Previous Page | Next Page

## Syntax: ESTIMATE Statement

```
ESTIMATE <"label"> estimate-specification <(divisor=n)> <, <"label"> estimate-specification <(divisor=n)> > <, ...> </ options> ;
```

The basic element of the ESTIMATE statement is the **estimate-specification**, which consists of model effects and their coefficients. A **estimate-specification** takes the general form

effect name <effect values ...>

The following variables can appear in the ESTIMATE statement:

#### label

is an optional label that identifies the particular row of the estimate in the output.

#### effect

identifies an effect that appears in the MODEL statement. The keyword INTERCEPT can be used as an effect when an intercept is fitted in the model. You do not need to include all effects that are in the MODEL statement.

#### values

are constants that are elements of the  ${\bf L}$  matrix and are associated with the fixed and random effects. There are two basic methods of specifying the entries of the  ${\bf L}$  matrix. The traditional representation—also known as the positional syntax—relies on entering coefficients in the position they assume in the  ${\bf L}$  matrix. For example, in the following statements the elements of  ${\bf L}$  that are associated with the  ${\it b}$  main effect receive a  ${\bf l}$  in the first position and a  ${\bf -l}$  in the second position:

```
class a b;
model y = a b a*b;
estimate 'B at A2' b 1 -1 a*b 0 0 1 -1;
```

The elements that are associated with the interaction receive a  $\mathbb{I}$  in the third position and a  $-\mathbb{I}$  in the fourth position. In order to specify coefficients correctly for the interaction term, you need to know how the levels of a and b vary in the interaction, which is governed by the order of the variables in the CLASS statement. The nonpositional syntax is designed to make it easier to enter coefficients for interactions and is necessary to enter coefficients for effects that are constructed with the EFFECT statement. In square brackets you enter the coefficient followed by the associated levels of the CLASS variables. If B has two levels and A has three levels, the previous ESTIMATE statement, by using nonpositional syntax for the interaction term, becomes the following statement:

```
estimate 'B at A2' b 1 -1 a*b [1, 2 1] [-1, 2 2];
```

The previous statement assigns value  ${\bf l}$  to the interaction where A is at level 2 and B is at level 1, and it assigns  $-{\bf l}$  to the interaction where both classification variables are at level 2. The comma that separates the entry for the  ${\bf l}$  matrix from the level indicators is optional. Further details about the nonpositional contrast syntax and its use with constructed effects can be found in the section <u>Positional and Nonpositional Syntax for Coefficients in Linear Functions</u>.

Based on the **estimate-specifications** in your ESTIMATE statement, the procedure constructs the matrix  $\mathbf{L}$  to test the hypothesis  $H: \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ . The procedure supports nonpositional syntax for the coefficients of model effects in the ESTIMATE statement. For details see the section <u>Positional and Nonpositional Syntax for Coefficients in Linear Functions</u>.

The procedure then produces for each row  $\mathbf{l}$  of  $\mathbf{L}$  an approximate t test of the hypothesis  $H: \mathbf{l}\boldsymbol{\beta} = 0$ . You can also obtain multiplicity-adjusted  $\boldsymbol{p}$ -values and confidence limits for multirow estimates with the  $\underline{\mathsf{ADJUST}}$  option.

Note that multirow estimates are permitted. Unlike releases prior to SAS 9.22, you do not need to specify a '*label'* for every row of the estimate; the procedure constructs a default label if a label is not specified.

If the procedure finds the estimate to be nonestimable, then it displays "Non-est" for the estimate entry.

<u>Table 19.17</u> summarizes important options in the ESTIMATE statement. All ESTIMATE options are subsequently discussed in alphabetical order.

Table 19.17 Important ESTIMATE Statement Options

	,
Option	Description
Construction and Computation of Estimable Functions	
<b>DIVISOR=</b>	Specifies a list of values to divide the coefficients
<u>NOFILL</u>	Suppresses the automatic fill-in of coefficients for higher-order effects
SINGULAR=	Tunes the estimability checking difference
Degrees of Freedom and <b>p</b> -values	
ADJUST=	Determines the method for multiple comparison adjustment of estimates
$ALPHA=\alpha$	Determines the confidence level $(1-lpha)$
<b>LOWER</b>	Performs one-sided, lower-tailed inference
<u>STEPDOWN</u>	Adjusts multiplicity-corrected <b>p</b> -values further in a step-down fashion
TESTVALUE=	Specifies values under the null hypothesis for tests
<u>UPPER</u>	Performs one-sided, upper-tailed inference
Statistical Output	
<u>CL</u>	Constructs confidence limits
<u>CORR</u>	Displays the correlation matrix of estimates
COV	Displays the covariance matrix of estimates
<u>E</u>	Prints the L matrix
<u>JOINT</u>	Produces a joint $F$ or chi-square test for the estimable functions
PLOTS=	Requests ODS statistical graphics if the analysis is sampling-based
SEED=	Specifies the seed for computations that depend on random numbers
Generalized Linear Modeling	
CATEGORY=	Specifies how to construct estimable functions with multinomial data
<u>EXP</u>	Exponentiates and displays estimates
<u>ILINK</u>	Computes and displays estimates and standard errors on the inverse linked scale

You can specify the following options in the ESTIMATE statement after a slash (/).

## ADJDFE=SOURCE ADJDFE=ROW

specifies how denominator degrees of freedom are determined when **p**-values and confidence limits are adjusted for multiple comparisons with the <u>ADJUST=</u> option. When you do not specify the ADJDFE= option, or when you specify ADJDFE=SOURCE, the denominator degrees of freedom for multiplicity-adjusted results are the denominator degrees of freedom for the final effect that is listed in the ESTIMATE statement from the "Type III" table.

The ADJDFE=ROW setting is useful if you want multiplicity adjustments to take into account that denominator degrees of freedom are not constant across estimates. For example, this can be the case when the denominator degrees of freedom are computed by the Satterthwaite method or according to Kenward and Roger (1997).

The ADJDFE= option has an effect only in mixed models that use these degree-of-freedom methods. It is not supported by the procedures that perform chi-square-based inference (LOGISTIC, PHREG, and

SURVEYLOGISTIC).

ADJUST=BON
ADJUST=SCHEFFE
ADJUST=SIDAK
ADJUST=SIMULATE<(simoptions)>
ADJUST=T

requests a multiple comparison adjustment for the **p**-values and confidence limits for the estimates. The adjusted quantities are produced in addition to the unadjusted quantities. Adjusted confidence limits are produced if the <u>CL</u> or <u>ALPHA=</u> option is in effect. For a description of the adjustments, see Chapter 39, <u>The GLM Procedure</u>, and Chapter 58, <u>The MULTTEST Procedure</u>, and the documentation for the <u>ADJUST=</u> option in the <u>LSMEANS</u> statement.

If the STEPDOWN option is in effect, the p-values are further adjusted in a step-down fashion.

#### ALPHA=number

requests that a t type confidence interval be constructed with confidence level 1 - number. The value of number must be between 0 and 1; the default is 0.05. If the "Estimates" table shows infinite degrees of freedom, then the confidence interval is a z type interval.

## CATEGORY=category-options

specifies how to construct estimates and multiplicity corrections for models with multinomial data (ordinal or nominal). This option is also important for constructing sets of estimable functions for **F** or chi-square tests with the <u>JOINT</u> option.

The *category-options* are used to indicate how response variable levels are treated in constructing the estimable functions. Possible values for the *category-options* are the following:

### **JOINT**

computes the estimable functions for every nonredundant category and treats them as a set. For example, a three-row ESTIMATE statement in a model with three response categories leads to six estimable functions.

## **SEPARATE**

computes the estimable functions for every nonredundant category in turn. For example, a three-row ESTIMATE statement in a model with three response categories leads to two sets of three estimable functions.

## quoted-value-list

computes the estimable functions only for the list of values given. The list must consist of formatted values of the response categories.

Consider the following ESTIMATE statements in the LOGISTIC procedure for an ordinal model with response categories 'vg', 'g', 'm', 'b', and 'vb'. Because there are five response categories, there are four nonredundant categories for the cumulative link model.

run;

The first ESTIMATE statement requests a two-row estimable function. The result is produced for two of the four nonredundant response categories. The second ESTIMATE statement produces four *t* tests, one for each nonredundant category. The multiplicity adjustment with *p*-value computation by simulation treats the four estimable functions as a unit for family-wise Type I error protection. The third ESTIMATE statement computes a two-row estimable function and reports its results separately for all nonredundant categories. The Bonferroni adjustment in this statement applies to a family of two tests that correspond to the two-row estimable function. Four Bonferroni adjustments for sets of size two are performed.

adjust=bon;

The CATEGORY= option is supported only by the procedures that support generalized linear modeling (LOGISTIC and SURVEYLOGISTIC) and by PROC PLM when it is used to perform statistical analyses on item stores created by these procedures.

#### CHISQ

requests that chi-square tests be performed in addition to **F** tests, when you request an **F** test with the <u>JOINT</u> option. This option has no effect in procedures that produce chi-square statistics by default.

#### CL

requests that t type confidence limits be constructed. If the procedure shows the degrees of freedom in the "Estimates" table as infinite, then the confidence limits are z intervals. The confidence level is 0.95 by default, and you can change the confidence level with the <u>ALPHA=</u> option. The confidence intervals are adjusted for multiplicity when you specify the ADJUST= option. However, if a step-down p-value adjustment is requested with the <u>STEPDOWN</u> option, only the p-values are adjusted for multiplicity.

### **CORR**

displays the estimated correlation matrix of the linear combination of the parameter estimates.

## COV

displays the estimated covariance matrix of the linear combination of the parameter estimates.

## DF=number

specifies the degrees of freedom for the *t* test and confidence limits. This option is not supported by the procedures that perform chi-square-based inference (LOGISTIC, PHREG, and SUVEYLOGISTIC).

## DIVISOR=value-list

specifies a list of values by which to divide the coefficients so that fractional coefficients can be entered as integer numerators. If you do not specify *value-list*, a default value of 1.0 is assumed. Missing values in the *value-list* are converted to 1.0.

If the number of elements in *value-list* exceeds the number of rows of the estimate, the extra values are ignored. If the number of elements in *value-list* is less than the number of rows of the estimate, the last value in *value-list* is copied forward.

If you specify a row-specific divisor as part of the specification of the estimate row, this value multiplies the corresponding divisor that is implied by the *value-list*. For example, the following statement divides the coefficients in the first row by 8, and the coefficients in the third and fourth row by 3:

Coefficients in the second row are not altered.

## Ε

requests that the L matrix coefficients be displayed.

## **EXP**

requests exponentiation of the estimate. When you model data with the logit, cumulative logit, or generalized logit link functions, and the estimate represents a log odds ratio or log cumulative odds ratio, the EXP option produces an odds ratio. In proportional hazards model, this option produces estimates of hazard ratios. If you specify the <u>CL</u> or <u>ALPHA=</u> option, the (adjusted) confidence bounds are also exponentiated.

The EXP option is supported only by PROC PHREG, PROC SURVEYPHREG, the procedures that support generalized linear modeling (LOGISTIC and SURVEYLOGISTIC), and by PROC PLM when it is used to perform statistical analyses on item stores created by these procedures.

### **ILINK**

requests that the estimate and its standard error also be reported on the scale of the mean (the inverse linked scale). The computation of the inverse linked estimate depends on the estimation mode. For example, if the analysis is based on a posterior sample when a BAYES statement is present, the inversely linked estimate is the average of the inversely linked values across the sample of posterior parameter estimates. If the analysis is not based on a sample of parameter estimates, the procedure computes the value on the mean scale by applying the inverse link to the estimate.

The interpretation of this quantity depends on the *effect values* specified in your ESTIMATE statement and on the link function. For example, in a model for binary data with logit link the following statements compute

$$\frac{1}{1+\exp\{-(\alpha_1-\alpha_2)\}}$$

where  $\alpha_1$  and  $\alpha_2$  are the fixed-effects solutions that are associated with the first two levels of the classification effect A:

```
class A;
model y = A / dist=binary link=logit;
estimate 'A one vs. two' A 1 -1 / ilink;
```

This quantity is not the difference of the probabilities that are associated with the two levels,

$$\pi_1 - \pi_2 = \frac{1}{1 + \exp\{-\beta_0 - \alpha_1\}} - \frac{1}{1 + \exp\{-\beta_0 - \alpha_2\}}$$

The standard error of the inversely linked estimate is based on the delta method. If you also specify the <u>CL</u> option, the procedure computes confidence limits for the estimate on the mean scale. In multinomial models for nominal data, the limits are obtained by the delta method. In other models they are obtained from the inverse link transformation of the confidence limits for the estimate. The ILINK option is specific to an ESTIMATE statement.

The ILINK option is supported only by the procedures that support generalized linear modeling (LOGISTIC and SURVEYLOGISTIC) and by PROC PLM when it is used to perform statistical analyses on item stores created by these procedures.

## JOINT<(joint-test-options)>

requests that a joint  $\mathbf{F}$  or chi-square test be produced for the rows of the estimate. The JOINT option in the ESTIMATE statement essentially replaces the CONTRAST statement.

When the <u>LOWERTAILED</u> or the <u>UPPERTAILED</u> options are in effect, or if the BOUNDS option described below is in effect, the JOINT option produces the chi-bar-square statistic according to Silvapulle and Sen (2004). This statistic uses a simulation-based approach to compute *p*-values in situations where the alternative hypotheses of the estimable functions are not simple two-sided hypotheses. See the section <u>Joint Hypothesis Tests with Complex Alternatives</u>, the <u>Chi-Bar-Square Statistic</u> for more information about this test statistic.

You can specify the following *joint-test-options* in parentheses:

## ACC=γ

specifies the accuracy radius for determining the necessary sample size in the simulation-based approach of Silvapulle and Sen (2004) for tests with order restrictions. The value of  $\gamma$  must be strictly between 0 and 1; the default value is 0.005.

 $EPS=\varepsilon$ 

specifies the accuracy confidence level for determining the necessary sample size in the simulation-based approach of Silvapulle and Sen (2004) for tests with order restrictions. The value of  $\varepsilon$  must be strictly between 0 and 1; the default value is 0.01.

#### LABEL='label'

assigns an identifying label to the joint test. If you do not specify a label, the first non-default label for the ESTIMATE rows is used to label the joint test.

## NOEST ONLY

performs only the **F** or chi-square test and suppresses other results from the ESTIMATE statement. This option is useful for emulating the CONTRAST statement that is available in other procedures.

#### NSAMP=n

specifies the number of samples for the simulation-based method of Silvapulle and Sen (2004). If n is not specified, it is constructed from the values of the ALPHA= $\alpha$ , the ACC= $\gamma$ , and the EPS= $\varepsilon$  options. With the default values for  $\gamma$ ,  $\varepsilon$ , and  $\alpha$  (0.005, 0.01, and 0.05, respectively), NSAMP=12,604 by default.

#### CHISQ

adds a chi-square test if the procedure produces an F test by default.

## **BOUNDS=***value-list*

specifies boundary values for the estimable linear function. The null value of the hypothesis is always zero. If you specify a positive boundary value z, the hypotheses are  $H\colon\theta=0$ ,  $H_a\colon\theta>0$  with the added constraint that  $\theta< z$ . The same is true for negative boundary values. The alternative hypothesis is then  $H_a\colon\theta<0$  subject to the constraint  $\theta>-|z|$ . If you specify a missing value, the hypothesis is assumed to be two-sided. The BOUNDS option enables you to specify sets of one- and two-sided joint hypotheses. If all values in *value-list* are set to missing, the procedure performs a simulation-based p-value calculation for a two-sided test.

## LOWER LOWERTAILED

requests that the p-value for the t test be based only on values that are less than the test statistic. A two-tailed test is the default. A lower-tailed confidence limit is also produced if you specify the CL or ALPHA= option.

Note that for ADJUST=SCHEFFE the one-sided adjusted confidence intervals and one-sided adjusted p-values are the same as the corresponding two-sided statistics, because this adjustment is based on only the right tail of the F distribution.

If you request a joint test with the <u>JOINT</u> option, then a one-sided left-tailed order restriction is applied to all estimable functions, and the corresponding chi-bar-square statistic of Silvapulle and Sen (2004) is computed in addition to the two-sided, standard, *F* or chi-square statistic. See the <u>JOINT</u> option for how to control the computation of the simulation-based chi-bar-square statistic.

## **NOFILL**

suppresses the automatic fill-in of coefficients of higher-order effects.

## PLOTS=plot-options

produces ODS statistical graphics of the distribution of estimable functions if the procedure performs the analysis in a sampling-based mode. For example, this is the case when procedures support a BAYES statement and perform a Bayesian analysis. The estimable functions are then computed for each of the posterior parameter estimates, and the "Estimates" table reports simple descriptive statistics for the evaluated functions. The PLOTS= option enables you in this situation to visualize the distribution of the estimable function. The following *plot-options* are available:

## ALL

produces all possible plots with their default settings.

## BOXPLOT<(boxplot-options)>

produces box plots of the distribution of the estimable function across the posterior sample. A separate box is generated for each estimable function, and all boxes appear on a single graph by default. You can affect the appearance of the box plot graph with the following options:

### ORIENTATION=VERTICAL|HORIZONTAL

#### ORIENT=VERT|HORIZ

specifies the orientation of the boxes. The default is vertical orientation of the box plots.

## **NPANELPOS=**number

specifies how to break the series of box plots across multiple panels. If the NPANELPOS option is not specified, or if *number* equals zero, then all box plots are displayed in a single graph; this is the default. If a negative number is specified, then exactly up to *number* of box plots are displayed per panel. If *number* is positive, then the number of boxes per panel is balanced to achieve small variation in the number of box plots per graph.

# DISTPLOT<(distplot-options)> DIST<(distplot-options)>

generates panels of histograms with a kernel density overlaid. A separate plot in each panel contains the results for each estimable function. You can specify the following *distplot-options* in parentheses:

### **BOX**|**NOBOX**

controls the display of a horizontal box plot of the estimable function's distribution across the posterior sample below the graph. The BOX option is enabled by default.

### **HIST | NOHIST**

controls the display of the histogram of the estimable function's distribution across the posterior sample. The HIST option is enabled by default.

### NORMAL | NONORMAL

controls the display of a normal density estimate on the graph. The NONORMAL option is enabled by default

## KERNEL | NOKERNEL

controls the display of a kernel density estimate on the graph. The KERNEL option is enabled by default.

#### NROWS=number

specifies the highest number of rows in a panel. The default is 3.

#### NCOLS=number

specifies the highest number of columns in a panel. The default is 3.

#### **UNPACK**

unpacks the panel into separate graphics.

#### **NONE**

does not produce any plots.

#### SEED=number

specifies the seed for the sampling-based components of the computations for the ESTIMATE statement (for example, chi-bar-square statistics and simulated **p**-values). **number** specifies an integer that is used to start the pseudo-random number generator for the simulation. If you do not specify a seed, or if you specify a value less than or equal to zero, the seed is generated from reading the time of day from the computer clock. There could be multiple ESTIMATE statements with SEED= specifications and there could be other statements that can supply a random number seed. Since the procedure has only one random number stream, the initial seed is shown in the SAS log.

## SINGULAR=number

tunes the estimability checking. If  $\mathbf{v}$  is a vector, define ABS( $\mathbf{v}$ ) to be the largest absolute value of the elements of  $\mathbf{v}$ . If ABS( $\mathbf{L} - \mathbf{L} \mathbf{T}$ ) is greater than  $c^*$ *number* for any row of  $\mathbf{L}$  in the contrast, then  $\mathbf{L} \boldsymbol{\beta}$  is declared nonestimable. Here,  $\mathbf{T}$  is the Hermite form matrix ( $\mathbf{X}'\mathbf{X}$ ) $^-\mathbf{X}'\mathbf{X}$ , and c is ABS( $\mathbf{L}$ ), except when it equals 0, and then c is 1. The value for *number* must be between 0 and 1; the default is 1E-4.

## STEPDOWN<(step-down-options)>

requests that multiplicity adjustments for the **p**-values of estimates be further adjusted in a step-down fashion. Step-down methods increase the power of multiple testing procedures by taking advantage of the fact that a **p**-value is never declared significant unless all smaller **p**-values are also declared significant. The STEPDOWN adjustment combined with <u>ADJUST=BON</u> corresponds to the methods of Holm (1979) and "Method 2" of Shaffer (1986); this is the default. Using step-down-adjusted **p**-values combined with <u>ADJUST=SIMULATE</u> corresponds to the method of Westfall (1997).

If the ESTIMATE statement is applied with a STEPDOWN option in a mixed model where the degrees-of-freedom method is that of Kenward and Roger (1997) or of Satterthwaite, then step-down-adjusted **p**-values are produced only if the <u>ADJDFE</u>=ROW option is in effect.

Also, the STEPDOWN option affects only **p**-values, not confidence limits. For <u>ADJUST=SIMULATE</u>, the generalized least squares hybrid approach of Westfall (1997) is used to increase Monte Carlo accuracy. You can specify the following **step-down-options** in parentheses after the STEPDOWN option:

#### MAXTIME=n

specifies the time (in seconds) to be spent computing the maximal logically consistent sequential subsets of equality hypotheses for TYPE=LOGICAL. The default is MAXTIME=60. If the MAXTIME value is exceeded, the adjusted tests are not computed. When this occurs, you can try increasing the MAXTIME value. However, note that there are common multiple comparisons problems for which this computation requires a huge amount of time—for example, all pairwise comparisons between more than 10 groups. In such cases, try to use TYPE=FREE (the default) or TYPE=LOGICAL(n) for small n.

# ORDER=PVALUE ORDER=ROWS

specifies the order in which the step-down tests to be performed. ORDER=PVALUE is the default, with estimates being declared significant only if all estimates with smaller (unadjusted) **p**-values are significant. If you specify ORDER=ROWS, then significances are evaluated in the order in which they are specified in the syntax.

### **REPORT**

specifies that a report on the step-down adjustment be displayed, including a listing of the sequential subsets (Westfall 1997) and, for <u>ADJUST=SIMULATE</u>, the step-down simulation results.

## TYPE=LOGICAL<(n)> TYPE=FREE

specifies how step-down adjustment are made. If you specify TYPE=LOGICAL, the step-down adjustments are computed by using maximal logically consistent sequential subsets of equality hypotheses (Shaffer 1986, Westfall 1997). Alternatively, for TYPE=FREE, sequential subsets are computed ignoring logical constraints. The TYPE=FREE results are more conservative than those for TYPE=LOGICAL, but they can be much more efficient to produce for many estimates. For example, it is not feasible to take logical constraints between all pairwise comparisons of more than about 10 groups. For this reason, TYPE=FREE is the default.

However, you can reduce the computational complexity of taking logical constraints into account by limiting the depth of the search tree used to compute them, specifying the optional depth parameter as a number n in parentheses after TYPE=LOGICAL. As with TYPE=FREE, results for TYPE=LOGICAL(n) are conservative relative to the true TYPE=LOGICAL results. But even for TYPE=LOGICAL(n) they can be appreciably less conservative than TYPE=FREE, and they are computationally feasible for much larger numbers of estimates. If you do not specify n or if n = -1, the full search tree is used.

TESTVALUE=value-list TESTMEAN=value-list specifies the value under the null hypothesis for testing the estimable functions in the ESTIMATE statement. The rules for specifying the *value-list* are very similar to those for specifying the divisor list in the <u>DIVISOR=</u> option. If no TESTVALUE= is specified, all tests are performed as  $H: \mathbf{L}\boldsymbol{\beta} = 0$ . Missing values in the *value-list* also are translated to zeros. If you specify fewer values than rows in the ESTIMATE statement, the last value in *value-list* is carried forward.

The TESTVALUE= option affects only *p*-values from individual, joint, and multiplicity-adjusted tests. It does not affect confidence intervals.

The TESTVALUE option is not available for the multinomial distribution, and the values are ignored when you perform a sampling-based (Bayesian) analysis.

## UPPER UPPERTAILED

requests that the p-value for the t test be based only on values that are greater than the test statistic. A two-tailed test is the default. An upper-tailed confidence limit is also produced if you specify the  $\underline{CL}$  or  $\underline{ALPHA}$ = option.

Note that for ADJUST=SCHEFFE the one-sided adjusted confidence intervals and one-sided adjusted p-values are the same as the corresponding two-sided statistics, because this adjustment is based on only the right tail of the F distribution.

If you request a joint test with the <u>JOINT</u> option, then a one-sided right-tailed order restriction is applied to all estimable functions, and the corresponding chi-bar-square statistic of Silvapulle and Sen (2004) is computed in addition to the two-sided, standard, *F* or chi-square statistic. See the <u>JOINT</u> option for how to control the computation of the simulation-based chi-bar-square statistic.

Previous Page | Next Page | Top of Page