


Tutorial 4 - Custom Analysis Dialog Box

Introduction

This tutorial shows you how to use the many features available through the Custom Analysis dialog box. These include:

- Selecting a subset of the data for analysis
- Specifying a title for your data
- Specifying labels for your data points
- Special features for handling violations of assumptions

To perform a custom analysis, click the Custom Analysis button  at the bottom of the Data window. The following popup menu will appear:

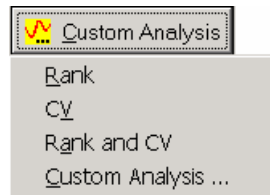


Figure 1: Custom Analysis Popup Menu

There are short cut menu items for two of the custom analysis options: to analyze the ranks of the data and to analyze the coefficient of variation (CV) rather than the standard deviation. CV is also commonly referred to as the relative standard deviation (RSD). Otherwise, selecting the last menu item displays the Custom Analysis dialog box where the complete set of custom analysis options can be specified including rank and CV. All of the available options are explained below.

Selecting the Data to be Analyzed

To illustrate the use of the Custom Analysis dialog box, the trade deficit data in the file Trade Deficit.cpa will be used. Open this file. The Data window should appear as in Figure 1. Column B has been selected. At this point you could analyze all the data in Column B by clicking the Fast Analysis button.

| | 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 | | | | | | | | |
| 16 | Feb '88 | 11.4 | | | | | | | | |
| 17 | Mar '88 | 7.9 | | | | | | | | |
| 18 | Apr '88 | 9.5 | | | | | | | | |
| 19 | May '88 | 8 | | | | | | | | |

Figure 1: Data Window for Trade Deficit Data

Suppose you only wanted to analyze the data from 1987 (rows 3-14). You can select a subset of the data for analysis using the Custom Analysis dialog box. Click on the Custom Analysis button at the bottom of the Data Window and then select the last menu item. The Custom Analysis dialog box shown in Figure 2 will appear.

Custom Analysis

Data To Analyze | Labels | Handling Violations of Assumptions

Data is In:
 Columns Rows

Location of Data:

Data appears in: Start Column: B

Number Values: 1

Time Period to Analyze: First Row: 1 Last Row: 65536

Name: Trade Deficit (appears in title and on y-axis)

OK Cancel Advanced Help

Figure 2: Tab 1 of Custom Analysis Dialog Box

The Custom Analysis dialog box contains three tabs. The first tab, titled Data to Analyze, is used to select the data. The controls on this tab are initialized based on the cell or cells selected in the Data window. The first control is the **Data Is In** radio box. It is used to specify whether the data resides in a column or row. The Columns button is checked indicating the data is in a column.

The next control is the **Start Column** box. It is used to specify the column containing the data. Column B has been selected. The next control is the **Number of Values** box. When there is only a single value per time period, this control should be set to 1. This is the case with the trade deficit data. However, if there are multiple values per time period, as was the case with the burst data in Tutorial 2, the number of values per time period should be entered. If a value of 3 were entered, Change-Point Analyzer would assume these values were in columns B-D.

The next 2 controls are the **Start Row** and **End Row** boxes. These are used to select the range of values to be analyzed. They are initially set to 1 and 65,536 respectively. This represents the entire column. Since blank values are ignored, this results in all 24 values being used in the analysis. To select just the data from 1987, set the **Start Row** box to 3 and the **End Row** box to 14. Once the data to be analyzed is selected, click the OK button to perform the analysis. The results are shown in Figure 3. Only the data for 1987 is used to perform the analysis.

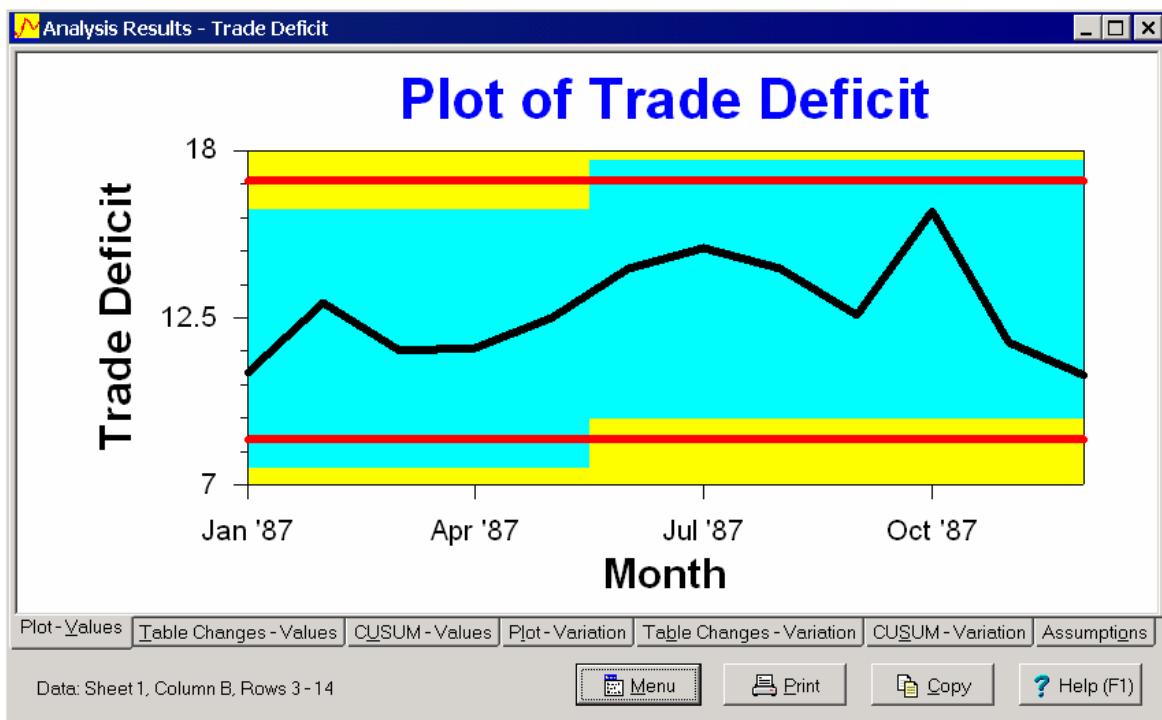


Figure 3: Change Point Analysis Using 1987 Data

You can also select a subset of the data using the Data window. Select only the values you want to analyze by dragging the mouse over these values. The exact steps are:

1. Move the mouse cursor over the cell in Row 3 of column B.
2. Press and hold down the left mouse button.
3. Move the mouse cursor down to the cell in Row 14 of column B.

4. Release the left mouse button.

Just the data from 1987 will be highlighted. You can now analyze the 1987 data by clicking the Fast Analysis button.

Specifying a Title for Your Data

The data in Figure 3 has the title Trade Deficit. This title appears both in the plot title and on the left axis. Change-Point Analyzer choose this title because it found this text in Column B. If column B did not contain any text, Change-Point Analyzer would have chosen the title Column B. You can use the **Name box** on the first tab of the Custom Analysis dialog box to specify some other title.

Specifying Labels for Your Data Points

The second tab, shown in Figure 4, is used to specify the column or row to be used for labels. The **Column Containing Labels** box is used to select the column containing the labels. It is initialized to column A since the Label button was used previously to mark column A as labels. The **Name for Labels** box is used to enter a name for the labels. This name is displayed on the bottom axis of the plots. When column A is selected, the label name is automatically initialized to Month because Change-Point Analyzer found this extra text in column A.

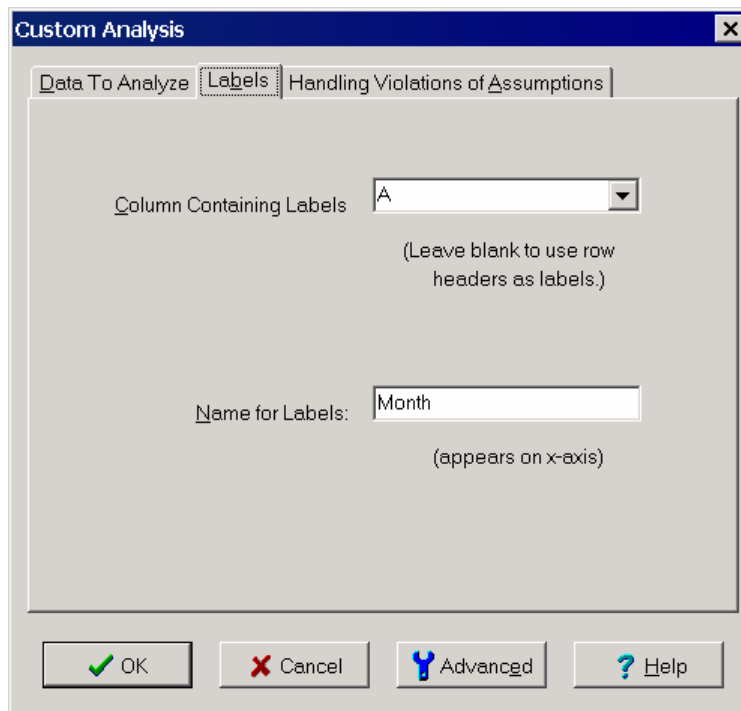


Figure 4: Tab 2 of Custom Analysis Dialog Box

Handling Violations of Assumptions

The third tab of the Custom Analysis dialog box, shown in Figure 5, allows several variations in the analysis which are sometimes necessary to handle outliers and violations of assumptions. Tutorials 5 and 6 show how to use this tab for handling these two complications. Change-Point Analyzer automatically checks for outliers and violations of assumptions and notifies you if either is found along with advice on how to proceed. You don't need to read these two tutorials unless you are notified of one of these complications.

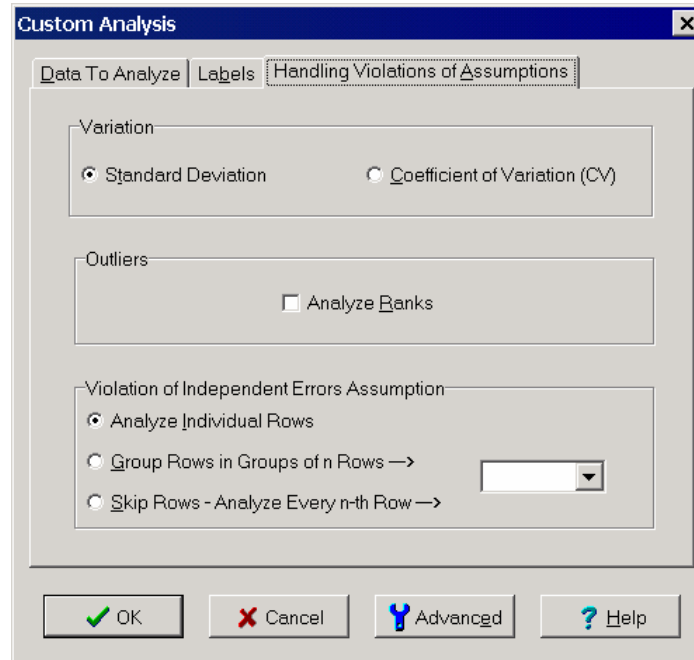


Figure 5: Tab 3 of Custom Analysis Dialog Box

There are occasions you might want to use these analysis features even when the program has not indicated outliers or a violation of assumptions. The analyze ranks feature replaces the data values with their corresponding ranks before performing the analysis. This makes the data robust to outliers giving it the ability to detect shifts despite the presence of the added noise created by the outliers. Further details are given in Tutorial 5. This feature may improve the analysis any time there are unusual values even when they do not trigger the outlier test. An example of such data is particulate counts where as many as 5-10% of the points might appear unusually high and thus fool the outlier test looking for 1 or 2 unusual values.

Another useful feature is the ability to analyze the coefficient of variation (CV) instead of the standard deviation. Certain types of data exhibit the behavior that the variation tends to increase proportional to the average. If the average shifts upwards by 50% so will the standard deviation. Some common examples are filling process, measurements systems and accuracy of cannons. Such processes are better characterized by the ratio of the standard deviation to the average, called the CV:

$$CV = 100 (\text{Standard Deviation}) / \text{Average}$$

CV is also frequently referred to as the relative standard deviation (RSD). A CV of 3% means the standard deviation is 3% of the average. In this case, a change-point analysis of the standard deviation will show shifts that mimic the shifts in the average and provides little additional information. An analysis of the CV instead would reveal if there was a change in the ratio between the standard deviation and average.