Chapter 26
VBAR Statement

Chapter Table of Contents

Chapter 26
VBAR Statement

Overview

The VBAR statement creates a Pareto chart with vertical bars representing the frequencies of problems in a process or operation. A vertical Pareto chart has one horizontal axis on which the Pareto categories are listed. The primary vertical axis appears on the left side of the chart and is used to read the lengths of the bars on the chart. The secondary vertical axis is on the right of the chart and is used to read the cumulative percent curve.

Getting Started

The examples in this section illustrate basic features of the VBAR statement. Complete syntax for the VBAR statement is presented in the "Syntax" section on page 807.

Creating a Pareto Chart from Raw Data

See PARETO4 in the SAS/QC Sample Library

In the fabrication of integrated circuits, common causes of failures include improper doping, corrosion, surface contamination, silicon defects, metallization, and oxide defects. The causes of 31 failures were recorded in a SAS data set called FAILURE1.

```
data failure1;
   length cause $ 16;
   label  cause = 'Cause of Failure';
   input  cause;
   datalines;
Corrosion
Oxide Defect
Contamination
Oxide Defect
Oxide Defect
Miscellaneous
Oxide Defect
Contamination
Metallization
Oxide Defect
Contamination
Contamination
Oxide Defect
Contamination
Contamination
Contamination
Corrosion
Silicon Defect
Miscellaneous
Contamination
Contamination
Contamination
Miscellaneous
Contamination
Contamination
Doping
Oxide Defect
Oxide Defect
Metallization
Contamination
Contamination
;
```

Each of the 31 observations corresponds to a different circuit, and the value of CAUSE provides the cause for the failure. These are raw data in the sense that there is more than one observation with the same value of CAUSE, and the observations are not sorted by CAUSE. The following statements produce a basic Pareto chart for the failures:

```
proc pareto data=failure1;
   vbar cause;
run;
```

The PARETO procedure is invoked with the first statement, referred to as the PROC statement. You specify the process variable to be analyzed in the VBAR statement.

The Pareto chart is shown in Figure 26.1.

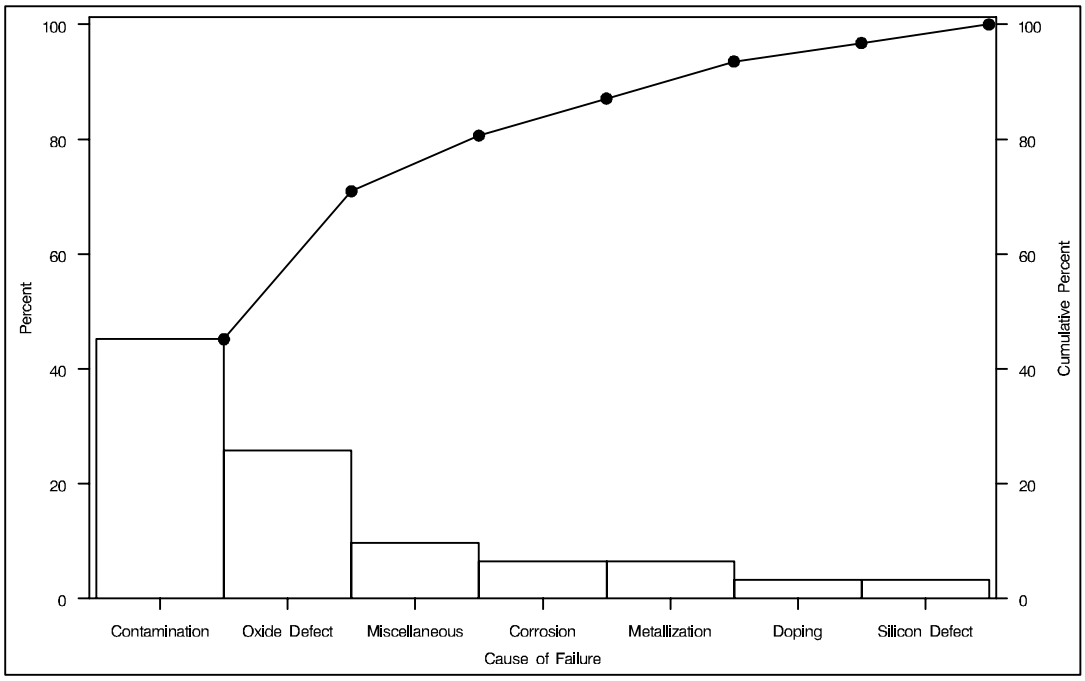

Figure 26.1. Pareto Chart for IC Failures in the Data Set FAILURE1

The procedure has classified the values of CAUSE into seven distinct categories (levels). The bars represent the percent of failures in each category, and they are arranged in decreasing order. Thus, the most frequently occurring category is *Contamination*, which accounts for 45% of the failures. The Pareto curve indicates the cumulative percent of failures from left to right; for example, *Contamination* and *Oxide* together account for 71% of the failures.

If there is sufficient space, the procedure labels the bars along the horizontal axis as in Figure 26.1. Otherwise, as in Figure 26.2, the procedure numbers the bars from left to right and adds a legend identifying the categories.

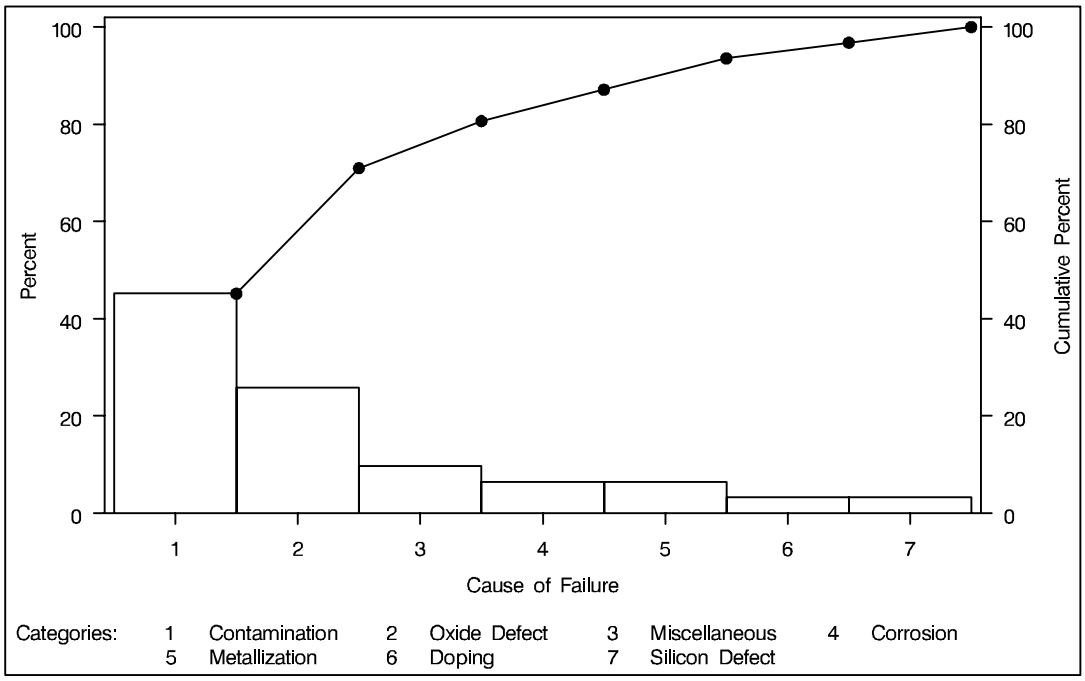

Figure 26.2. Pareto Chart with Category Legend

A category legend is likely to be introduced when

- the number of categories is large
- the category labels are lengthy (as in this example). Category labels can be up to 32 characters.
- a large text height is used. You can specify the height with the HEIGHT= option in the VBAR statement or with the HTEXT= option in a GOPTIONS statement (not shown here).

Creating a Pareto Chart Using Frequency Data

See PARETO5 in the SAS/QC Sample Library

In some situations, a count (frequency) is available for each category, or you can compress a large data set by creating a frequency variable for the categories before applying the PARETO procedure.

For example, you can use the FREQ procedure to obtain the compressed data set FAILURE2 from the data set FAILURE1.

```
proc freq data=failure1;
   tables cause / noprint out=failure2;

proc print;
run;
```

A listing of FAILURE2 is shown in Figure 26.3.

Obs	cause	COUNT	PERCENT
1	Contamination	14	45.1613
2	Corrosion	2	6.4516
3	Doping	1	3.2258
4	Metallization	2	6.4516
5	Miscellaneous	3	9.6774
6	Oxide Defect	8	25.8065
7	Silicon Defect	1	3.2258

Figure 26.3. The Data Set FAILURE2 Created Using PROC FREQ

The following statements produce a Pareto chart for the data in FAILURE2:

```
title 'Analysis of IC Failures';
symbol value=dot;
proc pareto data=failure2;
   vbar cause / freq       = count
                scale      = count
                interbar   = 1.0
                last       = 'Miscellaneous'
                nlegend    = 'Total Circuits'
                cframenleg = empty ;
run;
```

The chart is displayed in Figure 26.4.

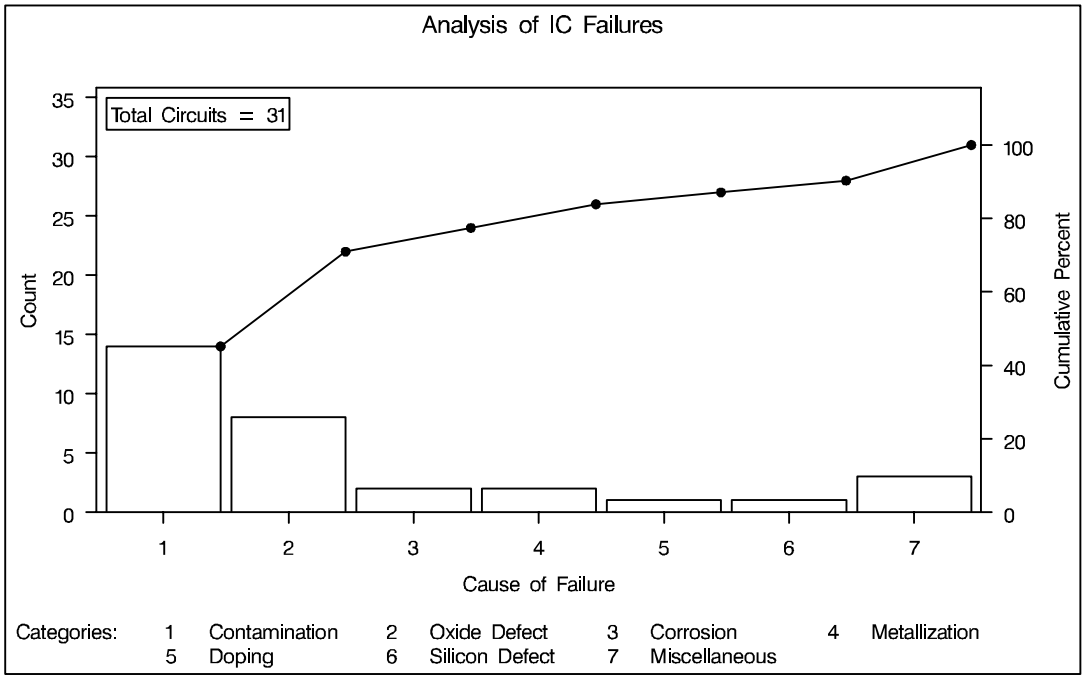

Figure 26.4. Pareto Chart with Frequency Scale

A slash (/) is used to separate the process variable CAUSE from the options specified in the VBAR statement. The frequency variable COUNT is specified with the FREQ= option. Specifying the keyword COUNT with the SCALE= option requests a frequency scale for the vertical axis.

The INTERBAR= option inserts a small space between the bars, and specifying LAST='Miscellaneous' causes the category *Miscellaneous* to be displayed last regardless of its frequency. The NLEGEND= option adds a sample size legend labeled *Total Circuits*, and the CFRAMENLEG= option frames the legend. The SYMBOL statement marks points on the curve with dots.

In the preceding statements, adding the keyword LINEPRINTER to the PROC PARETO statement requests a line printer version of the chart, which is displayed in Figure 26.5. *

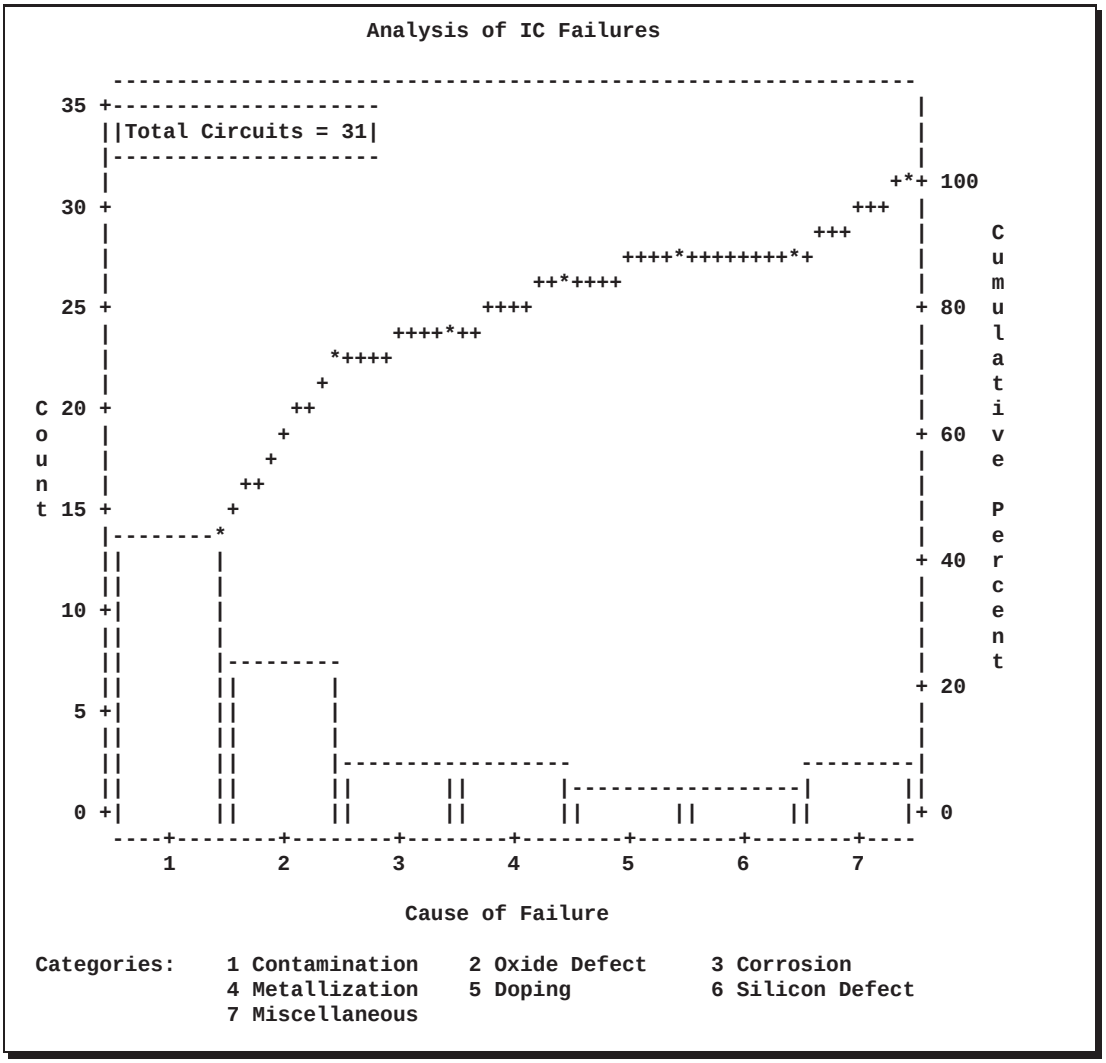

Figure 26.5. Pareto Chart for IC Failures in the Data Set FAILURE2

*In Release 6.12 and previous releases of SAS/QC software, the keyword GRAPHICS was required in the PROC PARETO statement to specify that the chart be created with a graphics device. In Version 7 and later releases, you can specify the LINEPRINTER option to request line printer plots.

There are two sets of tied categories in this example; *Corrosion* and *Metallization* each occur twice, and *Doping* and *Silicon Defect* each occur once. The procedure displays tied categories in alphabetical order of their formatted values. Thus, *Corrosion* appears before *Metallization*, and *Doping* appears before *Silicon Defect* in Figure 26.4 and Figure 26.5. This is simply a convention, and no practical significance should be attached to the order in which tied categories are arranged.

Restricting the Number of Pareto Categories

See PARETO2 in the SAS/QC Sample Library

Unlike the previous examples, some applications involve too many categories to display on a chart. The solution presented here is to create a restricted Pareto chart that displays only the most frequently occurring categories.

The following statements create a Pareto chart for the five most frequently occurring levels of CAUSE in the data set FAILURE2, which is listed in Figure 26.3:

```
title 'IC Manufacturing Problems' ;
symbol value=dot;
proc pareto data=failure2;
   vbar cause / freq       = count
                scale      = count
                maxncat    = 5;
run;
```

The MAXNCAT= option specifies the number of categories to be displayed. The chart, shown in Figure 26.6, does not display the categories *Doping* and *Silicon Defect*.

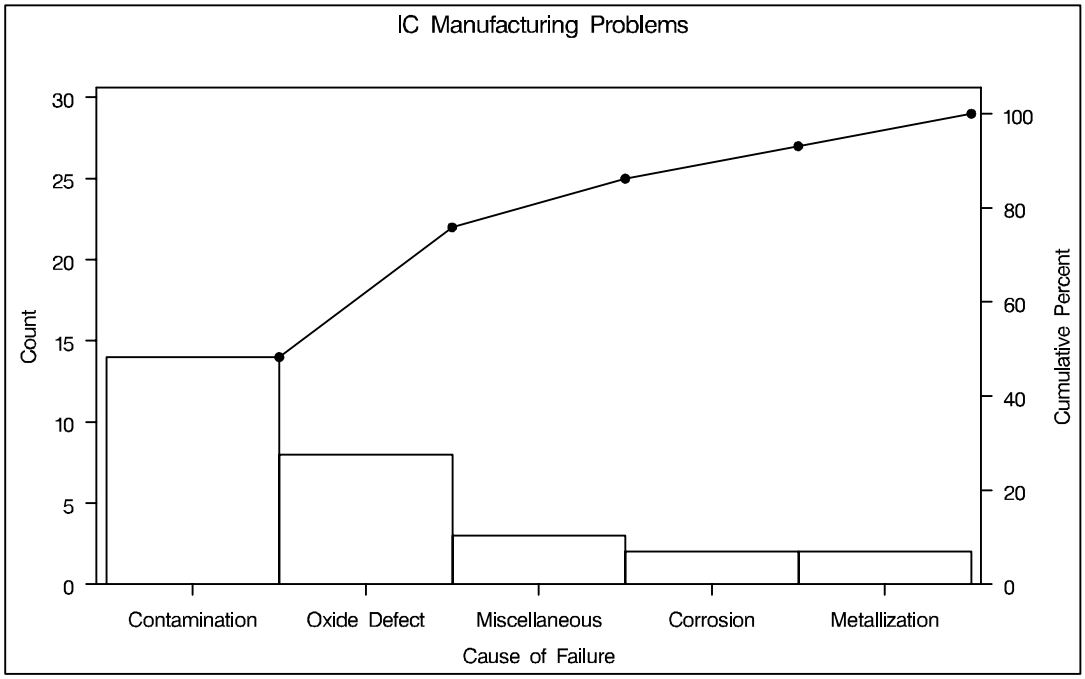

Figure 26.6. Restricted Pareto Chart

You can also display the most frequently occurring categories and merge the remaining categories into a single *other* category that is represented by a bar. You can specify the name for the new category with the OTHER= option. If, in addition, you specify the name with the LAST= option, the category is positioned at the far right of the chart. The following statements illustrate both options:

```
title 'IC Manufacturing Problems' ;
symbol value=dot;
proc pareto data=failure2;
   vbar cause / freq       = count
                scale      = count
                maxncat    = 5
                other      = 'Others'
                last       = 'Others';
run;
```

The chart is shown in Figure 26.7.

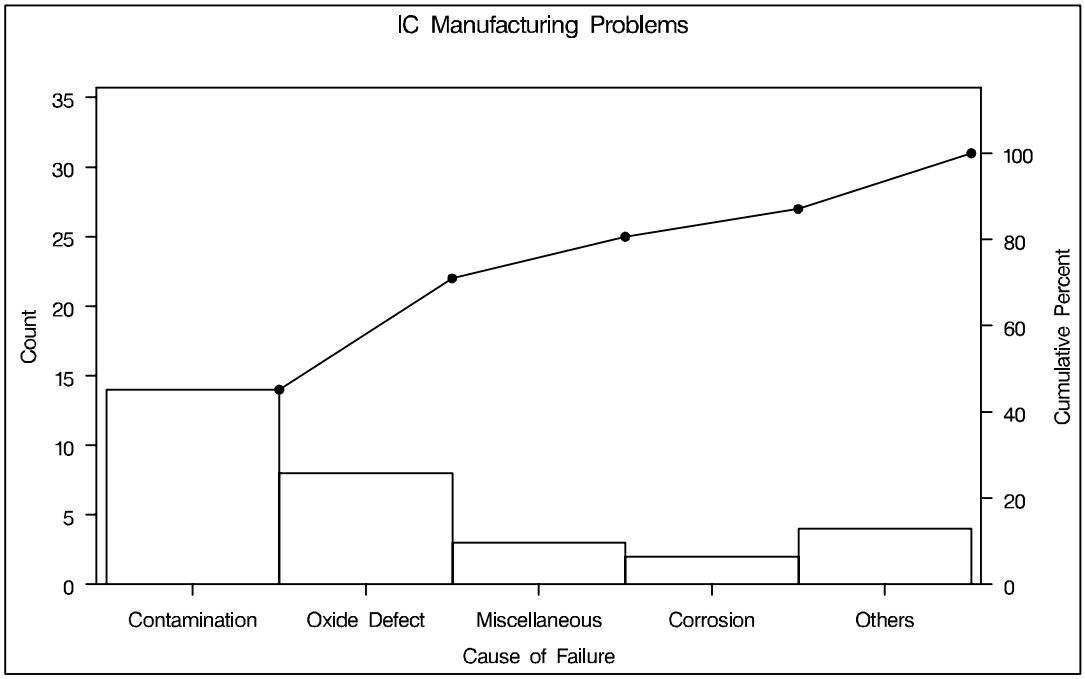

Figure 26.7. Restricted Pareto Chart with *Other* Category

The number of categories displayed is five, which is the number specified with the MAXNCAT= option. The first four categories are the four most frequently occurring problems in FAILURE2, and the fifth category merges the remaining problems.

Note that *Corrosion* and *Metallization* both have a frequency of two. When the MAXNCAT= option is applied to categories with tied frequencies, the procedure breaks the tie by using the order of the formatted values. Thus *Corrosion* is displayed, whereas *Metallization* is merged into the *Other* category. The MAXNCAT= and related options are described in "Restricted Pareto Charts" on page 888.

Syntax

The syntax for the VBAR statement is as follows:

VBAR (*variable-list***) < /** *options* **> ;**

You can use any number of VBAR statements in the PARETO procedure. If you specify two or more *variables* in the VBAR statement, they must be enclosed in parentheses. The components of the VBAR statement are described as follows.

options
specify the layout and features of the chart, and they are listed after a slash (/) that follows the variables to be analyzed.

The "Summary of Options" section, which follows, provides summary tables of options organized by function. The "Dictionary of Options" on page 813 describes the *options* in detail.

variable-list
specifies the process variables to be analyzed. A chart is created for each *variable*, and the values of each *variable* determine the Pareto categories for that chart. A list of two or more *variables* must be enclosed in parentheses.

The *variables* can be numeric or character, and the maximum length of a character *variable* is 32. Formatted values are used to determine the categories and are displayed in labels and legends. The maximum format length is 32.

Summary of Options

The following tables list the VBAR statement options by function. For complete descriptions, see "Dictionary of Options" on page 813.

Table 26.1. Options for the Cumulative Percent Curve

ANCHOR=*keyword*	specifies corner of leftmost bar to which curve is anchored
CCONNECT=*color*	specifies color for curve
CMPCTLABEL	labels curve points with their values
CONNECTCHAR= *'character'*	specifies plot character for curve segments
NOCURVE	suppresses curve
NOVLABEL2	suppresses secondary vertical axis label
NOVTICK2	suppresses secondary vertical axis tick marks and tick mark labels
SYMBOLCHAR= *'character'*	specifies plot character for points on curve

Table 26.2. Data Processing Options

CFRAMENLEG=*color*	frames the NLEGEND legend and fills the frame with the specified color
FREQ=*variable*	specifies frequency variable
MISSING	specifies that missing values of the process variable be treated as a Pareto category
MISSING1	specifies that missing values of the first CLASS= variable be analyzed as a level
MISSING2	specifies that missing values of the second CLASS= variable be analyzed as a level
NLEGEND	requests sample size legend
NLEGEND='*label*'| (*variable*)	requests sample size legend with specified label
OUT=*SAS-data-set*	creates output data set that saves information displayed in the Pareto chart
WEIGHT=*variable-list*	specifies weight variables used to weight frequencies

Table 26.3. Options for Restricting the Number of Categories

COTHER=*color*	specifies color for OTHER= bar
LOTHER='*label*'	specifies label for OTHER= bar
MAXCMPCT=*percent*	displays only the categories with cumulative percentage less than the *percent* specified
MAXNCAT=n	displays only the categories with the n highest values
MINPCT=*percent*	displays only the categories with percents greater than the *percent* specified
OTHER='*category*'	merges all categories not displayed
OTHERCVAL='*value*'	specifies an OUT= data set character variable value for the OTHER= category
OTHERNVAL=*value*	specifies an OUT= data set numeric variable value for the OTHER= category
POTHER=*pattern*	specifies pattern for OTHER= bar

Table 26.4. Options to Enhance Plots Produced on Graphics Devices

ANNOTATE= *SAS-data-set*	specifies annotate data set with primary vertical axis data units
ANNOTATE2= *SAS-data-set*	specifies annotate data set with secondary vertical axis data units
CTEXT=*color*	specifies color for text
CTEXTSIDE=*color*	specifies color for row labels
CTEXTTOP=*color*	specifies color for column labels
DESCRIPTION='*string*'	specifies description for graphics catalog member
FONT=*font*	specifies font for text
HEIGHT=*value*	specifies text height in percent screen units
HTML=(*variable*)	specifies a variable whose values are URLs to be associated with bars
INFONT=*font*	specifies font for text inside frame
INHEIGHT=*value*	specifies text height in percent screen units for text inside frame
NAME='*string*'	specifies name for graphics catalog member

Table 26.5. Options for Reference Lines

CHREF=*color*	specifies color for HREF= lines
CVREF=*color*	specifies color for VREF= and VREF2= lines
HREF=*value-list*	requests reference lines perpendicular to horizontal axis
HREFCHAR='*character*'	specifies plot character for HREF= lines
HREFLABELS= ('*label1*'...'*labeln*')	specifies labels for HREF= lines
HREFLABPOS=n	specifies position of HREFLABELS= labels
LHREF=*linetype*	specifies line type for HREF= lines
LVREF=*linetype*	specifies line type for VREF= and VREF2= lines
VREF=*value-list*	requests reference lines perpendicular to primary vertical axis
VREF2=*value-list*	requests reference lines perpendicular to secondary vertical axis
VREFCHAR='*character*'	specifies plot character for VREF= lines
VREFLABELS= ('*label1*'...'*labeln*')	specifies labels for VREF= lines
VREF2LABELS= ('*label1*'...'*labeln*')	specifies labels for VREF2= lines
VREFLABPOS=n	specifies position of VREFLABELS= and VREF2LABELS= labels

Table 26.6. Options for Comparative Pareto Charts

ANNOKEY	applies annotation only to the key cell
CFRAMESIDE=*color*	specifies frame color for row labels
CFRAMETOP=*color*	specifies frame color for column labels
CLASS=(*variable-list*)	specifies classification variables
CLASSKEY= (*'value1' 'value2'*)	specifies the key cell
CPROP=*color*	specifies color for proportion-of-frequency bar
CTILES=(*variable*)	specifies colors for tile backgrounds
INTERTILE=*value*	specifies distance in percent screen units between tiles
MISSING1	specifies that missing values of the first CLASS= variable be analyzed as a level
MISSING2	specifies that missing values of the second CLASS= variable be analyzed as a level
NCOLS=n	specifies number of columns
NOKEYMOVE	suppresses the placement of the key cell in the top left corner
NROWS=n	specifies number of rows
ORDER1=*keyword*	specifies the order in which values of the first CLASS= variable are displayed
ORDER2=*keyword*	specifies the order in which values of the second CLASS= variable are displayed
TILELEGEND=(*variable*)	specifies legend for CTILES= colors
TILELEGLABEL=*'label'*	specifies label for TILELEGEND= legend

Table 26.7. Options for Grids

CGRID=*color*	specifies color for GRID lines
CGRID2=*color*	specifies color for GRID2 lines
GRID	adds grid corresponding to primary vertical axis
GRID2	adds grid corresponding to secondary vertical axis
LGRID=*linetype*	specifies line type for GRID lines
LGRID2=*linetype*	specifies line type for GRID2 lines
WGRID=*n*	specifies width of GRID lines
WGRID2=*n*	specifies width of GRID2 lines

Table 26.8. Options for Controlling Axes

ANGLE=*value*	rotates horizontal axis tick mark labels
AXISFACTOR=*value*	specifies distance factor between the tallest bar and the upper frame
CAXIS=*color*	specifies axis color
CAXIS2=*color*	specifies color for secondary vertical axis and tick marks
CFRAME=*color*	specifies color for area enclosed by axes and frame
HOFFSET=*value*	specifies horizontal axis offset in percent screen units
NOCHART	suppresses Pareto chart
NOFRAME	suppresses axis frame
NOHLABEL	suppresses horizontal axis label
NOVLABEL	suppresses primary vertical axis label
NOVLABEL2	suppresses secondary vertical axis label
NOVTICK	suppresses tick marks and tick mark labels for primary vertical axis
NOVTICK2	suppresses tick marks and tick mark labels for secondary vertical axis
SCALE=*keyword*	specifies units in which primary vertical axis is scaled
TURNVLABEL	turns and strings vertically the characters in the primary and secondary vertical axis labels
VAXIS=*value-list*	specifies tick mark values for primary vertical axis
VAXISLABEL=*'label'*	labels primary vertical axis
VAXIS2=*value-list*	specifies tick mark values for secondary vertical axis
VAXIS2LABEL=*'label'*	labels secondary vertical axis
VOFFSET=*value*	specifies vertical axis offset in percent screen units
WAXIS=n	specifies width in pixels for the axes and frame

Table 26.9. Options for Displaying a Sample Size Legend

CFRAMENLEG=*color*	frames the NLEGEND legend and fills the frame with the specified color
NLEGEND	requests sample size legend
NLEGEND=*'label'*|(*variable*)	requests sample size legend with specified label

Table 26.10. Options for Displaying Bars

BARLABEL=*keyword* (*variable-list*)	displays labels for bars
BARLABPOS=*keyword*	specifies position of BARLABEL= option
BARLEGEND= (*variable-list*)	displays legend for CBARS= colors or PBARS= patterns
BARLEGLABEL=*'label'*	displays label for BARLEGEND= legend
BARWIDTH=*value*	specifies width (vertical dimension) in percent screen units of the bars
CATLEGLABEL=*'label'*	specifies label for Pareto categories legend
CBARLINE=*color*	specifies color for bar outlines
CBARS=*color*	specifies color for bars
CBARS=*variable-list*	specifies variable that provides bar colors
CHIGH(n)=*color*	specifies color for bars with the n highest values
CLOW(n)=*color*	specifies color for bars with the n lowest values
HLLEGLABEL=*'label'*	displays label for the legend that describes colors and patterns of highest or lowest bars
INTERBAR=*value*	specifies distance between bars in percent screen units
LABOTHER= *'other-label'*	specifies label for "other" category
LAST=*'category'*	specifies bottommost category
NOHLLEG	suppresses legend describing colors and patterns of highest or lowest bars
PBARS=*pattern*	specifies pattern for the bars
PBARS=(*variable-list*)	specifies variable that provides bar patterns
PHIGH(n)=*pattern*	specifies pattern for bars with the n highest values
PLOW(n)=*pattern*	specifies pattern for bars with the n lowest values
WBARLINE=*n*	specifies width for bar outlines

Dictionary of Options

The following entries provide detailed descriptions of options you can specify after the slash (/) in the VBAR statement. For example, to request that the bars of your Pareto chart be colored red, use the CBARS= option, as follows:

```
proc pareto data=failure;
   vbar cause / cbars = red ;
run;
```

The marginal notes *Graphics* and *Line Printer* identify options that apply only to charts displayed with graphics devices and line printers, respectively.

ANCHOR=BC | BL | TC | TR

specifies where the Pareto curve is anchored to the leftmost bar on the chart. The following table lists the possible positions.

Keyword	Anchoring position
BC	bottom center
BL	bottom left corner
TC	top center
TR	top right corner

See Output 29.2.1 on page 901 for an illustration. The default is TR.

ANGLE=*value*

specifies an angle in degrees for rotating the Pareto category labels on the horizontal axis. The *value* is the angle between the baseline of the label and the horizontal axis. See Output 29.1.1 on page 896 and Output 29.1.2 on page 896 for an illustration. The *value* must be greater than or equal to -90 and less than 90. If you are using a line printer, only ANGLE=0 and ANGLE=-90 are applicable. If you are using a graphics device and you specify the ANGLE= option, you should also specify a software font with the FONT= option in the VBAR statement or the FTEXT= option in a GOPTIONS statement. The default *value* is zero.

ANNOKEY

Graphics

specifies that annotation requested with the ANNOTATE= and ANNOTATE2= options is to be applied only to the key cell in a comparative Pareto chart. By default, annotation is applied to all of the cells.

ANNOTATE=*SAS-data-set*
ANNO=*SAS-data-set*

Graphics

specifies an input data set that contains annotate variables as described in *SAS/GRAPH Software: Reference*. You can use this data set to customize charts. This data set is associated with the primary vertical axis. If the annotation is based on data coordinates, you must use the same units as the primary vertical axis. Features provided in the ANNOTATE= data set are added to every chart produced with the VBAR statement.

ANNOTATE2=*SAS-data-set*
ANNO2=*SAS-data-set*

Graphics

specifies an input data set that contains annotate variables as described in *SAS/GRAPH Software: Reference*. You can use this data set to customize charts. This data set is associated with the secondary vertical axis. If the annotation is based on data coordinates, you must use the same units as the secondary vertical axis. Features provided in the ANNOTATE2= data set are added to every chart produced with the VBAR statement.

AXISFACTOR=*value*

specifies a factor used in scaling the primary vertical axis. This factor determines (approximately) the ratio of the length of the axis to the length of the tallest bar, and it is used to provide space for the cumulative percent curve. The *value* must be greater than one.

By default, the factor is chosen so that the curve will be anchored at the top right corner of the leftmost bar (see also the ANCHOR= option). If this causes the bars to be flattened excessively, however, a smaller default factor is used.

The AXISFACTOR= option is not applicable if the curve is suppressed with the NOCURVE option.

BARLABEL = VALUE | CMPCT | (*variable***)**

specifies that a label is to be displayed above each bar. If you specify BARLABEL=VALUE, the label indicates the height of the bar in the units used by the primary vertical axis. See Example 29.8 on page 914 for an illustration.

If you specify BARLABEL=CMPCT, the label indicates the cumulative percent for that bar. An alternative to BARLABEL=CMPCT is the CMPCTLABEL option, which labels points on the cumulative percent curve with the cumulative percents.

If you specify BARLABEL= (*variable*), the label indicates the values of the variable specified in parentheses. The variable can have a formatted length less than or equal to 32. If a format is associated with the variable, then the formatted value is displayed. The values must be consistent within observations corresponding to a particular Pareto category. The variable is saved in the OUT= data set.

BARLABPOS=VFIT | HCENTER | VBAR | HLJUST

specifies the position for labels requested with the BARLABEL= option.

- BARLABPOS=VFIT displays the labels vertically on or above the bars, depending on the available space (the labels are truncated if necessary).
- BARLABPOS=HCENTER centers the labels horizontally above the bars.
- BARLABPOS=VBAR displays the labels vertically on the bars (the labels are truncated if necessary).
- BARLABPOS=HLJUST left-justifies the labels horizontally above the bars.

By default, the labels are horizontally centered above the bars, with a reduction in text height if necessary. Reduction is not applied with the BARLABPOS= options.

BARLEGEND=(*variable-list***)**

Graphics

specifies that a legend is to be added to the chart to explain colors for bars specified with the CBARS=(*variable-list*) option or patterns for bars specified with the PBARS=(*variable-list*) option. The *variable-list* must be enclosed in parentheses even if only one *variable* is specified. See Output 29.4.1 on page 908 for an illustration.

The values of the BARLEGEND= variable provide the explanatory labels used in the legend. The variable can have a formatted length that does not exceed 32. If a format is associated with the variable, then the formatted value is displayed.

The BARLEGEND= option is not applicable unless you specify CBARS=(*variable-list*) or PBARS=(*variable-list*) or both. In the DATA= data set, the values of the BARLEGEND= variable must be identical in observations for which the value of the CBARS= variable or the PBARS= variable (or the combination of these two values) is the same. This ensures that the legend derived from the BARLEGEND= variable is consistent.

If you specify more than one process variable in the chart statement and a corresponding list of CBARS= or PBARS= variables, you can specify a list of BARLEGEND= variables. The number of BARLEGEND= variables should be less than or equal to the number of process variables. The lists of variables are matched so that the first BARLEGEND= variable is applied to the first process variable and the first CBARS= or PBARS= variable, the second BARLEGEND= variable is applied to the second process variable and the second CBARS= or PBARS= variable, and so forth. If the list of process variables is longer than the *variable-list* in the BARLEGEND= option, the charts for the extra process variables will not display a bar legend.

BARLEGLABEL=*'label'*

Graphics

specifies the label displayed to the left of the legend created with the BARLEGEND= option. See Output 29.4.1 on page 908 for an illustration.

The BARLEGLABEL= option is applicable only in conjunction with CBARS= or PBARS= variables. The label can be up to 16 characters and must be enclosed in quotes.

If no label is specified with the BARLEGLABEL= option, the label associated with the BARLEGEND= variable is displayed (unless the label is longer than 16 characters, in which case the variable name is displayed). If the BARLEGLABEL= option is not specified and no label is associated with the BARLEGEND= variable, no legend label is displayed. If both labels are specified, the BARLEGLABEL= label takes precedence over the variable label.

BARWIDTH=*value*

Graphics

specifies the width of the bars in percent screen units. By default, the bars are as wide as possible.

CATLEGLABEL=*'label'*

specifies a label for the category legend that is added when there is insufficient space to label the categories along the horizontal axis. The *label* can be up to 16 characters and must be enclosed in quotes. The default label is *Categories:*. See Example 29.3

on page 905 for an illustration. The CATLEGLABEL= option is ignored if it is unnecessary to add the legend.

CAXIS=*color*
CAXES=*color*
CA=*color*

Graphics

specifies the color for the axis line and tick marks on the chart. The default color is the first color in the device color list.

CAXIS2=*color*

Graphics

specifies the color for the tick mark labels and axis label associated with the secondary axis. By default, the color specified with the CTEXT= option (or its default) is used.

CBARLINE=*color*

Graphics

specifies the color for bar outlines.

CBARS=*color*
CBARS=(*variable-list***)**

Graphics

specifies how the bars of the Pareto chart are to be colored. You can use one of the following approaches:

- You can specify a single color to be used for all the bars with CBARS=*color*. You can use this option in conjunction with the CHIGH and CLOW options. See Output 29.2.1 on page 901 for an illustration.
- You can specify a distinct color for each bar (or combination of bars) by providing the colors as values of a CBARS= variable. This variable must be a character variable of length eight. You can use the special value `EMPTY` to indicate that a bar is not to be colored. Note that the variable name must be enclosed in parentheses. You cannot specify a CBARS= variable in conjunction with the CHIGH and CLOW options. See Output 29.3.1 on page 906 and Output 29.4.1 on page 908 for examples.

If you specify more than one process variable, you can specify more than one CBARS= variable. The number of CBARS= variables should be less than or equal to the number of process variables. The two lists of variables are paired in order of specification. If a CBARS= variable is not provided for a process variable, the bars for that chart are not colored.

If you specify one or more CBARS= variables, you can also use the BARLEGEND= option to add a legend to the chart that explains the significance of each color. Furthermore, you can use the PBARS= option to specify patterns in conjunction with the CBARS= option. See Output 29.4.1 on page 908 and Output 29.5.1 on page 909 for examples.

CCONNECT=*color*

Graphics

specifies the color for the line segments connecting the points on the cumulative percent curve. You can specify the color for the points themselves with the COLOR= option in the SYMBOL statement; this is the default color.

CFRAME=*color*

Graphics

specifies the color for filling the area enclosed by the axes and the frame. By default, this area is not filled. The CFRAME= option cannot be used in conjunction with the NOFRAME option or the CTILES= option.

CFRAMENLEG=*color*

Graphics

specifies that the legend requested with the NLEGEND option be framed and that the frame be filled with the color indicated. If you specify CFRAMENLEG=EMPTY, a frame is drawn but is not filled with a color. See Figure 26.4 on page 803 and Output 29.1.4 on page 898 for illustrations.

CFRAMESIDE=*color*

Graphics

specifies the color for filling the frame area for the row labels displayed along the left side of a comparative Pareto chart requested with the CLASS= option. If a label is associated with the classification variable, this color is also used to fill the frame area for this label. By default, these areas are not filled.

CFRAMETOP=*color*

Graphics

specifies the color for filling the frame area for the column labels displayed across the top of a comparative Pareto chart requested with the CLASS= option. If a label is associated with the classification variable, this color is also used to fill the frame area for this label. By default, these areas are not filled.

CGRID=*color*

Graphics

specifies the color for grid lines requested with the GRID option. The default color is the first color in the device color list. If you specify the CGRID= option, you do not need to specify the GRID option.

CGRID2=*color*

Graphics

specifies the color for grid lines requested with the GRID2 option. The default color is the first color in the device color list. If you specify the CGRID2= option, you do not need to specify the GRID2 option.

CHIGH(n**)=***color*

Graphics

specifies the color used to fill the bars with the n highest values. You cannot use the CHIGH option in conjunction with a CBARS= variable, but you can use the CHIGH(n)= option together with the CLOW(n)= and CBARS=*color* options. See Output 29.3.1 on page 906 for an illustration. By default, the bars are empty.

CHREF=*color*

Graphics

specifies the color for lines requested with the HREF= option. The default is the first color in the device color list.

CLASS=*variable*

CLASS=(*variable1 variable2***)**

creates a comparative Pareto chart using the levels of the *variables*. You must enclose two *variables* in parentheses. See Example 29.1 on page 895 and Example 29.2 on page 899.

If you specify a single CLASS= *variable*, the observations in the input data set are classified by the formatted values (levels) of the *variable*. A Pareto chart is created for the process variable values in each level, and these component charts (referred to

as cells) are arranged in an array. The cells are labeled with the levels, and uniform horizontal and vertical axes are used to facilitate comparisons.

If you specify two CLASS= *variables*, the observations in the input data set are cross-classified by the values (levels) of the *variables*. A Pareto chart is created for the process variable values in each cell of the cross-classification, and these charts are arranged in a *matrix*. The levels of the first CLASS= *variable* label the rows, and the levels of the second CLASS= *variable* label the columns. Uniform horizontal and vertical axes are used to facilitate comparisons.

Note that the array or matrix comparative Pareto chart is displayed only if the chart is produced on a graphics device. If the chart is produced on a line printer, the cells are printed separately.

The CLASS= *variables* can be numeric or character. The maximum length of a character *variable* is 32. If a format is associated with a CLASS= *variable*, the formatted values determine the levels. Only the first 32 characters of the formatted value are used to determine the levels. You can specify whether missing values are to be treated as a level with the MISSING1 and MISSING2 options.

If a label is associated with a CLASS= *variable*, the label is displayed on the chart. On charts produced with a graphics device, the *variable* label is displayed parallel to the column (or row) labels. On charts produced with a line printer, the *variable* label is displayed at the top of the chart.

CLASSKEY=*'value'*
CLASSKEY=(*'value1' 'value2'***)**

specifies the *key cell* in a comparative histogram requested with the CLASS= option. The bin size and midpoints are first determined for the key cell, and then the midpoint list is extended to accommodate the data ranges for the remaining cells. Thus the choice of the key cell determines the uniform horizontal axis used for all cells.

If you specify CLASS=*variable*, you can specify CLASSKEY=*'value'* to identify the key cell as the level for which *variable* is equal to *value*. The *value* can have up to 32 characters, and you must specify a formatted *value*. By default, the levels are sorted in the order determined by the ORDER1= option, and the key cell is the level that occurs first in this order. The cells are displayed in this order from top to bottom (or left to right), and consequently the key cell is displayed at the top or at the left. If you specify a different key cell with the CLASSKEY= option, this cell is displayed at the top or at the left unless you also specify the NOKEYMOVE option.

If you specify CLASS=(*variable1 variable2*), you can specify CLASSKEY=(*'value1' 'value2'*) to identify the key cell as the level for which *variable1* is equal to *value1* and *variable2* is equal to *value2*. Here, *value1* and *value2* must be formatted values, and they must be enclosed in quotes. By default, the levels of *variable1* are sorted in the order determined by the ORDER1= option, and then within each of these levels, the levels of *variable2* are sorted in the order determined by the ORDER2= option. The default key cell is the combination of levels of *variable1* and *variable2* that occurs first in this order. The cells are displayed in order of *variable1* from top to bottom and in order of *variable2* from left to right. Consequently, the default key cell is displayed in the upper left corner. If you specify a different key cell with the

CLASSKEY= option, this cell is displayed in the upper left corner unless you also specify the NOKEYMOVE option.

For an example of the use of the CLASSKEY= option, see Output 29.1.3 on page 897.

CLOW(*n*)=*color***

specifies the color used to fill the bars with the n lowest values. You cannot use the CLOW(n)= option in conjunction with a CBARS=*variable*, but you can use the CLOW(n)= option together with the CBARS=*color* and CHIGH options. See Output 29.3.1 on page 906 for an illustration of the CHIGH(n)= option. By default, the bars are empty. *Graphics*

CMPCTLABEL

labels points on the cumulative percent curve with their values. By default, the points are not labeled.

CONNECTCHAR=*'character'*
CCHAR=*'character'*

specifies the plot character for line segments that connect points on the cumulative percent curve. The default character is a plus sign (+). *Line Printer*

COTHER=*color*

specifies the color for the bar defined by the OTHER= option. By default, this bar is not filled with a color. The COTHER= option is not applicable unless a CBARS=*variable* is specified. *Graphics*

CPROP=*color*

specifies the color for a proportion-of-frequency bar that is displayed horizontally across the top of each tile in a comparative Pareto chart. The length of the bar relative to the width of the tile indicates the proportion of the total frequency count in the chart that is represented by the tile. You can use the bars to visualize the distribution of frequency count by tile. See Output 29.1.4 on page 898 for an illustration. *Graphics*

The CPROP= option provides a graphical alternative to the NLEGEND options, which display the actual count. The CPROP= option is applicable only with comparative Pareto charts. Empty bars are displayed if you specify CPROP=EMPTY. Bars are not displayed if the CPROP= option is not specified.

CTEXT=*color*
CT=*color*

specifies the color for text, such as tick mark labels, axis labels, and legends. The default is the value specified for the CTEXT= option in the GOPTIONS statement. *Graphics*

CTEXTSIDE=*color*

specifies the color for row labels displayed along the left side of a comparative Pareto chart requested with the CLASS= option. The default color is the color specified with the CTEXT= option in the VBAR statement or the CTEXT= option in the GOPTIONS statement. *Graphics*

CTEXTTOP=*color*

specifies the color for column labels displayed across the top of a comparative Pareto chart requested with the CLASS= option. The default color is the color specified *Graphics*

with the CTEXT= option in the VBAR statement or the CTEXT= option in the GOPTIONS statement.

CTILES=(*variable***)**

Graphics

specifies a character variable of length eight whose values are the fill colors for the tiles in a comparative Pareto chart. The CTILES= option generalizes the CFRAME= option, which provides a single color for all of the tiles. The *variable* must be enclosed in parentheses. The values of the *variable* must be identical for all observations with the same level of the CLASS= variables. You can use the same color to fill more than one tile. You can use the special value EMPTY to indicate that a tile is not to be filled.

The CTILES= option cannot be used in conjunction with the NOFRAME option or the CFRAME= option. You can use the TILELEGEND= option in conjunction with the CTILES= option to add an explanatory legend for the CTILES= colors at the bottom of the chart. See Output 29.5.1 on page 909 for an illustration. By default, the tiles are not filled.

CVREF=*color*

Graphics

specifies the color for lines requested with the VREF= and VREF2= options. The default color is the first color in the device color list.

DESCRIPTION=*'string'*
DES=*'string'*

Graphics

specifies a descriptive string, up to 40 characters, that appears in the description field of the PROC GREPLAY master menu.

FONT=*font*

Graphics

specifies a software font for text used in labels and legends. The FONT= option takes precedence over the FTEXT= option in the GOPTIONS statement.

FREQ=*variable*

specifies a frequency variable whose value provides the counts (numbers of occurrences) of the values of the process variable. Specifying a FREQ= variable is equivalent to replicating the observations in the input data set. The FREQ= variable must be a numeric variable with nonnegative integer values. If you specify more than one process variable in the chart statement, the FREQ= variable values are used with each process variable. See “26.2” on page 802 for an illustration. If you do not specify a FREQ= variable, each value of the process variable is counted exactly once.

GRID

adds a grid to the Pareto chart corresponding to the primary vertical axis. Grid lines are horizontal lines positioned at tick marks on the primary vertical axis. The lines are useful for comparing the heights of the bars.

GRID2

adds a grid to the Pareto chart corresponding to the secondary vertical axis. Grid lines are horizontal lines positioned at tick marks on the secondary vertical axis. The lines are useful for reading the cumulative percent curve.

HEIGHT=*value*

specifies the height in percent screen units of text for labels and legends. This option should be used only in conjunction with the FONT= option. The HEIGHT= option takes precedence over the HTEXT= option in a GOPTIONS statement. *Graphics*

HLLEGLABEL=*'label'*

specifies a label displayed to the left of the legend that is automatically created when you use a combination of the CHIGH, CLOW, PHIGH, and PLOW options. See Output 29.3.1 on page 906 for an illustration. The *label* can be up to 16 characters and must be enclosed in quotes. The default *label* is *Bars:*. *Graphics*

HOFFSET=*value*

specifies the length in percent screen units of the offset at both ends of the horizontal axis. You can eliminate the offset by specifying HOFFSET=0. *Graphics*

HREF=*'value-list'*

specifies where reference lines perpendicular to the horizontal (Pareto category) axis are to appear on the chart. Character values can be up to 32 characters and must be enclosed in quotes. The values must be values of the process variable even when the bars are numbered and a category legend is introduced.

HREFCHAR=*'character'*

specifies the plot character used to form the lines requested with the HREF= option. The default character is a vertical bar (|). *Line Printer*

HREFLABELS=*'label1'...'labeln'*

specifies labels for the lines requested with the HREF= option. The number of labels must equal the number of lines requested. Labels can be up to 16 characters and must be enclosed in quotes.

HREFLABPOS=*n*

specifies the vertical positioning of the HREFLABELS= labels. HREFLABPOS=1 positions the labels along the top of the chart. HREFLABPOS=2 staggers the labels from top to bottom. HREFLABPOS=3 positions the labels along the bottom. By default, HREFLABPOS=1.

HTML=*variable*

specifies URLs as values of the specified character variable (or formatted values of a numeric variable). These URLs are associated with bars on the Pareto chart when high resolution graphics output is directed into HTML. The value of the HTML= variable should be the same for each observation with a given value of the subgroup variable.

INFONT=*font*

specifies a software font for text used inside the frame of the chart, such as sample size legends. The INFONT= option takes precedence over the FONT= option and the FTEXT= option in the GOPTIONS statement. *Graphics*

INHEIGHT=*value*

specifies the height in percent screen units of text used inside the frame of the chart, such as sample size legends and bar labels. This option should be used in conjunction with the INFONT= option. *Graphics*

INTERBAR=*value*

specifies the distance in percent screen units between bars on the chart. By default, the bars are contiguous. See Figure 26.4 on page 803 for an illustration.

INTERTILE=*value*

Graphics

specifies the distance in horizontal percent screen units between tiles (cells) in a comparative Pareto chart. By default, the tiles are contiguous. See Output 29.1.3 on page 897 for an illustration.

LABOTHER = *'other-label'*

is used in conjunction with the BARLABEL=(variable) option and specifies a label for the "other" category that is optionally specified with the OTHER= option.

LAST=*'category'*

specifies that the bar corresponding to the *category* is to be displayed at the right end of the chart regardless of the percent associated with this category. The *category* must be a formatted value of the process variable and must be enclosed in quotes. The *category* can be up to 32 characters. See Figure 26.7 on page 806 for an illustration.

LGRID=*line-type*

Graphics

specifies the line type for the grid requested with the GRID option. The default *line-type* is 1, which produces a solid line. If you specify the LGRID= option, you do not need to specify the GRID option.

LGRID2=*line-type*

Graphics

specifies the line type for the grid requested with the GRID2 option. The default *line-type* is 1, which produces a solid line. If you specify the LGRID2= option, you do not need to specify the GRID2 option.

LHREF=*line-type*

LH=*line-type*

Graphics

specifies the line type for lines requested with the HREF= option. The default *line-type* is 2, which produces a dashed line.

LOTHER=*'label'*

specifies a label for the bar defined with the OTHER= option. This label appears in the legend created with the BARLEGEND= option. The *label* must be enclosed in quotes and can be up to 32 characters. The default *label* is the value specified with the OTHER= option. The LOTHER= option is applicable only when a BARLEGEND= variable is specified.

LVREF=*line-type*

LV=*line-type*

Graphics

specifies the line type for lines requested with the VREF= and VREF2= options. See Output 29.2.3 on page 903 for an illustration. The default *line-type* is 2, which produces a dashed line.

MAXCMPCT=*percent*

specifies that only the Pareto categories with the n highest frequency counts are to be displayed, where the sum of the n corresponding percents is less than or equal to the specified *percent*. For example, if you specify

```
proc pareto data=failure;
   vbar cause / maxcmpct = 90 ;
```

the chart displays only the n most frequently occurring categories that account for no more than 90 percent of the total frequency.

You can use the OTHER= option in conjunction with the MAXCMPCT= option to create and display a new category that combines those categories that are not selected with the MAXCMPCT= option. For example, if you specify

```
proc pareto data=failure;
   vbar cause / maxcmpct = 90
                other    = 'Others' ;
```

the chart displays the categories that account for no more than 90 percent of the total frequency, together with a category labeled *Others* that merges the remaining categories. The MAXCMPCT= option is an alternative to the MINPCT= and MAXNCAT= options.

MAXNCAT=n

specifies that only the Pareto categories with the n highest frequencies are to be displayed. For example, if you specify

```
proc pareto data=failure;
   vbar cause / maxncat = 20 ;
```

the chart displays only the categories with the 20 highest frequencies. If the total number of categories is less than 20, all the categories are displayed.

You can use the OTHER= option in conjunction with the MAXNCAT= option to create and display a new category that combines those categories that are not selected with the MAXNCAT= option. For example, if you specify

```
proc pareto data=failure;
   vbar cause / maxncat = 20
                other= 'Others' ;
```

the chart displays the categories with the 19 highest frequencies, together with a category labeled *Others* that merges the remaining categories. See Figure 26.6 on page 805 for another illustration.

The MAXNCAT= option is an alternative to the MINPCT= and MAXCMPCT= options.

MINPCT=*percent*

specifies that only the Pareto categories with frequency percents greater than or equal to the specified *percent* are to be displayed. For example, if you specify

```
proc pareto data=failure;
   vbar cause / minpct = 5 ;
```

the chart displays only those categories with at least five percent of the total frequency.

You can use the OTHER= option in conjunction with the MINPCT= option to create and display a new category that combines those categories that are not selected with the MINPCT= option. The merged category created by the OTHER= option is displayed even if its total percent is less than the *percent* specified with the MINPCT= option. For example, if you specify

```
proc pareto data=failure;
   vbar cause / minpct = 5
                other  = 'Others' ;
```

the chart displays the categories with percents greater than or equal to five percent, together with a category labeled *Others* that merges the remaining categories.

The MINPCT= option is an alternative to the MAXNCAT= and MAXCMPCT= options.

MISSING

specifies that missing values of the process variable are to be treated as a Pareto category represented with a bar on the chart. If the process variable is a character variable, a missing value is defined as a blank internal (unformatted) value. If the process variable is numeric, a missing value is defined as any of the SAS missing values. If you do not specify the MISSING option, missing values are excluded from the analysis.

MISSING1

specifies that missing values of the first CLASS= variable are to be treated as a level of the CLASS= variable. If the first CLASS= variable is a character variable, a missing value is defined as a blank internal (unformatted) value. If the first CLASS= variable is numeric, a missing value is defined as any of the SAS missing values. If you do not specify MISSING1, observations in the DATA= data set for which the first CLASS= variable is missing are excluded from the analysis.

MISSING2

specifies that missing values of the second CLASS= variable are to be treated as a level of the CLASS= variable. If the second CLASS= variable is a character variable, a missing value is defined as a blank internal (unformatted) value. If the second CLASS= variable is numeric, a missing value is defined as any of the SAS missing values. If you do not specify MISSING2, observations in the DATA= data set for which the second CLASS= variable is missing are excluded from the analysis.

NAME=*'string'*

Graphics

specifies a name for the chart, up to eight characters, that appears in the PROC GREPLAY master menu. The default name is 'PARETO'.

NCOLS=*n*
NCOL=*n*

specifies the number of columns in a comparative Pareto chart. You can use the NCOLS= option in conjunction with the NROWS= option. See Output 29.2.3 (page 903) and Output 29.2.4 (page 904) for an illustration. By default, NCOLS=1 and NROWS=2 if one CLASS= variable is specified, and NCOLS=2 and NROWS=2 if two CLASS= variables are specified.

NLEGEND
NLEGEND='*label***'**
NLEGEND=(*variable***)**

requests a sample size legend and specifies its form as follows:

- If you specify the NLEGEND option, the form is *N*=*n*, where n is the total count for the Pareto categories. In a comparative Pareto chart, a legend is displayed in each tile, and n is the total count for that particular cell. See Output 29.2.1 on page 901 for an illustration.
- If you specify the NLEGEND='*label*' option, the form is *label*=*n*, where n is the total count for the Pareto categories. The label can be up to 32 characters and must be enclosed in quotes. For an illustration, see Figure 26.4 on page 803 or Output 29.1.4 on page 898.
- If you specify the NLEGEND=(*variable*) option, the legend is the value of the *variable*, which must be a variable in the DATA= data set whose formatted length does not exceed 32. If a format is associated with the variable, then the formatted value is displayed. This option is intended for use with comparative Pareto charts and enables you to display a customized legend inside each tile (this legend need not provide total count). It is assumed that the values of the *variable* are identical for all observations in a particular class.

By default, the legend is placed in the upper-left corner of the chart. If the NOCURVE option is specified, the legend is placed in the upper-right corner of the chart. You can use the CFRAMENLEG= option to frame the sample size legend. No legend is displayed if you do not specify an NLEGEND option.

NOCHART

suppresses the creation of a Pareto chart. This option is useful when you are simply creating an output data set.

NOCURVE

suppresses the display of the cumulative percent curve and the secondary vertical axis. Compare Output 29.2.1 (page 901) and Output 29.2.2 (page 902) for an illustration.

NOFRAME

suppresses the frame that is drawn around the chart by default. The NOFRAME option cannot be specified in conjunction with the CFRAME= or CTILES= options.

NOHLABEL
suppresses the label for the horizontal axis. This is useful for avoiding clutter in situations where the meaning of the horizontal axis is apparent from the labels for the Pareto categories. See Output 29.2.2 on page 902 for an illustration.

NOHLLEG
suppresses the legend generated by the CHIGH(n)=, CLOW(n)=, PHIGH(n)=, and PLOW(n)= options.

NOKEYMOVE
Graphics
suppresses the rearrangement of cells within a comparative Pareto chart that occurs when you use the CLASSKEY= option. The key cell appears in the top left corner of a comparative Pareto chart unless you use the CLASSKEY= option together with the NOKEYMOVE option.

NOVLABEL
suppresses the label for the primary vertical axis.

NOVLABEL2
suppresses the label for the secondary vertical axis. This is useful for avoiding clutter on comparative Pareto charts.

NOVTICK
suppresses the primary vertical axis label, tick marks, and tick mark labels.

NOVTICK2
suppresses the secondary vertical axis label, tick marks, and tick mark labels.

NROWS=n
NROW=n
Graphics
specifies the number of rows in a comparative Pareto chart. You can use the NROWS= option in conjunction with the NCOLS= option. See Output 29.2.3 on page 903 and Output 29.2.4 on page 904 for an illustration. By default, NROWS=2.

ORDER1=INTERNAL | FORMATTED | DATA | FREQ
specifies the display order for the values of the first CLASS= variable. The levels of the first CLASS= variable are always constructed using the formatted values of the variable, and the formatted values are always used to label the rows (columns) of a comparative Pareto chart.

If you specify ORDER1=INTERNAL, the rows (columns) are displayed from top to bottom (left to right) in increasing order of the internal, or unformatted, values of the first CLASS= variable. If there are two or more distinct internal values with the same formatted value, the order is determined by the internal value that occurs first in the input data set. For example, suppose that you use a numeric CLASS= variable called DAY (with values 1, 2, and 3) to create a one-way comparative Pareto chart. Suppose also that you use the FORMAT procedure to associate the formatted values 1 = 'Wednesday', 2 = 'Thursday', and 3 = 'Friday' with the variable DAY. If you specify ORDER1=INTERNAL, the rows of the comparative chart will appear in chronological order (*Wednesday*, *Thursday*, *Friday*) from top to bottom.

If you specify ORDER1=FORMATTED, the rows (columns) are displayed from top to bottom (left to right) in increasing order of the formatted values of the first CLASS= variable. For instance, in the previous illustration, if you specify ORDER1=FORMATTED, the rows will appear in alphabetical order (*Friday*, *Thursday*, *Wednesday*) from top to bottom.

If you specify ORDER1=DATA, the rows (columns) are displayed from top to bottom (left to right) in the order in which the values of the first CLASS= variable first appear in the input data set.

If you specify ORDER1=FREQ, the rows (columns) are displayed from top to bottom (left to right) in order of *decreasing* frequency count. If two or more classes have the same frequency count, the order is determined by the formatted values.

By default, ORDER1=INTERNAL.

ORDER2=INTERNAL | FORMATTED | DATA | FREQ

specifies the display order for the values of the second CLASS= variable. The levels of the second CLASS= variable are always constructed using the formatted values of the variable, and the formatted values are always used to label the columns of a two-way comparative Pareto chart.

The PARETO procedure determines the layout of a two-way comparative Pareto chart by first using the ORDER1= option to obtain the order of the rows from top to bottom (recall that ORDER1=INTERNAL by default). Then the ORDER2= option is applied to the observations corresponding to the first row to obtain the order of the columns from left to right. If any columns remain unordered (that is, the categories are unbalanced), the ORDER2= option is applied to the observations in the second row, and so on until all the columns have been ordered.

The values of the ORDER2= option are interpreted as described for the ORDER1= option. By default, ORDER2=INTERNAL.

OTHER=*'category'*

specifies a new category that merges all categories not selected with the MAXNCAT=, MINPCT=, or MAXCMPCT= options. See Figure 26.7 on page 806 for an illustration.

The *category* should be specified as a formatted value of the process variable. The *category* can be up to 32 characters and must be enclosed in quotes. If you specify an OUT= data set, you should also specify an internal value corresponding to the *category* with the OTHERCVAL= option or the OTHERNVAL= option.

The OTHER= option is not applicable unless you specify the MAXNCAT=, MINPCT=, or MAXCMPCT= option. You can use the COTHER=, LOTHER=, POTHER=, OTHERCVAL=, and OTHERNVAL= options with the OTHER= option.

OTHERCVAL=*'value'*

specifies the internal (unformatted) value for a character process variable in the OUT= data set that corresponds to the category created with the OTHER= option. The *category* can be up to 32 characters and must be enclosed in quotes.

The OTHERCVAL= option is not applicable unless you specify the OTHER= and OUT= options. If you specify the OTHER= option but not the OTHERCVAL= option, the default *value* is the *value* specified with the OTHER= option.

OTHERNVAL=*value*
specifies the internal (unformatted) value for a numeric process variable in the OUT= data set that corresponds to the category created with the OTHER= option. The OTHERNVAL= option is not applicable unless you specify the OTHER= and OUT= options. If you specify the OTHER= option but not the OTHERNVAL= option, *value* is assigned a missing value.

OUT=*SAS-data-set*
creates an output data set that contains the information displayed in the Pareto chart. This is useful if you want to create a report to accompany your chart. See Example 29.8 on page 914 for an illustration.

PBARS=*pattern*
PBARS=(*variable-list***)**

Graphics

specifies pattern fills for the bars. You can use one of two approaches:

- You can specify a single pattern to be used for all the bars with the PBARS=*pattern* option. You can use this option in conjunction with the PHIGH and PLOW options. See Output 29.2.1 on page 901 for an illustration.
- You can specify a distinct pattern for *each* bar (or combination of bars) by providing the patterns as values of a PBARS= variable. For example, you might use the solid pattern (S) to indicate severe problems and the empty pattern (E) for all other problems. The variable must be a character variable of length eight, and the variable name must be enclosed in parentheses. You cannot specify a PBARS= variable in conjunction with the PHIGH and PLOW options. See Output 29.4.1 on page 908 and Output 29.5.1 on page 909 for illustrations.

If you specify more than one process variable in the chart statement, you can provide more than one PBARS= variable. The number of PBARS= variables should be less than or equal to the number of process variables. The two lists of variables are paired in order of specification. If a PBARS= variable is not provided for a process variable, the bars for that chart are not filled.

If you specify one or more variables with the PBARS= option, you can also use the BARLEGEND= option to add a legend to the chart that explains the significance of each pattern. Furthermore, you can use the CBARS= option to specify colors in conjunction with the PBARS= option. See Output 29.4.1 on page 908 and Output 29.5.1 on page 909 for illustrations.

PHIGH(*n***)=***pattern*

Graphics

specifies the pattern used to fill the bars with the n highest values. You cannot specify the PHIGH option in conjunction with a PBARS= variable, but you can specify the PHIGH(n)= option together with the PLOW(n)= and PBARS=*pattern* options. See Output 29.3.1 on page 906 for an illustration. By default, the bars are empty.

PLOW(*n*)=*pattern***

Graphics

specifies the pattern used to fill the bars with the n lowest values. You cannot specify the PLOW option in conjunction with a PBARS= variable, but you can use the PLOW(n)= option together with the PHIGH(n)= and PBARS=*pattern* options. See Output 29.3.1 on page 906 for an illustration of the PHIGH(n)= option. By default, the bars are empty.

POTHER=*pattern*

Graphics

specifies the pattern used for the bar defined by the OTHER= option. By default, this bar is empty. The POTHER= option is not applicable unless a PBARS= variable is specified.

SCALE=PERCENT | COUNT | WEIGHT

specifies the scale for the primary vertical axis.

If you specify SCALE=PERCENT, the scale is percent of total frequency. If a WEIGHT= variable is used, the scale is percent of total weight.

If you specify SCALE=COUNT, the scale is counts. See Output 29.1.4 on page 898 for an illustration. This option is not applicable if a WEIGHT= variable is used. You can specify SCALE=FREQUENCY instead of SCALE=COUNT.

If you specify SCALE=WEIGHT, the vertical axis is scaled in the same units as the WEIGHT= variable. This option is not applicable unless you use a WEIGHT= variable.

By default, SCALE=PERCENT. See Output 29.8.1 on page 915 for an example. Regardless of how SCALE= is specified, the secondary axis is scaled in cumulative percent units.

SYMBOLCHAR=*'character'*

Line Printer

specifies the plot character for points on the cumulative percent curve. The default character is an asterisk (*).

TILELEGEND=(*variable***)**

Graphics

specifies a variable used to add a legend for CTILES= colors. The variable can have a formatted length less than or equal to 32. If a format is associated with the variable, then the formatted value is displayed. The TILELEGEND= option must be used in conjunction with the CTILES= option for filling the tiles in a comparative Pareto chart. If CTILES= is specified and TILELEGEND= is not specified, a color legend is not displayed.

The values of the CTILES= and TILELEGEND= variables should be consistent for all observations with the same level of the CLASS= variables. The value of the TILELEGEND= variable is used to identify the corresponding color value of the CTILES= variable in the legend. See Output 29.5.1 on page 909 for an illustration.

TILELEGLABEL=*'label'*

Graphics

specifies a label displayed to the left of the legend that is created when you specify a TILELEGEND= variable. The *label* can be up to 16 characters and must be enclosed in quotes. The default *label* is *Tiles:*. See Output 29.5.1 on page 909 for an illustration.

TURNVLABEL
TURNVLABELS

Graphics

turns and strings out vertically the characters in the labels for the primary and secondary vertical axes. This happens by default when a hardware font is used.

VAXIS=*value-list*

specifies tick mark values for the primary vertical axis. The values must be equally spaced and in increasing order, and the first *value* must be zero. You must scale the values in the same units as the bars (see the SCALE= option), and the last *value* must be greater than or equal to the height of the largest bar.

VAXISLABEL=*'label'*

specifies a label, up to 40 characters, for the primary vertical axis. The default label depends on the value of the SCALE= option, or it is the label associated with the WEIGHT= variable.

VAXIS2=*value-list*

specifies tick mark values for the secondary vertical axis. The values must be equally spaced and in increasing order, and the first *value* must be zero. You must scale the values in percent units, and the last *value* must be greater than or equal to 100.

VAXIS2LABEL=*'label'*

specifies a label, up to 40 characters, for the secondary vertical axis. The default label is *Cumulative Percent* or *Cm Pct*, depending on the space available.

VOFFSET=*value*

Graphics

specifies the length in percent screen units of the offset at the upper end of the primary vertical axis.

VREF=*value-list*

specifies where reference lines perpendicular to the primary vertical axis are to appear on the chart. You must specify the values in the same units used to scale the primary axis. By default, the primary axis is scaled in percent units, but you can specify other units with the SCALE= option. See Output 29.2.3 on page 903 for an illustration.

VREF2=*value-list*

specifies where reference lines perpendicular to the secondary vertical axis are to appear on the chart. You must specify the values in cumulative percent units.

VREFCHAR=*'character'*

Line Printer

specifies the character used to form the lines requested with the VREF= and VREF2= options. The default character is a dash (-).

VREFLABELS=*'label1'...'labeln'*

specifies labels for the lines requested with the VREF= option. The number of labels must equal the number of lines requested. Enclose the labels in quotes. Labels can be up to 16 characters.

VREF2LABELS=*'label1'...'labeln'*

specifies labels for the lines requested with the VREF2= option. The number of labels must equal the number of lines requested. Enclose the labels in quotes. Labels can be up to 16 characters.

VREFLABPOS=*n*
specifies the vertical positioning of the VREFLABELS= and VREF2LABELS= labels. If you specify VREFLABPOS=1, the labels are positioned at the left of the chart, and if you specify VREFLABPOS=2, the labels are positioned at the right. By default, n=1.

WAXIS=*n*
specifies the line thickness (in pixels) for the axes and frame. By default, $n = 1$. *Graphics*

WBARLINE=*n*
specifies the width for bar outlines.

WEIGHT=*variable-list*
specifies weight variables used to construct weighted Pareto charts. The WEIGHT= variables are paired with the process variables in order of specification. The WEIGHT= variables must be numeric, and their values must be nonnegative (noninteger values are permitted). If a WEIGHT= variable is not provided for a process variable, the weights applied to that process variable are assumed to be one. See "Weighted Pareto Charts" on page 888 for computational details.

A WEIGHT= variable is particularly useful for carrying out a Pareto analysis based on *cost* rather than frequency of occurrence. See Example 29.8 on page 914 for an illustration.

WGRID=*n*
specifies the width of the primary chart grid lines. If the WGRID= option is specified the GRID option is not required.

WGRID2=*n*
specifies the width of the secondary chart grid lines. If the WGRID2= option is specified the GRID2 option is not required.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/QC® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1994 pp.

SAS/QC® User's Guide, Version 8

ISBN 1–58025–493–4

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999