Chapter 9

PROBPLOT Statement

Chapter Table of Contents

Chapter 9
PROBPLOT Statement

Overview

The PROBPLOT statement creates a probability plot, which compares ordered values of a variable with percentiles of a specified theoretical distribution such as the normal. If the data distribution matches the theoretical distribution, the points on the plot form a linear pattern. Thus, you can use a probability plot to determine how well a theoretical distribution models a set of measurements.

You can specify one of the following theoretical distributions with the PROBPLOT statement:

- beta
- exponential
- gamma
- three-parameter lognormal
- normal
- two-parameter Weibull
- three-parameter Weibull

You can use options in the PROBPLOT statement to

- specify or estimate shape parameters for the theoretical distribution
- display a reference line corresponding to specified or estimated location and scale parameters for the theoretical distribution
- request graphical enhancements

Note: Probability plots are similar to Q-Q plots, which you can create with the QQPLOT statement (see Chapter 10, “QQPLOT Statement”). Probability plots are preferable for graphical estimation of percentiles, whereas Q-Q plots are preferable for graphical estimation of distribution parameters and capability indices.

Getting Started

The following examples illustrate the basic syntax of the PROBPLOT statement. For complete details of the PROBPLOT statement, see the “Syntax” section on page 281. Advanced examples are provided on the “Examples” section on page 302.

Creating a Normal Probability Plot

See CAPPROB1 in the SAS/QC Sample Library

The diameters of 50 steel rods are measured and saved as values of the variable DISTANCE in the following data set:*

```
data measures;
   input diameter @@;
   label diameter='Diameter in mm';
   datalines;
5.501  5.251  5.404  5.366  5.445
5.576  5.607  5.200  5.977  5.177
5.332  5.399  5.661  5.512  5.252
5.404  5.739  5.525  5.160  5.410
5.823  5.376  5.202  5.470  5.410
5.394  5.146  5.244  5.309  5.480
5.388  5.399  5.360  5.368  5.394
5.248  5.409  5.304  6.239  5.781
5.247  5.907  5.208  5.143  5.304
5.603  5.164  5.209  5.475  5.223
;
```

The process producing the rods is in statistical control, and as a preliminary step in a capability analysis of the process, you decide to check whether the diameters are normally distributed. The following statements create the normal probability plot shown in Figure 9.1:

```
title 'Normal Probability Plot for Diameters';
proc capability data=measures noprint;
   probplot diameter;
run;
```

If you specify the LINEPRINTER option in the PROC CAPABILITY statement, the plot is created using a line printer, as shown in Figure 9.2. Note that the PROBPLOT statement creates a normal probability plot for DIAMETER by default.

The nonlinearity of the point pattern indicates a departure from normality. Since the point pattern is curved with slope increasing from left to right, a theoretical distribution that is skewed to the right, such as a lognormal distribution, should provide a better fit than the normal distribution. This possibility is explored in the next example.

*This data set is analyzed using quantile-quantile plots in Example 10.1 on page 336 and Example 10.2 on page 337.

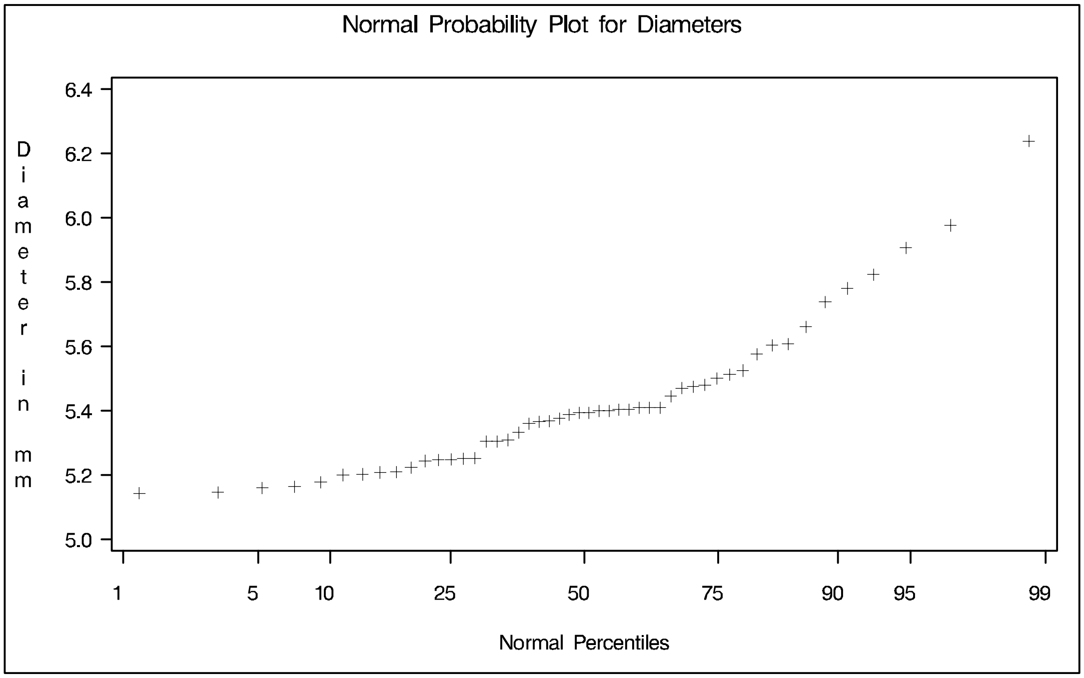


Figure 9.1. Normal Probability Plot Created with Graphics Device

```
                     Normal Probability Plot for Diameters

       --------------------------------------------------------------------
  6.4 +                                                                    |
      |                                                                    |
      |                                                                  + |
  6.2 +                                                                    |
D     |                                                                    |
i     |                                                                    |
a 6.0 +                                                            +       |
m     |                                                         +          |
e     |                                                                    |
t 5.8 +                                                     + +            |
e     |                                                   +                |
r     |                                                  +                 |
  5.6 +                                              +++                   |
i     |                                           +++                      |
n     |                                        +++                         |
  5.4 +                               +++++++++                            |
m     |                          +++++                                     |
m     |                      +++++                                         |
  5.2 +              + ++ +++                                              |
      | +     +  + +                                                       |
      |                                                                    |
  5.0 +                                                                    |
       +---------+----+--------+---------+--------+--------+----+---------+
       1         5   10        25        50       75       90   95        99

                               Normal Percentiles

Observations:    +   (5 Hidden)
```

Figure 9.2. Normal Probability Plot Created with Line Printer

Creating Lognormal Probability Plots

See CAPPROB3 in the SAS/QC Sample Library

When you request a lognormal probability plot, you must specify the shape parameter σ for the lognormal distribution (see Table 9.13 on page 299 for the equation). The value of σ must be positive, and typical values of σ range from 0.1 to 1.0. Alternatively, you can specify that σ is to be estimated from the data.

The following statements illustrate the first approach by creating a series of three lognormal probability plots for the variable DIAMETER introduced in the preceding example:

```
title 'Lognormal Probability Plot for Diameters';
proc capability data=measures noprint;
   probplot diameter / lognormal( sigma = 0.2 0.5 0.8 )
                       href  = 95
                       lhref = 1
                       square ;
run;
```

The LOGNORMAL option requests plots based on the lognormal family of distributions, and the SIGMA= option requests plots for σ equal to 0.2, 0.5, and 0.8. These plots are displayed in Figure 9.3, Figure 9.4, and Figure 9.5, respectively. The value $\sigma = 0.5$ in Figure 9.4 produces the most linear pattern.

The SQUARE option displays the probability plot in a square format, the HREF= option requests a reference line at the 95th percentile, and the LHREF= option specifies the line type for the reference line.

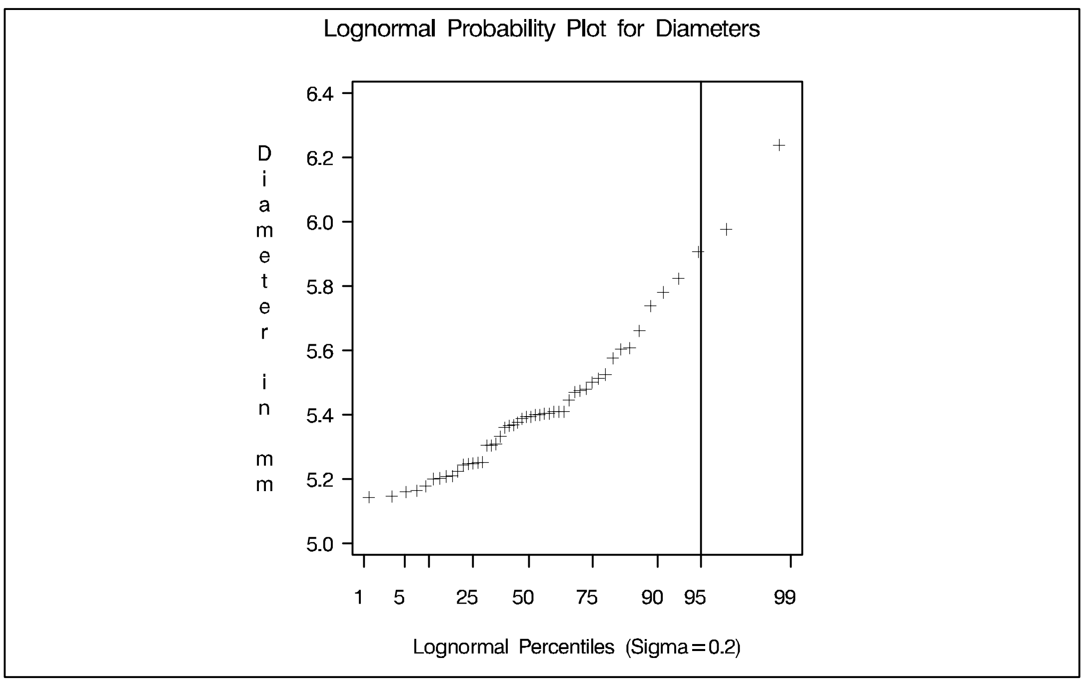


Figure 9.3. Probability Plot Based on Lognormal Distribution with $\sigma = 0.2$

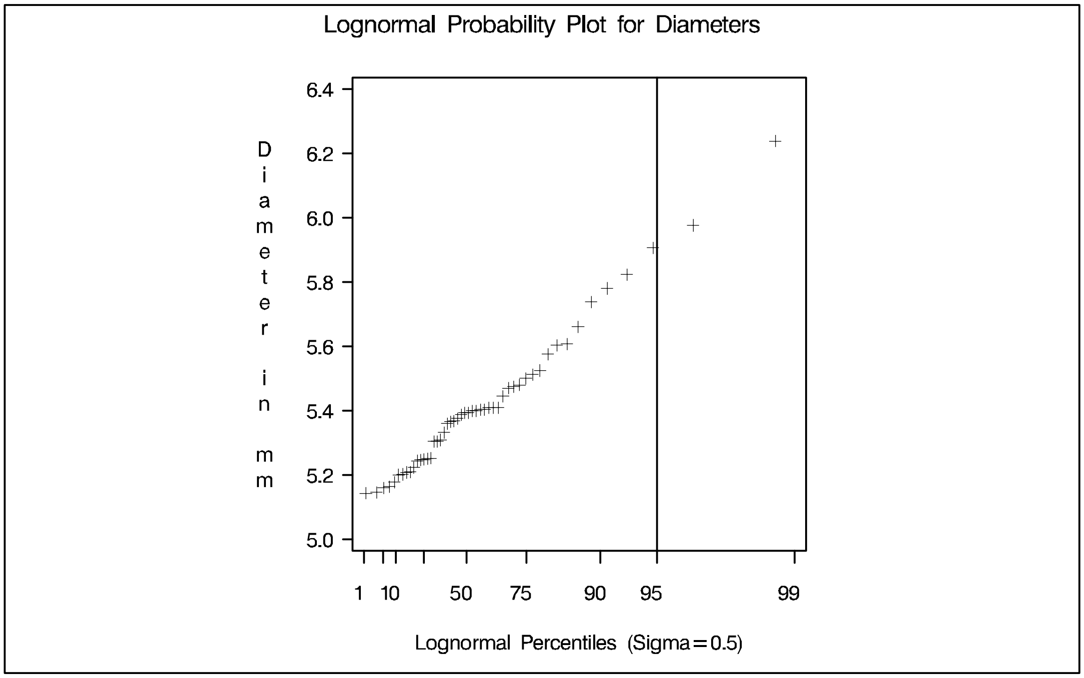


Figure 9.4. Probability Plot Based on Lognormal Distribution with $\sigma = 0.5$

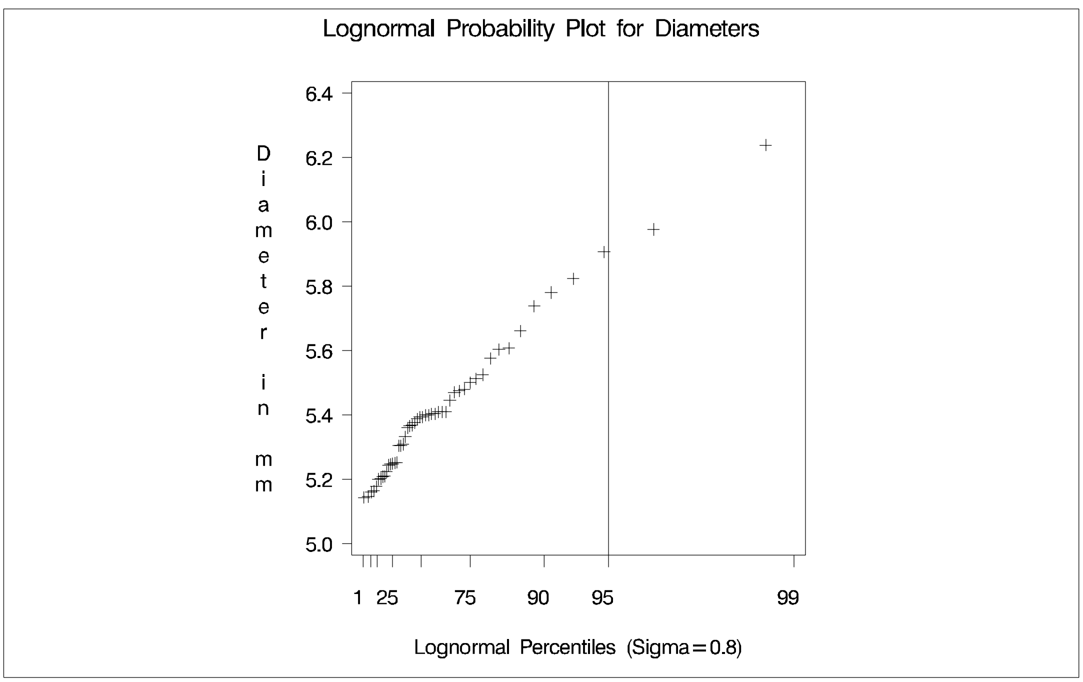


Figure 9.5. Probability Plot Based on Lognormal Distribution with $\sigma = 0.8$

Based on Figure 9.4, the 95^{th} percentile of the diameter distribution is approximately 5.9 mm, since this is the value corresponding to the intersection of the point pattern with the reference line.

The following statements illustrate how you can create a lognormal probability plot for DIAMETER using a local maximum likelihood estimate for σ.

```
title 'Lognormal Probability Plot for Diameters';
proc capability data=measures noprint;
   probplot diameter / lognormal( sigma = est )
                       href  = 95
                       lhref = 1
                       square ;
run;
```

The plot is displayed in Figure 9.6. Note that the maximum likelihood estimate of σ (in this case 0.041) does not necessarily produce the most linear point pattern. This example is continued in Example 9.2 on page 303.

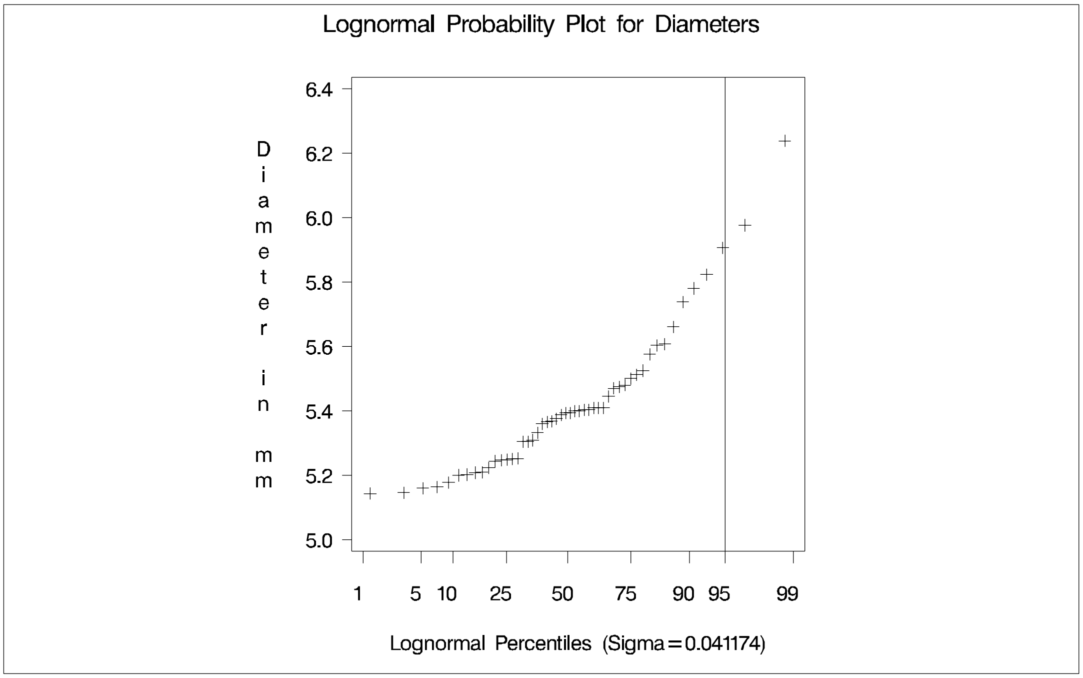


Figure 9.6. Probability Plot Based on Lognormal Distribution with Estimated σ

Syntax

The syntax for the PROBPLOT statement is as follows:

PROBPLOT<*variables* > < **/** *options* >**;**

You can specify the keyword PROB as an alias for PROBPLOT, and you can use any number of PROBPLOT statements in the CAPABILITY procedure. The components of the PROBPLOT statement are described as follows.

variables

are the process variables for which to create probability plots. If you specify a VAR statement, the *variables* must also be listed in the VAR statement. Otherwise, the *variables* can be any numeric variables in the input data set. If you do not specify a list of *variables*, then by default the procedure creates a probability plot for each variable listed in the VAR statement, or for each numeric variable in the DATA= data set if you do not specify a VAR statement. For example, each of the following PROBPLOT statements produces two probability plots, one for LENGTH and one for WIDTH:

```
proc capability data=measures;
   var length width;
   probplot;
run;

proc capability data=measures;
   probplot length width;
run;
```

options

specify the theoretical distribution for the plot or add features to the plot. If you specify more than one variable, the *options* apply equally to each variable. Specify all *options* after the slash (/) in the PROBPLOT statement. You can specify only one *option* naming the distribution in each PROBPLOT statement, but you can specify any number of other *options*. The distributions available are the beta, exponential, gamma, lognormal, normal, two-parameter Weibull, and three-parameter Weibull. By default, the procedure produces a plot for the normal distribution.

In the following example, the NORMAL option requests a normal probability plot for each variable, while the MU= and SIGMA= *normal-options* request a distribution reference line corresponding to the normal distribution with $\mu = 10$ and $\sigma = 0.3$. The SQUARE option displays the plot in a square frame, and the CTEXT= option specifies the text color.

```
proc capability data=measures;
   probplot length1 length2 / normal(mu=10 sigma=0.3)
                              square
                              ctext=blue;
run;
```

Summary of Options

The following tables list the PROBPLOT statement *options* by function. For complete descriptions, see the "Dictionary of Options" section on page 285.

Distribution Options

Table 9.1 summarizes the options for requesting a specific theoretical distribution.

Table 9.1. Keywords to Select a Theoretical Distribution

BETA(*beta-options*)	specifies beta probability plot for shape parameters α, β specified with mandatory ALPHA= and BETA= *beta-options*
EXPONENTIAL(*exponential-options*)	specifies exponential probability plot
GAMMA(*gamma-options*)	specifies gamma probability plot for shape parameter α specified with mandatory ALPHA= *gamma-option*
LOGNORMAL(*lognormal-options*)	specifies lognormal probability plot for shape parameter σ specified with mandatory SIGMA= *lognormal-option*
NORMAL(*normal-options*)	specifies normal probability plot
WEIBULL(*Weibull-options*)	specifies three-parameter Weibull probability plot for shape parameter c specified with mandatory C= *Weibull-option*
WEIBULL2(*Weibull2-options*)	specifies two-parameter Weibull probability plot

Table 9.2 through Table 9.9 summarize options that specify distribution parameters and control the display of a distribution reference line. Specify these options in parentheses after the distribution option. For example, the following statements use the NORMAL option to request a normal probability plot with a distribution reference line:

```
proc capability data=measures;
   probplot length / normal(mu=10 sigma=0.3 color=red);
run;
```

The MU= and SIGMA= *normal-options* display a distribution reference line that corresponds to the normal distribution with mean $\mu_0 = 10$ and standard deviation $\sigma_0 = 0.3$, and the COLOR= *normal-option* specifies the color for the line.

Table 9.2. Reference Line Options Available with All Distributions

COLOR=*color*	specifies color of distribution reference line
L=*linetype*	specifies line type of distribution reference line
SYMBOL='*character*'	specifies plotting character for line printer
W=n	specifies width of distribution reference line

Table 9.3. Beta-Options

ALPHA=*value-list*|EST	specifies mandatory shape parameter α
BETA=*value-list*|EST	specifies mandatory shape parameter β
SIGMA=*value*|EST	specifies σ_0 for distribution reference line
THETA=*value*|EST	specifies θ_0 for distribution reference line

Table 9.4. Exponential-Options

SIGMA=*value*|EST	specifies σ_0 for distribution reference line
THETA=*value*|EST	specifies θ_0 for distribution reference line

Table 9.5. Gamma-Options

ALPHA=*value-list*|EST	specifies mandatory shape parameter α
SIGMA=*value*|EST	specifies σ_0 for distribution reference line
THETA=*value*|EST	specifies θ_0 for distribution reference line

Table 9.6. Lognormal-Options

SIGMA=*value-list*|EST	specifies mandatory shape parameter σ
SLOPE=*value*|EST	specifies slope of distribution reference line
THETA=*value*|EST	specifies θ_0 for distribution reference line
ZETA=*value*|EST	specifies ζ_0 for distribution reference line (slope is $\exp(\zeta_0)$)

Table 9.7. Normal-Options

MU=*value*|EST	specifies μ_0 for distribution reference line
SIGMA=*value*|EST	specifies σ_0 for distribution reference line

Table 9.8. Weibull-Options

C=*value-list*|EST	specifies mandatory shape parameter c
SIGMA=*value*|EST	specifies σ_0 for distribution reference line
THETA=*value*|EST	specifies θ_0 for distribution reference line

Table 9.9. Weibull2-Options

C=*value*|EST	specifies c_0 for distribution reference line (slope is $1/c_0$)
SIGMA=*value*|EST	specifies σ_0 for distribution reference line (intercept is $\log(\sigma_0)$)
SLOPE=*value*|EST	specifies slope of distribution reference line
THETA=*value*	specifies known lower threshold θ_0

General Options

Table 9.10 through Table 9.12 list options that control the appearance of the plots.

Table 9.10. General Plot Layout Options

GRID	specifies reference lines perpendicular to the percentile axis at major tick marks
HREF=*value-list*	specifies reference lines perpendicular to the horizontal axis
HREFLABELS= ’*label1*’ . . . ’*labeln*’	specifies line labels for HREF= lines
LEGEND=*name* | NONE	identifies LEGEND statement
NADJ=*value*	adjusts sample size (N) when computing percentiles
NOFRAME	suppresses frame around plotting area
NOLEGEND	suppresses legend
NOLINELEGEND	suppresses distribution reference line information in legend
NOSPECLEGEND	suppresses specifications information in legend
PCTLMINOR	requests minor tick marks for percentile axis
PCTLORDER=*value-list*	specifies tick mark labels for percentile axis
RANKADJ=*value*	adjusts ranks when computing percentiles
ROTATE	switches horizontal and vertical axes
SQUARE	displays plot in square format
VREF=*value-list*	specifies reference lines perpendicular to the vertical axis
VREFLABELS= ’*label1*’ . . . ’*labeln*’	specifies line labels for VREF= lines

Table 9.11. Options to Enhance Plots Produced on Line Printers

GRIDCHAR='*character*'	specifies character for GRID lines
HREFCHAR='*character*'	specifies character for HREF= lines
NOOBSLEGEND	suppresses legend for hidden points
PROBSYMBOL='*character*'	specifies character for plotted points
VREFCHAR='*character*'	specifies character for VREF= lines

Table 9.12. Options to Enhance Plots Produced on Graphics Devices

ANNOTATE=*SAS-data-set*	provides an annotate data set
CAXIS=*color*	specifies color for axis
CFRAME=*color*	specifies color for frame
CHREF=*color*	specifies color for HREF= lines
CTEXT=*color*	specifies color for text
CVREF=*color*	specifies color for VREF= lines
DESCRIPTION='*string*'	specifies description for graphics catalog member
FONT=*font*	specifies software font for text
HAXIS=*name*	identifies AXIS statement for horizontal axis
HMINOR=n	specifies number of minor tick marks on horizontal axis
LGRID=*linetype*	specifies line type for GRID lines
LHREF=*linetype*	specifies line type for HREF= lines
LVREF=*linetype*	specifies line type for VREF= lines
NAME='*string*'	specifies name for plot in graphics catalog
VAXIS=*name*	identifies AXIS statement for vertical axis
VMINOR=*value*	specifies number of minor tick marks on vertical axis

Dictionary of Options

The following entries provide detailed descriptions of options for the PROBPLOT statement. The marginal notes *Graphics* and *Line Printer* identify options that apply only to graphics devices and line printers, respectively.

ALPHA=*value-list***|EST**

specifies values for a mandatory shape parameter α $(\alpha > 0)$ for probability plots requested with the BETA and GAMMA options. A plot is created for each value specified. For examples, see the entries for the BETA and GAMMA options. If you specify ALPHA=EST, a maximum likelihood estimate is computed for α.

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

Graphics

specifies an input data set containing annotate variables as described in *SAS/GRAPH Software: Reference*. You can use this data set to add features to the plot. The ANNOTATE= data set specified in the PROBPLOT statement is used for all plots created by the statement. You can also specify an ANNOTATE= data set in the PROC CAPABILITY statement to enhance all plots created by the procedure; for more information, see "ANNOTATE= Data Sets" on page 31.

BETA(ALPHA=*value-list***|EST BETA=***value-list***|EST** <*beta-options* >**)**

creates a beta probability plot for each combination of the shape parameters α and β given by the mandatory ALPHA= and BETA= options. If you specify ALPHA=EST and BETA=EST, a plot is created based on maximum likelihood estimates for α and β. In the following examples, the first PROBPLOT statement produces one plot, the second statement produces four plots, the third statement produces six plots, and the fourth statement produces one plot:

```
proc capability data=measures;
   probplot width / beta(alpha=2 beta=2);
   probplot width / beta(alpha=2 3 beta=1 2);
   probplot width / beta(alpha=2 to 3 beta=1 to 2 by 0.5);
   probplot width / beta(alpha=est beta=est);
run;
```

To create the plot, the observations are ordered from smallest to largest, and the i^{th} ordered observation is plotted against the quantile $B^{-1}_{\alpha\beta}\left(\frac{i-0.375}{n+0.25}\right)$, where $B^{-1}_{\alpha\beta}(\cdot)$ is the inverse normalized incomplete beta function, n is the number of nonmissing observations, and α and β are the shape parameters of the beta distribution. The horizontal axis is scaled in percentile units.

The point pattern on the plot for ALPHA=α and BETA=β tends to be linear with intercept* θ and slope σ if the data are beta distributed with the specific density function

$$p(x) = \begin{cases} \frac{(x-\theta)^{\alpha-1}(\theta+\sigma-x)^{\beta-1}}{B(\alpha,\beta)\sigma^{(\alpha+\beta-1)}} & \text{for } \theta < x < \theta+\sigma \\ 0 & \text{for } x \le \theta \text{ or } x \ge \theta+\sigma \end{cases}$$

where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and

$\theta =$ lower threshold parameter
$\sigma =$ scale parameter $(\sigma > 0)$
$\alpha =$ first shape parameter $(\alpha > 0)$
$\beta =$ second shape parameter $(\beta > 0)$

To obtain graphical estimates of α and β, specify lists of values for the ALPHA= and BETA= options, and select the combination of α and β that most nearly linearizes the point pattern.

*The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, "QQPLOT Statement."

To assess the point pattern, you can add a diagonal distribution reference line corresponding to θ_0 and σ_0 with the *beta-options* THETA=θ_0 and SIGMA=σ_0. Alternatively, you can add a line corresponding to estimated values of θ_0 and σ_0 with the *beta-options* THETA=EST and SIGMA=EST. Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   probplot width / beta(alpha=2 beta=3 theta=4 sigma=5);
run;
```

Agreement between the reference line and the point pattern indicates that the beta distribution with parameters α, β, θ_0 and σ_0 is a good fit. You can specify the SCALE= option as an alias for the SIGMA= option and the THRESHOLD= option as an alias for the THETA= option.

BETA=*value-list***|EST**

specifies values for the shape parameter β $(\beta > 0)$ for probability plots requested with the BETA distribution option. A plot is created for each value specified with the BETA= option. If you specify BETA=EST, a maximum likelihood estimate is computed for β. For examples, see the preceding entry for the BETA option.

C=*value(-list)***|EST**

specifies the shape parameter c $(c > 0)$ for probability plots requested with the WEIBULL and WEIBULL2 options. You must specify C= as a *Weibull-option* with the WEIBULL option; in this situation it accepts a list of values, or if you specify C=EST, a maximum likelihood estimate is computed for c. You can optionally specify C=*value* or C=EST as a *Weibull2-option* with the WEIBULL2 option to request a distribution reference line; in this situation, you must also specify SIGMA=*value* or SIGMA=EST.

For example, the first PROBPLOT statement below creates three three-parameter Weibull plots corresponding to the shape parameters $c = 1$, $c = 2$, and $c = 3$. The second PROBPLOT statement creates a single three-parameter Weibull plot corresponding to an estimated value of c. The third PROBPLOT statement creates a single two-parameter Weibull plot with a distribution reference line corresponding to $c_0 = 2$ and $\sigma_0 = 3$.

```
proc capability data=measures;
   probplot width / weibull(c=1 2 3);
   probplot width / weibull(c=est);
   probplot width / weibull2(c=2 sigma=3);
run;
```

CAXIS=*color*
CAXES=*color*

specifies the color used for the axes. This option overrides any COLOR= specifications in an AXIS statement. The default is the first color in the device color list. *Graphics*

CFRAME=*color*

specifies the fill color for the area enclosed by the axes and frame. This area is not filled by default. *Graphics*

CHREF=*color*

Graphics

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

COLOR=*color*

Graphics

specifies the color for a diagonal distribution reference line. Specify the COLOR= option in parentheses following a distribution option keyword. The default is the first color in the device color list.

CTEXT=*color*

Graphics

specifies the color for tick mark values and axis labels. The default is the color specified for the CTEXT= option in the most recent GOPTIONS statement.

CVREF=*color*

Graphics

specifies the color for reference lines requested by the VREF= option. The default is the first color in the device color list.

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

Graphics

specifies a description, up to 40 characters, that appears in the PROC GREPLAY master menu. The default string is the variable name.

EXPONENTIAL<**(***exponential-options***)**>
EXP(<*exponential-options*>**)**

creates an exponential probability plot. To create the plot, the observations are ordered from smallest to largest, and the i^{th} ordered observation is plotted against the quantile $-\log\left(1 - \frac{i-0.375}{n+0.25}\right)$, where n is the number of nonmissing observations. The horizontal axis is scaled in percentile units.

The point pattern on the plot tends to be linear with intercept* θ and slope σ if the data are exponentially distributed with the specific density function

$$p(x) = \begin{cases} \frac{1}{\sigma}\exp\left(-\frac{x-\theta}{\sigma}\right) & \text{for } x \geq \theta \\ 0 & \text{for } x < \theta \end{cases}$$

where θ is a threshold parameter, and σ is a positive scale parameter.

To assess the point pattern, you can add a diagonal distribution reference line corresponding to θ_0 and σ_0 with the *exponential-options* THETA=θ_0 and SIGMA=σ_0. Alternatively, you can add a line corresponding to estimated values of θ_0 and σ_0 with the *exponential-options* THETA=EST and SIGMA=EST. Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   probplot width / exponential(theta=4 sigma=5);
run;
```

Agreement between the reference line and the point pattern indicates that the exponential distribution with parameters θ_0 and σ_0 is a good fit. You can specify the

*The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, "QQPLOT Statement."

SCALE= option as an alias for the SIGMA= option and the THRESHOLD= option as an alias for the THETA= option.

FONT=*font*

specifies a software font for horizontal and vertical reference line labels and axis labels. You can also specify fonts for axis labels in an AXIS statement. The FONT= font takes precedence over the FTEXT= font you specify in the GOPTIONS statement. Hardware characters are used by default. *Graphics*

GAMMA(ALPHA=*value-list***|EST** <*gamma-options*> **)**

creates a gamma probability plot for each value of the shape parameter α given by the mandatory ALPHA= option. If you specify ALPHA=EST, a plot is created based on a maximum likelihood estimate for α.

For example, the first PROBPLOT statement below creates three plots corresponding to $\alpha = 0.4$, $\alpha = 0.5$, and $\alpha = 0.6$. The second PROBPLOT statement creates a single plot.

```
proc capability data=measures;
   probplot width / gamma(alpha=0.4 to 0.6 by 0.2);
   probplot width / gamma(alpha=est);
run;
```

To create the plot, the observations are ordered from smallest to largest, and the i^{th} ordered observation is plotted against the quantile $G_\alpha^{-1}\left(\frac{i-0.375}{n+0.25}\right)$, where $G_\alpha^{-1}(\cdot)$ is the inverse normalized incomplete gamma function, n is the number of nonmissing observations, and α is the shape parameter of the gamma distribution. The horizontal axis is scaled in percentile units.

The point pattern on the plot for ALPHA=α tends to be linear with intercept* θ and slope σ if the data are gamma distributed with the specific density function

$$p(x) = \begin{cases} \frac{1}{\sigma\Gamma(\alpha)}\left(\frac{x-\theta}{\sigma}\right)^{\alpha-1}\exp\left(-\frac{x-\theta}{\sigma}\right) & \text{for } x > \theta \\ 0 & \text{for } x \le \theta \end{cases}$$

where

θ = threshold parameter
σ = scale parameter $(\sigma > 0)$
α = shape parameter $(\alpha > 0)$

To obtain a graphical estimate of α, specify a list of values for the ALPHA= option, and select the value that most nearly linearizes the point pattern.

To assess the point pattern, you can add a diagonal distribution reference line corresponding to θ_0 and σ_0 with the *gamma-options* THETA=θ_0 and SIGMA=σ_0. Alternatively, you can add a line corresponding to estimated values of θ_0 and σ_0 with the

*The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, “QQPLOT Statement.”

gamma-options THETA=EST and SIGMA=EST. Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   probplot width / gamma(alpha=2 theta=3 sigma=4);
run;
```

Agreement between the reference line and the point pattern indicates that the gamma distribution with parameters α, θ_0 and σ_0 is a good fit. You can specify the SCALE= option as an alias for the SIGMA= option and the THRESHOLD= option as an alias for the THETA= option.

GRID
draws reference lines perpendicular to the percentile axis at major tick marks.

GRIDCHAR='*character***'**
Line Printer specifies the character used to form the lines requested by the GRID option for a line printer. The default is the vertical bar (|).

HAXIS=*name*
Graphics specifies the name of an AXIS statement describing the horizontal axis.

HMINOR=*n*
HM=*n*
Graphics specifies the number of minor tick marks between each major tick mark on the horizontal axis. Minor tick marks are not labeled. The default is 0.

HREF=*value-list*
draws reference lines perpendicular to the horizontal axis at the values specified. For an example, see Output 9.2.1 on page 304.

HREFCHAR='*character***'**
Line Printer specifies the character used to form the reference lines requested by the HREF= option for a line printer. The default is the vertical bar (|).

HREFLABELS='*label1***'** . . . **'***labeln***'**
HREFLABEL='*label1***'** . . . **'***labeln***'**
HREFLAB='*label1***'** . . . **'***labeln***'**
specifies labels for the reference lines requested by the HREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters. For an example, see Output 9.2.1 on page 304.

L=*linetype*
Graphics specifies the line type for a diagonal distribution reference line. Specify the L= option in parentheses after a distribution option keyword, as illustrated in the entry for the LOGNORMAL option. The default is 1, which produces a solid line.

LEGEND=*name* **| NONE**
specifies the name of a LEGEND statement describing the legend for specification limit reference lines and fitted curves. Specifying LEGEND=NONE is equivalent to specifying the NOLEGEND option.

LGRID=*linetype*

specifies the line type for the reference lines requested by the GRID option. The default is 1, which produces solid lines. *Graphics*

LHREF=*linetype*
LH=*linetype*

specifies the line type for reference lines requested by the HREF= option. For an example, see Output 9.2.1 on page 304. The default is 2, which produces a dashed line. *Graphics*

LOGNORMAL(SIGMA=*value-list***|EST** $<$*lognormal-options* $>$**)**
LNORM(SIGMA=*value-list***|EST** $<$*lognormal-options* $>$**)**

creates a lognormal probability plot for each value of the shape parameter σ given by the mandatory SIGMA= option or its alias, the SHAPE= option. If you specify SIGMA=EST, a plot is created based on a maximum likelihood estimate for σ.

For example, the first PROBPLOT statement below produces two plots, and the second PROBPLOT statement produces a single plot:

```
proc capability data=measures;
   probplot width / lognormal(sigma=1.5 2.5 l=2);
   probplot width / lognormal(sigma=est);
run;
```

To create the plot, the observations are ordered from smallest to largest, and the i^{th} ordered observation is plotted against the quantile $\exp\left(\sigma\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)\right)$, where $\Phi^{-1}(\cdot)$ is the inverse standard cumulative normal distribution, n is the number of nonmissing observations, and σ is the shape parameter of the lognormal distribution. The horizontal axis is scaled in percentile units.

The point pattern on the plot for SIGMA=σ tends to be linear with intercept* θ and slope $\exp(\zeta)$ if the data are lognormally distributed with the specific density function

$$p(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}(x-\theta)} \exp\left(-\frac{(\log(x-\theta)-\zeta)^2}{2\sigma^2}\right) & \text{for } x > \theta \\ 0 & \text{for } x \le \theta \end{cases}$$

where

$\theta =$ threshold parameter
$\zeta =$ scale parameter
$\sigma =$ shape parameter $(\sigma > 0)$

To obtain a graphical estimate of σ, specify a list of values for the SIGMA= option, and select the value that most nearly linearizes the point pattern.

To assess the point pattern, you can add a diagonal distribution reference line corresponding to θ_0 and ζ_0 with the *lognormal-options* THETA=θ_0 and ZETA=ζ_0. Alternatively, you can add a line corresponding to estimated values of θ_0 and ζ_0 with the *lognormal-options* THETA=EST and ZETA=EST.

*The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, “QQPLOT Statement.”

Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   probplot width / lognormal(sigma=2 theta=3 zeta=0);
run;
```

Agreement between the reference line and the point pattern indicates that the lognormal distribution with parameters σ, θ_0, and ζ_0 is a good fit. See Example 9.2 on page 303 for an example.

You can specify the THRESHOLD= option as an alias for the THETA= option and the SCALE= option as an alias for the ZETA= option.

LVREF=*linetype*

Graphics

specifies the line type for reference lines requested by the VREF= option. For an example, see Output 9.2.1 on page 304. The default is 2, which produces a dashed line.

MU=*value***|EST**

specifies the mean μ_0 for a normal probability plot requested with the NORMAL option. The MU= and SIGMA= *normal-options* must be specified together, and they request a distribution reference line as illustrated in Example 9.1 on page 302. Specify MU=EST to request a distribution reference line with μ_0 equal to the sample mean.

NADJ=*value*

specifies the adjustment value added to the sample size in the calculation of theoretical percentiles. The default is $\frac{1}{4}$, as recommended by Blom (1958). Also refer to Chambers and others (1983) for additional information.

NAME='*string***'**

Graphics

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

NOFRAME

suppresses the frame around the area bounded by the axes.

NOLEGEND
LEGEND=NONE

suppresses legends for specification limits, fitted curves, distribution lines, and hidden observations.

NOLINELEGEND
NOLINEL

suppresses the legend for the optional distribution reference line.

NOOBSLEGEND
NOOBSL

Line Printer

suppresses the legend that indicates the number of hidden observations.

NORMAL$<$**(***normal-options***)**$>$
NORM$<$**(***normal-options***)**$>$

creates a normal probability plot. This is the default if you do not specify a distribution option. To create the plot, the observations are ordered from smallest to largest, and the i^{th} ordered observation is plotted against the quantile $\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)$, where $\Phi^{-1}(\cdot)$ is the inverse cumulative standard normal distribution, and n is the number of nonmissing observations. The horizontal axis is scaled in percentile units.

The point pattern on the plot tends to be linear with intercept* μ and slope σ if the data are normally distributed with the specific

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{for all } x$$

where μ is the mean and σ is the standard deviation ($\sigma > 0$).

To assess the point pattern, you can add a diagonal distribution reference line corresponding to μ_0 and σ_0 with the *normal-options* MU=μ_0 and SIGMA=σ_0. Alternatively, you can add a line corresponding to estimated values of μ_0 and σ_0 with the *normal-options* THETA=EST and SIGMA=EST; the estimates of μ_0 and $]sigma_0$ are the sample mean and sample standard deviation.

Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   probplot length / normal(mu=10 sigma=0.3);
   probplot length / normal(mu=est sigma=est);
run;
```

Agreement between the reference line and the point pattern indicates that the normal distribution with parameters μ_0 and σ_0 is a good fit.

NOSPECLEGEND
NOSPECL

suppresses the legend for specification limit reference lines.

PCTLMINOR

requests minor tick marks for the percentile axis. See Output 9.2.1 on page 304 for an example.

PCTLORDER=*value-list*

specifies the tick mark values labeled on the theoretical percentile axis. Since the values are percentiles, the labels must be between 0 and 100, exclusive. The values must be listed in increasing order and must cover the plotted percentile range. Otherwise, a default list is used. For example, consider the following:

```
proc capability data=measures;
   probplot length / pctlorder=1 10 25 50 75 90 99;
run;
```

Note that the ORDER= option in the AXIS statement is not supported by the PROBPLOT statement.

*The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, “QQPLOT Statement.”

PROBSYMBOL='*character***'**

Line Printer

specifies the character used to mark the points when the plot is produced on a line printer. The default is the plus sign (+).

RANKADJ=*value*

specifies the adjustment value added to the ranks in the calculation of theoretical percentiles. The default is $-\frac{3}{8}$, as recommended by Blom (1958). Also refer to Chambers and others (1983) for additional information.

ROTATE

Graphics

switches the horizontal and vertical axes so that the theoretical percentiles are plotted vertically while the data are plotted horizontally. Regardless of whether the plot has been rotated, horizontal axis options (such as HAXIS=) still refer to the horizontal axis, and vertical axis options (such as VAXIS=) still refer to the vertical axis. All other options that depend on axis placement adjust to the rotated axes.

SCALE=*value***|EST**

is an alias for the SIGMA= option with the BETA, EXPONENTIAL, GAMMA, WEIBULL and WEIBULL2 options and for the ZETA= option with the LOGNORMAL option. See the entries for the SIGMA= and ZETA= options.

SHAPE=*value-list***|EST**

is an alias for the ALPHA= option with the GAMMA option, for the SIGMA= option with the LOGNORMAL option, and for the C= option with the WEIBULL and WEIBULL2 options. See the entries for the ALPHA=, C=, and SIGMA= options.

SIGMA=*value-list***|EST**

specifies the value of the parameter σ, where $\sigma > 0$. Alternatively, you can specify SIGMA=EST to request a maximum likelihood estimate for σ_0. The interpretation and use of the SIGMA= option depend on the distribution option with which it is specified, as indicated by the following table:

Distribution Option	Use of the SIGMA= Option
BETA EXPONENTIAL GAMMA WEIBULL	THETA=θ_0 and SIGMA=σ_0 request a distribution reference line corresponding to θ_0 and σ_0.
LOGNORMAL	SIGMA=$\sigma_1 \ldots \sigma_n$ requests n probability plots with shape parameters $\sigma_1 \ldots \sigma_n$. The SIGMA= option must be specified.
NORMAL	MU=μ_0 and SIGMA=σ_0 request a distribution reference line corresponding to μ_0 and σ_0. SIGMA=EST requests a line with σ_0 equal to the sample standard deviation.
WEIBULL2	SIGMA=σ_0 and C=c_0 request a distribution reference line corresponding to σ_0 and c_0.

In the following example, the first PROBPLOT statement requests a normal plot with a distribution reference line corresponding to $\mu_0 = 5$ and $\sigma_0 = 2$, and the second PROBPLOT statement requests a lognormal plot with shape parameter $\sigma = 3$:

```
proc capability data=measures;
   probplot length / normal(mu=5 sigma=2);
   probplot width  / lognormal(sigma=3);
run;
```

SLOPE=*value***|EST**

specifies the slope* for a distribution reference line requested with the LOGNORMAL and WEIBULL2 options.

When you use the SLOPE= option with the LOGNORMAL option, you must also specify a threshold parameter value θ_0 with the THETA= *lognormal-option* to request the line. The SLOPE= option is an alternative to the ZETA= *lognormal-option* for specifying ζ_0, since the slope is equal to $\exp(\zeta_0)$.

When you use the SLOPE= option with the WEIBULL2 option, you must also specify a scale parameter value σ_0 with the SIGMA= *Weibull2-option* to request the line. The SLOPE= option is an alternative to the C= *Weibull2-option* for specifying c_0, since the slope is equal to $1/c_0$. See "Location and Scale Parameters" on page 300.

For example, the first and second PROBPLOT statements below produce the same set of probability plots as the third and fourth PROBPLOT statements:

```
proc capability data=measures;
   probplot width / lognormal(sigma=2 theta=0 zeta=0);
   probplot width / weibull2(sigma=2 theta=0 c=0.25);
   probplot width / lognormal(sigma=2 theta=0 slope=1);
   probplot width / weibull2(sigma=2 theta=0 slope=4);
run;
```

SQUARE

displays the probability plot in a square frame. For an example, see Output 9.2.1 on page 304. The default is a rectangular frame.

SYMBOL='*character***'**

specifies the character used to display the distribution reference line when the plot is created using a line printer. The default character is the first letter of the distribution option keyword. Line Printer

THETA=*value***|EST**

specifies the lower threshold parameter θ for plots requested with the BETA, EXPONENTIAL, GAMMA, LOGNORMAL, WEIBULL, and WEIBULL2 options. When used with the WEIBULL2 option, the THETA= option specifies the known lower threshold θ_0, for which the default is 0. When used with the other distribution options, the THETA= option specifies θ_0 for a distribution reference line; alternatively in this situation, you can specify THETA=EST to request a maximum likelihood estimate for θ_0. To request the line, you must also specify a scale parameter. See Output 9.2.1 on page 304 for an example of the THETA= option with a lognormal probability plot.

*The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, "QQPLOT Statement."

THRESHOLD=*value*
is an alias for the THETA= option.

VAXIS=*name*
Graphics
specifies the name of an AXIS statement describing the vertical axis, as illustrated by Output 9.1.1 on page 303.

VMINOR=*n*
VM=*n*
Graphics
specifies the number of minor tick marks between each major tick mark on the vertical axis. Minor tick marks are not labeled. The default is 0.

VREF=*value-list*
draws reference lines perpendicular to the vertical axis at the values specified. See Output 9.2.1 on page 304 for an example.

VREFCHAR='*character***'**
Line Printer
specifies the character used to form the lines requested by the VREF= option for a line printer. The default is the hyphen (-).

VREFLABELS='*label1***'** . . . **'***labeln***'**
VREFLABEL='*label1***'** . . . **'***labeln***'**
VREFLAB='*label1***'** . . . **'***labeln***'**
specifies labels for the lines requested by the VREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

W=*n*
Graphics
specifies the width in pixels for a diagonal distribution reference line. Specify the W= option in parentheses after a distribution option keyword. For an example, see the entry for the WEIBULL option. The default is 1.

WEIBULL(C=*value-list***|EST** *<Weibull-options >***)**
WEIB(C=*value-list <Weibull-options >***)**
creates a three-parameter Weibull probability plot for each value of the shape parameter c given by the mandatory C= option or its alias, the SHAPE= option. If you specify C=EST, a plot is created based on a maximum likelihood estimate for c. In the following example, the first PROBPLOT statement creates four plots, and the second PROBPLOT statement creates a singlel plot:

```
proc capability data=measures;
   probplot width / weibull(c=1.8 to 2.4 by 0.2 w=2);
   probplot width / weibull(c=est);
run;
```

To create the plot, the observations are ordered from smallest to largest, and the i^{th} ordered observation is plotted against the quantile $\left(-\log\left(1-\frac{i-0.375}{n+0.25}\right)\right)^{\frac{1}{c}}$, where n is the number of nonmissing observations, and c is the Weibull distribution shape parameter. The horizontal axis is scaled in percentile units.

The point pattern on the plot for C=c tends to be linear with intercept[*] θ and slope σ if the data are Weibull distributed with the specific density function

$$p(x) = \begin{cases} \frac{c}{\sigma}\left(\frac{x-\theta}{\sigma}\right)^{c-1} \exp\left(-\left(\frac{x-\theta}{\sigma}\right)^{c}\right) & \text{for } x > \theta \\ 0 & \text{for } x \leq \theta \end{cases}$$

where

$\theta =$ threshold parameter
$\sigma =$ scale parameter $(\sigma > 0)$
$c =$ shape parameter $(c > 0)$

To obtain a graphical estimate of c, specify a list of values for the C= option, and select the value that most nearly linearizes the point pattern.

To assess the point pattern, you can add a diagonal distribution reference line corresponding to θ_0 and σ_0 with the *Weibull-options* THETA=θ_0 and SIGMA=σ_0. Alternatively, you can add a line corresponding to estimated values of θ_0 and σ_0 with the *Weibull-options* THETA=EST and SIGMA=EST. Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   probplot width / weibull(c=2 theta=3 sigma=4);
run;
```

Agreement between the reference line and the point pattern indicates that the Weibull distribution with parameters c, θ_0, and σ_0 is a good fit. You can specify the SCALE= option as an alias for the SIGMA= option and the THRESHOLD= option as an alias for the THETA= option.

WEIBULL2$<$**(***Weibull2-options***)**$>$
W2$<$**(***Weibull2-options***)**$>$

creates a two-parameter Weibull probability plot. You should use the WEIBULL2 option when your data have a *known* lower threshold θ_0. You can specify the threshold value θ_0 with the THETA= *Weibull2-option* or its alias, the THRESHOLD= *Weibull2-option*. The default is $\theta_0 = 0$.

To create the plot, the observations are ordered from smallest to largest, and the log of the shifted i^{th} ordered observation $x_{(i)}$, denoted by $\log(x_{(i)} - \theta_0)$, is plotted against the quantile $\log\left(-\log\left(1 - \frac{i-0.375}{n+0.25}\right)\right)$, where n is the number of nonmissing observations. The horizontal axis is scaled in percentile units. Note that the C= shape parameter option is not mandatory with the WEIBULL2 option.

The point pattern on the plot for THETA=θ_0 tends to be linear with intercept $\log(\sigma)$ and slope $\frac{1}{c}$ if the data are Weibull distributed with the specific density function

$$p(x) = \begin{cases} \frac{c}{\sigma}\left(\frac{x-\theta_0}{\sigma}\right)^{c-1} \exp\left(-\left(\frac{x-\theta_0}{\sigma}\right)^{c}\right) & \text{for } x > \theta_0 \\ 0 & \text{for } x \leq \theta_0 \end{cases}$$

[*]The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, “QQPLOT Statement.”

where

θ_0 = known lower threshold
σ = scale parameter ($\sigma > 0$)
c = shape parameter ($c > 0$)

An advantage of the two-parameter Weibull plot over the three-parameter Weibull plot is that the parameters c and σ can be estimated from the slope and intercept of the point pattern. A disadvantage is that the two-parameter Weibull distribution applies only in situations where the threshold parameter is known.

To assess the point pattern, you can add a diagonal distribution reference line corresponding to σ_0 and c_0 with the *Weibull2-options* SIGMA=σ_0 and C=c_0. Alternatively, you can add a distribution reference line corresponding to estimated values of σ_0 and c_0 with the *Weibull2-options* SIGMA=EST and C=EST. Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   probplot width / weibull2(theta=3 sigma=4 c=2);
run;
```

Agreement between the distribution reference line and the point pattern indicates that the Weibull distribution with parameters c_0, θ_0 and σ_0 is a good fit. You can specify the SCALE= option as an alias for the SIGMA= option and the SHAPE= option as an alias for the C= option.

ZETA=*value***|EST**
specifies a value for the scale parameter ζ for lognormal probability plots requested with the LOGNORMAL option. Specify THETA=θ_0 and ZETA=ζ_0 to request a distribution reference line with intercept θ_0 and slope $\exp(\zeta_0)$. See Output 9.2.1 on page 304 for an example.

Details

This section provides details on the following topics:

- distributions supported by the PROBPLOT statement
- SYMBOL statement options

Summary of Theoretical Distributions

You can use the PROBPLOT statement to request probability plots based on the theoretical distributions summarized in the following table:

Table 9.13. Distributions and Parameters

Distribution	Density Function $p(x)$	Range	Parameters		
			Location	Scale	Shape
Beta	$\frac{(x-\theta)^{\alpha-1}(\theta+\sigma-x)^{\beta-1}}{B(\alpha,\beta)\sigma^{(\alpha+\beta-1)}}$	$\theta < x < \theta + \sigma$	θ	σ	α, β
Exponential	$\frac{1}{\sigma} \exp\left(-\frac{x-\theta}{\sigma}\right)$	$x \geq \theta$	θ	σ	
Gamma	$\frac{1}{\sigma\Gamma(\alpha)} \left(\frac{x-\theta}{\sigma}\right)^{\alpha-1} \exp\left(-\frac{x-\theta}{\sigma}\right)$	$x > \theta$	θ	σ	α
Lognormal (3-parameter)	$\frac{1}{\sigma\sqrt{2\pi}(x-\theta)} \exp\left(-\frac{(\log(x-\theta)-\zeta)^2}{2\sigma^2}\right)$	$x > \theta$	θ	ζ	σ
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	all x	μ	σ	
Weibull (3-parameter)	$\frac{c}{\sigma} \left(\frac{x-\theta}{\sigma}\right)^{c-1} \exp\left(-\left(\frac{x-\theta}{\sigma}\right)^{c}\right)$	$x > \theta$	θ	σ	c
Weibull (2-parameter)	$\frac{c}{\sigma} \left(\frac{x-\theta_0}{\sigma}\right)^{c-1} \exp\left(-\left(\frac{x-\theta_0}{\sigma}\right)^{c}\right)$	$x > \theta_0$	θ_0 (known)	σ	c

You can request these distributions with the BETA, EXPONENTIAL, GAMMA, LOGNORMAL, NORMAL, WEIBULL, and WEIBULL2 options, respectively. If you do not specify a distribution option, a normal probability plot is created.

Shape Parameters

Some of the distribution options in the PROBPLOT statement require you to specify one or two shape parameters in parentheses after the distribution keyword. These are summarized in Table 9.14.

Table 9.14. Shape Parameter Options for the PROBPLOT Statement

Distribution Keyword	Mandatory Shape Parameter Option	Range
BETA	ALPHA=α, BETA=β	$\alpha > 0, \beta > 0$
EXPONENTIAL	None	
GAMMA	ALPHA=α	$\alpha > 0$
LOGNORMAL	SIGMA=σ	$\sigma > 0$
NORMAL	None	
WEIBULL	C=c	$c > 0$
WEIBULL2	None	

You can visually estimate the value of a shape parameter by specifying a list of values for the shape parameter option. The PROBPLOT statement produces a separate plot for each value. You can then use the value of the shape parameter producing the most nearly linear point pattern. Alternatively, you can request that the plot be created using an estimated shape parameter. For an example, see "Creating Lognormal Probability Plots" on page 278.

Location and Scale Parameters

If you specify the location and scale parameters for a distribution (or if you request estimates for these parameters), a diagonal distribution reference line is displayed on the plot. (An exception is the two-parameter Weibull distribution, for which a line is displayed when you specify or estimate the scale and shape parameters.) Agreement between this line and the point pattern indicates that the distribution with these parameters is a good fit. For illustrations, see Example 9.1 on page 302 and Example 9.2 on page 303.

The following table shows how the specified parameters determine the intercept* and slope of the line:

Table 9.15. Intercept and Slope of Distribution Reference Line

	Parameters			Linear Pattern	
Distribution	Location	Scale	Shape	Intercept	Slope
Beta	θ	σ	α , β	θ	σ
Exponential	θ	σ		θ	σ
Gamma	θ	σ	α	θ	σ
Lognormal	θ	ζ	σ	θ	$\exp(\zeta)$
Normal	μ	σ		μ	σ
Weibull (3-parameter)	θ	σ	c	θ	σ
Weibull (2-parameter)	θ_0 (known)	σ	c	$\log(\sigma)$	$\frac{1}{c}$

*The intercept and slope are based on the quantile scale for the horizontal axis, which is displayed on a Q-Q plot; see Chapter 10, "QQPLOT Statement."

For the LOGNORMAL and WEIBULL2 options, you can specify the slope directly with the SLOPE= option. That is, for the LOGNORMAL option, specifying THETA=θ_0 and SLOPE=$\exp(\zeta_0)$ displays the same line as specifying THETA=θ_0 and ZETA=ζ_0. For the WEIBULL2 option, specifying SIGMA=σ_0 and SLOPE=$\frac{1}{c_0}$ displays the same line as specifying SIGMA=σ_0 and C=c_0.

SYMBOL Statement Options

In earlier releases of SAS/QC software, graphical features of lower and upper specification lines and diagonal distribution reference lines were controlled with options in the SYMBOL2, SYMBOL3, and SYMBOL4 statements, respectively. These options are still supported, although they have been superseded by options in the PROBPLOT and SPEC statements. The following table summarizes the two sets of options:

Table 9.16. SYMBOL Statement Options

Feature	Statement and Options	Alternative Statement and Options
Symbol markers	SYMBOL1 Statement	
character	VALUE=*special-symbol*	
color	COLOR=*color*	
font	FONT=*font*	
height	HEIGHT=*value*	
Lower specification line	SPEC Statement	SYMBOL2 Statement
position	LSL=*value*	
color	CLSL=*color*	COLOR=*color*
line type	LLSL=*linetype*	LINE=*linetype*
width	WLSL=*value*	WIDTH=*value*
Upper specification line	SPEC Statement	SYMBOL3 Statement
position	USL=*value*	
color	CUSL=*color*	COLOR=*color*
line type	LUSL=*linetype*	LINE=*linetype*
width	WUSL=*value*	WIDTH=*value*
Target reference line	SPEC Statement	
position	TARGET=*value*	
color	CTARGET=*color*	
line type	LTARGET=*linetype*	
width	WTARGET=*value*	
Distribution reference line	PROBPLOT Statement	SYMBOL4 Statement
color	COLOR=*color*	COLOR=*color*
line type	LINE=*linetype*	LINE=*linetype*
width	WIDTH=*value*	WIDTH=*value*

For an illustration of these options, see Example 9.1 on page 302.

Examples

This section provides advanced examples of the PROBPLOT statement.

Example 9.1. Displaying a Normal Reference Line

See CAPPROB4 in the SAS/QC Sample Library

Measurements of the distance between two holes cut into 50 steel sheets are saved as values of the variable DISTANCE in the following data set:

```
data sheets;
   input distance @@;
   label distance='Hole Distance in cm';
   datalines;
 9.80 10.20 10.27  9.70  9.76
10.11 10.24 10.20 10.24  9.63
 9.99  9.78 10.10 10.21 10.00
 9.96  9.79 10.08  9.79 10.06
10.10  9.95  9.84 10.11  9.93
10.56 10.47  9.42 10.44 10.16
10.11 10.36  9.94  9.77  9.36
 9.89  9.62 10.05  9.72  9.82
 9.99 10.16 10.58 10.70  9.54
10.31 10.07 10.33  9.98 10.15
;
```

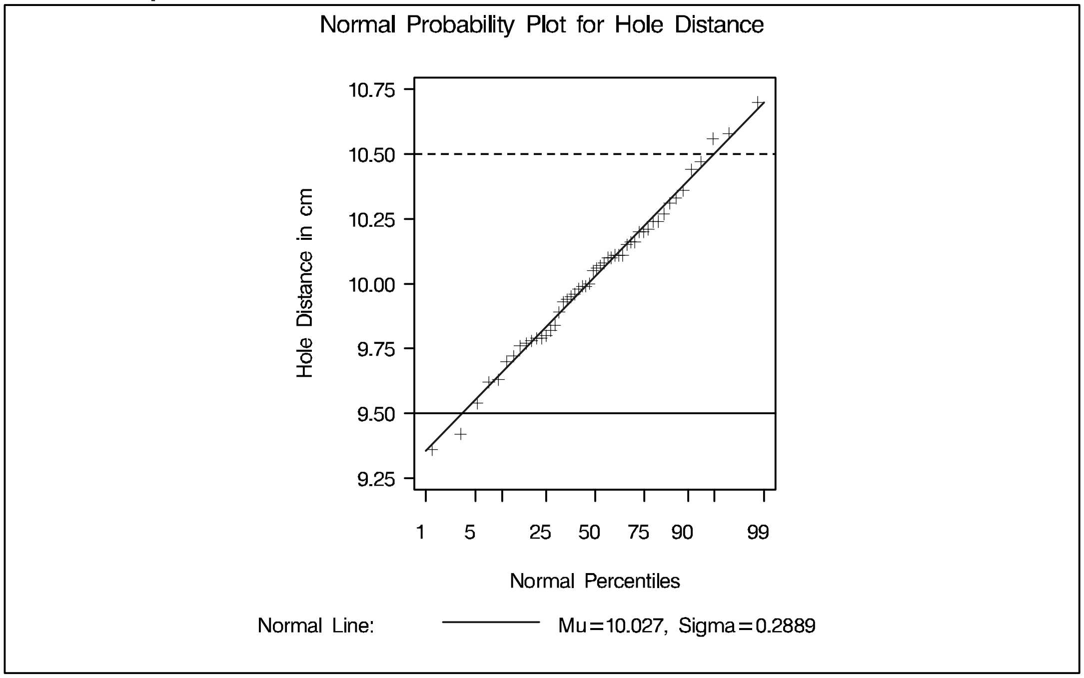
The cutting process is in control, and you decide to check whether the process distribution is normal. The following statements create a normal probability plot for DISTANCE with lower and upper specification lines at 9.5 cm and 10.5 cm:

```
symbol v=dot;
title 'Normal Probability Plot for Hole Distance';

proc capability data=sheets noprint;
   spec  lsl=9.5    llsl=1  clsl=black
         usl=10.5   lusl=2  cusl=black;
   probplot distance / normal(mu=est sigma=est color=blue)
                       square
                       nospeclegend
                       vaxis=axis1;
   axis1 label=(a=90 r=0);
run;
```

The plot is shown in Output 9.1.1. The MU= and SIGMA= *normal-options* request the diagonal reference line that corresponds to the normal distribution with estimated parameters $\hat{\mu} = 10.027$ and $\hat{\sigma} = 0.2889$. The LSL= and USL= SPEC statement options request the lower and upper specification lines, and the LLSL=, LUSL=, CLSL=, and CUSL= options specify the line types and colors. The SYMBOL statement specifies the symbol marker for the plotted points, and the AXIS1 statement specifies the angle and rotation for the vertical axis label.

Output 9.1.1. Normal Reference Line



Example 9.2. Displaying a Lognormal Reference Line

See CAPPROB3 in the SAS/QC Sample Library

This example is a continuation of "Creating Lognormal Probability Plots" on page 278. Figure 9.4 shows that a lognormal distribution with shape parameter $\sigma = 0.5$ is a good fit for the distribution of DIAMETER in the data set MEASURES.

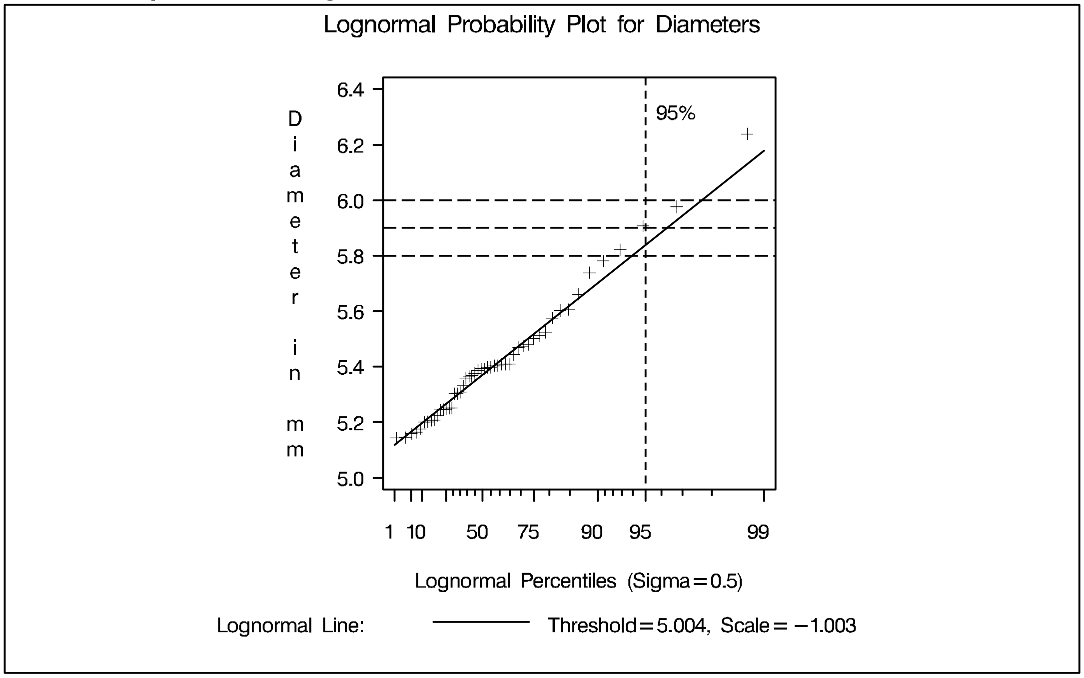
The lognormal distribution involves two other parameters: a threshold parameter θ and a scale parameter ζ. See Table 9.13 on page 299 for the equation of the lognormal density function. The following statements illustrate how you can request a diagonal distribution reference line whose slope and intercept are determined by estimates of θ and ζ.

```
symbol v=dot;
title 'Lognormal Probability Plot for Diameters';

proc capability data=measures noprint;
   probplot diameter / lognormal(sigma=0.5 theta=est zeta=est)
                       square
                       pctlminor
                       href      = 95
                       lhref     = 2
                       hreflabel = '95%'
                       vref      = 5.8 to 6.0 by 0.1
                       lvref     = 3;
run;
```

The plot is shown in Output 9.2.1.

Output 9.2.1. Lognormal Reference Line



The close agreement between the diagonal reference line and the point pattern indicates that the specific lognormal distribution with $\hat{\sigma} = 0.5$, $\hat{\theta} = 5.004$, and $\hat{\zeta} = -1.003$ is a good fit for the diameter measurements.

Specifying HREF=95 adds a reference line indicating the 95^{th} percentile of the lognormal distribution. The LHREF= and HREFLABEL= options specify the line type and a label for this line. The PCTLMINOR option displays minor tick marks on the percentile axis. The VREF= option adds reference lines indicating diameter values of 5.8, 5.9, and 6.0, and the LVREF= option specifies their line type.

Based on the intersection of the diagonal reference line with the HREF= line, the estimated 95^{th} percentile of the diameter distribution is 5.85 mm.

Note that you could also construct a similar plot in which all three parameters are estimated by substituting SIGMA=EST for SIGMA=0.5 in the preceding statements.

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

The Institute is a private company devoted to the support and further development of its software and related services.