

Tutorial 1 - Getting Started

Introduction

Before starting this tutorial, install Change-Point Analyzer on your computer. This tutorial will walk you through the steps of using this software to analyze a simple data set. We will be using the trade deficit data that appears in Donald Wheeler's book: *Understanding Variation – The Key to Managing Chaos*. Table 1 shows the US trade deficit each month for 1987 and 1988 in billions of dollars.

Table 1: US Trade Deficits 1987-1988 (\$ billions)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1987	10.7	13.0	11.4	11.5	12.5	14.1	14.8	14.1	12.6	16.0	11.7	10.6
1988	10.0	11.4	7.9	9.5	8.0	11.8	10.5	11.2	9.2	10.1	10.4	10.5

Step 1: Start the Program

Start Change-Point Analyzer. Your screen should appear as in Figure 1. The blank Data window is used to enter the data.

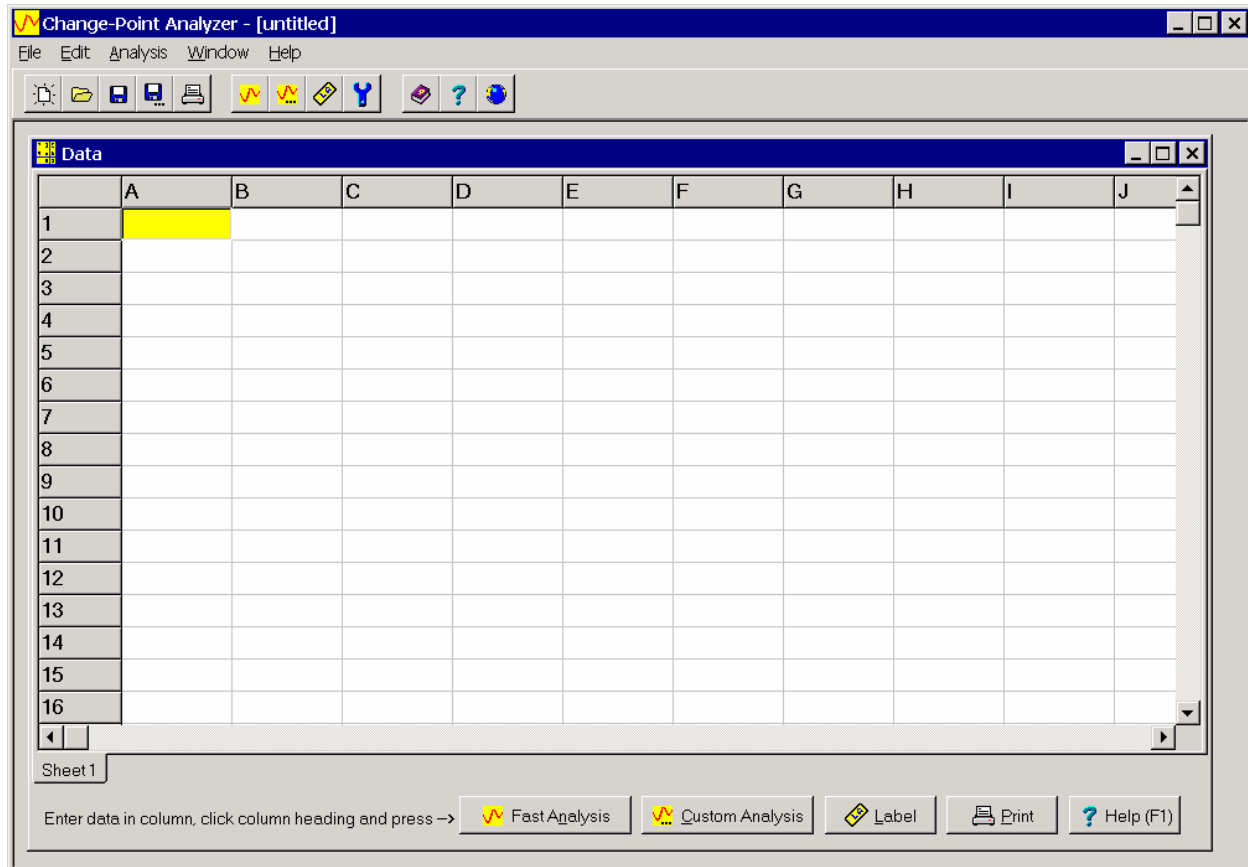
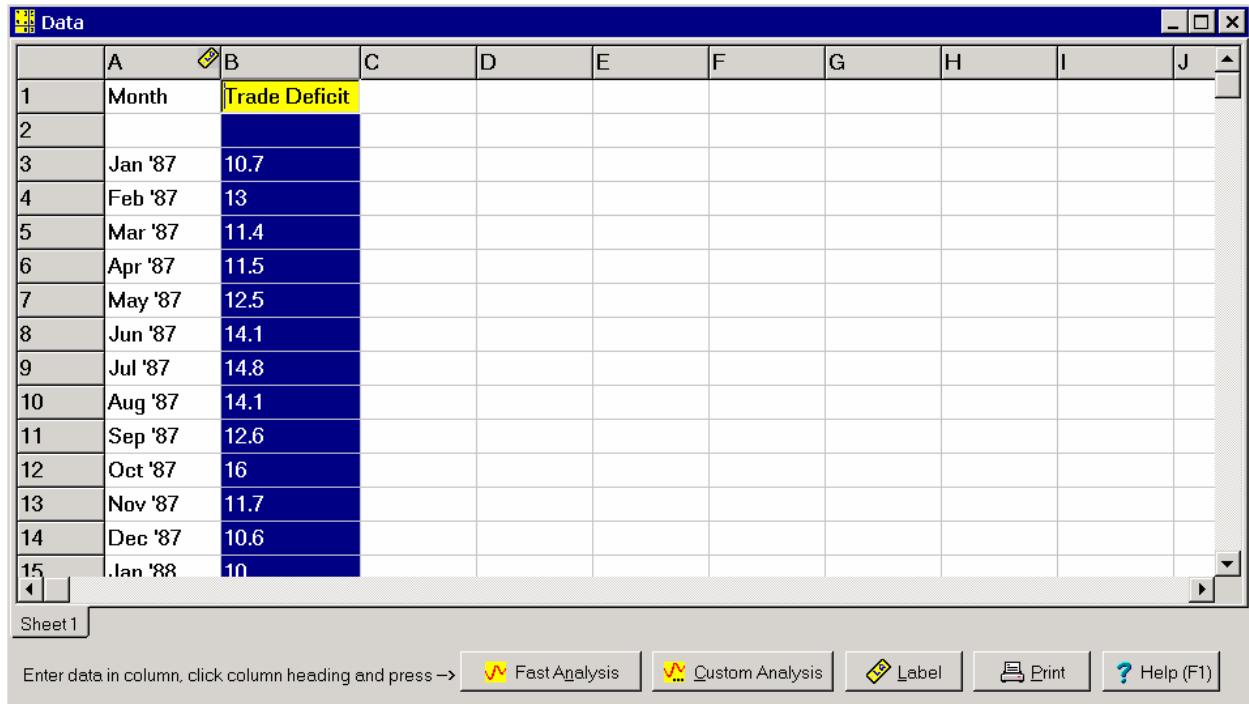


Figure 1: New Session

Step 2: Enter the Data and Labels

The data is generally entered in columns. You can also enter labels for the data points. We will enter both labels and data as shown in Figure 2. Start entering the labels by clicking on cell A1 (Column A and Row 1). Cell A1 will be highlighted in yellow. Type the word “Month” and press the Enter key twice. This causes cell A3 to be highlighted. Type the first label, “Jan ‘87”. Continue typing the rest of the labels, pressing the Enter key after each label to move to the next cell. Next enter the data. Start by typing “Trade Deficit” into cell B1. Use the scroll bars to move to this cell. Move to cell B3 and start entering the data.



	A	B	C	D	E	F	G	H	I	J
1	Month	Trade Deficit								
2										
3	Jan '87	10.7								
4	Feb '87	13								
5	Mar '87	11.4								
6	Apr '87	11.5								
7	May '87	12.5								
8	Jun '87	14.1								
9	Jul '87	14.8								
10	Aug '87	14.1								
11	Sep '87	12.6								
12	Oct '87	16								
13	Nov '87	11.7								
14	Dec '87	10.6								
15	Jan '88	10								

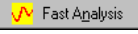
Figure 2: Data Window with Data Entered

Once the data has been entered, you can print a copy of the data for your records by clicking on the Print button. You can also paste data into the Data window using the clipboard. This allows you to easily analyze data you have already entered elsewhere. Change-Point Analyzer also comes with an Add-In for Excel that, once installed, allows you to initiate the analysis from within Excel. Tutorial 3 describes how to use this Add-In. The Trade Deficit data contains a single observation per time period. Change-Point Analyzer can also handle data sets containing multiple observations per time period. Tutorial 2 describes how handle this type of data.

Step 3: Select the Label Column

To select column A as labels, click the header for column A. This will highlight the entire column. Then click the Label button. The label icon will appear in the column header indicating the column has been selected. Labels are optional. If no labels are specified, the row numbers are used as labels.

Step 4: Perform the Analysis

Before analyzing the data, you must first select the data you want to analyze. To do this, click on the header for column B. This will highlight the entire column. Next click the Fast Analysis button  at the bottom of the Data window. Now sit back as the analysis is performed.

As the analysis proceeds, the Analysis Status dialog box shown in Figure 3 is displayed. The purpose of this dialog box is to assure you that the analysis is progressing. Don't worry about understanding the status messages. Anything of importance will be displayed in the Analysis Results window once the analysis is completed. Analyses take from less than a second to several minutes depending on the speed of your machine and the amount of data you are analyzing. For the trade deficit data, it should only take 1-2 seconds to complete the analysis.

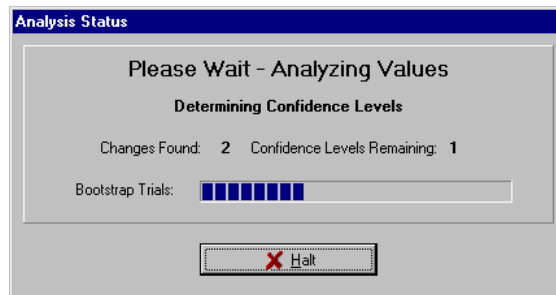


Figure 3: Analysis Status Dialog Box

Change-Point Analyzer automatically checks for outliers and a violation of assumptions. If either is detected, a message appears describing the problem and recommending how to proceed. A custom analysis is required. Tutorial 4 describes the general procedure for performing custom analyses while Tutorials 5 and 6 describe how to handle outliers and a violation of assumptions.

Step 5: Interpret the Results

Once the analysis has been completed, the results are shown in the Analysis Results window in Figure 4. The results are displayed on seven tabs. We will examine each of these seven tabs.

Step 5.1: Plot - Values Tab

The first tab shows a plot of the data (wavy black line). It also summarizes the results of the change point analysis in blue. The blue region shifts twice, once around June '87 and again around Nov '87. This indicates that the change-point analysis detected two changes in the trade deficit data. The second tab will provide further details.

Also shown are two red lines, which are control limits. They represent the maximum range that the values are expected to vary over assuming no change has occurred. Points outside the control limits indicate a change has occurred. These control limits assume the values come from the normal distribution and may not be appropriate for all sets of data. The fact that point 12 is above the upper control limit also indicates that some sort of change occurred. Control charting is an alternative approach to detecting changes. As we will see, a change-point analysis

has many advantages over control charting and is the preferred approach when analyzing historical data. However, Change-Point Analyzer also provides control charts of the data.

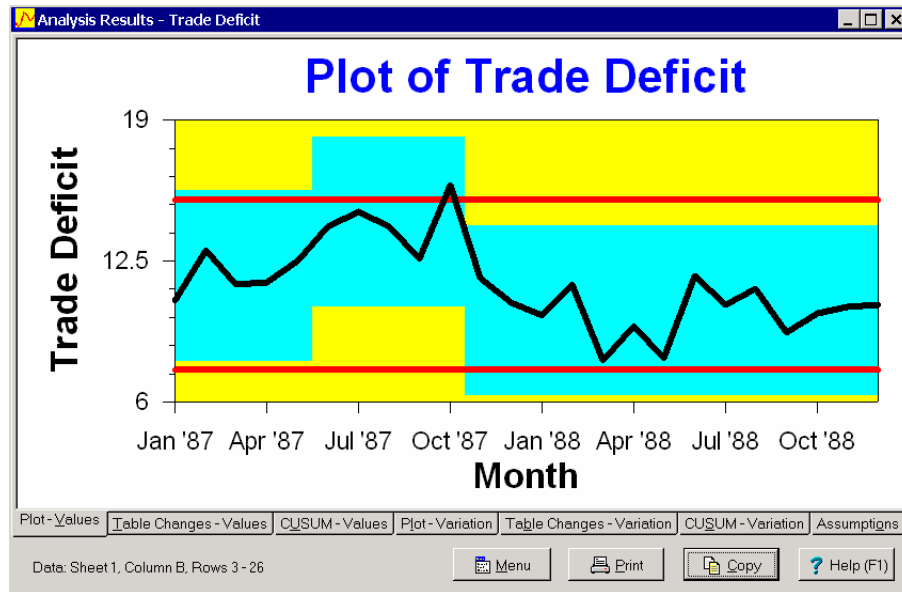



Figure 4: Analysis Results Window Showing Tab 1

You can hide the control limits, not display the blue region, represent the data with points rather than a line, change the title, scale and colors and much more. To do so, right click the mouse over the plot. You can also click the Menu button . This displays the popup menu in Figure 5 for making these changes. The other tabs also have popup menus associated with them for customizing their display.

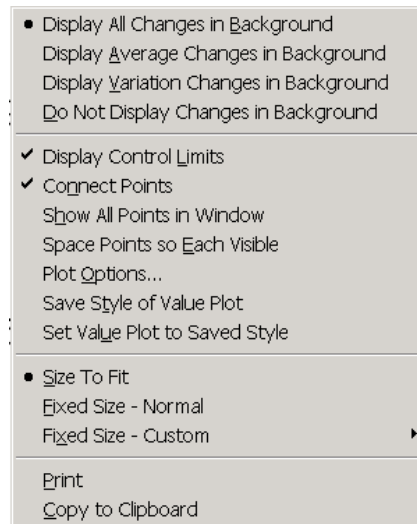




Figure 5: Popup Menu for Modifying Plot

Take a minute to make some changes to your plot. When you are done, you can print it by clicking the Print button  on the bottom of the Data Analysis window. You can also copy the plot to the clipboard by clicking the Copy button . This allows you to paste the plot into your word processor using the Paste menu item on its Edit menu.

Step 5.2: Table Changes - Values Tab

The second tab shows the results of the change-point analysis in table form (Figure 6). Each change detected is listed along with further information describing the change.

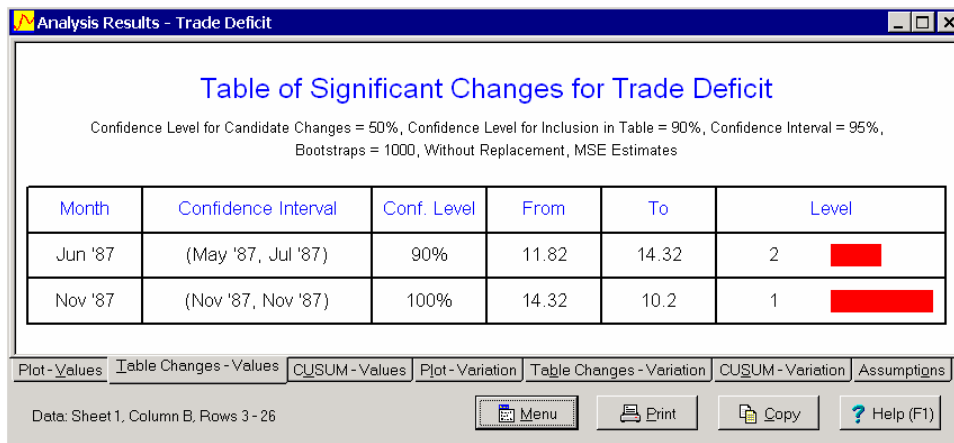


Figure 6: Analysis Results Window Showing Tab 2

The analysis detects two changes. The first change is estimated to have occurred around Jun '87. This point represents the first month following the change. The second change is estimated to have occurred around Nov '87. Associated with each change is a confidence level indicating how confident the analysis is that the change actually occurred. The first change occurred with 90% confidence. The second change occurred with 100% confidence. We are much more confident about the second change.

Also associated with each change is a confidence interval for the time of the change indicating how well the time of the change has been pinpointed. 95% confidence is used for all confidence intervals. With 95% confidence, the first change occurred between May '87 and Jul '87. With 95% confidence, the second change occurred at Nov '87. The fact that the confidence interval for the first change is wider indicates that the time of the first change cannot be as accurately pinpointed as the second change.

The second tab also gives additional information about each change. The table indicates that prior to the first change the average trade deficit was 11.82 billion dollars while after the first change it was 14.32 billion dollars. Tab 2 also gives a level associated with each change. The level is an indication of the importance of the change. The level 1 change is the first change detected and that which is most visibly apparent in the plot in Figure 4. Level 2 changes are detected on a second pass through the data. Any number of levels can exist dependent on the number of changes found.

Figure 4 represented these two changes by the shifts in the blue background. The blue background represents a region expected to contain all the values based on the current model that two changes occurred. Since all points fall within this region, this model fully explains the variation in the data.

While the control chart in Figure 4 barely detected any change had occurred (one point barely out), the change-point analysis detected two changes. It also provided additional details

including confidence levels and confidence intervals. This example illustrates two of the benefits of a change-point analysis: it is more powerful at detecting smaller sustained changes and it better characterizes such changes. When used to analyze historical data for trends and changes, a change-point analysis provides far more useful information than a control chart. For such data, the best approach is to perform a change-point analysis. However, this does not prevent one from also control charting the data. The results of both approaches can be displayed in a single plot as in Figure 4.

Step 5.3: CUSUM - Values Tab

The third tab shows a cumulative sum chart (CUSUM) of the data (Figure 7). Change-Point Analyzer uses the CUSUM charts to identify the changes reported on Tab 2. You can ignore these plots if you desire. All the relevant results are given in Tab 2.

Interpreting a CUSUM chart takes practice. Here are the basic rules:

- A period where the CUSUM increases represents a period of time where the data is above the overall average.
- A period where the CUSUM decreases represents a period of time where the data is below the overall average.
- A straight-line segment represents a period of time where no change occurred.
- A sudden change in direction indicates the values have shifted.

Applying these rules to the CUSUM in Figure 7, we see the trade deficit was above the overall average up to point 13 and thereafter was below the overall average. Around Nov '87 there is a sudden change in direction indicating a change. There is also an indication of a smaller change around June '87. The changes in Tab 2 are represented by changes in the background color. You can read more about CUSUM charts in the technical articles.

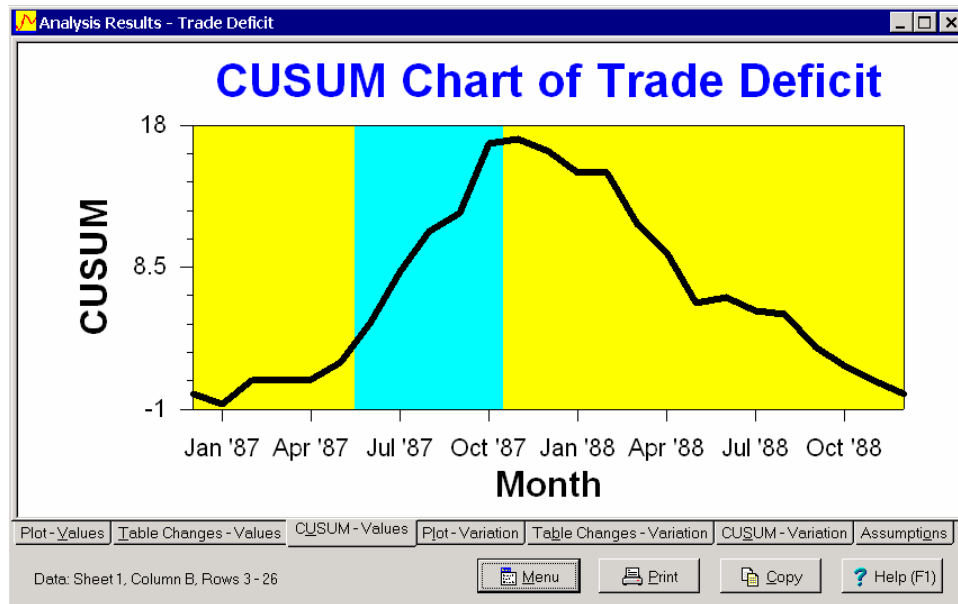


Figure 7: Analysis Results Window Showing Tab 3

Step 5.4: Plot - Variation Tab

Tabs 1-3 analyze the data for changes or shifts in the average. Tabs 4-6 do the same for the variation. For the trade deficit data, there is only a single value per month. This presents a special problem in that the standard deviation cannot be calculated for each month. Change-Point Analyzer groups consecutive months together to form pairs and then estimates the standard deviation of each pair. The first pair is Jan '87 and Feb '87. The second pair is Mar '87 and Apr '87. There are 24 data points resulting in 12 pairs. The standard deviation of each pair is calculated yielding 12 standard deviations. These are the values plotted on Tab 4 (Figure 8).

Changes are shown by changes in the background color. In this case there are no color changes shown in the background, which indicates that the change-point analysis did not detect a change. The plot also contains control limits in red. These control limits assume the values come from the normal distribution and may not be appropriate for all sets of data. As all points are inside the control limits, the control chart also indicates no change occurred.

Changes in the variation are also displayed on the plot on Tab 1 (Figure 4). The width of the blue region is six times the standard deviation. If a change in the variation is detected, the blue region will change height. For the trade deficit data, the blue region shifted up and down twice indicating two changes in the average. However, the blue region stayed the same height. This indicates no change in the variation was detected.

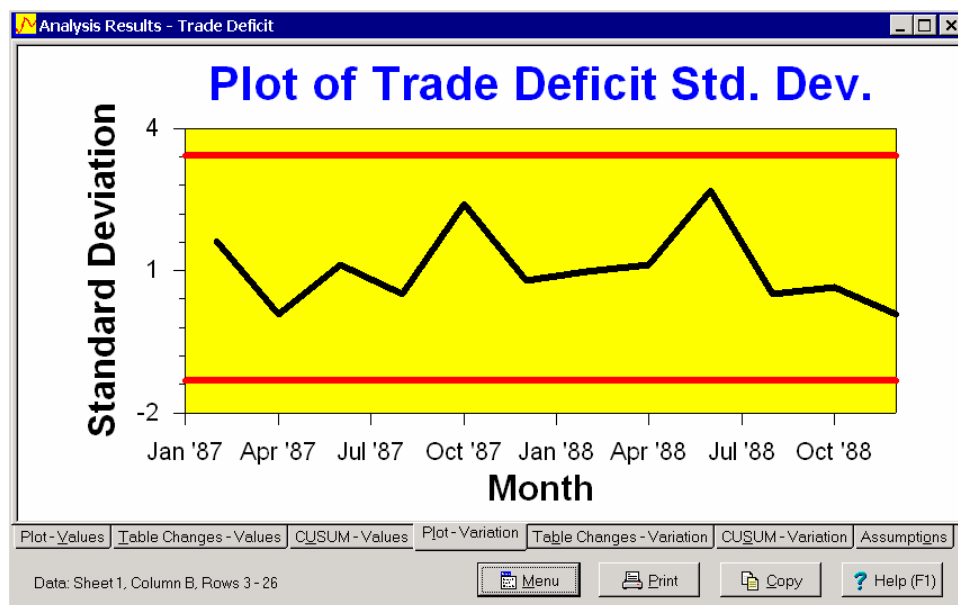


Figure 8: Analysis Results Window Showing Tab 4

Step 5.5: Table Changes - Variation Tab

The fifth tab shows the table of variation changes (Figure 9). It is interpreted similar to the table on Tab 2. For the trade deficit data no changes in the variation were found, so the table is empty.

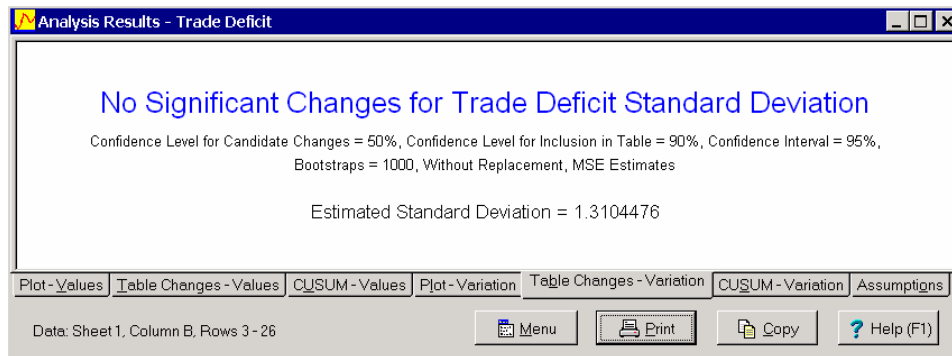


Figure 9: Analysis Results Window Showing Tab 5

Step 5.6: CUSUM - Variation Tab

The sixth tab shows a CUSUM chart of the variation (Figure 10). See Step 5.3 for information on interpreting a CUSUM chart. Changes are shown by changes in the background color. In this case there are no color changes shown in the background, which indicates that the change-point analysis did not detect a change.

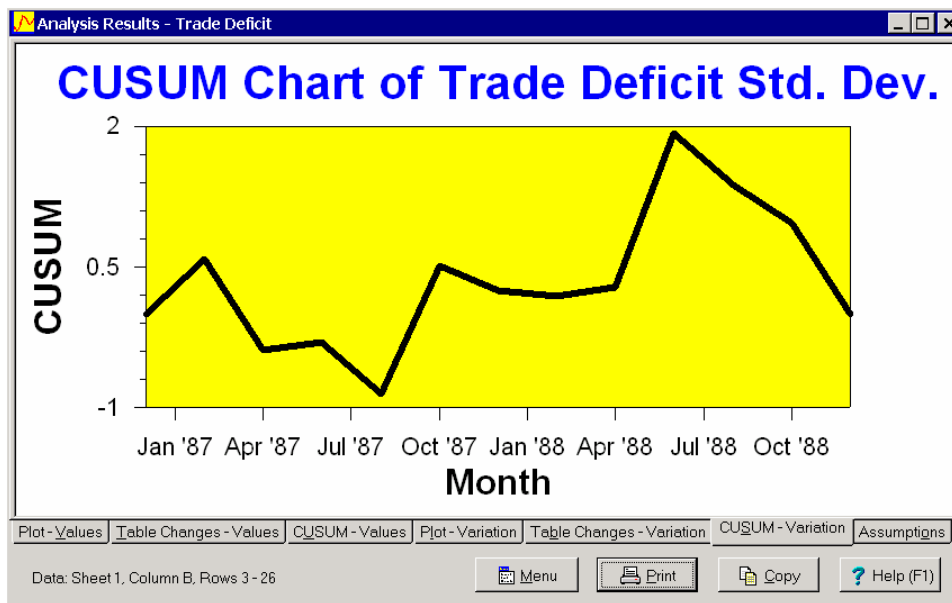


Figure 10: Analysis Results Window Showing Tab 6

Step 5.7: Assumptions Tab

The seventh tab provides instructions on how to handle outliers or a violation of assumptions if they are detected (Figure 11). In this case, neither was found. Tutorials 5 and 6 show how to deal with these two complications.

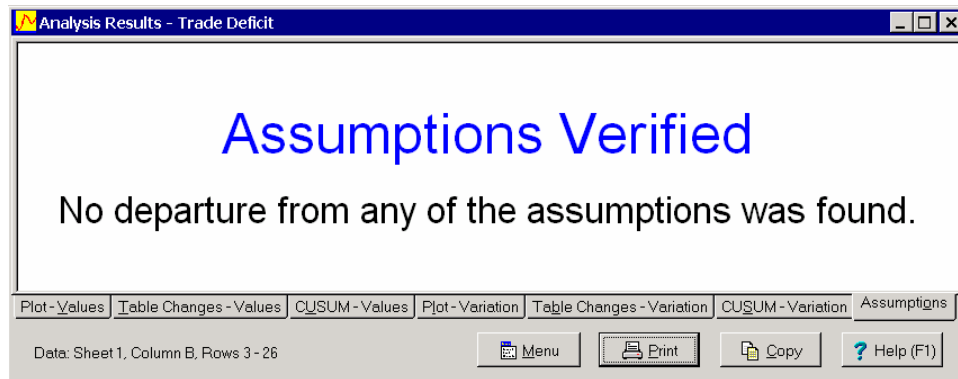


Figure 11: Analysis Results Window Showing Tab 7

Conclusion

This tutorial demonstrates two important advantages of a change-point analysis:

1. It frequently detects changes missed by control charts and is capable of determining that multiple changes have occurred.
2. It provides a more detailed description of each change including confidence levels and confidence intervals.

When analyzing historical data, a change-point analysis is superior to control charting. However, the greatest advantage of a change-point analysis is its ease of use. The same analysis can be used for all types of data including measurements, pass/fail data and counts. No longer do you have to deal with a confusing array of control charts such as individual charts, p-charts, u-charts, \bar{X} -R charts and more. Just enter the data, select the column and click on the Fast Analysis button. The program automatically checks for outliers and violation of assumptions and then displays the results in an easy to understand fashion.

Other Tutorials

In this tutorial, the data consisted of a single value per time period. Tutorial 2 shows how to handle multiple observations per time period. Tutorial 3 shows how to perform a change-point analysis directly from Excel using the Excel Add-In. Tutorial 4 shows how to select a subset of the data, change the title and more using the Custom Analysis dialog box.

The final 2 tutorials show you how to handle complications that can arise: Tutorial 5 deals with outliers and Tutorial 6 deals with a violation of assumptions. Change-Point Analyzer detects and notifies you of either of these situations. You don't need to read the last 2 tutorials until you encounter one of these complications.