8.3 Using Process Capability Ratios

For an On-Center (On-Target) Process

- In Chapter 6, we defined $C_p = \dfrac{USL - LSL}{6\sigma}$ where USL and LSL are the upper and lower specification limits, and 6σ is the width of the specification band.
- When the process standard deviation σ is unknown, we replace σ with an estimate $\widehat{\sigma}$. Thus, we have

$$\widehat{C}_p = \frac{USL - LSL}{6\widehat{\sigma}} \tag{20}$$

- Recall an interpretation of C_p and $\widehat{C}_p$: $P = 100\left(\frac{1}{C_p}\right)$ and $P \approx 100\left(\frac{1}{\widehat{C}_p}\right)$ where P is the percentage of the specification band used by the process.
- For one-sided specifications, we define C_p as:

$$C_{pu} = \text{upper specification only}$$

$$C_{pl} = \text{lower specification only}$$

- Replace μ and σ with estimates to get $\widehat{C}_{pu}$ and $\widehat{C}_{pl}$.
- Recall that C_p is a measure of the ability of the process to meet specifications assuming
  - The actual process mean is the target mean μ.
  - The quality characteristic follows a normal (or near-normal) distribution.
- These assumptions are essential to the accuracy of the process capability ratio. The following table shows several C_p values and the process "fallout" in defective parts per million. If the assumptions are not true, then the values in this table are not accurate.

For an Off-Center (Off-Target) Process

- The process capability ratio C_p does not take into account where the process mean is located relative to the specifications. It simply measures the spread of the specifications relative to the 6σ spread in the process.
- A deficiency of using C_p is shown in the figure following the table. Many distributions can have the same C_p but the amount of defective product produced can vary significantly.
- This led to a process capability ratio that also takes into account the center of the process:

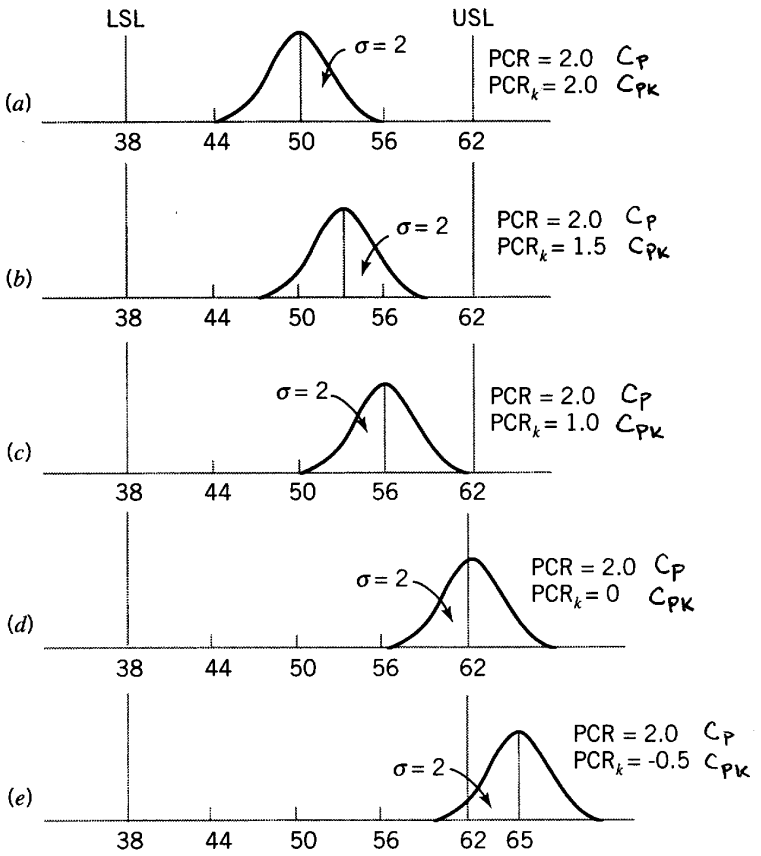
$$C_{pk} = \tag{21}$$

- Thus, C_{pk} is the one-sided C_p for the specification limit nearest to the process mean.
  - If $C_p = C_{pk}$ then the process is on-target (centered).
  - If $C_p > C_{pk}$ then the process is off-target (off-center).

- It is common to say the C_p measures the **potential capability** of the process while C_{pk} measures the **actual capability** of the process (assuming the distribution is normal).

Table Values of the Process-Capability Ratio (C_p) and Associated Process Fallout for a Normally Distributed Process (in Defective PPM)

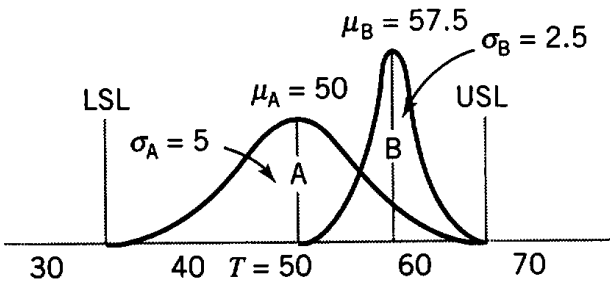
	Process Fallout (in defective PPM)	
C_p	One-Sided Specifications	Two-Sided Specifications
0.25	226,628	453,255
0.50	66,807	133,614
0.60	35,931	71,861
0.70	17,865	35,729
0.80	8,198	16,395
0.90	3,467	6,934
1.00	1,350	2,700
1.10	484	967
1.20	159	318
1.30	48	96
1.40	14	27
1.50	4	7
1.60	1	2
1.70	0.17	0.34
1.80	0.03	0.06
2.00	0.0009	0.0018



- Warning: Many quality 'experts' advise against routine use or dependency on process capability ratios because they tend to oversimplify the complexities of a process.
  - Any statistic that combines information about location and variability and which requires a normality assumption for meaningful interpretation is likely to be misused.
  - Also, process capability ratios are point estimates, and, therefore, are virtually useless when computed from small samples (because of the large variability of the estimate).
- One approach to deal with nonnormality is to transform the data so the transformed measurement is approximately normal. The specification limits will be transformed accordingly.

More on Process Centering

- C_{pk} was developed because C_p does not adequately deal with the case when the process is not centered within specifications.
- C_{pk}, however, has its flaws. That is, it does not adequately address all issues regarding an off-target process.
- For example, consider the following figure. Two distributions can have equal C_{pk} values yet be centered in different locations.



- In general, for any fixed value of μ in the interval (LSL, USL), C_{pk} depends inversely on σ. That is, C_{pk} becomes large as $\sigma \to 0$.
- Thus, C_{pk} can be highly influenced by σ and not the location of the process. A large C_{pk} value in itself does not tell us anything about the location of the mean within (LSL, USL).
- To address this problem, a third process capability ratio C_{pm} was developed. If $T = (USL - LSL)/2$ is the target value, then

$$C_{pm} = \frac{USL - LSL}{6\tau} \tag{22}$$

where τ is the square root of the expected squared deviation from target T. That is:

$$\begin{aligned} \tau^2 &= E[(x-T)^2] \\ &= \\ &= \\ &= \\ &= \end{aligned}$$

- Let $\xi = \dfrac{\mu - T}{\sigma}$. We can then rewrite (22) as

$$\begin{aligned} C_{pm} &= \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} \\ &= \\ &= \\ &= \end{aligned}$$

- An estimate of C_{pm} is $\widehat{C}_{pm} =$ where $V = \dfrac{T - \overline{x}}{s}$.

- The behavior of the process capability ratios:
  - If $\mu = T$, then $C_p = C_{pk} = C_{pm}$.
  - C_{pk} and C_{pm} decrease as μ moves away from T.
  - $C_{pk} < 0$ for $\mu > USL$ and $\mu < LSL$.
  - $C_{pm} \to 0$ as $|\mu - T| \to \infty$.

8.4 Process Capability Indices Assuming Non-normality

- A simple approach for assessing process capability with non-normal data is to transform the data so that the transformed values are approximately normal, or, at the very least, closer to being normal than the original data.

- For any transformation considered, the appropriateness of the transformation should be checked using graphical methods and goodness-of-fit tests.

- If a goodness-of-fit test has a large p-value (and plots indicate no problems), then it is reasonable to estimate process capability indices under that transformation.

- Although this approach may seem appealing, many statisticians/quality engineers avoid transformations so that results (especially endpoints of confidence intervals) do not have to be transformed back to the original scale.

- The most common alternative that avoids finding and using a transformation is to model the data with a probability distribution (e.g., Weibull, lognormal, gamma, beta, ...), and analogous to the transformation approach, the appropriateness of a fitted distribution should be checked using graphical methods and goodness-of-fit tests.

- Once an acceptable distribution is found, then quantile estimates are used to estimate "generalized" process capability indices. That is, estimate generalizations of the C_p, C_{pk}, and C_{pm} indices to a particular distribution.

- Let P_α be the $100\alpha^{\text{th}}$ percentile for a distribution and T be the process target.

- Generalized indices use the fact that for the normal distribution, the 3σ limits are the lower .135 percentile ($P_{.00135}$) and the upper 99.865 percentile ($P_{.99865}$). Or, 3σ is the distance from the median $P_{.5}$ to the control limits.

- We will apply the same criterion to any distribution yielding the **generalized capability indices**:

C_p = C_{pu} = C_{pl} =

C_{pk} = $\min(C_{pl}, C_{pu})$ C_{pm} =

  If the data are normally distributed, then $P_{.5} = \mu$, $P_{.99865} = \mu + 3\sigma$, and $P_{.00135} = \mu - 3\sigma$. These formulas then reduce to the ones defined earlier under the normality assumption.

- SAS and other software programs can be used to fit various distributions. For example, SAS has options for fitting Weibull, beta, exponential, gamma, lognormal, normal, gumbel, inverse gaussian, pareto, rayleigh, power function, and Johnson-type distributions.

- One problem with this approach is the potential for fitting distributions that are inconsistent with natural boundaries. For example, for a fitted distribution, there may be a positive probability associated with a set of values that are below or above any possible value that can be observed from that process. Generally, this will not be a major problem as long as the set of those impossible values has a very small probability assigned from the fitted distribution.

- Thus, in general for the non-normal data case, the accuracy of the generalized indices in assessing process capability is contingent on how well the fitted distribution performs as a model.

- For another approach, Clements (1989) proposed the use of Pearson curves which are claimed to provide improved quantile estimates, which in turn, are used to calculate generalized capability indices.

Example: A random sample of 50 measurements was collected. The data represents the gap between a plate and a substrate in an industrial process. The target value is 1.55 with specification limits at LSL=1.2 and USL=1.9. The following SAS code will:

- Generate output for assessing the normality of the data.

- Fit the data assuming a two-parameter lognormal distribution. That is, the support is $(0, \infty)$.

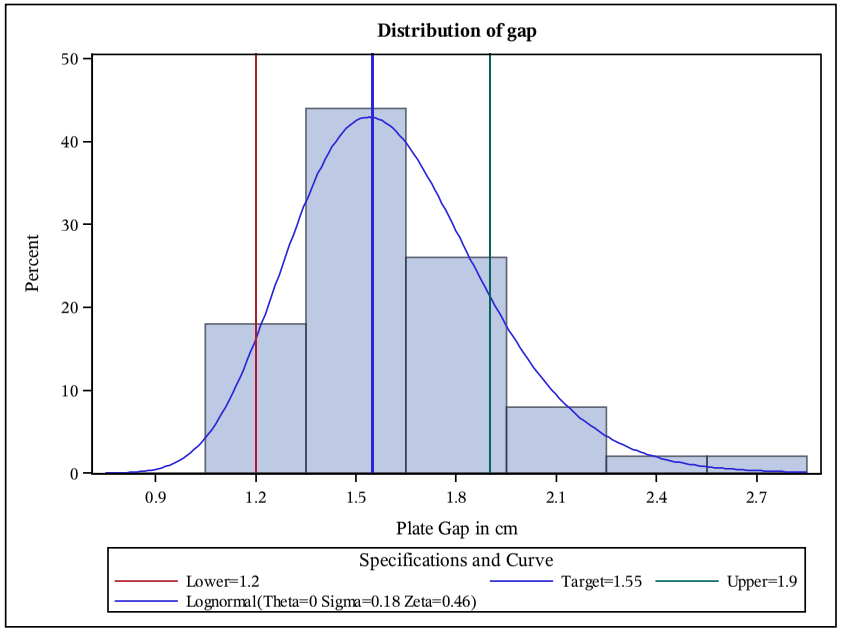
- Fit the data assuming three-parameter lognormal distributions with threshold values of $\theta =$ 0.5, 1.0, and 1.2. That is, the support is (θ, ∞).

- Fit the data assuming three-parameter lognormal distributions but let SAS estimate the threshold value of $\theta = 1.126$.

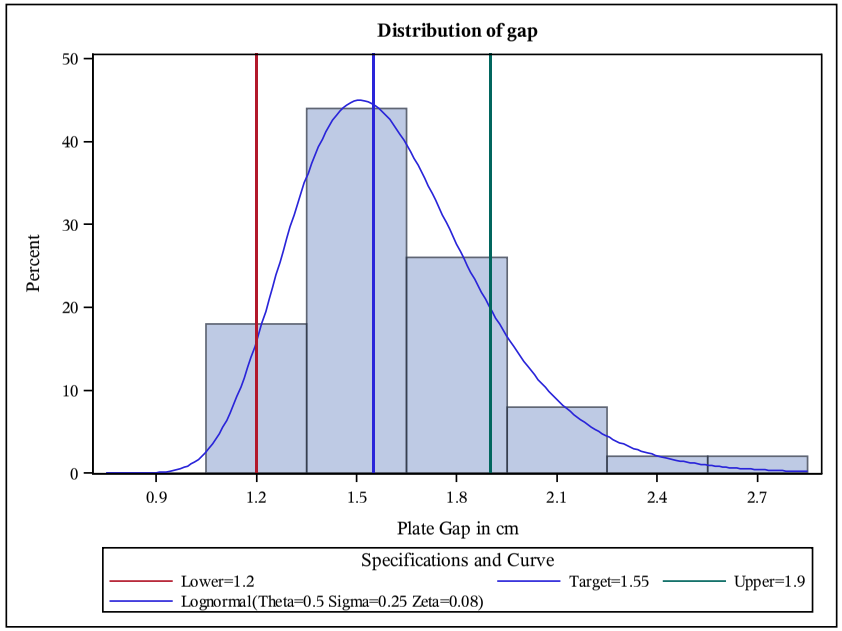
SAS Code for Plate Gap Data

```
DM 'LOG; CLEAR; OUT; CLEAR;';
* ODS PRINTER PDF file='C:\COURSES\ST528\SAS\cpcp.PDF';
ODS LISTING;
OPTIONS PS=80 LS=76 NODATE NONUMBER;

DATA plates;
    LABEL gap='Plate Gap in cm';
    INPUT gap @@;
LINES;
 1.746  1.357  1.376  1.327  1.485  2.741  1.241  1.777  1.768  1.409
 1.252  1.512  1.534  2.456  1.742  1.378  1.714  2.021  1.597  1.231
 1.541  1.805  1.682  1.418  1.506  1.501  1.247  1.922  1.880  1.344
 1.519  2.102  1.275  1.601  1.388  1.450  1.845  1.319  1.486  1.529
 2.247  1.690  1.676  1.314  1.736  1.643  1.483  1.352  1.636  1.980
 ;
SYMBOL1 VALUE=dot WIDTH=3 L=1;


PROC CAPABILITY data=plates ;
    SPECS LSL=1.2 USL=1.9 TARGET=1.55 ;
    HISTOGRAM gap / lognormal(indices);
    QQPLOT gap / lognormal(threshold=0 sigma=est);
 TITLE 'Capability Analysis of Plate Gaps -- Threshold=0';


 PROC CAPABILITY data=plates ;
   SPECS LSL=1.2 USL=1.9 TARGET=1.55 ;
   HISTOGRAM gap / lognormal(theta=.5 indices);
   QQPLOT gap / lognormal(threshold=.5 sigma=est);
 TITLE 'Capability Analysis of Plate Gaps -- Threshold=.5';


 PROC CAPABILITY data=plates ;
   SPECS LSL=1.2 USL=1.9 TARGET=1.55 ;
   HISTOGRAM gap / lognormal(theta=1  indices);
   QQPLOT gap / lognormal(threshold=1 sigma=est);
 TITLE 'Capability Analysis of Plate Gaps -- Threshold=1';


 PROC CAPABILITY data=plates ;
   SPECS LSL=1.2 USL=1.9 TARGET=1.55 ;
   HISTOGRAM gap / lognormal(theta=1.2  indices);
   QQPLOT gap / lognormal(threshold=1.2 sigma=est);
 TITLE 'Capability Analysis of Plate Gaps -- Threshold=1.2';


 PROC CAPABILITY data=plates ;
   SPECS LSL=1.2 USL=1.9 TARGET=1.55 ;
   HISTOGRAM gap / lognormal(theta=est  indices);
   QQPLOT gap / lognormal(threshold=est sigma=est);
 TITLE 'Capability Analysis of Plate Gaps -- Threshold Estimated';
RUN;
```

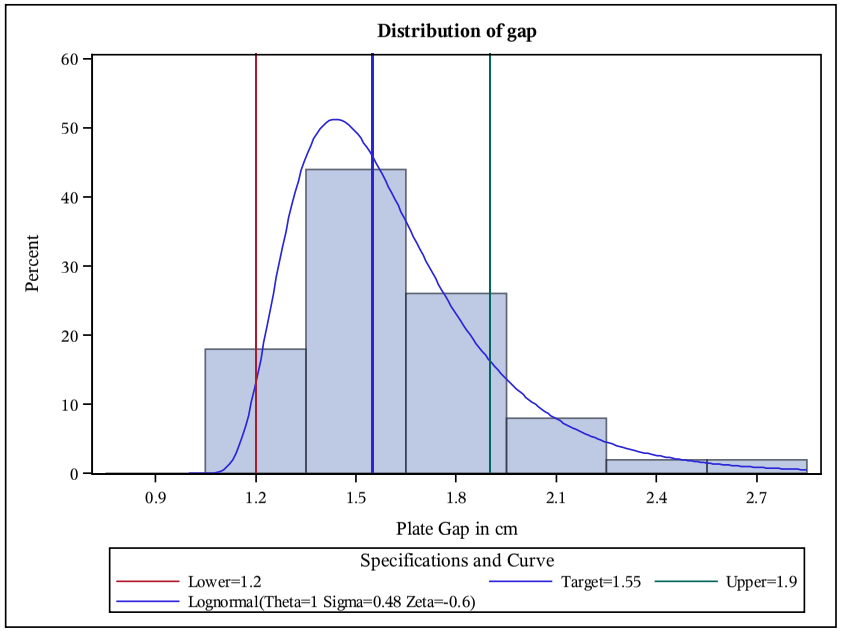
Capability Analysis of Plate Gaps -- Threshold=0



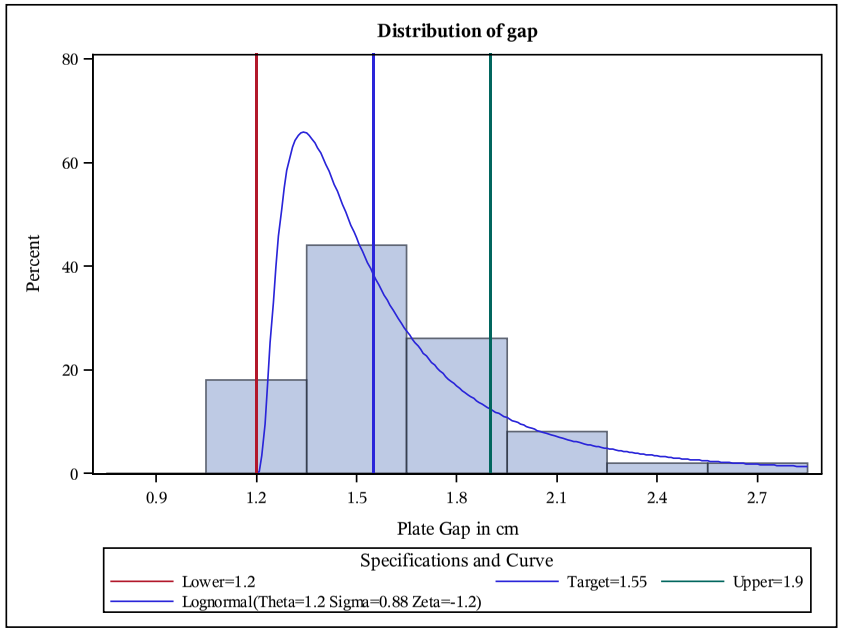
Capability Analysis of Plate Gaps -- Threshold=.5



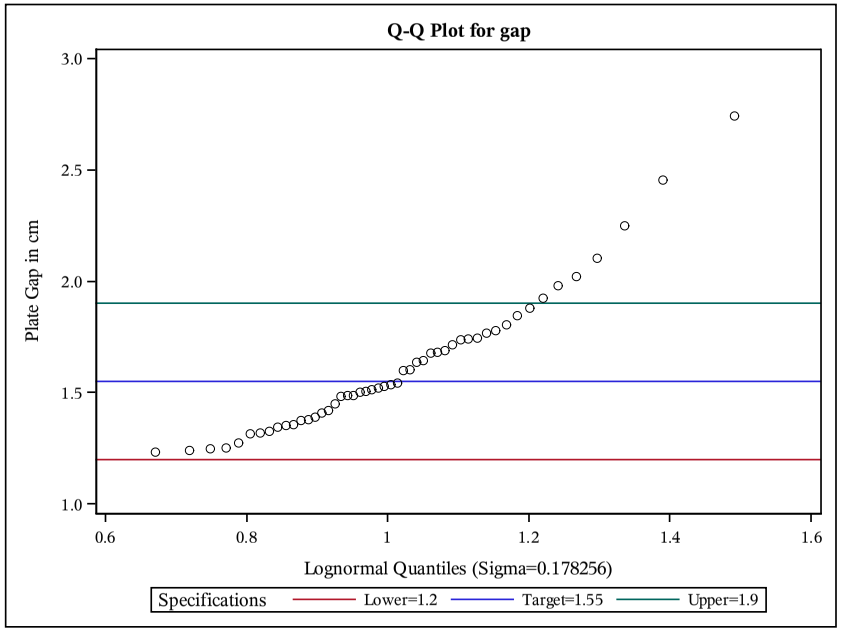
Capability Analysis of Plate Gaps -- Threshold=1



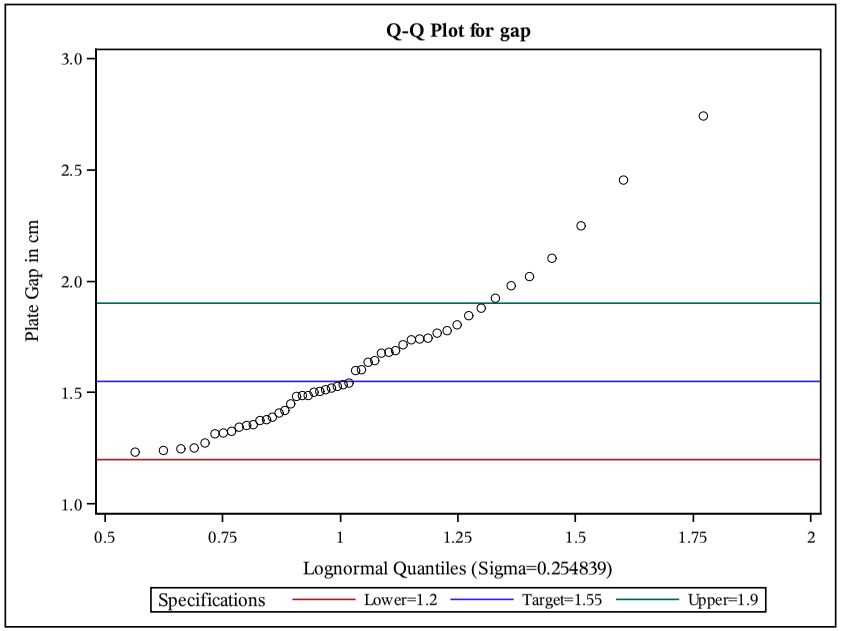
Capability Analysis of Plate Gaps -- Threshold=1.2



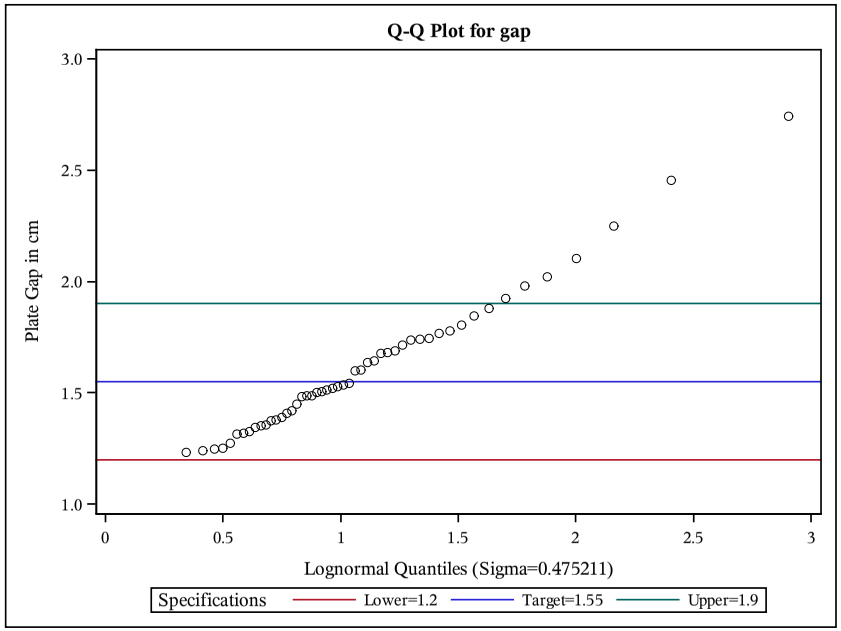
Capability Analysis of Plate Gaps -- Threshold=0



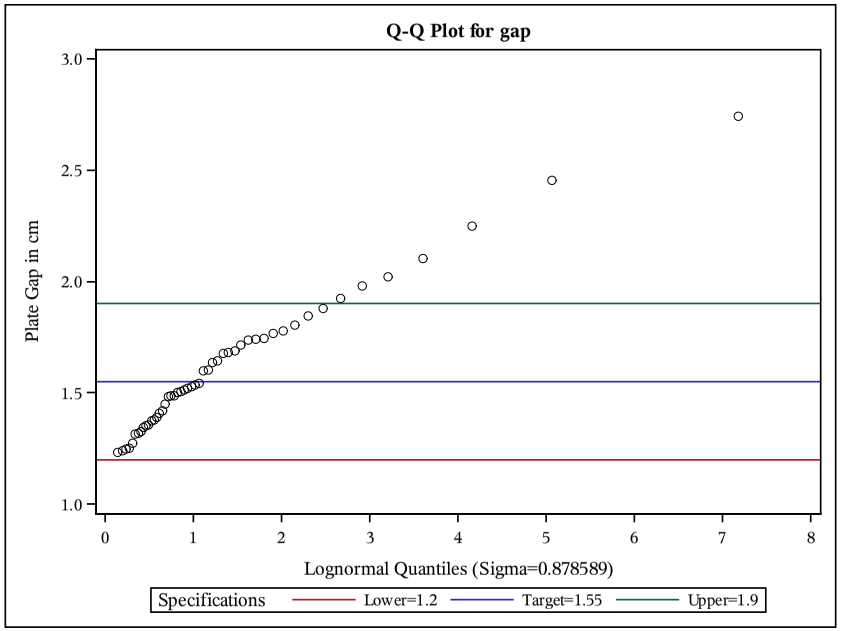
Capability Analysis of Plate Gaps -- Threshold=.5



Capability Analysis of Plate Gaps -- Threshold=1



Capability Analysis of Plate Gaps -- Threshold=1.2



- The small p-values for the Tests for Normality in (2A) all indicate rejection of the null hypothesis that the data come from a normal distribution. Therefore, the process capability indices (assuming normality) will not be reliable.

- The lower threshold θ defines the support of the three-parameter lognormal pdf. That is, the support is (θ, ∞). Or, in other words, if $\theta \neq 0$, we have included a shift parameter making it a 3-parameter lognormal distribution.

- For example, suppose we set the threshold at $\theta = .5$. This is equivalent to subtracting .5 from each data point and then fitting a two-parameter lognormal with the transformed $(Y - .5)$ data.

- Note that the smallest data value is 1.231. The two-parameter lognormal distribution, however, assumes the minimum is 0. Therefore, it is reasonable to also consider three-parameter lognormal distributions with a lower-threshold $\theta > 0$.

- Five threshold cases of a lognormal distribution were considered: (2B) $\theta = 0$, (2C) $\theta = 0.5$, (2D) $\theta = 1.0$, (2E) $\theta = 1.2$, and (2F) $\theta = 1.126$.

- The previous two pages contain plots of histograms with the fitted lognormal pdfs superimposed as well as the qq plots for threshold values of $\theta = 0.5,\ 1.0$, and 1.2.

- Based on these plots, I would not consider $\theta = 1.2$ as a good choice for the lognormal threshold parameter. This decision is supported by the following table of p-values summarizes the results of the goodness-of-fit test for lognormality.

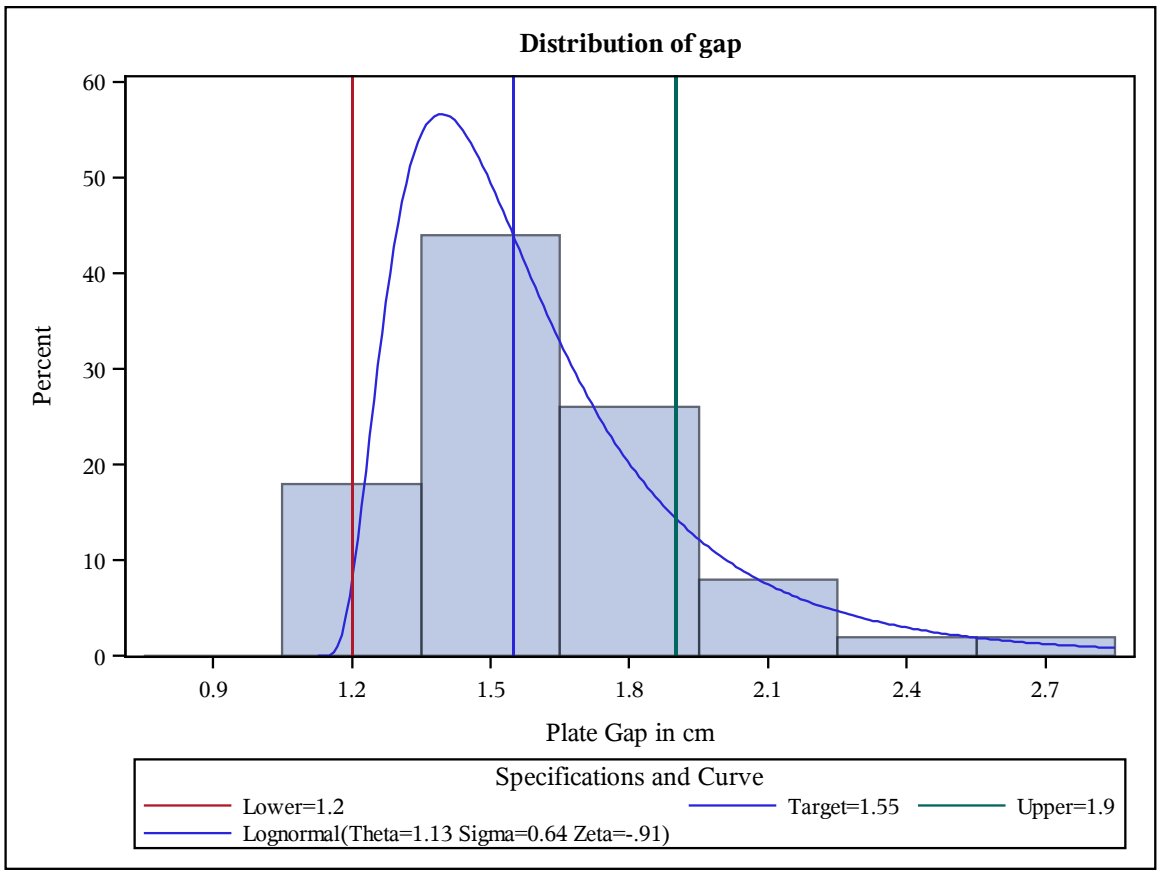
Test	(2B) $\theta = 0$	(2C) $\theta = .5$	(2D) $\theta = 1$	(2E) $\theta = 1.2$	(2F) $\theta = 1.126$
Kolmogorov-Smirnov	0.143	$> .150$	$> .150$	.044	$> .250$
Cramer-von Mises	0.196	.460	$> .500$	.085	$> .500$
Anderson-Darling	0.113	.355	$> .500$	.069	$> .500$

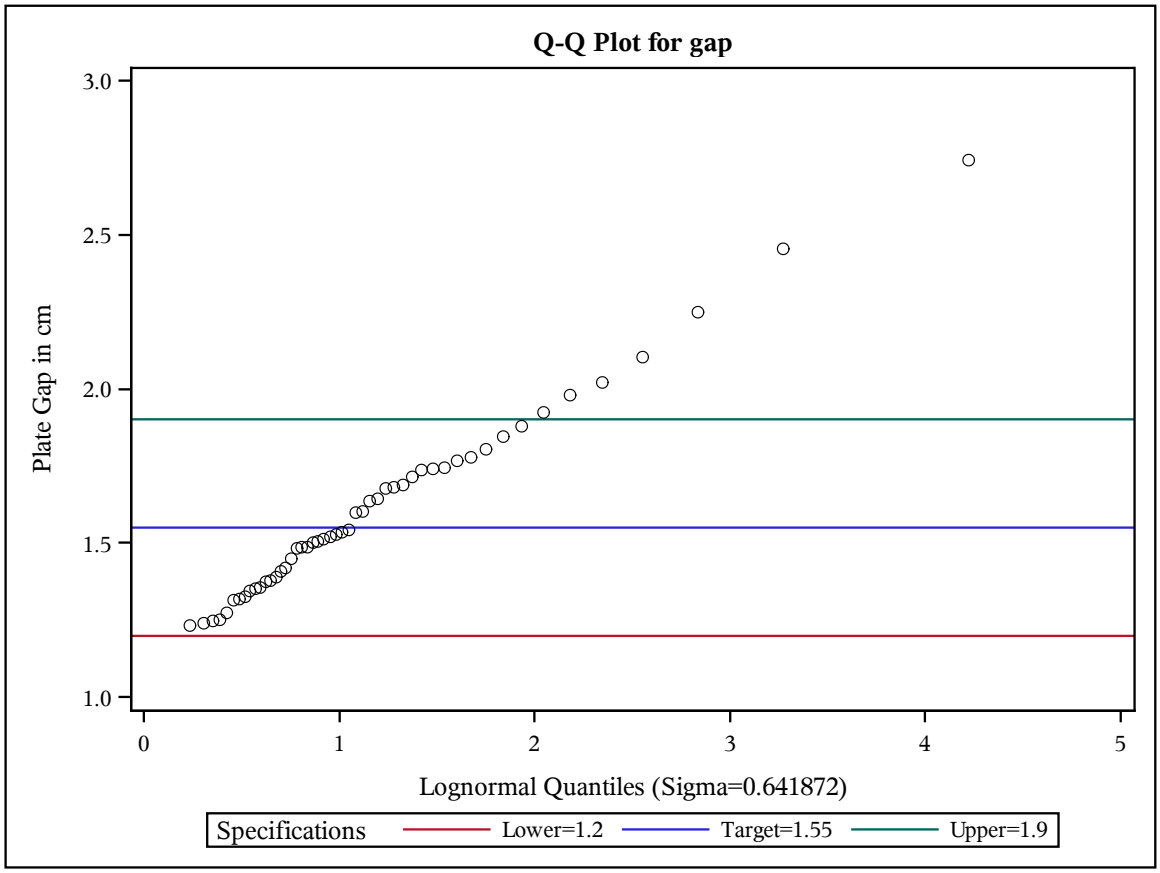
- All of goodness-of-fit tests for the lognormality have relatively large p-values (except for the $\theta = 1.2$ case). However, for $\theta = 1.126$, the p-values are consistently large indicating this distribution is a good choice.

- The histogram and qq plot on the next page are for the SAS-estimated threshold $\theta = 1.126$. Note the good fit in the histgram and the linearity in the qq plot.

- Following output (2A)-(2F) are several pages summarizing the options for fitting beta, exponential, gamma, lognormal, normal, and Weibull distributions in SAS.

Capability Analysis of Plate Gaps -- Threshold Estimated





(2A) Capability Analysis of Plate Gaps

The CAPABILITY Procedure
Variable: gap (Plate Gap in cm)

Moments

N	50	Sum Weights	50
Mean	1.61562	Sum Observations	80.781
Std Deviation	0.31347557	Variance	0.09826693
Skewness	1.46899085	Kurtosis	2.84054912
Uncorrected SS	135.326479	Corrected SS	4.81507978
Coeff Variation	19.4028034	Std Error Mean	0.04433214

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.886977	Pr < W	0.0002
Kolmogorov-Smirnov	D	0.134075	Pr > D	0.0235
Cramer-von Mises	W-Sq	0.194892	Pr > W-Sq	0.0059
Anderson-Darling	A-Sq	1.314974	Pr > A-Sq	<0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	2.7410
99%	2.7410
95%	2.2470
90%	2.0005
75% Q3	1.7460
50% Median	1.5315
25% Q1	1.3780
10%	1.2945
5%	1.2470
1%	1.2310
0% Min	1.2310

Specification Limits

--------Limit--------		------Percent-------	
Lower (LSL)	1.200000	% < LSL	0.00000
Target	1.550000	% Between	86.00000
Upper (USL)	1.900000	% > USL	14.00000

Process Capability Indices

Index	Value	95% Confidence Limits	
Cp	0.372171	0.298661	0.445536
CPL	0.441948	0.313632	0.567803
CPU	0.302395	0.191282	0.411300
Cpk	0.302395	0.192299	0.412490
Cpm	0.364276	0.295955	0.439536

Warning: Normality is rejected for alpha = 0.05 using the Shapiro-Wilk test

(2B) Capability Analysis of Plate Gaps -- Threshold=0

The CAPABILITY Procedure
Fitted Lognormal Distribution for gap (Plate Gap in cm)

Parameters for Lognormal Distribution

Parameter	Symbol	Estimate
Threshold	Theta	0
Scale	Zeta	0.463328
Shape	Sigma	0.178256
Mean		1.614808
Std Dev		0.290152

Goodness-of-Fit Tests for Lognormal Distribution

Test	----Statistic-----		DF	------p Value------	
Kolmogorov-Smirnov	D	0.10880269		Pr > D	0.143
Cramer-von Mises	W-Sq	0.08243290		Pr > W-Sq	0.196
Anderson-Darling	A-Sq	0.60336046		Pr > A-Sq	0.113
Chi-Square	Chi-Sq	4.50372694	3	Pr > Chi-Sq	0.212

Percent Outside Specifications for Lognormal Distribution

Lower Limit		Upper Limit	
LSL	1.200000	USL	1.900000
Obs Pct < LSL	0	Obs Pct > USL	14.000000
Est Pct < LSL	5.746477	Est Pct > USL	15.828997

Capability Indices Based on Lognormal Distribution

Cp	0.392803
CPL	0.591452
CPU	0.276434
Cpk	0.276434
Cpm	0.303964

Quantiles for Lognormal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	1.23100	1.04984
5.0	1.24700	1.18545
10.0	1.29450	1.26476
25.0	1.37800	1.40930
50.0	1.53150	1.58935
75.0	1.74600	1.79241
90.0	2.00050	1.99725
95.0	2.24700	2.13088
99.0	2.74100	2.40612

(2C) Capability Analysis of Plate Gaps -- Threshold=.5

The CAPABILITY Procedure
Fitted Lognormal Distribution for gap (Plate Gap in cm)

Parameters for Lognormal Distribution

Parameter	Symbol	Estimate
Threshold	Theta	0.5
Scale	Zeta	0.075773
Shape	Sigma	0.254839
Mean		1.61432
Std Dev		0.288645

Goodness-of-Fit Tests for Lognormal Distribution

Test	----Statistic-----		DF	------p Value------	
Kolmogorov-Smirnov	D	0.09553564		Pr > D	>0.150
Cramer-von Mises	W-Sq	0.05414127		Pr > W-Sq	0.460
Anderson-Darling	A-Sq	0.40946795		Pr > A-Sq	0.355
Chi-Square	Chi-Sq	2.68244162	3	Pr > Chi-Sq	0.443

Percent Outside Specifications for Lognormal Distribution

Lower Limit		Upper Limit	
LSL	1.200000	USL	1.900000
Obs Pct < LSL	0	Obs Pct > USL	14.000000
Est Pct < LSL	4.485352	Est Pct > USL	15.315432

Capability Indices Based on Lognormal Distribution

Cp	0.385711
CPL	0.656913
CPU	0.259451
Cpk	0.259451
Cpm	0.275875

Quantiles for Lognormal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	1.23100	1.09626
5.0	1.24700	1.20935
10.0	1.29450	1.27817
25.0	1.37800	1.40836
50.0	1.53150	1.57872
75.0	1.74600	1.78102
90.0	2.00050	1.99535
95.0	2.24700	2.14041
99.0	2.74100	2.45154

(2D) Capability Analysis of Plate Gaps -- Threshold=1

The CAPABILITY Procedure
Fitted Lognormal Distribution for gap (Plate Gap in cm)

Parameters for Lognormal Distribution

Parameter	Symbol	Estimate
Threshold	Theta	1
Scale	Zeta	-0.59778
Shape	Sigma	0.475211
Mean		1.615779
Std Dev		0.30995

Goodness-of-Fit Tests for Lognormal Distribution

Test	----Statistic-----		DF	------p Value------	
Kolmogorov-Smirnov	D	0.05389670		Pr > D	>0.150
Cramer-von Mises	W-Sq	0.02211931		Pr > W-Sq	>0.500
Anderson-Darling	A-Sq	0.15891242		Pr > A-Sq	>0.500
Chi-Square	Chi-Sq	0.93296884	3	Pr > Chi-Sq	0.817

Percent Outside Specifications for Lognormal Distribution

Lower Limit		Upper Limit	
LSL	1.200000	USL	1.900000
Obs Pct < LSL	0	Obs Pct > USL	14.000000
Est Pct < LSL	1.663282	Est Pct > USL	15.005189

Capability Indices Based on Lognormal Distribution

Cp	0.324646
CPL	0.837741
CPU	0.201320
Cpk	0.201320
Cpm	0.196952

Quantiles for Lognormal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	1.23100	1.18208
5.0	1.24700	1.25172
10.0	1.29450	1.29916
25.0	1.37800	1.39920
50.0	1.53150	1.55003
75.0	1.74600	1.75786
90.0	2.00050	2.01129
95.0	2.24700	2.20186
99.0	2.74100	2.66151

(2E) Capability Analysis of Plate Gaps -- Threshold=1.2

The CAPABILITY Procedure
Fitted Lognormal Distribution for gap (Plate Gap in cm)

Parameters for Lognormal Distribution

Parameter	Symbol	Estimate
Threshold	Theta	1.2
Scale	Zeta	-1.18899
Shape	Sigma	0.878589
Mean		1.647971
Std Dev		0.483293

Goodness-of-Fit Tests for Lognormal Distribution

Test	----Statistic-----		DF	------p Value------	
Kolmogorov-Smirnov	D	0.12674509		Pr > D	0.044
Cramer-von Mises	W-Sq	0.10939392		Pr > W-Sq	0.085
Anderson-Darling	A-Sq	0.69385261		Pr > A-Sq	0.069
Chi-Square	Chi-Sq	2.62518061	3	Pr > Chi-Sq	0.453

Percent Outside Specifications for Lognormal Distribution

Lower Limit		Upper Limit	
LSL	1.200000	USL	1.900000
Obs Pct < LSL	0	Obs Pct > USL	14.000000
Est Pct < LSL	0	Est Pct > USL	17.173615

Capability Indices Based on Lognormal Distribution

Cp	0.165579
CPL	1.077196
CPU	0.100249
Cpk	0.100249
Cpm	0.086954

Quantiles for Lognormal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	1.23100	1.23944
5.0	1.24700	1.27178
10.0	1.29450	1.29877
25.0	1.37800	1.36837
50.0	1.53150	1.50453
75.0	1.74600	1.75080
90.0	2.00050	2.13892
95.0	2.24700	2.49197
99.0	2.74100	3.55118

(2F) Capability Analysis of Plate Gaps -- Threshold Estimated

The CAPABILITY Procedure
Fitted Lognormal Distribution for gap (Plate Gap in cm)

Parameters for Lognormal Distribution

Parameter	Symbol	Estimate
Threshold	Theta	1.125768
Scale	Zeta	-0.9055
Shape	Sigma	0.641872
Mean		1.622604
Std Dev		0.354754

Goodness-of-Fit Tests for Lognormal Distribution

Test	Statistic		DF	p Value	
Kolmogorov-Smirnov	D	0.08348421		Pr > D	>0.250
Cramer-von Mises	W-Sq	0.03743301		Pr > W-Sq	>0.500
Anderson-Darling	A-Sq	0.23546790		Pr > A-Sq	>0.500
Chi-Square	Chi-Sq	1.14271615	2	Pr > Chi-Sq	0.565

Percent Outside Specifications for Lognormal Distribution

Lower Limit		Upper Limit	
LSL	1.200000	USL	1.900000
Obs Pct < LSL	0	Obs Pct > USL	14.000000
Est Pct < LSL	0.413543	Est Pct > USL	15.575504

Capability Indices Based on Lognormal Distribution

Cp	0.257866
CPL	0.955746
CPU	0.156126
Cpk	0.156126
Cpm	0.144730

Quantiles for Lognormal Distribution

Percent	Quantile	
	Observed	Estimated
1.0	1.23100	1.21660
5.0	1.24700	1.26645
10.0	1.29450	1.30339
25.0	1.37800	1.38802
50.0	1.53150	1.53011
75.0	1.74600	1.74917
90.0	2.00050	2.04621
95.0	2.24700	2.28794
99.0	2.74100	2.92565

Distribution-Options

In addition, each of the *distribution-options* has additional options that are listed in parentheses after the *distribution-option*. These additional options allow you to control the fitted curve and enhance the plot.

The *distribution-options* are as follows:

BETA<(*beta-options*)>
superimposes a beta density curve on the histogram. The equation of the fitted density curve is

$$p(x) = \frac{100h\%\ (x-\theta)^{\alpha-1}(\sigma+\theta-x)^{\beta-1}}{B(\alpha,\beta)\sigma^{(\alpha+\beta-1)}} \qquad \text{where}$$

θ=lower threshold parameter (lower endpoint parameter)
σ=scale parameter ($\sigma>0$)
α=shape parameter ($\alpha>0$)
β=shape parameter ($\beta>0$)
h=width of a histogram interval.
$B(\alpha,\beta)=\Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

The notation of the density equation is chosen to be consistent with notation used in the rest of this chapter. Many texts, including Johnson and Kotz (1970), write the beta density function as

$$\frac{(x-a)^{p-1}(b-x)^{q-1}}{B(p,q)(b-a)^{p+q-1}}$$

where the two notations are related as follows:

$b-a=\sigma$ $\qquad$ $p=\alpha$
$a=\theta$ $\qquad$ $q=\beta$

The beta distribution involves two threshold parameters, or endpoint parameters: a lower threshold parameter $\theta=a$ and an upper threshold parameter $\theta+\sigma=b$. If you specify a fitted beta curve, the data must lie between these two endpoints, which you can provide by specifying the THRESHOLD= and SCALE= *beta-options*. By default, SCALE=1 and THRESHOLD=0. In addition, you can specify the shape parameters α and β with the ALPHA= and BETA= *beta-options*, respectively. By default, the procedure calculates maximum-likelihood estimates for α and β. To fit a beta density curve to a set of data bounded below by 32 and bounded above by 212, and use the default maximum likelihood estimates for α and β, use the following statement:

```
histogram / beta(theta=32 sigma=180);
```

The beta distributions are also referred to as Pearson Type I or II distributions. These include the *power-function* distribution ($\beta=1$), the *arc-sine* distribution ($\alpha=\beta=1/2$), and the *generalized arc-sine* distributions ($\alpha+\beta=1$, $\beta\neq 1/2$). Properties of beta distributions are discussed in Johnson and Kotz (1970).

BETA can appear only once in a HISTOGRAM statement. The following *beta-options* can be used with BETA. Enclose the *beta-options* in parentheses immediately after the word BETA.

ALPHA=α
A=α
specifies the first shape parameter α for the fitted curve. By default, the procedure calculates a maximum-likelihood estimate for α.

BETA=β
B=β
specifies the second shape parameter β for the fitted curve. By default, the procedure calculates a maximum-likelihood estimate for β.

MIDPERCENTS
prints a table of the midpoints of the histogram intervals, the percent of the population in each interval, and the estimated percent of the population in each interval (estimated from the fitted beta distribution).

NOPRINT
suppresses printed statistics and quantiles for the fitted beta curve.

PERCENTS=*percents*
PERCENT=*percents*
lists percents for which quantiles calculated from the data and quantiles estimated from the fitted curve are tabulated. The *percents* must be between 0 and 100. The default *percents* are 1, 5, 10, 25, 50, 75, 90, 95, and 99.

SCALE=σ
SIGMA=σ
specifies the scale parameter σ for the fitted curve. By default, SCALE=1. For the beta distribution, the upper threshold is $\theta+\sigma$. In the default case, the maximum data value must be less than 1, and in general the maximum must be less than $\theta+\sigma$.

SYMBOL='*character*'
specifies the *character* used to plot the beta density curve if the histogram is produced on a line printer. By default, SYMBOL='B'.

SYMBOL= is ignored when you specify the GRAPHICS option in the PROC CAPABILITY statement. To control the color and style of the beta curve when the chart is produced on a graphics device, see the section on the SYMBOL statement later in this chapter.

THRESHOLD=θ
THETA=θ
specifies the lower threshold parameter θ for the fitted curve. By default, THRESHOLD=0. θ must be less than the minimum data value.

EXPONENTIAL<(*exponential-options*)>
EXP<(*exponential-options*)>
superimposes an exponential density curve on the histogram. The equation of the fitted density curve is

$$p(x) = ((100h\%) / \sigma)\exp(-(x - \theta) / \sigma)$$

if $x \geq \theta$ and 0 otherwise, where

θ=threshold parameter
σ=scale parameter (σ>0)
h=width of a histogram interval.

The minimum data value must be greater than or equal to the threshold parameter θ, which you can specify with the THRESHOLD= *exponential-option*. By default, θ=0. In addition, you can specify σ with the SCALE= *exponential-option*. By default, σ is estimated. Note that some authors define the scale parameter as 1/σ. Properties of exponential distributions are discussed in Johnson and Kotz (1970).

EXPONENTIAL can appear only once in a HISTOGRAM statement. The following *exponential-options* can be used. Enclose the *exponential-options* in parentheses immediately after the word EXPONENTIAL.

MIDPERCENTS
prints a table of the midpoints of the histogram intervals, the percent of the population in each interval, and the estimated percent of the population in each interval (estimated from the fitted exponential distribution).

NOPRINT
suppresses printed statistics and quantiles for the fitted exponential curve.

PERCENTS=*percents*
PERCENT=*percents*
lists percents for which quantiles calculated from the data and quantiles estimated from the fitted curve are tabulated. The *percents* must be between 0 and 100. The default *percents* are 1, 5, 10, 25, 50, 75, 90, 95, and 99.

SCALE=σ
SIGMA=σ
specifies the scale parameter σ for the fitted curve. By default, the procedure calculates a maximum-likelihood estimate for σ.

SYMBOL='*character*'
specifies the *character* used to plot the exponential density if the histogram is produced on a line printer. By default, SYMBOL='E'.

SYMBOL= is ignored when you specify the GRAPHICS option in the PROC CAPABILITY statement. See the section on the SYMBOL statement for details on controlling the color and style of the curve when the plot is produced on a graphics device.

THRESHOLD=θ
THETA=θ
specifies the threshold parameter θ for the fitted curve. By default, THRESHOLD=0. θ must be less than or equal to the minimum data value.

GAMMA<(*gamma-options*)>
superimposes a gamma density curve on the histogram. The equation of the fitted density curve is

$$p(x) = ((100h\%) / (\Gamma(\alpha)\sigma)))((x - \theta)/\sigma)^{\alpha-1}\exp(-(x - \theta)/\sigma)$$

if $x > \theta$ and 0 otherwise, where

θ=threshold parameter
σ=scale parameter (σ>0)
α=shape parameter (α>0)
h=width of a histogram interval.

The minimum data value must be greater than the threshold parameter θ, which you can specify with the THRESHOLD= *gamma-option*. By default, θ=0. In addition, you can specify σ and α with the SCALE= and ALPHA= *gamma-options*. By default, the procedure calculates maximum-likelihood estimates for these parameters.

The gamma distributions are also referred to as Pearson Type III distributions, and they include the chi-square distributions. For a chi-square distribution, the equation of the fitted curve is

$$p(x) = ((100h\%) / 2\Gamma(\nu/2))\ [(u/2)^{(\nu/2)-1} \exp(-u/2)]$$

where $u=x^2$. Notice that this is a Gamma distribution with $\alpha=\nu/2$, $\sigma=2$, and $\theta=0$.

Properties of gamma distributions are discussed in Johnson and Kotz (1970).

GAMMA can appear only once in a HISTOGRAM statement. The following *gamma-options* can be used. Enclose *gamma-options* in parentheses immediately after the word GAMMA.

* MAXITER=*n*
specifies the maximum number of iterations in the Newton-Raphson approximation of the maximum-likelihood estimate of α. By default, MAXITER=20.

* MIDPERCENTS
prints a table of the midpoints of the histogram intervals, the percent of observations in each interval, and the estimated percent of the population in each interval (estimated from the fitted gamma distribution).

NOPRINT
suppresses printed statistics and quantiles for the fitted gamma curve.

PERCENTS=*percents*
PERCENT=*percents*
lists percents for which quantiles calculated from the data and quantiles estimated from the fitted curve are tabulated. The *percents* must be between 0 and 100. The default *percents* are 1, 5, 10, 25, 50, 75, 90, 95, and 99.

* SCALE=σ
SIGMA=σ
specifies the scale parameter σ for the fitted curve. By default, a maximum-likelihood estimate is calculated for σ.

* SHAPE=α
ALPHA=α
specifies the shape parameter α for the fitted curve. By default, a maximum-likelihood estimate is calculated for α.

SYMBOL='*character*'
specifies the character used to plot the gamma density if the histogram is produced on a line printer. By default, SYMBOL='G'.

SYMBOL= is ignored when you specify the GRAPHICS option in the PROC CAPABILITY statement. See the section on the SYMBOL statement for details on controlling the color and style of the gamma density when the chart is produced on a graphics device.

* THRESHOLD=θ
THETA=θ
specifies the threshold parameter θ for the fitted curve. By default, THRESHOLD=0. θ must be less than the minimum data value.

LOGNORMAL<(*lognormal-options*)>
LNORM<(*lognormal-options*)>
superimposes a lognormal density curve on the histogram. The equation of the fitted density curve is

$$p(x) = ((100h\%) / \sigma\sqrt{2\pi})\ \exp(-(\log(x-\theta)-\zeta)^2 / 2\sigma^2) / (x-\theta)$$

if $x>\theta$ and 0 otherwise, where

θ=threshold parameter
ζ=scale parameter
σ=shape parameter ($\sigma>0$)
h=width of a histogram interval.

The minimum data value must be greater than the threshold parameter θ, which you can specify with the THRESHOLD= *lognormal-option*. By default, $\theta=0$. You can specify the ζ and σ parameters with the SCALE= and SHAPE= *lognormal-options*, respectively. By default, the procedure calculates maximum-likelihood estimates for these parameters.

Note that σ denotes the shape parameter of the lognormal distribution, whereas σ denotes the scale parameter of the beta, exponential, gamma, normal, and Weibull distributions (discussed elsewhere in this section). The use of σ to denote the lognormal shape parameter is a convention that follows from the fact that $(\log(X-\theta)-\zeta)/\sigma$ has a standard normal distribution if X is lognormally distributed. Also note that the lognormal scale parameter ζ can be negative.

Properties of lognormal distributions are discussed in Johnson and Kotz (1970).

LOGNORMAL can appear only once in a HISTOGRAM statement. The following *lognormal-options* can be used. Enclose the *lognormal-options* in parentheses immediately after the word LOGNORMAL.

MIDPERCENTS
prints a table of the midpoints of the histogram intervals, the percent of observations in each interval, and the estimated percent of the population in each interval (estimated from the fitted lognormal distribution).

NOPRINT
suppresses printed statistics and quantiles for the fitted lognormal curve.

PERCENTS=*percents*
PERCENT=*percents*
lists percents for which quantiles calculated from the data and quantiles estimated from the fitted curve are tabulated. The *percents* must be between zero and 100. The default *percents* are 1, 5, 10, 25, 50, 75, 90, 95, and 99.

SCALE=ζ
ZETA=ζ
specifies the scale parameter ζ for the fitted curve. By default, a maximum-likelihood estimate is calculated for ζ.

SHAPE=σ
SIGMA=σ
specifies the shape parameter σ for the fitted curve. By default, a maximum-likelihood estimate is calculated for σ.

SYMBOL='*character*'
specifies the *character* used to plot the lognormal curve if the histogram is produced on a line printer. By default, SYMBOL='L'.

SYMBOL= is ignored when you specify the GRAPHICS option in the PROC CAPABILITY statement. To control the color and style of the lognormal density when the plot is produced on a graphics device, see the section on the SYMBOL statement later in this chapter.

THRESHOLD=θ
THETA=θ
specifies the threshold parameter θ for the fitted curve. By default, THRESHOLD=0. θ must be less than the minimum data value.

NORMAL<(*normal-options*)>
NORM<(*normal-options*)>
superimposes a normal density curve on the histogram. The equation of the fitted density curve is

$$p(x) = ((100h\%) / \sigma\sqrt{2\pi}) \exp(-((x - \mu) / \sigma)^2 / 2)$$

where

μ=mean
σ=standard deviation (σ>0)
h=width of a histogram interval.

By default, the procedure calculates maximum-likelihood estimates for μ and σ. You can specify these parameters with the MU= and SIGMA= *normal-options*, respectively. Properties of normal distributions are discussed in Johnson and Kotz (1970).

NORMAL can appear only once in a HISTOGRAM statement. The following *normal-options* can be used. Enclose *normal-options* in parentheses immediately after the word NORMAL.

MIDPERCENTS
prints a table of the midpoints of the histogram intervals, the percent of observations in each interval, and the estimated percent of the population in each interval (estimated from the fitted normal distribution).

MU=μ
specifies the parameter μ for the fitted curve. The default value of μ is the sample mean.

NOPRINT
suppresses printed statistics and quantiles for the fitted normal curve.

PERCENTS=*percents*
PERCENT=*percents*
lists percents for which quantiles calculated from the data and quantiles estimated from the fitted curve are tabulated. The *percents* must be between zero and 100. The default *percents* are 1, 5, 10, 25, 50, 75, 90, 95, and 99.

SIGMA=σ
specifies the parameter σ for the fitted curve. The default value of σ is the sample standard deviation.

SYMBOL='*character*'
specifies the character used to plot the normal density if the histogram is produced on a line printer. By default, SYMBOL='N'.

SYMBOL= is ignored when you specify the GRAPHICS option in the PROC CAPABILITY statement. To control the color and style of the normal density when the plot is produced on a graphics device, see the section on the SYMBOL statement later in this chapter.

WEIBULL<(*Weibull-options*)>
WEIB<(*Weibull-options*)>
superimposes a Weibull density curve on the histogram. The equation of the fitted density curve is

$$p(x) = ((100ch\%) / \sigma)((x - \theta) / \sigma)^{c-1} \exp(-((x - \theta) / \sigma)^c)$$

if $x > \theta$ and 0 otherwise, where

θ=threshold parameter
σ=scale parameter ($\sigma > 0$)
c=shape parameter ($c > 0$)
h=width of a histogram interval.

The minimum data value must be greater than the threshold parameter θ, which you can specify with the THRESHOLD= *Weibull-option*. By default, $\theta = 0$. You can specify σ and c with the SCALE= and SHAPE= *Weibull-options*, respectively. By default, the procedure calculates maximum-likelihood estimates for these parameters. Properties of Weibull distributions are discussed in Johnson and Kotz (1970).

WEIBULL can appear only once in a HISTOGRAM statement. The following *Weibull-options* can be used. Enclose *Weibull-options* in parentheses immediately after the word WEIBULL.

C=*c*
See SHAPE= later in this list.

CDELTA=*value*
specifies the change in c for which to stop iterations in the Newton-Raphson approximation of the maximum-likelihood estimate of c. Iteration continues until the change in *c* between consecutive steps is less than the value specified or the number of iterations exceeds the value of MAXITER= (see below). By default, CDELTA=0.00001.

CINITIAL=*value*
specifies the initial value for *c* in the Newton-Raphson approximation of the maximum likelihood equation for *c* (Johnson and Kotz 1970, 255). By default, CINITIAL=1.8.

MAXITER=*value*
specifies the maximum number of iterations in the Newton-Raphson approximation of the maximum-likelihood estimate of *c*. By default, MAXITER=20.

MIDPERCENTS
prints a table of the midpoints of the histogram intervals, the percent of observations in each interval, and the estimated percent of the population in each interval (estimated from the fitted Weibull distribution).

NOPRINT
suppresses printed statistics and quantiles for the fitted Weibull curve.

PERCENTS=*percents*
PERCENT=*percents*
lists percents for which quantiles calculated from the data and quantiles estimated from the fitted curve are tabulated. The *percents* must be between 0 and 100. The default *percents* are 1, 5, 10, 25, 50, 75, 90, 95, and 99.

SCALE=σ
SIGMA=σ
specifies the scale parameter σ for the fitted curve. By default, a maximum-likelihood estimate is calculated for σ.

SHAPE=*c*
C=*c*
specifies the shape parameter *c* for the fitted curve. By default, the procedure calculates a maximum-likelihood estimate for *c*.

SYMBOL='*character*'
specifies the character used to plot the Weibull density if the histogram is produced on a line printer. By default, SYMBOL='W'.

SYMBOL= is ignored when you specify the GRAPHICS option in the PROC CAPABILITY statement. To control the color and style of the Weibull density when the plot is produced on a graphics device, see the section on the SYMBOL statement.

THRESHOLD= θ
THETA=θ
specifies the threshold parameter θ for the fitted curve. By default, THRESHOLD=0. The value of θ must be less than the minimum data value.