

CURVEFIT

```
CURVEFIT [VARIABLES=] varname [WITH varname]
  [/MODEL= [LINEAR**] [LOGARITHMIC] [INVERSE]
    [QUADRATIC] [CUBIC] [COMPOUND]
    [POWER] [S] [GROWTH] [EXPONENTIAL]
    [LGSTIC] [ALL]]
  [/CIN={95** }
    {value}]
  [/UPPERBOUND={NO**}
    {n } ]
  [/ {CONSTANT† }
    {NOCONSTANT}]
  [/PLOT={FIT**}
    {NONE } ]
  [/ID = varname]
  [/PRINT=ANOVA]
  [/SAVE=[PRED] [RESID] [CIN]]
  [/APPLY [= 'model name'] [{SPECIFICATIONS}]
    {FIT } ]
```

**Default if the subcommand is omitted.

†Default if the subcommand is omitted and there is no corresponding specification on the TSET command.

Example

```
CURVEFIT VARY
  /MODEL=CUBIC.
```

Overview

CURVEFIT fits selected curves to a line plot, allowing you to examine the relationship between one or more dependent variables and one independent variable. CURVEFIT also fits curves to time series and produces forecasts, forecast errors, lower confidence limits, and upper confidence limits. You can choose curves from a variety of regression models.

Options

Model Specification. There are 11 regression models available on the MODEL subcommand. You can fit any or all of these to the data. The keyword ALL is available to fit all 11 models. You can control whether the regression equation includes a constant term using the CONSTANT or NOCONSTANT subcommand.

Upperbound Value. You can specify the upperbound value for the logistic model using the `UPPERBOUND` subcommand.

Output. You can produce an analysis-of-variance summary table using the `PRINT` subcommand. You can suppress the display of the curve-fitting plot using the `PLOT` subcommand.

New Variables. To evaluate the regression statistics without saving predicted and residual variables, specify `TSET NEWVAR=NONE` prior to `CURVEFIT`. To save the new variables and replace the variables saved earlier, use `TSET NEWVAR=CURRENT` (the default). To save the new variables without erasing variables saved earlier, use `TSET NEWVAR=ALL` or the `SAVE` subcommand on `CURVEFIT`.

Forecasting. When used with the `PREDICT` command, `CURVEFIT` can produce forecasts and confidence limits beyond the end of the series. For more information, see “`PREDICT`” on p. 1313.

Basic Specification

The basic specification is one or more dependent variables. If the variables are not time series, you must also specify the keyword `WITH` and an independent variable.

- By default, the `LINEAR` model is fit.
- A 95% confidence interval is used unless it is changed by a `TSET CIN` command prior to the procedure.
- `CURVEFIT` produces a plot of the curve, a regression summary table displaying the type of curve used, the R^2 coefficient, degrees of freedom, overall F test and significance level, and the regression coefficients.
- For each variable and model combination, `CURVEFIT` creates four variables: fit/forecast values, residuals, lower confidence limits, and upper confidence limits. These variables are automatically labeled and added to the active dataset unless `TSET NEWVAR=NONE` is specified prior to `CURVEFIT`. For more information, see “`SAVE Subcommand`” on p. 434.

Subcommand Order

- Subcommands can be specified in any order.

Syntax Rules

- `VARIABLES` can be specified only once.
- Other subcommands can be specified more than once, but only the last specification of each one is executed.

Operations

- When `CURVEFIT` is used with the `PREDICT` command to forecast values beyond the end of a time series, the original and residual series are assigned the system-missing value after the last case in the original series.
- If a model requiring a log transformation (`COMPOUND`, `POWER`, `S`, `GROWTH`, `EXPONENTIAL`, or `LGSTIC`) is requested and there are values in the dependent variable(s) less than or equal to 0, the model cannot be fit because nonpositive values cannot be log-transformed.

- CURVEFIT uses listwise deletion of missing values. Whenever one dependent variable is missing a value for a particular case or observation, that case or observation will not be included in any computations.
- For the models QUADRATIC and CUBIC, a message is issued if the tolerance criterion is not met. (See TSET for information on changing the tolerance criterion.)
- Since CURVEFIT automatically generates four variables for each dependent variable and model combination, the ALL specification after MODEL should be used cautiously to avoid creating and adding to the active dataset many more variables than are necessary.
- The residual variable is always reported in the original metric. To compute the logged residual (which should be used for diagnostic checks) for the models COMPOUND, POWER, S, GROWTH, and EXPONENTIAL, specify

```
COMPUTE NEWVAR = LN(VAR) - LN(FIT#n) .
```

where *NEWVAR* is the logged residual, *VAR* is the name of the dependent variable or observed series, and *FIT#n* is the name of the fitted variable generated by CURVEFIT.

For the LGSTIC (logistic) model, the logged residual can be obtained by

```
COMPUTE NEWERR = LN(VAR) - LN(1/FIT#n) .
```

or, if upperbound value *u* is specified on the UPPERBOUND subcommand, by

```
COMPUTE NEWVAR = LN(1/VAR - 1/u) - LN(1/FIT#n) .
```

- CURVEFIT obeys the WEIGHT command when there is an independent variable. The WEIGHT specification is ignored if no independent variable is specified.

Limitations

- A maximum of 1 VARIABLES subcommand. There is no limit on the number of dependent variables or series named on the subcommand.
- A maximum of 1 independent variable can be specified after the keyword WITH.

Example

```
CURVEFIT VARY
/MODEL=CUBIC .
```

- This example fits a cubic curve to the series *VARY*.

VARIABLES Subcommand

VARIABLES specifies the variables and is the only required subcommand. The actual keyword VARIABLES can be omitted.

- If the dependent variables specified are not time series, you must also specify the keyword WITH and an independent variable.

MODEL Subcommand

MODEL specifies the model or models to be fit to the data. The default model is LINEAR.

- You can fit any or all of the 11 available models.
- Model name keywords can be abbreviated to the first three characters.
- You can use the keyword ALL to fit all models.
- When the LGSTIC model is specified, the upperbound value is included in the output.

The following table lists the available models and their regression equations. The linear transformations for the last six models are also shown.

Keyword	Equation	Linear equation
LINEAR	$Y = b_0 + b_1 t$	
LOGARITHMIC	$Y = b_0 + b_1 \ln(t)$	
INVERSE	$Y = b_0 + b_1/t$	
QUADRATIC	$Y = b_0 + b_1 t + b_2 t^2$	
CUBIC	$Y = b_0 + b_1 t + b_2 t^2 + b_3 t^3$	
COMPOUND	$Y = b_0 b_1^t$	$\ln(Y) = \ln(b_0) + t \ln(b_1)$
POWER	$Y = b_0 t^{b_1}$	$\ln(Y) = \ln(b_0) + b_1 \ln(t)$
S	$Y = e^{b_0 + b_1/t}$	$\ln(Y) = b_0 + b_1/t$
GROWTH	$Y = e^{b_0 + b_1 t}$	$\ln(Y) = b_0 + b_1 t$
EXPONENTIAL	$Y = b_0 e^{b_1 t}$	$\ln(Y) = \ln(b_0) + b_1 t$
LGSTIC (logistic)	$Y = (1/u + b_0 b_1^t)_{-1}$	$\ln(1/Y - 1/u) = \ln(b_0) + t \ln(b_1)$

where

b_0 = a constant

b_n = regression coefficient

t = independent variable or time value

\ln = the natural logarithm

u = upperbound value for LGSTIC

Example

CURVEFIT VARX.

- This command fits a curve to *VARX* using the linear regression model (the default).

Example

```
CURVEFIT VARY
/MODEL=GROWTH EXPONENTIAL.
```

- This command fits two curves to *VARY*, one using the growth model and the other using the exponential model.

UPPERBOUND Subcommand

UPPERBOUND is used with the logistic model (keyword LGSTIC) to specify an upper boundary value to be used in the regression equation.

- The specification on UPPERBOUND must be a positive number and must be greater than the largest data value in any of the specified dependent variables.
- The default UPPERBOUND value is infinity, so that $1/u = 0$ and is dropped from the equation.
- You can specify UPPERBOUND NO to reset the value to infinity when applying a previous model.
- If you specify UPPERBOUND without LGSTIC, it is ignored.
- Note that UPPERBOUND is a subcommand and cannot be used within a MODEL subcommand. For example, the following specification is *not* valid:

```
/MODEL=CUBIC LGSTIC /UPPER=99 LINEAR
```

The correct specification is:

```
/MODEL=CUBIC LGSTIC LINEAR
/UPPER=99
```

CONSTANT and NOCONSTANT Subcommands

CONSTANT and NOCONSTANT indicate whether a constant term should be estimated in the regression equation. The specification overrides the corresponding setting on the TSET command.

- CONSTANT indicates that a constant should be estimated. It is the default unless changed by TSET NOCONSTANT prior to the current procedure.
- NOCONSTANT eliminates the constant term from the model.

Example

```
CURVEFIT Y1
/MODEL=COMPOUND
/NOCONSTANT.
```

- In this example, a compound curve is fit to *Y1* with no constant term in the model.

CIN Subcommand

CIN controls the size of the confidence interval.

- The specification on CIN must be greater than 0 and less than 100.
- The default confidence interval is 95.
- The CIN subcommand overrides the TSET CIN setting.

PLOT Subcommand

PLOT specifies whether the curve-fitting plot is displayed. If PLOT is not specified, the default is FIT. The curve-fitting plot is displayed. PLOT=FIT is generally used with an APPLY subcommand to turn off a PLOT=NONE specification in the applied model.

FIT	<i>Display the curve-fitting plot.</i>
NONE	<i>Do not display the plot.</i>

ID Subcommand

ID specifies an identification variable. When in point selection mode, you can click on an individual chart point to display the value of the ID variable for the selected case.

SAVE Subcommand

SAVE saves the values of predicted, residual, and/or confidence interval variables generated during the current session in the active dataset.

- SAVE saves the specified variables with default names: *FIT_n* for predicted values, *ERR_n* for residuals, *LCL_n* for the lower confidence limit, and *UCL_n* for the upper confidence limit, where *n* increments each time any variable is saved for a model.
- SAVE overrides the CURRENT or NONE setting on TSET NEWVARS (see TSET).

PRED	<i>Predicted variable.</i>
RESID	<i>Residual variable.</i>
CIN	<i>Confidence interval.</i>

PRINT Subcommand

PRINT is used to produce an additional analysis-of-variance table for each model and variable.

- The only specification on PRINT is the keyword ANOVA.

APPLY Subcommand

APPLY allows you to use a previously defined CURVEFIT model without having to repeat the specifications.

- The specifications on APPLY can include the name of a previous model in quotes and one of two keywords. All of these specifications are optional.
- If a model name is not specified, the model specified on the previous CURVEFIT command is used.
- To change one or more of the specifications of the model, specify the subcommands of only those portions you want to change after the subcommand APPLY.
- If no variables or series are specified on the CURVEFIT command, the dependent variables that were originally specified with the model being reapplied are used.
- To change the dependent variables used with the model, enter new variable names before or after the APPLY subcommand.

The keywords available for APPLY on CURVEFIT are:

SPECIFICATIONS	<i>Use only the specifications from the original model. This is the default.</i>
FIT	<i>Use the coefficients estimated for the original model in the equation.</i>

Example

```
CURVEFIT X1
  /MODEL=QUADRATIC.
CURVEFIT Z1
  /APPLY.
```

- The first command fits a quadratic curve to *X1*.
- The second command fits the same type of curve to *Z1*.

Example

```
CURVEFIT X1 Y1 Z1
  /MODEL=QUADRATIC.
CURVEFIT APPLY
  /MODEL=CUBIC.
```

- The first command fits quadratic curves to *X1*, *Y1*, and *Z1*.
- The second command fits curves to the same three series using the cubic model.

References

- Abraham, B., and J. Ledolter. 1983. *Statistical methods of forecasting*. New York: John Wiley and Sons.
- Draper, N. R., and H. Smith. 1981. *Applied regression analysis*, 2nd ed. New York: John Wiley and Sons.

Montgomery, D. C., and E. A. Peck. 1982. *Introduction to linear regression analysis*. New York: John Wiley and Sons.

DATA LIST

```
DATA LIST [FILE=file] [{FIXED}] [RECORDS={1}] [SKIP={n}] [{TABLE  }]
                                     {n}                {NOTABLE}

                                     {FREE}  [{"delimiter", "delimiter,..., TAB}]]
                                     {LIST}

/{1   } varname {col location [(format)]} [varname ...]
 {rec #}          {(FORTRAN-like format)  }

[/ {2   } ...] [/ ...]
 {rec #}
```

Numeric and string input formats:

Type	Column-style format	FORTRAN-like format
Numeric (default)	d or F, d	Fw.d
Restricted numeric	N, d	Nw.d
Scientific notation	E, d	Ew.d
Numeric with commas	COMMA, d	COMMAw.d
Numeric with dots	DOT, d	DOTw.d
Numeric with commas and dollar sign	DOLLAR, d	DOLLARw.d
Numeric with percent sign	PCT, d	PCTw.d
Zoned decimal	Z, d	Zw.d
String	A	Aw

Format elements to skip columns:

Type	Column-style format	FORTRAN-like format
Tab to column <i>n</i>		Tn
Skip <i>n</i> columns		nX

Date and time input formats:

Type	Data input	Format	FORTRAN-like format
International date	dd-mmm-yyyy	DATE	DATEw
American date	mm/dd/yyyy	ADATE	ADATEw
European date	dd/mm/yy	EDATE	EDATEw
Julian date	yyddd	JDATE	JDATEw
Sorted date	yy/mm/dd	SDATE	SDATEw
Quarter and year	qQyyyy	QYR	QYRw
Month and year	mm/yyyy	MOYR	MOYRw
Week and year	wkWKyyyy	WKYR	WKYRw