Statistics** to summarize the **mean**, **standard deviation** and **coefficient of variation** of ***Moisture*** by each ***Roll Position***.

Store Descriptive Statistics

- Select **Stat** → **Basic Statistics** → **Store Descriptive Statistics…**
- Select ***Moisture*** as the **Variables**.
- Select ***Roll Position*** as the **By variable**.

Store Descriptive Statistics

Click the **Statistics…** button → select **Mean**, **Standard deviation** and **Coefficient of variation** → click **OK** once.

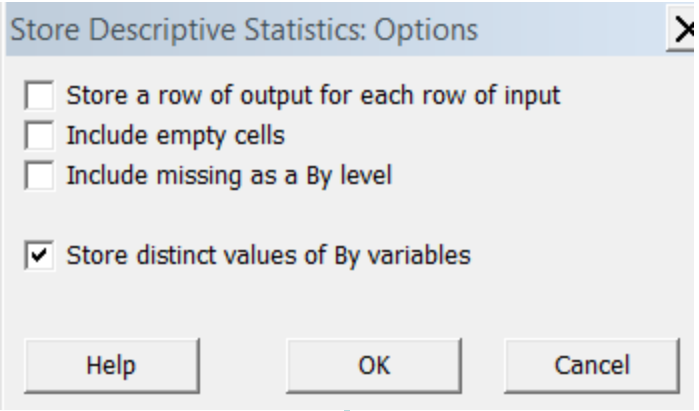


Store Descriptive Statistics

- Sometimes when you are summarizing data, the grouping variables are redundant with each other in some way.
  - This will cause Minitab to output rows with all missing values except for the group variable identifier.

Store Descriptive Statistics

- To avoid this, click the **Options…** button and ***uncheck*** the ***Include empty cells*** option.
- Click **OK** → click **OK**
- The following new variables appear in the worksheet.



ByVar1	Mean1	StDev1	CVariation1	N1
A	7.58062	0.946430	12.4849	50
B	7.98162	0.869969	10.8997	50
C	7.61903	0.806833	10.5897	50

Store Descriptive Statistics

- What I usually do is create a new worksheet within this Minitab project.
- I then cut and paste these new variables that contain the descriptive stats into the new worksheet.
- Then the variables can be re-named to something useful.

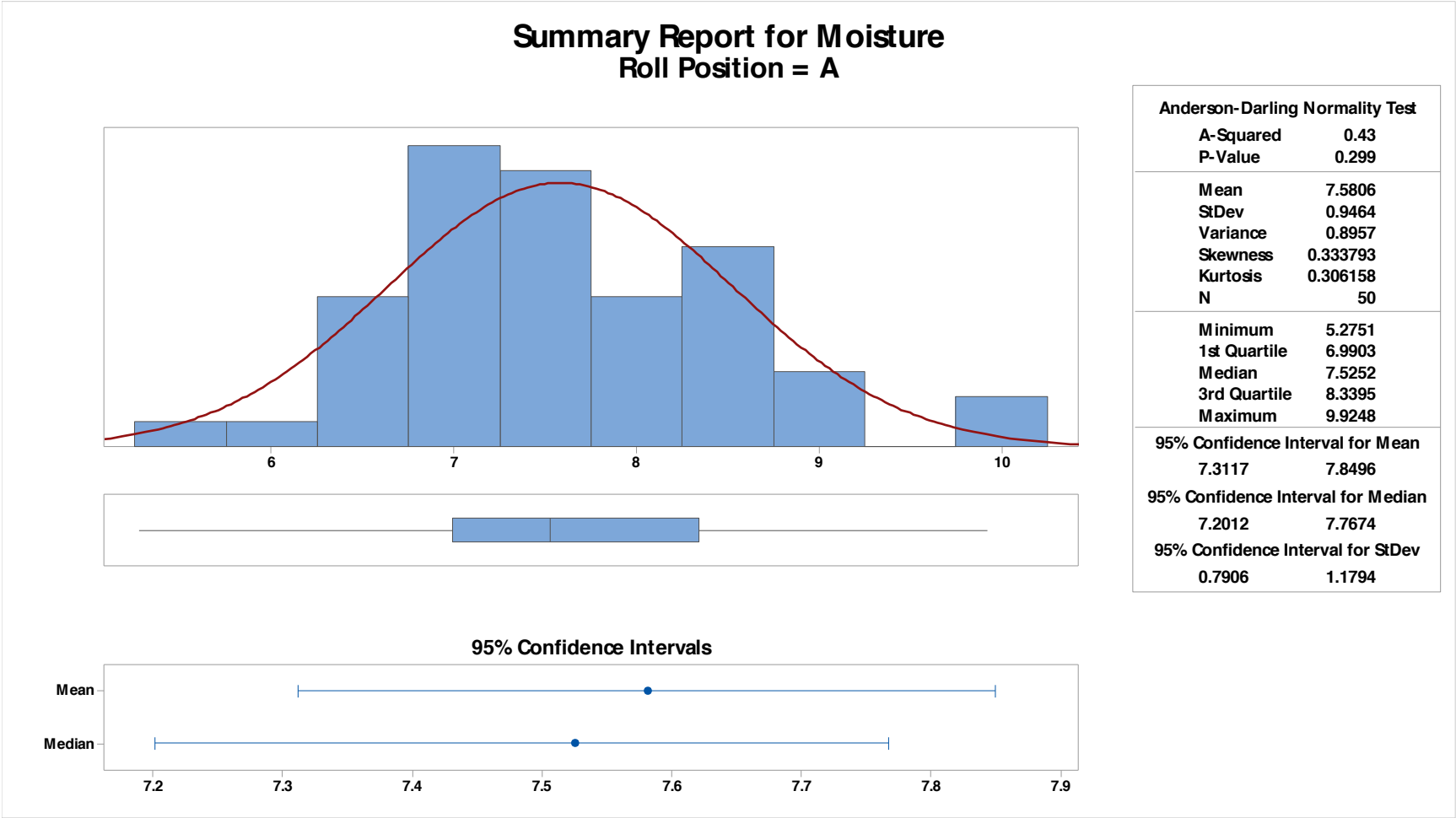
Graphical Summary

- Another way to display descriptive statistics in Minitab is the **Graphical Summary dialog**.
- This analysis shows the summary statistics, confidence intervals, histogram with a Normal curve, box plot and a test for Normality.

Graphical Summary

- To create the **Graphical Summary** select **Stat → Basic Statistics → Graphical Summary**.
- Select ***Moisture*** as the "Variables:" and ***Roll Position*** as the "By variables (optional):".
- Click **OK**.

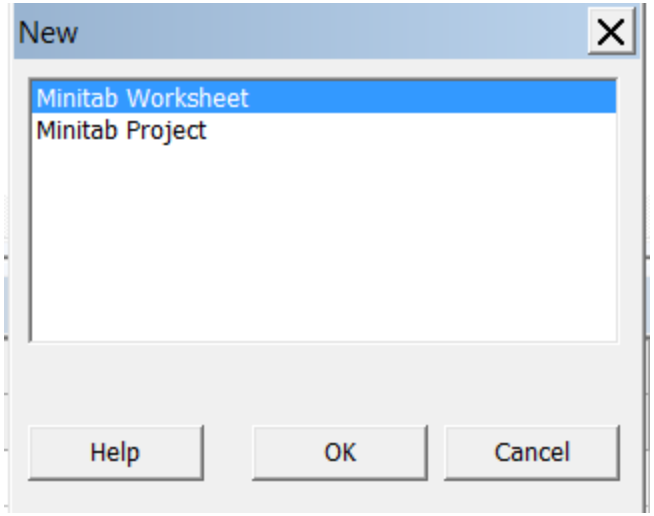
Graphical Summary



Anderson-Darling Normality Test	
A-Squared	0.43
P-Value	0.299
Mean	7.5806
StDev	0.9464
Variance	0.8957
Skewness	0.333793
Kurtosis	0.306158
N	50
Minimum	5.2751
1st Quartile	6.9903
Median	7.5252
3rd Quartile	8.3395
Maximum	9.9248
95% Confidence Interval for Mean	
7.3117	7.8496
95% Confidence Interval for Median	
7.2012	7.7674
95% Confidence Interval for StDev	
0.7906	1.1794

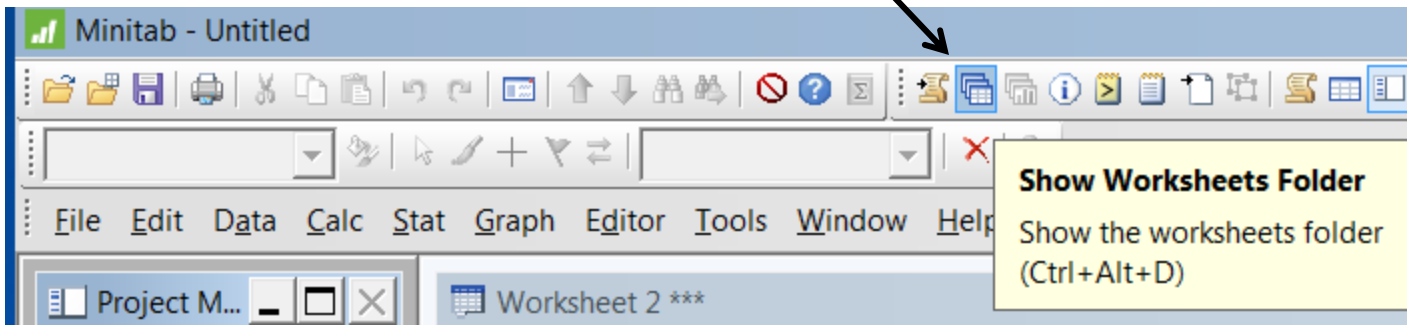
Create a New Worksheet

- Select **File** → **New** → select **Minitab Worksheet** → click **OK**
- A new worksheet should appear named **Worksheet 2** (or some other number)



Create a New Worksheet

- Now is the time to become familiar with some of the **shortcut** icons on the Minitab ribbon above menus.
- Click on the **Show Worksheets Folders** icon in the ribbon.



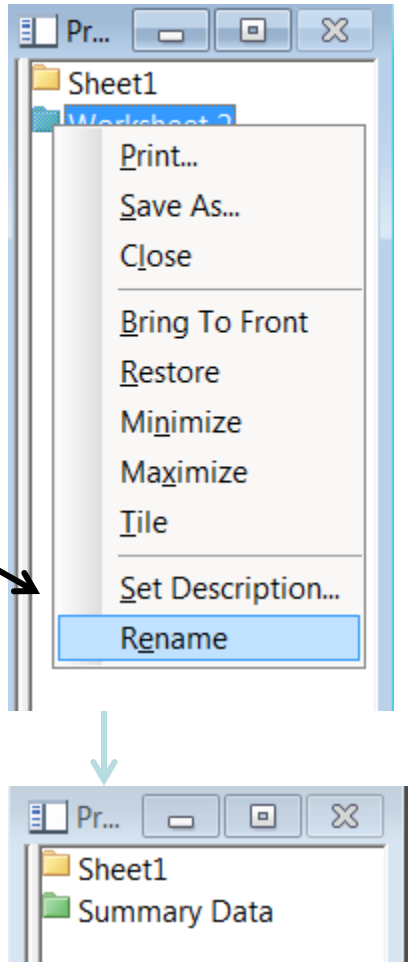
Create a New Worksheet

- You should now see **Project Manager** with the 2 entries: **Sheet1** and **Worksheet 2**
- It is a good idea to give the worksheets distinctive names once you have created more than one.

Create a New Worksheet

- Right click on ***Worksheet 2*** → select **Rename** → type in "***Summary Data***"

→press the **Enter** key.



Now try renaming **Sheet 1** to ***All Data***

Copying Data to a Worksheet

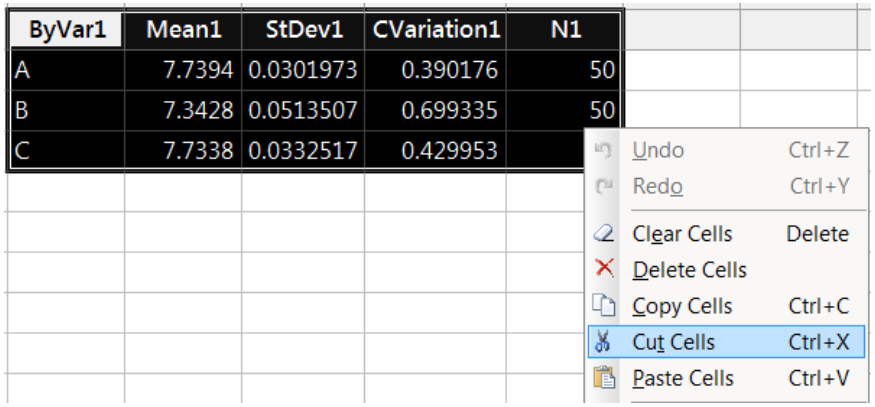
- To navigate to another **Worksheet** use the **Project Manager**, double-click on the **Worksheet** you want to make current.
  - Look at the bottom left corner of the Minitab screen and you will see the ***Current Worksheet*** listed.

Copying Data to a Worksheet

- Double-click on the ***All Data*** worksheet entry in the **Project Manager** to make it the current worksheet.
- Find the new columns ByVar1, Mean1, StDev1, CVariation1, N1

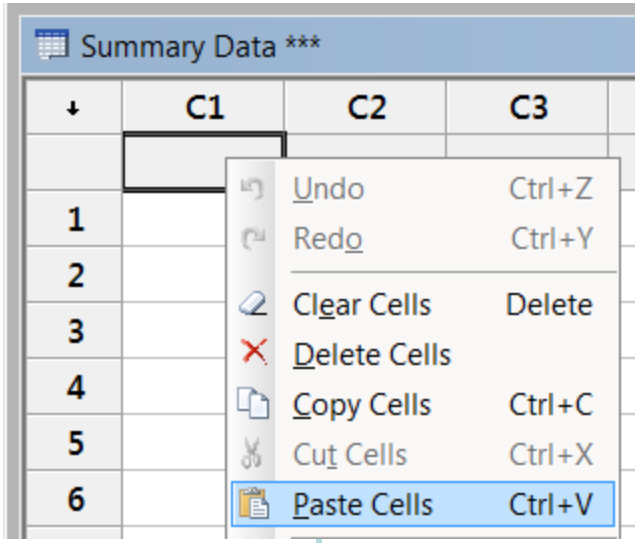
Copying Data to a Worksheet

- Right click in the column labeled **ByVar1** and hold down the button and drag the mouse so that the column labels to **N1** and data are highlighted.
- Right click on the highlighted cells and select **Cut Cells**.



Copying Data to a Worksheet

- Double-click on **Summary Data** entry in the **Project Manager** to make it the current worksheet.
- Right click in the blank box below the **C1** column label and select **Paste Cells**



Summary Data ***

	C1-T	C2	C3	C4	C5
	ByVar1	Mean1	StDev1	CVariation1	N1
1	A	7.7394	0.0301973	0.390176	50
2	B	7.3428	0.0513507	0.699335	50
3	C	7.7338	0.0332517	0.429953	50

Changing Column Names

- Select the newly created column **C1-T** by single-clicking in the box **C1-T** on the worksheet.
- Change **ByVar1** to **Roll Position** by typing **Roll Position** in the box → hit the **Enter** key.

Changing Column Names

- You should now rename each of the non-descript variable names with something meaningful, perhaps the following:
  - **Mean1** → **AvgMoisture**
  - **StDev1** → **StdDevMoisture**
  - **CVariation1** → **MoistureCoV**
- Save and close the project.

EXERCISE

- A trial was run to assess the effect on sheet formation of adding a certain chemical.
- The dataset ***Exercise 3.xls*** contains the results of this trial.
- Create some graphics and useful statistics to describe the outcome of the trial.