

Describe – Attributes and Counts

This procedure is designed to summarize data contained in either an *Attribute* column or a *Count* column. It creates barcharts and Pareto charts to summarize the frequency of occurrence of data that can be divided into categories. In the case of an *Attribute* column, the procedure finds all unique values in the column and determines how often each value occurs. In the case of a *Count* column, it simply displays the counts graphically.

The data displayed by this analysis consist of counts for m categories, where:

$$c_j = \text{frequency of occurrence of category } j, j = 1, 2, \dots, m$$

The charts created are:

- **Barchart** – plots the category frequencies as vertical bars in the order encountered in the data.
- **Pareto Chart** – plots the percentage of the data represented by each category, sorted from most frequent to least frequent. A cumulative line is drawn above the bars.

Access

Highlight: one *Count* column or one *Attribute* column. A *Labels* column may also be selected to supply labels for the categories. If not supplied, the labels will be automatically generated from the data.

Select: *Describe* from the main menu.

Output Page 1: A barchart.

Output Page 2: A Pareto chart.

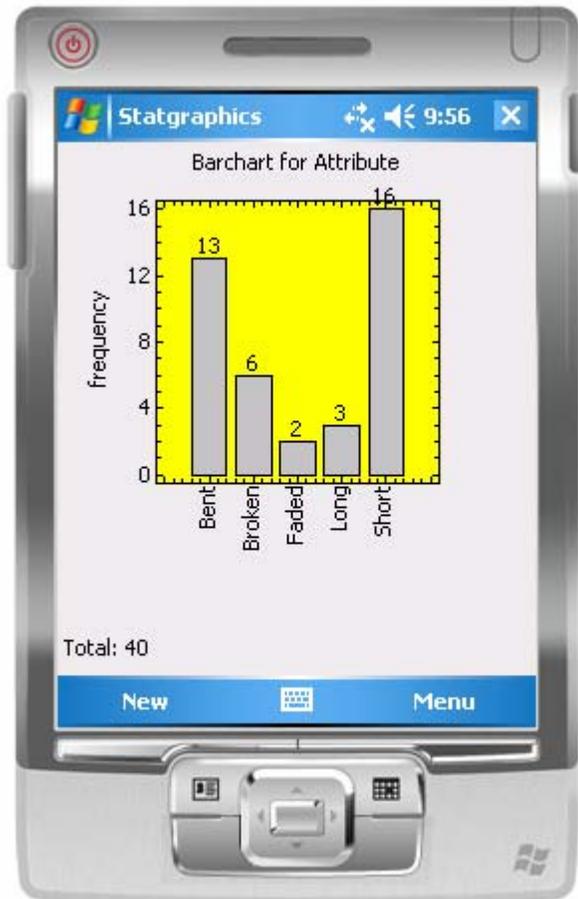
Sample Data

The file *defects.sgm* contains information about $n = 40$ items, each of which was found to be defective. A portion of the data is shown below:

<i>Row</i>	<i>Defect</i>
1	Short
2	Bent
3	Short
4	Broken
5	Bent
6	Short
7	Long
8	Faded

Barchart

If an *Attribute* column is selected, the procedure begins by finding each unique value in the column. For each of the m unique values found, it then determines the frequency of occurrence. The frequencies are displayed as a vertical barchart, with the height of each bar being proportional to the observed frequency. The frequencies c_j are indicated above each bar.



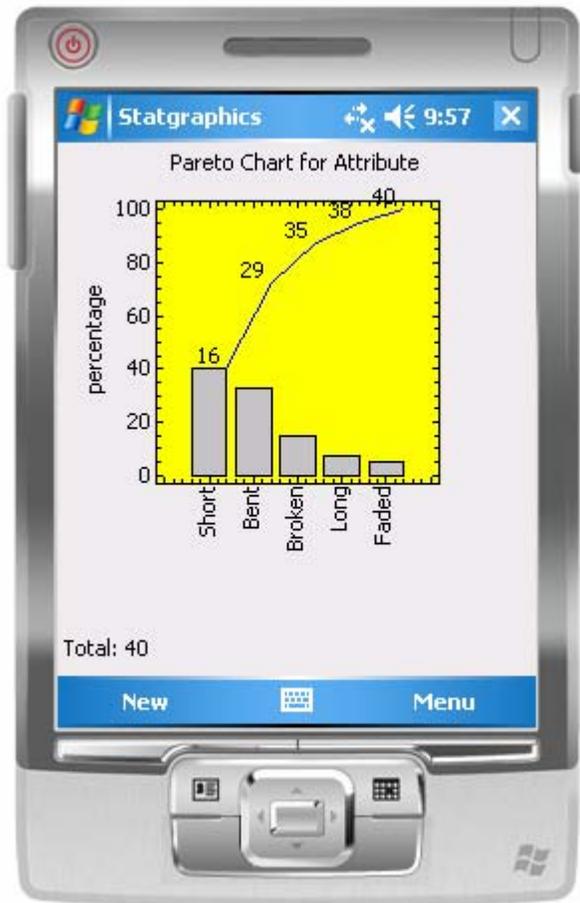
If a *Count* column is selected, the column is assumed to contain the frequencies c_j , which are simply displayed.

The labels for each column are generated automatically, unless a *Labels* variable is also selected. In that case, the values in the *Labels* column are used to label each bar.

The sum of the frequencies is also displayed.

Pareto Chart

When the frequencies represent different types of defects, the data are often displayed as a Pareto chart. In such a chart, the frequencies are sorted from most frequent to least frequent. A line is drawn above the bars representing the percentage of the total represented by each bar and any bars to the left.



Displayed above each bar is the cumulative frequency

$$S_j = \sum_{i=1}^j c_i \quad (1)$$