Chapter 193

Descriptive Statistics – Summary Tables
(Old Version)

Introduction

This procedure produces tables of means, medians, percentiles, standard deviations, coefficients of variation, sums, and counts for various combinations of grouping (break) variables. Nine tabular formats are available. The tables are similar in structure to those produced by cross tabulation.

This module is used to summarize data containing a combination of continuous and categorical variables. Large volumes of such data may be summarized in statistical tables of means, counts, or standard deviation. Discussions of these statistics are found in the Descriptive Statistics chapter and will not be reproduced here.

Types of Categorical Variables

Note that we will refer to two types of categorical variables: Categorical and Grouping. Grouping variables are used to split a database into subgroups. A separate table is generated for each unique set of values of the grouping variables. The values of a categorical variable are used to define the rows and columns of the tabulation table. Up to two categorical variables may be used per table.

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. Only the first 8 of the 150 observations are displayed.
Resale dataset (subset)

<table>
<thead>
<tr>
<th>State</th>
<th>Price</th>
<th>Bedrooms</th>
<th>TotalSqft</th>
<th>LotSize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nev</td>
<td>260000</td>
<td>2</td>
<td>2042</td>
<td>10173</td>
</tr>
<tr>
<td>Nev</td>
<td>66900</td>
<td>3</td>
<td>1392</td>
<td>13069</td>
</tr>
<tr>
<td>Vir</td>
<td>127900</td>
<td>2</td>
<td>1792</td>
<td>7065</td>
</tr>
<tr>
<td>Nev</td>
<td>181900</td>
<td>3</td>
<td>2645</td>
<td>8484</td>
</tr>
<tr>
<td>Nev</td>
<td>262100</td>
<td>2</td>
<td>2613</td>
<td>8355</td>
</tr>
<tr>
<td>Nev</td>
<td>147500</td>
<td>2</td>
<td>1935</td>
<td>7056</td>
</tr>
<tr>
<td>Nev</td>
<td>167200</td>
<td>2</td>
<td>1278</td>
<td>6116</td>
</tr>
<tr>
<td>Nev</td>
<td>395700</td>
<td>2</td>
<td>1455</td>
<td>14422</td>
</tr>
</tbody>
</table>

**Missing Values**

Observations with missing values in either the categorical variable or the continuous variable are ignored.

**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.

**Table Layout**

**Layout**

This option specifies the table layout. These layouts are defined in terms of the number of tables, the table rows, the sub-rows within a row, and the table columns.

- **1 TABLES: One, ROWS: Data, COLUMNS: Statistics**
  - **Tables**: only one table.
  - **Rows**: a row for each Data Variable.
  - **Columns**: a column for each Statistic.

  An example of this layout is:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count</th>
<th>Mean</th>
<th>Median</th>
<th>Std Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>122</td>
<td>12.04</td>
<td>12</td>
<td>4.527</td>
</tr>
<tr>
<td>X2</td>
<td>124</td>
<td>23.45</td>
<td>25</td>
<td>5.831</td>
</tr>
<tr>
<td>X3</td>
<td>133</td>
<td>34.16</td>
<td>38</td>
<td>6.094</td>
</tr>
<tr>
<td>X4</td>
<td>126</td>
<td>61.38</td>
<td>63</td>
<td>3.725</td>
</tr>
</tbody>
</table>
2 TABLES: One, ROWS: Categorical, COLUMNS: Data, SUBROWS: Statistics
Tables: only one table.
Rows: a set of rows for each category of the Row Variable.
Sub-Rows: a row for each Statistic.
Columns: a column for each Data Variable.

An example of this layout is:

<table>
<thead>
<tr>
<th>Row Variable</th>
<th>Data Variable</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Mean</td>
<td>12.36</td>
<td>23.77</td>
<td>51.78</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>57.62</td>
<td>62.17</td>
<td>79.18</td>
</tr>
<tr>
<td>Group 2</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
</tr>
</tbody>
</table>

3 TABLES: One, ROWS: Data, COLUMNS: Categorical, SUBROWS: Statistics
Tables: only one table.
Rows: a set of rows for each Data Variable.
Sub-Rows: a row for each Statistic.
Columns: a column for each category of the Column Variable.

An example of this layout is:

<table>
<thead>
<tr>
<th>Data Variables</th>
<th>Column Variable</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Mean</td>
<td>12.36</td>
<td>23.77</td>
<td>51.78</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>57.62</td>
<td>62.17</td>
<td>79.18</td>
</tr>
<tr>
<td>X2</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
</tr>
</tbody>
</table>

4 TABLES: Statistics, ROWS: Categorical, COLUMNS: Data
Tables: a separate table (and plot) for each Statistic.
Rows: a row for each category of the Row Variable.
Columns: a column for each Data Variable.

An example of this layout is:

<table>
<thead>
<tr>
<th>Table of Means</th>
<th>Data Variables</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row Variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>Mean</td>
<td>12.36</td>
<td>23.77</td>
<td>51.78</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>57.62</td>
<td>62.17</td>
<td>79.18</td>
</tr>
<tr>
<td>Group 2</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
</tr>
<tr>
<td>Group 3</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
</tr>
<tr>
<td>Group 4</td>
<td>Mean</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
</tr>
</tbody>
</table>

5 TABLES: Statistics, ROWS: Data, COLUMNS: Categorical
Tables: a separate table (and plot) for each Statistic.
Rows: a row for each Data Variable.
Columns: a column for each category of the Column Variable.

An example of this layout is:

<table>
<thead>
<tr>
<th>Table of Means</th>
<th>Data Variables</th>
<th>Column Variable</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Variables</td>
<td></td>
<td>Column Variable</td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
</tr>
<tr>
<td>X1</td>
<td>Mean</td>
<td>12.36</td>
<td>23.77</td>
<td>51.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>57.62</td>
<td>62.17</td>
<td>79.18</td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>Mean</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
<td></td>
</tr>
</tbody>
</table>
Descriptive Statistics – Summary Tables (Old Version)

- **6 TABLES: Statistics, ROWS: Categorical, COLUMNS: Categorical, SUBROWS: Data Tables**: a separate table for each Statistic.
  - **Rows**: a set of rows for each category of the Row Variable.
  - **Sub-Rows**: a row for each Data Variable.
  - **Columns**: a column for each category of the Column Variable.

An example of this layout is:

```
Table of Means

<table>
<thead>
<tr>
<th>Row Variable</th>
<th>Data Variable</th>
<th>Column Variable</th>
<th>Group1</th>
<th>Group2</th>
<th>Group3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level1</td>
<td>X1</td>
<td>12.36</td>
<td>23.77</td>
<td>51.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X2</td>
<td>57.62</td>
<td>62.17</td>
<td>79.18</td>
<td></td>
</tr>
<tr>
<td>Level2</td>
<td>X1</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X2</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
<td></td>
</tr>
</tbody>
</table>
```

- **7 TABLES: Data, ROWS: Categorical, COLUMNS: Categorical, SUBROWS: Statistics Tables**: a separate table for each Data Variable.
  - **Rows**: a set of rows for each category of the Row Variable.
  - **Sub-Rows**: a row for each Statistic.
  - **Columns**: a column for each category of the Column Variable.

An example of this layout is:

```
Summary for X1

<table>
<thead>
<tr>
<th>Row Variable</th>
<th>Column Variable</th>
<th>Group1</th>
<th>Group2</th>
<th>Group3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level1</td>
<td>Mean</td>
<td>12.36</td>
<td>23.77</td>
<td>51.78</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>57.62</td>
<td>62.17</td>
<td>79.18</td>
</tr>
<tr>
<td>Level2</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
</tr>
</tbody>
</table>
```

- **8 TABLES: Data and Statistics, ROWS: Categorical, COLUMNS: Categorical Tables**: a separate table (and plot) for each Data Variable and Statistic.
  - **Rows**: a row for each category of the Row Variable.
  - **Columns**: a column for each category of the Column Variable.

An example of this layout is:

```
Means of X1

<table>
<thead>
<tr>
<th>Row Variables</th>
<th>Column Variable</th>
<th>Group1</th>
<th>Group2</th>
<th>Group3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level1</td>
<td>Mean</td>
<td>12.36</td>
<td>23.77</td>
<td>51.78</td>
</tr>
<tr>
<td>Level2</td>
<td>Mean</td>
<td>57.62</td>
<td>62.17</td>
<td>79.18</td>
</tr>
<tr>
<td>Level3</td>
<td>Mean</td>
<td>87.65</td>
<td>54.32</td>
<td>43.21</td>
</tr>
<tr>
<td>Level4</td>
<td>Mean</td>
<td>42.89</td>
<td>22.33</td>
<td>64.87</td>
</tr>
</tbody>
</table>
```

- **9 Data Summary List**
  - An item-by-item list of the statistics. *Note that only six columns can be displayed.*
  - **Tables**: one table
  - **Rows**: a row for each Grouping Variable category and Row Variable category.
  - **Columns**: a column for each Grouping Variable, Row Variable, and statistic.

An example of this layout is:

```
Group Variable | Row Variable | Mean | StdDev
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Level1</td>
<td>Group1</td>
<td>12.36</td>
<td>23.77</td>
</tr>
<tr>
<td>Level1</td>
<td>Group2</td>
<td>57.62</td>
<td>62.17</td>
</tr>
<tr>
<td>Level2</td>
<td>Group1</td>
<td>54.32</td>
<td>43.21</td>
</tr>
<tr>
<td>Level2</td>
<td>Group2</td>
<td>22.33</td>
<td>64.87</td>
</tr>
</tbody>
</table>
```
**Numeric Data Variables**

**Data Variables**
Select one or more data variable. The statistics (means, standard deviations, etc.) generated will be for the values in these variables.

**Table Statistics: Count … 90th Pctile**
Check each statistic that is to be reported.

**Categorical Variables**

**Row (Column) Variable(s)**
Specify one or more categorical variables for use in table rows (columns). Each unique value in each variable will result in a separate row in the table. The data values themselves may be text (e.g. “Low, Med, High”) or numeric (e.g. “1, 2, 3”), but the data as a whole should be categorical. If more than one variable is entered, a separate table will be created for each variable.

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

**Create Other Row (Column) Variables from Numeric Data**
Check this box to create tables with rows 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 row (column) variables from numeric data, you do not have to enter a categorical row (column) variable in the input box above (but you can). If both numeric and categorical row (column) variables are entered, a separate table and analysis will be calculated for each variable.

**Numeric Variable(s) to Categorize for Use in Table Rows (Columns)**
Specify one or more variables that have only numeric values to be used in rows (columns) of the table. 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”.

**Group Numeric Data into Categories Using**
Choose the method by which numeric data will be combined into categories for use in table rows or columns. 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.
Descriptive Statistics – Summary Tables (Old Version)

**Range**
Integer ≥ 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 \( L1 < X \leq L2 \) 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.

**Grouping (Break) Variables and Frequency Variable**

**Number of Grouping Variables**
Select the number of grouping (break) variables. All reports and plots will be generated for each unique combination of the values of the grouping variables.

You can select up to 8 grouping variables.

**Grouping Variables**
Select an optional categorical grouping (or break) variable. All tables, statistical reports, and plots will be generated for each unique value of this variable.

If you specify more than one grouping variable, the tables, statistical reports, and plots generated for each unique combination of the values of the variables chosen.

**Frequency Variable**
This optional variable specifies the number of observations that each row represents. When omitted, each row represents a single observation. If your data is the result of previous summarization, you may want certain rows to represent several observations. Note that negative values are treated as a zero frequency and are omitted. Fractional values may be used. You may also think of this as a weighting variable.
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 Counts indicates that you want the number of missing values displayed.
Include in All indicates that you want the missing values treated just like any other category.

Missing Values
Specify individual missing values here.

Report Options Tab
The following options control the format of the reports.

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.

Show Total
Specify whether to show row and/or column total statistics for those reports that use a by (Table Row or Table Column) variable.

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.

Show Totals
Specify whether to show row and column total statistics for those reports and plots that use a categorical variable.

Use Short Statistical Names on Reports
Normally, the names of the statistical items in the reports are the complete names, such as Standard Deviation. Checking this option causes a shorter name, such as Std Dev, to be used so that more columns can be displayed on a single row.

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 of the statistics that may be displayed. Click the plot format button to change the plot settings.

Show Plots

Check this option to display plots corresponding to the reports that are displayed. These plots are only available when the Layout is set to 4, 5, or 8.

Example 1 – Layout 1: Variable Summary Report

The data used are found in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template Example 1 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.
   - On the menus, select Analysis, then Descriptive Statistics, then Descriptive Statistics - Summary Tables. The procedure will be displayed.
   - 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.
   - Set Layout to 1. TABLES: One, ROWS: Data, COLUMNS: Statistics.
   - Double-click in the Data Variables text box. This will bring up the variable selection window.
   - Select Price to LotSize from the list of variables and then click Ok. “Price-LotSize” will appear in the Data Variables box.

4 Specify the statistics.
   - In the Table Statistics section, check Count, Mean, Median, Std Dev, COV, and COD.

5 Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.
### Variable Summary Report

<table>
<thead>
<tr>
<th>Variables</th>
<th>Count</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Coef of Variation</th>
<th>Coef of Dispersion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>150</td>
<td>174392</td>
<td>158200</td>
<td>97656.81</td>
<td>0.55998</td>
<td>49.050</td>
</tr>
<tr>
<td>Year</td>
<td>150</td>
<td>1971.273</td>
<td>1973</td>
<td>13.84667</td>
<td>0.00702</td>
<td>0.572</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>150</td>
<td>2.42</td>
<td>2</td>
<td>0.8919476</td>
<td>0.36857</td>
<td>35.000</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>150</td>
<td>2.4</td>
<td>2.5</td>
<td>0.8047677</td>
<td>0.33532</td>
<td>24.800</td>
</tr>
<tr>
<td>Garage</td>
<td>150</td>
<td>1.266667</td>
<td>1</td>
<td>0.5636252</td>
<td>0.44497</td>
<td>36.000</td>
</tr>
<tr>
<td>Fireplace</td>
<td>150</td>
<td>0.96</td>
<td>1</td>
<td>0.6939818</td>
<td>0.72290</td>
<td>48.000</td>
</tr>
<tr>
<td>Quality</td>
<td>150</td>
<td>0.731666</td>
<td>0.75</td>
<td>0.35248</td>
<td>0.48175</td>
<td>39.778</td>
</tr>
<tr>
<td>Brick</td>
<td>150</td>
<td>0.503333</td>
<td>0.5</td>
<td>0.4157013</td>
<td>0.82590</td>
<td>68.667</td>
</tr>
<tr>
<td>TotalSqft</td>
<td>150</td>
<td>1893.38</td>
<td>1872.5</td>
<td>754.2496</td>
<td>0.39836</td>
<td>31.980</td>
</tr>
<tr>
<td>FinishSqft</td>
<td>150</td>
<td>1597.947</td>
<td>1496</td>
<td>672.1644</td>
<td>0.42064</td>
<td>35.156</td>
</tr>
<tr>
<td>LotSize</td>
<td>150</td>
<td>8366.913</td>
<td>8344.5</td>
<td>2376.334</td>
<td>0.28402</td>
<td>23.993</td>
</tr>
</tbody>
</table>

The definitions of these statistics are identical to those found in the Descriptive Statistics chapter. They will not be repeated here.

### Example 2 – Layout 3: Categorical Variable as Columns

The data used are found in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template **Example 2** 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.**
   - On the menus, select **Analysis**, then **Descriptive Statistics**, then **Descriptive Statistics - Summary Tables**. The procedure will be displayed.
   - 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**.
   - Double-click in the **Data Variables** text box. This will bring up the variable selection window.
   - Select **Price**, **TotalSqft**, and **LotSize** from the list of variables and then click **Ok**.
   - “Price,TotalSqft,LotSize” will appear in the Data Variables box.
   - Set **Layout to 3. Tables: One, Rows: Data, Columns: Categorical, Subrows: Statistics**.
   - In the **Table Statistics** section, check **Count**, **Mean**, and **Std Dev**.
   - Double-click in the **Column Variables** text box. This will bring up the variable selection window.
   - Select **State** from the list of variables and then click **Ok**. “State” will appear in the Column Variables box.
   - Set **Layout to 1. Tables: One, Rows: Data, Columns: Statistics**.

4. **Specify the report format.**
   - Click on the **Report Options tab**.
   - In Variable Names, select **Labels**.
   - In Value Labels, select **Value Labels**.

5. **Run the procedure.**
   - From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.
Statistical Summary

<table>
<thead>
<tr>
<th>Variables</th>
<th>State</th>
<th>Nevada</th>
<th>Virginia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>88</td>
<td>62</td>
<td>150</td>
</tr>
<tr>
<td>Sales Price</td>
<td>Count</td>
<td>170762.5</td>
<td>179543.5</td>
<td>174392</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>98665.72</td>
<td>96771.49</td>
<td>97656.81</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>974392</td>
<td>97656.81</td>
<td>97656.81</td>
</tr>
<tr>
<td>Total Area (Sqft)</td>
<td>Count</td>
<td>1881.33</td>
<td>1910.484</td>
<td>1893.38</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>788.569</td>
<td>708.6572</td>
<td>754.2496</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>788.569</td>
<td>708.6572</td>
<td>754.2496</td>
</tr>
<tr>
<td>Lot Size (Sqft)</td>
<td>Count</td>
<td>8571.454</td>
<td>8076.597</td>
<td>8366.913</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>2419.88</td>
<td>2301.226</td>
<td>2376.334</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>2419.88</td>
<td>2301.226</td>
<td>2376.334</td>
</tr>
</tbody>
</table>

The definitions of these statistics are identical to those found in the Descriptive Statistics chapter. They will not be repeated here.

Example 3 – Layout 5: Table of One Statistic

The data used are found in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template Example 3 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.
   - On the menus, select Analysis, then Descriptive Statistics, then Descriptive Statistics - Summary Tables. The procedure will be displayed.
   - 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.
   - Double-click in the Data Variables text box. This will bring up the variable selection window.
   - Select Bedrooms, Bathrooms, Garage, and Fireplace from the list of variables and then click Ok. “Bedrooms-Fireplace” will appear in the Data Variables box.
   - Set Layout to 5. TABLES: Statistics, ROWS: Data, COLUMNS: Categorical (Plots Possible)
   - In the Table Statistics section, check Mean.
   - Double-click in Column Variables text box. This will bring up the variable selection window.
   - Select State from the list of variables and then click Ok. “State” will appear in the Column Variables box.
   - Set Layout to 1. TABLES: One, ROWS: Data, COLUMNS: Statistics.

4. Specify the report format.
   - Click on the Report Options tab.
   - In Variable Names, select Labels.
   - In Value Labels, select Value Labels.
   - Set Show Total to On Reports and Plots

5. Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.
Table and Plot of One Statistic

### Table of Means

<table>
<thead>
<tr>
<th>Variables</th>
<th>Nevada</th>
<th>Virginia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bedrooms</td>
<td>2.352273</td>
<td>2.516129</td>
<td>2.42</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>2.409091</td>
<td>2.387097</td>
<td>2.4</td>
</tr>
<tr>
<td>Garage</td>
<td>1.261364</td>
<td>1.274194</td>
<td>1.266667</td>
</tr>
<tr>
<td>Fireplace</td>
<td>1.022727</td>
<td>0.8709677</td>
<td>0.96</td>
</tr>
</tbody>
</table>

### Plot of Means

The definitions of these statistics are identical to those found in the Descriptive Statistics chapter. They will not be repeated here.
Example 4 – Layout 6: Multiple Y’s, Two Categoricals, One Statistic

The data used are found in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template Example 4 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.
   - On the menus, select Analysis, then Descriptive Statistics, then Descriptive Statistics - Summary Tables. The procedure will be displayed.
   - 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.
   - Double-click in the Data Variables text box. This will bring up the variable selection window.
   - Select Price, FinishSqft, and LotSize from the list of variables and then click Ok. “Price,FinishSqft-LotSize” will appear in the Data Variables box.
   - Check the Create Other Row Variables from Numeric Data box.
   - Double-click in the Numeric Variables text box. This will bring up the variable selection window.
   - Select TotalSqft from the list of variables and then click Ok. “TotalSqft” will appear in the Numeric Variables box.
   - Set the Group Numeric Data into Categories Using option to List of Interval Upper Limits.
   - Set the List box to 1000 2000 3000.
   - Double-click in Column Variables text box. This will bring up the variable selection window.
   - Select State from the list of variables and then click Ok. “State” will appear in the Column Variables box.

4 Specify the report format.
   - Click on the Report Options tab.
   - In Variable Names, select Labels.
   - In Value Labels, select Value Labels.
   - Set Show Total to On Reports and Plots. Note that no plots are displayed with this table.
   - Set the Rows Decimal Places to Auto (Up to 7).

5 Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.
Multiple Y’s, Two Categoricals, One Statistic

<table>
<thead>
<tr>
<th>Total Area (Sqft)</th>
<th>State</th>
<th>Nevada</th>
<th>Virginia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up To 1000</td>
<td>Sales Price</td>
<td>160475</td>
<td>142850</td>
<td>152921.4</td>
</tr>
<tr>
<td></td>
<td>Finished Area (Sqft)</td>
<td>738.125</td>
<td>739.6667</td>
<td>738.7857</td>
</tr>
<tr>
<td></td>
<td>Lot Size (Sqft)</td>
<td>8816</td>
<td>9857.833</td>
<td>9262.5</td>
</tr>
<tr>
<td>1000 To 2000</td>
<td>Sales Price</td>
<td>153293.3</td>
<td>172992.9</td>
<td>160849.3</td>
</tr>
<tr>
<td></td>
<td>Finished Area (Sqft)</td>
<td>1234.311</td>
<td>1247.179</td>
<td>1239.247</td>
</tr>
<tr>
<td></td>
<td>Lot Size (Sqft)</td>
<td>9094.8</td>
<td>7674.286</td>
<td>8549.945</td>
</tr>
<tr>
<td>2000 To 3000</td>
<td>Sales Price</td>
<td>197200</td>
<td>186461.5</td>
<td>192029.6</td>
</tr>
<tr>
<td></td>
<td>Finished Area (Sqft)</td>
<td>1974.214</td>
<td>2086.077</td>
<td>2028.074</td>
</tr>
<tr>
<td></td>
<td>Lot Size (Sqft)</td>
<td>7503.179</td>
<td>8129.808</td>
<td>7804.889</td>
</tr>
<tr>
<td>Over 3000</td>
<td>Sales Price</td>
<td>189071.4</td>
<td>291400</td>
<td>211811.1</td>
</tr>
<tr>
<td></td>
<td>Finished Area (Sqft)</td>
<td>3375.143</td>
<td>2871</td>
<td>3263.111</td>
</tr>
<tr>
<td></td>
<td>Lot Size (Sqft)</td>
<td>9200.714</td>
<td>7673.5</td>
<td>8861.333</td>
</tr>
<tr>
<td>Total</td>
<td>Sales Price</td>
<td>170762.5</td>
<td>179543.5</td>
<td>174392</td>
</tr>
<tr>
<td></td>
<td>Finished Area (Sqft)</td>
<td>1594.92</td>
<td>1602.242</td>
<td>1597.947</td>
</tr>
<tr>
<td></td>
<td>Lot Size (Sqft)</td>
<td>8571.454</td>
<td>8076.597</td>
<td>8366.913</td>
</tr>
</tbody>
</table>

The definitions of these statistics are identical to those found in the Descriptive Statistics chapter. They will not be repeated here.
Example 5 – Layout 7: Complete Summary for each Data Variable

The data used are found in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template Example 5 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.
   • On the menus, select Analysis, then Descriptive Statistics, then Descriptive Statistics - Summary Tables. The procedure will be displayed.
   • 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.
   • Double-click in the Data Variables text box. This will bring up the variable selection window.
   • Select Price from the list of variables and then click Ok. “Price” will appear in the Data Variables box.
   • Set Layout to 7. TABLES: Data, ROWS: Categorical, COLUMNS: Categorical, SUBROWS: Statistics.
   • Check the Create Other Row Variables from Numeric Data box.
   • Double-click in the Numeric Variables text box. This will bring up the variable selection window.
   • Select TotalSqft from the list of variables and then click Ok. “TotalSqft” will appear in the Numeric Variables box.
   • Set the Group Numeric Data into Categories Using option to List of Interval Upper Limits.
   • Set the List box to 1000 2000 3000.
   • Double-click in Column Variables text box. This will bring up the variable selection window.
   • Select State from the list of variables and then click Ok. “State” will appear in the Column Variables box.

4 Specify the report format.
   • Click on the Report Options tab.
   • In Variable Names, select Labels.
   • In Value Labels, select Value Labels.
   • Set Show Total to On Reports and Plots. Note that no plots are displayed with this table.
   • Set the Row Decimal Places to Auto (Up to 7).

5 Run the procedure.
   • From the Run menu, select Run Procedure. Alternatively, just click the green Run button.
## Complete Summary for Each Data Variable

### Statistical Summary of Sales Price

<table>
<thead>
<tr>
<th>Total Area (Sqft)</th>
<th>State</th>
<th>Nevada</th>
<th>Virginia</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>8</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Up To 1000</td>
<td>Mean</td>
<td>160475</td>
<td>142850</td>
<td>152921.4</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>136050</td>
<td>85200</td>
<td>110500</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>110945.7</td>
<td>107838.2</td>
<td>105747.5</td>
</tr>
<tr>
<td>1000 To 2000</td>
<td>Count</td>
<td>45</td>
<td>28</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>153293.3</td>
<td>172992.9</td>
<td>160849.3</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>123400</td>
<td>163000</td>
<td>150100</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>91336.91</td>
<td>71798.73</td>
<td>84405.74</td>
</tr>
<tr>
<td>2000 To 3000</td>
<td>Count</td>
<td>28</td>
<td>26</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>197200</td>
<td>186461.5</td>
<td>192029.6</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>182850</td>
<td>145550</td>
<td>176250</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>106136.7</td>
<td>111024.2</td>
<td>107621.7</td>
</tr>
<tr>
<td>Over 3000</td>
<td>Count</td>
<td>7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>189071.4</td>
<td>291400</td>
<td>211811.1</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>150900</td>
<td>291400</td>
<td>168500</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>94037.06</td>
<td>173806.8</td>
<td>111554.4</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>88</td>
<td>62</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>170762.5</td>
<td>179543.5</td>
<td>174392</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>151050</td>
<td>162800</td>
<td>158200</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>98665.72</td>
<td>96771.49</td>
<td>97656.81</td>
</tr>
</tbody>
</table>

The definitions of these statistics are identical to those found in the Descriptive Statistics chapter. They will not be repeated here.
Example 6 – Layout 8: One Data Variable and Statistic, Two Categories

The data used are found in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template Example 6 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.
   - On the menus, select Analysis, then Descriptive Statistics, then Descriptive Statistics - Summary Tables. The procedure will be displayed.
   - 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.
   - Double-click in the Data Variables text box. This will bring up the variable selection window.
   - Select Price from the list of variables and then click Ok. “Price” will appear in the Data Variables box.
   - Set Layout to 8. TABLES: Data and Statistics, ROWS: Categorical, COLUMNS: Categorical (Plots Possible).
   - Double-click in Row Variables text box. This will bring up the variable selection window.
   - Select State from the list of variables and then click Ok. “State” will appear in the Row Variables box.
   - Check the Create Other Column Variables from Numeric Data box.
   - Double-click in the Numeric Variables text box. This will bring up the variable selection window.
   - Select TotalSqft from the list of variables and then click Ok. “TotalSqft” will appear in the Numeric Variables box.
   - Set the Group Numeric Data into Categories Using option to List of Interval Upper Limits.
   - Set the List box to 1000 2000 3000.

4 Change the report format.
   - Click on the Report Options tab.
   - In Variable Names, select Labels.
   - In Value Labels, select Value Labels.
   - Set Show Total to On Reports and Plots. Note that no plots are displayed with this table.
   - Set the Row Decimal Places to Auto (Up to 7).
   - Set the Mean, Sum Decimal Places to 0.

5 Change the plot format.
   - Click on the Plots tab.
   - Check Interactive box in the upper, right-hand corner of the Plot Format button.

6 Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.
   - When the Bar Chart Format window is displayed, click the Group Axis tab.
   - In the Ticks section of the window, click the Layout icon of the Lower Axis Labels. This button is the second icon to the right.
   - Set the Rotation Angle to 39. Click OK twice.
# One Data Variable and Statistic, Two Categories

## Means of Sales Price

<table>
<thead>
<tr>
<th>State</th>
<th>Up To 1000</th>
<th>1000 To 2000</th>
<th>2000 To 3000</th>
<th>Over 3000</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nevada</td>
<td>160475</td>
<td>153293</td>
<td>197200</td>
<td>189071</td>
<td>170763</td>
</tr>
<tr>
<td>Virginia</td>
<td>142850</td>
<td>172993</td>
<td>186462</td>
<td>291400</td>
<td>179544</td>
</tr>
<tr>
<td>Total</td>
<td>152921</td>
<td>160849</td>
<td>192030</td>
<td>211811</td>
<td>174392</td>
</tr>
</tbody>
</table>

## Plot of Means

![Plot of Means](image)
Example 7 – List Format

The data used are found in the Resale dataset. You may follow along here by making the appropriate entries or load the completed template Example 7 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.
   •  On the menus, select Analysis, then Descriptive Statistics, then Descriptive Statistics - Summary Tables. The Descriptive Statistics - Summary Tables procedure will be displayed.
   •  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.
   •  In Table Layout, select 9. DATA SUMMARY LIST.
   •  Double-click in the Data Variables text box. This will bring up the variable selection window.
   •  Select Price from the list of variables and then click Ok. “Price” will appear in the Data Variables box.
   •  Double-click in the Row Variables text box. This will bring up the variable selection window.
   •  Select Neighborhood from the list of variables and then click Ok. “Neighborhood” will appear in the Row Variables box.
   •  Set the Number of Grouping Variables to 2.
   •  Double-click in the Grouping Variable 1 text box. This will bring up the variable selection window.
   •  Select State from the list of variables and then click Ok.
   •  Double-click in the Grouping Variable 2 text box. This will bring up the variable selection window.
   •  Select City from the list of variables and then click Ok.

4  Click on the Report Options tab.
   •  In Variable Names, select Names.
   •  In Value Labels, select Data Values.

5  Run the procedure.
   •  From the Run menu, select Run Procedure. Alternatively, just click the green Run button.
List Format Report

The definitions of these statistics are identical to those found in the Descriptive Statistics chapter. They will not be repeated here.

This format is especially useful for creating a database containing only summary information such as the means, standard deviations, etc. To create a summary database, take the following steps:

1. Run this report on the data, summarizing across the categorical variables of interest.
2. Copy the output report to the clipboard.
3. Open a new database (or spreadsheet).
4. Paste the data from the clipboard to this new database by placing the cursor in the upper-left cell and pasting. The paste can use the Ctrl-V key or Paste from the Edit menu.
5. Label the columns in the Variable Info sheet.

<table>
<thead>
<tr>
<th>State</th>
<th>City</th>
<th>Neighborhood</th>
<th>Price Count</th>
<th>Price Mean</th>
<th>Price StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nev</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>203727.3</td>
<td>105805.4</td>
</tr>
<tr>
<td>Nev</td>
<td>1</td>
<td>2</td>
<td>16</td>
<td>183625</td>
<td>105754.7</td>
</tr>
<tr>
<td>Nev</td>
<td>2</td>
<td>3</td>
<td>16</td>
<td>135018.8</td>
<td>94628.04</td>
</tr>
<tr>
<td>Nev</td>
<td>2</td>
<td>4</td>
<td>13</td>
<td>156192.3</td>
<td>93304.72</td>
</tr>
<tr>
<td>Nev</td>
<td>2</td>
<td>5</td>
<td>20</td>
<td>192190</td>
<td>100400.5</td>
</tr>
<tr>
<td>Nev</td>
<td>3</td>
<td>6</td>
<td>12</td>
<td>151125</td>
<td>88063.07</td>
</tr>
<tr>
<td>Vir</td>
<td>4</td>
<td>7</td>
<td>13</td>
<td>197307.7</td>
<td>80288.13</td>
</tr>
<tr>
<td>Vir</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>168700</td>
<td>86626.27</td>
</tr>
<tr>
<td>Vir</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>178716.7</td>
<td>107857.3</td>
</tr>
<tr>
<td>Vir</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>159511.1</td>
<td>132957.2</td>
</tr>
<tr>
<td>Vir</td>
<td>5</td>
<td>11</td>
<td>9</td>
<td>150488.9</td>
<td>70977.03</td>
</tr>
<tr>
<td>Vir</td>
<td>6</td>
<td>12</td>
<td>11</td>
<td>212963.6</td>
<td>112784.7</td>
</tr>
</tbody>
</table>