

## Chapter 201

# Descriptive Statistics – Summary Tables

---

## Introduction

This procedure is used to summarize continuous data. Large volumes of such data may be easily summarized in statistical tables of means, counts, standard deviations, etc. Categorical group variables may be used to calculate summaries for individual groups. The tables are similar in structure to those produced by cross tabulation.

This procedure produces tables of the following summary statistics:

- **Count**
- **Missing Count**
- **Sum**
- **Mean**
- **Standard Deviation (Std Dev)**
- **Standard Error (Std Error)**
- **Lower 95% Confidence Limit for the Mean (95% LCL)**
- **Upper 95% Confidence Limit for the Mean (95% UCL)**
- **Median**
- **Minimum**
- **Maximum**
- **Range**
- **Interquartile Range (IQR)**
- **10th Percentile (10th Pctile)**
- **25th Percentile (25th Pctile)**
- **75th Percentile (75th Pctile)**
- **90th Percentile (90th Pctile)**
- **Variance**
- **Mean Absolute Deviation (MAD)**
- **Mean Absolute Deviation from the Median (MADM)**
- **Coefficient of Variation (COV)**
- **Coefficient of Dispersion (COD)**
- **Skewness**
- **Kurtosis**

---

## Types of Categorical Variables

Note that we will refer to two types of categorical variables: *Group Variables* and *Break Variables*.

The values of a *Group Variable* are used to define the rows, sub rows, and columns of the summary table. Up to two *Group Variables* may be used per table. *Group Variables* are not required.

*Break Variables* are used to split a database into subgroups. A separate report is generated for each unique set of values of the break variables.

## Data Structure

The data below are a subset of the Resale dataset provided with the software. This (computer simulated) data gives the selling price, the number of bedrooms, the total square footage (finished and unfinished), and the size of the lots for 150 residential properties sold during the last four months in two states. This data is representative of the type of data that may be analyzed with this procedure. Only the first 8 of the 150 observations are displayed.

### Resale dataset (subset)

State	Price	Bedrooms	TotalSqft	LotSize
Nev	260000	2	2042	10173
Nev	66900	3	1392	13069
Vir	127900	2	1792	7065
Nev	181900	3	2645	8484
Nev	262100	2	2613	8355
Nev	147500	2	1935	7056
Nev	167200	2	1278	6116
Nev	395700	2	1455	14422

## Missing Values

Observations with missing values in either the group variables or the continuous data variables are ignored. The procedure also allows you to specify up to 5 additional values to be considered as missing in categorical group variables.

## Summary Statistics

The following sections outline the summary statistics that are available in this procedure.

### Count

The number of non-missing data values,  $n$ . If no frequency variable was specified, this is the number of rows with non-missing values.

### Missing Count

The number of missing data values. If no frequency variable was specified, this is the number of rows with missing values.

### Sum

The sum (or total) of the data values.

$$Sum = \sum_{i=1}^n x_i$$

## Descriptive Statistics – Summary Tables

---

### Mean

The average of the data values.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

---

### Variance

The sample variance,  $s^2$ , is a popular measure of dispersion. It is an average of the squared deviations from the mean.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

---

### Standard Deviation (Std Dev)

The sample standard deviation,  $s$ , is a popular measure of dispersion. It measures the average distance between a single observation and the mean. It is equal to the square root of the sample variance.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

---

### Standard Error (Std Error)

The standard error of the mean, a measure of the variation of the sample mean about the population mean, is computed by dividing the sample standard deviation by the square root of the sample size.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

---

### 95% Confidence Interval for the Mean (95% LCL & 95% UCL)

This is the upper and lower values of a 95% confidence interval estimate for the mean based on a  $t$  distribution with  $n - 1$  degrees of freedom. This interval estimate assumes that the population standard deviation is not known and that the data for this variable are normally distributed.

$$95\% \text{ CI} = \bar{x} \pm t_{\alpha/2, n-1} s_{\bar{x}}$$

---

### Minimum

The smallest data value.

---

### Maximum

The largest data value.

## Descriptive Statistics – Summary Tables

---

### Range

The difference between the largest and smallest data values.

$$\text{Range} = \text{Maximum} - \text{Minimum}$$

---

### Percentiles

The  $100p^{\text{th}}$  percentile is the value below which  $100p\%$  of data values may be found (and above which  $100p\%$  of data values may be found). The  $100p^{\text{th}}$  percentile is computed as

$$Z_{100p} = (1-g)X_{[k1]} + gX_{[k2]}$$

where  $k1$  equals the integer part of  $p(n+1)$ ,  $k2=k1+1$ ,  $g$  is the fractional part of  $p(n+1)$ , and  $X_{[k]}$  is the  $k^{\text{th}}$  observation when the data are sorted from lowest to highest.

---

### Median

The median (or 50th percentile) is the “middle number” of the sorted data values.

$$\text{Median} = Z_{50}$$

---

### Interquartile Range (IQR)

The difference between the 75th and 25th percentiles (the 3rd and 1st quartiles). This represents the range of the middle 50% of the data. It serves as a robust measure of the variation in the data.

$$\text{IQR} = Z_{75} - Z_{25}$$

---

### Mean Absolute Deviation (MAD)

A measure of dispersion that is not affected by outliers as much as the standard deviation and variance. It measures the average absolute distance between a single observation and the mean.

$$\text{MAD} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

---

### Mean Absolute Deviation from the Median (MADM)

A measure of dispersion that is even more robust to outliers than the mean absolute deviation (MAD) since the median is used as the center point of the distribution. It measures the average absolute distance between a single observation and the median.

$$\text{MADM} = \frac{\sum_{i=1}^n |x_i - \text{Median}|}{n}$$

## Descriptive Statistics – Summary Tables

### Coefficient of Variation (COV)

A relative measure of dispersion used to compare the amount of variation in two samples. It is calculated by dividing the standard deviation by the mean. Sometimes it is referred to as COV or CV.

$$COV = \frac{s}{\bar{x}}$$

### Coefficient of Dispersion (COD)

A robust, relative measure of dispersion. It is calculated by dividing the robust mean absolute deviation from the median (MADM) by the median. It is frequently used in real estate or tax assessment applications.

$$COD = \frac{MADM}{Median} = \frac{\left( \frac{\sum_{i=1}^n |x_i - Median|}{n} \right)}{Median}$$

### Skewness

Measures the direction and degree of asymmetry in the data distribution.

$$Skewness = \frac{m_3}{m_2^{3/2}}$$

where

$$m_r = \frac{\sum_{i=1}^n (x_i - \bar{x})^r}{n}$$

### Kurtosis

Measures the heaviness of the tails in the data distribution.

$$Kurtosis = \frac{m_4}{m_2^2}$$

where

$$m_r = \frac{\sum_{i=1}^n (x_i - \bar{x})^r}{n}$$

---

## Procedure Options

This section describes the options available in this procedure. To find out more about using a procedure, turn to the Procedures chapter.

---

### Variables Tab

This panel specifies the variables that will be used in the analysis and the summary table contents and layout.

---

#### Summary Table Contents and Layout

##### Data Variable(s)

Specify one or more variables whose descriptive statistics are to be calculated. These statistics, selected from those available, will be computed for each combination of the values in the categorical group variables (if any) that you have selected. The data in these variables must be numeric. Text values will be skipped in the calculations.

##### Table Position for Data Variable(s)

Select the position for Data Variable(s) in the summary table(s). See *Layout Preview for Items in the Table*.

The options are

- **Auto**  
The position for the data variable(s) is automatically determined based on the location of other items in the table.
- **Rows**  
Each data variable is listed as a separate row in the table.
- **Columns**  
Each data variable is listed as a separate column in the table.
- **Sub Rows**  
Each data variable is listed as a separate sub row in the table.
- **Tables**  
A separate table is created for each data variable.

The positions for Data Variable(s) and for Statistics can both be set to “Tables” only if there is at least one group variable. In this case, a separate table is created for each Data Variable/Statistic combination.

##### Plots

Plots are created with a table’s row item on the group axis and the column item as the legend variable. When there is only one column, the sub row item is used as the legend variable. For combined plots from tables with rows, sub rows, and columns, the rows and sub rows are combined onto the group axis and the columns are included in the legend.

##### Statistics

Select one or more statistics to be included in the table(s) and plot(s). The statistics are computed separately for each Data Variable. If one or more Group Variables are entered, the statistics are computed for each combination of the group values. The order of the statistics on the table(s) and plot(s) can be changed using the up/down arrow buttons to the right.

## Descriptive Statistics – Summary Tables

### Table Position for Statistics

Select the position for Statistics in the summary table(s). See *Layout Preview for Items in the Table*.

The options are

- **Auto**  
The position for the statistics is automatically determined based on the location of other items in the table.
- **Rows**  
Each statistic is listed as a separate row in the table.
- **Columns**  
Each statistic is listed as a separate column in the table.
- **Sub Rows**  
Each statistic is listed as a separate sub row in the table.
- **Tables**  
A separate table is created for each statistic.

The positions for Data Variable(s) and for Statistics can both be set to “Tables” only if there is at least one group variable. In this case, a separate table is created for each Data Variable/Statistic combination.

### Plots

Plots are created with a table’s row item on the group axis and the column item as the legend variable. When there is only one column, the sub row item is used as the legend variable. For combined plots from tables with rows, sub rows, and columns, the rows and sub rows are combined onto the group axis and the columns are included in the legend.

### Include Group Variable 1, 2

Check to include a categorical group variable in the table. When only Group Variable 1 is used, statistics are computed for each unique group value. When both Group Variable 1 and Group Variable 2 are used, statistics are computed for each combination of group values.

### Group Variable 1, 2

Specify one or more categorical group variables. The categories may be text (e.g. “Low, Med, High”) or numeric (e.g. “1, 2, 3”). When only Group Variable 1 is used, statistics are computed for each unique group value. When both Group Variable 1 and Group Variable 2 are used, statistics are computed for each combination of group values. If more than one variable is entered for Group Variable 1, a separate table will be created for each variable entered.

The data values in each variable will be sorted alpha-numerically before being listed in the table. If you want the values to be displayed in a different order, specify a custom value order for the data columns entered here using the Column Info Table on the Data Window.

## Descriptive Statistics – Summary Tables

### Table Position for Group Variable 1, 2

Select the position for the values of the Group Variable in the summary table(s). See *Layout Preview for Items in the Table*.

The options are

- **Auto**  
The position for the group variable is automatically determined based on the location of other items in the table.
- **Rows**  
Each group variable is listed as a separate row in the table.
- **Columns**  
Each group variable is listed as a separate column in the table.
- **Sub Rows**  
Each group variable is listed as a separate sub row in the table.

### Plots

Plots are created with a table's row item on the group axis and the column item as the legend variable. When there is only one column, the sub row item is used as the legend variable. For combined plots from tables with rows, sub rows, and columns, the rows and sub rows are combined onto the group axis and the columns are included in the legend.

### Create Other Group Variables from Numeric Data

Check this box to create group variables from numeric data. When checked, additional options will be displayed to specify how the numeric data will be classified into categorical variables.

If you choose to create group variables from numeric data, you do not have to enter a categorical group variable in the input box above (but you can).

If both numeric and categorical group variables are entered, a separate table and analysis will be calculated for each variable.

### Numeric Variable(s) to Categorize

Specify one or more variables that have only numeric values. Numeric values from these variables will be combined into a set of categories using the categorization options that follow. If more than one variable is entered, a separate table will be created for each variable.

For example, suppose you want to tabulate a variable containing individual income values into four categories: "Below 10000", "10000 to 40000", "40000 to 80000", and "Over 80000". You could select the income variable here, set **Group Numeric Data into Categories Using** to "List of Interval Upper Limits" and set the **List** to "10000 40000 80000".



## Descriptive Statistics – Summary Tables

### Group Numeric Data into Categories Using

Choose the method by which numeric data will be combined into categories.

The choices are:

- **Number of Intervals, Minimum, and/or Width**

This option allows you to specify the categories by entering any combination of the three parameters:

Number of Intervals

Minimum

Width

All three are optional.

#### Number of Intervals

This is the number of intervals into which the values of the numeric variables are categorized. If not enough intervals are specified to reach past the maximum data value, more will be added.

#### *Range*

Integer  $\geq 2$

#### Minimum

This value is used in conjunction with the Number of Intervals and Width values to construct a set of intervals into which the numeric variables are categorized. This is the minimum value of the first interval.

#### *Range*

This value must be less than the minimum data value.

#### Width

This value is used in conjunction with the Number of Intervals and Minimum values to construct a set of intervals into which the numeric variables are categorized. All intervals will have a width equal to this value. A data value  $X$  is in this interval if

$$\text{Lower Limit} < X \leq \text{Upper Limit}.$$

- **List of Interval Upper Limits**

This option allows you to specify the categories for the numeric variable by entering a list of interval boundaries directly, separated by blanks or commas. An interval of the form  $LI < X \leq L2$  is generated for each interval. The actual number of intervals is one more than the number of items specified here.

For example, suppose you want to tabulate a variable containing individual income values into four categories: “Below 10000”, “10000 to 40000”, “40000 to 80000”, and “Over 80000”. You would set **List of Interval Upper Limits** to “10000 40000 80000”. Note that 10000 would be included in the “Below 10000” interval, but not the “10000 to 40000” interval. Also, 80000 would be included in the “40000 to 80000” interval, not the “Over 80000” interval.

---

### Layout Preview for Items in the Table

This section provides a representation of how the results will be laid out in the table based on your item position choices. Colors are used only to highlight the position of items in the table. The actual output will all be black.

## Breaks, Frequencies Tab

This panel lets you specify up to eight break variables and a frequency variable. These variables are completely optional.

### Break Variables

Enter up to 8 categorical break variables. The values in these variables are used to break the output up into separate reports and plots. A separate set of reports is generated for each unique value (or unique combination of values if multiple break variables are specified). The break variables will be applied in the order that they are listed here.

### Frequency Variable

Specify an optional frequency (count) variable. This data column contains integers that represent the number of observations (frequency) associated with each row of the dataset. If this option is left blank, each dataset row has a frequency of one. This variable lets you modify that frequency. This may be useful when your data are tabulated and you want to enter counts.

---

## Missing Values Tab

This panel lets you specify up to five missing values (besides the default of blank). For example, '0', '9', or 'NA' may be missing values in your database.

### Missing Value Inclusion

Specifies whether to include observations with missing values in the tables.

*Delete All* indicates that you want the missing values totally ignored.

*Include in All* indicates that you want the missing values treated just like any other category.

### Missing Values

Specify up to five individual missing values here, one per box.

---

## Report Options Tab

The following options control the format of the reports.

---

### Group Variable Marginal Totals

#### Display Group Variable Marginal Total on the Summary Tables

Check to display marginal totals for group variables (if used) on the tables. If no group variables are being used then this option is ignored.

---

### Report Options

#### Variable Names

This option lets you select whether to display only variable names, variable labels, or both.

#### Value Labels

This option lets you select whether to display only values, value labels, or both. Use this option if you want the table to automatically attach labels to the values (like 1=Yes, 2=No, etc.). See the section on specifying *Value Labels* elsewhere in this manual.

---

## Summary Table Formatting

### Column Justification

Specify whether data columns in the tables will be left or right justified.

### Column Widths

Specify how the widths of columns in the contingency tables will be determined.

The options are

- **Autosize to Minimum Widths**

Each data column is individually resized to the smallest width required to display the data in the column. This usually results in columns with different widths. This option produces the most compact table possible, displaying the most data per page.

- **Autosize to Equal Minimum Width**

The smallest width of each data column is calculated and then all columns are resized to the width of the widest column. This results in the most compact table possible where all data columns have the same width. This is the default setting.

- **Custom (User-Specified)**

Specify the widths (in inches) of the columns directly instead of having the software calculate them for you.

### Custom Widths

Enter one or more values for the widths (in inches) of columns in the contingency tables.

- **Single Value**

If you enter a single value, that value will be used as the width for all data columns in the table.

- **List of Values**

Enter a list of values separated by spaces corresponding to the widths of each column. The first value is used for the width of the first data column, the second for the width of the second data column, and so forth. Extra values will be ignored. If you enter fewer values than the number of columns, the last value in your list will be used for the remaining columns.

Type the word “Autosize” for any column to cause the program to calculate its width for you. For example, enter “1 Autosize 0.7” to make column 1 be 1 inch wide, column 2 be sized by the program, and column 3 be 0.7 inches wide.

### Wrap Column Headings onto Two Lines

Check this option to make column headings wrap onto two lines. Use this option to condense your table when your data are spaced too far apart because of long column headings.

### Use Short Statistical Names on Reports and Plots

Normally, the names of the statistical items in the reports and plots are complete names, such as “Standard Deviation.” Checking this option causes a shorter name, such as “SD”, to be used instead so that more columns can be displayed together in tables and so that plot titles and labels are not so long. A maximum of 13 columns can be displayed on a single row.

## Descriptive Statistics – Summary Tables

---

### Decimal Places

#### Item Decimal Places

These decimal options allow the user to specify the number of decimal places for items in the output. Your choice here will not affect calculations; it will only affect the format of the output.

- **Auto**

If one of the “Auto” options is selected, the ending zero digits are not shown. For example, if “Auto (0 to 7)” is chosen,

0.0500 is displayed as 0.05

1.314583689 is displayed as 1.314584

The output formatting system is not designed to accommodate “Auto (0 to 13)”, and if chosen, this will likely lead to lines that run on to a second line. This option is included, however, for the rare case when a very large number of decimals is needed.

---

### Plots Tab

The options on this panel control the appearance of the plots that may be displayed. Click the plot format button to change the plot settings.

#### Show Separate Plots of each Statistic for each Table

Check to display plots for each statistic in each table. This may result in several plots being created for each table. Plots are created with a table’s row item on the group axis and the column item as the legend variable. When there is only one column, the sub row item is used as the legend variable.

#### Show a Combined Plot for each Table

Check to display a single plot for each table that contains all of the information in the table. Plots are created with a table’s row item on the group axis and the column item as the legend variable. When there is only one column, the sub row item is used as the legend variable. For combined plots from tables with rows, sub rows, and columns, the rows and sub rows are combined onto the group axis and the columns are included in the legend.

#### Display Break Variables Values as Subtitles on the Plots

Specify whether to display the values of the break variables as the second title line on the plots.

#### Display Group Variable Marginal Totals on the Plots

Check to display marginal totals for group variables (if used) on the plots. If no group variables are being used then this option is ignored.

## Example 1 – Basic Variable Summary Report (No Group Variables)

The data used in this example are in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template **Example 1a** by clicking on Open Example Template from the File menu of the Descriptive Statistics – Summary Tables window.

### 1 Open the Resale dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Resale.NCSS**.
- Click **Open**.

### 2 Open the Descriptive Statistics – Summary Tables window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Descriptive Statistics – Summary Tables** procedure
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

### 3 Specify the variables.

- Select the **Variables tab**.
- For **Data Variable(s)**, select **Price, Bedrooms, Bathrooms, Garage, and TotalSqft**.

### 4 Specify the statistics.

- In the Table Statistics section, check **Count, Mean, Std Dev, 95% LCL, and 95% UCL**.

### 5 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green **Run** button.

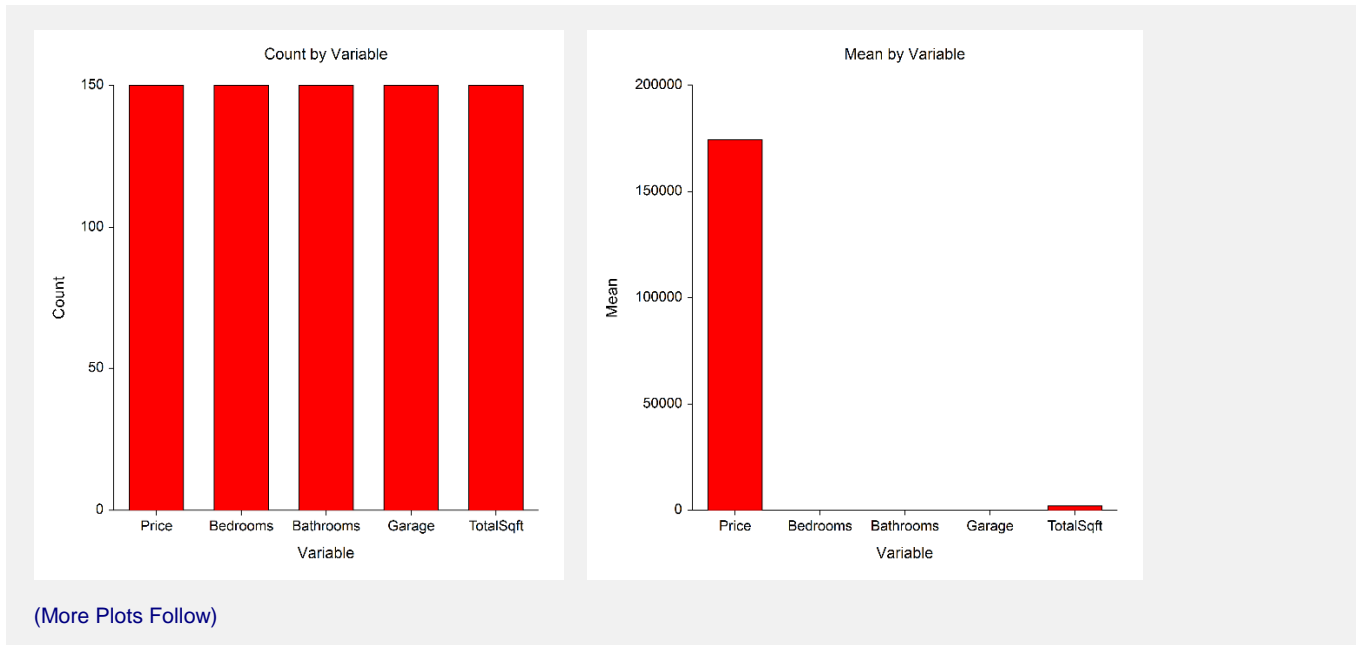
## Summary Table

Statistic	Variable				
	Price	Bedrooms	Bathrooms	Garage	TotalSqft
Count	150	150	150	150	150
Mean	174392	2.42	2.4	1.266667	1893.38
Standard Deviation	97656.81	0.8919476	0.8047677	0.5636252	754.2496
Lower 95% CL Mean	158636	2.276093	2.270158	1.175731	1771.689
Upper 95% CL Mean	190148	2.563908	2.529842	1.357602	2015.071

The table is created with the statistics as rows and the data variables as columns when the positions are both set to “Auto”.

## Descriptive Statistics – Summary Tables

## Plots of Each Statistic



The plots are not very informative because the variables have vastly different scales.

### Example 1b – Adjust Item Table Positions (Data Variables in Rows and Statistics in Columns)

To rotate the table, all we have to do is change the position of one of the items. Go back to the Descriptive Statistics – Summary Tables procedure window and change the position for Data Variable(s) to “Rows” (or load the completed template **Example 1b** by clicking on Open Example Template from the File menu) and run the procedure again to get the results.

#### 6 Specify the group variable.

- For **Data Variable(s) Position**, select **Rows** and re-run the procedure.

<u>Variable</u>	<u>Statistic</u>				
	<b>Count</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Lower 95% CL Mean</b>	<b>Upper 95% CL Mean</b>
<b>Price</b>	150	174392	97656.81	158636	190148
<b>Bedrooms</b>	150	2.42	0.8919476	2.276093	2.563908
<b>Bathrooms</b>	150	2.4	0.8047677	2.270158	2.529842
<b>Garage</b>	150	1.266667	0.5636252	1.175731	1.357602
<b>TotalSqft</b>	150	1893.38	754.2496	1771.689	2015.071

The table is now rotated with the data variables as rows and the statistics as columns. Notice that the actual summary statistic values are exactly the same.

---

## Example 2 – Variable Summary Report (One Group Variable)

The data used in this example are in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template **Example 2a** by clicking on Open Example Template from the File menu of the Descriptive Statistics – Summary Tables window.

### 1 Open the Resale dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Resale.NCSS**.
- Click **Open**.

### 2 Open the Descriptive Statistics – Summary Tables window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Descriptive Statistics – Summary Tables** procedure
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

### 3 Specify the variables.

- Select the **Variables** tab.
- For **Data Variable(s)**, select **Price**, **TotalSqft**, and **LotSize**.

### 4 Specify the statistics.

- In the Table Statistics section, check **Count**, **Mean**, and **Std Dev**.

### 5 Specify the group variable.

- Check **Group Variable 1** and select **State** as the variable.

### 6 Specify the report format.

- Click on the **Report Options** tab.
- For **Variable Names**, select **Labels**.
- For **Value Labels**, select **Value Labels**.

### 7 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green **Run** button.

## Descriptive Statistics – Summary Tables

## Summary Table

		<u>Variable</u>		
<u>State</u>	<u>Statistic</u>	<u>Sales Price</u>	<u>Total Area (Sqft)</u>	<u>Lot Size (Sqft)</u>
<b>Nevada</b>	Count	88	88	88
	Mean	170762.5	1881.33	8571.454
	Standard Deviation	98665.72	788.569	2419.88
<b>Virginia</b>	Count	62	62	62
	Mean	179543.5	1910.484	8076.597
	Standard Deviation	96771.49	708.6572	2301.226
<b>Total</b>	Count	150	150	150
	Mean	174392	1893.38	8366.913
	Standard Deviation	97656.81	754.2496	2376.334

The table displays the group variable values as the rows, the statistics as the subrows, and the data variables as the columns. The plots are not shown because they are not very informative because the variables have vastly different scales. Totals are given for the group variable.

### Example 2b – Adjust Item Table Positions (Data Variables in Rows, Statistics in Sub Rows, and Group Variable in Columns)

To rotate the table, simply change the position the items. Go back to the Descriptive Statistics – Summary Tables procedure window and change the position for Data Variable(s) to “Rows” (or load the completed template **Example 2b** by clicking on Open Example Template from the File menu) and run the procedure again to get the results.

#### 8 Specify the group variable.

- For **Data Variable(s) Position**, select **Rows** and re-run the procedure.

		<u>State</u>		
<u>Variable</u>	<u>Statistic</u>	<u>Nevada</u>	<u>Virginia</u>	<u>Total</u>
<b>Sales Price</b>	Count	88	62	150
	Mean	170762.5	179543.5	174392
	Standard Deviation	98665.72	96771.49	97656.81
<b>Total Area (Sqft)</b>	Count	88	62	150
	Mean	1881.33	1910.484	1893.38
	Standard Deviation	788.569	708.6572	754.2496
<b>Lot Size (Sqft)</b>	Count	88	62	150
	Mean	8571.454	8076.597	8366.913
	Standard Deviation	2419.88	2301.226	2376.334

The table is now rotated with the data variables as rows and the group variable values as columns. Notice that the actual summary statistic values are exactly the same.



## Descriptive Statistics – Summary Tables

**Example 2c – Adjust Item Table Positions (Data Variables in Rows, Group Variable in Sub Rows, and Statistics in Columns)**

To change the table so that statistics are presented as columns with the group variable as subrows and the data variables as rows, change the position of Statistics to “Columns” with the position for Data Variable(s) still set to “Rows” (or load the completed template **Example 2c** by clicking on Open Example Template from the File menu) and run the procedure again to get the results.

**9 Specify the statistics.**

- For **Statistics Position**, select **Columns** and re-run the procedure.

<u>Variable</u>	<u>State</u>	<u>Count</u>	<u>Statistic</u>	
			<u>Mean</u>	<u>Standard Deviation</u>
<b>Sales Price</b>	Nevada	88	170762.5	98665.72
	Virginia	62	179543.5	96771.49
	Total	150	174392	97656.81
<b>Total Area (Sqft)</b>	Nevada	88	1881.33	788.569
	Virginia	62	1910.484	708.6572
	Total	150	1893.38	754.2496
<b>Lot Size (Sqft)</b>	Nevada	88	8571.454	2419.88
	Virginia	62	8076.597	2301.226
	Total	150	8366.913	2376.334

The table now has the data variables as rows and the group variable values as subrows with the statistics as columns.

## Example 3 – Variable Summary Report (Two Group Variables)

The data used in this example are in the Pain dataset. In this example we'll show you how to make even more customizations to adjust the appearance of the tables and plots and how easy it is to make position adjustments. You may follow along here by making the appropriate entries or load the completed template **Example 3a** by clicking on Open Example Template from the File menu of the Descriptive Statistics – Summary Tables window.

### 1 Open the Pain dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **Pain.NCSS**.
- Click **Open**.

### 2 Open the Descriptive Statistics – Summary Tables window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Descriptive Statistics – Summary Tables** procedure
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

### 3 Specify the variables.

- Select the **Variables** tab.
- For **Data Variable(s)**, select **Pain**.

### 4 Specify the statistics.

- In the Table Statistics section, click the **Uncheck All** button to uncheck all selected statistics.
- In the Table Statistics section, check **Mean, Median, Minimum, Maximum, 25<sup>th</sup> Pctile, and 75<sup>th</sup> Pctile**.
- Use the up/down arrow buttons to move the checked statistics so that the order is **Mean, Minimum, 25<sup>th</sup> Pctile, Median, 75<sup>th</sup> Pctile, then Maximum**. It does not matter if there are unchecked statistics between checked statistics. Checked statistics will be output in the order they are listed.

### 5 Specify the group variables.

- Check **Group Variable 1** and select **Drug** as the variable.
- Check **Group Variable 2** and select **Time** as the variable.

### 6 Specify the report options and format.

- Click on the **Report Options** tab.
- Uncheck **Display Group Variable Marginal Total on the Summary Tables** so totals are not displayed.
- Check **Use Short Statistical Names on Reports and Plots** to fit more columns on one row of the table.
- Change the **Decimal Places for Sum, Mean, CI Limits** to **2** to limit the width of the displayed values.

### 7 Specify the plots.

- Click on the **Plots** tab.
- Click on the **Separate Plot... Format** button.
- On the **Bar Chart Format** window, select the **Numeric Axis** tab and enter **100** for **Max** under **Boundaries**. This will make it so that all statistic plots are on the same scale for easy comparison among plots. Click **OK** to save the plot settings.
- Check **Show a Combined Plot for Each Table**.
- Click on the **Combined Plot Format** button.
- On the **Bar Chart Format** window, select the **Group Axis** tab and click on the **Layout** button for the **Lower Axis Tick Label**.
- Change **Alignment** to **Right**, **Rotation Angle** to **45**, and **Margin Above the Text** to **10**.
- Click **OK** to save the layout settings and **OK** again to save the plot settings.

Descriptive Statistics – Summary Tables

8 Run the procedure.

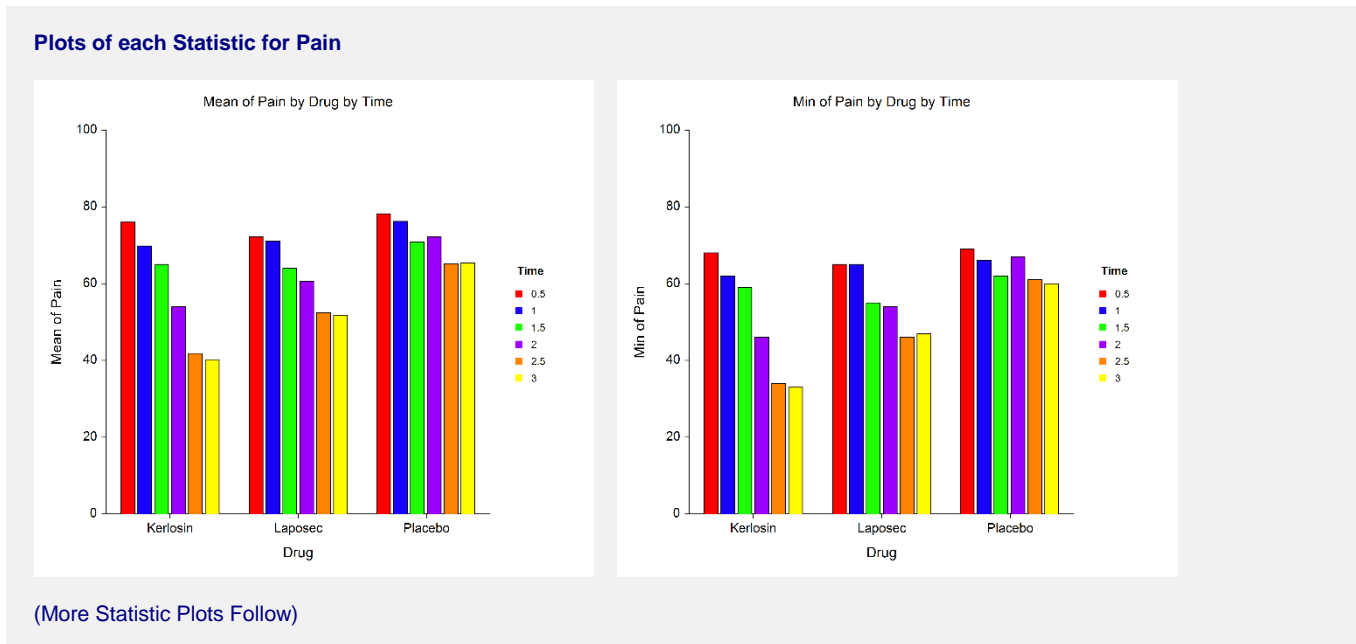
- From the Run menu, select **Run Procedure**. Alternatively, just click the green **Run** button.

Output

Summary Table of Pain

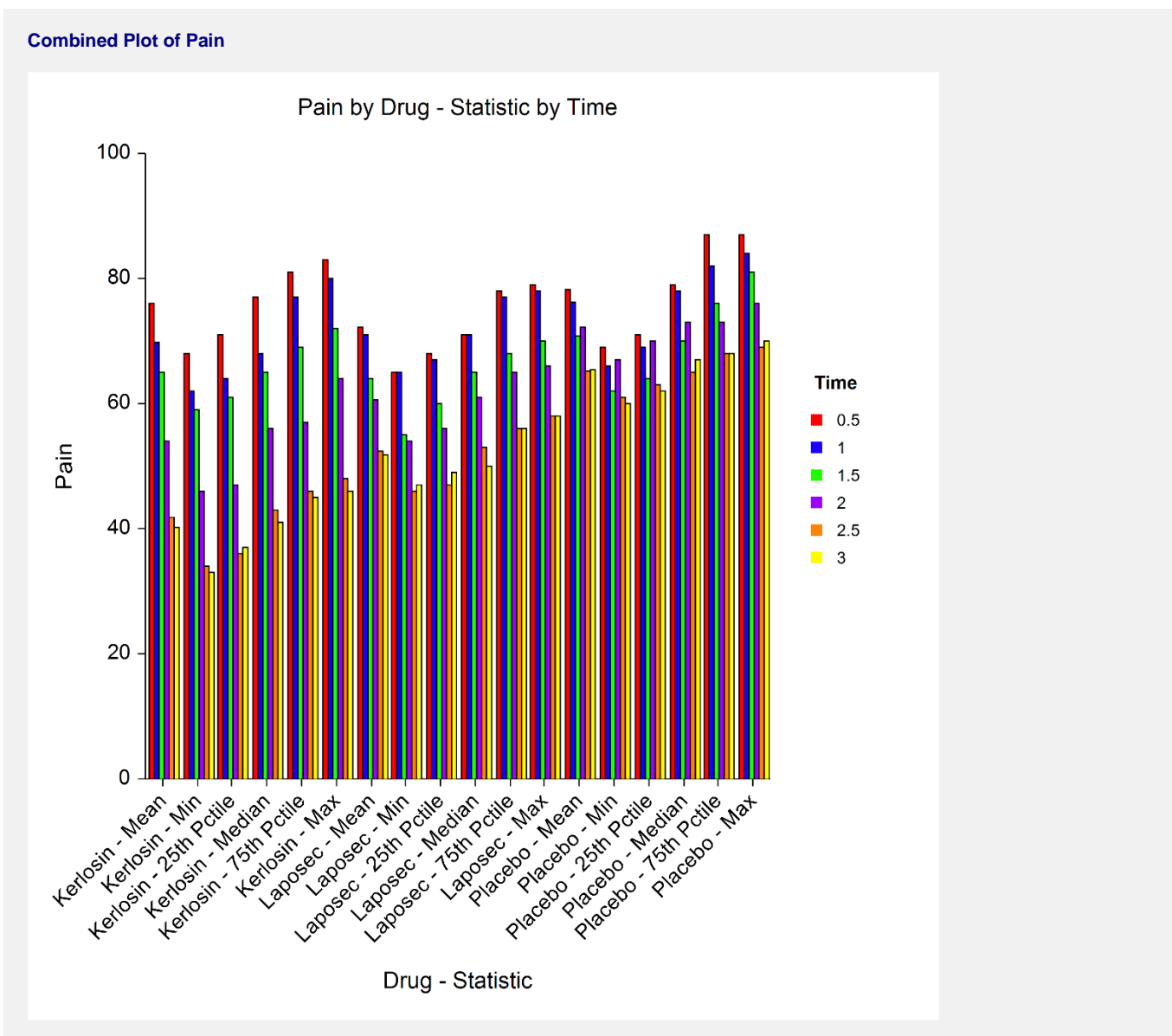
Drug	Statistic	Time					
		0.5	1	1.5	2	2.5	3
Kerlosin	Mean	76.00	69.86	65.00	54.00	41.71	40.14
	Min	68	62	59	46	34	33
	25th Pctile	71	64	61	47	36	37
	Median	77	68	65	56	43	41
	75th Pctile	81	77	69	57	46	45
	Max	83	80	72	64	48	46
Laposec	Mean	72.29	71.00	64.00	60.57	52.43	51.71
	Min	65	65	55	54	46	47
	25th Pctile	68	67	60	56	47	49
	Median	71	71	65	61	53	50
	75th Pctile	78	77	68	65	56	56
	Max	79	78	70	66	58	58
Placebo	Mean	78.14	76.29	70.86	72.14	65.14	65.43
	Min	69	66	62	67	61	60
	25th Pctile	71	69	64	70	63	62
	Median	79	78	70	73	65	67
	75th Pctile	87	82	76	73	68	68
	Max	87	84	81	76	69	70

The table displays Group Variable 1 (Drug) values as the rows, the statistics as the subrows, and Group Variable 2 (Time) values as the columns.



Individual plots are created with the table row item (Group Variable 1 --- “Drug”) on the group (X) axis and the table column item (Group Variable 2 --- “Time”) as the legend variable. A separate plot is created for each statistic. These plots are very useful for seeing overall trends. From the plots shown here, it is apparent that the average and minimum pain response is lower for both drugs than for placebo and that the pain control is better over time. Kerlosin appears to control pain the best from these results. Statistical tests would need to be performed, however, to assert statistical significance in the differences.

## Descriptive Statistics – Summary Tables



The combined plot displays all of the information in the table. We rotated the group axis labels so they would not overlap and be readable. The table row item (Group Variable 1 --- “Drug”) and table sub row item (Statistic) are combined on the group (X) axis. The table column item (Group Variable 2 --- “Time”) is the legend variable.

### Example 3b – Adjust Item Table Positions (Group 2 Variable in Rows, Group 1 Variable in Sub Rows, and Statistics in Columns)

To change the orientation on the tables and plots, simply change the position the items. Let’s display the **Statistics as the columns** and **Time as the rows**. This will put **Drug as the sub row**. Go back to the Descriptive Statistics – Summary Tables procedure window and change the position for Statistics to “Columns” and the position for Group Variable 2 (Time) to “Rows” (or load the completed template **Example 3b** by clicking on Open Example Template from the File menu) and run the procedure again to get the results.

#### 9 Specify the statistics and group variables.

- For **Statistics Position**, select **Columns**.
- For **Group Variable 2 Position**, select **Rows** and re-run the procedure.

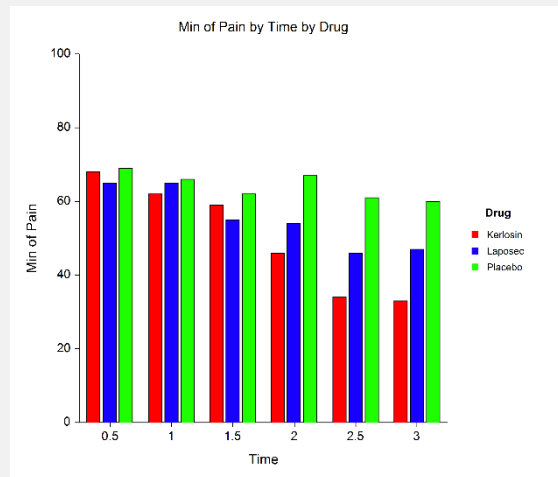
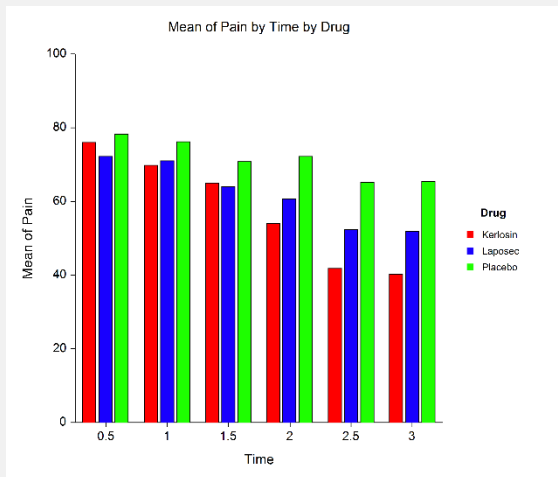
### Descriptive Statistics – Summary Tables

Summary Table of Pain

Time	Drug	Statistic					
		Mean	Min	25th Pctile	Median	75th Pctile	Max
0.5	Kerlosin	76.00	68	71	77	81	83
	Laposec	72.29	65	68	71	78	79
	Placebo	78.14	69	71	79	87	87
1	Kerlosin	69.86	62	64	68	77	80
	Laposec	71.00	65	67	71	77	78
	Placebo	76.29	66	69	78	82	84
1.5	Kerlosin	65.00	59	61	65	69	72
	Laposec	64.00	55	60	65	68	70
	Placebo	70.86	62	64	70	76	81
2	Kerlosin	54.00	46	47	56	57	64
	Laposec	60.57	54	56	61	65	66
	Placebo	72.14	67	70	73	73	76
2.5	Kerlosin	41.71	34	36	43	46	48
	Laposec	52.43	46	47	53	56	58
	Placebo	65.14	61	63	65	68	69
3	Kerlosin	40.14	33	37	41	45	46
	Laposec	51.71	47	49	50	56	58
	Placebo	65.43	60	62	67	68	70

The table displays Group Variable 2 (Time) values as the rows, Group Variable 1 (Drug) values as the subrows, and the statistics as the columns.

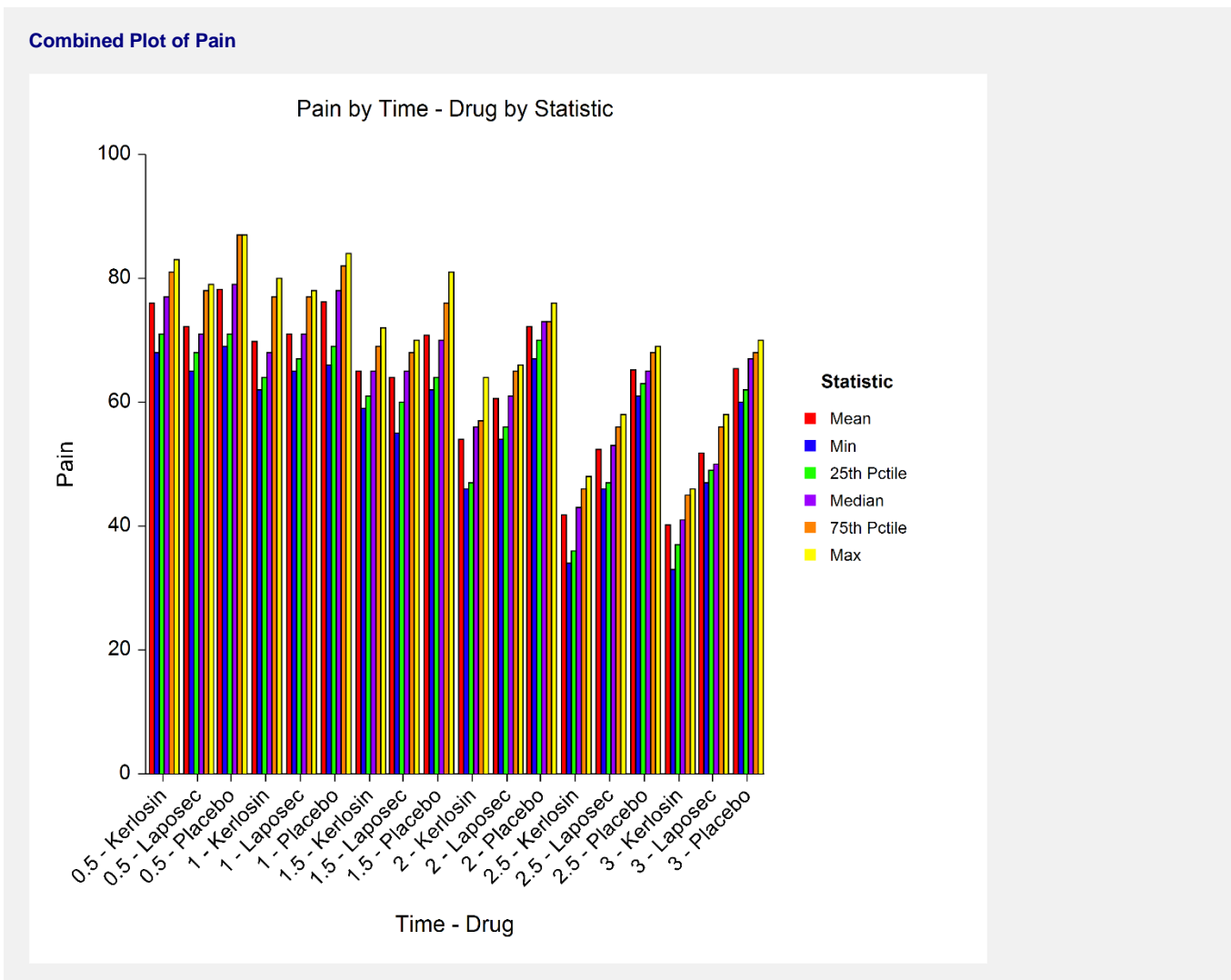
Plots of each Statistic for Pain



(More Statistic Plots Follow)

The individual plots are different now with the table row item (Group Variable 2 --- “Time”) on the group (X) axis and the table column item (Group Variable 1 --- “Drug”) as the legend variable. A separate plot is created for each statistic. These plots are again useful for seeing overall trends. There is a very distinct reduction in pain over time.

## Descriptive Statistics – Summary Tables



Again, the combined plot displays all of the information in the table. The table row item (Group Variable 2 --- “Time”) and table sub row item (Group Variable 1 --- “Drug”) are combined on the group (X) axis. The table column item (Statistic) is the legend variable.

### Example 3c – Adjust Item Table Positions (Creating a Separate Table for each Data Variable and Statistic Combination)

It’s easy to create a separate table for each data variable and statistic combination (this can only be done when there is at least one group variable). Let’s display a separate table for each statistic with **Time as the rows** and **Drug as the columns**. There will be no sub row item. Go back to the Descriptive Statistics – Summary Tables procedure window and change the position for Data Variable(s) to “Tables” and the position for Statistics to “Tables” as well. Leave the position for Group Variable 2 (Time) as “Rows” (or load the completed template **Example 3c** by clicking on Open Example Template from the File menu) and run the procedure again to get the results.

#### 10 Specify the statistics and group variables.

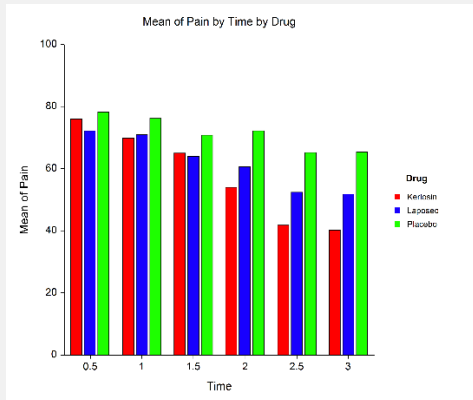
- For **Data Variable(s) Position**, select **Tables**.
- For **Statistics Position**, select **Tables**.
- Leave **Group Variable 2 Position** as **Rows** and re-run the procedure.

Descriptive Statistics – Summary Tables

Summary Table of Mean of Pain

Time	Drug		
	Kerlosin	Laposec	Placebo
0.5	76.00	72.29	78.14
1	69.86	71.00	76.29
1.5	65.00	64.00	70.86
2	54.00	60.57	72.14
2.5	41.71	52.43	65.14
3	40.14	51.71	65.43

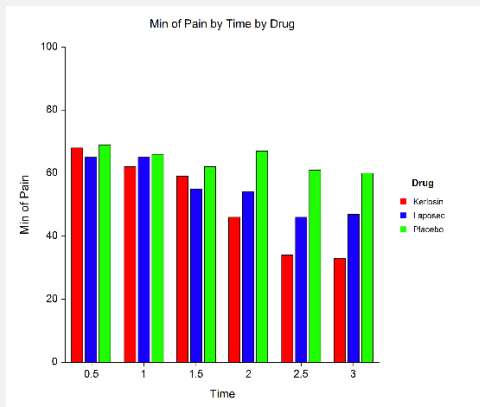
Plot of Mean of Pain



Summary Table of Min of Pain

Time	Drug		
	Kerlosin	Laposec	Placebo
0.5	68	65	69
1	62	65	66
1.5	59	55	62
2	46	54	67
2.5	34	46	61
3	33	47	60

Plot of Min of Pain



(Report continues with table and plot for each Data Variable/Statistic combination)

A separate table is created for each Data Variable/Statistic combination. If more than one data variable were entered, the report would be even longer. There is no combined plot in the output because the combined plot is the same as the individual plot in this case.