Tabulation

The Tabulation procedure is designed to summarize a single column of attribute data. It tabulates the frequency of occurrence of each unique value within that column. The frequencies are displayed both in tabular form and graphically as a barchart or piechart.

Sample StatFolio: tabulation.sgp

Sample Data:
The file 93cars.sgd contains information on 26 variables for \( n = 93 \) makes and models of automobiles, taken from Lock (1993). The table below shows a partial list of 3 columns from that file:

<table>
<thead>
<tr>
<th>Make</th>
<th>Model</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acura</td>
<td>Integra</td>
<td>Small</td>
</tr>
<tr>
<td>Acura</td>
<td>Legend</td>
<td>Midsize</td>
</tr>
<tr>
<td>Audi</td>
<td>90</td>
<td>Compact</td>
</tr>
<tr>
<td>Audi</td>
<td>100</td>
<td>Midsize</td>
</tr>
<tr>
<td>BMW</td>
<td>535i</td>
<td>Midsize</td>
</tr>
<tr>
<td>Buick</td>
<td>Century</td>
<td>Midsize</td>
</tr>
<tr>
<td>Buick</td>
<td>LeSabre</td>
<td>Large</td>
</tr>
<tr>
<td>Buick</td>
<td>Roadmaster</td>
<td>Large</td>
</tr>
<tr>
<td>Buick</td>
<td>Riviera</td>
<td>Midsize</td>
</tr>
<tr>
<td>Cadillac</td>
<td>DeVille</td>
<td>Large</td>
</tr>
<tr>
<td>Cadillac</td>
<td>Seville</td>
<td>Midsize</td>
</tr>
<tr>
<td>Chevrolet</td>
<td>Cavalier</td>
<td>Compact</td>
</tr>
</tbody>
</table>

Data Input
The data input dialog box specifies the column containing the data to be tabulated.

- **Data**: numeric or non-numeric column with the data to be tabulated.
Select: subset selection.

Analysis Summary
The Analysis Summary shows the number of rows with non-missing data and the number of unique values in the column.

Tabulation - Type
Data variable: Type
Number of observations: 93
Number of unique values: 6

Frequency Table
This pane presents the results in tabular form.

<table>
<thead>
<tr>
<th>Class</th>
<th>Value</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Cumulative Frequency</th>
<th>Cum. Rel. Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Compact</td>
<td>16</td>
<td>0.1720</td>
<td>16</td>
<td>0.1720</td>
</tr>
<tr>
<td>2</td>
<td>Large</td>
<td>11</td>
<td>0.1183</td>
<td>27</td>
<td>0.2903</td>
</tr>
<tr>
<td>3</td>
<td>Midsize</td>
<td>22</td>
<td>0.2366</td>
<td>49</td>
<td>0.5269</td>
</tr>
<tr>
<td>4</td>
<td>Small</td>
<td>21</td>
<td>0.2258</td>
<td>70</td>
<td>0.7527</td>
</tr>
<tr>
<td>5</td>
<td>Sporty</td>
<td>14</td>
<td>0.1505</td>
<td>84</td>
<td>0.9032</td>
</tr>
<tr>
<td>6</td>
<td>Van</td>
<td>9</td>
<td>0.0968</td>
<td>93</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

This table displays the following information for each unique value or “class”:

- **Value** - the value of the data variable.
- **Frequency** - the frequency of occurrence $f_j$ of each value in the data column.
- **Relative Frequency** - the proportion of times each value occurred, given by $f_j/n$.
- **Cumulative Frequency** - the number of observations lying in the current or previous classes:

$$\sum_{i=1}^{j} f_i$$  \hspace{1cm} (1)

- **Cumulative Relative Frequency** - the proportion of observations lying in the current or previous classes:

$$\frac{\sum_{i=1}^{j} f_i}{n}$$  \hspace{1cm} (2)
**Barchart**

The *Barchart* plots the frequency associated with each unique value as a vertical or horizontal bar.

![Barchart Image]

**Pane Options**

- **Chart Type**: The bars may be clustered side by side as shown in the example or stacked one upon the other.

- **Scaling**: whether the axis scale shows the frequencies $f_j$ or the percentages given by

  $$ p_j = 100 \frac{f_j}{n} \% $$  

(3)

- **Direction**: whether the bars extend horizontally or vertically.

© 2009 by StatPoint Technologies, Inc.
- **Baseline**: the value from which the bars extend.

**Example: Vertical Barchart with Non-Zero Baseline**

The following chart plots vertical bars extending from a baseline of 12.

**Piechart**

The *Piechart* plots each unique value using slices of a pie to represent their relative magnitude.
Pane Options

Piechart Options

- **Legends**: the desired information to be placed in the legend block to the right of the pie.
- **Labels**: the desired information to be placed next to each slice of the pie.
- **Diameter**: the size of the pie relative to the size of the plotting area. Reducing this value makes the pie smaller.
- **Offset #**: number of a slice between 1 and \( k \) to offset from the rest of the pie.
- **Lines**: whether to extend lines from each slice to its label.

Example: Piechart with Offset Slice

![Piechart for Type](image)
Save Results
The following results can be saved to the datasheet:

1. *Class Frequencies* – the frequencies of occurrence of each unique value.

2. *Class Labels* – the labels associated with each unique value.