Chapter 10
QQPLOT Statement

Chapter Table of Contents

Chapter 10
QQPLOT Statement

Overview

The QQPLOT statement creates a quantile-quantile plot (Q-Q plot), which compares ordered values of a variable with quantiles 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 Q-Q plot to determine how well a theoretical distribution models a set of measurements.

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

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

You can use options in the QQPLOT statement to

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

Note: Q-Q plots are similar to probability plots, which you can create with the PROBPLOT statement (see Chapter 9, “PROBPLOT Statement”). Q-Q plots are preferable for graphical estimation of distribution parameters and capability indices, whereas probability plots are preferable for graphical estimation of percentiles.

Getting Started

The following examples illustrate the basic syntax of the QQPLOT statement. For complete details of the QQPLOT statement, see the “Syntax” section on page 311. Advanced examples are provided on the “Examples” section on page 336.

Creating a Normal Quantile-Quantile Plot

See CAPQQ1 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 Q-Q plot for DISTANCE, shown in Figure 10.1, with lower and upper specification lines at 9.5 cm and 10.5 cm:*

```
title 'Normal Quantile-Quantile Plot for Hole Distance';
proc capability data=sheets noprint;
   spec lsl=9.5 usl=10.5;
   qqplot distance;
run;
```

The plot compares the ordered values of DISTANCE with quantiles of the normal distribution. The linearity of the point pattern indicates that the measurements are normally distributed. Note that a normal Q-Q plot is created by default. If you specify the LINEPRINTER option in the PROC CAPABILITY statement, the plot is created using a line printer, as shown in Figure 10.2. The specification lines are requested with the LSL= and USL= options in the SPEC statement.

*For a P-P plot using these data, see Figure 8.1 on page 253. For a probability plot using these data, see Example 9.2 on page 303.

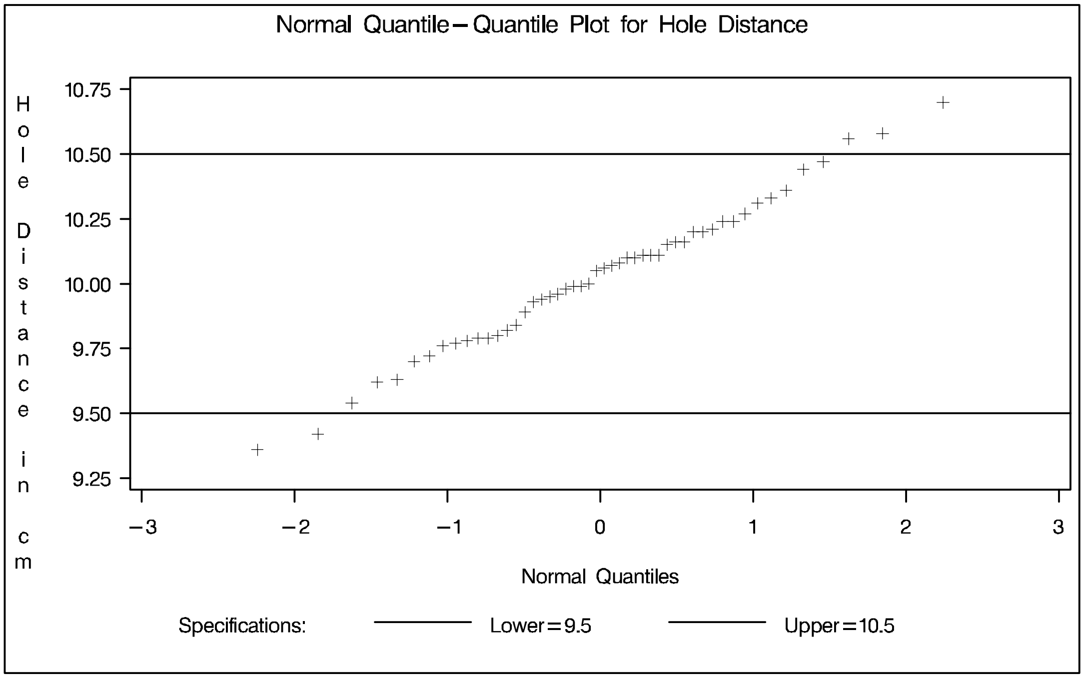


Figure 10.1. Normal Quantile-Quantile Plot Created with Graphics Device

```
               Normal Quantile-Quantile Plot for Hole Distance

         -------------------------------------------------------------------
H 10.75 +                                                                   |
o       |                                                          +        |
l       |                                                   + +             |
e 10.50 +UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU+UUUUUUUUUUUUUUUUU|
        |                                                +                  |
D       |                                            +++                    |
i 10.25 +                                         +++                       |
s       |                                      +++                          |
t       |                                 +++++                             |
a 10.00 +                              +++                                  |
n       |                            ++                                     |
c       |                          ++                                       |
e  9.75 +                     +++++                                         |
        |                  + +                                              |
i       |                 +                                                 |
n  9.50 +LLLLLLLLLLLLLLL+LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL|
        |             +                                                     |
c       |        +                                                          |
m  9.25 +                                                                   |
         +----------+----------+----------+----------+----------+----------+
        -3         -2         -1          0          1          2          3

                                  Normal Quantiles

Specifications:   LLL Lower = 9.5    UUU Upper = 10.5
Observations:     +   (13 Hidden)
```

Figure 10.2. Normal Quantile-Quantile Plot Created with Line Printer

Adding a Distribution Reference Line

See CAPQQ1 in the SAS/QC Sample Library

In a normal Q-Q plot, the normal distribution with mean μ_0 and standard deviation σ_0 is represented by a reference line with intercept μ_0 and slope σ_0. The following statements reproduce the Q-Q plot in Figure 10.1, adding the line for which μ_0 and σ_0 are estimated by the sample mean and standard deviation:

```
title 'Normal Quantile-Quantile Plot for Hole Distance';
proc capability data=sheets noprint;
   spec  lsl=9.5   llsl=2   clsl=blue
         usl=10.5  lusl=20  cusl=blue;
   qqplot distance / normal(mu=est sigma=est color=red l=2)
                     square
                     nospeclegend;
run;
```

The plot is displayed in Figure 10.3.

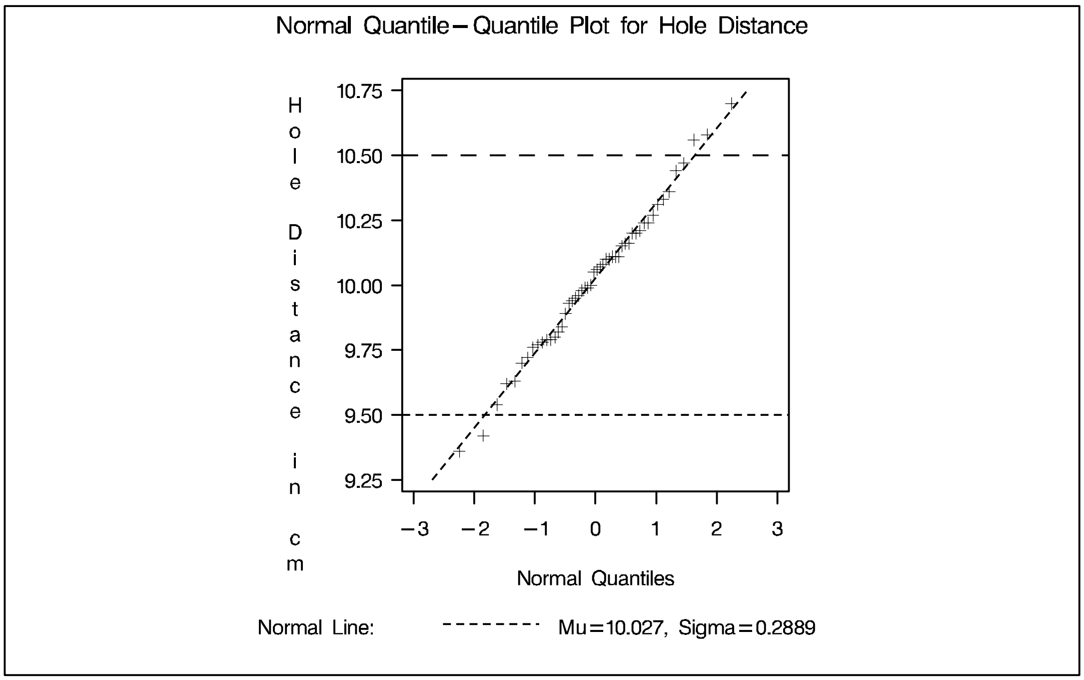


Figure 10.3. Adding a Distribution Reference Line to a Q-Q Plot

Specifying MU=EST and SIGMA=EST with the NORMAL option requests the reference line (alternatively, you can specify numeric values for μ_0 and σ_0 with the MU= and SIGMA= options). The COLOR= and L= options specify the color of the line and the line type. The SQUARE option displays the plot in a square format, and the NOSPECLEGEND option suppresses the legend for the specification lines. The LLSL=, LUSL=, CLSL=, and CUSL= options in the SPEC statement specify line types and colors for the specification limits.

Syntax

The syntax for the QQPLOT statement is as follows:

QQPLOT<*variables* > < **/** *options* >**;**

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

variables

are the process variables for which to create Q-Q 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 Q-Q 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 QQPLOT statements produces two Q-Q plots, one for LENGTH and one for WIDTH:

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

proc capability data=measures;
   qqplot 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 QQPLOT statement. You can specify only one *option* naming the distribution in each QQPLOT 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 Q-Q plot for each variable. The MU= and SIGMA= *normal-options* request a distribution reference line with intercept 10 and slope 0.3 for each plot, corresponding to a normal distribution with mean $\mu = 10$ and standard deviation $\sigma = 0.3$. The SQUARE option displays the plot in a square frame, and the CTEXT= option specifies the text color.

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

Summary of Options

The following tables list the QQPLOT statement *options* by function. For complete descriptions, see “Dictionary of Options” on page 315.

Distribution Options

Table 10.1 summarizes the options for requesting a specific theoretical distribution.

Table 10.1. Keywords to Select a Theoretical Distribution

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

Table 10.2 through Table 10.9 summarize options that specify parameter values for theoretical distributions and that 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 Q-Q plot with a specific distribution reference line. The MU= and SIGMA= *normal-options* display a distribution reference line with intercept 10 and slope 0.3. The COLOR= *normal-option* draws the line in red.

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

Table 10.2. 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 10.3. Beta-Options

ALPHA=*value-list*|EST	specifies mandatory shape parameter α
BETA=*value-list*|EST	specifies mandatory shape parameter β
SIGMA=*value*|EST	specifies reference line slope σ
THETA=*value*|EST	specifies reference line intercept θ

Table 10.4. Exponential-Options

SIGMA=*value*|EST	specifies reference line slope σ
THETA=*value*|EST	specifies reference line intercept θ

Table 10.5. Gamma-Options

ALPHA=*value-list*|EST	specifies mandatory shape parameter α
SIGMA=*value*|EST	specifies reference line slope σ
THETA=*value*|EST	specifies reference line intercept θ

Table 10.6. Lognormal-Options

SIGMA=*value-list*|EST	specifies mandatory shape parameter σ
SLOPE=*value*|EST	specifies reference line slope
THETA=*value*|EST	specifies reference line intercept θ
ZETA=*value*|EST	specifies reference line slope $\exp(\zeta_0)$

Table 10.7. Normal-Options

CPKREF	specifies vertical reference lines at intersection of specification limits with distribution reference line
CPKSCALE	rescales horizontal axis in C_{pk} units
MU=*value*|EST	specifies reference line intercept μ
SIGMA=*value*|EST	specifies reference line slope σ

Table 10.8. Weibull-Options

C=*value-list*|EST	specifies mandatory shape parameter c
SIGMA=*value*|EST	specifies reference line slope σ
THETA=*value*|EST	specifies reference line intercept θ

Table 10.9. Weibull2-Options

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

General Options

Table 10.10 through Table 10.12 list options that control the appearance of the plots.

Table 10.10. General Plot Layout Options

HREF=*value-list*	specifies reference lines perpendicular to the horizontal axis
HREFLABELS= ’*label1*’ . . . ’*labeln*’	specifies labels for HREF= lines
LEGEND=*name* | NONE	specifies LEGEND statement
NADJ=*value*	adjusts sample size (N) when computing quantiles
NOFRAME	suppresses frame around plotting area
NOLEGEND	suppresses legend
NOLINELEGEND	suppresses distribution reference line information in legend
NOSPECLEGEND	suppresses specifications information in legend
PCTLAXIS(*axis-options*)	adds a nonlinear percentile axis
PCTLMINOR	adds minor tick marks to percentile axis
PCTLSCALE	replaces theoretical quantiles with percentiles
RANKADJ=*value*	adjusts ranks when computing quantiles
ROTATE	switches horizontal and vertical axes
SQUARE	displays Q-Q plot in square format
VREF=*value-list*	specifies reference lines perpendicular to the vertical axis
VREFLABELS= ’*label1*’ . . . ’*labeln*’	specifies labels for VREF= lines

Table 10.11. Options to Enhance Plots Produced on Line Printers

HREFCHAR=’*character*’	specifies line character for HREF= lines
NOOBSLEGEND	suppresses legend for hidden points
QQSYMBOL=’*character*’	specifies character for plotted points
VREFCHAR=’*character*’	specifies character for VREF= lines

Table 10.12. 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
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 QQPLOT statement. The marginal notes *Graphics* and *Line Printer* identify options that apply to graphics devices and line printers, respectively.

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

specifies values for a mandatory shape parameter α $(\alpha > 0)$ for Q-Q 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*

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 QQPLOT 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. *Graphics*

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

creates a beta Q-Q 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 example, the first QQPLOT 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;
   qqplot width / beta(alpha=2 beta=2);
   qqplot width / beta(alpha=2 3 beta=1 2);
   qqplot width / beta(alpha=2 to 3 beta=1 to 2 by 0.5);
   qqplot 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 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. To assess the point pattern, you can add a diagonal distribution reference line with intercept θ_0 and slope 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 slope 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;
   qqplot 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 Q-Q 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 distribution option.

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

specifies the shape parameter c ($c > 0$) for Q-Q 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 an example, see Output 10.3.1 on page 342.

CAXIS=*color*
CAXES=*color*

specifies the color 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*
CFR=*color*

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

CHREF=*color*
CH=*color*

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

COLOR=*color*

specifies the color for a distribution reference line. Specify the COLOR= option in parentheses following a distribution option keyword. For an example, see Figure 10.3 on page 310. The default is the fourth color in the device color list. *Graphics*

CPKREF

draws reference lines extending from the intersections of the specification limits with the distribution reference line to the quantile axis in plots requested with the NORMAL option. Specify CPKREF in parentheses after the NORMAL option. You can use the CPKREF option with the CPKSCALE option for graphical estimation of the capability indices *CPU*, *CPL*, and C_{pk}, as illustrated in Output 10.4.1 on page 344. *Graphics*

CPKSCALE

rescales the quantile axis in C_{pk} units for plots requested with the NORMAL option. Specify CPKSCALE in parentheses after the NORMAL option. You can use the CPKSCALE option with the CPKREF option for graphical estimation of the capability indices *CPU*, *CPL*, and C_{pk}, as illustrated in Output 10.4.1 on page 344.

CTEXT=*color*

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. In the *Graphics* absence of a GOPTIONS statement, the default color is the first color in the device color list.

CVREF=*color*
CV=*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 Q-Q 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 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 the threshold parameter, and σ is the scale parameter $(\sigma > 0)$.

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

```
proc capability data=measures;
   qqplot 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 SCALE= option as an alias for the SIGMA= option and the THRESHOLD= option as an alias for the THETA= option.

FONT=*font*

Graphics

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.

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

creates a gamma Q-Q plot for each value of the shape parameter α given by the mandatory ALPHA= option or its alias, the SHAPE= option. The following example produces three probability plots:

```
proc capability data=measures;
   qqplot width / gamma(alpha=0.4 to 0.6 by 0.1);
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 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 with intercept θ_0 and slope 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 *gamma-options* THETA=EST and SIGMA=EST. Specify these options in parentheses, as in the following example:

```
proc capability data=measures;
   qqplot 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.

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

HMINOR=*n*
HM=*n*
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. *Graphics*

HREF=*value-list*
draws reference lines perpendicular to the horizontal axis at the values specified. See Example 10.3 on page 341 for illustrations. Related options include the HREFCHAR=, CHREF=, and LHREF= options.

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.

L=*linetype*

Graphics

specifies the line type for a distribution reference line. Specify the L= option in parentheses following a distribution option keyword. 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.

LHREF=*linetype*
LH=*linetype*

Graphics

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

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

creates a lognormal Q-Q plot for each value of the shape parameter σ given by the mandatory SIGMA= option or its alias, the SHAPE= option. For example,

```
proc capability data=measures;
   qqplot width/ lognormal(shape=1.5 2.5);
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 cumulative standard normal distribution, n is the number of nonmissing observations, and σ is the shape parameter of the lognormal distribution.

The 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. For an illustration, see Example 10.2 on page 337.

To assess the point pattern, you can add a diagonal distribution reference line corresponding to the threshold parameter θ_0 and the scale parameter 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. This line has intercept θ_0 and slope $\exp(\zeta_0)$. Agreement between the line and the point pattern indicates that the lognormal distribution with parameters σ, θ_0, and ζ_0 is a good fit. See Output 10.2.4 on page 339 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.

You can also display the reference line by specifying THETA=θ_0, and you can specify the slope with the SLOPE= option. For example, the following two QQPLOT statements produce charts with identical reference lines:

```
proc capability data=measures;
   qqplot width / lognormal(sigma=2 theta=3 zeta=1);
   qqplot width / lognormal(sigma=2 theta=3 slope=2.718);
run;
```

LVREF=*linetype*
LV=*linetype*
specifies the line type for reference lines requested by the VREF= option. The default is 2, which produces a dashed line. *Graphics*

MU=*value***|EST**
specifies a value for the mean μ for a normal Q-Q plot requested with the NORMAL option. Specify MU=μ_0 and SIGMA=σ_0 to request a distribution reference line with intercept μ_0 and slope σ_0. Specify MU=EST to request a distribution reference line with intercept equal to the sample mean, as illustrated in Figure 10.3 on page 310.

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

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

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. For an example, see Output 10.4.1 on page 344.

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 Q-Q 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 pattern on the plot tends to be linear with intercept μ and slope σ if the data are normally distributed with the specific density function

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

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 with intercept μ_0 and slope 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;
   qqplot length / normal(mu=10 sigma=0.3);
run;
```

For an example, see “Adding a Distribution Reference Line” on page 310. 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. You can specify MU=EST and SIGMA=EST to request a distribution reference line with the sample mean and sample standard deviation as the intercept and slope.

Other *normal-options* include CPKREF and CPKSCALE. The CPKREF option draws reference lines extending from the intersections of specification limits with the distribution reference line to the theoretical quantile axis. The CPKSCALE option rescales the theoretical quantile axis in C_{pk} units. You can use the CPKREF option with the CPKSCALE option for graphical estimation of the capability indices *CPU*, *CPL*, and C_{pk}, as illustrated in Output 10.4.1 on page 344.

NOSPECLEGEND
NOSPECL
suppresses the legend for specification limit reference lines. For an example, see Figure 10.3 on page 310.

PCTLAXIS(*axis-options***)**

adds a nonlinear percentile axis along the frame of the Q-Q plot opposite the theoretical quantile axis. The added axis is identical to the axis for probability plots produced with the PROBPLOT statement. When using the PCTLAXIS option, you must specify HREF= values in quantile units, and you cannot use the NOFRAME option. You can specify the following *axis-options*:

GRID	draws vertical grid lines at major percentiles
GRIDCHAR='*character*'	specifies grid line plotting character on line printer
LABEL='*string*'	specifies label for percentile axis
LGRID=*linetype*	specifies line type for grid

For example, the following statements display the plot in Figure 10.4:

See CAPQQ1 in the SAS/QC Sample Library

```
title 'Normal Quantile-Quantile Plot for Hole Distance';
proc capability data=sheets noprint;
   qqplot distance /
      normal(mu=est sigma=est color=blue)
      pctlaxis(grid lgrid=35 label='Normal Percentiles')
      nolegend;
run;
```

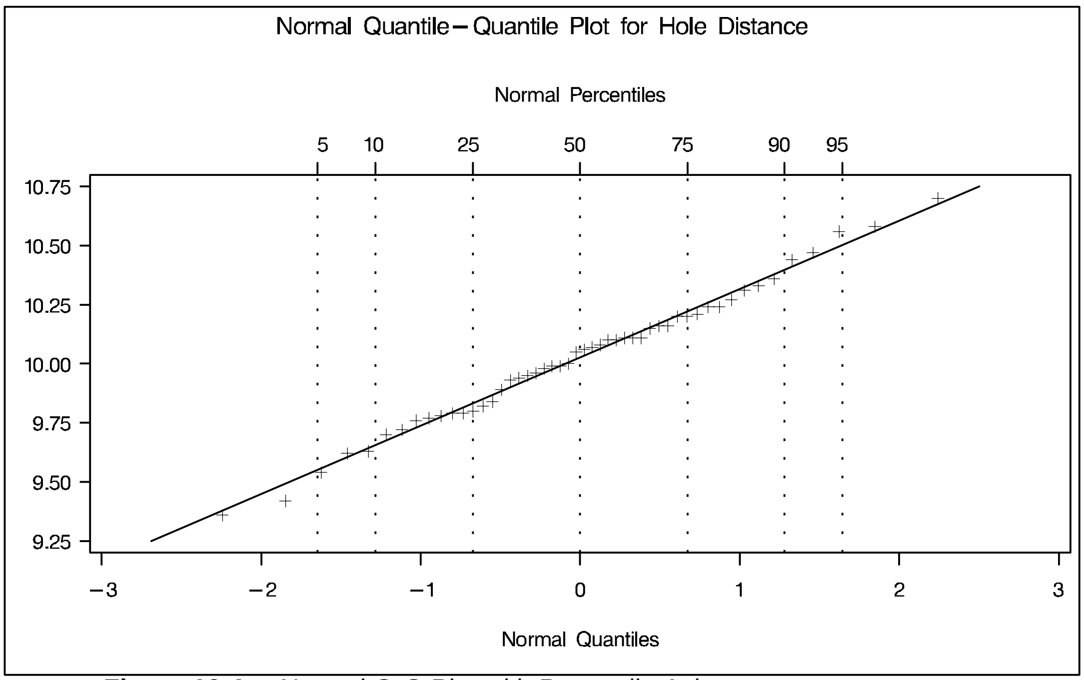


Figure 10.4. Normal Q-Q Plot with Percentile Axis

PCTLMINOR

requests minor tick marks for the percentile axis displayed when you use the PCTLAXIS option. See the entry for the PCTLAXIS option for an example.

PCTLSCALE

requests scale labels for the theoretical quantile axis in percentile units, resulting in a nonlinear axis scale. Tick marks are drawn uniformly across the axis based on the quantile scale. In all other respects, the plot remains the same, and you must specify HREF= values in quantile units. For a true nonlinear axis, use the PCTLAXIS option or use the PROBPLOT statement. For example, the following statements display the plot in Figure 10.5:

See CAPQQ1 in the SAS/QC Sample Library

```
title 'Normal Quantile-Quantile Plot for Hole Distance';
proc capability data=sheets noprint;
   spec  lsl=9.5   llsl=2   clsl=blue
         usl=10.5  lusl=20  cusl=blue;
   qqplot distance / normal(mu=est sigma=est cpkref)
                     pctlscale
                     pctlaxis(grid lgrid=35)
                     nolegend;
run;
```

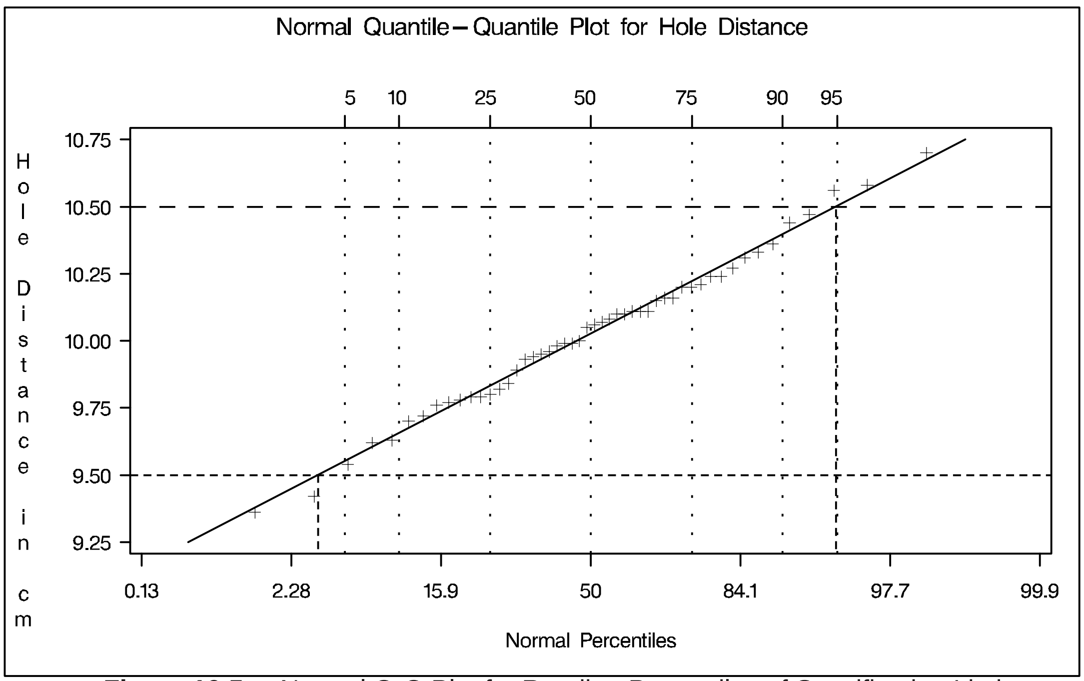


Figure 10.5. Normal Q-Q Plot for Reading Percentiles of Specification Limits

QQSYMBOL='*character*'

Line Printer

specifies the character used to plot the Q-Q points 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 quantiles. The default is $-\frac{3}{8}$, as described 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=) refer to the horizontal axis, and vertical axis options (such as VAXIS=) 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 distribution parameter σ, where $\sigma > 0$. Alternatively, you can specify SIGMA=EST to request a maximum likelihood estimate for σ_0. The use of the SIGMA= option depends on the distribution option 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 with intercept θ_0 and slope σ_0.
LOGNORMAL	SIGMA=$\sigma_1 \ldots \sigma_n$ requests n Q-Q plots with shape parameters $\sigma_1 \ldots \sigma_n$. The SIGMA= option is mandatory.
NORMAL	MU=μ_0 and SIGMA=σ_0 request a distribution reference line with intercept μ_0 and slope σ_0. SIGMA=EST requests a slope equal to the sample standard deviation.
WEIBULL2	SIGMA=σ_0 and C=c_0 request a distribution reference line with intercept $\log(\sigma_0)$ and slope $\frac{1}{c_0}$.

For an example using SIGMA=EST, see Output 10.4.1 on page 344. For an example of lognormal plots using the SIGMA= option, see Example 10.2 on page 337.

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= option. Specifying the SLOPE= option is an alternative to specifying ZETA=ζ_0, which requests a slope of $\exp(\zeta_0)$. See Output 10.2.4 on page 339 for an example.

When you use the SLOPE= option with the WEIBULL2 option, you must also specify a scale parameter value σ_0 with the SIGMA= option. Specifying the SLOPE= option is an alternative to specifying C=c_0, which requests a slope of $\frac{1}{c_0}$.

For example, the first and second QQPLOT statements that follow produce plots identical to those produced by the third and fourth QQPLOT statements:

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

For more information, see “Graphical Estimation” on page 332.

SQUARE
displays the Q-Q plot in a square frame. Compare Figure 10.1 on page 309 with Figure 10.3 on page 310. The default is a rectangular frame.

SYMBOL='*character*'

Line Printer

specifies the character used to plot a distribution reference line when the plot is produced on a line printer. The default character is the first letter of the distribution option keyword.

THETA=*value***|EST**
specifies the lower threshold parameter θ for Q-Q 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. See Output 10.3.2 on page 343 for an example.

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 10.2.4 on page 339 for an example of the THETA= option with a lognormal Q-Q plot.

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

VAXIS=*name*

Graphics

specifies the name of an AXIS statement describing the vertical axis. For an example, see Example 10.1 on page 336.

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. For illustrations, see Output 10.2.4 on page 339 or Example 10.3 on page 341. Related options include the VREFCHAR=, CVREF=, and LVREF= options.

VREFCHAR=’*character***’**

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

VREFLABELS=’*label1***’** . . . **’***labeln***’**
VREFLABEL=’*label1***’** . . . **’***labeln***’**
VREFLAB=’*label1***’** . . . **’***labeln***’**

specifies labels for the reference 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*

specifies the width in pixels for a distribution reference line, as in the following example. The default is 1. *Graphics*

```
proc capability data=measures;
   qqplot length / normal(mu=5 sigma=2 w=2);
run;
```

WEIBULL(C=*value-list***|EST** $<$*Weibull-options* $>$**)**
WEIB(C=*value-list* $<$*Weibull-options* $>$**)**

creates a three-parameter Weibull Q-Q plot for each value of the shape parameter c given by the mandatory C= option or its alias, the SHAPE= option. For example,

```
proc capability data=measures;
   qqplot width / weibull(c=1.8 to 2.4 by 0.2);
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 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

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

where θ is the threshold parameter, σ is the scale parameter $(\sigma > 0)$, and c is the 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. For an illustration, see Example 10.3 on page 341. To assess the point pattern, you can add a diagonal distribution reference line with intercept θ_0 and slope 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;
   qqplot 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 Q-Q 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= option or its alias, the THRESHOLD= option. If you are uncertain of the lower threshold value, you can estimate θ_0 graphically by specifying a list of values for the THETA= option. Select the value that most linearizes the point pattern. 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 $x_{(i)}$, $\log(x_{(i)} - \theta_0)$, 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. Unlike the three-parameter Weibull quantile, the preceding expression is free of distribution parameters. This is why the C= shape parameter option is not mandatory with the WEIBULL2 option.

The 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

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

where θ_0 is a known lower threshold parameter, σ is a scale parameter $(\sigma > 0)$, and c is a shape parameter $(c > 0)$.

The advantage of a two-parameter Weibull plot over a three-parameter Weibull plot is that you can visually estimate the shape parameter c and the scale parameter σ from the slope and intercept of the point pattern; see Example 10.3 on page 341 for an illustration of this method. The disadvantage is that the two-parameter Weibull distribution applies only in situations where the threshold parameter is known. See “Graphical Estimation” on page 332 for more information.

To assess the point pattern, you can add a diagonal distribution reference line corresponding to the scale parameter σ_0 and shape parameter 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. This line has intercept $\log(\sigma_0)$ and slope $\frac{1}{c_0}$. Agreement between the 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.

You can also display the reference line by specifying SIGMA=σ_0, and you can specify the slope with the SLOPE= option. For example, the following QQPLOT statements produce identical plots:

```
proc capability data=measures;
   qqplot width / weibull2(theta=3 sigma=4 c=2);
   qqplot width / weibull2(theta=3 sigma=4 slope=0.5);
run;
```

ZETA=*value***|EST**

specifies a value for the scale parameter ζ for lognormal Q-Q 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 $\exp(\zeta_0)$.

Details

This section provides details on the following topics:

- construction of Q-Q plots
- interpretation of Q-Q plots
- distributions supported by the QQPLOT statement
- graphical estimation of shape parameters, location and scale parameters, theoretical percentiles, and capability indices
- SYMBOL statement options

Construction of Quantile-Quantile and Probability Plots

Figure 10.6 illustrates how a Q-Q plot is constructed. First, the n nonmissing values of the variable are ordered from smallest to largest:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

Then the i^{th} ordered value $x_{(i)}$ is represented on the plot by a point whose y-coordinate is $x_{(i)}$ and whose x-coordinate is $F^{-1}\left(\frac{i-0.375}{n+0.25}\right)$, where $F(\cdot)$ is the theoretical distribution with zero location parameter and unit scale parameter.

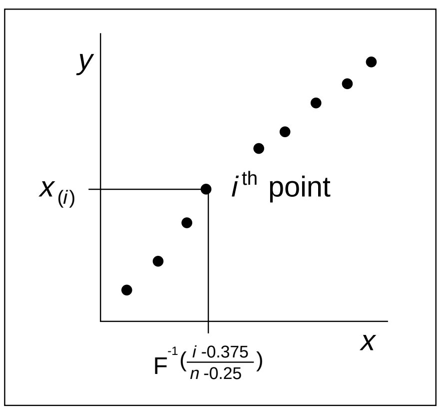


Figure 10.6. Construction of a Q-Q Plot

You can modify the adjustment constants -0.375 and 0.25 with the RANKADJ= and NADJ= options. This default combination is recommended by Blom (1958). For additional information, refer to Chambers and others (1983). Since $x_{(i)}$ is a quantile of the empirical cumulative distribution function (ecdf), a Q-Q plot compares quantiles of the ecdf with quantiles of a theoretical distribution. Probability plots (see Chapter 9, “PROBPLOT Statement”) are constructed the same way, except that the x-axis is scaled nonlinearly in percentiles.

Interpretation of Quantile-Quantile and Probability Plots

The following properties of Q-Q plots and probability plots make them useful diagnostics of how well a specified theoretical distribution fits a set of measurements:

- If the quantiles of the theoretical and data distributions agree, the plotted points fall on or near the line $y = x$.
- If the theoretical and data distributions differ only in their location or scale, the points on the plot fall on or near the line $y = ax + b$. The slope a and intercept b are visual estimates of the scale and location parameters of the theoretical distribution.

Q-Q plots are more convenient than probability plots for graphical estimation of the location and scale parameters since the x-axis of a Q-Q plot is scaled linearly. On the other hand, probability plots are more convenient for estimating percentiles or probabilities.

There are many reasons why the point pattern in a Q-Q plot may not be linear. Chambers and others (1983) and Fowlkes (1987) discuss the interpretations of commonly encountered departures from linearity, and these are summarized in the following table.

Table 10.13. Quantile-Quantile Plot Diagnostics

Description of Point Pattern	Possible Interpretation
All but a few points fall on a line	Outliers in the data
Left end of pattern is below the line; right end of pattern is above the line	Long tails at both ends of the data distribution
Left end of pattern is above the line; right end of pattern is below the line	Short tails at both ends of the data distribution
Curved pattern with slope increasing from left to right	Data distribution is skewed to the right
Curved pattern with slope decreasing from left to right	Data distribution is skewed to the left
Staircase pattern (plateaus and gaps)	Data have been rounded or are discrete

In some applications, a nonlinear pattern may be more revealing than a linear pattern. However, Chambers and others (1983) note that departures from linearity can also be due to chance variation.

Summary of Theoretical Distributions

You can use the QQPLOT statement to request Q-Q plots based on the theoretical distributions summarized in the following table:

Table 10.14. QQPLOT Statement Distribution Options

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 Q-Q plot is created.

Graphical Estimation

You can use Q-Q plots to estimate shape, location, and scale parameters and to estimate percentiles. If you are working with a normal Q-Q plot, you can also estimate certain capability indices.

Shape Parameters

Some distribution options in the QQPLOT statement require that you specify one or two shape parameters in parentheses after the distribution keyword. These are summarized in Table 10.15.

You can visually estimate a shape parameter by specifying a list of values for the shape parameter option. A separate plot is displayed for each value, and you can then select the value that linearizes the point pattern. Alternatively, you can request that the plot be created using an estimated shape parameter. See the entries for the distribution options in "Dictionary of Options" for details on specification of shape parameters. Example 10.2 on page 337 and Example 10.3 on page 341 illustrate shape parameter estimation with lognormal and Weibull Q-Q plots.

Note that for Q-Q plots requested with the WEIBULL2 option, you can estimate the shape parameter c from a linear pattern using the fact that the slope of the pattern is $\frac{1}{c}$. For an illustration, see Example 10.3 on page 341.

Table 10.15. Shape Parameter Options for the QQPLOT 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	

Location and Scale Parameters

When the point pattern on a Q-Q plot is linear, its intercept and slope provide estimates of the location and scale parameters. (An exception to this rule is the two-parameter Weibull distribution, for which the intercept and slope are related to the scale and shape parameters.) Table 10.16 shows how the intercept and slope are related to the parameters for each distribution supported by the QQPLOT statement.

Table 10.16. Intercept and Slope of Linear Q-Q Plots

	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	$\log(\sigma)$	$\frac{1}{c}$

You can enhance a Q-Q plot with a diagonal *distribution reference line* by specifying the parameters that determine the slope and intercept of the line; alternatively, you can request estimates for these parameters. This line is an aid to checking the linearity of the point pattern, and it facilitates parameter estimation. For instance, specifying MU=3 and SIGMA=2 with the NORMAL option requests a line with intercept 3 and slope 2. Specifying SIGMA=1 and C=2 with the WEIBULL2 option requests a line with intercept $\log(1) = 0$ and slope $\frac{1}{2}$.

With 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=$\exp(\zeta_0)$ gives the same reference line as specifying THETA=θ_0 and ZETA=ζ_0. For the WEIBULL2 option, specifying SIGMA=σ_0 and SLOPE=$\frac{1}{c_0}$ gives the same reference line as specifying SIGMA=σ_0 and C=c_0.

For an example of parameter estimation using a normal Q-Q plot, see "Adding a Distribution Reference Line" on page 310. Example 10.2 on page 337 illustrates parameter estimation using a lognormal plot, and Example 10.3 on page 341 illustrates estimation using two-parameter and three-parameter Weibull plots.

Theoretical Percentiles

There are two ways to estimate percentiles from a Q-Q plot:

- Specify the PCTLAXIS option, which adds a percentile axis opposite the theoretical quantile axis. The scale for the percentile axis ranges between 0 and 100 with tick marks at percentile values such as 1, 5, 10, 25, 50, 75, 90, 95, and 99. See Figure 10.4 on page 323 for an example.
- Specify the PCTLSCALE option, which relabels the horizontal axis tick marks with their percentile equivalents but does not alter their spacing. For example, on a normal Q-Q plot, the tick mark labeled "0" is relabeled as "50" since the 50^{th} percentile corresponds to the zero quantile. See Figure 10.5 on page 324 for an example.

You can also estimate percentiles using probability plots created with the PROBPLOT statement. See Output 9.2.1 on page 304 for an example.

Capability Indices

When the point pattern on a normal Q-Q plot is linear, you can estimate the capability indices CPU, CPL, and C_{pk} from the plot, as explained by Rodriguez (1992). This method exploits the fact that the horizontal axis of a Q-Q plot indicates the distance in standard deviation units (multiple of σ) between a measurement or specification limit and the process average.

In particular, one-third the standardized distance between an upper specification limit and the mean is the one-sided capability index *CPU*.

$$CPU = \frac{USL - \mu}{3\sigma}$$

Likewise, one-third the standardized distance between a lower specification limit and the mean is the one-sided capability index *CPL*.

$$CPL = \frac{\mu - LSL}{3\sigma}$$

Consequently, if you *rescale* the quantile axis of a normal Q-Q plot by a factor of three, you can read *CPU* and *CPL* from the horizontal coordinates of the points at which the upper and lower specification lines intersect the point pattern. Since C_{pk} is defined as the minimum of *CPU* and *CPL*, this method also provides a graphical estimate of C_{pk}. For an illustration, see Example 10.4 on page 343.

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 QQPLOT and SPEC statements. The following table summarizes the two sets of options:

Table 10.17. 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	QQPLOT Statement	SYMBOL4 Statement
color	COLOR=*color*	COLOR=*color*
line type	LINE=*linetype*	LINE=*linetype*
width	WIDTH=*value*	WIDTH=*value*

Examples

This section provides advanced examples of the QQPLOT statement.

Example 10.1. Interpreting a Normal Q-Q Plot of Nonnormal Data

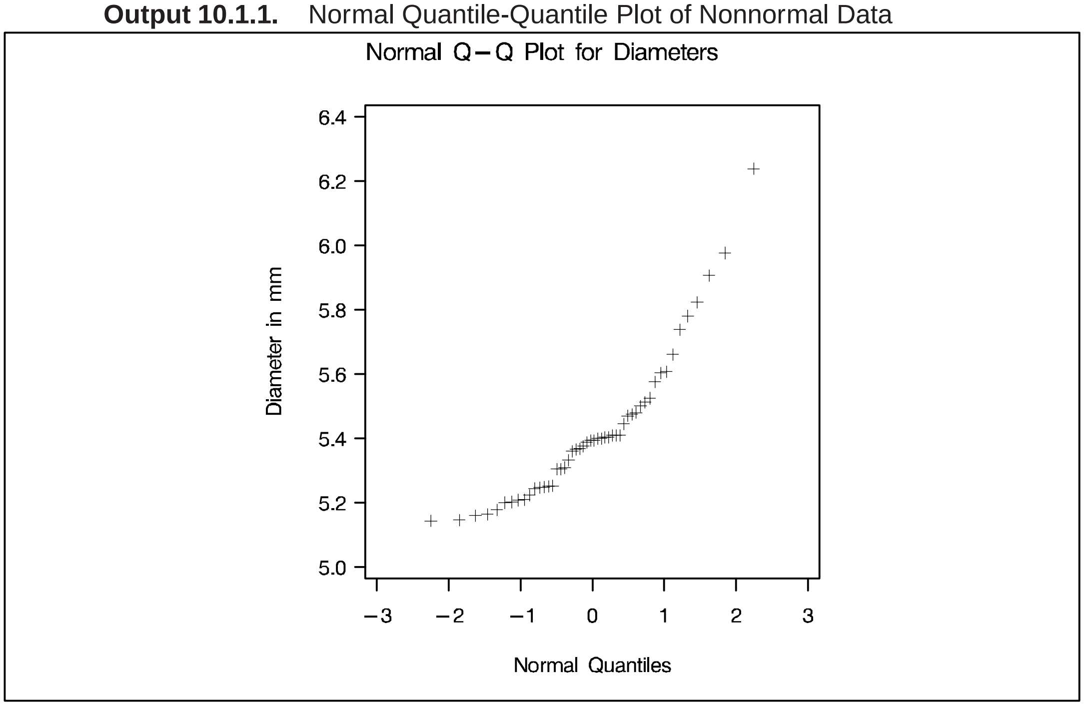
See CAPQQ2 in the SAS/QC Sample Library

The following statements produce the normal Q-Q plot in Output 10.1.1:

```
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
;

title 'Normal Q-Q Plot for Diameters';
proc capability data=measures noprint;
   qqplot diameter / normal square vaxis=axis1;
   axis1 label=(a=90 r=0);
run;
```

Output 10.1.1. Normal Quantile-Quantile Plot of Nonnormal Data



The nonlinearity of the points in Output 10.1.1 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. The mild curvature suggests that you should examine the data with a series of lognormal Q-Q plots for small values of the shape parameter, as illustrated in the next example.

Example 10.2. Estimating Parameters from Lognormal Plots

This example, which is a continuation of Example 10.1, demonstrates techniques for estimating the shape parameter, location and scale parameters, and theoretical percentiles for a lognormal distribution.

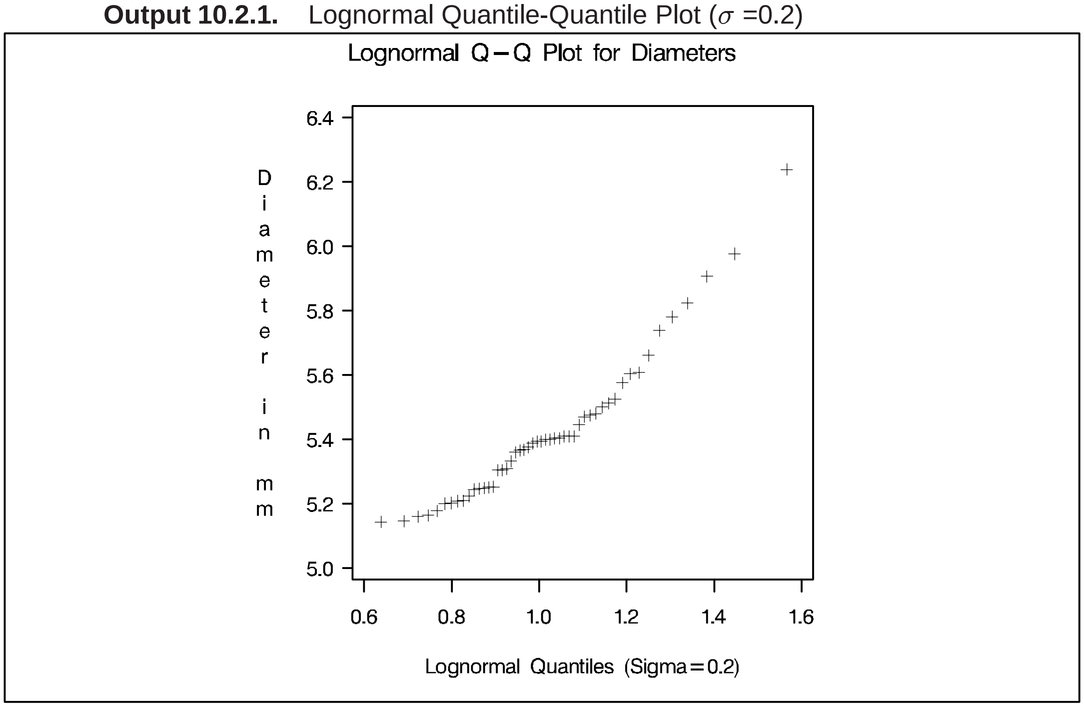
Three-Parameter Lognormal Plots

The three-parameter lognormal distribution depends on a threshold parameter θ, a scale parameter ζ, and a shape parameter σ. You can estimate σ from a series of lognormal Q-Q plots with different values of σ. The estimate is the value of σ that linearizes the point pattern. You can then estimate the threshold and scale parameters from the intercept and slope of the point pattern. The following statements create the series of plots in Output 10.2.1 through Output 10.2.3 for σ values of 0.2, 0.5, and 0.8:

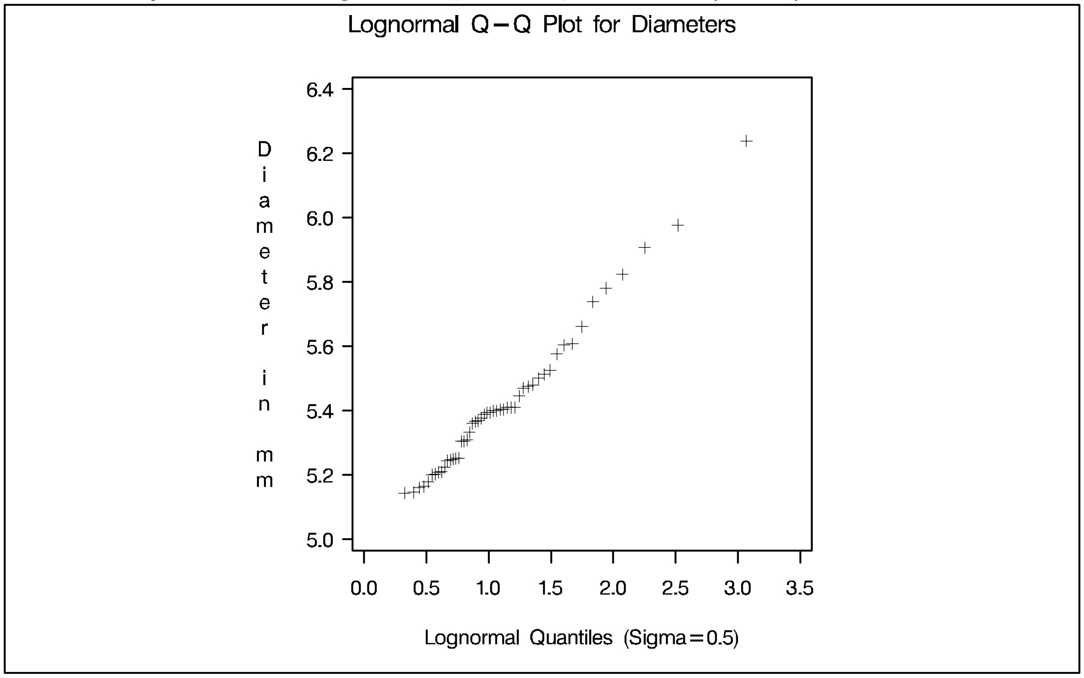
See CAPQQ2 in the SAS/QC Sample Library

```
title 'Lognormal Q-Q Plot for Diameters';
proc capability data=measures noprint;
   qqplot diameter / lognormal(sigma=0.2 0.5 0.8)
                     square;
run;
```

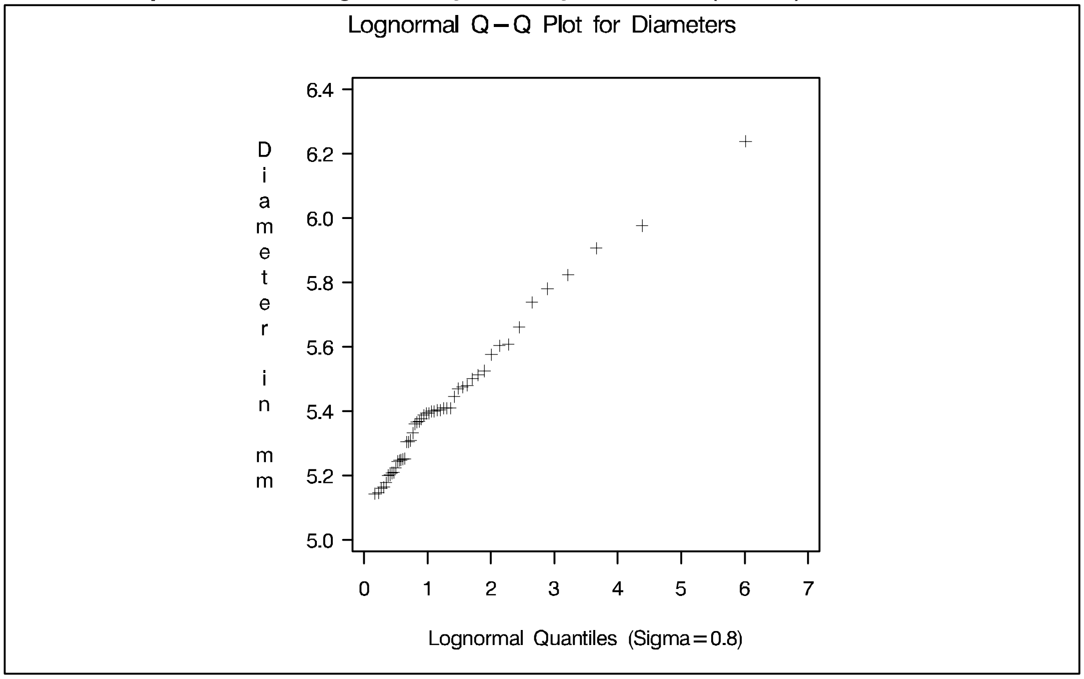
Output 10.2.1. Lognormal Quantile-Quantile Plot (σ =0.2)



Output 10.2.2. Lognormal Quantile-Quantile Plot (σ =0.5)



Output 10.2.3. Lognormal Quantile-Quantile Plot (σ =0.8)



Note: You must specify a value for the shape parameter σ for a lognormal Q-Q plot with the SIGMA= option or its alias, the SHAPE= option.

The plot in Output 10.2.2 displays the most linear point pattern, indicating that the lognormal distribution with $\sigma = 0.5$ provides a reasonable fit for the data distribution.

Data with this particular lognormal distribution have the density function

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

The points in the plot fall on or near the line with intercept θ and slope $\exp(\zeta)$. Based on Output 10.2.2, $\theta \approx 5$ and $\exp(\zeta) \approx \frac{1.2}{3} = 0.4$, giving $\zeta \approx \log(0.4) \approx -0.92$.

Estimating Percentiles

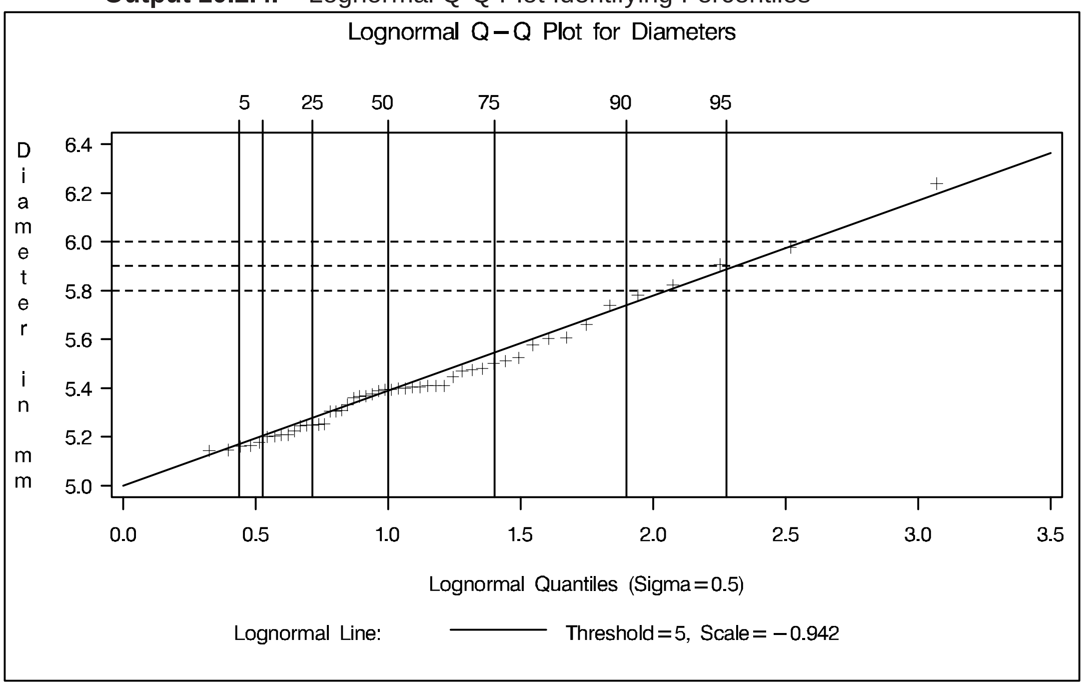
You can use a Q-Q plot to estimate percentiles such as the 95^{th} percentile of the lognormal distribution.*

See CAPQQ2 in the SAS/QC Sample Library

The point pattern in Output 10.2.2 has a slope of approximately 0.39 and an intercept of 5. The following statements reproduce this plot, adding a lognormal reference line with this slope and intercept. The result is shown in Output 10.2.4.

```
proc capability data=measures noprint;
   qqplot diameter / lognormal(sigma=0.5 theta=5 slope=0.39)
                     pctlaxis(grid)
                     vref=5.8 5.9 6.0;
run;
```

Output 10.2.4. Lognormal Q-Q Plot Identifying Percentiles



The PCTLAXIS option labels the major percentiles, and the GRID option draws percentile axis reference lines. The 95^{th} percentile is 5.9, since the intersection of the distribution reference line and the 95^{th} reference line occurs at this value on the vertical axis.

*You can also use a probability plot for this purpose. See Output 9.2.1 on page 304.

Alternatively, you can compute this percentile from the estimated lognormal parameters. The $100\alpha^{\text{th}}$ percentile of the lognormal distribution is

$$P_\alpha = \exp(\sigma\Phi^{-1}(\alpha) + \zeta) + \theta$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative standard normal distribution. Consequently,

$$P_{0.95} \approx \exp\left(\frac{1}{2}\Phi^{-1}(0.95) + \log(0.39)\right) + 5 \approx \exp\left(\frac{1}{2} \times 1.645 - 0.94\right) + 5 \approx 5.89$$

Two-Parameter Lognormal Plots

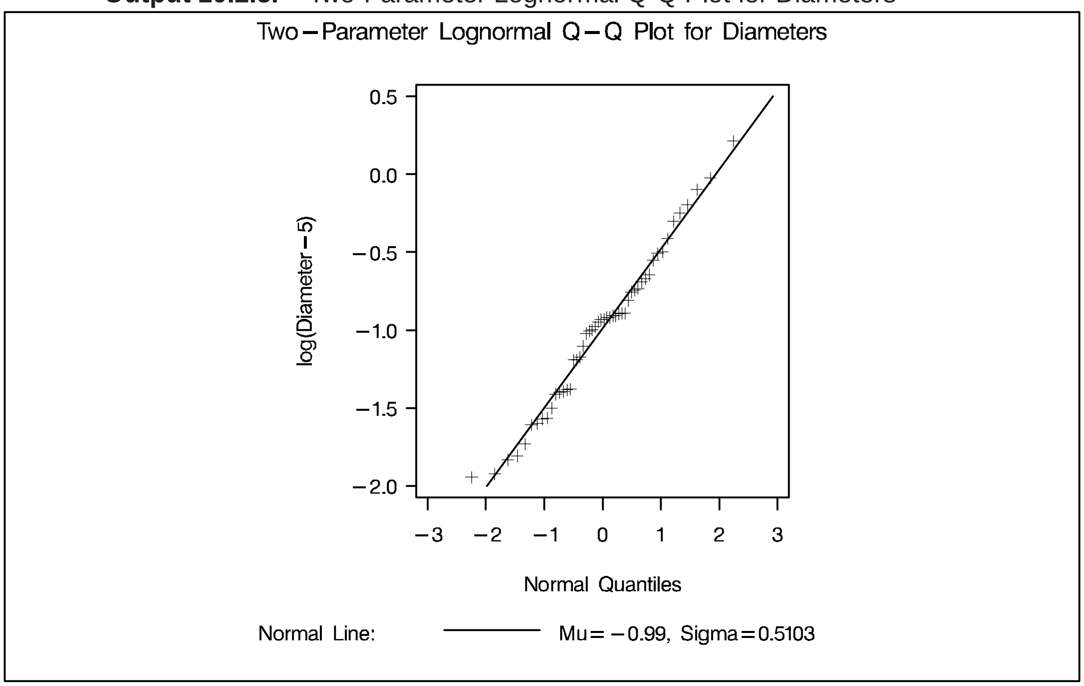
See CAPQQ2 in the SAS/QC Sample Library

If a known threshold parameter is available, you can construct a two-parameter lognormal Q-Q plot by subtracting the threshold from the data and requesting a normal Q-Q plot. The following statements create this plot for DIAMETER, assuming a known threshold of five:

```
data measures;
   set measures;
   label logdiam = 'log(Diameter-5)';
   logdiam = log( diameter - 5 );

title 'Two-Parameter Lognormal Q-Q Plot for Diameters';
proc capability data=measures noprint;
   qqplot logdiam / normal( mu=est sigma=est )
                    vaxis=axis1
                    square;
axis1 label=(a=90 r=0);
run;
```

Output 10.2.5. Two-Parameter Lognormal Q-Q Plot for Diameters



Because the point pattern in Output 10.2.5 is linear, you can estimate the lognormal parameters ζ and σ as the normal plot estimates of μ and σ, which are -0.99 and 0.51. These values correspond to the previous estimates of -0.92 for ζ and 0.5 for σ.

Example 10.3. Comparing Weibull Q-Q Plots

This example compares the use of three-parameter and two-parameter Weibull Q-Q plots for the failure times in months for 48 integrated circuits. The times are assumed to follow a Weibull distribution.

```
data failures;
   input time @@;
   label time='Time in Months';
   datalines;
 29.42 32.14 30.58 27.50 26.08 29.06 25.10 31.34
 29.14 33.96 30.64 27.32 29.86 26.28 29.68 33.76
 29.32 30.82 27.26 27.92 30.92 24.64 32.90 35.46
 30.28 28.36 25.86 31.36 25.26 36.32 28.58 28.88
 26.72 27.42 29.02 27.54 31.60 33.46 26.78 27.82
 29.18 27.94 27.66 26.42 31.00 26.64 31.44 32.52
 ;
```

Three-Parameter Weibull Plots

See CAPQQ3 in the SAS/QC Sample Library

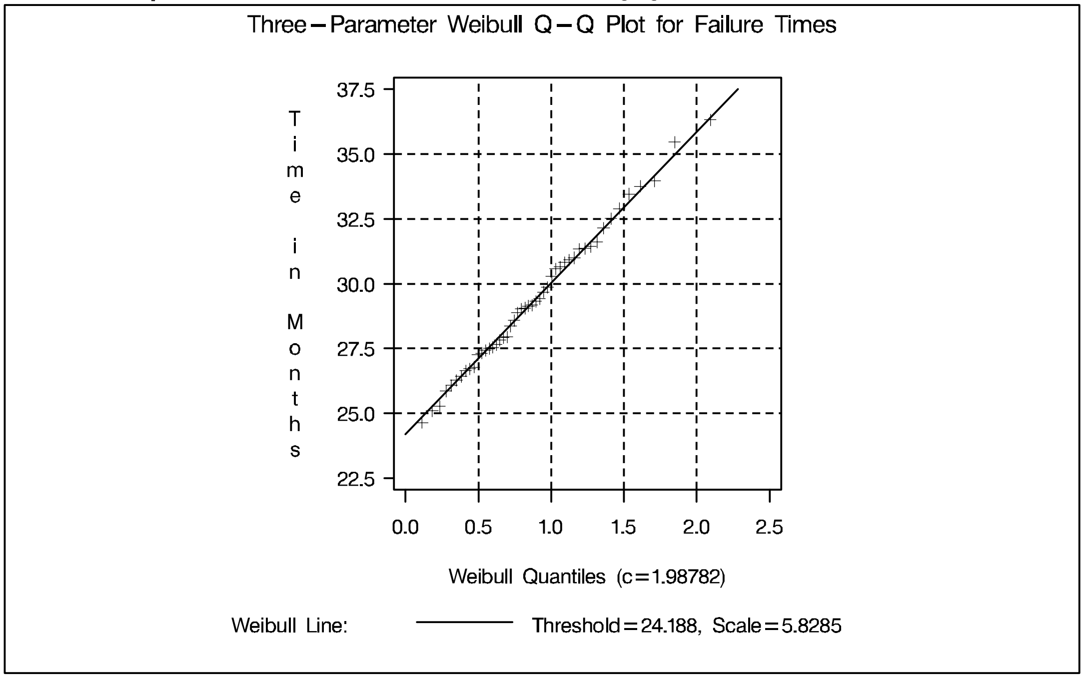
If no assumption is made about the parameters of this distribution, you can use the WEIBULL option to request a three-parameter Weibull plot. As in the previous example, you can visually estimate the shape parameter c by requesting plots for different values of c and choosing the value of c that linearizes the point pattern. Alternatively, you can request a maximum likelihood estimate for c, as illustrated in the following statements produce Weibull plots for c =1, 2 and 3:

```
title 'Three-Parameter Weibull Q-Q Plot for Failure Times';
proc capability data=failures noprint;
   qqplot time / weibull(c=est theta=est sigma=est)
                 square
                 href=0.5 1 1.5 2
                 vref=25 27.5 30 32.5 35;
run;
```

Note: When using the WEIBULL option, you must either specify a list of values for the Weibull shape parameter c with the C= option, or you must specify C=EST.

Output 10.3.1 displays the plot for the estimated value $c = 1.99$. The reference line corresponds to the estimated values for the threshold and scale parameters of ($\hat{\theta}_0$=24.19 and $\hat{\sigma}_0$=5.83, respectively.

Output 10.3.1. Three-Parameter Weibull Q-Q Plot for $c = 2$



Two-Parameter Weibull Plots

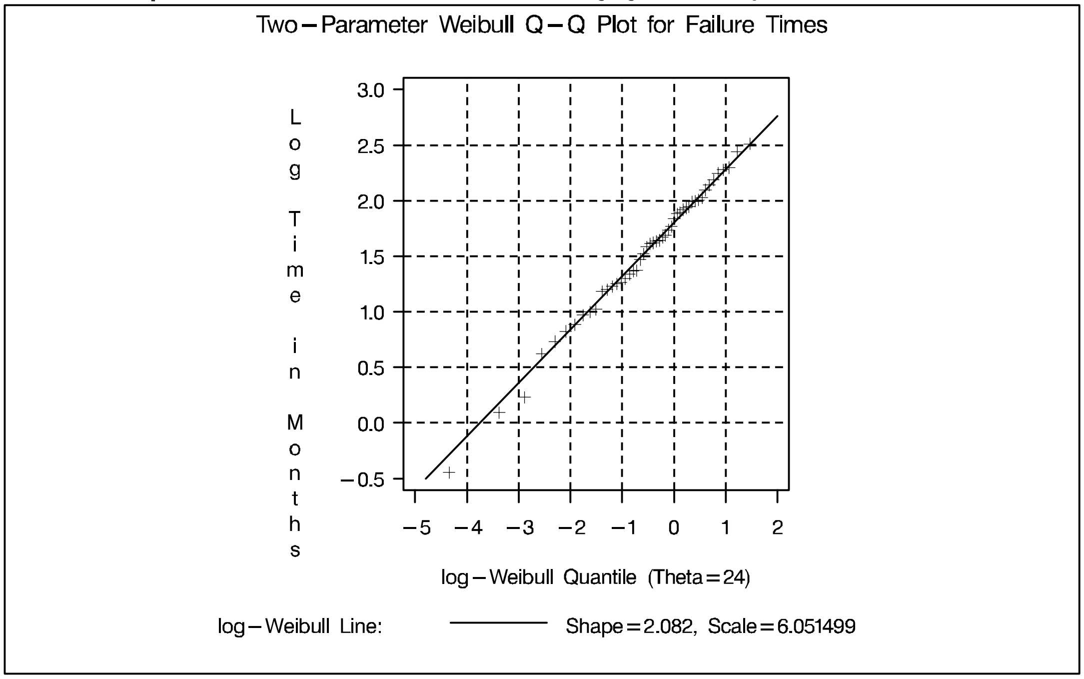
See CAPQQ3 in the SAS/QC Sample Library

Now, suppose it is known that the circuit lifetime is at least 24 months. The following statements use the threshold value $\theta_0 = 24$ to produce the two-parameter Weibull Q-Q plot shown in Output 10.3.2:

```
title 'Two Parameter Weibull Q-Q Plot for Failure Times';
proc capability data=failures noprint;
   qqplot time / weibull2(theta=24 c=est sigma=est)
                 square
                 href= -4 to 1
                 vref= 0 to 2.5 by 0.5;
run;
```

The reference line is based on maximum likelihood estimates $\hat{c}$=2.08 and $\hat{\sigma}$=6.05. These estimates agree with those of the previous example.

Output 10.3.2. Two-Parameter Weibull Q-Q Plot for $\theta_0 = 24$



Example 10.4. Estimating Cpk from a Normal Q-Q Plot

See CAPQQ1 in the SAS/QC Sample Library

This example illustrates how you can use a normal Q-Q plot to estimate the capability index C_{pk}. The data used here are the distance measurements provided in the "Creating a Normal Quantile-Quantile Plot" section on page 308.

The linearity of the point pattern in Figure 10.3 on page 310 indicates that the measurements are normally distributed (recall that normality should be checked when process capability indices are reported). Furthermore, Figure 10.3 shows that the upper specification limit is about 1.7 standard deviation units above the mean, and the lower specification limit is about 1.8 standard deviation units below the mean. Since *CPU* is defined as

$$CPU = \frac{USL - \mu}{3\sigma}$$

and *CPL* is defined as

$$CPL = \frac{\mu - LSL}{3\sigma}$$

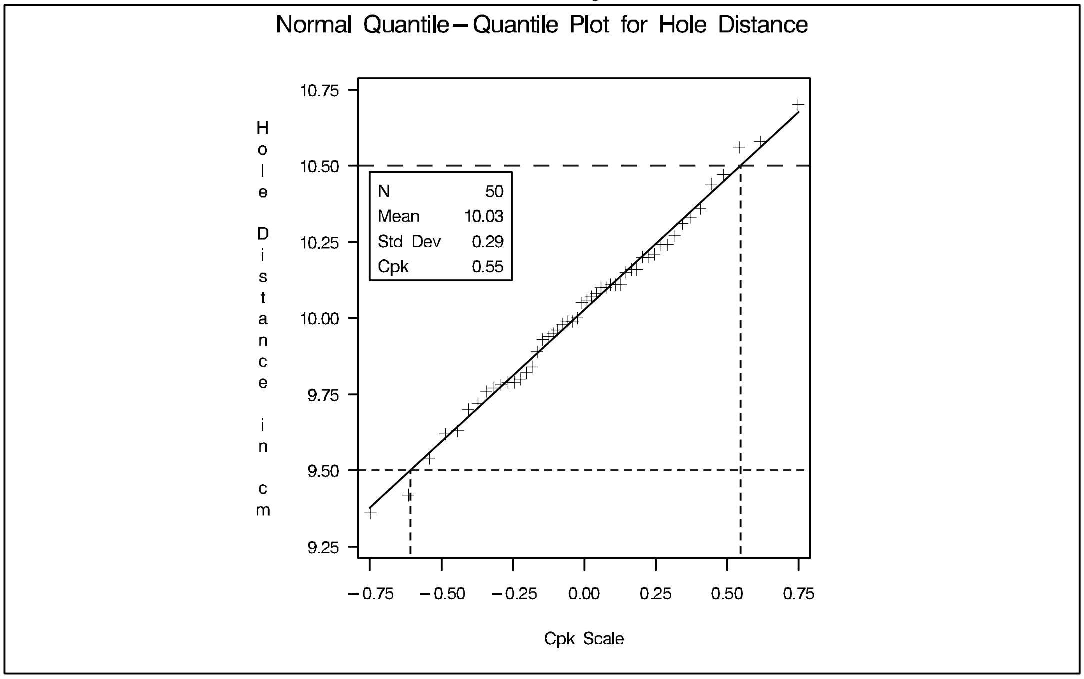
it follows that an estimate of *CPU* is $1.7/3 = 0.57$, and an estimate of *CPL* is $1.8/3 = 0.6$. Thus, except for a factor of three, you can estimate *CPU* and *CPL* from the points of intersection between the specification lines and the point pattern.

The following statements facilitate this type of estimation by creating a Q-Q plot, displayed in Output 10.4.1, in which the horizontal axis is rescaled by a factor of three:

```
title 'Normal Quantile-Quantile Plot for Hole Distance';
proc capability data=sheets noprint;
   spec  lsl=9.5   llsl=2   clsl=blue
         usl=10.5  lusl=20  cusl=blue;
   qqplot distance / normal(mu=est sigma=est cpkscale cpkref)
                     nolegend
                     square;
   inset n mean (5.2) std='Std Dev' (4.2) cpk (4.2) /
         pos=(-0.75,10.48) data refpoint=tl;
run;
```

The CPKSCALE option rescales the horizontal axis, and the CPKREF option adds reference lines indicating the intersections of the distribution reference line and the specification limits.

Output 10.4.1. Normal Q-Q Plot With C_{pk} Scaling



Using this display, you can estimate *CPU* and *CPL* directly from the horizontal axis as 0.55 and 0.60, respectively (the negative sign for -0.60 is ignored). The minimum of these values (0.55) is an estimate of C_{pk}. Note that this estimate agrees with the numerically obtained estimate for C_{pk} that is displayed on the plot with the INSET statement.

See Rodriguez (1992) for further discussion concerning the use of Q-Q plots in process capability analysis.

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.