Chapter 2
CDFPLOT Statement

Chapter Table of Contents

Chapter 2
CDFPLOT Statement

Overview

The CDFPLOT statement plots the observed cumulative distribution function (*cdf*) of a variable, defined as

$$
\begin{aligned}
F_N(x) &= \text{percent of nonmissing values} \leq x \\
&= \frac{\text{number of values} \leq x}{N} \times 100\%
\end{aligned}
$$

where N is the number of nonmissing observations. The cdf is an increasing step function that has a vertical jump of $\frac{1}{N}$ at each value of x equal to an observed value. The cdf is also referred to as the empirical cumulative distribution function (ecdf).

You can use options in the CDFPLOT statement to

- superimpose specification limits

- superimpose fitted theoretical distributions (beta, exponential, gamma, lognormal, normal, and Weibull)

- specify graphical enhancements (such as color or text height)

Getting Started

Creating a Cumulative Distribution Plot

This section introduces the CDFPLOT statement with a simple example. A company that produces fiber optic cord is interested in the breaking strength of the cord. The following statements create a data set named CORD, which contains 50 breaking strengths measured in pounds per square inch (psi), and they display the cdf plot in Figure 2.1. The plot shows a symmetric distribution with observations concentrated 6.9 and 7.1. The plot also shows that only a small percentage ($< 5\%$) of the observations are below the lower specification limit of 6.8.

```
data cord;
    label strength = 'Breaking Strength (psi)';
    input strength @@;
datalines;
6.94 6.97 7.11 6.95 7.12 6.70 7.13 7.34 6.90 6.83
7.06 6.89 7.28 6.93 7.05 7.00 7.04 7.21 7.08 7.01
7.05 7.11 7.03 6.98 7.04 7.08 6.87 6.81 7.11 6.74
6.95 7.05 6.98 6.94 7.06 7.12 7.19 7.12 7.01 6.84
6.91 6.89 7.23 6.98 6.93 6.83 6.99 7.00 6.97 7.01
;

title 'Cumulative Distribution Function of Breaking Strength';
proc capability data=cord noprint;
    var strength;
    spec lsl=6.8 llsl=2;
    cdfplot strength;
run;
```
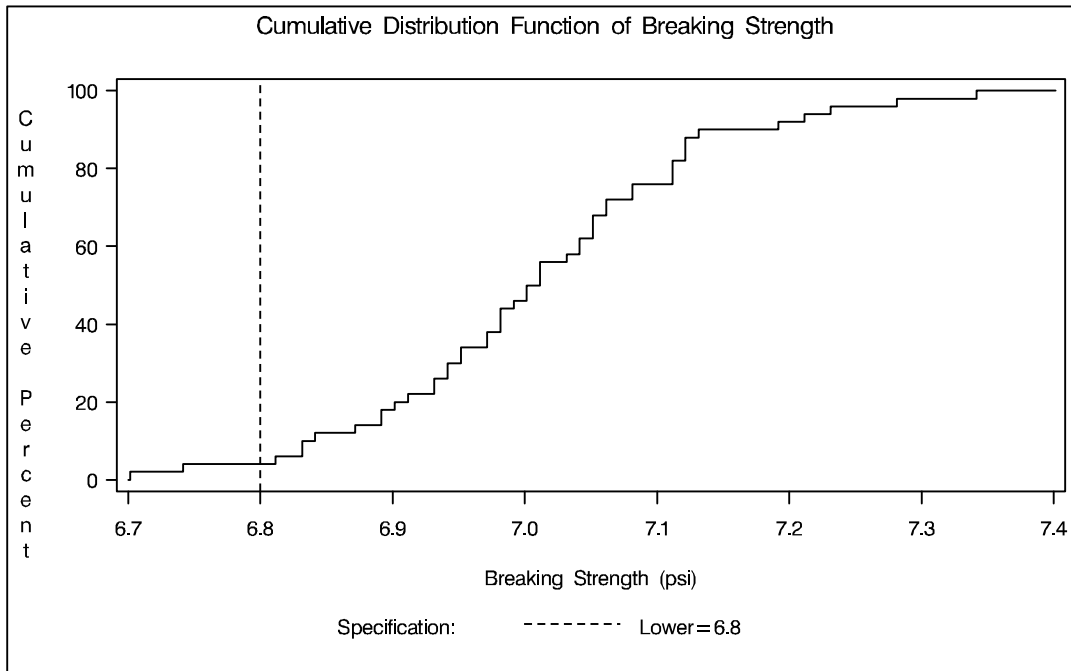

Figure 2.1. Cumulative Distribution Function

Syntax

The syntax for the CDFPLOT statement is as follows:

CDFPLOT *<variables> < **/** options >*;

You can specify the keyword CDF as an alias for CDFPLOT. You can specify any number of CDFPLOT statements after a PROC CAPABILITY statement. The components of the CDFPLOT statement are described as follows:

variables

specify variables for which to create cdf 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 *variables* in a CDFPLOT statement, then a cdf plot is created for each variable listed in the VAR statement, or for each numeric variable in the input data set if you do not use a VAR statement.

For example, suppose a data set named STEEL contains exactly three numeric variables, LENGTH, WIDTH and HEIGHT. The following statements create a cdf plot for each of the three variables:

```
proc capability data=steel;
   cdfplot;
run;
```

The following statements create a cdf plot for LENGTH and a cdf plot for WIDTH:

```
proc capability data=steel;
   var length width;
   cdfplot;
run;
```

The following statements create a cdf plot for WIDTH:

```
proc capability data=steel;
   var length width;
   cdfplot width;
run;
```

By default, the horizontal axis of a cdf plot is labeled with the *variable* name. If you specify a label for a variable, however, the label is used. The default vertical axis label is *Cumulative Percent*, and the axis is scaled in percent of observations.

If you specify a SPEC statement or a SPEC= data set in addition to the CDFPLOT statement, then the specification limits for each *variable* are displayed as reference lines and are identified in a legend.

67

options

add features to plots. All *options* appear after the slash (/) in the CDFPLOT statement. In the following example, the NORMAL option superimposes a normal cdf on the plot, and the CTEXT= option specifies the color of the text.

```
proc capability data=steel;
   cdfplot length / normal
                    ctext = yellow;
run;
```

Summary of Options

The following tables list all *options* by function. The "Dictionary of Options" on page 71 describes each option in detail.

Distribution Options

You can use the options listed in Table 2.1 to superimpose a fitted theoretical distribution function on your cdf plot.

Table 2.1. Main Distribution Options

BETA(*beta-options*)	plots two-parameter beta distribution function, parameters θ and σ assumed known
EXPONENTIAL(*exponential-options*)	plots one-parameter exponential distribution function, parameter θ assumed known
GAMMA(*gamma-options*)	plots two-parameter gamma distribution function, parameter θ assumed known
LOGNORMAL(*lognormal-options*)	plots two-parameter lognormal distribution function, parameter θ assumed known
NORMAL(*normal-options*)	plots normal distribution function
WEIBULL(*Weibull-options*)	plots two-parameter Weibull distribution function, parameter θ assumed known

You can specify options in parentheses after each distribution option to control features of the theoretical distribution function. For example, the following statements use the NORMAL option to superimpose a normal distribution:

```
proc capability;
   cdfplot / normal(mu=10 sigma=0.5 color=red);
run;
```

The COLOR= option specifies the color for the curve, and the *normal-options* MU= and SIGMA= specify the parameters $\mu = 10$ and $\sigma = 0.5$ for the distribution function. If you do not specify these parameters, maximum likelihood estimates are computed.

Table 2.2. Options Used with All Distribution Options

COLOR=*color*	specifies color of theoretical distribution function
L=*linetype*	specifies line type of theoretical distribution function
SYMBOL='*character*'	specifies *character* used to plot theoretical distribution function if cdf plot is produced on a line printer
W=*n*	specifies width of theoretical distribution function

Table 2.3. Beta-Options

ALPHA=*value*	specifies first shape parameter α for beta distribution function
BETA=*value*	specifies second shape parameter β for beta distribution function
SIGMA=*value*	specifies scale parameter σ for beta distribution function
THETA=*value*	specifies lower threshold parameter θ for beta distribution function

Table 2.4. Exponential-Options

SIGMA=*value*	specifies scale parameter σ for exponential distribution function
THETA=*value*	specifies threshold parameter θ for exponential distribution function

Table 2.5. Gamma-Options

ALPHADELTA=*value*	specifies change in successive estimates of α at which the Newton-Raphson approximation of $\hat{\alpha}$ terminates
ALPHAINITIAL=*value*	specifies initial value for α in the Newton-Raphson approximation of $\hat{\alpha}$
MAXITER=*n*	specifies maximum number of iterations in the Newton-Raphson approximation of $\hat{\alpha}$
SIGMA=*value*	specifies scale parameter σ for gamma distribution function
ALPHA=*value*	specifies shape parameter α for gamma distribution function
THETA=*value*	specifies threshold parameter θ for gamma distribution function

Table 2.6. Lognormal-Options

ZETA=*value*	specifies scale parameter ζ for lognormal distribution function
SIGMA=*value*	specifies shape parameter σ for lognormal distribution function
THETA=*value*	specifies threshold parameter θ for lognormal distribution function

Table 2.7. Normal-Options

MU=*value*	specifies mean μ for normal distribution function
SIGMA=*value*	specifies standard deviation σ for normal distribution function

Table 2.8. Weibull-Options

C=*value*	specifies shape parameter c for Weibull distribution function
CDELTA=*value*	specifies change in successive estimates of c at which the Newton-Raphson approximation of $\hat{c}$ terminates
CINITIAL=*value*	specifies initial value for c in the Newton-Raphson approximation of $\hat{c}$
MAXITER=*value*	specifies maximum number of iterations in the Newton-Raphson approximation of $\hat{c}$
SIGMA=*value*	specifies scale parameter σ for Weibull distribution function
THETA=*value*	specifies threshold parameter θ for Weibull distribution function

General Options

Table 2.9. Options to Enhance Plots Produced on Graphics Devices

ANNOTATE= *SAS-data-set*	specifies 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*	specifies AXIS statement for horizontal axis
HMINOR=*n*	specifies number of horizontal minor tick marks
LEGEND=*name* | NONE	identifies LEGEND statement
LHREF=*linetype*	specifies line style for HREF= lines
LVREF=*linetype*	specifies line style for VREF= lines
NAME='*string*'	specifies name for plot in graphics catalog
VAXIS=*name*	specifies AXIS statement for vertical axis
VMINOR=*n*	specifies number of vertical minor tick marks

Table 2.10. Options to Enhance Plots Produced on Line Printers

CDFSYMBOL='*character*'	specifies character for plotted points
HREFCHAR='*character*'	specifies line character for HREF= lines
VREFCHAR='*character*'	specifies line character for VREF= lines

70

Table 2.11. General Plot Layout Options

HREF=*value-list*	specifies reference lines perpendicular to the horizontal axis
HREFLABELS= 'label1'...'labeln'	specifies labels for HREF= lines
NOCDFLEGEND	suppresses legend for superimposed theoretical cdf
NOECDF	suppresses plot of empirical (observed) distribution function
NOFRAME	suppresses frame around plotting area
NOLEGEND	suppresses legend
NOSPECLEGEND	suppresses specifications legend
VREF=*value-list*	specifies reference lines perpendicular to the vertical axis
VREFLABELS= 'label1'...'labeln'	specifies labels for VREF= lines
VSCALE=PERCENT | PROPORTION	specifies scale for vertical axis

Dictionary of Options

The following entries provide detailed descriptions of the *options* in the CDFPLOT statement. The marginal notes *Graphics* and *Line Printer* identify options that can be used only with graphics devices and line printers, respectively.

ALPHA=*value*

specifies the shape parameter α for distribution functions requested with the BETA and GAMMA options. Enclose the ALPHA= option in parentheses after the BETA or GAMMA keywords. If you do not specify a value for α, the procedure calculates a maximum likelihood estimate. For examples, see the entries for the BETA and GAMMA options.

ALPHADELTA=*value*

specifies the change in successive estimates of $\hat{\alpha}$ at which iteration terminates in the Newton-Raphson approximation of the maximum likelihood estimate of α for curves requested by the GAMMA option. Enclose the ALPHADELTA= option in parentheses after the GAMMA keyword. Iteration continues until the change in α is less than the value specified or the number of iterations exceeds the value of the MAXITER= option (see page 76). The default value is 0.00001.

ALPHAINITIAL=*value*

specifies the initial value for $\hat{\alpha}$ in the Newton-Raphson approximation of the maximum likelihood estimate of α for fitted gamma distributions requested with the GAMMA option. Enclose the ALPHAINITIAL= option in parentheses after the GAMMA keyword. The default value is Thom's approximation of the estimate of α (refer to Johnson *et al.* (1995).

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

specifies an annotate data set, as described in *SAS/GRAPH Software: Reference*, that allows you to add features to the cdf plot. The ANNOTATE= data set you specify in the CDFPLOT statement is used for all plots created by the statement. You can also specify an ANNOTATE= data set in the PROC CAPABILITY statement, which provides annotate information used for all plots created by the procedure (see "AN-NOTATE= Data Sets" on page 31).

BETA<**(***beta-options* **)**>

displays a fitted beta distribution function on the cdf plot. The equation of the fitted cdf is

$$
F(x) = \begin{cases} 0 & \text{for } x \leq \theta \\ I_{\frac{x-\theta}{\sigma}}(\alpha, \beta) & \text{for } \theta < x < \theta + \sigma \\ 1 & \text{for } x \geq \sigma + \theta \end{cases}
$$

where $I_y(\alpha, \beta)$ is the incomplete beta function, and

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

The beta distribution is bounded below by the parameter θ and above by the value $\theta + \sigma$. You can specify θ and σ using the THETA= and SIGMA= *beta-options*, as illustrated in the following statements, which fit a beta distribution bounded between 50 and 75. The default values for θ and σ are 0 and 1, respectively.

```
proc capability;
    cdfplot / beta(theta=50 sigma=25);
run;
```

The beta distribution has two shape parameters, α and β. If these parameters are known, you can specify their values with the ALPHA= and BETA= *beta-options*. If you do not specify values for α and β, the procedure calculates maximum likelihood estimates.

The BETA option can appear only once in a CDFPLOT statement. Table 2.2 on page 69 and Table 2.3 on page 69 list options you can specify with the BETA distribution option.

BETA=*value*
B=*value*

specifies the second shape parameter β for beta distribution functions requested by the BETA option. Enclose the BETA= option in parentheses after the BETA keyword. If you do not specify a value for β, the procedure calculates a maximum likelihood estimate. For examples, see the preceding entry for the BETA option.

C=*value*

specifies the shape parameter *c* for Weibull distribution functions requested with the WEIBULL option. Enclose the C= option in parentheses after the WEIBULL keyword. If you do not specify a value for *c*, the procedure calculates a maximum likelihood estimate. You can specify the SHAPE= option as an alias for the C= option.

CAXIS=*color*
CAXES=*color*

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

CDELTA=*value*

specifies the change in successive estimates of *c* at which iterations terminate in the Newton-Raphson approximation of the maximum likelihood estimate of *c* for fitted Weibull curves requested by the WEIBULL option. Enclose the CDELTA= option in parentheses after the WEIBULL keyword. Iteration continues until the change in *c* between consecutive steps is less than the *value* specified or until the number of iterations exceeds the value of the MAXITER= option (see page 76). The default value is 0.00001.

CDFSYMBOL='*character***'**

specifies the character used to plot the points when the cdf plot is produced on a line printer. The default is the plus sign (+). Use the SYMBOL statement to control the plotting symbol when the plot is produced on a graphics device. *Line Printer*

CFRAME=*color*
CFR=*color*

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

CHREF=*color*
CH=*color*

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

CINITIAL=*value*

specifies the initial value for $\hat{c}$ in the Newton-Raphson approximation of the maximum likelihood estimate of *c* for Weibull distributions requested by the WEIBULL option. The default value is 1.8 (refer to Johnson *et al.* 1995).

COLOR=*color*

specifies the color of the fitted distribution curve. Enclose the COLOR= option in parentheses after a distribution option. For a syntax example, see page 68. *Graphics*

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. *Graphics*

CVREF=*color*
CV=*color*

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

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

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

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

displays a fitted exponential distribution function on the cdf plot. The equation of the fitted cdf is

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

where

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

The parameter θ must be less than or equal to the minimum data value. You can specify θ with the THETA= *exponential-option*. The default value for θ is 0. You can specify σ with the SIGMA= *exponential-option*. By default, a maximum likelihood estimate is computed for σ. For example, the following statements fit an exponential distribution with $\theta = 10$ and a maximum likelihood estimate for σ:

```
proc capability;
    cdfplot / exponential(theta=10 l=2 color=green);
run;
```

The exponential curve is green and has a line type of 2.

The EXPONENTIAL option can appear only once in a CDFPLOT statement. Table 2.2 on page 69 and Table 2.4 on page 69 list the options you can specify with the EXPONENTIAL option.

FONT=*font*

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

GAMMA<**(***gamma-options***)**>

displays a fitted gamma distribution function on the cdf plot. The equation of the fitted cdf is

$$F(x) = \begin{cases} 0 & \text{for } x \leq \theta \\ \frac{1}{\Gamma(\alpha)\sigma} \int_{\theta}^{x} \left(\frac{t-\theta}{\sigma}\right)^{\alpha-1} \exp\left(-\frac{t-\theta}{\sigma}\right) dt & \text{for } x > \theta \end{cases}$$

74

where

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

The parameter θ for the gamma distribution must be less than the minimum data value. You can specify θ with the THETA= *gamma-option*. The default value for θ is 0. In addition, the gamma distribution has a shape parameter α and a scale parameter σ. You can specify these parameters with the ALPHA= and SIGMA= *gamma-options*. By default, maximum likelihood estimates are computed for α and σ. For example, the following statements fit a gamma distribution function with $\theta = 4$ and maximum likelihood estimates for α and σ:

```
proc capability;
    cdfplot / gamma(theta=4);
run;
```

Note that the maximum likelihood estimate of α is calculated iteratively using the Newton-Raphson approximation. The *gamma-options* ALPHADELTA=, ALPHAINITIAL=, and MAXITER= control the approximation.

The GAMMA option can appear only once in a CDFPLOT statement. Table 2.2 on page 69 and Table 2.5 on page 69 list the options you can specify with the GAMMA 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 hori- | *Graphics*
zontal 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. See Output 2.2.1 on page 83 for an example that uses the similar VREF= option. See also the entries for the CHREF=, HREFCHAR=, and LHREF= options.

HREFCHAR=*'character'*

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

HREFLABELS=*'label1' . . . 'labeln'*
HREFLABEL=*'label1' . . . 'labeln'*
HREFLAB=*'label1' . . . 'labeln'*

specifies labels for the 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. See Output 2.2.1 on page 83 for an example that uses the similar VREFLABELS= option.

75

LEGEND=_name_ **| NONE**

specifies the name of a LEGEND statement describing the legend for specification limit reference lines and superimposed distribution functions. Specifying LEGEND=NONE, which suppresses all legend information, is equivalent to specifying the NOLEGEND option.

LHREF=_linetype_

LH=_linetype_

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

LOGNORMAL<**(**_lognormal-options_**)**>

displays a fitted lognormal distribution function on the cdf plot. The equation of the fitted cdf is

$$
F(x) = \begin{cases} 0 & \text{for } x \le \theta \\ \Phi\left(\frac{\log(x-\theta)-\zeta}{\sigma}\right) & \text{for } x > \theta \end{cases}
$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and

$\theta =$ threshold parameter

$\zeta =$ scale parameter

$\sigma =$ shape parameter $(\sigma > 0)$

The parameter θ for the lognormal distribution must be less than the minimum data value. You can specify θ with the THETA= _lognormal-option_. The default value for θ is 0. In addition, the lognormal distribution has a shape parameter σ and a scale parameter ζ. You can specify these parameters with the SIGMA= and ZETA= _lognormal-options_. By default, maximum likelihood estimates are computed for σ and ζ. For example, the following statements fit a lognormal distribution function with $\theta = 10$ and maximum likelihood estimates for σ and ζ:

```
proc capability;
   cdfplot / lognormal(theta = 10);
run;
```

The LOGNORMAL option can appear only once in a CDFPLOT statement. Table 2.2 on page 69 and Table 2.6 on page 69 list options that you can specify with the LOGNORMAL option.

LVREF=_linetype_

LV=_linetype_

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

MAXITER=_n_

specifies the maximum number of iterations in the Newton-Raphson approximation of the maximum likelihood estimate of α for fitted gamma distributions requested with the GAMMA option and c for fitted Weibull distributions requested with the WEIBULL option. Enclose the MAXITER= option in parentheses after the GAMMA or WEIBULL keywords. The default value of n is 20.

MU=*value*

    specifies the parameter μ for normal distribution functions requested with the NOR-MAL option. Enclose the MU= option in parentheses after the NORMAL keyword. The default value is the sample mean. For an example, see the entry for the NOR-MAL option.

NAME=*'string'*

    specifies a name for the plot, up to eight characters, that appears in the PROC GRE-PLAY master menu. The default is 'CAPABILI'.   `| Graphics |`

NOCDFLEGEND

    suppresses the legend for the superimposed theoretical cumulative distribution function.

NOECDF

    suppresses the observed distribution function (the empirical cumulative distribution function) of the variable, which is drawn by default. This option allows you to create theoretical cdf plots without displaying the data distribution. The NOECDF option can be used only with a theoretical distribution (such as the NORMAL option).

NOFRAME

    suppresses the frame around the subplot area.

NOLEGEND

    suppresses legends for specification limits, theoretical distribution functions, and hidden observations. Specifying the NOLEGEND option is equivalent to specifying LEGEND=NONE.

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

    displays a fitted normal distribution function on the cdf plot. The equation of the fitted cdf is

$$F(x) = \; \Phi\left(\tfrac{x-\mu}{\sigma}\right) \quad \text{for } -\infty < x < \infty$$

    where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and

    $\mu = \text{mean}$
    $\sigma = \text{standard deviation } (\sigma > 0)$

    You can specify known values for μ and σ with the MU= and SIGMA= *normal-options*, as shown in the following statements:

```
proc capability;
   cdfplot / normal(mu=14 sigma=.05);
run;
```

    By default, the sample mean and sample standard deviation are calculated for μ and σ. The NORMAL option can appear only once in a CDFPLOT statement. For an example, see Output 2.1.1 on page 81. Table 2.2 on page 69 and Table 2.7 on page 69 list options that you can specify with the NORMAL option.

NOSPECLEGEND
NOSPECL

    suppresses the portion of the legend for specification limit reference lines.

SCALE=*value*

    is an alias for the SIGMA= option for distributions requested by the BETA, EXPO-NENTIAL, GAMMA, and WEIBULL options and for the ZETA= option for distributions requested by the LOGNORMAL option.

SHAPE=*value*

    is an alias for the ALPHA= option for distributions requested by the GAMMA option, for the SIGMA= option for distributions requested by the LOGNORMAL option, and for the C= option for distributions requested by the WEIBULL option.

SIGMA=*value*

    specifies the parameter σ for distribution functions requested by the BETA, EXPONENTIAL, GAMMA, LOGNORMAL, NORMAL, and WEIBULL options. Enclose the SIGMA= option in parentheses after the distribution keyword. The following table summarizes the use of the SIGMA= option:

Distribution Option	SIGMA= Specifies	Default Value	Alias
BETA	scale parameter σ	1	SCALE=
EXPONENTIAL GAMMA WEIBULL	scale parameter σ	maximum likelihood estimate	SCALE=
LOGNORMAL	shape parameter σ	maximum likelihood estimate	SHAPE=
NORMAL	scale parameter σ	standard deviation	

SYMBOL='*character***'**

    specifies the *character* used to plot the theoretical distribution function if the cdf plot is produced on a line printer. Enclose the SYMBOL= option in parentheses after the distribution option. The default character is the first letter of the distribution option keyword.

THETA=*value*

    specifies the lower threshold parameter θ for theoretical cumulative distribution functions requested with the BETA, EXPONENTIAL, GAMMA, LOGNORMAL, and WEIBULL options. Enclose the THETA= option in parentheses after the distribution keyword. The default *value* is 0.

THRESHOLD=*value*

    is an alias for the THETA= option. See the preceding entry for the THETA= option.

VAXIS=*name*

    specifies the name of an AXIS statement describing the vertical axis. See Output 2.1.1 on page 81 for an example.

VMINOR=*n*
VM=*n*

    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 2.2.1 on page 83 for an example. See also the entries for the CVREF=, LVREF=, and VREFCHAR= options.

VREFCHAR=*'character'*

specifies the character used to form the 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 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. See Output 2.2.1 on page 83 for an example.

VSCALE=PERCENT | PROPORTION

specifies the scale of the vertical axis. The value PERCENT scales the data in units of percent of observations per data unit. The value PROPORTION scales the data in units of proportion of observations per data unit. The default is PERCENT.

W=*n*

specifies the width in pixels of the superimposed theoretical distribution. Enclose the W= option in parentheses after the distribution option. For example, the following statements display an exponential distribution with a width of 3. The default is 1.   *Graphics*

```
proc capability;
    cdfplot / exponential(w=3);
run;
```

WEIBULL<**(***Weibull-options***)**>

displays a fitted Weibull distribution function on the cdf plot. The equation of the fitted cdf is

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

where

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

The parameter θ must be less than the minimum data value. You can specify θ with the THETA= *Weibull-option*. The default value for θ is 0. In addition, the Weibull distribution has a shape parameter c and a scale parameter σ. You can specify these parameters with the SIGMA= and C= *Weibull-options*. By default, maximum likelihood estimates are computed for c and σ. For example, the following statements fit a Weibull distribution function with $\theta = 15$ and maximum likelihood estimates for σ and c:

<center>79</center>

```
proc capability;
   cdfplot / weibull(theta=15);
run;
```

Note that the maximum likelihood estimate of c is calculated iteratively using the Newton-Raphson approximation. The *Weibull-options* CDELTA=, CINITIAL=, and MAXITER= control the approximation.

The WEIBULL option can appear only once in a CDFPLOT statement. Table 2.2 on page 69 and Table 2.8 on page 70 list options that you can specify with the WEIBULL option.

ZETA=*value*

specifies a value for the scale parameter ζ for a lognormal distribution function requested with the LOGNORMAL option. Enclose the ZETA= option in parentheses after the LOGNORMAL keyword. If you do not specify a *value* for ζ, a maximum likelihood estimate is computed. You can specify the SCALE= option as an alias for the ZETA= option.

80

Examples

This section illustrates how to display a fitted distribution function, inset tables, and display reference lines on your cdf plot.

Example 2.1. Fitting a Normal Distribution

You can use the CDFPLOT statement to fit any of six theoretical distributions (beta, exponential, gamma, lognormal, normal, and Weibull) and superimpose them on the cdf plot. The following statements use the NORMAL option to display a fitted normal distribution function on a cdf plot of breaking strengths. The data set CORD is given in Figure 2.1 on page 66, and the plot is shown in Output 2.1.1.

See CAPCDF1 in the SAS/QC Sample Library

```
title 'Cumulative Distribution Function of Breaking Strength';
proc capability data=cord;
   spec lsl=6.8 llsl=2;
   cdfplot strength / normal
                      vaxis = axis1;
   inset mean std pctlss / cfill  = blank
                           format = 5.2
                           header = 'Summary Statistics';
   axis1 label=(a=90 r=0);
run;
```

Output 2.1.1.   Superimposed Normal Distribution Function
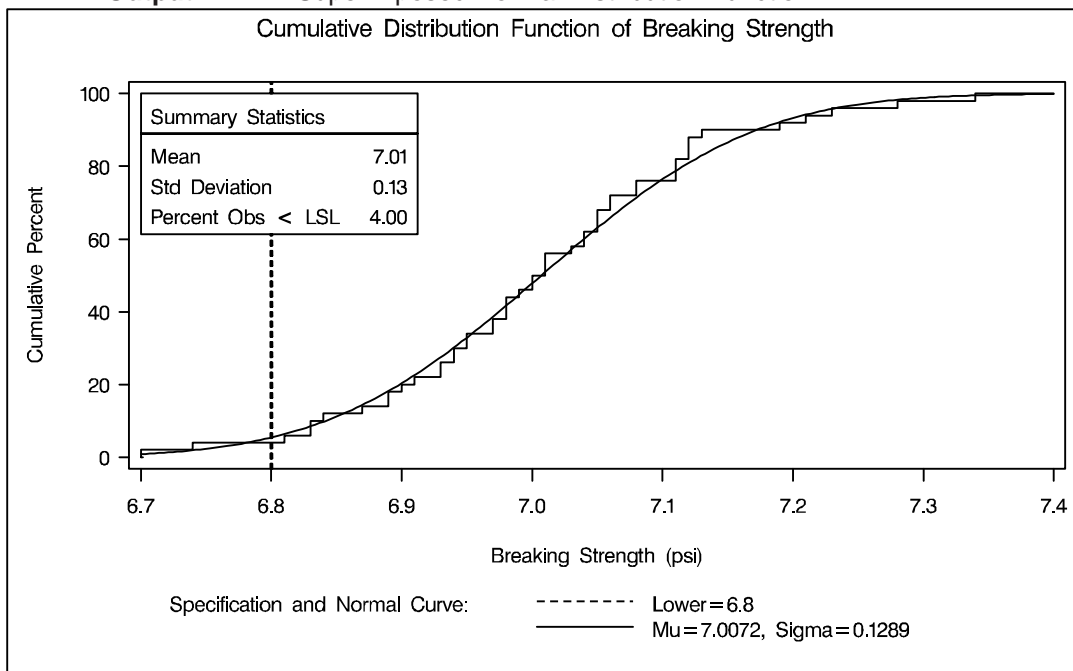
The NORMAL option requests the fitted curve. The VAXIS= option specifies the AXIS statement controlling the vertical axis. The AXIS1 statement is used to rotate the vertical axis label *Cumulative Percent*. The INSET statement requests an inset containing the mean, the standard deviation, and the percent of observations below the lower specification limit. For more information about the INSET statement, see

81

Chapter 5, "INSET Statement" starting on page 191. The SPEC statement requests a lower specification limit at 6.8 with a line type of 2 (a dashed line). For more information about the SPEC statement, see "Syntax for the SPEC Statement" on page 26.

The agreement between the empirical and the normal distribution functions in Output 2.1.1 is evidence that the normal distribution is an appropriate model for the distribution of breaking strengths.

The CAPABILITY procedure provides a variety of other tools for assessing goodness of fit. Goodness-of-fit tests (see "Printed Output" on page 157) provide a quantitative assessment of a proposed distribution. Probability and Q-Q plots, created with the PROBPLOT (page 275), QQPLOT (page 307), and PPPLOT (page 251) statements, provide effective graphical diagnostics.

Example 2.2. Using Reference Lines with CDF Plots

See CAPCDF1
in the SAS/QC
Sample Library

Customer requirements dictate that the breaking strengths in the previous example have upper and lower specification limits of 7.2 and 6.8 psi, respectively. Moreover, less than 5% of the cords can have breaking strengths outside the limits.

The following statements create a cdf plot with reference lines at the 5% and 95% cumulative percent levels:

```
title 'Cumulative Distribution Function of Breaking Strength';
proc capability data=cord noprint;
   var strength;
   spec lsl=6.8 llsl=2 usl=7.2 lusl=2;
   cdfplot strength /  vref = 5 95
                       vreflabels = '5%' '95%'
                       vaxis =axis1;
   inset pctgtr pctlss  / cfill  = blank
                          format = 5.2
                          pos    = e
                          header = 'Summary Statistics';
   axis1 label=(a=90 r=0);
run;
```

The INSET statement requests an inset with the percentages of measurements above the upper limit and below the lower limit. For more information about the INSET statement, see Chapter 5, "INSET Statement" starting on page 191.

In Output 2.2.1, the empirical cdf is below the intersection between the lower specification limit line and the 5% line, so less than 5% of the measurements are below the lower limit. The ecdf, however, is *also* below the intersection between the upper specification limit line and the 95% line, implying that *more* than 5% of the measurements are greater than the upper limit. Thus, the goal of having less than 5% of the measurements above the upper specification limit has not been met.

Output 2.2.1. Reference Lines with a Cumulative Distribution Function Plot
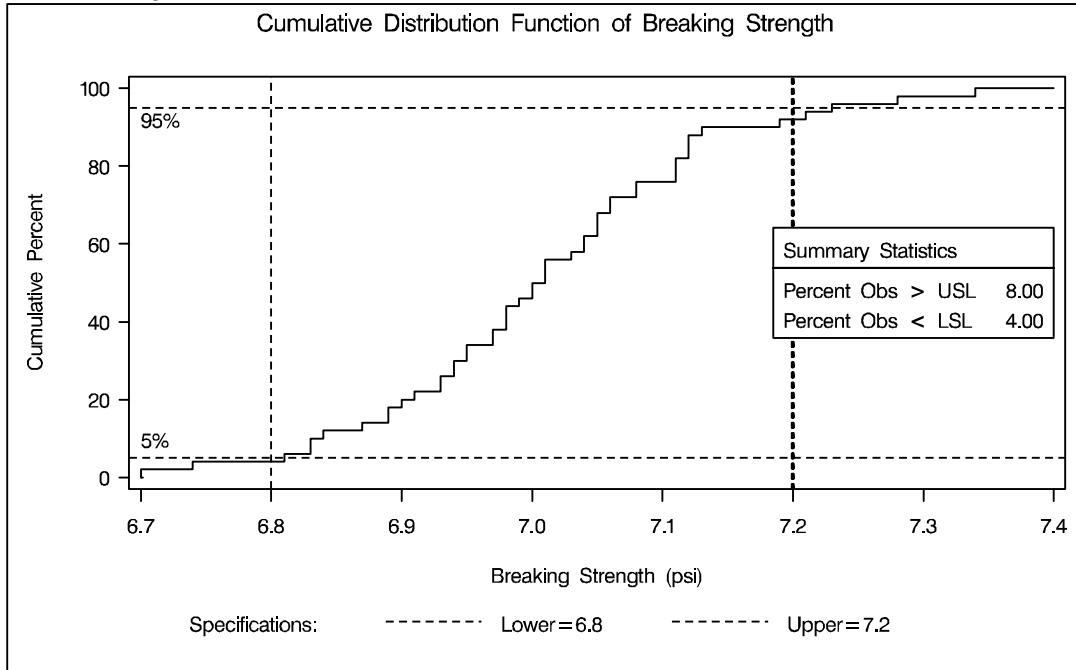
SAS/QC® User's Guide, Version 8