

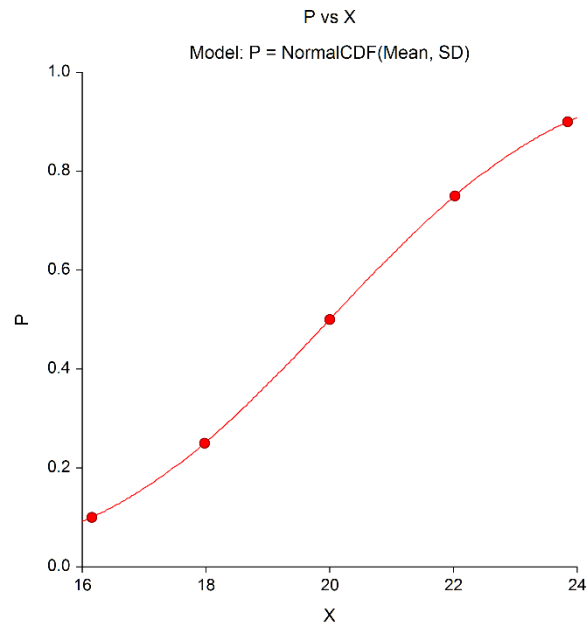
## Chapter 353

# Curve Fitting – CDF

## Introduction

This procedure estimates the parameters of several probability distributions using paired data that consist of an  $X$  variable's value and its corresponding cumulative distribution function ( $CDF$ ) value, labelled  $P$ .  $CDF(x)$  is the probability that  $X$  is less than a particular value of  $X$  which is labelled  $x$ . The probability distribution functions that are available in this procedure include Beta, Gamma, Logistic, Normal, and Weibull.

Curve fitting refers to estimating the distribution's parameters using nonlinear regression. An introduction to curve fitting and nonlinear regression is provided in the chapter entitled *Introduction to Curve Fitting*. Below is an example of fitting a CDF that can be accomplished with this procedure.



One application that could use this procedure to supply appropriate parameter estimates is the sample size planning technique (Bayesian) Assurance. This technique, which is available in our companion product **PASS**, calculates the Expected Power of a given design model using data from an *elicitation step*. The elicitation step provides an expert's educated guess of certain quantiles of each parameter's distribution. These quantiles are then smoothed by finding a good-fitting probability distribution.

## Available CDF Models

The following probability distribution models are estimated by this procedure.

### 1. Beta Distribution: $P = \text{BetaCDF}(A, B)$

This model estimates the parameters of the CDF of the Beta distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

#### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$0 \leq X \leq 1$	
A	Shape	$A > 0$	$A = B$ implies symmetrical
B	Shape	$B > 0$	

### 2. CDF of the Gamma Distribution: $P = \text{GammaCDF}(A, B)$

This model estimates the parameters of the CDF of the Gamma distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

#### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$X > 0$	
A	Shape	$A > 0$	If $A = 1$ , $X$ is the Exponential with rate $1/B$
B	Scale	$B > 0$	

### 3. CDF of the Logistic Distribution: $P = \text{LogisticCDF}(A, B)$

This model estimates the parameters of the CDF of the Logistic distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

#### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$-\infty < X < \infty$	
A	Location	$-\infty < B < \infty$	
B	Scale	$B > 0$	$B = q\sigma$ , where $q = \sqrt{3}/\pi$

---

#### 4. CDF of the Lognormal Distribution: $P = \text{LognormalCDF}(A, B)$

This model estimates the parameters of the CDF of the Lognormal distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

If  $X$  is lognormally distributed, then  $Z = \log(X)$  is normally distributed.

##### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$-\infty < X < \infty$	
A	Mean	$-\infty < A < \infty$	A is the mean of $\log(X)$
B	Standard Deviation	$B > 0$	B is the SD of $\log(X)$ .

---

#### 5. CDF of the Normal Distribution: $P = \text{NormalCDF}(A, B)$

This model estimates the parameters of the CDF of the Normal distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

##### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$-\infty < X < \infty$	
A	Mean	$-\infty < A < \infty$	
B	Standard Deviation	$B > 0$	

---

#### 6. CDF of the Student's T Distribution: $P = \text{StudentsTCDF}(A, B, C)$

This model estimates the parameters of the CDF of Student's  $t$  distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

##### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$-\infty < X < \infty$	
A	Mean	$-\infty < A < \infty$	
B	Standard Deviation	$B > 0$	
C	Degrees of Freedom	$C > 0$	Often an integer

---

## 7. CDF of the Triangle Distribution: $P = \text{TriangleCDF}(A, B, C)$

This model estimates the parameters of the CDF of the Triangle distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$B \leq X \leq C$	
A	Mode	$B \leq A \leq C$	
B	Minimum	$B < C$	
C	Maximum	$C > B$	

---

## 8. CDF of the Uniform Distribution: $P = \text{UniformCDF}(A, B)$

This model estimates the parameters of the CDF of the Triangle distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$A \leq X \leq B$	
A	Minimum	$A < B$	
B	Maximum	$B > A$	

---

## 9. CDF of the Weibull Distribution: $P = \text{WeibullCDF}(A, B)$

This model estimates the parameters of the CDF of the Normal distribution from a set of data consisting of two columns. One column gives the value of a random variable  $X$ . The other column gives the corresponding probability that  $X$  will take a value less than or equal to  $X$ .

### Arguments

Argument	Type	Range	Comment
P	Dependent Variable	$0 \leq P \leq 1$	CDF Value
X	Random Variable	$-\infty < X < \infty$	
A	Scale	$A > 0$	
B	Shape	$B > 0$	$B = 1$ results in the Exponential distribution

---

## Confidence Intervals

Two methods are used to calculate confidence intervals of the regression parameters and predicted values. The first method is based on the usual normality and constant variance of residuals assumptions. When the data follow these assumptions, standard expressions for the confidence intervals are used based on the Student's  $t$  distribution. Unfortunately, nonlinear regression datasets rarely follow these assumptions.

The second method is called the *bootstrap* method. This is a modern, computer-intensive method that has only become available in recent years as extensive computer power has become available. Unfortunately, this method requires a moderate sample size.

---

### Bootstrap Confidence Intervals

*Bootstrapping* provides standard errors and confidence intervals for nonlinear-regression parameter, predicted means, and predicted values. The method is simple in concept, but it requires extensive computation time.

Bootstrap confidence intervals are based on the assumption that your sample is actually representative of the population. Beginning with this assumption,  $B$  samples are drawn ( $B$  is over 1000) of size  $N$  from your original sample with replacement. With replacement sampling means that each observation may be selected more than once. For each bootstrap sample, the nonlinear-regression results are computed and stored.

Suppose you want the standard error and a confidence interval of a regression parameter. The bootstrap sampling process provides  $B$  estimates of this parameter. The standard deviation of these  $B$  estimates is the bootstrap estimate of the standard error of the parameter. The bootstrap confidence interval is found by arranging the  $B$  values in sorted order and selecting the appropriate percentiles from the list. For example, a 90% bootstrap confidence interval for the parameter is given by fifth and ninety-fifth percentiles of the bootstrap parameter values.

The main assumption made when using the bootstrap is that your sample approximates the population. Because of this assumption, bootstrapping does not work well for small samples in which there is little likelihood that the sample is representative of the population. Bootstrapping should only be used in medium to large samples.

---

### Bootstrap Prediction Intervals

Bootstrap confidence intervals for the mean of  $Y$  given  $X$  are generated from the bootstrap sample in the usual way. To calculate prediction intervals for the predicted value (not the mean) of  $Y$  given  $X$  requires a modification to the predicted value of  $Y$  to be made to account for the variation of  $Y$  about its mean. This modification of the predicted  $Y$  values in the bootstrap sample, suggested by Davison and Hinkley, is as follows.

$$\hat{y}_i = \hat{y}_i + e_r^*$$

where  $e_r^*$  is a randomly selected modified residual (see below). By adding the residual, we have added an appropriate amount of variation to represent the variance of individual  $Y$ 's about their mean value.

## Modified Residuals

Davison and Hinkley (1999) page 279 recommend the use of a special rescaling of the residuals when bootstrapping to keep results unbiased. Because of the high amount of computing involved in bootstrapping, these modified residuals are calculated using

$$e_j^* = \frac{e_j}{\sqrt{1 - \frac{1}{N}}} - \bar{e}$$

where

$$\bar{e} = \frac{\sum_{j=1}^N e_j}{N}$$

Note that there is a different rescaling than Davison and Hinkley recommended. We have used this rescaling because it is much quicker to calculate.

## Hypothesis Testing

When curves are fit to two or more groups, it is often of interest to test whether certain regression parameters are equal and whether the fitted curves coincide. Although some approximate results have been obtained using indicator variables, these are asymptotic results, and little is known about their appropriateness in small samples. We provide a test of the hypothesis that all group curves coincide using an  $F$ -test that compares the residual sum of squares obtained when the grouping is

ignored with the total of the residual sum of squares obtained for each group. This test is routinely used in the analysis of variance associated with linear models and its application to nonlinear models has occasionally been suggested. However, it is based on naive assumptions that seldom occur.

Because of the availability of fast computing speed in recent years, a second method of hypothesis testing, called the *randomization test*, is now available. This test will be discussed next.

## Randomization Test

Randomization testing is discussed by Edgington (1987). The details of the randomization test are simple: all possible permutations of the group variable while leaving the dependent and independent variables in their original order are investigated. For each permutation, the difference between the estimated group parameters is calculated. The number of permutations with a magnitude greater than or equal to that of the actual sample is counted. Dividing this count by the number of permutations gives the significance level of the test.

The randomization test is suggested because an exact test is achieved without making unrealistic assumptions about the data such as constant variance, normality, or model accuracy. The test was not used in the past because the number of computations was prohibitive. In fact, the randomization test was originally proposed by Fisher and he chose his  $F$ -test because its distribution close approximated the randomization distribution.

The only assumption that a randomization test makes is that the data values are *exchangeable* under the null hypothesis.

For even moderate sample sizes, the total number of permutations is in the trillions, so a Monte Carlo approach is used in which the permutations are found by random selection rather than enumeration. Using

## Curve Fitting – CDF

this approach, a reasonable approximation to the test's probability level may be found by considering only a few thousand permutations rather than the trillions needed for complete enumeration. Edgington suggests that at least 1000 permutations be computed. We suggest that this be increased to 10000 for important results.

The program tests two types of hypotheses using randomization tests. The first is that each of the estimated model parameters is equal. The second is that the individual fitted curves coincide across all groups.

### Randomization Statistics for Testing Parameter Equivalence

The test statistic for comparing a model parameter is formed by summing the difference between the group parameter estimates for each pair of groups. If there are  $G$  groups, the test statistic is computed using the formula

$$B_{RT} = \sum_{i=1}^{G-1} \sum_{j=i+1}^G |\hat{\beta}_i - \hat{\beta}_j|$$

### Randomization Statistics for Testing Curve Equivalence

The test statistic for comparing the whole curve is formed by summing the difference between the estimated predicted values for each pair of groups at several points along the curve. If there are  $G$  groups and  $K$  equally spaced test points, the test statistic is computed using the formula

$$C_{RT} = \sum_{k=1}^K \sum_{i=1}^{G-1} \sum_{j=i+1}^G |\hat{y}_{ki} - \hat{y}_{kj}|$$

---

## Data Structure

The data are entered in two variables: one cdf variable and one independent variable. Additionally, you may specify a frequency variable containing the observation count for each row and a group variable that is used to partition the data into independent groups.

---

## Missing Values

Rows with missing values in the variables being analyzed are ignored in the calculations. When only the value of the dependent variable is missing, predicted values are generated.

## Example 1 – Fitting a Normal Distribution

This section presents an example of how to fit a Normal cdf to the data in the table below. These data were generated using the Normal distribution on **NCSS's** Probability Calculator with the mean set to 20 and the standard deviation set to 3. This will allow us to determine how well the results conform to the original values.

P	X
0.10	16.155
0.25	17.977
0.50	20.000
0.75	22.023
0.90	23.845

These data are recorded in the *Normal* dataset. In this example, the dependent variable is P, and the independent variable is X.

### Setup

To run this example, complete the following steps:

#### 1 Open the Normal example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Normal** and click **OK**.

#### 2 Specify the Curve Fitting – CDF procedure options

- Find and open the **Curve Fitting – CDF** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

---

P (Prob X < x) Variable ..... **P**  
 X Variable ..... **X**  
 CDF Model..... **P = NormalCDF(Mean, SD)**  
 Parameter 1 Name ..... **Mean**  
     Min Start Max..... **Automatic**  
 Parameter 2 Name ..... **SD**  
     Min Start Max..... **Automatic**  
 Bootstrap Confidence Intervals..... **Unchecked**  
 Randomization Hypothesis Tests ..... **Unchecked**



Reports Tab

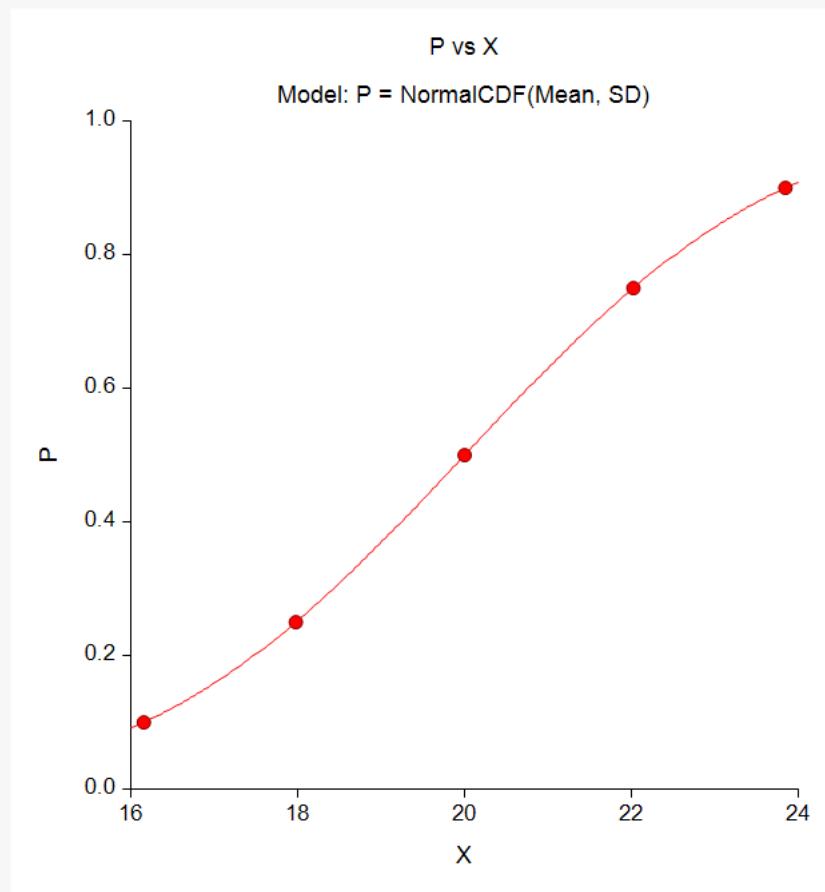
Combined Summary Report .....	<b>Unchecked</b>
Group Difference F-Test .....	<b>Unchecked</b>
Iteration Summary Report.....	<b>Unchecked</b>
Iteration Detail Report.....	<b>Checked</b>
Model Estimation Report .....	<b>Checked</b>
Analysis of Variance Report .....	<b>Checked</b>
Parameter Correlations Report.....	<b>Checked</b>
Predicted Values Report.....	<b>Checked</b>
Residual Report.....	<b>Checked</b>
Predict P at these X Values .....	<b>10 20 30 40 50</b>

**3 Run the procedure**

- Click the **Run** button to perform the calculations and generate the output.

**Function Plot**

Function Plot



This plot displays the data and fitted curve, allowing you to quickly assess the quality of the results. In this example the data fits perfectly.

## Iteration Detail

### Iteration Detail

Iteration Number	Residual Sum of Squares	Mean	SD
1	0.0005748274	20.00001	2.84003
2	2.992327E-07	20.00000	2.99611
3	4.984121E-09	20.00000	2.99981
4	4.983957E-09	20.00000	2.99981
5	4.983957E-09	20.00000	2.99981
6	4.983957E-09	20.00000	2.99981

This report displays the progress of the search algorithm as it iterates toward a solution. It allows you to assess whether the algorithm had indeed converged or whether the program should be re-run with the Maximum Iterations increased or the search parameters changed.

## Model Estimation

### Model Estimation

Parameter Name	Parameter Estimate	Asymptotic Standard Error	95% Confidence Interval Limits for the Parameter Estimate*	
			Lower	Upper
Mean	19.999999974772	0.00019	19.99940	20.00060
SD	2.99981053180615	0.00028	2.99892	3.00070

### Estimation Summary

Iterations	6	Rows Read	5
R-Squared	1.000000	Rows Used	5
Random Seed	5279034	Total Count	5

\* These confidence intervals assume that the observations are independent. This is often not true for the observations fitted by this procedure. Also, the true sample size may not be properly represented.

This report displays the details of the estimation of the model parameters.

### Parameter Name

The name of the parameter whose results are shown on this line.

### Parameter Estimate

The estimated value of this parameter. Note that it is always provided in double precision.

### 95% Confidence Interval Limits for the Parameter Estimate (Lower and Upper)

The lower and upper values of a 95% confidence interval for this parameter. This is a large sample (at least 25 observations for each parameter) confidence limit.

## Asymptotic Standard Error

An estimate of the standard error of the parameter based on asymptotic (large sample) results.

## Iterations

The number of iterations that were completed before the nonlinear algorithm terminated. If the number of iterations is equal to the Maximum Iterations that you set, the algorithm did not converge, but was aborted.

## R-Squared

There is no direct R-squared defined for nonlinear regression. This is a pseudo R-squared constructed to approximate the usual R-squared value used in multiple regression. We use the following generalization of the usual R-squared formula:

$$R\text{-Squared} = (ModelSS - MeanSS)/(TotalSS - MeanSS)$$

where *MeanSS* is the sum of squares due to the mean, *ModelSS* is the sum of squares due to the model, and *TotalSS* is the total (uncorrected) sum of squares of P (the dependent variable).

This version of R-squared tells you how well the model performs after removing the influence of the mean of P. Since many nonlinear models do not explicitly include a parameter for the mean of P, this R-squared may be negative (in which case we set it to zero) or difficult to interpret. However, if you think of it as a direct extension of the R-squared that you use in multiple regression, it will serve well for comparative purposes.

## Random Seed

This is the value of the random seed that was used when running the bootstrap confidence intervals and randomization tests. If you want to duplicate your results exactly, enter this random seed into the Random Seed box under the Simulation tab.

## Rows Read

This is the number of dataset rows considered after using any filters that are set.

## Rows Used

This is the number of dataset rows actually used during the estimation process.

## Analysis of Variance Table

**Analysis of Variance Table**

Source	DF	Sum of Squares	Mean Square
Mean	1	1.25	1.25
Model	2	1.695	0.8475
Model (Adjusted)	1	0.445	0.445
Error	3	4.983957E-09	1.661319E-09
Total (Adjusted)	4	0.445	
Total	5	1.695	

This report displays goodness of fit results in the common ANOVA table format.

### Source

The labels of the various sources of variation.

### DF

The degrees of freedom.

### Sum of Squares

The sum of squares associated with this term. Note that these sums of squares are based on Y, the dependent variable. Individual terms are defined as follows:

<b>Mean</b>	The sum of squares associated with the mean of Y. This may or may not be a part of the model. It is presented since it is the amount used to adjust the other sums of squares.
<b>Model</b>	The sum of squares associated with the model.
<b>Model (Adjusted)</b>	The model sum of squares minus the mean sum of squares.
<b>Error</b>	The sum of the squared residuals. This is often called the sum of squares error or just "SSE."
<b>Total</b>	The sum of the squared Y values.
<b>Total (Adjusted)</b>	The sum of the squared Y values minus the mean sum of squares.

### Mean Square

The sum of squares divided by the degrees of freedom. The Mean Square for Error is an estimate of the underlying variation in the data.

## Asymptotic Correlation Matrix of Parameters

### Asymptotic Correlation Matrix of Parameters

	Mean	SD
Mean	1.000000	-0.007778
SD	-0.007778	1.000000

This report displays the asymptotic correlations of the parameter estimates. When these correlations are high (absolute value greater than 0.98), the precision of the parameter estimates is suspect.

## Predicted Values for Specified X Values

### Predicted Values for Specified X Values

X	Predicted Value of P	95% Prediction Interval Limits	
		Lower	Upper
10.00000	0.00043	0.00030	0.00056
20.00000	0.50000	0.49985	0.50015
30.00000	0.99957	0.99944	0.99970
40.00000	1.00000	0.99987	1.00013
50.00000	1.00000	0.99987	1.00013

This section shows the predicted mean values and asymptotic (large sample) prediction intervals for the X values that were specified. Note that these are prediction limits for a new value, not confidence limits for the mean of the values.

## Predicted Values and Residuals Section

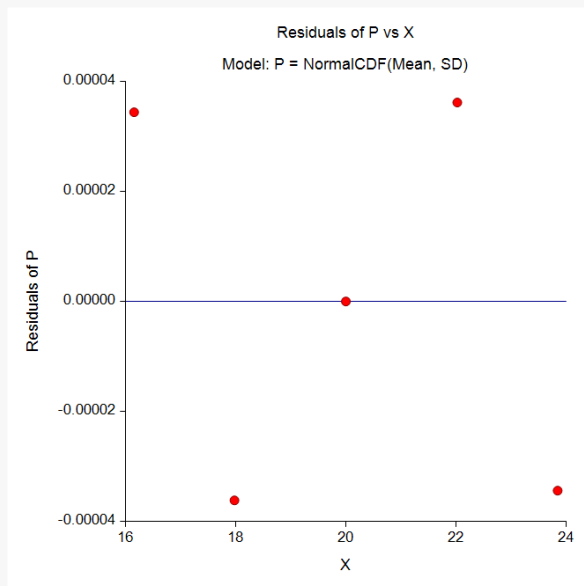
### Predicted Values and Residuals

Row No.	X	P		95% Prediction Interval Limits		Residual
		Observed	Predicted	Lower	Upper	
1	16.15500	0.10000	0.09997	0.09982	0.10012	0.00003
2	17.97700	0.25000	0.25004	0.24988	0.25019	-0.00004
3	20.00000	0.50000	0.50000	0.49985	0.50015	0.00000
4	22.02300	0.75000	0.74996	0.74981	0.75012	0.00004
5	23.84500	0.90000	0.90003	0.89988	0.90018	-0.00003

This section shows the values of the predicted values, prediction limits, and residuals. If you have observations in which the independent variable is given, but the dependent (Y) variable is blank, a predicted value and prediction limits will be generated and displayed in this report.

## Residual Plot(s)

### Residual Plot(s)



This is a scatter plot of the residuals versus the independent variable, X. The preferred pattern is a rectangular shape or point cloud. Any nonrandom pattern may require a redefinition of the model.

## Example 2 – Fitting a Normal Distribution with Groups

This section presents an example of how to fit a normal cdf to the two groups of data in the table below.

P	X	Type
0.10	16	A
0.25	18	A
0.50	20	A
0.75	22	A
0.90	24	A
0.10	15	B
0.25	17	B
0.50	19	B
0.75	21	B
0.90	23	B

These data are recorded in the *Normal with Groups* dataset. In this example, the dependent variable is P, and the independent variable is X. The groups are defined by the values of Type.

### Setup

To run this example, complete the following steps:

**1 Open the Normal with Groups example dataset**

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Normal with Groups** and click **OK**.

**2 Specify the Curve Fitting – CDF procedure options**

- Find and open the **Curve Fitting – CDF** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

```

P (Prob X < x) Variable .....P
X Variable .....X
Group Variable.....Type
CDF Model.....P = NormalCDF(Mean, SD)
Parameter 1 Name .....Mean
  Min Start Max.....Automatic
Parameter 2 Name .....SD
  Min Start Max.....Automatic
Bootstrap Confidence Intervals.....Unchecked
Randomization Hypothesis Tests .....Unchecked
    
```

## Reports Tab

Combined Summary Report .....	<b>Checked</b>
Group Difference F-Test .....	<b>Checked</b>
Iteration Summary Report.....	<b>Unchecked</b>
Iteration Detail Report.....	<b>Unchecked</b>
Model Estimation Report .....	<b>Checked</b>
Analysis of Variance Report .....	<b>Unchecked</b>
Parameter Correlations Report.....	<b>Unchecked</b>
Predicted Values Report.....	<b>Unchecked</b>
Residual Report.....	<b>Unchecked</b>

**3 Run the procedure**

- Click the **Run** button to perform the calculations and generate the output.

**Parameter Estimates for All Groups****Parameter Estimates for All Groups**

Type	Count	Iterations	R2	Mean	SD
A	5	5	0.99972	20.00000	3.04592
B	5	10	0.99972	19.00000	3.04596
Combined	10	14	0.97436	19.50000	3.11238

This report displays a summary of the results for each group and then for the case in which all groups are combined into one group.

**Group Name (Type)**

This column, headed by the name of the Group Variable, lists the group value that is displayed on this line.

**Count**

This is the number of observations used by the nonlinear regression algorithm.

**Iterations**

This is the number of iterations used by the nonlinear regression algorithm to find the estimates. You should note whether the maximum number of iterations has been reached (in which case the algorithm did not converge).

**R2**

This is the value of the pseudo R-squared value. A value near one indicates that the model fits the data well. A value near zero indicates that the model does not fit the data well.

**Mean SD**

The final values of the estimated parameters are displayed so that you may compare them across groups.



## Analysis of Variance Across Groups

### Analysis of Variance Across Groups

Type	Count	Iterations	Model R2	Error DF	Sum Squares Error	Mean Square Error
A	5	5	0.99972	3	0.0001246508	4.155028E-05
B	5	10	0.99972	3	0.0001246509	4.155031E-05
Combined	10	14	0.97436	8	0.0228172	0.00285215
Ignored			0.99972	6	0.0002493018	4.155029E-05

This report displays goodness of fit results for each group and then for the case in which all groups are combined into one dataset. The final row of the report, labeled 'Ignored', gives the goodness of fit statistics for the model in which a separate curve is fit for each group.

### Group Name (Type)

This column, headed by the name of the Group Variable, lists the group value that is displayed on this line.

### Count

This is the number of observations used by the nonlinear regression algorithm.

### Iterations

This is the number of iterations used by the nonlinear regression algorithm to find the estimates. You should note whether the maximum number of iterations has been reached (in which case the algorithm did not converge).

### R2

This is the value of the pseudo R-squared value. A value near one indicates that the model fits the data well. A value near zero indicates that the model does not fit the data well. Note

### Error DF

The degrees of freedom are the number of observations minus the number of parameters fit.

### Sum Squares Error

This is the sum of the squared residuals for this group.

### Mean Square Error

This is a rough estimate of the variance of the residuals for this group.

## Curve Inequality F-Test

### Curve Inequality F-Test

Curves Tested	DF	Mean Square	F-Ratio	F-Test P-Value
All	2	0.01128395	271.5732	0.00000
Error	6	4.155029E-05		

This report displays an F-Test of whether all of the group curves are equal. This test compares the residual sum of squares obtained when the grouping is ignored with the total of the residual sum of squares obtained for each group. This test is routinely used in the analysis of linear models. Its application to nonlinear models has occasionally been suggested. However, it is based on normality assumptions which seldom occur. When testing curve coincidence is important, we suggest you use a randomization test.

### Curves Tested

This column indicates the term presented on this row.

### DF

The degrees of freedom of this term.

### Mean Square

The mean square associated with this term.

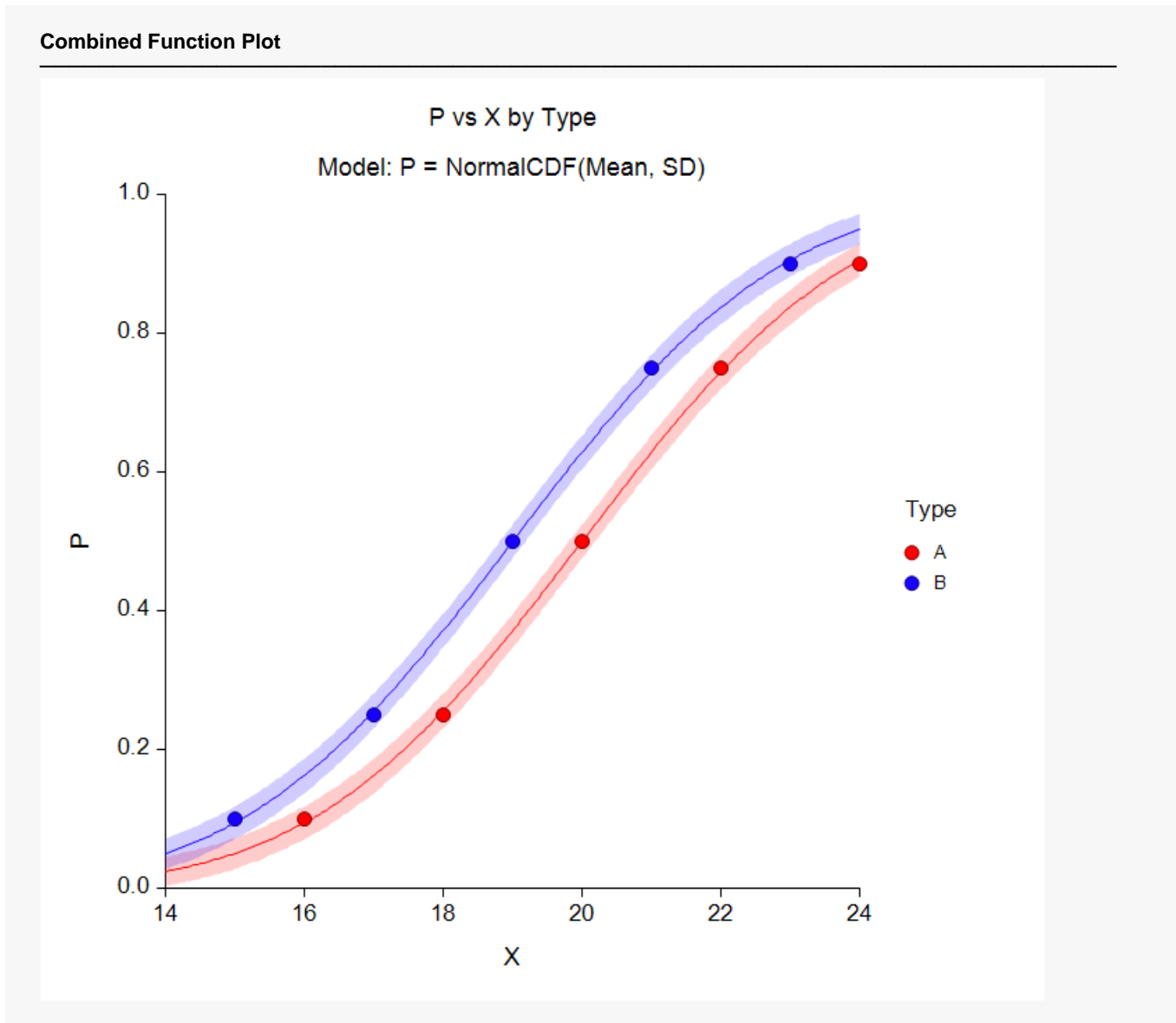
### F-Ratio

The F-ratio for testing the hypothesis that all curves coincide.

### F-Test P-Value

This is the probability level of the F-ratio. When this value is less than 0.05 (a common value for alpha), the test is 'significant' meaning that the hypothesis of equal curves is rejected. If this value is larger than the nominal level (0.05), the null hypothesis cannot be rejected. We do not have enough evidence to reject.

## Combined Function Plot



This plot displays all of the data and fitted curves, allowing you to quickly assess the quality of the results.

## Model Estimation Section

### Model Estimation for Type = A

Parameter Name	Parameter Estimate	Asymptotic Standard Error	95% Confidence Interval Limits for the Parameter Estimate*	
			Lower	Upper
Mean	19.9999991438107	0.03020	19.90389	20.09611
SD	3.04591773878094	0.04547	2.90121	3.19063

#### Estimation Summary

Iterations	5	Rows Read	5
R-Squared	0.999720	Rows Used	5
Random Seed	5369465	Total Count	5

\* These confidence intervals assume that the observations are independent. This is often not true for the observations fitted by this procedure. Also, the true sample size may not be properly represented.

### Model Estimation for Type = B

Parameter Name	Parameter Estimate	Asymptotic Standard Error	95% Confidence Interval Limits for the Parameter Estimate*	
			Lower	Upper
Mean	18.9999995927047	0.03020	18.90389	19.09611
SD	3.04595729775831	0.04547	2.90124	3.19067

#### Estimation Summary

Iterations	10	Rows Read	5
R-Squared	0.999720	Rows Used	5
Random Seed	5369465	Total Count	5

\* These confidence intervals assume that the observations are independent. This is often not true for the observations fitted by this procedure. Also, the true sample size may not be properly represented.

This report displays the details of the estimation of the model parameters for each group.

## Example 3 – Fitting a Weibull Distribution with Groups

This section presents an example of how to fit a Weibull cdf to the two groups of data in the table below.

P	X	Type	Weibull
0.10	16	A	0.945
0.25	18	A	1.32
0.50	20	A	1.77
0.75	22	A	2.23
0.90	24	A	2.64
0.10	15	B	1.00
0.25	17	B	1.30
0.50	19	B	1.80
0.75	21	B	2.20
0.90	23	B	2.60

These data are recorded in the *Normal with Groups* dataset. In this example, the dependent variable is *P*, and the independent variable is *Weibull*. The groups are defined by the values of *Type*.

### Setup

To run this example, complete the following steps:

**1 Open the Normal with Groups example dataset**

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Normal with Groups** and click **OK**.

**2 Specify the Curve Fitting – CDF procedure options**

- Find and open the **Curve Fitting – CDF** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

---

P (Prob X < x) Variable ..... **P**

X Variable ..... **Weibull**

Group Variable..... **Type**

CDF Model..... **P = WeibullCDF(Scale, Shape)**

Parameter 1 Name ..... **Scale**

    Min Start Max..... **Automatic**

Parameter 2 Name ..... **Shape**

    Min Start Max..... **Automatic**

Bootstrap Confidence Intervals..... **Unchecked**

Randomization Hypothesis Tests ..... **Unchecked**

Reports Tab

---

Combined Summary Report .....	<b>Checked</b>
Group Difference F-Test .....	<b>Checked</b>
Iteration Summary Report.....	<b>Unchecked</b>
Iteration Detail Report.....	<b>Unchecked</b>
Model Estimation Report .....	<b>Checked</b>
Analysis of Variance Report .....	<b>Unchecked</b>
Parameter Correlations Report.....	<b>Unchecked</b>
Predicted Values Report.....	<b>Checked</b>
Residual Report.....	<b>Unchecked</b>
Predict P at these X Values .....	<b>1 1.5 2 2.5 3</b>

### 3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

## Parameter Estimates for All Groups

**Parameter Estimates for All Groups**

---

Type	Count	Iterations	R2	Scale	Shape
A	5	3	1.00000	0.50005	3.00084
B	5	3	0.99830	0.50137	3.11936
Combined	10	4	0.99899	0.50074	3.05982

---

This report displays a summary of the results for each group and then for the case in which all groups are combined into one group.

## Analysis of Variance Across Groups

**Analysis of Variance Across Groups**

---

Type	Count	Iterations	Model R2	Error DF	Sum Squares Error	Mean Square Error
A	5	3	1.00000	3	6.630716E-08	2.210239E-08
B	5	3	0.99830	3	0.0007573197	0.0002524399
Combined	10	4	0.99899	8	0.0008951991	0.0001118999
Ignored			0.99915	6	0.000757386	0.000126231

---

This report displays goodness of fit results for each group and then for the case in which all groups are combined into one dataset. The final row of the report, labeled 'Ignored', gives the goodness of fit statistics for the model in which a separate curve is fit for each group.

## Curve Inequality F-Test

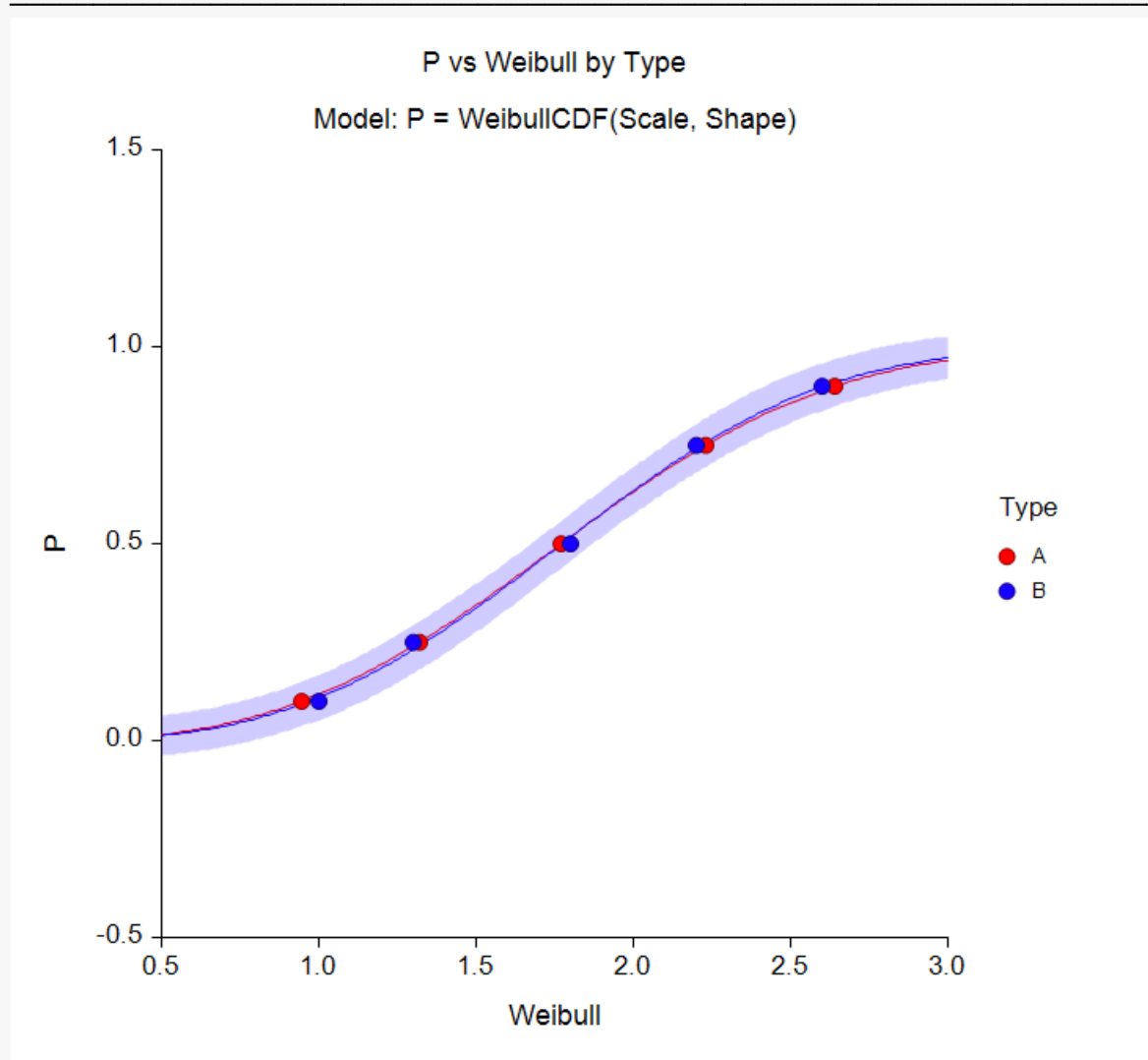
### Curve Inequality F-Test

Curves Tested	DF	Mean Square	F-Ratio	F-Test P-Value
All	2	6.890652E-05	0.5459	0.60561
Error	6	0.000126231		

This report displays an F-Test of whether all of the group curves are equal.

## Combined Function Plot

### Combined Function Plot



This plot displays all of the data and fitted curves, allowing you to quickly assess the quality of the results.

## Model Estimation Section

### Model Estimation for Type = A

Parameter Name	Parameter Estimate	Asymptotic Standard Error	95% Confidence Interval Limits for the Parameter Estimate*	
			Lower	Upper
Scale	0.500046637865741	0.00004	0.49991	0.50018
Shape	3.00084001655523	0.00106	2.99748	3.00420

### Estimation Summary

Iterations	3	Rows Read	5
R-Squared	1.000000	Rows Used	5
Random Seed	5395629	Total Count	5

\* These confidence intervals assume that the observations are independent. This is often not true for the observations fitted by this procedure. Also, the true sample size may not be properly represented.

### Predicted Values for Specified X Values for Type = A

Weibull	Predicted Value of P	95% Prediction Interval Limits	
		Lower	Upper
1.00000	0.11747	0.11693	0.11801
1.50000	0.34419	0.34362	0.34477
2.00000	0.63222	0.63166	0.63278
2.50000	0.85830	0.85772	0.85887
3.00000	0.96585	0.96534	0.96636

### Model Estimation for Type = B

Parameter Name	Parameter Estimate	Asymptotic Standard Error	95% Confidence Interval Limits for the Parameter Estimate*	
			Lower	Upper
Scale	0.501370283595292	0.00436	0.48748	0.51526
Shape	3.11936317911747	0.11713	2.74662	3.49211

### Estimation Summary

Iterations	3	Rows Read	5
R-Squared	0.998298	Rows Used	5
Random Seed	5395629	Total Count	5

\* These confidence intervals assume that the observations are independent. This is often not true for the observations fitted by this procedure. Also, the true sample size may not be properly represented.



## Curve Fitting – CDF

**Predicted Values for Specified X Values for Type = B**

<b>Weibull</b>	<b>Predicted Value of P</b>	<b>95% Prediction Interval Limits</b>	
		<b>Lower</b>	<b>Upper</b>
1.00000	0.10958	0.05237	0.16678
1.50000	0.33710	0.27581	0.39839
2.00000	0.63526	0.57553	0.69499
2.50000	0.86775	0.80662	0.92887
3.00000	0.97192	0.91810	1.02574

This report displays the details of the estimation of the model parameters for each group.