## Chapter 94
The PROBIT Procedure

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview: PROBIT Procedure</td>
<td>7572</td>
</tr>
<tr>
<td>Getting Started: PROBIT Procedure</td>
<td>7573</td>
</tr>
<tr>
<td>Estimating the Natural Response Threshold Parameter</td>
<td>7573</td>
</tr>
<tr>
<td>Syntax: PROBIT Procedure</td>
<td>7577</td>
</tr>
<tr>
<td>PROC PROBIT Statement</td>
<td>7577</td>
</tr>
<tr>
<td>BY Statement</td>
<td>7582</td>
</tr>
<tr>
<td>CDFPLOT Statement</td>
<td>7583</td>
</tr>
<tr>
<td>CLASS Statement</td>
<td>7591</td>
</tr>
<tr>
<td>EFFECTPLOT Statement</td>
<td>7593</td>
</tr>
<tr>
<td>ESTIMATE Statement</td>
<td>7594</td>
</tr>
<tr>
<td>INSET Statement</td>
<td>7595</td>
</tr>
<tr>
<td>IPPPLOT Statement</td>
<td>7597</td>
</tr>
<tr>
<td>LPREDPLOT Statement</td>
<td>7605</td>
</tr>
<tr>
<td>LSMEANS Statement</td>
<td>7613</td>
</tr>
<tr>
<td>LSMESTIMATE Statement</td>
<td>7614</td>
</tr>
<tr>
<td>MODEL Statement</td>
<td>7615</td>
</tr>
<tr>
<td>OUTPUT Statement</td>
<td>7621</td>
</tr>
<tr>
<td>PREDPPLOT Statement</td>
<td>7622</td>
</tr>
<tr>
<td>SLICE Statement</td>
<td>7631</td>
</tr>
<tr>
<td>STORE Statement</td>
<td>7631</td>
</tr>
<tr>
<td>TEST Statement</td>
<td>7631</td>
</tr>
<tr>
<td>WEIGHT Statement</td>
<td>7631</td>
</tr>
<tr>
<td>Details: PROBIT Procedure</td>
<td>7632</td>
</tr>
<tr>
<td>Missing Values</td>
<td>7632</td>
</tr>
<tr>
<td>Response Level Ordering</td>
<td>7632</td>
</tr>
<tr>
<td>Computational Method</td>
<td>7633</td>
</tr>
<tr>
<td>Distributions</td>
<td>7634</td>
</tr>
<tr>
<td>INEST= SAS-data-set</td>
<td>7635</td>
</tr>
<tr>
<td>Model Specification</td>
<td>7635</td>
</tr>
<tr>
<td>Lack-of-Fit Tests</td>
<td>7636</td>
</tr>
<tr>
<td>Rescaling the Covariance Matrix</td>
<td>7637</td>
</tr>
<tr>
<td>Tolerance Distribution</td>
<td>7637</td>
</tr>
<tr>
<td>Inverse Confidence Limits</td>
<td>7638</td>
</tr>
<tr>
<td>OUTEST= SAS-data-set</td>
<td>7639</td>
</tr>
<tr>
<td>XDATA= SAS-data-set</td>
<td>7639</td>
</tr>
<tr>
<td>Traditional High-Resolution Graphics</td>
<td>7640</td>
</tr>
</tbody>
</table>
Overview: PROBIT Procedure

The PROBIT procedure calculates maximum likelihood estimates of regression parameters and the natural (or threshold) response rate for quantal response data from biological assays or other discrete event data. This includes probit, logit, ordinal logistic, and extreme value (or gompit) regression models.

Probit analysis developed from the need to analyze qualitative (dichotomous or polytomous) dependent variables within the regression framework. Many response variables are binary by nature (yes/no), while others are measured ordinally rather than continuously (degree of severity). Researchers have shown ordinary least squares (OLS) regression to be inadequate when the dependent variable is discrete (Collett 2003; Agresti 2002). Probit or logit analyses are more appropriate in this case.

The PROBIT procedure computes maximum likelihood estimates of the parameters \( \beta \) and \( C \) of the probit equation by using a modified Newton-Raphson algorithm. When the response \( Y \) is binary, with values 0 and 1, the probit equation is

\[
p = \Pr(Y = 0) = C + (1 - C)F(x' \beta)
\]

where

- \( \beta \) is a vector of parameter estimates
- \( F \) is a cumulative distribution function (normal, logistic, or extreme value)
- \( x \) is a vector of explanatory variables
- \( p \) is the probability of a response
- \( C \) is the natural (threshold) response rate

Notice that PROC PROBIT, by default, models the probability of the lower response levels. The choice of the distribution function \( F \) (normal for the probit model, logistic for the logit model, and extreme value or Gompertz for the gompit model) determines the type of analysis. For most problems, there is relatively little difference between the normal and logistic specifications of the model. Both distributions are symmetric about the value zero. The extreme value (or Gompertz) distribution, however, is not symmetric, approaching 0 on the left more slowly than it approaches 1 on the right. You can use the extreme value distribution where such asymmetry is appropriate.
For ordinal response models, the response, $Y$, of an individual or an experimental unit can be restricted to one of a (usually small) number, $k + 1 (k \geq 1)$, of ordinal values, denoted for convenience by $1, \ldots, k, k + 1$. For example, the severity of coronary disease can be classified into three response categories as $1=$no disease, $2=$angina pectoris, and $3=$myocardial infarction. The PROBIT procedure fits a common slopes cumulative model, which is a parallel-lines regression model based on the cumulative probabilities of the response categories rather than on their individual probabilities. The cumulative model has the form

$$
\Pr(Y \leq 1 \mid x) = F(x' \beta) \\
\Pr(Y \leq i \mid x) = F(\alpha_i + x' \beta), \quad 2 \leq i \leq k
$$

where $\alpha_2, \ldots, \alpha_k$ are $k - 1$ intercept parameters. By default, the covariate vector $x$ contains an overall intercept term.

You can set or estimate the natural (threshold) response rate $C$. Estimation of $C$ can begin either from an initial value that you specify or from the rate observed in a control group. By default, the natural response rate is fixed at zero.

An observation in the data set analyzed by the PROBIT procedure might contain the response and explanatory values for one subject. Alternatively, it might provide the number of observed events from a number of subjects at a particular setting of the explanatory variables. In this case, PROC PROBIT models the probability of an event.

The PROBIT procedure uses ODS Graphics to create graphs as part of its output. For general information about ODS Graphics, see Chapter 21, “Statistical Graphics Using ODS.” For specific information about the graphics available in the PROBIT procedure, see the section “ODS Graphics” on page 7643.

---

**Getting Started: PROBIT Procedure**

The following example illustrates how you can use the PROBIT procedure to compute the threshold response rate and regression parameter estimates for quantal response data.

**Estimating the Natural Response Threshold Parameter**

Suppose you want to test the effect of a drug at 12 dosage levels. You randomly divide 180 subjects into 12 groups of 15—one group for each dosage level. You then conduct the experiment and, for each subject, record the presence or absence of a positive response to the drug. You summarize the data by counting the number of subjects responding positively in each dose group. Your data set is as follows:

```plaintext
data study;
  input Dose Respond @@;
  Number = 15;
datalines;
  0 3 1.1 4 1.3 4 2.0 3 2.2 5 2.8 4
  3.7 5 3.9 9 4.4 8 4.8 11 5.9 12 6.8 13;
;```

```plaintext
```
The variable `dose` represents the amount of drug administered. The first group, receiving a dose level of 0, is the control group. The variable `number` represents the number of subjects in each group. All groups are equal in size; hence, `number` has the value 15 for all observations. The variable `respond` represents the number of subjects responding to the associated drug dosage.

You can model the probability of positive response as a function of dosage by using the following statements:

```plaintext
ods graphics on;

proc probit data=study log10 optc plots=(predpplot ippplot);
   model respond/number=dose;
   output out=new p=p_hat;
run;
```

The DATA= option specifies that PROC PROBIT analyzes the SAS data set `study`. The LOG10 option replaces the first continuous independent variable (`dose`) with its common logarithm. The OPTC option estimates the natural response rate. When you use the LOG10 option with the OPTC option, any observations with a dose value less than or equal to zero are used in the estimation as a control group.

The PLOTS= option in the PROC PROBIT statement, together with the ODS GRAPHICS statement, requests two plots for the estimated probability values and dosage levels. For general information about ODS Graphics, see Chapter 21, “Statistical Graphics Using ODS.” For specific information about the graphics available in the PROBIT procedure, see the section “ODS Graphics” on page 7643.

The MODEL statement specifies a proportional response by using the variables `respond` and `number` in `events/trials` syntax. The variable `dose` is the stimulus or explanatory variable.

The OUTPUT statement creates a new data set, `new`, that contains all the variables in the original data set, and a new variable, `p_hat`, that represents the predicted probabilities.

The results from this analysis are displayed in the following figures. 

Figure 94.1 displays background information about the model fit. Included are the name of the input data set, the response variables used, and the number of observations, events, and trials. The last line in Figure 94.1 shows the final value of the log-likelihood function.

Figure 94.2 displays the table of parameter estimates for the model. The parameter $C$, which is the natural response threshold or the proportion of individuals responding at zero dose, is estimated to be 0.2409. Since both the intercept and the slope coefficient have significant $p$-values (0.0020, 0.0010), you can write the model for

$$\Pr(\text{response}) = C + (1 - C)F(x'\beta)$$

as

$$\Pr(\text{response}) = 0.2409 + 0.7591(\Phi(-4.1439 + 6.2308 \times \log_{10}(\text{dose})))$$

where $\Phi$ is the normal cumulative distribution function.

Finally, PROC PROBIT specifies the resulting tolerance distribution by providing the mean $MU$ and scale parameter $SIGMA$ as well as the covariance matrix of the distribution parameters in Figure 94.3.
Figure 94.1 Model Fitting Information for the PROBIT Procedure

The Probit Procedure

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Events Variable</td>
</tr>
<tr>
<td>Trials Variable</td>
</tr>
<tr>
<td>Number of Observations</td>
</tr>
<tr>
<td>Number of Events</td>
</tr>
<tr>
<td>Number of Trials</td>
</tr>
<tr>
<td>Number of Events In Control Group</td>
</tr>
<tr>
<td>Number of Trials In Control Group</td>
</tr>
<tr>
<td>Name of Distribution</td>
</tr>
<tr>
<td>Log Likelihood</td>
</tr>
</tbody>
</table>

Figure 94.2 Model Parameter Estimates for the PROBIT Procedure

Analysis of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-4.1438</td>
<td>1.3415</td>
<td>-6.7731 -1.5146</td>
<td>9.54</td>
<td>0.0020</td>
</tr>
<tr>
<td>Log10(Dose)</td>
<td>1</td>
<td>6.2308</td>
<td>1.8996</td>
<td>2.5076 9.9539</td>
<td>10.76</td>
<td>0.0010</td>
</tr>
<tr>
<td>C_</td>
<td>1</td>
<td>0.2409</td>
<td>0.0523</td>
<td>0.1385 0.3433</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 94.3 Tolerance Distribution Estimates for the PROBIT Procedure

Estimated Covariance Matrix for Tolerance Parameters

<table>
<thead>
<tr>
<th>MU</th>
<th>SIGMA</th>
<th>C_</th>
</tr>
</thead>
<tbody>
<tr>
<td>MU</td>
<td>0.001158</td>
<td>-0.000493</td>
</tr>
<tr>
<td>SIGMA</td>
<td>-0.000493</td>
<td>0.002394</td>
</tr>
<tr>
<td>C_</td>
<td>0.000954</td>
<td>-0.000999</td>
</tr>
</tbody>
</table>

The PLOT=PREDPLOT option creates the plot in Figure 94.4, showing the relationship between dosage level, observed response proportions, and estimated probability values. The dashed lines represent pointwise confidence bands for the fitted probabilities, and a reference line is plotted at the estimated threshold value of 0.24.

The PLOT=IPPPLOT option creates the plot in Figure 94.5, showing the inverse relationship between dosage level and observed response proportions/estimated probability values. The dashed lines represent pointwise fiducial limits for the predicted values of the dose variable, and a reference line is also plotted at the estimated threshold value of 0.24.

The two plot options can be put together with the PLOTS= option, as shown in the PROC PROBIT statement.
**Figure 94.4** Plot of Observed and Fitted Probabilities versus Dose Level

**Figure 94.5** Inverse Predicted Probability Plot with Fiducial Limits
Syntax: PROBIT Procedure

The following statements are available in the PROBIT procedure:

PROC PROBIT <options>;
  BY variables;
  CDFPLOT <VAR=variable> <options>;
  CLASS variables <(options)> . . . <variable <(options)>> </options>;
  ESTIMATE <'label'> estimate-specification <(divisor=n)> . . . <'label'> estimate-specification <(divisor=n)> </options>;
  EFFECTPLOT <plot-type <(plot-definition-options)>> </options>;
  INSET <keyword-list> </options>;
  IPPPLOT <VAR=variable> <options>;
  LPREDPLOT <VAR=variable> <options>;
  LSMEANS <model-effects> </options>;
  LSMESTIMATE model-effect <'label'> values <(divisor=n)> . . . <'label'> values <(divisor=n)> </options>;
  MODEL response <(response _options)> = independents </options>;
  OUTPUT <OUT=SAS-data-set> <options>;
  PREDPPLOT <VAR=variable> <options>;
  SLICE model-effect </options>;
  STORE <OUT=item-store-name </LABEL='label'>; 
  TEST <model-effects> </options>;
  WEIGHT variable;

A MODEL statement is required. Only a single MODEL statement can be used with one invocation of the PROBIT procedure. If multiple MODEL statements are present, only the last one is used. Main effects and higher-order terms can be specified in the MODEL statement, as in the GLM procedure. If a CLASS statement is used, it must precede the MODEL statement.

The CDFPLOT, INSET, IPPPLOT, LPREDPLOT, and PREDPPLOT statements are used to produce graphical output. You can use any appropriate combination of the graphical statements after the MODEL statement.

The ESTIMATE, EFFECTPLOT, LSMEANS, LSMESTIMATE, SLICE, STORE, and TEST statements are common to many procedures. Summary descriptions of functionality and syntax for these statements are also given after the PROC PROBIT statement in alphabetical order, and full documentation about them is available in Chapter 19, “Shared Concepts and Topics.”

PROC PROBIT Statement

PROC PROBIT <options>;

The PROC PROBIT statement invokes the PROBIT procedure. Table 94.1 summarizes the options available in the PROC PROBIT statement.
Table 94.1 PROC PROBIT Statement Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COVOUT</td>
<td>Writes the parameter estimate covariance matrix to the OUTEST= data set</td>
</tr>
<tr>
<td>C=</td>
<td>Controls how the natural response is handled</td>
</tr>
<tr>
<td>DATA=</td>
<td>Specifies the SAS data set to be used</td>
</tr>
<tr>
<td>GOUT=</td>
<td>Specifies a graphics catalog in which to save graphics output</td>
</tr>
<tr>
<td>HPROB=</td>
<td>Specifies a minimum probability level for the Pearson’s chi-square</td>
</tr>
<tr>
<td>INEST=</td>
<td>Specifies an input SAS data set that contains initial estimates</td>
</tr>
<tr>
<td>INVERSECL</td>
<td>Computes confidence limits</td>
</tr>
<tr>
<td>LACKFIT</td>
<td>Performs two goodness-of-fit tests</td>
</tr>
<tr>
<td>LOG</td>
<td>Replaces the first continuous independent variable with its natural logarithm</td>
</tr>
<tr>
<td>LOG10</td>
<td>Replaces the first continuous independent variable with log to the base 10</td>
</tr>
<tr>
<td>NAMELEN=</td>
<td>Specifies the length of effect names to be n characters</td>
</tr>
<tr>
<td>NOPRINT</td>
<td>Suppresses the display of all output including graphics</td>
</tr>
<tr>
<td>OPTC</td>
<td>Controls how the natural response is handled</td>
</tr>
<tr>
<td>ORDER=</td>
<td>Specifies the sort order for the levels of the classification variables</td>
</tr>
<tr>
<td>OUTEST=</td>
<td>Specifies a SAS data set to contain the parameter estimates</td>
</tr>
<tr>
<td>PLOT</td>
<td>PLOTS</td>
</tr>
<tr>
<td>XDATA=</td>
<td>Specifies an input SAS data set that contains values for all the independent variables</td>
</tr>
</tbody>
</table>

You can specify the following **options** in the PROC PROBIT statement.

**COVOUT**

writes the parameter estimate covariance matrix to the OUTEST= data set.

**C=rate**

**OPTC**

controls how the natural response is handled. Specify the OPTC option to request that the natural response rate \( C \) be estimated. Specify the **C=rate** option to set the natural response rate or to provide the initial estimate of the natural response rate. The natural response rate value must be a number between 0 and 1.

- If you specify neither the OPTC nor the **C=** option, a natural response rate of zero is assumed.
- If you specify both the OPTC and the **C=** option, the **C=** option should be a reasonable initial estimate of the natural response rate. For example, you could use the ratio of the number of responses to the number of subjects in a control group.
- If you specify the **C=** option but not the OPTC option, the natural response rate is set to the specified value and not estimated.
- If you specify the OPTC option but not the **C=** option, PROC PROBIT’s action depends on the response variable, as follows:
  - If you specify either the LN or LOG10 option and some subjects have the first independent variable (dose) values less than or equal to zero, these subjects are treated as a control group. The initial estimate of \( C \) is then the ratio of the number of responses to the number of subjects in this group.
– If you do not specify the LN or LOG10 option or if there is no control group, then one of the following occurs:

  · If all responses are greater than zero, the initial estimate of the natural response rate is the minimal response rate (the ratio of the number of responses to the number of subjects in a dose group) across all dose levels.
  
  · If one or more of the responses is zero (making the response rate zero in that dose group), the initial estimate of the natural rate is the reciprocal of twice the largest number of subjects in any dose group in the experiment.

**DATA=SAS-data-set**

specifies the SAS data set to be used by PROC PROBIT. By default, the procedure uses the most recently created SAS data set.

**GOUT=graphics-catalog**

specifies a graphics catalog in which to save graphics output.

**HPROB=p**

specifies a minimum probability level for the Pearson’s chi-square to indicate a good fit. The default value is 0.10. The LACKFIT option must also be specified for this option to have any effect. For Pearson’s goodness-of-fit chi-square values with probability greater than the HPROB= value, the fiducial limits, if requested with the INVERSECL option, are computed by using a critical value of 1.96. For chi-square values with probability less than the value of the HPROB= option, the critical value is a 0.95 two-sided quantile value taken from the $t$ distribution with degrees of freedom equal to $(k - 1) \times m - q$, where $k$ is the number of levels for the response variable, $m$ is the number of different sets of independent variable values, and $q$ is the number of parameters fit in the model. Note that the HPROB= option can also appear in the MODEL statement.

**INEST=SAS-data-set**

specifies an input SAS data set that contains initial estimates for all the parameters in the model. See the section “INEST= SAS-data-set” on page 7635 for a detailed description of the contents of the INEST= data set.

**INVERSECL< (PROB=rates) >**

computes confidence limits for the values of the first continuous independent variable (such as dose) that yield selected response rates. You can optionally specify a list of response rates as rates. The response rates must be between zero and one, and can be a list separated by blanks, commas, or in the form of a DO list.

For example,

```
PROB = .1 TO .9 by .1
PROB = .1 .2 .3 .4
PROB = .01, .25, .75, .9
```

are valid lists of response rates.

If the algorithm fails to converge (this can happen when $C$ is nonzero), missing values are reported for the confidence limits. See the section “Inverse Confidence Limits” on page 7638 for details. Note that the INVERSECL option can also appear in the MODEL statement.
Chapter 94: The PROBIT Procedure

LACKFIT
performs two goodness-of-fit tests (a Pearson’s chi-square test and a log-likelihood ratio chi-square test) for the fitted model.

To compute the test statistics, proper grouping of the observations into subpopulations is needed. You can use the AGGREGATE or AGGREGATE= option for this end. See the entry for the AGGREGATE and AGGREGATE= options under the MODEL statement. If neither AGGREGATE nor AGGREGATE= is specified, PROC PROBIT assumes each observation is from a separate subpopulation and computes the goodness-of-fit test statistics only for the events/trials syntax.

**NOTE:** This test is not appropriate if the data are very sparse, with only a few values at each set of the independent variable values.

If the Pearson’s chi-square test statistic is significant, then the covariance estimates and standard error estimates are adjusted. See the section “Lack-of-Fit Tests” on page 7636 for a description of the tests. Note that the LACKFIT option can also appear in the MODEL statement.

LOG
LN
analyzes the data by replacing the first continuous independent variable with its natural logarithm. This variable is usually the level of some treatment such as dosage. In addition to the usual output given by the INVERSECL option, the estimated dose values and 95% fiducial limits for dose are also displayed. If you specify the OPTC option, any observations with a dose value less than or equal to zero are used in the estimation as a control group. If you do not specify the OPTC option with the LOG or LN option, then any observations with the first continuous independent variable values less than or equal to zero are ignored.

LOG10
specifies an analysis like that of the LN or LOG option, except that the common logarithm (log to the base 10) of the dose value is used rather than the natural logarithm.

NAMELEN=n
specifies the length of effect names in tables and output data sets to be n characters, where n is a value between 20 and 200. The default length is 20 characters.

NOPRINT
suppresses the display of all output including graphics. Note that this option temporarily disables the Output Delivery System (ODS). For more information, see Chapter 20, “Using the Output Delivery System.”

OPTC
controls how the natural response is handled. See the description of the C= option on page 7578 for details.

ORDER=DATA | FORMATTED | FREQ | INTERNAL
specifies the sort order for the levels of the classification variables (which are specified in the CLASS statement).

This option applies to the levels for all classification variables, except when you use the (default) ORDER=FORMATTED option with numeric classification variables that have no explicit format. In that case, the levels of such variables are ordered by their internal value.

The ORDER= option can take the following values:
<table>
<thead>
<tr>
<th>Value of ORDER=</th>
<th>Levels Sorted By</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA</td>
<td>Order of appearance in the input data set</td>
</tr>
<tr>
<td>FORMATTED</td>
<td>External formatted value, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) value</td>
</tr>
<tr>
<td>FREQ</td>
<td>Descending frequency count; levels with the most observations come first in the order</td>
</tr>
<tr>
<td>INTERNAL</td>
<td>Unformatted value</td>
</tr>
</tbody>
</table>

By default, ORDER=FORMATTED. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.

This order also applies to the levels of the response variable. Response level ordering is important because PROC PROBIT always models the probability of response levels at the beginning of the ordering. See the section “Response Level Ordering” on page 7632 for further details.

For more information about sort order, see the chapter on the SORT procedure in the *Base SAS Procedures Guide* and the discussion of BY-group processing in *SAS Language Reference: Concepts*.

**OUTEST=SAS-data-set**
specifies a SAS data set to contain the parameter estimates and, if the COVOUT option is specified, their estimated covariances. If you omit this option, the output data set is not created. The contents of the data set are described in the section “OUTEST= SAS-data-set” on page 7639.

**PLOT | PLOTS <=plot-request >**
specifies options that control details of the plots created by ODS Graphics. These plots are related to a dose variable, which is identified as the first single continuous independent variable in the MODEL statement. If there are interaction terms with this variable in the model, the PROBIT procedure will not produce any plot.

You can specify more than one plot request within the parentheses after PLOTS=. For a single plot request, you can omit the parentheses.

ODS Graphics must be enabled before plots can be requested. For example:

```sas
proc probit plots=predplot;
   model r/n = dose;
run;
```

For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 607 in Chapter 21, “Statistical Graphics Using ODS.”

The following *plot-requests* are available.

**ALL**
creates all appropriate plots.
CDFPLOT< (LEVEL=(character-list)) >
requests the plot of predicted cumulative distribution function (CDF) of the multinomial response variable as a function of a single continuous independent variable (dose variable). This single continuous independent variable must be the first single continuous independent variable listed in the MODEL statement. You can request this plot only with a multinomial model.

The LEVEL= suboption specifies the levels of the multinomial response variable for which the CDF curves are requested. There are \( k - 1 \) curves for a \( k \)-level multinomial response variable (for the highest level, it is the constant line 1). You can specify any of them to be plotted by the LEVEL= suboption.

IPPPLOT
requests the inverse plot of the predicted probability against the first single continuous variable (dose variable) in the MODEL statement for the binomial model. You can request this plot only with a binomial model. The confidence limits for the predicted values of the dose variable are the computed fiducial limits, not the inverse of the confidence limits of the predicted probabilities. See the section “Inverse Confidence Limits” on page 7638 for more details.

LPREDPLOT< (LEVEL=(character-list)) >
requests the plot of the linear predictor \( x'b \) against the first single continuous variable (dose variable) in the MODEL statement for either the binomial model or the multinomial model. The confidence limits for the predicted values are available only for the binomial model.

For the multinomial model, you can use the LEVEL= suboption to specify the levels for which the linear predictor lines are plotted.

NONE
suppresses all plots.

PREDPPLOT< (LEVEL=(character-list)) >
requests the plot of the predicted probability against the first single continuous variable (dose variable) in the MODEL statement for both the binomial model and the multinomial model. Confidence limits are available only for the binomial model.

For the multinomial model, you can use the LEVEL= suboption to specify the levels for which the linear predictor lines are plotted.

XDATA=SAS-data-set
specifies an input SAS data set that contains values for all the independent variables in the MODEL statement and variables in the CLASS statement. If there are covariates specified in a MODEL statement, you specify fixed values for the effects in the MODEL statement by the XDATA= data set when predicted values and/or fiducial limits for a single continuous variable (dose variable) are required. These specified values for the effects in the MODEL statement are also used for generating plots. See the section “XDATA= SAS-data-set” on page 7639 for a detailed description of the contents of the XDATA= data set.

BY Statement

BY variables ;
You can specify a BY statement with PROC PROBIT to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement for the PROBIT procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

For more information about BY-group processing, see the discussion in SAS Language Reference: Concepts. For more information about the DATASETS procedure, see the discussion in the Base SAS Procedures Guide.

CDFPLOT Statement

CDFPLOT <VAR=variable> <options> ;

The CDFPLOT statement plots the predicted cumulative distribution function (CDF) of the multinomial response variable as a function of a single continuous independent variable (dose variable). You can use this statement only after a multinomial model statement.

VAR=variable

specifies a single continuous variable (dose variable) in the independent variable list of the MODEL statement. If a VAR= variable is not specified, the first single continuous variable in the independent variable list of the MODEL statement is used. If such a variable does not exist in the independent variable list of the MODEL statement, an error is reported.

The predicted cumulative distribution function is defined as

\[ \hat{F}_j(x) = C + (1 - C)F(\hat{a}_j + x^T\hat{b}) \]

where \( j = 1, \ldots, k \) are the indexes of the \( k \) levels of the multinomial response variable, \( F \) is the CDF of the distribution used to model the cumulative probabilities, \( \hat{b} \) is the vector of estimated parameters, \( x \) is the covariate vector, \( \hat{a}_j \) are estimated ordinal intercepts with \( \hat{a}_1 = 0 \), and \( C \) is the threshold parameter, either known or estimated from the model. Let \( x_1 \) be the covariate corresponding to the dose variable and \( x_{-1} \) be the vector of the rest of the covariates. Let the corresponding estimated parameters be \( \hat{b}_1 \) and \( \hat{b}_{-1} \). Then

\[ \hat{F}_j(x) = C + (1 - C)F(\hat{a}_j + x_1\hat{b}_1 + x_{-1}^T\hat{b}_{-1}) \]

To plot \( \hat{F}_j \) as a function of \( x_1 \), \( x_{-1} \) must be specified. You can use the XDATA= option to provide the values of \( x_{-1} \) (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow the rules:
• If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
• If the effect is a single classification variable, the highest level of the variable is used.

options

specify the levels of the multinomial response variable for which the CDF curves are requested, and add features to the plot. There are \( k - 1 \) curves for a \( k \)-level multinomial response variable (for the highest level, it is the constant line 1). You can specify any of them to be plotted by the LEVEL= option in the CDFPLOT statement. See the LEVEL= option for how to specify the levels.

An attached box on the right side of the plot is used to label these curves with the names of their levels. You can specify the color of this box by using the CLABBOX= option.

You can use options in the CDFPLOT statement to do the following:

• superimpose specification limits
• specify the levels for which the CDF curves are requested
• specify graphical enhancements (such as color or text height)

Summary of Options

Table 94.2 through Table 94.8 summarize the options available in the CDFPLOT statement. The “Dictionary of Options” on page 7586 describes each option in detail.

CDF Options

Table 94.2  Options for CDFPLOT

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEVEL=(character-list)</td>
<td>Specifies the names of the levels for which the CDF curves are requested</td>
</tr>
<tr>
<td>NOTHRESH</td>
<td>Suppresses the threshold line</td>
</tr>
<tr>
<td>THRESHLABPOS=value</td>
<td>Specifies the position for the label of the threshold line</td>
</tr>
</tbody>
</table>

General Options

Table 94.3  Color Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAXIS=color</td>
<td>Specifies color for axis</td>
</tr>
<tr>
<td>CFIT=color</td>
<td>Specifies color for fitted curves</td>
</tr>
<tr>
<td>CFRAME=color</td>
<td>Specifies color for frame</td>
</tr>
<tr>
<td>CGRID=color</td>
<td>Specifies color for grid lines</td>
</tr>
<tr>
<td>CHREF=color</td>
<td>Specifies color for HREF= lines</td>
</tr>
<tr>
<td>CLABBOX=color</td>
<td>Specifies color for label box</td>
</tr>
<tr>
<td>CTEXT=color</td>
<td>Specifies color for text</td>
</tr>
<tr>
<td>CVREF=color</td>
<td>Specifies color for VREF= lines</td>
</tr>
</tbody>
</table>
### Table 94.4  Options to Enhance Plots Produced on Graphics Devices

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANNOTATE=</td>
<td>Specifies an Annotate data set</td>
</tr>
<tr>
<td>INBORDER</td>
<td>Requests a border around plot</td>
</tr>
<tr>
<td>LFIT=linetype</td>
<td>Specifies line style for fitted curves</td>
</tr>
<tr>
<td>LGRID=linetype</td>
<td>Specifies line style for grid lines</td>
</tr>
<tr>
<td>NOFRAME</td>
<td>Suppresses the frame around plotting areas</td>
</tr>
<tr>
<td>NOGRID</td>
<td>Suppresses grid lines</td>
</tr>
<tr>
<td>NOFIT</td>
<td>Suppresses CDF curves</td>
</tr>
<tr>
<td>NOHLABEL</td>
<td>Suppresses horizontal labels</td>
</tr>
<tr>
<td>NOHTICK</td>
<td>Suppresses horizontal ticks</td>
</tr>
<tr>
<td>NOVTICK</td>
<td>Suppresses vertical ticks</td>
</tr>
<tr>
<td>TURNVLABELS</td>
<td>Vertically strings out characters in vertical labels</td>
</tr>
<tr>
<td>WFIT=n</td>
<td>Specifies thickness for fitted curves</td>
</tr>
<tr>
<td>WGRID=n</td>
<td>Specifies thickness for grids</td>
</tr>
<tr>
<td>WREFL=n</td>
<td>Specifies thickness for reference lines</td>
</tr>
</tbody>
</table>

### Table 94.5  Axis Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAXIS=value1 to value2</td>
<td>Specifies tick mark values for horizontal axis</td>
</tr>
<tr>
<td>HOFFSET=value</td>
<td>Specifies offset for horizontal axis</td>
</tr>
<tr>
<td>HLOWER=value</td>
<td>Specifies lower limit on horizontal axis scale</td>
</tr>
<tr>
<td>HUPPER=value</td>
<td>Specifies upper limit on horizontal axis scale</td>
</tr>
<tr>
<td>NHTICK=n</td>
<td>Specifies number of ticks for horizontal axis</td>
</tr>
<tr>
<td>NVTICK=n</td>
<td>Specifies number of ticks for vertical axis</td>
</tr>
<tr>
<td>VAXIS=value1 to value2</td>
<td>Specifies tick mark values for vertical axis</td>
</tr>
<tr>
<td>VAXISLABEL='label'</td>
<td>Specifies label for vertical axis</td>
</tr>
<tr>
<td>VOFFSET=value</td>
<td>Specifies offset for vertical axis</td>
</tr>
<tr>
<td>VLOWER=value</td>
<td>Specifies lower limit on vertical axis scale</td>
</tr>
<tr>
<td>VUPPER=value</td>
<td>Specifies upper limit on vertical axis scale</td>
</tr>
<tr>
<td>WAXIS=n</td>
<td>Specifies thickness for axis</td>
</tr>
</tbody>
</table>

### Table 94.6  Graphics Catalog Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIPTION='string'</td>
<td>Specifies description for graphics catalog member</td>
</tr>
<tr>
<td>NAME='string'</td>
<td>Specifies name for plot in graphics catalog</td>
</tr>
</tbody>
</table>
Table 94.7   Options for Text Enhancement

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FONT=font</td>
<td>Specifies software font for text</td>
</tr>
<tr>
<td>HEIGHT=value</td>
<td>Specifies height of text outside framed areas</td>
</tr>
<tr>
<td>INHEIGHT=value</td>
<td>Specifies height of text inside framed areas</td>
</tr>
</tbody>
</table>

Table 94.8   Options for Reference Lines

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HREF&lt; (INTERSECT) &gt;=value</td>
<td>Requests horizontal reference line</td>
</tr>
<tr>
<td>HREFLABELS=('label1',...,'labeln')</td>
<td>Specifies labels for HREF= lines</td>
</tr>
<tr>
<td>HREFLABPOS=n</td>
<td>Specifies vertical position of labels for HREF= lines</td>
</tr>
<tr>
<td>LHREF=linetype</td>
<td>Specifies line style for HREF= lines</td>
</tr>
<tr>
<td>LVREF=linetype</td>
<td>Specifies line style for VREF= lines</td>
</tr>
<tr>
<td>VREF&lt; (INTERSECT) &gt;=value</td>
<td>Requests vertical reference line</td>
</tr>
<tr>
<td>VREFLABELS=('label1',...,'labeln')</td>
<td>Specifies labels for VREF= lines</td>
</tr>
<tr>
<td>VREFLABPOS=n</td>
<td>Specifies horizontal position of labels for VREF= lines</td>
</tr>
</tbody>
</table>

Dictionary of Options

The following entries provide detailed descriptions of the options in the CDFPLOT statement.

**ANNOTATE=SAS-data-set**

specifies an Annotate data set, as described in SAS/GRAPH: Reference, that enables you to add features to the CDF plot. The ANNOTATE= data set you specify in the CDFPLOT statement is used for all plots created by the statement.

**CAXIS=color**

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

**CFIT=color**

specifies the color for the fitted CDF curves. The default is the first color in the device color list.

**CFRAME=color**

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

CGRID=color
   specifies the color for grid lines. The default is the first color in the device color list.

CLABBOX=color
   specifies the color for the area enclosed by the label box for CDF curves. This area is not shaded by default.

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

CTEXT=color
   specifies the color for tick mark values and axis labels. The default is the color specified for the CTEXT= option in the most recent GOPTIONS statement.

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

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

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

HAXIS=value1 to value2 < by value3>
   specifies tick mark values for the horizontal axis; value1, value2, and value3 must be numeric, and value1 must be less than value2. The lower tick mark is value1. Tick marks are drawn at increments of value3. The last tick mark is the greatest value that does not exceed value2. If value3 is omitted, a value of 1 is used.

   Examples of HAXIS= lists follow:

   haxis = 0 to 10
   haxis = 2 to 10 by 2
   haxis = 0 to 200 by 10

HEIGHT=value
   specifies the height of text used outside framed areas. The default value is 3.846 (in percentage).

HLOWER=value
   specifies the lower limit on the horizontal axis scale. The HLOWER= option specifies value as the lower horizontal axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the HAXIS= option is used.
**OFFSET=value**
specifies offset for horizontal axis. The default value is 1.

**HUPPER=value**
specifies value as the upper horizontal axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the HAXIS= option is used.

**HREF < (INTERSECT) > =value-list**
requests reference lines perpendicular to the horizontal axis. If (INTERSECT) is specified, a second reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the intersecting vertical axis reference line is labeled with the vertical axis value. See also the CHREF=, HREFLABELS=, and LHREF= options.

**HREFLABELS=’label1’, . . . , ’labeln’**
**HREFLABEL=’label1’, . . . , ’labeln’**
**HREFLAB=’label1’, . . . , ’labeln’**
specifies labels for the lines requested by the HREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

**HREFLABPOS=n**
specifies the vertical position of labels for HREF= lines. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Top</td>
</tr>
<tr>
<td>2</td>
<td>Staggered from top</td>
</tr>
<tr>
<td>3</td>
<td>Bottom</td>
</tr>
<tr>
<td>4</td>
<td>Staggered from bottom</td>
</tr>
<tr>
<td>5</td>
<td>Alternating from top</td>
</tr>
<tr>
<td>6</td>
<td>Alternating from bottom</td>
</tr>
</tbody>
</table>

**INBORDER**
requests a border around CDF plots.

**INHEIGHT=value**
Specifies height of text inside framed areas.

**LEVEL=(character-list)**
**ORDINAL=(character-list)**
specifies the names of the levels for which CDF curves are requested. Names should be quoted and separated by space. If there is no correct name provided, no CDF curve is plotted.

**LFIT=linetype**
specifies a line style for fitted curves. By default, fitted curves are drawn by connecting solid lines (linetype = 1).
LGRID=linetype
specifies a line style for all grid lines. linetype is between 1 and 46. The default is 35.

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

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

NAME='string'
specifies a name for the plot, up to eight characters, that appears in the PROC GREPLAY master menu. The default is 'PROBIT'.

NHTICK=n
Specifies number of ticks for horizontal axis.

NVTICK=n
Specifies number of ticks for vertical axis.

NOFIT
suppresses the fitted CDF curves.

NOFRAME
suppresses the frame around plotting areas.

NOGRID
suppresses grid lines.

NOHLABEL
suppresses horizontal labels.

NOHTICK
suppresses horizontal tick marks.

NOTHRESH
suppresses the threshold line.

NOVLABEL
suppresses vertical labels.

TURNV LABELS
vertically strings out characters in vertical labels.

NOVTICK
suppresses vertical tick marks.
THRESHLABPOS=n
specifies the horizontal position of labels for the threshold line. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left</td>
</tr>
<tr>
<td>2</td>
<td>Right</td>
</tr>
</tbody>
</table>

VAXIS=value1 to value2 < by value3>
specifies tick mark values for the vertical axis; value1, value2, and value3 must be numeric, and value1 must be less than value2. The lower tick mark is value1. Tick marks are drawn at increments of value3. The last tick mark is the greatest value that does not exceed value2. This method of specification of tick marks is not valid for logarithmic axes. If value3 is omitted, a value of 1 is used.

Examples of VAXIS= lists follow:

```
  vaxis = 0 to 10
  vaxis = 0 to 2 by .1
```

VAXISLABEL='string'
specifies a label for the vertical axis.

VLOWER=value
specifies the lower limit on the vertical axis scale. The VLOWER= option specifies value as the lower vertical axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the VAXIS= option is used.

VOFFSET=value
specifies the offset for the vertical axis.

VREF=value-list
requests reference lines perpendicular to the vertical axis. If (INTERSECT) is specified, a second reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting horizontal axis reference line is labeled with the horizontal axis value. See also the CVREF=, LVREF=, and VREFLABELS= options.

VREFLABELS='label1', . . . , 'labeln'
VREFLABEL='label1', . . . , 'labeln'
VREFLAB='label1', . . . , 'labeln'
specifies labels for the lines requested by the VREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

VREFLABPOS=n
specifies the horizontal position of labels for VREF= lines. The following table shows valid values for n and the corresponding label placements.
CLASS Statement

<table>
<thead>
<tr>
<th>( n )</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left</td>
</tr>
<tr>
<td>2</td>
<td>Right</td>
</tr>
</tbody>
</table>

**VUPPER=value**
specifies the upper limit on the vertical axis scale. The VUPPER= option specifies value as the upper vertical axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the VAXIS= option is used.

**WAXIS=n**
specifies line thickness for axes and frame. The default value is 1.

**WFIT=n**
specifies line thickness for fitted curves. The default value is 1.

**WGRID=n**
specifies line thickness for grids. The default value is 1.

**WREFL=n**
specifies line thickness for reference lines. The default value is 1.

---

**CLASS Statement**

```
CLASS variable < (options) > . . . < variable < (options) > > < / global-options > ;
```

The CLASS statement names the classification variables to be used as explanatory variables in the analysis. The CLASS statement must precede the MODEL statement. Most options can be specified either as individual variable options or as global-options. You can specify options for each variable by enclosing the options in parentheses after the variable name. You can also specify global-options for the CLASS statement by placing them after a slash (/). Global-options are applied to all the variables specified in the CLASS statement. If you specify more than one CLASS statement, the global-options specified in any one CLASS statement apply to all CLASS statements. However, individual CLASS variable options override the global-options. You can specify the following values for either an option or a global-option:

**CPREFIX=n**
specifies that, at most, the first \( n \) characters of a CLASS variable name be used in creating names for the corresponding design variables. The default is \( 32 - \min(32, \max(2, f)) \), where \( f \) is the formatted length of the CLASS variable.

**DESCENDING**
**DESC**
reverses the sort order of the classification variable. If both the DESCENDING and ORDER= options are specified, PROC PROBIT orders the categories according to the ORDER= option and then reverses that order.
LPREFIX=n
specifies that, at most, the first \( n \) characters of a CLASS variable label be used in creating labels for the corresponding design variables. The default is \( 256 - \min(256, \max(2, f)) \), where \( f \) is the formatted length of the CLASS variable.

MISSING
treats missing values (., ._, .A, ., . . . , .Z for numeric variables and blanks for character variables) as valid values for the CLASS variable.

ORDER=DATA | FORMATTED | FREQ | INTERNAL
specifies the sort order for the levels of classification variables. This ordering determines which parameters in the model correspond to each level in the data, so the ORDER= option can be useful when you use the CONTRAST statement. By default, ORDER=FORMATTED. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent. When ORDER=FORMATTED is in effect for numeric variables for which you have supplied no explicit format, the levels are ordered by their internal values.

The following table shows how PROC PROBIT interprets values of the ORDER= option.

<table>
<thead>
<tr>
<th>Value of ORDER=</th>
<th>Levels Sorted By</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA</td>
<td>Order of appearance in the input data set</td>
</tr>
<tr>
<td>FORMATTED</td>
<td>External formatted values, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) values</td>
</tr>
<tr>
<td>FREQ</td>
<td>Descending frequency count; levels with more observations come earlier in the order</td>
</tr>
<tr>
<td>INTERNAL</td>
<td>Unformatted value</td>
</tr>
</tbody>
</table>

For more information about sort order, see the chapter on the SORT procedure in the Base SAS Procedures Guide and the discussion of BY-group processing in SAS Language Reference: Concepts.

PARAM=GLM
specifies the parameterization method for the classification variable or variables. The default is GLM parameterization. Only GLM parameterization is currently available. The REF= option in the CLASS statement indirectly determines the reference level for a singular GLM parameterization through the order of levels.

REF='level' | keyword
For PARAM=GLM, the REF= option specifies a level of the classification variable to be put at the end of the list of levels. This level thus corresponds to the reference level in the usual interpretation of the linear estimates with a singular parameterization.

For an individual variable REF= option (but not for a global REF= option), you can specify the level of the variable to use as the reference level. Specify the formatted value of the variable if a format is assigned. For a global or individual variable REF= option, you can use one of the following keywords.

FIRST designates the first ordered level as reference.
LAST designates the last ordered level as reference.

By default, REF=LAST.
TRUNCATE< =n>
specifies the length $n$ of CLASS variable values to use in determining CLASS variable levels. The
default is to use the full formatted length of the CLASS variable. If you specify TRUNCATE without
the length $n$, the first 16 characters of the formatted values are used. When formatted values are longer
than 16 characters, you can use this option to revert to the levels as determined in releases before SAS
9. The TRUNCATE option is available only as a global option.

**EFFECTPLOT Statement**

**EFFECTPLOT** < plot-type < (plot-definition-options) > > < / options > ;

The EFFECTPLOT statement produces a display of the fitted model and provides options for changing and
enhancing the displays. Table 94.9 describes the available *plot-types* and their *plot-definition-options*.

**Table 94.9  Plot-Types and Plot-Definition-Options**

<table>
<thead>
<tr>
<th>Plot-Type and Description</th>
<th>Plot-Definition-Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOX</td>
<td>PLOTBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>X= CLASS variable or effect</td>
</tr>
<tr>
<td>CONTOUR</td>
<td>PLOTBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>X= continuous variable</td>
</tr>
<tr>
<td></td>
<td>Y= continuous variable</td>
</tr>
<tr>
<td>FIT</td>
<td>PLOTBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>X= continuous variable</td>
</tr>
<tr>
<td>INTERACTION</td>
<td>PLOTBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>SLICEBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>X= CLASS variable or effect</td>
</tr>
<tr>
<td>MOSAIC</td>
<td>PLOTBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>X= CLASS effects</td>
</tr>
<tr>
<td>SLICEFIT</td>
<td>PLOTBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>SLICEBY= variable or CLASS effect</td>
</tr>
<tr>
<td></td>
<td>X= continuous variable</td>
</tr>
</tbody>
</table>

For full details about the syntax and options of the EFFECTPLOT statement, see the section “EFFECTPLOT
Statement” on page 414 in Chapter 19, “Shared Concepts and Topics.”
**ESTIMATE Statement**

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

The ESTIMATE statement provides a mechanism for obtaining custom hypothesis tests. Estimates are formed as linear estimable functions of the form $L\hat{\beta}$. You can perform hypothesis tests for the estimable functions, construct confidence limits, and obtain specific nonlinear transformations.

Table 94.10 summarizes the options available in the ESTIMATE statement.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction and Computation of Estimable Functions</strong></td>
<td></td>
</tr>
<tr>
<td>DIVISOR=</td>
<td>Specifies a list of values to divide the coefficients</td>
</tr>
<tr>
<td>NOFILL</td>
<td>Suppresses the automatic fill-in of coefficients for higher-order effects</td>
</tr>
<tr>
<td>SINGULAR=</td>
<td>Tunes the estimability checking difference</td>
</tr>
<tr>
<td><strong>Degrees of Freedom and $p$-values</strong></td>
<td></td>
</tr>
<tr>
<td>ADJUST=</td>
<td>Determines the method for multiple comparison adjustment of estimates</td>
</tr>
<tr>
<td>ALPHA=$\alpha$</td>
<td>Determines the confidence level $(1 - \alpha)$</td>
</tr>
<tr>
<td>LOWER</td>
<td>Performs one-sided, lower-tailed inference</td>
</tr>
<tr>
<td>STEPDOWN</td>
<td>Adjusts multiplicity-corrected $p$-values further in a step-down fashion</td>
</tr>
<tr>
<td>TESTVALUE=</td>
<td>Specifies values under the null hypothesis for tests</td>
</tr>
<tr>
<td>UPPER</td>
<td>Performs one-sided, upper-tailed inference</td>
</tr>
<tr>
<td><strong>Statistical Output</strong></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td>Constructs confidence limits</td>
</tr>
<tr>
<td>CORR</td>
<td>Displays the correlation matrix of estimates</td>
</tr>
<tr>
<td>COV</td>
<td>Displays the covariance matrix of estimates</td>
</tr>
<tr>
<td>E</td>
<td>Prints the $L$ matrix</td>
</tr>
<tr>
<td>JOINT</td>
<td>Produces a joint $F$ or chi-square test for the estimable functions</td>
</tr>
<tr>
<td>PLOTS=</td>
<td>Requests ODS statistical graphics if the analysis is sampling-based</td>
</tr>
<tr>
<td>SEED=</td>
<td>Specifies the seed for computations that depend on random numbers</td>
</tr>
<tr>
<td><strong>Generalized Linear Modeling</strong></td>
<td></td>
</tr>
<tr>
<td>CATEGORY=</td>
<td>Specifies how to construct estimable functions with multinomial data</td>
</tr>
<tr>
<td>EXP</td>
<td>Exponentiates and displays estimates</td>
</tr>
<tr>
<td>ILINK</td>
<td>Computes and displays estimates and standard errors on the inverse linked scale</td>
</tr>
</tbody>
</table>
For details about the syntax of the ESTIMATE statement, see the section “ESTIMATE Statement” on page 442 in Chapter 19, “Shared Concepts and Topics.”

**INSET Statement**

\[
\text{INSET < keyword-list> </ options> ;}
\]

The box or table of summary information produced on plots made with the CDFPLOT, IPPPLOT, LPREDPLOT, or PREDPPLOT statement is called an *inset*. You can use the INSET statement to customize both the information that is printed in the inset box and the appearance of the inset box. To supply the information that is displayed in the inset box, you specify *keywords* corresponding to the information you want shown. For example, the following statements produce a predicted probability plot with the number of trials, the number of events, the name of the distribution, and the estimated optimum natural threshold in the inset.

```plaintext
proc probit data=epidemic;
  model r/n = dose;
  predpplot;
  inset nobs ntrials nevents dist optc;
run;
```

By default, inset entries are identified with appropriate labels. However, you can provide a customized label by specifying the *keyword* for that entry followed by the equal sign (=) and the label in quotes.

For example, the following INSET statement produces an inset containing the number of observations and the name of the distribution, labeled “Sample Size” and “Distribution” in the inset.

```plaintext
inset nobs='Sample Size' dist='Distribution';
```

If you specify a *keyword* that does not apply to the plot you are creating, then the *keyword* is ignored.

The *options* control the appearance of the box.

If you specify more than one INSET statement, only the first one is used.

**Keywords Used in the INSET Statement**

Table 94.11 and Table 94.12 list *keywords* available in the INSET statement to display summary statistics, distribution parameters, and distribution fitting information.

**Table 94.11** Summary Statistics

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOBS</td>
<td>Number of observations</td>
</tr>
<tr>
<td>NTRIALS</td>
<td>Number of trials</td>
</tr>
<tr>
<td>NEVENTS</td>
<td>Number of events</td>
</tr>
<tr>
<td>C</td>
<td>User-input threshold</td>
</tr>
<tr>
<td>OPTC</td>
<td>Estimated natural threshold</td>
</tr>
<tr>
<td>NRESPLEV</td>
<td>Number of levels of the response variable</td>
</tr>
</tbody>
</table>
Table 94.12 General Information

| CONFIDENCE | Confidence coefficient for all confidence intervals |
| DIST | Name of the distribution |

Options Used in the INSET Statement

Table 94.13 and Table 94.14 list the options available in the INSET statement.

Table 94.13 Color and Pattern Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFILL=</td>
<td>Specifies color for filling box</td>
</tr>
<tr>
<td>CFILLH=</td>
<td>Specifies color for filling box header</td>
</tr>
<tr>
<td>CFRA =</td>
<td>Specifies color for frame</td>
</tr>
<tr>
<td>CHEADER=</td>
<td>Specifies color for text in header</td>
</tr>
<tr>
<td>CTEXT=</td>
<td>Specifies color for text</td>
</tr>
</tbody>
</table>

Table 94.14 General Appearance Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FONT=</td>
<td>Specifies software font for text</td>
</tr>
<tr>
<td>HEIGHT=</td>
<td>Specifies height of text</td>
</tr>
<tr>
<td>HEADER=</td>
<td>Specifies text for header or box title</td>
</tr>
<tr>
<td>NOFRAME</td>
<td>Omits frame around box</td>
</tr>
<tr>
<td>POS=</td>
<td>Determines the position of the inset. The value can be a compass point (N, NE, E, SE, S, SW, W, NW) or a pair of coordinates (x, y) enclosed in parentheses. The coordinates can be specified in axis percentage units or axis data units.</td>
</tr>
<tr>
<td>REFPOINT=</td>
<td>Specifies the reference point for an inset that is positioned by a pair of coordinates with the POS= option. You use the REFPOINT= option in conjunction with the POS= coordinates. The REFPOINT= option specifies which corner of the inset frame you have specified with coordinates (x, y), and it can take the value of BR (bottom right), BL (bottom left), TR (top right), or TL (top left). The default is REFPOINT=BL. If the inset position is specified as a compass point, then the REFPOINT= option is ignored.</td>
</tr>
</tbody>
</table>
IPPPLOT Statement

**IPPPLOT** < variable > < options > ;

The IPPPLOT statement plots the inverse of the predicted probability (IPP) against a single continuous variable (dose variable) in the MODEL statement for the binomial model. You can only use this statement after a binomial model statement. The confidence limits for the predicted values of the dose variable are the computed fiducial limits, not the inverse of the confidence limits of the predicted probabilities. See the section “Inverse Confidence Limits” on page 7638 for more details.

**VAR=** variable

specifies a single continuous variable (dose variable) in the independent variable list of the MODEL statement. If a VAR= variable is not specified, the first single continuous variable in the independent variable list of the MODEL statement is used. If such a variable does not exist in the independent variable list of the MODEL statement, an error is reported.

For the binomial model, the response variable is a probability. An estimate of the dose level $\hat{x}_1$ needed for a response of $p$ is given by

$$\hat{x}_1 = (F^{-1}(p) - x'_{-1} \hat{b}_{-1})/\hat{b}_1$$

where $F$ is the cumulative distribution function used to model the probability, $x_{-1}$ is the vector of the rest of the covariates, $\hat{b}_{-1}$ is the vector of the estimated parameters corresponding to $x_{-1}$, and $\hat{b}_1$ is the estimated parameter for the dose variable of interest.

To plot $\hat{x}_1$ as a function of $p$, $x_{-1}$ must be specified. You can use the XDATA= option to provide the values of $x_{-1}$ (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow the rules:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.

**options**

add features to the plot.

You can use **options** in the IPPPLOT statement to do the following:

- superimpose specification limits
- suppress or add the observed data points on the plot
- suppress or add the fiducial limits on the plot
- specify graphical enhancements (such as color or text height)

**Summary of Options**

Table 94.15 through Table 94.21 summarize the **options** available in the IPPPLOT statement. The “Dictionary of Options” on page 7600 describes each option in detail.
**IPP Options**

**Table 94.15** Plot Layout Options for IPPPLOT

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOCONF</td>
<td>Suppresses fiducial limits</td>
</tr>
<tr>
<td>NODATA</td>
<td>Suppresses observed data points on the plot</td>
</tr>
<tr>
<td>NOTHRESH</td>
<td>Suppresses the threshold line</td>
</tr>
<tr>
<td>THRESHLABPOS=value</td>
<td>Specifies the position for the label of the threshold line</td>
</tr>
</tbody>
</table>

**General Options**

**Table 94.16** Color Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAXIS=color</td>
<td>Specifies color for axis</td>
</tr>
<tr>
<td>CFIT=color</td>
<td>Specifies color for fitted curves</td>
</tr>
<tr>
<td>CFRAME=color</td>
<td>Specifies color for frame</td>
</tr>
<tr>
<td>CGRID=color</td>
<td>Specifies color for grid lines</td>
</tr>
<tr>
<td>CHREF=color</td>
<td>Specifies color for HREF= lines</td>
</tr>
<tr>
<td>CTEXT=color</td>
<td>Specifies color for text</td>
</tr>
<tr>
<td>CVREF=color</td>
<td>Specifies color for VREF= lines</td>
</tr>
</tbody>
</table>

**Table 94.17** Options to Enhance Plots Produced on Graphics Devices

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANNOTATE=SAS-data-set</td>
<td>Specifies an Annotate data set</td>
</tr>
<tr>
<td>INBORDER</td>
<td>Requests a border around plot</td>
</tr>
<tr>
<td>LFIT=linetype</td>
<td>Specifies line style for fitted curves and confidence limits</td>
</tr>
<tr>
<td>LGRID=linetype</td>
<td>Specifies line style for grid lines</td>
</tr>
<tr>
<td>NOFRAME</td>
<td>Suppresses the frame around plotting areas</td>
</tr>
<tr>
<td>NOGRID</td>
<td>Suppresses grid lines</td>
</tr>
<tr>
<td>NOFIT</td>
<td>Suppresses fitted curves</td>
</tr>
<tr>
<td>NOHLABEL</td>
<td>Suppresses horizontal labels</td>
</tr>
<tr>
<td>NOHTICK</td>
<td>Suppresses horizontal ticks</td>
</tr>
<tr>
<td>NOVTICK</td>
<td>Suppresses vertical ticks</td>
</tr>
<tr>
<td>TURNVLABELS</td>
<td>Vertically strings out characters in vertical labels</td>
</tr>
<tr>
<td>WFIT=n</td>
<td>Specifies thickness for fitted curves</td>
</tr>
<tr>
<td>WGRID=n</td>
<td>Specifies thickness for grids</td>
</tr>
<tr>
<td>WREFL=n</td>
<td>Specifies thickness for reference lines</td>
</tr>
</tbody>
</table>
Table 94.18  Axis Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAXIS=\textit{value1} to \textit{value2} \texttt{&lt; by \textit{value3}&gt;}</td>
<td>Specifies tick mark values for horizontal axis</td>
</tr>
<tr>
<td>HOFFSET=\textit{value}</td>
<td>Specifies offset for horizontal axis</td>
</tr>
<tr>
<td>HLOWER=\textit{value}</td>
<td>Specifies lower limit on horizontal axis scale</td>
</tr>
<tr>
<td>HUPPER=\textit{value}</td>
<td>Specifies upper limit on horizontal axis scale</td>
</tr>
<tr>
<td>NHTICK=\textit{n}</td>
<td>Specifies number of ticks for horizontal axis</td>
</tr>
<tr>
<td>NVTICK=\textit{n}</td>
<td>Specifies number of ticks for vertical axis</td>
</tr>
<tr>
<td>VAXIS=\textit{value1} to \textit{value2} \texttt{&lt; by \textit{value3}&gt;}</td>
<td>Specifies tick mark values for vertical axis</td>
</tr>
<tr>
<td>VAXISLABEL=\textit{'label'}</td>
<td>Specifies label for vertical axis</td>
</tr>
<tr>
<td>VOFFSET=\textit{value}</td>
<td>Specifies offset for vertical axis</td>
</tr>
<tr>
<td>VLOWER=\textit{value}</td>
<td>Specifies lower limit on vertical axis scale</td>
</tr>
<tr>
<td>VUPPER=\textit{value}</td>
<td>Specifies upper limit on vertical axis scale</td>
</tr>
<tr>
<td>WAXIS=\textit{n}</td>
<td>Specifies thickness for axis</td>
</tr>
</tbody>
</table>

Table 94.19  Options for Reference Lines

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HREF&lt; (INTERSECT) &gt; \texttt{=} \textit{value-list}</td>
<td>Requests horizontal reference line</td>
</tr>
<tr>
<td>HREFLABELS= \texttt{('label1'\ldots'labeln')}</td>
<td>Specifies labels for HREF= lines</td>
</tr>
<tr>
<td>HREFLABPOS=\textit{n}</td>
<td>Specifies vertical position of labels for HREF= lines</td>
</tr>
<tr>
<td>LHREF=\textit{linetype}</td>
<td>Specifies line style for HREF= lines</td>
</tr>
<tr>
<td>LVREF=\textit{linetype}</td>
<td>Specifies line style for VREF= lines</td>
</tr>
<tr>
<td>VREF&lt; (INTERSECT) &gt; \texttt{=} \textit{value-list}</td>
<td>Requests vertical reference line</td>
</tr>
<tr>
<td>VREFLABELS= \texttt{('label1'\ldots'labeln')}</td>
<td>Specifies labels for VREF= lines</td>
</tr>
<tr>
<td>VREFLABPOS=\textit{n}</td>
<td>Specifies horizontal position of labels for VREF= lines</td>
</tr>
</tbody>
</table>

Table 94.20  Graphics Catalog Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIPTION=\texttt{'}\textit{string}'</td>
<td>Specifies description for graphics catalog member</td>
</tr>
<tr>
<td>NAME=\texttt{'}\textit{string}'</td>
<td>Specifies name for plot in graphics catalog</td>
</tr>
</tbody>
</table>
Chapter 94: The PROBIT Procedure

**Table 94.21 Options for Text Enhancement**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FONT=font</td>
<td>Specifies software font for text</td>
</tr>
<tr>
<td>HEIGHT=value</td>
<td>Specifies height of text used outside framed areas</td>
</tr>
<tr>
<td>INHEIGHT=value</td>
<td>Specifies height of text inside framed areas</td>
</tr>
</tbody>
</table>

**Dictionary of Options**

The following entries provide detailed descriptions of the *options* in the IPPPLOT statement.

- **ANNOTATE=SAS-data-set**
  - Specifies an Annotate data set, as described in *SAS/GRAPH: Reference*, that enables you to add features to the IPP plot. The ANNOTATE= data set you specify in the IPPPLOT statement is used for all plots created by the statement.

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

- **CFIT=color**
  - Specifies the color for the fitted IPP curves. The default is the first color in the device color list.

- **CFRAME=color**
  - Specifies the color for the area enclosed by the axes and frame. This area is not shaded by default.

- **CGRID=color**
  - Specifies the color for grid lines. The default is the first color in the device color list.

- **CHREF=color**
  - Specifies the color for lines requested by the HREF= option. The default is the first color in the device color list.

- **CTEXT=color**
  - Specifies the color for tick mark values and axis labels. The default is the color specified for the CTEXT= option in the most recent GOPTIONS statement.

- **CV=color**
  - Specifies the color for lines requested by the VREF= option. The default is the first color in the device color list.
DESCRIPTION = 'string'

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

FONT = font

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

HAXIS = value1 to value2 < by value3 >

Specifies tick mark values for the horizontal axis; value1, value2, and value3 must be numeric, and value1 must be less than value2. The lower tick mark is value1. Tick marks are drawn at increments of value3. The last tick mark is the greatest value that does not exceed value2. If value3 is omitted, a value of 1 is used.

Examples of HAXIS= lists follow:

\[
\begin{align*}
  \text{haxis} &= 0 \text{ to } 10 \\
  \text{haxis} &= 2 \text{ to } 10 \text{ by } 2 \\
  \text{haxis} &= 0 \text{ to } 200 \text{ by } 10
\end{align*}
\]

HEIGHT = value

Specifies the height of text used outside framed areas. The default value is 3.846 (in percentage).

HLOWER = value

Specifies the lower limit on the horizontal axis scale. The HLOWER= option specifies value as the lower horizontal axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the HAXIS= option is used.

HOFFSET = value

Specifies offset for horizontal axis. The default value is 1.

HUPPER = value

Specifies value as the upper horizontal axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the HAXIS= option is used.

HREF < (INTERSECT) > = value-list

Requests reference lines perpendicular to the horizontal axis. If (INTERSECT) is specified, a second reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the intersecting vertical axis reference line is labeled with the vertical axis value. See also the CHREF=, HREFLABELS=, and LHREF= options.

HREFLABELS = 'label1', . . . , 'labeln'

HREFLABEL = 'label1', . . . , 'labeln'

HREFLAB = 'label1', . . . , 'labeln'

Specifies labels for the lines requested by the HREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.
HREFLABPOS=n
 specifies the vertical position of labels for HREF= lines. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Top</td>
</tr>
<tr>
<td>2</td>
<td>Staggered from top</td>
</tr>
<tr>
<td>3</td>
<td>Bottom</td>
</tr>
<tr>
<td>4</td>
<td>Staggered from bottom</td>
</tr>
<tr>
<td>5</td>
<td>Alternating from top</td>
</tr>
<tr>
<td>6</td>
<td>Alternating from bottom</td>
</tr>
</tbody>
</table>

INBORDER
 requests a border around IPP plots.

INHEIGHT=value
 Specifies height of text inside framed areas.

LFIT=linetype
 specifies a line style for fitted curves and confidence limits. By default, fitted curves are drawn by connecting solid lines (linetype = 1) and confidence limits are drawn by connecting dashed lines (linetype = 3).

LGRID=linetype
 specifies a line style for all grid lines. The value for linetype must be between 1 and 46. The default is 35.

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

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

NAME='string'
 specifies a name for the plot, up to eight characters, that appears in the PROC GREPLAY master menu. The default is 'PROBIT'.

NHTICK=n
 Specifies number of ticks for horizontal axis.

NVTICK=n
 Specifies number of ticks for vertical axis.

NOCONF
 suppresses fiducial limits from the plot.
NODATA
suppresses observed data points from the plot.

NOFIT
suppresses the fitted IPP curves.

NOFRAME
suppresses the frame around plotting areas.

NOGRID
suppresses grid lines.

NOHLABEL
suppresses horizontal labels.

NOHTICK
suppresses horizontal tick marks.

NOTHRESH
suppresses the threshold line.

NOVLABEL
suppresses vertical labels.

TURNVLABELS
vertically strings out characters in vertical labels.

NOVTICK
suppresses vertical tick marks.

THRESHLABPOS=n
specifies the vertical position of labels for the threshold line. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Top</td>
</tr>
<tr>
<td>2</td>
<td>Bottom</td>
</tr>
</tbody>
</table>

VAXIS=\text{value1 to value2 < by value3>}
specifies tick mark values for the vertical axis; value1, value2, and value3 must be numeric, and value1 must be less than value2. The lower tick mark is value1. Tick marks are drawn at increments of value3. The last tick mark is the greatest value that does not exceed value2. This method of specification of tick marks is not valid for logarithmic axes. If value3 is omitted, a value of 1 is used.

Examples of VAXIS= lists follow:

\begin{align*}
vaxis & = 0 \text{ to } 10 \\
vaxis & = 0 \text{ to } 2 \text{ by } .1
\end{align*}
VAXISLABEL=’string’
specifies a label for the vertical axis.

VLOWER=value
specifies the lower limit on the vertical axis scale. The VLOWER= option specifies value as the lower vertical axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the VAXIS= option is used.

VOFFSET=value
specifies the offset for the vertical axis.

VREF=value-list
requests reference lines perpendicular to the vertical axis. If (INTERSECT) is specified, a second reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting horizontal axis reference line is labeled with the horizontal axis value. See also the CVREF=, LVREF=, and VREFLABELS= options.

VREFLABELS=’label1’, . . . , ’labeln’

VREFLABEL=’label1’, . . . , ’labeln’

VREFLAB=’label1’, . . . , ’labeln’
specifies labels for the lines requested by the VREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

VREFLABPOS=n
specifies the horizontal position of labels for VREF= lines. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left</td>
</tr>
<tr>
<td>2</td>
<td>Right</td>
</tr>
</tbody>
</table>

VUPPER=value
specifies the upper limit on the vertical axis scale. The VUPPER= option specifies value as the upper vertical axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the VAXIS= option is used.

WAXIS=n
specifies line thickness for axes and frame. The default value is 1.

WFIT=n
specifies line thickness for fitted curves. The default value is 1.

WGRID=n
specifies line thickness for grids. The default value is 1.

WREFL=n
specifies line thickness for reference lines. The default value is 1.
**LPREDPLOT Statement**

```
LPREDPLOT < VAR=variable > < options > ;
```

The LPREDPLOT statement plots the linear predictor (LPRED) $x'\mathbf{b}$ against a single continuous variable (dose variable) in the MODEL statement for either the binomial model or the multinomial model. The confidence limits for the predicted values are available only for the binomial model.

**VAR= variable**

specifies a single continuous variable (dose variable) in the independent variable list of the MODEL statement for which the linear predictor plot is plotted. If a VAR= variable is not specified, the first single continuous variable in the independent variable list of the MODEL statement is used. If such a variable does not exist in the independent variable list of the MODEL statement, an error is reported.

Let $x_1$ be the covariate of the dose variable, $x_{-1}$ be the vector of the rest of the covariates, $\hat{\mathbf{b}}_{-1}$ be the vector of estimated parameters corresponding to $x_{-1}$, and $\hat{b}_1$ be the estimated parameter for the dose variable of interest.

To plot $\hat{x}'\mathbf{b}$ as a function of $x_1$, $x_{-1}$ must be specified. You can use the XDATA= option to provide the values of $x_{-1}$ (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow these rules:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.

**options**

add features to the plot.

For the multinomial model, you can use the LEVEL= option to specify the levels for which the linear predictor lines are plotted. The lines are labeled by the names of their levels in the middle.

You can use `options` in the LPREDPLOT statement to do the following:

- superimpose specification limits
- suppress or add the observed data points on the plot for the binomial model
- suppress or add the confidence limits for the binomial model
- specify the levels for which the linear predictor lines are requested for the multinomial model
- specify graphical enhancements (such as color or text height)

**Summary of Options**

Table 94.22 through Table 94.28 list all `options` by function. The “Dictionary of Options” on page 7608 describes each option in detail.
**LPRED Options**

**Table 94.22**  Plot Layout Options for LPREDPLOT

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEVEL= (character-list)</td>
<td>Specifies the names of the levels for which the linear predictor lines are requested (only for the multinomial model)</td>
</tr>
<tr>
<td>NOCONF</td>
<td>Suppresses fiducial limits (only for the binomial model)</td>
</tr>
<tr>
<td>NODATA</td>
<td>Suppresses observed data points on the plot (only for the binomial model)</td>
</tr>
<tr>
<td>NOTHRESH</td>
<td>Suppresses the threshold line</td>
</tr>
<tr>
<td>THRESHLABPOS=value</td>
<td>Specifies the position for the label of the threshold line</td>
</tr>
</tbody>
</table>

**General Options**

**Table 94.23**  Color Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAXIS= color</td>
<td>Specifies color for axis</td>
</tr>
<tr>
<td>CFIT= color</td>
<td>Specifies color for fitted curves</td>
</tr>
<tr>
<td>CFRAME= color</td>
<td>Specifies color for frame</td>
</tr>
<tr>
<td>CGRID= color</td>
<td>Specifies color for grid lines</td>
</tr>
<tr>
<td>CHREF= color</td>
<td>Specifies color for HREF= lines</td>
</tr>
<tr>
<td>CTEXT= color</td>
<td>Specifies color for text</td>
</tr>
<tr>
<td>CVREF= color</td>
<td>Specifies color for VREF= lines</td>
</tr>
</tbody>
</table>

**Table 94.24**  Options to Enhance Plots Produced on Graphics Devices

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANNOTATE= SAS-data-set</td>
<td>Specifies an Annotate data set</td>
</tr>
<tr>
<td>INBORDER</td>
<td>Requests a border around plot</td>
</tr>
<tr>
<td>LFIT=linetype</td>
<td>Specifies line style for fitted curves and confidence limits</td>
</tr>
<tr>
<td>LGRID=linetype</td>
<td>Specifies line style for grid lines</td>
</tr>
<tr>
<td>NOFRAME</td>
<td>Suppresses the frame around plotting areas</td>
</tr>
<tr>
<td>NOGRID</td>
<td>Suppresses grid lines</td>
</tr>
<tr>
<td>NOFIT</td>
<td>Suppresses fitted curves</td>
</tr>
<tr>
<td>NOHLABEL</td>
<td>Suppresses horizontal labels</td>
</tr>
<tr>
<td>NOHTICK</td>
<td>Suppresses horizontal ticks</td>
</tr>
<tr>
<td>NOVTICK</td>
<td>Suppresses vertical ticks</td>
</tr>
<tr>
<td>TURNVLBLABELS</td>
<td>Vertically strings out characters in vertical labels</td>
</tr>
<tr>
<td>WFIT=n</td>
<td>Specifies thickness for fitted curves</td>
</tr>
<tr>
<td>WGRID=n</td>
<td>Specifies thickness for grids</td>
</tr>
<tr>
<td>WREFL=n</td>
<td>Specifies thickness for reference lines</td>
</tr>
</tbody>
</table>
### Table 94.25  Axis Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAXIS=value1 to value2 &lt; by value3 &gt;</td>
<td>Specifies tick mark values for horizontal axis</td>
</tr>
<tr>
<td>HOFFSET=value</td>
<td>Specifies offset for horizontal axis</td>
</tr>
<tr>
<td>HLOWER=value</td>
<td>Specifies lower limit on horizontal axis scale</td>
</tr>
<tr>
<td>HUPPER=value</td>
<td>Specifies upper limit on horizontal axis scale</td>
</tr>
<tr>
<td>NHTICK=n</td>
<td>Specifies number of ticks for horizontal axis</td>
</tr>
<tr>
<td>NVTICK=n</td>
<td>Specifies number of ticks for vertical axis</td>
</tr>
<tr>
<td>VAXIS=value1 to value2 &lt; by value3 &gt;</td>
<td>Specifies tick mark values for vertical axis</td>
</tr>
<tr>
<td>VAXISLABEL='label'</td>
<td>Specifies label for vertical axis</td>
</tr>
<tr>
<td>VOFFSET=value</td>
<td>Specifies offset for vertical axis</td>
</tr>
<tr>
<td>VLOWER=value</td>
<td>Specifies lower limit on vertical axis scale</td>
</tr>
<tr>
<td>VUPPER=value</td>
<td>Specifies upper limit on vertical axis scale</td>
</tr>
<tr>
<td>WAXIS=n</td>
<td>Specifies thickness for axis</td>
</tr>
</tbody>
</table>

### Table 94.26  Graphics Catalog Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIPTION='string'</td>
<td>Specifies description for graphics catalog member</td>
</tr>
<tr>
<td>NAME='string'</td>
<td>Specifies name for plot in graphics catalog</td>
</tr>
</tbody>
</table>

### Table 94.27  Options for Text Enhancement

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FONT=font</td>
<td>Specifies software font for text</td>
</tr>
<tr>
<td>HEIGHT=value</td>
<td>Specifies height of text used outside framed areas</td>
</tr>
<tr>
<td>INHEIGHT=value</td>
<td>Specifies height of text inside framed areas</td>
</tr>
</tbody>
</table>
Table 94.28 Options for Reference Lines

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HREF&lt;INTERSECT&gt; value</td>
<td>Requests horizontal reference line</td>
</tr>
<tr>
<td>HREFLABELS label</td>
<td>Specifies labels for HREF= lines</td>
</tr>
<tr>
<td>HREFLABPOS n</td>
<td>Specifies vertical position of labels for HREF= lines</td>
</tr>
<tr>
<td>LHREF linetype</td>
<td>Specifies line style for HREF= lines</td>
</tr>
<tr>
<td>LVREF linetype</td>
<td>Specifies line style for VREF= lines</td>
</tr>
<tr>
<td>VREF&lt;INTERSECT&gt; value</td>
<td>Requests vertical reference line</td>
</tr>
<tr>
<td>VREFLABELS label</td>
<td>Specifies labels for VREF= lines</td>
</tr>
<tr>
<td>VREFLABPOS n</td>
<td>Specifies horizontal position of labels for VREF= lines</td>
</tr>
</tbody>
</table>

Dictionary of Options

The following entries provide detailed descriptions of the options in the LPREDPLOT statement.

ANNOTATE=SAS-data-set

ANNO=SAS-data-set

specifies an Annotate data set, as described in SAS/GRAPH: Reference, that enables you to add features to the LPRED plot. The ANNOTATE= data set you specify in the LPREDPLOT statement is used for all plots created by the statement.

CAXIS=color

CAXES=color

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

CFIT=color

specifies the color for the fitted LPRED lines. The default is the first color in the device color list.

CFRAME=color

CFR=color

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

CGRID=color

specifies the color for grid lines. The default is the first color in the device color list.

CHREF=color

CH=color

specifies the color for lines requested by the HREF= option. The default is the first color in the device color list.
CTEXT=color
  specifies the color for tick mark values and axis labels. The default is the color specified for the
  CTEXT= option in the most recent GOPTIONS statement.

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

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

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

HAXIS=value1 to value2 < by value3>
  specifies tick mark values for the horizontal axis; value1, value2, and value3 must be numeric, and
  value1 must be less than value2. The lower tick mark is value1. Tick marks are drawn at increments
  of value3. The last tick mark is the greatest value that does not exceed value2. If value3 is omitted, a
  value of 1 is used.

Examples of HAXIS= lists follow:

  haxis = 0 to 10
  haxis = 2 to 10 by 2
  haxis = 0 to 200 by 10

HEIGHT=value
  specifies the height of text used outside framed areas. The default value is 3.846 (in percentage).

HLOWER=value
  specifies the lower limit on the horizontal axis scale. The HLOWER= option specifies value as the
  lower horizontal axis tick mark. The tick mark interval and the upper axis limit are determined
  automatically. This option has no effect if the HAXIS= option is used.

HOFFSET=value
  specifies offset for horizontal axis. The default value is 1.

HUPPER=value
  specifies value as the upper horizontal axis tick mark. The tick mark interval and the lower axis limit
  are determined automatically. This option has no effect if the HAXIS= option is used.

HREF < (INTERSECT) > =value-list
  requests reference lines perpendicular to the horizontal axis. If (INTERSECT) is specified, a second
  reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same
  point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the
  intersecting vertical axis reference line is labeled with the vertical axis value. See also the CHREF=,
  HREFLABELS=, and LHREF= options.
**HREFLABELS=**'label1', . . . , 'labeln'

**HREFLABELS=**'label1', . . . , 'labeln'

**HREFLABEL=**'label1', . . . , 'labeln'

specifies labels for the lines requested by the HREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

**HREFLABPOS=**n

specifies the vertical position of labels for HREF= lines. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Top</td>
</tr>
<tr>
<td>2</td>
<td>Staggered from top</td>
</tr>
<tr>
<td>3</td>
<td>Bottom</td>
</tr>
<tr>
<td>4</td>
<td>Staggered from bottom</td>
</tr>
<tr>
<td>5</td>
<td>Alternating from top</td>
</tr>
<tr>
<td>6</td>
<td>Alternating from bottom</td>
</tr>
</tbody>
</table>

**INBORDER**

requests a border around LPRED plots.

**INHEIGHT=**value

Specifies height of text inside framed areas.

**LEVEL=**(character-list)

**ORDINAL=**(character-list)

specifies the names of the levels for which linear predictor lines are requested. Names should be quoted and separated by space. If there is no correct name provided, no LPRED line is plotted.

**LFIT=**linetype

specifies a line style for fitted curves and confidence limits. By default, fitted curves are drawn by connecting solid lines (linetype = 1) and confidence limits are drawn by connecting dashed lines (linetype = 3).

**LGRID=**linetype

specifies a line style for all grid lines. The value for linetype is between 1 and 46. The default is 35.

**LHREF=**linetype

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

**LVREF=**linetype

specifies the line type for lines requested by the VREF= option. The default is 2, which produces a dashed line.
NAME='string'
specifies a name for the plot, up to eight characters, that appears in the PROC GREPLAY master menu. The default is 'PROBIT'.

NHTICK=n
specifies number of ticks for horizontal axis.

NVTICK=n
specifies number of ticks for vertical axis.

NOCONF
suppresses confidence limits from the plot. This works only for the binomial model. Confidence limits are not plotted for the multinomial model.

NODATA
suppresses observed data points from the plot. This works only for the binomial model. Data points are not plotted for the multinomial model.

NOFIT
suppresses the fitted LPRED lines.

NOFRAME
suppresses the frame around plotting areas.

NOGRID
suppresses grid lines.

NOHLABEL
suppresses horizontal labels.

NOHTICK
suppresses horizontal tick marks.

NOTHRESH
suppresses the threshold line.

NOVLABEL
suppresses vertical labels.

NOVTICK
suppresses vertical tick marks.

TURNVLABELS
vertically strings out characters in vertical labels.

THRESHLABPOS=n
specifies the horizontal position of labels for the threshold line. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left</td>
</tr>
<tr>
<td>2</td>
<td>Right</td>
</tr>
</tbody>
</table>
**VAXIS=**<br>specifies tick mark values for the vertical axis; `value1`, `value2`, and `value3` must be numeric, and `value1` must be less than `value2`. The lower tick mark is `value1`. Tick marks are drawn at increments of `value3`. The last tick mark is the greatest value that does not exceed `value2`. This method of specification of tick marks is not valid for logarithmic axes. If `value3` is omitted, a value of 1 is used.<br><br>Examples of VAXIS= lists follow:<br><br>`vaxis = 0 to 10`
`vaxis = 0 to 2 by .1`

**VAXISLABEL=**`'string'`
specifies a label for the vertical axis.

**VLOWER=**`value`
specifies the lower limit on the vertical axis scale. The VLOWER= option specifies `value` as the lower vertical axis tick mark. The tick mark interval and the upper axis limit are determined automatically. This option has no effect if the VAXIS= option is used.

**VOFFSET=**`value`
specifies the offset for the vertical axis.

**VREF=**`value-list`
requests reference lines perpendicular to the vertical axis. If (INTERSECT) is specified, a second reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting horizontal axis reference line is labeled with the horizontal axis value. See also the CVREF=, LVREF=, and VREFLABELS= options.

**VREFLABELS=**`'label1', . . . , 'labeln'`
**VREFLABEL=**`'label1', . . . , 'labeln'`
**VREFLAB=**`'label1', . . . , 'labeln'`
specifies labels for the lines requested by the VREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

**VREFLABPOS=**`n`
specifies the horizontal position of labels for VREF= lines. The following table shows valid values for `n` and the corresponding label placements.

<table>
<thead>
<tr>
<th><code>n</code></th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left</td>
</tr>
<tr>
<td>2</td>
<td>Right</td>
</tr>
</tbody>
</table>

**VUPPER=**`number`
specifies the upper limit on the vertical axis scale. The VUPPER= option specifies `number` as the upper vertical axis tick mark. The tick mark interval and the lower axis limit are determined automatically. This option has no effect if the VAXIS= option is used.
The LSMEANS statement computes and compares least squares means (LS-means) of fixed effects. LS-means are predicted population margins—that is, they estimate the marginal means over a balanced population. In a sense, LS-means are to unbalanced designs as class and subclass arithmetic means are to balanced designs.

Table 94.29 summarizes the options available in the LSMEANS statement.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction and Computation of LS-Means</strong></td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>Modifies the covariate value in computing LS-means</td>
</tr>
<tr>
<td>BYLEVEL</td>
<td>Computes separate margins</td>
</tr>
<tr>
<td>DIFF</td>
<td>Requests differences of LS-means</td>
</tr>
<tr>
<td>OM=</td>
<td>Specifies the weighting scheme for LS-means computation as determined by the input data set</td>
</tr>
<tr>
<td>SINGULAR=</td>
<td>Tunes estimability checking</td>
</tr>
<tr>
<td><strong>Degrees of Freedom and p-values</strong></td>
<td></td>
</tr>
<tr>
<td>ADJUST=</td>
<td>Determines the method for multiple-comparison adjustment of LS-means differences</td>
</tr>
<tr>
<td>ALPHA=α</td>
<td>Determines the confidence level ( 1 - \alpha )</td>
</tr>
<tr>
<td>STEPDOWN</td>
<td>Adjusts multiple-comparison ( p )-values further in a step-down fashion</td>
</tr>
<tr>
<td><strong>Statistical Output</strong></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td>Constructs confidence limits for means and mean differences</td>
</tr>
<tr>
<td>CORR</td>
<td>Displays the correlation matrix of LS-means</td>
</tr>
<tr>
<td>COV</td>
<td>Displays the covariance matrix of LS-means</td>
</tr>
<tr>
<td>E</td>
<td>Prints the ( L ) matrix</td>
</tr>
<tr>
<td>LINES</td>
<td>Produces a “Lines” display for pairwise LS-means differences</td>
</tr>
<tr>
<td>MEANS</td>
<td>Prints the LS-means</td>
</tr>
<tr>
<td>PLOTS=</td>
<td>Requests graphs of means and mean comparisons</td>
</tr>
</tbody>
</table>
Table 94.29  continued

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEED=</td>
<td>Specifies the seed for computations that depend on random numbers</td>
</tr>
</tbody>
</table>

Generalized Linear Modeling

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>Exponentiates and displays estimates of LS-means or LS-means differences</td>
</tr>
<tr>
<td>ILINK</td>
<td>Computes and displays estimates and standard errors of LS-means (but not differences) on the inverse linked scale</td>
</tr>
<tr>
<td>ODDSRATIO</td>
<td>Reports (simple) differences of least squares means in terms of odds ratios if permitted by the link function</td>
</tr>
</tbody>
</table>

Table 94.30 summarizes the options available in the LSMESTIMATE statement.

Table 94.30  LSMESTIMATE Statement Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction and Computation of LS-Means</td>
<td></td>
</tr>
<tr>
<td>AT</td>
<td>Modifies covariate values in computing LS-means</td>
</tr>
<tr>
<td>BYLEVEL</td>
<td>Computes separate margins</td>
</tr>
<tr>
<td>DIVISOR=</td>
<td>Specifies a list of values to divide the coefficients</td>
</tr>
<tr>
<td>OM=</td>
<td>Specifies the weighting scheme for LS-means computation as determined by a data set</td>
</tr>
<tr>
<td>SINGULAR=</td>
<td>Tunes estimability checking</td>
</tr>
<tr>
<td>Degrees of Freedom and $p$-values</td>
<td></td>
</tr>
<tr>
<td>ADJUST=</td>
<td>Determines the method for multiple-comparison adjustment of LS-means differences</td>
</tr>
<tr>
<td>ALPHA=$\alpha$</td>
<td>Determines the confidence level $(1 - \alpha)$</td>
</tr>
<tr>
<td>LOWER</td>
<td>Performs one-sided, lower-tailed inference</td>
</tr>
<tr>
<td>STEPDOWN</td>
<td>Adjusts multiple-comparison $p$-values further in a step-down fashion</td>
</tr>
</tbody>
</table>

For details about the syntax of the LSMEANS statement, see the section “LSMEANS Statement” on page 458 in Chapter 19, “Shared Concepts and Topics.”

LSMEANS Statement

```
LSMEANS model-effect < 'label' > values < divisor=n >
   < , ... < 'label' > values < divisor=n > >
   </ options > ;
```

The LSMESTIMATE statement provides a mechanism for obtaining custom hypothesis tests among least squares means.

Table 94.30 summarizes the options available in the LSMESTIMATE statement.
Table 94.30  continued

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TESTVALUE=</td>
<td>Specifies values under the null hypothesis for tests</td>
</tr>
<tr>
<td>UPPER</td>
<td>Performs one-sided, upper-tailed inference</td>
</tr>
</tbody>
</table>

**Statistical Output**

- **CL**: Constructs confidence limits for means and mean differences
- **CORR**: Displays the correlation matrix of LS-means
- **COV**: Displays the covariance matrix of LS-means
- **E**: Prints the L matrix
- **ELSM**: Prints the K matrix
- **JOINT**: Produces a joint F or chi-square test for the LS-means and LS-means differences
- **PLOTS=**: Requests graphs of means and mean comparisons
- **SEED=**: Specifies the seed for computations that depend on random numbers

**Generalized Linear Modeling**

- **CATEGORY=**: Specifies how to construct estimable functions with multinomial data
- **EXP**: Exponentiates and displays LS-means estimates
- **ILINK**: Computes and displays estimates and standard errors of LS-means (but not differences) on the inverse linked scale

For details about the syntax of the LSMESTIMATE statement, see the section “LSMESTIMATE Statement” on page 477 in Chapter 19, “Shared Concepts and Topics.”

**MODEL Statement**

```sas
< label: > MODEL response< (response_options) > = effects < / options > ;
< label: > MODEL events/trials = effects < / options > ;
```

The MODEL statement names the variables used as the response and the independent variables. Additionally, you can specify the distribution used to model the response, as well as other options. Only a single MODEL statement can be used with one invocation of the PROBIT procedure. If multiple MODEL statements are present, only the last is used. Main effects and interaction terms can be specified in the MODEL statement, as in the GLM procedure.

The optional `label`, which must be a valid SAS name, is used to label output from the matching MODEL statement.

The `response` can be a single variable whose value indicates the level of the observed response. The `response` variable can be numeric or character. For example, the `response` might be a variable called `Symptoms` that takes the value ‘None,’ ‘Mild,’ or ‘Severe.’ Variable options specific to the response variable can be specified in parentheses immediately after the response variable. Identifying the event level for binomial responses and
ordering of response levels for multinomial responses is critical in these models. You can use the response variable options to do this.

Alternatively, the response can be specified as a pair of variable names separated by a slash (/). The value of the first variable, events, is the number of positive responses (or events). The value of the second variable, trials, is the number of trials. Both variables must be numeric and nonnegative, and the ratio of the first variable value to the second variable value must be between 0 and 1, inclusive. For example, the variables might be hits, a variable containing the number of hits for a baseball player, and AtBats, a variable containing the number of times at bat. A model for hitting proportion (batting average) as a function of age could be specified as

```
model hits/AtBats=age;
```

The effects following the equal sign are the covariates in the model. Higher-order effects, such as interactions and nested terms, are allowed in the list, as in the GLM procedure. Variable names and combinations of variable names representing higher-order terms are allowed to appear in this list. Classification variables can be used as effects, and indicator variables are generated for the class levels. If you do not specify any covariates following the equal sign, an intercept-only model is fit.

Table 94.31 summarizes the options available in the MODEL statement.

**Table 94.31  MODEL Statement Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGGREGATE</td>
<td>Specifies the subpopulations</td>
</tr>
<tr>
<td>ALPHA=</td>
<td>Sets the significance level</td>
</tr>
<tr>
<td>CONVERGE=</td>
<td>Specifies the convergence criterion</td>
</tr>
<tr>
<td>CORRB</td>
<td>Displays the estimated correlation matrix</td>
</tr>
<tr>
<td>COVB</td>
<td>Displays the estimated covariance matrix</td>
</tr>
<tr>
<td>DESCENDING</td>
<td>Reverses the order of the response categories</td>
</tr>
<tr>
<td>DISTRIBUTION=</td>
<td>Specifies the cumulative distribution function</td>
</tr>
<tr>
<td>EVENT=</td>
<td>Specifies the event category for the binary response model</td>
</tr>
<tr>
<td>HPROB=</td>
<td>Specifies a minimum probability level</td>
</tr>
<tr>
<td>INITIAL=</td>
<td>Sets initial values for the parameters</td>
</tr>
<tr>
<td>INTERCEPT=</td>
<td>Initializes the intercept parameter</td>
</tr>
<tr>
<td>INVERSECL</td>
<td>Computes confidence limits</td>
</tr>
<tr>
<td>IPRINT</td>
<td>Displays the iteration history, the final evaluation of the gradient, and the second derivative matrix</td>
</tr>
<tr>
<td>LACKFIT</td>
<td>Performs two goodness-of-fit tests</td>
</tr>
<tr>
<td>MAXITER=</td>
<td>Specifies the maximum number of iterations</td>
</tr>
<tr>
<td>NOINT</td>
<td>Fits a model with no intercept parameter</td>
</tr>
<tr>
<td>ORDER=</td>
<td>Specifies the sort order for the levels of the response variable</td>
</tr>
<tr>
<td>REFERENCE=</td>
<td>Specifies the reference category for the binary response model</td>
</tr>
<tr>
<td>SCALE=</td>
<td>Specifies the method for estimating the dispersion parameter</td>
</tr>
<tr>
<td>SINGULAR=</td>
<td>Specifies the singularity criterion</td>
</tr>
</tbody>
</table>
Response Variable Options

DESCENDING | DESC
reverses the order of the response categories. If both the DESCENDING and ORDER= options are specified, PROC PROBIT orders the response categories according to the ORDER= option and then reverses that order. See the section “Response Level Ordering” on page 7632 for more detail.

EVENT='category' | keyword
specifies the event category for the binary response model. PROC PROBIT models the probability of the event category. The EVENT= option has no effect when there are more than two response categories. You can specify the value (formatted if a format is applied) of the event category in quotation marks, or you can specify one of the following keywords.

FIRST
designates the first ordered category as the event.

LAST
designates the last ordered category as the event.

By default, EVENT=FIRST.

One of the most common sets of response levels is \{0,1\}, where 1 represents the event for which the probability is to be modeled. Consider the example where Y takes the values 1 and 0 for event and nonevent, respectively, and Exposure is the explanatory variable. To specify the value 1 as the event category, use the following MODEL statement:

```
model Y(event='1') = Exposure;
```

ORDER=DATA | FORMATTED | FREQ | INTERNAL
specifies the sort order for the levels of the response variable. The following table displays the available ORDER= options:

<table>
<thead>
<tr>
<th>ORDER=</th>
<th>Levels Sorted By</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA</td>
<td>order of appearance in the input data set</td>
</tr>
<tr>
<td>FORMATTED</td>
<td>external formatted value, except for numeric variables with no explicit format, which are sorted by their unformatted (internal) value</td>
</tr>
<tr>
<td>FREQ</td>
<td>descending frequency count; levels with the most observations come first in the order</td>
</tr>
<tr>
<td>INTERNAL</td>
<td>unformatted value</td>
</tr>
</tbody>
</table>

By default, ORDER=FORMATTED. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent. When ORDER=FORMATTED is in effect for numeric variables for which you have supplied no explicit format, the levels are ordered by their internal values.

For more information about sort order, see the chapter on the SORT procedure in the Base SAS Procedures Guide and the discussion of BY-group processing in SAS Language Reference: Concepts.
REFERENCE='category' | keyword

specifies the reference category for the binary response model. Specifying one response category as the reference is the same as specifying the other response category as the event category. You can specify the value (formatted if a format is applied) of the reference category in quotation marks, or you can specify one of the following keywords:

FIRST designates the first ordered category as the reference.
LAST designates the last ordered category as the reference.

By default, REF=LAST.

Model Options

The following options are available in the MODEL statement.

AGGREGATE

AGGREGATE=variable-list

specifies the subpopulations on which the Pearson’s chi-square test statistic and the log-likelihood ratio chi-square test statistic (deviance) are calculated if the LACKFIT option is specified. See the section “Rescaling the Covariance Matrix” on page 7637 for details of Pearson’s chi-square and deviance calculations.

Observations with common values in the given list of variables are regarded as coming from the same subpopulation. Variables in the list can be any variables in the input data set. Specifying the AGGREGATE option is equivalent to specifying the AGGREGATE= option with a variable list that includes all independent variables in the MODEL statement. The PROBIT procedure sorts the input data set according to the variables specified in this list. Information for the sorted data set is reported in the “Response-Covariate Profile” table.

The deviance and Pearson’s goodness-of-fit statistics are calculated if the LACKFIT option is specified in the MODEL statement. The calculated results are reported in the “Goodness-of-Fit” table. If the Pearson’s chi-square test is significant with the test level specified by the HPROB= option, the fiducial limits, if required with the INVERSECL option in the MODEL statement, are modified (see the section “Inverse Confidence Limits” on page 7638 for details). Also, the covariance matrix is rescaled by the dispersion parameter when the SCALE= option is specified.

ALPHA=value

sets the significance level for the confidence intervals for regression parameters, fiducial limits for the predicted values, and confidence intervals for the predicted probabilities. The value must be between 0 and 1. The default value is ALPHA=0.05.

CONVERGE=value

specifies the convergence criterion. Convergence is declared when the maximum change in the parameter estimates between Newton-Raphson steps is less than the value specified. The change is a relative change if the parameter is greater than 0.01 in absolute value; otherwise, it is an absolute change.

By default, CONVERGE=1.0E–8.
CORRB
displays the estimated correlation matrix of the parameter estimates.

COVB
displays the estimated covariance matrix of the parameter estimates.

**DISTRIBUTION**=distribution-type

**DIST**=distribution-type

**D**=distribution-type

specifies the cumulative distribution function used to model the response probabilities. The distributions are described in the section “Details: PROBIT Procedure” on page 7632. Valid values for distribution-type are as follows:

- **NORMAL**  
  the normal distribution for the probit model

- **LOGISTIC**  
  the logistic distribution for the logit model

- **EXTREMEVALUE | EXTREME | GOMPERTZ**  
  the extreme value, or Gompertz distribution for the gompit model

By default, DISTRIBUTION=NORMAL.

**HPROB**=p

specifies a minimum probability level for the Pearson’s chi-square to indicate a good fit. The default value is 0.10. The LACKFIT option must also be specified for this option to have any effect. For Pearson’s goodness-of-fit chi-square values with probability greater than the HPROB= value, the fiducial limits, if requested with the INVERSECL option, are computed by using a critical value of 1.96. For chi-square values with probability less than the value of the HPROB= option, the critical value is a 0.95 two-sided quantile value taken from the $t$ distribution with degrees of freedom equal to $(k - 1) \times m - q$, where $k$ is the number of levels for the response variable, $m$ is the number of different sets of independent variable values, and $q$ is the number of parameters fit in the model. If you specify the HPROB= option in both the PROC PROBIT and MODEL statements, the MODEL statement option takes precedence.

**INITIAL**=values

sets initial values for the parameters in the model other than the intercept. The values must be given in the order in which the variables are listed in the MODEL statement. If some of the independent variables listed in the MODEL statement are classification variables, then there must be as many values given for that variable as there are classification levels minus 1. The INITIAL option can be specified as follows.

<table>
<thead>
<tr>
<th>Type of List</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>List separated by blanks</td>
<td>initial=3 4 5</td>
</tr>
<tr>
<td>List separated by commas</td>
<td>initial=3,4,5</td>
</tr>
</tbody>
</table>

By default, all parameters have initial estimates of zero.

**NOTE:** The INITIAL= option is overwritten by the INEST= option in the PROC PROBIT statement.
INTERCEPT=value
initializes the intercept parameter to value. By default, INTERCEPT=0.

INVERSECL<(PROB=rates)>
computes confidence limits for the values of the first continuous independent variable (such as dose) that yield selected response rates. You can optionally specify a list of response rates as rates. The response rates must be between zero and one; they can be a list separated by blanks, commas, or in the form of a DO list. For example, the following expressions are all valid lists of response rates:

\[
\text{PROB = .1 TO .9 by .1} \\
\text{PROB = .1 .2 .3 .4} \\
\text{PROB = .01, .25, .75, .9}
\]

If the algorithm fails to converge (this can happen when C is nonzero), missing values are reported for the confidence limits. See the section “Inverse Confidence Limits” on page 7638 for details.

ITPRINT
displays the iteration history, the final evaluation of the gradient, and the second derivative matrix (Hessian).

LACKFIT
performs two goodness-of-fit tests (a Pearson’s chi-square test and a log-likelihood ratio chi-square test) for the fitted model.

To compute the test statistics, proper grouping of the observations into subpopulations is needed. You can use the AGGREGATE or AGGREGATE= option for this purpose. See the entry for the AGGREGATE and AGGREGATE= options under the MODEL statement. If neither AGGREGATE nor AGGREGATE= is specified, PROC PROBIT assumes each observation is from a separate subpopulation and computes the goodness-of-fit test statistics only for the events/trials syntax.

NOTE: This test is not appropriate if the data are very sparse, with only a few values at each set of the independent variable values.

If the Pearson’s chi-square test statistic is significant, then the covariance estimates and standard error estimates are adjusted. See the section “Lack-of-Fit Tests” on page 7636 for a description of the tests. Note that the LACKFIT option can also appear in the PROC PROBIT statement. See the section “PROC PROBIT Statement” on page 7577 for details.

MAXITER=value
MAXIT=value
specifies the maximum number of iterations to be performed in estimating the parameters. By default, MAXITER=50.

NOINT
fits a model with no intercept parameter. If the INTERCEPT= option is also specified, the intercept is fixed at the specified value; otherwise, it is set to zero. This is most useful when the response is binary. When the response has k levels, then k – 1 intercept parameters are fit. The NOINT option sets the intercept parameter corresponding to the lowest response level equal to zero. A Lagrange multiplier, or score, test for the restricted model is computed when the NOINT option is specified.
SCALE=scale

enables you to specify the method for estimating the dispersion parameter. To correct for overdispersion or underdispersion, the covariance matrix is multiplied by the estimate of the dispersion parameter. Valid values for scale are as follows:

D | DEVIANCE

specifies that the dispersion parameter be estimated by the deviance divided by its degrees of freedom.

P | PEARSON

specifies that the dispersion parameter be estimated by the Pearson’s chi-square statistic divided by its degrees of freedom. This is set as the default method for estimating the dispersion parameter.

You can use the AGGREGATE= option to define the subpopulations for calculating the Pearson’s chi-square statistic and the deviance.

The “Goodness-of-Fit” table includes the Pearson’s chi-square statistic, the deviance, their degrees of freedom, the ratio of each statistic divided by its degrees of freedom, and the corresponding p-value.

SINGULAR=value

specifies the singularity criterion for determining linear dependencies in the set of independent variables. The sum of squares and crossproducts matrix of the independent variables is formed and swept. If the relative size of a pivot becomes less than the value specified, then the variable corresponding to the pivot is considered to be linearly dependent on the previous set of variables considered. By default, value=1E–12.

**OUTPUT Statement**

```
OUTPUT < OUT=SAS-data-set keyword=name ...keyword=name > ;
```

The OUTPUT statement creates a new SAS data set containing all variables in the input data set and, optionally, the fitted probabilities, the estimate of $\hat{\beta}$, and the estimate of its standard error. Estimates of the probabilities, $\hat{\beta}$, and the standard errors are computed for observations with missing response values as long as the values of all the explanatory variables are nonmissing. This enables you to compute these statistics for additional settings of the explanatory variables that are of interest but for which responses are not observed.

You can specify multiple OUTPUT statements. Each OUTPUT statement creates a new data set and applies only to the preceding MODEL statement. If you want to create a SAS data set in a permanent library, you must specify a two-level name. For more information about permanent libraries and SAS data sets, see SAS Language Reference: Concepts.

Details on the specifications in the OUTPUT statement are as follows:

- **keyword=name** specifies the statistics to include in the output data set and assigns names to the new variables that contain the statistics. Specify a keyword for each desired statistic (see the following list of keywords), an equal sign, and the variable to contain the statistic.

The **keywords** allowed and the statistics they represent are as follows:
The PROBIT Procedure

**PROB | P** cumulative probability estimates

\[ p = C + (1 - C) F(a_j + x' \beta) \]

**STD** standard error estimates of \(a_j + x' b\)

**XBETA** estimates of \(a_j + x' \beta\)

**OUT=SAS-data-set** names the output data set. By default, the new data set is named by using the DATA convention.

When the *single variable response* syntax is used, the _LEVEL_ variable is added to the output data set, and there are \(k - 1\) output observations for each input observation, where \(k\) is the number of response levels. There is no observation output corresponding to the highest response level. For each of the \(k - 1\) observations, the PROB variable contains the fitted probability of obtaining a response level up to the level indicated by the _LEVEL_ variable, the XBETA variable contains \(a_j + x' b\), where \(j\) references the levels \((a_1 = 0)\), and the STD variable contains the standard error estimate of the XBETA variable. See the section “Details: PROBIT Procedure” on page 7632 for the formulas for the parameterizations.

**PREDPLOT Statement**

```
PREDPLOT < VAR=variable > < options > ;
```

The PREDPLOT statement plots the predicted probability against a single continuous variable (dose variable) in the MODEL statement for both the binomial model and the multinomial model. Confidence limits are available only for the binomial model. An attached box on the right side of the plot is used to label predicted probability curves with the names of their levels for the multinomial model. You can specify the color of this box by using the CLABBOX= option.

**VAR=** specifies a single continuous variable (dose variable) in the independent variable list of the MODEL statement. If a VAR= variable is not specified, the first single continuous variable in the independent variable list of the MODEL statement is used. If such a variable does not exist in the independent variable list of the MODEL statement, an error is reported.

The predicted probability is

\[ \hat{p} = C + (1 - C) F(x' \hat{b}) \]

for the binomial model and

\[
\begin{align*}
\hat{p}_1 &= C + (1 - C) F(x' \hat{b}) \\
\hat{p}_j &= (1 - C)(F(\hat{a}_j + x' \hat{b}) - F(\hat{a}_{j-1} + x' \hat{b})) \quad j = 2, \ldots, k - 1 \\
\hat{p}_k &= (1 - C)(1 - F(\hat{a}_{k-1} + x' \hat{b}))
\end{align*}
\]
for the multinomial model with \( k \) response levels, where \( F \) is the cumulative distribution function used to model the probability, \( x' \) is the vector of the covariates, \( \hat{a}_j \) are the estimated ordinal intercepts with \( \hat{a}_1 = 0 \), \( C \) is the threshold parameter, either known or estimated from the model, and \( \hat{b}' \) is the vector of estimated parameters.

To plot \( \hat{p} \) (or \( \hat{p}_j \)) as a function of a continuous variable \( x_1 \), the remaining covariates \( x_{-1} \) must be specified. You can use the XDATA= option to provide the values of \( x_{-1} \) (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow these rules:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.

**options**

enable you to plot the observed data and add features to the plot.

You can use **options** in the PREDPPLOT statement to do the following:

- superimpose specification limits
- suppress or add observed data points for the binomial model
- suppress or add confidence limits for the binomial model
- specify the levels for which predicted probability curves are requested for the multinomial model
- specify graphical enhancements (such as color or text height)

**Summary of Options**

Table 94.32 through Table 94.38 list **options** by function. The "Dictionary of Options" on page 7626 describes each option in detail.

**PREDPPLOT Options**

<table>
<thead>
<tr>
<th><strong>LEVEL=(character-list)</strong></th>
<th>Specifies the names of the levels for which the predicted probability curves are requested (only for the multinomial model)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NOCONF</strong></td>
<td>Suppresses confidence limits</td>
</tr>
<tr>
<td><strong>NODATA</strong></td>
<td>Suppresses observed data points on the plot</td>
</tr>
<tr>
<td><strong>NOTHRESH</strong></td>
<td>Suppresses the threshold line</td>
</tr>
<tr>
<td><strong>THRESHLABPOS=value</strong></td>
<td>Specifies the position for the label of the threshold line</td>
</tr>
</tbody>
</table>
General Options

Table 94.33  Color Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAXIS=color</td>
<td>Specifies color for the axes</td>
</tr>
<tr>
<td>CFIT=color</td>
<td>Specifies color for fitted curves</td>
</tr>
<tr>
<td>CFRAME=color</td>
<td>Specifies color for frame</td>
</tr>
<tr>
<td>CGRID=color</td>
<td>Specifies color for grid lines</td>
</tr>
<tr>
<td>CHREF=color</td>
<td>Specifies color for HREF= lines</td>
</tr>
<tr>
<td>CLABBOX=color</td>
<td>Specifies color for label box</td>
</tr>
<tr>
<td>CTEXT=color</td>
<td>Specifies color for text</td>
</tr>
<tr>
<td>CVREF=color</td>
<td>Specifies color for VREF= lines</td>
</tr>
</tbody>
</table>

Table 94.34  Options to Enhance Plots Produced on Graphics Devices

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANNOTATE=SAS-data-set</td>
<td>Specifies an Annotate data set</td>
</tr>
<tr>
<td>INBORDER</td>
<td>Requests a border around plot</td>
</tr>
<tr>
<td>LFIT=linetype</td>
<td>Specifies line style for fitted curves and confidence limits</td>
</tr>
<tr>
<td>LGRID=linetype</td>
<td>Specifies line style for grid lines</td>
</tr>
<tr>
<td>NOFRAME</td>
<td>Suppresses the frame around plotting areas</td>
</tr>
<tr>
<td>NOGRID</td>
<td>Suppresses grid lines</td>
</tr>
<tr>
<td>NOFIT</td>
<td>Suppresses fitted curves</td>
</tr>
<tr>
<td>NOHLABEL</td>
<td>Suppresses horizontal labels</td>
</tr>
<tr>
<td>NOHTICK</td>
<td>Suppresses horizontal ticks</td>
</tr>
<tr>
<td>NOVTICK</td>
<td>Suppresses vertical ticks</td>
</tr>
<tr>
<td>TURNVLABELS</td>
<td>Vertically strings out characters in vertical labels</td>
</tr>
<tr>
<td>WFIT=n</td>
<td>Specifies thickness for fitted curves</td>
</tr>
<tr>
<td>WGRID=n</td>
<td>Specifies thickness for grids</td>
</tr>
<tr>
<td>WREFL=n</td>
<td>Specifies thickness for reference lines</td>
</tr>
</tbody>
</table>
### Table 94.35  Axis Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAXIS=\textit{value1 to value2} \textit{&lt; by value3&gt; }</td>
<td>Specifies tick mark values for horizontal axis</td>
</tr>
<tr>
<td>HOFFSET=\textit{value}</td>
<td>Specifies offset for horizontal axis</td>
</tr>
<tr>
<td>HLOWER=\textit{value}</td>
<td>Specifies lower limit on horizontal axis scale</td>
</tr>
<tr>
<td>HUPPER=\textit{value}</td>
<td>Specifies upper limit on horizontal axis scale</td>
</tr>
<tr>
<td>NHTICK=\textit{n}</td>
<td>Specifies number of ticks for horizontal axis</td>
</tr>
<tr>
<td>VAXIS=\textit{value1 to value2} \textit{&lt; by value3&gt; }</td>
<td>Specifies tick mark values for vertical axis</td>
</tr>
<tr>
<td>VAXISLABEL=’\textit{label}’</td>
<td>Specifies label for vertical axis</td>
</tr>
<tr>
<td>VOFFSET=\textit{value}</td>
<td>Specifies offset for vertical axis</td>
</tr>
<tr>
<td>VLOWER=\textit{value}</td>
<td>Specifies lower limit on vertical axis scale</td>
</tr>
<tr>
<td>VUPPER=\textit{value}</td>
<td>Specifies upper limit on vertical axis scale</td>
</tr>
<tr>
<td>WAXIS=\textit{n}</td>
<td>Specifies thickness for axis</td>
</tr>
</tbody>
</table>

### Table 94.36  Graphics Catalog Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIPTION=’\textit{string}’</td>
<td>Specifies description for graphics catalog member</td>
</tr>
<tr>
<td>NAME=’\textit{string}’</td>
<td>Specifies name for plot in graphics catalog</td>
</tr>
</tbody>
</table>

### Table 94.37  Options for Text Enhancement

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FONT=\textit{font}</td>
<td>Specifies software font for text</td>
</tr>
<tr>
<td>HEIGHT=\textit{value}</td>
<td>Specifies height of text used outside framed areas</td>
</tr>
<tr>
<td>INHEIGHT=\textit{value}</td>
<td>Specifies height of text inside framed areas</td>
</tr>
</tbody>
</table>
Table 94.38  Options for Reference Lines

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HREF&lt;(INTERSECT) &gt;</td>
<td>Requests horizontal reference line</td>
</tr>
<tr>
<td>=value-list</td>
<td></td>
</tr>
<tr>
<td>HREFLABELS=</td>
<td>Specifies labels for HREF= lines</td>
</tr>
<tr>
<td>('label1', . . . , 'labeln')</td>
<td></td>
</tr>
<tr>
<td>HREFLABPOS=n</td>
<td>Specifies vertical position of labels for HREF= lines</td>
</tr>
<tr>
<td>LHREF=linetype</td>
<td>Specifies line style for HREF= lines</td>
</tr>
<tr>
<td>LVREF=linetype</td>
<td>Specifies line style for VREF= lines</td>
</tr>
<tr>
<td>VREF&lt;(INTERSECT) &gt;</td>
<td>Requests vertical reference line</td>
</tr>
<tr>
<td>=value-list</td>
<td></td>
</tr>
<tr>
<td>VREFLABELS=</td>
<td>Specifies labels for VREF= lines</td>
</tr>
<tr>
<td>('label1', . . . , 'labeln')</td>
<td></td>
</tr>
<tr>
<td>VREFLABPOS=n</td>
<td>Specifies horizontal position of labels for VREF= lines</td>
</tr>
</tbody>
</table>

**Dictionary of Options**

The following entries provide detailed descriptions of the *options* in the PREDPLOT statement.

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

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

specifies an Annotate data set, as described in *SAS/GRAPH: Reference*, that enables you to add features to the predicted probability plot. The ANNOTATE= data set you specify in the PREDPLOT statement is used for all plots created by the statement.

**CAXIS=** *color*

**CAXES=** *color*

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

**CFIT=** *color*

specifies the color for the fitted predicted probability curves. The default is the first color in the device color list.

**CFRAME=** *color*

**CFR=** *color*

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

**CGRID=** *color*

specifies the color for grid lines. The default is the first color in the device color list.

**CHREF=** *color*

**CH=** *color*

specifies the color for lines requested by the HREF= option. The default is the first color in the device color list.
CTEXT=color
    specifies the color for tick mark values and axis labels. The default is the color specified for the
    CTEXT= option in the most recent GOPTIONS statement.

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

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

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

HAXIS=value1 to value2 < by value3>
    specifies tick mark values for the horizontal axis; value1, value2, and value3 must be numeric, and
    value1 must be less than value2. The lower tick mark is value1. Tick marks are drawn at increments
    of value3. The last tick mark is the greatest value that does not exceed value2. If value3 is omitted, a
    value of 1 is used.

Examples of HAXIS= lists follow:
    haxis = 0 to 10
    haxis = 2 to 10 by 2
    haxis = 0 to 200 by 10

HEIGHT=value
    specifies the height of text used outside framed areas.

HLOWER=value
    specifies the lower limit on the horizontal axis scale. The HLOWER= option specifies value as the
    lower horizontal axis tick mark. The tick mark interval and the upper axis limit are determined
    automatically. This option has no effect if the HAXIS= option is used.

HOFFSET=value
    specifies the offset for the horizontal axis. The default value is 1.

HUPPER=value
    specifies value as the upper horizontal axis tick mark. The tick mark interval and the lower axis limit
    are determined automatically. This option has no effect if the HAXIS= option is used.

HREF < (INTERSECT) > =value-list
    requests reference lines perpendicular to the horizontal axis. If (INTERSECT) is specified, a second
    reference line perpendicular to the vertical axis is drawn that intersects the fit line at the same
    point as the horizontal axis reference line. If a horizontal axis reference line label is specified, the
    intersecting vertical axis reference line is labeled with the vertical axis value. See also the CHREF=,
    HREFLABELS=, and LHREF= options.
HREFLABELS=’label1’, . . . , ’labeln’ 
HREFLABEL=’label1’, . . . , ’labeln’ 
HREFLAB=’label1’, . . . , ’labeln’ 

specifies labels for the lines requested by the HREF= option. The number of labels must equal the number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

HREFLABPOS=n 

specifies the vertical position of labels for HREF= lines. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Top</td>
</tr>
<tr>
<td>2</td>
<td>Staggered from top</td>
</tr>
<tr>
<td>3</td>
<td>Bottom</td>
</tr>
<tr>
<td>4</td>
<td>Staggered from bottom</td>
</tr>
<tr>
<td>5</td>
<td>Alternating from top</td>
</tr>
<tr>
<td>6</td>
<td>Alternating from bottom</td>
</tr>
</tbody>
</table>

INBORDER 
requests a border around predicted probability plots.

INHEIGHT=value 

Specifies height of text inside framed areas.

LEVEL=(character-list) 
ORDINAL= (character-list) 

specifies the names of the levels for which predicted probability curves are requested. Names should be quoted and separated by space. If there is no correct name provided, no fitted probability curve is plotted.

LFIT=linetype 

specifies a line style for fitted curves and confidence limits. By default, fitted curves are drawn by connecting solid lines (linetype = 1) and confidence limits are drawn by connecting dashed lines (linetype = 3).

LGRID=linetype 

specifies a line style for all grid lines. The value for linetype is between 1 and 46. The default is 35.

LHREF=linetype 

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

LVREF=linetype 
LV=linetype 

specifies the line type for lines requested by the VREF= option. The default is 2, which produces a dashed line.
NAME=’string’
    specifies a name for the plot, up to eight characters, that appears in the PROC GREPLAY master menu. The default is ’PROBIT’.

NHTICK=n
    Specifies number of ticks for horizontal axis.

NVTICK=n
    Specifies number of ticks for vertical axis.

NOCONF
    suppresses confidence limits from the plot. This works only for the binomial model. Confidence limits are not plotted for the multinomial model.

NODATA
    suppresses observed data points from the plot. This works only for the binomial model. The data points are not plotted for the multinomial model.

NOFIT
    suppresses the fitted predicted probability curves.

NOFRAME
    suppresses the frame around plotting areas.

NOGRID
    suppresses grid lines.

NOHLABEL
    suppresses horizontal labels.

NOHTICK
    suppresses horizontal tick marks.

NOTHRESH
    suppresses the threshold line.

NOVLABEL
    suppresses vertical labels.

NOVTICK
    suppresses vertical tick marks.

TURNVLABELS
    vertically strings out characters in vertical labels.

THRESHLABPOS=n
    specifies the horizontal position of labels for the threshold line. The following table shows valid values for n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left</td>
</tr>
<tr>
<td>2</td>
<td>Right</td>
</tr>
</tbody>
</table>
**VAXIS=** value1 to value2 < by value3>
specifies tick mark values for the vertical axis; value1, value2, and value3 must be numeric, and value1
must be less than value2. The lower tick mark is value1. Tick marks are drawn at increments of value3.
The last tick mark is the greatest value that does not exceed value2. This method of specification of
tick marks is not valid for logarithmic axes. If value3 is omitted, a value of 1 is used.

Examples of VAXIS= lists follow:

\[
\begin{align*}
\text{vaxis} & = 0 \text{ to } 10 \\
\text{vaxis} & = 0 \text{ to } 2 \text{ by } .1
\end{align*}
\]

**VAXISLABEL=** ’string’
specifies a label for the vertical axis.

**VLOWER=** value
specifies the lower limit on the vertical axis scale. The VLOWER= option specifies value as the lower
vertical axis tick mark. The tick mark interval and the upper axis limit are determined automatically.
This option has no effect if the VAXIS= option is used.

**VOFFSET=** value
specifies the offset for the vertical axis.

**VREF=** value-list
requests reference lines perpendicular to the vertical axis. If (INTERSECT) is specified, a second
reference line perpendicular to the horizontal axis is drawn that intersects the fit line at the same point
as the vertical axis reference line. If a vertical axis reference line label is specified, the intersecting
horizontal axis reference line is labeled with the horizontal axis value. See also the CVREF=, LVREF=,
and VREFLABELS= options.

**VREFLABELS=** ’label1’, . . . , ’labeln’
**VREFLABEL=** ’label1’, . . . , ’labeln’
**VREFLAB=** ’label1’, . . . , ’labeln’
specifies labels for the lines requested by the VREF= option. The number of labels must equal the
number of lines. Enclose each label in quotes. Labels can be up to 16 characters.

**VREFLABPOS=** n
specifies the horizontal position of labels for VREF= lines. The following table shows valid values for
n and the corresponding label placements.

<table>
<thead>
<tr>
<th>n</th>
<th>Label Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left</td>
</tr>
<tr>
<td>2</td>
<td>Right</td>
</tr>
</tbody>
</table>

**VUPPER=** value
specifies the upper limit on the vertical axis scale. The VUPPER= option specifies value as the upper
vertical axis tick mark. The tick mark interval and the lower axis limit are determined automatically.
This option has no effect if the VAXIS= option is used.
WAXIS=n
  specifies line thickness for axes and frame. The default value is 1.

WFIT=n
  specifies line thickness for fitted curves. The default value is 1.

WGRID=n
  specifies line thickness for grids. The default value is 1.

WREFL=n
  specifies line thickness for reference lines. The default value is 1.

---

SLICE Statement

SLICE model-effect < / options > ;

The SLICE statement provides a general mechanism for performing a partitioned analysis of the LS-means for an interaction. This analysis is also known as an analysis of simple effects.

The SLICE statement uses the same options as the LSMEANS statement, which are summarized in Table 19.21. For details about the syntax of the SLICE statement, see the section “SLICE Statement” on page 506 in Chapter 19, “Shared Concepts and Topics.”

---

STORE Statement

STORE < OUT= > item-store-name < / LABEL='label' > ;

The STORE statement requests that the procedure save the context and results of the statistical analysis. The resulting item store has a binary file format that cannot be modified. The contents of the item store can be processed with the PLM procedure. For details about the syntax of the STORE statement, see the section “STORE Statement” on page 509 in Chapter 19, “Shared Concepts and Topics.”

---

TEST Statement

TEST < model-effects> < / options > ;

The TEST statement enables you to perform chi-square tests for model effects that test Type I, Type II, or Type III hypotheses. By default, the Type III tests are performed. For more information, see Chapter 19, “Shared Concepts and Topics.”

---

WEIGHT Statement

WEIGHT variable ;

A WEIGHT statement can be used with PROC PROBIT to weight each observation by the value of the variable specified. The contribution of each observation to the likelihood function is multiplied by the value of the weight variable. Observations with zero, negative, or missing weights are not used in model estimation.
Details: PROBIT Procedure

Missing Values

PROC PROBIT does not use any observations having missing values for any of the independent variables, the response variables, or the weight variable. If only the response variables are missing, statistics requested in the OUTPUT statement are computed.

Response Level Ordering

For binary response data, PROC PROBIT fits the following model by default:

\[ \Phi^{-1} \left( \frac{p - C}{1 - C} \right) = x' \beta \]

where \( p \) is the probability of the response level identified as the first level in the “Response Profile” table in the output and \( \Phi \) is the normal cumulative distribution function. By default, the covariate vector \( x \) contains an intercept term. This is sometimes called Abbot’s formula.

Because of the symmetry of the normal (and logistic) distribution, the effect of reversing the order of the two response values is to change the signs of \( \beta \) in the preceding equation.

By default, response levels appear in ascending, sorted order (that is, the lowest level appears first, and then the next lowest, and so on). There are a number of ways that you can control the sort order of the response categories and, therefore, which level is assigned the first ordered level. One of the most common sets of response levels is \{0,1\}, with 1 representing the event with the probability that is to be modeled.

Consider the example where \( Y \) takes the values 1 and 0 for event and nonevent, respectively, and EXPOSURE is the explanatory variable. By default, PROC PROBIT assigns the first ordered level to response level 0, causing the probability of the nonevent to be modeled. There are several ways to change this.

Besides recoding the variable \( Y \), you can do the following:

- Explicitly state which response level is to be modeled by using the response variable option EVENT= in the MODEL statement:

  \[
  \text{model } Y(\text{event='1'}) = \text{Exposure};
  \]

- Specify the nonevent category for the response variable in the response variable option REF= in the MODEL statement:

  \[
  \text{model } Y(\text{ref='0'}) = \text{Exposure};
  \]

- Specify the response variable option DESCENDING in the MODEL statement to assign the lowest ordered value to \( Y=1 \):
model Y(descending)=Exposure;

- Assign a format to Y such that the first formatted value (when the formatted values are put in sorted order) corresponds to the event. For the following example, Y=0 could be assigned formatted value ‘nonevent’ and Y=1 could be assigned formatted value ‘event.’ Since ORDER=FORMATTED by default, Y=1 becomes the first ordered level. See Example 94.3 for an illustration of this method.

```
proc format;
    value disease 1='event' 0='nonevent';
run;
proc probit;
    model y=exposure;
    format y disease.;
run;
```

- Arrange the input data set so that Y=1 appears first and use the ORDER=DATA option in the PROC PROBIT statement. Because ORDER=DATA sorts levels in order of their appearance in the data set, Y=1 becomes the first ordered level. Note that this option causes classification variables to be sorted by their order of appearance in the data set, also.

---

**Computational Method**

The log-likelihood function is maximized by means of a ridge-stabilized Newton-Raphson algorithm. Initial regression parameter estimates are set to zero. The INITIAL= and INTERCEPT= options in the MODEL statement can be used to give nonzero initial estimates.

The log-likelihood function, \( L \), is computed as

\[
L = \sum_i w_i \ln(p_i)
\]

where the sum is over the observations in the data set, \( w_i \) is the weight for the \( i \)th observation, and \( p_i \) is the modeled probability of the observed response. In the case of the events/trials syntax in the MODEL statement, each observation contributes two terms corresponding to the probability of the event and the probability of its complement:

\[
L = \sum_i w_i [r_i \ln(p_i) + (n_i - r_i) \ln(1 - p_i)]
\]

where \( r_i \) is the number of events and \( n_i \) is the number of trials for observation \( i \). This log-likelihood function differs from the log-likelihood function for a binomial or multinomial distribution by additive terms consisting of the log of binomial or multinomial coefficients. These terms are parameter-independent and do not affect the model estimation or the standard errors and tests.

The estimated covariance matrix, \( V \), of the parameter estimates is computed as the negative inverse of the information matrix of second derivatives of \( L \) with respect to the parameters evaluated at the final parameter estimates. Thus, the estimated covariance matrix is derived from the observed information matrix rather
than the expected information matrix (these are generally not the same). The standard error estimates for the parameter estimates are taken as the square roots of the corresponding diagonal elements of $V$.

If convergence of the maximum likelihood estimates is attained, a Type III chi-square test statistic is computed for each effect, testing whether there is any contribution from any of the levels of the effect. This statistic is computed as a quadratic form in the appropriate parameter estimates by using the corresponding submatrix of the asymptotic covariance matrix estimate. See Chapter 47, “The GLM Procedure,” and Chapter 15, “The Four Types of Estimable Functions,” for more information about Type III estimable functions.

The asymptotic covariance matrix is computed as the inverse of the observed information matrix. Note that if the NOINT option is specified and classification variables are used, the first classification variable contains a contribution from an intercept term. The results are displayed in an ODS table named “Type3Analysis”.

Chi-square tests for individual parameters are Wald tests based on the observed information matrix and the parameter estimates. If an effect has a single degree of freedom in the parameter estimates table, the chi-square test for this parameter is equivalent to the Type III test for this effect.

Prior to SAS 8.2, a multiple-degrees-of-freedom statistic was computed for each effect to test for contribution from any level of the effect. In general, the Type III test statistic in a main-effect-only model (no interaction terms) will be equal to the previously computed effect statistic, unless there are collinearities among the effects. If there are collinearities, the Type III statistic will adjust for them, and the value of the Type III statistic and the number of degrees of freedom might not be equal to those of the previous effect statistic.

The theory behind these tests assumes large samples. If the samples are not large, it might be better to base the tests on log-likelihood ratios. These changes in log likelihood can be obtained by fitting the model twice, once with all the parameters of interest and once leaving out the parameters to be tested. See Cox and Oakes (1984) for a discussion of the merits of some possible test methods.

If some of the independent variables are perfectly correlated with the response pattern, then the theoretical parameter estimates can be infinite. Although fitted probabilities of 0 and 1 are not especially pathological, infinite parameter estimates are required to yield these probabilities. Due to the finite precision of computer arithmetic, the actual parameter estimates are not infinite. Indeed, since the tails of the distributions allowed in the PROBIT procedure become small rapidly, an argument to the cumulative distribution function of around 20 becomes effectively infinite. In the case of such parameter estimates, the standard error estimates and the corresponding chi-square tests are not trustworthy.

### Distributions

The distributions, $F(x)$, allowed in the PROBIT procedure are specified with the DISTRIBUTION= option in the MODEL statement. The cumulative distribution functions for the available distributions are

<table>
<thead>
<tr>
<th>Cumulative Distribution Function</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$</td>
<td>Normal</td>
</tr>
<tr>
<td>$\frac{1}{1+e^{-x}}$</td>
<td>Logistic</td>
</tr>
<tr>
<td>$1 - e^{-e^x}$</td>
<td>Extreme value or Gompertz</td>
</tr>
</tbody>
</table>

The variances of these three distributions are not all equal to 1, and their means are not all equal to zero. Their means and variances are shown in the following table, where $\gamma$ is the Euler constant.
When comparing parameter estimates by using different distributions, you need to take into account the different scalings and, for the extreme value (or Gompertz) distribution, a possible shift in location. For example, if the fitted probabilities are in the neighborhood of 0.1 to 0.9, then the parameter estimates from the logistic model should be about $\pi / \sqrt{3}$ larger than the estimates from the probit model.

### INEST= SAS-data-set

The INEST= data set names a SAS data set that specifies initial estimates for all the parameters in the model. The INEST= data set must contain the intercept variables (named Intercept for binary response model and Intercept, Intercept2, Intercept3, and so forth, for multinomial response models) and all independent variables in the MODEL statement.

If BY processing is used, the INEST= data set should also include the BY variables, and there must be at least one observation for each BY group. If there is more than one observation in a BY group, the first one read is used for that BY group.

If the INEST= data set also contains the _TYPE_ variable, only observations with the _TYPE_ value “PARMS” are used as starting values. Combining the INEST= data set and the option MAXIT= in the MODEL statement, partial scoring can be done, such as predicting on a validation data set by using the model built from a training data set.

You can specify starting values for the iterative algorithm in the INEST= data set. This data set overwrites the INITIAL= option in the MODEL statement, which is a little difficult to use for models with multilevel interaction effects. The INEST= data set has the same structure as the “OUTEST= SAS-data-set” on page 7639, but it is not required to have all the variables or observations that appear in the OUTEST= data set. One simple use of the INEST= option is passing the previous OUTEST= data set directly to the next model as an INEST= data set, assuming that the two models have the same parameterization.

### Model Specification

For a two-level response, the probability that the lesser response occurs is modeled by the probit equation as

$$p = C + (1 - C) F(x'b)$$

The probability of the other (complementary) event is $1 - p$.

For a multilevel response with outcomes labeled $l_i$ for $i = 1, 2, \ldots, k$, the probability, $p_j$, of observing level $l_j$ is as follows:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Logistic</td>
<td>0</td>
<td>$\pi^2/3$</td>
</tr>
<tr>
<td>Extreme value or Gompertz</td>
<td>$-\gamma$</td>
<td>$\pi^2/6$</td>
</tr>
</tbody>
</table>
Thus, for a $k$-level response, there are $k - 2$ additional parameters, $a_2, a_3, \ldots, a_{k-1}$, estimated. These parameters are denoted by $\text{Intercept}_j, j = 2, 3, \ldots, k - 1$, in the output.

An intercept parameter is always added to the set of independent variables as the first term in the model unless the NOINT option is specified in the MODEL statement. If a classification variable taking on $k$ levels is used as one of the independent variables, a set of $k$ indicator variables is generated to model the effect of this variable. Because of the presence of the intercept term, there are at most $k - 1$ degrees of freedom for this effect in the model.

**Lack-of-Fit Tests**

Two goodness-of-fit tests can be requested from the PROBIT procedure: a Pearson’s chi-square test and a log-likelihood ratio chi-square test.

To compute the test statistics, you can use the AGGREGATE or AGGREGATE= option grouping the observations into subpopulations. If neither AGGREGATE nor AGGREGATE= is specified, PROC PROBIT assumes that each observation is from a separate subpopulation and computes the goodness-of-fit test statistics only for the $\text{events/trials}$ syntax.

If the Pearson’s goodness-of-fit chi-square test is requested and the $p$-value for the test is too small, variances and covariances are adjusted by a heterogeneity factor (the goodness-of-fit chi-square divided by its degrees of freedom) and a critical value from the $t$ distribution is used to compute the fiducial limits. The Pearson’s chi-square test statistic is computed as

$$
\chi^2_P = \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(r_{ij} - n_i \hat{p}_{ij})^2}{n_i \hat{p}_{ij}}
$$

where the sum on $i$ is over grouping, the sum on $j$ is over levels of response, $r_{ij}$ is the frequency of response level $j$ for the $i$th grouping, $n_i$ is the total frequency for the $i$th grouping, and $\hat{p}_{ij}$ is the fitted probability for the $j$th level at the $i$th grouping.

The likelihood ratio chi-square test statistic is computed as

$$
\chi^2_D = 2 \sum_{i=1}^{m} \sum_{j=1}^{k} r_{ij} \ln \left( \frac{r_{ij}}{n_i \hat{p}_{ij}} \right)
$$
This quantity is sometimes called the deviance. If the modeled probabilities fit the data, these statistics should be approximately distributed as chi-square with degrees of freedom equal to \((k - 1) \times m - q\), where \(k\) is the number of levels of the multinomial or binomial response, \(m\) is the number of sets of independent variable values (covariate patterns), and \(q\) is the number of parameters fit in the model.

In order for the Pearson’s statistic and the deviance to be distributed as chi-square, there must be sufficient replication within the groupings. When this is not true, the data are sparse, and the \(p\)-values for these statistics are not valid and should be ignored. Similarly, these statistics, divided by their degrees of freedom, cannot serve as indicators of overdispersion. A large difference between the Pearson’s statistic and the deviance provides some evidence that the data are too sparse to use either statistic.

---

**Rescaling the Covariance Matrix**

One way of correcting overdispersion is to multiply the covariance matrix by a dispersion parameter. You can supply the value of the dispersion parameter directly, or you can estimate the dispersion parameter based on either the Pearson’s chi-square statistic or the deviance for the fitted model.

The Pearson’s chi-square statistic \(\chi^2_P\) and the deviance \(\chi^2_D\) are defined in the section “Lack-of-Fit Tests” on page 7636. If the SCALE= option is specified in the MODEL statement, the dispersion parameter is estimated by

\[
\hat{\sigma}^2 = \begin{cases} 
\chi^2_P / (m(k - 1) - q) & \text{SCALE=PEARSON} \\
\chi^2_D / (m(k - 1) - q) & \text{SCALE=DEVIANACE} \\
(constant)^2 & \text{SCALE=constant}
\end{cases}
\]

In order for the Pearson’s statistic and the deviance to be distributed as chi-square, there must be sufficient replication within the subpopulations. When this is not true, the data are sparse, and the \(p\)-values for these statistics are not valid and should be ignored. Similarly, these statistics, divided by their degrees of freedom, cannot serve as indicators of overdispersion. A large difference between the Pearson’s statistic and the deviance provides some evidence that the data are too sparse to use either statistic.

You can use the AGGREGATE (or AGGREGATE=) option to define the subpopulation profiles. If you do not specify this option, each observation is regarded as coming from a separate subpopulation. For events/trials syntax, each observation represents \(n\) Bernoulli trials, where \(n\) is the value of the trials variable; for single-trial syntax, each observation represents a single trial. Without the AGGREGATE (or AGGREGATE=) option, the Pearson’s chi-square statistic and the deviance are calculated only for events/trials syntax.

Note that the parameter estimates are not changed by this method. However, their standard errors are adjusted for overdispersion, affecting their significance tests.

---

**Tolerance Distribution**

For a single independent variable, such as a dosage level, the models for the probabilities can be justified on the basis of a population with mean \(\mu\) and scale parameter \(\sigma\) of tolerances for the subjects. Then, given a dose \(x\), the probability, \(P\), of observing a response in a particular subject is the probability that the subject’s tolerance is less than the dose or

\[
P = F \left( \frac{x - \mu}{\sigma} \right)
\]
Thus, in this case, the intercept parameter, $b_0$, and the regression parameter, $b_1$, are related to $\mu$ and $\sigma$ by

$$b_0 = -\frac{\mu}{\sigma}, \quad b_1 = \frac{1}{\sigma}$$

**NOTE:** The parameter $\sigma$ is not equal to the standard deviation of the population of tolerances for the logistic and extreme value distributions.

### Inverse Confidence Limits

In bioassay problems, estimates of the values of the independent variables that yield a desired response are often needed. For instance, the value yielding a 50% response rate (called the ED50 or LD50) is often used. The `INVERSECL` option requests that confidence limits be computed for the value of the independent variable that yields a specified response. These limits are computed only for the first continuous variable effect in the model. The other variables are set either at their mean values if they are continuous or at the reference (last) level if they are discrete variables. For a discussion of inverse confidence limits, see Hubert, Bohidar, and Peace (1988).

For the PROBIT procedure, the response variable is a probability. An estimate of the first continuous variable value needed to achieve a response of $p$ is given by

$$\hat{x}_1 = \frac{1}{b_1} \left( F^{-1}(p) - x^*b^* \right)$$

where $F$ is the cumulative distribution function used to model the probability, $x^*$ is the vector of independent variables excluding the first one, which can be specified by the `XDATA=` option described in the section “XDATA= SAS-data-set” on page 7639, $b^*$ is the vector of parameter estimates excluding the first one, and $b_1$ is the estimated regression coefficient for the independent variable of interest. This estimate assumes that there is no natural response rate ($C = 0$). When $C$ is nonzero, the quantiles and confidence limits for the independent variable correspond to the adjusted probability $C + (1 - C)p$, rather than to $p$. As a result, an estimate of the value yielding response rate $p$ is associated with the $(p - C)/(1 - C)$ quantile. For example, if $C = 0.1$ then an estimate of the LD50 is found corresponding to the 0.44 quantile. This value can be thought of as yielding 50% of the variable’s effect, but a 44% response rate. For both binary and ordinal models, the `INVERSECL` option provides estimates of the value of $x_1$, which yields $Pr(\text{first response level}) = p$, for various values of $p$.

This estimator is given as a ratio of random variables, such as $r = a/b$. Confidence limits for this ratio can be computed by using Fieller’s theorem. A brief description of this theorem follows. See Finney (1971) for a more complete description of Fieller’s theorem.

If the random variables $a$ and $b$ are thought to be distributed as jointly normal, then for any fixed value $r$ the following probability statement holds if $z$ is an $\alpha/2$ quantile from the standard normal distribution and $V$ is the variance-covariance matrix of $a$ and $b$:

$$Pr \left( (a - rb)^2 > z^2(V_{aa} - 2rV_{ab} + r^2V_{bb}) \right) = \alpha$$

Usually the inequality can be solved for $r$ to yield a confidence interval. The PROBIT procedure uses a value of 1.96 for $z$, corresponding to an $\alpha$ value of 0.05, unless the goodness-of-fit $p$-value is less than the specified value of the `HPROB=` option. When this happens, the covariance matrix is scaled by the heterogeneity factor, and a $t$ distribution quantile is used for $z$. 
It is possible for the roots of the equation for \( r \) to be imaginary or for the confidence interval to be all points outside of an interval. In these cases, the limits are set to missing by the PROBIT procedure.

Although the normal and logistic distribution give comparable fitted values of \( p \) if the empirically observed proportions are not too extreme, they can give appreciably different values when extrapolated into the tails. Correspondingly, the estimates of the confidence limits and dose values can be different for the two distributions even when they agree quite well in the body of the data. Extrapolation outside of the range of the actual data is often sensitive to model assumptions, and caution is advised if extrapolation is necessary.

### OUTEST= SAS-data-set

The OUTEST= data set contains parameter estimates and the log likelihood for the model. You can specify a label in the MODEL statement to distinguish between the estimates for different models used by the PROBIT procedure. If you specify the COVOUT option, the OUTEST= data set also contains the estimated covariance matrix of the parameter estimates.

The OUTEST= data set contains each variable used as a dependent or independent variable in any MODEL statement. One observation consists of parameter values for the model with the dependent variable having the value -1. If you specify the COVOUT option, there are additional observations containing the rows of the estimated covariance matrix. For these observations, the dependent variable contains the parameter estimate for the corresponding row variable. The following variables are also added to the data set:

- `_MODEL_` a character variable containing the label of the MODEL statement, if present, or blank otherwise
- `_NAME_` a character variable containing the name of the dependent variable for the parameter estimates observations or the name of the row for the covariance matrix estimates
- `_TYPE_` a character variable containing the type of the observation, either PARMS for parameter estimates or COV for covariance estimates
- `_DIST_` a character variable containing the name of the distribution modeled
- `_LNLIKE_` a numeric variable containing the last computed value of the log likelihood
- `_C_` a numeric variable containing the estimated threshold parameter
- `INTERCEPT` a numeric variable containing the intercept parameter estimates and covariances

Any BY variables specified are also added to the OUTEST= data set.

### XDATA= SAS-data-set

The XDATA= data set is used for specifying values for the effects in the MODEL statement when predicted values and/or fiducial limits for a single continuous variable (dose variable) are required. It is also used for plots specified by the CDFPLOT, IPPPLOT, LPREDPLOT, and PREDPLOT statement.

The XDATA= data names a SAS data set that contains user input values for all the independent variables in the MODEL statement and the variables in the CLASS statement. The XDATA= data set has the same structure as the DATA= data set but is not required to have all the variables or observations that appear in the DATA= data set.
The XDATA= data set must contain all the independent variables in the MODEL statement and variables in the CLASS statement. Even though variables in the CLASS statement are not used in the MODEL statement, valid values are required for these variables in the XDATA= data set. Missing values are not allowed. For independent variables in the MODEL statement, although the dose variable’s value is not used in the computing of predicted values and/or fiducial limits for the dose variable, missing values are not allowed in the XDATA= data set for any of the independent variables. Missing values are allowed for the dependent variables and other variables if they are included in the XDATA= data set and not listed in the CLASS statement.

If BY processing is used, the XDATA= data set should also include the BY variables, and there must be at least one valid observation for each BY group. If there is more than one valid observation in one BY group, the last one read is used for that BY group.

If there is no XDATA= data set in the PROC PROBIT statement, by default, the PROBIT procedure will use overall mean for effects containing continuous variable (or variables) and the highest level of a single classification variable as reference level. The rules are summarized as follows:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.

---

**Traditional High-Resolution Graphics**

This section provides examples of using syntax available with the traditional high-resolution plots. A more modern alternative is to use ODS Graphics. See the section “ODS Graphics” on page 7643 for details.

There are four plot statements that you can use to request traditional high-resolution plots: CDFPLOT, IPPPLOT, LPREDPLOT, and PREDPPLOT. Some of these statements apply only to either the binomial model or the multinomial model. Table 94.39 shows the availability of these statements for different models.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Binomial</th>
<th>Multinomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDFPLOT</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>IPPPLOT</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>LPREDPLOT</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PREDPPLOT</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The following example uses the data set *study* in the section “Estimating the Natural Response Threshold Parameter” on page 7573 to illustrate how to create high-resolution plots for the binomial model:

```plaintext
proc probit data=study log10 optc;
  model respond/number=dose;
  predpplot var=dose cfit=blue; inset;
  lpredplot var=dose cfit=blue; inset;
  ipppplot var=dose cfit=blue; inset/pos=se;
run;
```
All plot statements must follow the MODEL statement. The VAR= option specifies a continuous independent variable (dose variable) against which the predicted probability or the linear predictor is plotted. The INSET statement requests the inset box with summary information. See the section “INSET Statement” on page 7595 for more details.

The PREDPLOT statement creates a plot that shows the relationship between dosage level, observed response proportions, and estimated probability values. See the section “PREDPLOT Statement” on page 7622 for more details. The IPPPLOT statement creates a similar plot. See the section “IPPPLOT Statement” on page 7597 for details about this plot. The LPREDPLOT statement creates a linear predictor plot, which is described in the section “LPREDPLOT Statement” on page 7605.

The following example uses the data set multi from Example 94.2 to illustrate how to create high-resolution plots for the multinomial model:

```
proc probit data=multi order=data;
  class prep;
  model symptoms=prep ldose;
  cdfplot var=ldose level=("None" "Mild" "Severe")
    cfit=blue cframe=ligr noconf;
  lpredplot var=ldose level=("None" "Mild" "Severe")
    cfit=blue cframe=ligr;
  predpplot var=ldose level=("None" "Mild" "Severe")
    cfit=blue cframe=ligr;
  weight n;
run;
```

The CDFPLOT statement creates a plot that shows the relationship between the cumulative response probabilities and the dose levels. The multinomial model plots are similar to those with the binomial model.

---

**Displayed Output**

If you request the iteration history (ITPRINT), PROC PROBIT displays the following:

- the current value of the log likelihood
- the ridging parameter for the modified Newton-Raphson optimization process
- the current estimate of the parameters
- the current estimate of the parameter C for a natural (threshold) model
- the values of the gradient and the Hessian on the last iteration

If you include classification variables, PROC PROBIT displays the following:

- the numbers of levels for each classification variable
- the (ordered) values of the levels
- the number of observations used
After the model is fit, PROC PROBIT displays the following:

- the name of the input data set
- the name of the dependent variables
- the number of observations used
- the number of events and the number of trials
- the final value of the log-likelihood function
- the parameter estimates
- the standard error estimates of the parameter estimates
- approximate chi-square test statistics for the test

If you specify the COVB or CORRB options, PROC PROBIT displays the following:

- the estimated covariance matrix for the parameter estimates
- the estimated correlation matrix for the parameter estimates

If you specify the LACKFIT option, PROC PROBIT displays the following:

- a count of the number of levels of the response and the number of distinct sets of independent variables
- a goodness-of-fit test based on the Pearson’s chi-square
- a goodness-of-fit test based on the likelihood-ratio chi-square

If you specify only one independent variable, the normal distribution is used to model the probabilities, and the response is binary, then PROC PROBIT displays the following:

- the mean MU of the stimulus tolerance
- the scale parameter SIGMA of the stimulus tolerance
- the covariance matrix for MU, SIGMA, and the natural response parameter C

If you specify the INVERSECL options, PROC PROBIT also displays the following:

- the estimated dose along with the 95% fiducial limits for probability levels 0.01 to 0.10, 0.15 to 0.85 by 0.05, and 0.90 to 0.99
**ODS Table Names**

PROC PROBIT assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information about ODS, see Chapter 20, “Using the Output Delivery System.”

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClassLevels</td>
<td>Classification variable levels</td>
<td>CLASS</td>
<td>Default</td>
</tr>
<tr>
<td>ConvergenceStatus</td>
<td>Convergence status</td>
<td>MODEL</td>
<td>Default</td>
</tr>
<tr>
<td>CorrB</td>
<td>Parameter estimate correlation matrix</td>
<td>MODEL</td>
<td>CORRB</td>
</tr>
<tr>
<td>CovB</td>
<td>Parameter estimate covariance matrix</td>
<td>MODEL</td>
<td>COVB</td>
</tr>
<tr>
<td>CovTolerance</td>
<td>Covariance matrix for location and scale</td>
<td>MODEL</td>
<td>Default</td>
</tr>
<tr>
<td>GoodnessOfFit</td>
<td>Goodness-of-fit tests</td>
<td>MODEL</td>
<td>LACKFIT</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>Heterogeneity correction</td>
<td>MODEL</td>
<td>LACKFIT</td>
</tr>
<tr>
<td>IterHistory</td>
<td>Iteration history</td>
<td>MODEL</td>
<td>ITPRINT</td>
</tr>
<tr>
<td>LagrangeStatistics</td>
<td>Lagrange statistics</td>
<td>MODEL</td>
<td>NOINT</td>
</tr>
<tr>
<td>LastGrad</td>
<td>Last evaluation of the gradient</td>
<td>MODEL</td>
<td>ITPRINT</td>
</tr>
<tr>
<td>LastHess</td>
<td>Last evaluation of the Hessian</td>
<td>MODEL</td>
<td>ITPRINT</td>
</tr>
<tr>
<td>LogProbitAnalysis</td>
<td>Probit analysis for log dose</td>
<td>MODEL</td>
<td>INVERSECL</td>
</tr>
<tr>
<td>ModelInfo</td>
<td>Model information</td>
<td>MODEL</td>
<td>Default</td>
</tr>
<tr>
<td>MuSigma</td>
<td>Location and scale</td>
<td>MODEL</td>
<td>Default</td>
</tr>
<tr>
<td>NObs</td>
<td>Observations summary</td>
<td>PROC</td>
<td>Default</td>
</tr>
<tr>
<td>ParameterEstimates</td>
<td>Parameter estimates</td>
<td>MODEL</td>
<td>Default</td>
</tr>
<tr>
<td>ParmInfo</td>
<td>Parameter indices</td>
<td>MODEL</td>
<td>Default</td>
</tr>
<tr>
<td>ProbitAnalysis</td>
<td>Probit analysis for linear dose</td>
<td>MODEL</td>
<td>INVERSECL</td>
</tr>
<tr>
<td>ResponseLevels</td>
<td>Response-covariate profile</td>
<td>MODEL</td>
<td>LACKFIT</td>
</tr>
<tr>
<td>ResponseProfiles</td>
<td>Counts for ordinal data</td>
<td>MODEL</td>
<td>Default</td>
</tr>
<tr>
<td>Type3Analysis</td>
<td>Type III tests</td>
<td>MODEL</td>
<td>Default</td>
</tr>
</tbody>
</table>

**ODS Graphics**

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “Statistical Graphics Using ODS.”

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 607 in Chapter 21, “Statistical Graphics Using ODS.”

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “A Primer on ODS Statistical Graphics” on page 606 in Chapter 21, “Statistical Graphics Using ODS.”

These ODS graphs are controlled by the PLOTS= option in the PROC PROBIT statement. You can specify more than one graph request with the PLOTS= option. Table 94.41 summarizes these requests.
Table 94.41 Options for Plots

<table>
<thead>
<tr>
<th>Option</th>
<th>Plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>All appropriate plots</td>
</tr>
<tr>
<td>CDFPLOT</td>
<td>Estimated cumulative probability</td>
</tr>
<tr>
<td>IPPPLOT</td>
<td>Inverse predicted probability</td>
</tr>
<tr>
<td>LPREDPLOT</td>
<td>Linear predictor</td>
</tr>
<tr>
<td>NONE</td>
<td>No plot</td>
</tr>
<tr>
<td>PREDPPLOT</td>
<td>Predicted probability</td>
</tr>
</tbody>
</table>

The following subsections provide information about these graphs.

**ODS Graph Names**

PROC PROBIT assigns a name to each graph it creates using ODS. You can use these names to reference the graphs when using ODS. The names are listed in Table 94.42.

Table 94.42 Graphs Produced by PROC PROBIT

<table>
<thead>
<tr>
<th>ODS Graph Name</th>
<th>Plot Description</th>
<th>Statement</th>
<th>PLOTS= Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDFPlot</td>
<td>Estimated cumulative probability</td>
<td>PROC</td>
<td>CDFPLOT</td>
</tr>
<tr>
<td>IPPPlot</td>
<td>Inverse predicted probability</td>
<td>PROC</td>
<td>IPPPLOT</td>
</tr>
<tr>
<td>LPredPlot</td>
<td>Linear predictor</td>
<td>PROC</td>
<td>LPREDPLOT</td>
</tr>
<tr>
<td>PredPPlot</td>
<td>Predicted probability</td>
<td>PROC</td>
<td>PREDPPLOT</td>
</tr>
</tbody>
</table>

**CDF Plot**

For a multinominal model, the predicted cumulative distribution function is defined as

\[
\hat{F}_j(x) = C + (1 - C) F(\hat{a}_j + x^T \hat{b})
\]

where \( j = 1, \ldots, k \) are the indexes of the \( k \) levels of the multinominal response variable, \( F \) is the CDF of the distribution used to model the cumulative probabilities, \( \hat{b} \) is the vector of estimated parameters, \( x \) is the covariate vector, \( \hat{a}_j \) are estimated ordinal intercepts with \( \hat{a}_1 = 0 \), and \( C \) is the threshold parameter, either known or estimated from the model. Let \( x_1 \) be the covariate corresponding to the dose variable and \( x_{-1} \) be the vector of the rest of the covariates. Let the corresponding estimated parameters be \( \hat{b}_1 \) and \( \hat{b}_{-1} \). Then

\[
\hat{F}_j(x) = C + (1 - C) F(\hat{a}_j + x_1 \hat{b}_1 + x_{-1} \hat{b}_{-1})
\]

To plot \( \hat{F}_j \) as a function of \( x_1, x_{-1} \) must be specified. You can use the XDATA= option to provide the values of \( x_{-1} \) (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow these rules:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.
The LEVEL= suboption specify the levels of the multinomial response variable for which the CDF curves are requested. There are \( k - 1 \) curves for a \( k \)-level multinomial response variable (for the highest level, it is the constant line 1). You can specify any of them to be plotted by the LEVEL= suboption. See the plot in Output 94.2.6 for an example.

**Inverse Predicted Probability Plot**

For the binomial model, the response variable is a probability. An estimate of the dose level \( \hat{x}_1 \) needed for a response of \( p \) is given by

\[
\hat{x}_1 = (F^{-1}(p) - x'_{-1} \hat{b}_{-1}) / \hat{b}_1
\]

where \( F \) is the cumulative distribution function used to model the probability, \( x_{-1} \) is the vector of the rest of the covariates, \( \hat{b}_{-1} \) is the vector of the estimated parameters corresponding to \( x_{-1} \), and \( \hat{b}_1 \) is the estimated parameter for the dose variable of interest.

To plot \( \hat{x}_1 \) as a function of \( p, x_{-1} \) must be specified. You can use the XDATA= option to provide the values of \( x_{-1} \) (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow these rules:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.

Output 94.4.12 in Example 94.4 shows an inverse predicted probability plot.

**Linear Predictor Plot**

For both binomial models and multinomial models, the linear predictor \( x'b \) can be plotted against the first single continuous variable (dose variable) in the MODEL statement.

Let \( x_1 \) be the covariate of the dose variable, \( x_{-1} \) be the vector of the rest of the covariates, \( \hat{b}_{-1} \) be the vector of estimated parameters corresponding to \( x_{-1} \), and \( \hat{b}_1 \) be the estimated parameter for the dose variable of interest.

To plot \( \hat{x}'b \) as a function of \( x_1, x_{-1} \) must be specified. You can use the XDATA= option to provide the values of \( x_{-1} \) (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow these rules:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.

For the multinomial model, you can use the LEVEL= suboption to specify the levels for which the linear predictor lines are plotted.

The confidence limits for the predicted values are only available for the binomial model. Output 94.4.13 in Example 94.4 shows a linear predictor plot for a binomial model.
Predicted Probability Plot

The predicted probability is

\[ \hat{p} = C + (1 - C) F(x'\hat{b}) \]

for the binomial model and

\[
\begin{align*}
\hat{p}_1 &= C + (1 - C) F(x'\hat{b}) \\
\hat{p}_j &= (1 - C)(F(\hat{a}_j + x'\hat{b}) - F(\hat{a}_{j-1} + x'\hat{b})), \quad j = 2, \ldots, k - 1 \\
\hat{p}_k &= (1 - C)(1 - F(\hat{a}_{k-1} + x'\hat{b}))
\end{align*}
\]

for the multinomial model with \( k \) response levels, where \( F \) is the cumulative distribution function used to model the probability, \( x' \) is the vector of the covariates, \( \hat{a}_j \) are the estimated ordinal intercepts with \( \hat{a}_1 = 0 \), \( C \) is the threshold parameter, either known or estimated from the model, and \( \hat{b}' \) is the vector of estimated parameters.

To plot \( \hat{p} \) (or \( \hat{p}_j \)) as a function of a continuous variable \( x_1 \), the remaining covariates \( x_{-1} \) must be specified. You can use the XDATA= option to provide the values of \( x_{-1} \) (see the XDATA= option in the PROC PROBIT statement for details), or use the default values that follow these rules:

- If the effect contains a continuous variable (or variables), the overall mean of this effect is used.
- If the effect is a single classification variable, the highest level of the variable is used.

For the multinomial model, you can use the LEVEL= suboption to specify the levels for which the linear predictor lines are plotted.

Confidence limits are plotted only for the binomial model. Output 94.1.7 in Example 94.1 shows a predicted probability plot for a binomial model; and Output 94.2.3 in Example 94.2 shows a predicted probability plot for a multinomial model.

---

Examples: PROBIT Procedure

Example 94.1: Dosage Levels

In this example, Dose is a variable representing the level of a stimulus, N represents the number of subjects tested at each level of the stimulus, and Response is the number of subjects responding to that level of the stimulus. Both probit and logit response models are fit to the data. The LOG10 option in the PROC PROBIT statement requests that the log base 10 of Dose is used as the independent variable. The probability \( p \) of a positive response is modeled as

\[ p = Pr(\text{Response}) = F(b_0 + b_1 \times \log_{10}(\text{Dose})) \]
The probabilities are estimated first by using the normal distribution function (the default) and then by using the logistic distribution function. Note that, in this model specification, the natural rate is assumed to be zero.

The LACKFIT option specifies lack-of-fit tests and the INVERSECL option specifies inverse confidence limits.

In the DATA step that reads the data, a number of observations are generated that have a missing value for the response. Although the PROBIT procedure does not use the observations with the missing values to fit the model, it does give predicted values for all nonmissing sets of independent variables. These data points fill in the plot of fitted and observed values in the logistic model displayed in Output 94.1.7. The plot, requested with the PLOT=PREDPLOT option, displays the estimated logistic cumulative distribution function and the observed response rates.

The following statements produce Output 94.1.1:

```plaintext
data a;
  infile cards eof=eof;
  input Dose N Response @@;
  Observed= Response/N;
  output;
  return;
  eof: do Dose=0.5 to 7.5 by 0.25;
    output;
  end;
  datalines;
  1 10 1 2 12 2 3 10 4 4 10 5 5 12 8 6 10 8 7 10 10
  ;
ods graphics on;
proc probit log10;
  model Response/N=Dose / lackfit inversecl itprint;
  output out=B p=Prob std=std xbeta=xbeta;
run;
```

**Output 94.1.1** Probit Analysis with Normal Distribution

**The Probit Procedure**

<table>
<thead>
<tr>
<th>Iter</th>
<th>Ridge</th>
<th>Loglikelihood</th>
<th>Intercept</th>
<th>Log10(Dose)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-51.292891</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>-37.881166</td>
<td>-1.355817008</td>
<td>2.635206083</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-37.286169</td>
<td>-1.764939171</td>
<td>3.3408954936</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>-37.280389</td>
<td>-1.812147863</td>
<td>3.4172391614</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>-37.280388</td>
<td>-1.812704962</td>
<td>3.418117919</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>-37.280388</td>
<td>-1.812704962</td>
<td>3.418117919</td>
</tr>
</tbody>
</table>
### Output 94.1.1 continued

#### Model Information

<table>
<thead>
<tr>
<th>Model Information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
<td>WORK.A</td>
</tr>
<tr>
<td>Events Variable</td>
<td>Response</td>
</tr>
<tr>
<td>Trials Variable</td>
<td>N</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>7</td>
</tr>
<tr>
<td>Number of Events</td>
<td>38</td>
</tr>
<tr>
<td>Number of Trials</td>
<td>74</td>
</tr>
<tr>
<td>Name of Distribution</td>
<td>Normal</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-37.28038802</td>
</tr>
</tbody>
</table>

#### Last Evaluation of the Negative of the Gradient

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Log10(Dose)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4349069E-7</td>
<td>-2.09809E-8</td>
</tr>
</tbody>
</table>

#### Last Evaluation of the Negative of the Hessian

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Log10(Dose)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>36.005280383</td>
</tr>
<tr>
<td>Log10(Dose)</td>
<td>20.152675982</td>
</tr>
</tbody>
</table>

#### Goodness-of-Fit Tests

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>3.6497</td>
<td>5</td>
<td>0.7299</td>
<td>0.6009</td>
</tr>
<tr>
<td>L.R. Chi-Square</td>
<td>4.6381</td>
<td>5</td>
<td>0.9276</td>
<td>0.4616</td>
</tr>
</tbody>
</table>

#### Response-Covariate Profile

<table>
<thead>
<tr>
<th>Response Levels</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Covariate Values</td>
<td>7</td>
</tr>
</tbody>
</table>

#### Analysis of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.8127</td>
<td>0.4493</td>
<td>-2.6934</td>
<td>-0.9320</td>
<td>16.27</td>
</tr>
<tr>
<td>Log10(Dose)</td>
<td>1</td>
<td>3.4181</td>
<td>0.7455</td>
<td>1.9569</td>
<td>4.8794</td>
<td>21.02</td>
</tr>
</tbody>
</table>

#### Probit Model in Terms of Tolerance Distribution

<table>
<thead>
<tr>
<th>MU</th>
<th>SIGMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.53032254</td>
<td>0.29255866</td>
</tr>
</tbody>
</table>

#### Estimated Covariance Matrix for Tolerance Parameters

<table>
<thead>
<tr>
<th>MU</th>
<th>SIGMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MU</td>
<td>SIGMA</td>
</tr>
<tr>
<td>0.002418</td>
<td>-0.000409</td>
</tr>
<tr>
<td>0.000409</td>
<td>0.004072</td>
</tr>
</tbody>
</table>
The \textit{p}-values in the goodness-of-fit table of 0.6009 for the Pearson’s chi-square and 0.4616 for the likelihood ratio chi-square indicate an adequate fit for the model fit with the normal distribution.

Tolerance distribution parameter estimates for the normal distribution indicate a mean tolerance for the population of 0.5303.

Output 94.1.2 displays probit analysis with the logarithm of dose levels. The LD50 (ED50 for log dose) is 0.5303, the dose corresponding to a probability of 0.5. This is the same as the mean tolerance for the normal distribution.

\textbf{Output 94.1.2} Probit Analysis with Normal Distribution

\begin{center}
\textbf{The Probit Procedure}
\end{center}

\begin{table}[h]
\centering
\begin{tabular}{cccc}
\hline
Probability & Log10(Dose) & 95\% Fiducial Limits \\
\hline
0.01 & -1.5027 & -0.6951 & 0.0771 \\
0.02 & -0.7052 & -0.5576 & 0.1347 \\
0.03 & -0.0199 & -0.4706 & 0.1716 \\
0.04 & 0.0181 & -0.4053 & 0.1994 \\
0.05 & 0.0491 & -0.3523 & 0.2221 \\
0.06 & 0.0754 & -0.3073 & 0.2416 \\
0.07 & 0.0986 & -0.2679 & 0.2588 \\
0.08 & 0.1192 & -0.2327 & 0.2742 \\
0.09 & 0.1380 & -0.2008 & 0.2883 \\
0.10 & 0.1553 & -0.1714 & 0.3014 \\
0.15 & 0.2270 & -0.5086 & 0.3563 \\
0.20 & 0.2840 & 0.0436 & 0.4012 \\
0.25 & 0.3329 & 0.1234 & 0.4411 \\
0.30 & 0.3769 & 0.1934 & 0.4785 \\
0.35 & 0.4175 & 0.2568 & 0.5150 \\
0.40 & 0.4562 & 0.3142 & 0.5518 \\
0.45 & 0.4935 & 0.3675 & 0.5899 \\
0.50 & 0.5303 & 0.4169 & 0.6305 \\
0.55 & 0.5670 & 0.4629 & 0.6745 \\
0.60 & 0.6044 & 0.5061 & 0.7227 \\
0.65 & 0.6430 & 0.5473 & 0.7760 \\
0.70 & 0.6837 & 0.5874 & 0.8355 \\
0.75 & 0.7276 & 0.6277 & 0.9026 \\
0.80 & 0.7765 & 0.6699 & 0.9800 \\
0.85 & 0.8335 & 0.7167 & 1.0727 \\
0.90 & 0.9052 & 0.7731 & 1.1919 \\
0.91 & 0.9225 & 0.7864 & 1.2209 \\
0.92 & 0.9413 & 0.8008 & 1.2526 \\
0.93 & 0.9620 & 0.8165 & 1.2875 \\
0.94 & 0.9851 & 0.8339 & 1.3267 \\
0.95 & 1.0115 & 0.8536 & 1.3714 \\
0.96 & 1.0425 & 0.8766 & 1.4242 \\
0.97 & 1.0805 & 0.9048 & 1.4892 \\
0.98 & 1.1311 & 0.9418 & 1.5760 \\
0.99 & 1.2109 & 0.9998 & 1.7132 \\
\hline
\end{tabular}
\end{table}
Output 94.1.3 displays probit analysis with dose levels. The ED50 for dose is 3.39 with a 95% confidence interval of (2.61, 4.27).

Output 94.1.3  Probit Analysis with Normal Distribution

The Probit Procedure

<table>
<thead>
<tr>
<th>Probability</th>
<th>Dose</th>
<th>95% Fiducial Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.70750</td>
<td>0.20175 1.19427</td>
</tr>
<tr>
<td>0.02</td>
<td>0.85012</td>
<td>0.27691 1.36380</td>
</tr>
<tr>
<td>0.03</td>
<td>0.95517</td>
<td>0.33834 1.48444</td>
</tr>
<tr>
<td>0.04</td>
<td>1.04266</td>
<td>0.39324 1.58274</td>
</tr>
<tr>
<td>0.05</td>
<td>1.11971</td>
<td>0.44429 1.66793</td>
</tr>
<tr>
<td>0.06</td>
<td>1.18976</td>
<td>0.49282 1.74443</td>
</tr>
<tr>
<td>0.07</td>
<td>1.25478</td>
<td>0.53960 1.81473</td>
</tr>
<tr>
<td>0.08</td>
<td>1.31600</td>
<td>0.58515 1.88042</td>
</tr>
<tr>
<td>0.09</td>
<td>1.37427</td>
<td>0.62980 1.94252</td>
</tr>
<tr>
<td>0.10</td>
<td>1.43019</td>
<td>0.67380 2.00181</td>
</tr>
<tr>
<td>0.15</td>
<td>1.68696</td>
<td>0.88950 2.27147</td>
</tr>
<tr>
<td>0.20</td>
<td>1.92353</td>
<td>1.10584 2.51906</td>
</tr>
<tr>
<td>0.25</td>
<td>2.15276</td>
<td>1.32870 2.76161</td>
</tr>
<tr>
<td>0.30</td>
<td>2.38180</td>
<td>1.56128 3.01000</td>
</tr>
<tr>
<td>0.35</td>
<td>2.61573</td>
<td>1.80543 3.27374</td>
</tr>
<tr>
<td>0.40</td>
<td>2.85893</td>
<td>2.06200 3.56306</td>
</tr>
<tr>
<td>0.45</td>
<td>3.11573</td>
<td>2.33098 3.89038</td>
</tr>
<tr>
<td>0.50</td>
<td>3.39096</td>
<td>2.61175 4.27138</td>
</tr>
<tr>
<td>0.55</td>
<td>3.69051</td>
<td>2.90374 4.72619</td>
</tr>
<tr>
<td>0.60</td>
<td>4.02199</td>
<td>3.20759 5.28090</td>
</tr>
<tr>
<td>0.65</td>
<td>4.39594</td>
<td>3.52651 5.97077</td>
</tr>
<tr>
<td>0.70</td>
<td>4.82770</td>
<td>3.86765 6.84706</td>
</tr>
<tr>
<td>0.75</td>
<td>5.34134</td>
<td>4.24385 7.99189</td>
</tr>
<tr>
<td>0.80</td>
<td>5.97787</td>
<td>4.67724 9.55169</td>
</tr>
<tr>
<td>0.85</td>
<td>6.81617</td>
<td>5.20900 11.82480</td>
</tr>
<tr>
<td>0.90</td>
<td>8.03992</td>
<td>5.93105 15.55653</td>
</tr>
<tr>
<td>0.91</td>
<td>8.36704</td>
<td>6.11584 16.63320</td>
</tr>
<tr>
<td>0.92</td>
<td>8.73752</td>
<td>6.32165 17.89163</td>
</tr>
<tr>
<td>0.93</td>
<td>9.16385</td>
<td>6.55431 19.39034</td>
</tr>
<tr>
<td>0.94</td>
<td>9.66463</td>
<td>6.82245 21.21881</td>
</tr>
<tr>
<td>0.95</td>
<td>10.26925</td>
<td>7.13949 23.52275</td>
</tr>
<tr>
<td>0.96</td>
<td>11.02811</td>
<td>7.52816 26.56066</td>
</tr>
<tr>
<td>0.97</td>
<td>12.03830</td>
<td>8.03149 30.85201</td>
</tr>
<tr>
<td>0.98</td>
<td>13.52585</td>
<td>8.74763 37.67206</td>
</tr>
<tr>
<td>0.99</td>
<td>16.25233</td>
<td>9.99709 51.66627</td>
</tr>
</tbody>
</table>
The following statements request probit analysis of dosage levels with the logistic distribution:

```
proc probit log10 plot=predpplot;
    model Response/N=Dose / d=logistic inversecl;
    output out=B p=Prob std=std xbeta=xbeta;
run;
```

The regression parameter estimates in Output 94.1.4 for the logistic model of –3.22 and 5.97 are approximately \( \pi / \sqrt{3} \) times as large as those for the normal model.

**Output 94.1.4** Probit Analysis with Logistic Distribution

```
The Probit Procedure

Model Information

<table>
<thead>
<tr>
<th>Data Set</th>
<th>WORK.B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events Variable</td>
<td>Response</td>
</tr>
<tr>
<td>Trials Variable</td>
<td>N</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>7</td>
</tr>
<tr>
<td>Number of Events</td>
<td>38</td>
</tr>
<tr>
<td>Number of Trials</td>
<td>74</td>
</tr>
<tr>
<td>Name of Distribution</td>
<td>Logistic</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-37.11065336</td>
</tr>
</tbody>
</table>
```

Analysis of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.2246</td>
<td>0.8861</td>
<td>-4.9613 -1.4880</td>
<td>13.24</td>
<td>0.0003</td>
</tr>
<tr>
<td>Log10(Dose)</td>
<td>1</td>
<td>5.9702</td>
<td>1.4492</td>
<td>3.1299 8.8105</td>
<td>16.97</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
```

Output 94.1.5 and Output 94.1.6 show that both the ED50 and the LD50 are similar to those for the normal model.
### Output 94.1.5 Probit Analysis with Logistic Distribution

**The Probit Procedure**

<table>
<thead>
<tr>
<th>Probability</th>
<th>Log10(Dose)</th>
<th>Fiducial Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-0.22955</td>
<td>0.04234</td>
</tr>
<tr>
<td>0.02</td>
<td>-0.11175</td>
<td>0.12404</td>
</tr>
<tr>
<td>0.03</td>
<td>-0.04212</td>
<td>0.17265</td>
</tr>
<tr>
<td>0.04</td>
<td>0.00780</td>
<td>0.20771</td>
</tr>
<tr>
<td>0.05</td>
<td>0.04693</td>
<td>0.23533</td>
</tr>
<tr>
<td>0.06</td>
<td>0.07925</td>
<td>0.25826</td>
</tr>
<tr>
<td>0.07</td>
<td>0.10686</td>
<td>0.27796</td>
</tr>
<tr>
<td>0.08</td>
<td>0.13103</td>
<td>0.29530</td>
</tr>
<tr>
<td>0.09</td>
<td>0.15259</td>
<td>0.31085</td>
</tr>
<tr>
<td>0.10</td>
<td>0.17209</td>
<td>0.32498</td>
</tr>
<tr>
<td>0.15</td>
<td>0.24958</td>
<td>0.38207</td>
</tr>
<tr>
<td>0.20</td>
<td>0.30792</td>
<td>0.42645</td>
</tr>
<tr>
<td>0.25</td>
<td>0.35611</td>
<td>0.46451</td>
</tr>
<tr>
<td>0.30</td>
<td>0.39820</td>
<td>0.49932</td>
</tr>
<tr>
<td>0.35</td>
<td>0.43644</td>
<td>0.53275</td>
</tr>
<tr>
<td>0.40</td>
<td>0.47221</td>
<td>0.56619</td>
</tr>
<tr>
<td>0.45</td>
<td>0.50651</td>
<td>0.60089</td>
</tr>
<tr>
<td>0.50</td>
<td>0.54013</td>
<td>0.63807</td>
</tr>
<tr>
<td>0.55</td>
<td>0.57374</td>
<td>0.67894</td>
</tr>
<tr>
<td>0.60</td>
<td>0.60804</td>
<td>0.72474</td>
</tr>
<tr>
<td>0.65</td>
<td>0.64381</td>
<td>0.77673</td>
</tr>
<tr>
<td>0.70</td>
<td>0.68205</td>
<td>0.83637</td>
</tr>
<tr>
<td>0.75</td>
<td>0.72414</td>
<td>0.90582</td>
</tr>
<tr>
<td>0.80</td>
<td>0.77233</td>
<td>0.98876</td>
</tr>
<tr>
<td>0.85</td>
<td>0.83067</td>
<td>1.09242</td>
</tr>
<tr>
<td>0.90</td>
<td>0.90816</td>
<td>1.23343</td>
</tr>
<tr>
<td>0.91</td>
<td>0.92766</td>
<td>1.26931</td>
</tr>
<tr>
<td>0.92</td>
<td>0.94922</td>
<td>1.30912</td>
</tr>
<tr>
<td>0.93</td>
<td>0.97339</td>
<td>1.35391</td>
</tr>
<tr>
<td>0.94</td>
<td>1.00100</td>
<td>1.40523</td>
</tr>
<tr>
<td>0.95</td>
<td>1.03332</td>
<td>1.46546</td>
</tr>
<tr>
<td>0.96</td>
<td>1.07245</td>
<td>1.53864</td>
</tr>
<tr>
<td>0.97</td>
<td>1.12237</td>
<td>1.63228</td>
</tr>
<tr>
<td>0.98</td>
<td>1.19200</td>
<td>1.76329</td>
</tr>
<tr>
<td>0.99</td>
<td>1.30980</td>
<td>1.98569</td>
</tr>
</tbody>
</table>
Output 94.1.6  Probit Analysis with Logistic Distribution

The Probit Procedure

<table>
<thead>
<tr>
<th>Probability</th>
<th>Dose</th>
<th>95% Fiducial Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.58945</td>
<td>0.10607 1.10241</td>
</tr>
<tr>
<td>0.02</td>
<td>0.77312</td>
<td>0.17718 1.33058</td>
</tr>
<tr>
<td>0.03</td>
<td>0.90757</td>
<td>0.23978 1.48817</td>
</tr>
<tr>
<td>0.04</td>
<td>1.01813</td>
<td>0.29773 1.61327</td>
</tr>
<tr>
<td>0.05</td>
<td>1.11413</td>
<td>0.35266 1.71922</td>
</tr>
<tr>
<td>0.06</td>
<td>1.20018</td>
<td>0.40546 1.81244</td>
</tr>
<tr>
<td>0.07</td>
<td>1.27896</td>
<td>0.45670 1.89654</td>
</tr>
<tr>
<td>0.08</td>
<td>1.35218</td>
<td>0.50675 1.97379</td>
</tr>
<tr>
<td>0.09</td>
<td>1.42100</td>
<td>0.55588 2.04572</td>
</tr>
<tr>
<td>0.10</td>
<td>1.48625</td>
<td>0.60430 2.11339</td>
</tr>
<tr>
<td>0.15</td>
<td>1.77656</td>
<td>0.84038 2.41030</td>
</tr>
<tr>
<td>0.20</td>
<td>2.03199</td>
<td>1.07379 2.66961</td>
</tr>
<tr>
<td>0.25</td>
<td>2.27043</td>
<td>1.31046 2.91416</td>
</tr>
<tr>
<td>0.30</td>
<td>2.50152</td>
<td>1.55393 3.15736</td>
</tr>
<tr>
<td>0.35</td>
<td>2.73172</td>
<td>1.80652 3.40996</td>
</tr>
<tr>
<td>0.40</td>
<td>2.96627</td>
<td>2.06957 3.68292</td>
</tr>
<tr>
<td>0.45</td>
<td>3.21006</td>
<td>2.34345 3.98927</td>
</tr>
<tr>
<td>0.50</td>
<td>3.46837</td>
<td>2.62768 4.34578</td>
</tr>
<tr>
<td>0.55</td>
<td>3.74746</td>
<td>2.92138 4.77466</td>
</tr>
<tr>
<td>0.60</td>
<td>4.05546</td>
<td>3.22451 5.30573</td>
</tr>
<tr>
<td>0.65</td>
<td>4.40366</td>
<td>3.53961 5.98041</td>
</tr>
<tr>
<td>0.70</td>
<td>4.80891</td>
<td>3.87391 6.86079</td>
</tr>
<tr>
<td>0.75</td>
<td>5.29836</td>
<td>4.24155 8.05044</td>
</tr>
<tr>
<td>0.80</td>
<td>5.92009</td>
<td>4.66820 9.74455</td>
</tr>
<tr>
<td>0.85</td>
<td>6.77126</td>
<td>5.20365 12.37149</td>
</tr>
<tr>
<td>0.90</td>
<td>8.09391</td>
<td>5.96508 17.11715</td>
</tr>
<tr>
<td>0.91</td>
<td>8.46559</td>
<td>6.16800 18.59129</td>
</tr>
<tr>
<td>0.92</td>
<td>8.89644</td>
<td>6.39837 20.37592</td>
</tr>
<tr>
<td>0.93</td>
<td>9.40575</td>
<td>6.66469 22.58957</td>
</tr>
<tr>
<td>0.94</td>
<td>10.02317</td>
<td>6.97977 25.42292</td>
</tr>
<tr>
<td>0.95</td>
<td>10.79732</td>
<td>7.36428 29.20549</td>
</tr>
<tr>
<td>0.96</td>
<td>11.81534</td>
<td>7.85438 34.56521</td>
</tr>
<tr>
<td>0.97</td>
<td>13.25466</td>
<td>8.52173 42.88232</td>
</tr>
<tr>
<td>0.98</td>
<td>15.55972</td>
<td>9.53941 57.98207</td>
</tr>
<tr>
<td>0.99</td>
<td>20.40815</td>
<td>11.52549 96.75820</td>
</tr>
</tbody>
</table>
The PLOT=PREDPLOT option together with the ODS GRAPHICS statement creates the plot of observed and fitted probabilities in Output 94.1.7. The dashed line represent pointwise confidence bands for the probabilities.

Output 94.1.7  Plot of Observed and Fitted Probabilities
Example 94.2: Multilevel Response

In this example, two preparations, a standard preparation and a test preparation, are each given at several dose levels to groups of insects. The symptoms are recorded for each insect within each group, and two multilevel probit models are fit. Because the natural sort order of the three levels is not the same as the response order, the ORDER=DATA response variable option is specified in the MODEL statement to get the desired order.

The following statements fit two models:

```sas
data multi;
  input Prep $ Dose Symptoms $ N;
  LDose=log10(Dose);
  if Prep='test' then PrepDose=LDose;
  else PrepDose=0;
  datalines;
  stand  10   None    33
  stand  10   Mild    7
  stand  10   Severe  10
  stand  20   None    17
  stand  20   Mild    13
  stand  20   Severe  17
  stand  30   None    14
  stand  30   Mild    3
  stand  30   Severe  28
  stand  40   None    9
  stand  40   Mild    8
  stand  40   Severe  32
  test   10   None    44
  test   10   Mild    6
  test   10   Severe  0
  test   20   None    32
  test   20   Mild    10
  test   20   Severe  12
  test   30   None    23
  test   30   Mild    7
  test   30   Severe  21
  test   40   None    16
  test   40   Mild    6
  test   40   Severe  19
;
proc probit data=multi;
  class Prep;
  nonpara: model Symptoms(order=data)=Prep LDose PrepDose / lackfit;
  weight N;
run;

proc probit data=multi;
  class Prep;
  parallel: model Symptoms(order=data)=Prep LDose / lackfit;
  weight N;
run;
```
Results of these two models are shown in Output 94.2.1 and Output 94.2.2. The first model allows for nonparallelism between the dose response curves for the two preparations by inclusion of an interaction between Prep and LDose. The interaction term is labeled PrepDose in the “Analysis of Parameter Estimates” table. The results of this first model indicate that the parameter for the interaction term is not significant, having a Wald chi-square of 0.73. Also, since the first model is a generalization of the second, a likelihood ratio test statistic for this same parameter can be obtained by multiplying the difference in log likelihoods between the two models by 2. The value obtained, $2 \times (-345.94 - (-346.31))$, is 0.73. This is in close agreement with the Wald chi-square from the first model. The lack-of-fit test statistics for the two models do not indicate a problem with either fit.

**Output 94.2.1** Multilevel Response: Nonparallel Analysis

### The Probit Procedure

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Weight Variable</td>
</tr>
<tr>
<td>Number of Observations</td>
</tr>
<tr>
<td>Name of Distribution</td>
</tr>
<tr>
<td>Log Likelihood</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class Level Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>Prep</td>
</tr>
<tr>
<td>Symptoms</td>
</tr>
</tbody>
</table>

### Analysis of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>3.8080</td>
<td>0.6252</td>
<td>2.5827 - 5.0333</td>
<td>37.10</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept2</td>
<td>1</td>
<td>0.4684</td>
<td>0.0559</td>
<td>0.3589 - 0.5780</td>
<td>70.19</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Prep</td>
<td>1</td>
<td>-1.2573</td>
<td>0.8190</td>
<td>-2.8624 - 0.3479</td>
<td>2.36</td>
<td>0.1247</td>
</tr>
<tr>
<td>Prep</td>
<td>0</td>
<td>0.0000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LDose</td>
<td>1</td>
<td>-2.1512</td>
<td>0.3909</td>
<td>-2.9173 -1.3851</td>
<td>30.29</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>PrepDose</td>
<td>1</td>
<td>-0.5072</td>
<td>0.5945</td>
<td>-1.6724 0.6580</td>
<td>0.73</td>
<td>0.3935</td>
</tr>
</tbody>
</table>

**Output 94.2.2** Multilevel Response: Parallel Analysis

### The Probit Procedure

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Weight Variable</td>
</tr>
<tr>
<td>Number of Observations</td>
</tr>
<tr>
<td>Name of Distribution</td>
</tr>
<tr>
<td>Log Likelihood</td>
</tr>
</tbody>
</table>
The negative coefficient associated with LDose indicates that the probability of having no symptoms (Symptoms='None') or no or mild symptoms (Symptoms='None' or Symptoms='Mild') decreases as LDose increases; that is, the probability of a severe symptom increases with LDose. This association is apparent for both treatment groups.

The negative coefficient associated with the standard treatment group (Prep = stand) indicates that the standard treatment is associated with more severe symptoms across all Ldose values.
The following statements use the PLOTS= option to create the plot shown in Output 94.2.3 and Output 94.2.4. Output 94.2.3 is the plot of the probabilities of the response taking on individual levels as a function of LDose. Since there are two covariates, LDose and Prep, the value of the classification variable Prep is fixed at the highest level, test. Instead of individual response level probabilities, the CDFPLOT option creates the plot of the cumulative response probabilities with confidence limits shown in Output 94.2.4.

```sas
proc probit data=multi
   plots=(predpplot(level=("None" "Mild" "Severe"))
         cdfplot(level=("None" "Mild" "Severe")));
   class Prep;
   parallel: model Symptoms(order=data)=Prep LDose / lackfit;
   weight N;
run;
```

**Output 94.2.3** Plot of Predicted Probabilities for the Test Preparation Group
Example 94.2: Multilevel Response

Output 94.2.4 Plot of Predicted Cumulative Probabilities for the Test Preparation Group

The following statements use the XDATA= data set to create plots of predicted probabilities and cumulative probabilities with Prep set to the stand level. The resulting plots are shown in Output 94.2.5 and Output 94.2.6.

data xrow;
   input Prep $ Dose Symptoms $ N;
   LDose=log10(Dose);
datalines;
   stand  40   Severe   32
run;

proc probit data=multi xdata=xrow
   plots=(predpplot(level=("None" "Mild" "Severe"))
           cdfplot(level=("None" "Mild" "Severe")));
   class Prep;
   parallel: model Symptoms(order=data)=Prep LDose / lackfit;
   weight N;
run;
Output 94.2.5  Plot of Predicted Probabilities for the Standard Preparation Group

Output 94.2.6  Plot of Predicted Cumulative Probabilities for the Standard Preparation Group
Example 94.3: Logistic Regression and Scoring New Data

In this example, a series of people are asked whether or not they would subscribe to a new newspaper. For each person, the variables sex (Female, Male), age, and subs (1=yes, 0=no) are recorded. The PROBIT procedure is used to fit a logistic regression model to the probability of subscribing (subs = 1) as a function of the variables sex and age. Specifically, the probability of subscribing is modeled as

\[ p = \Pr(\text{subs} = 1) = F(b_0 + b_1 \times \text{sex} + b_2 \times \text{age}) \]

where \( F \) is the cumulative logistic distribution function.

By default, the PROBIT procedure models the probability of the lower response level. The following statements format the values of subs as 1 = ‘accept’ and 0 = ‘reject’, and model \( \Pr(\text{subs} = 1) \) by using the EVENT= response variable option. The STORE statement saves the fitted model in an item store named LogitModel. The results are shown in Output 94.3.1.

data news;
    input sex $ age subs @@;
datalines;
Female 35 0 Male 44 0
Male 45 1 Female 47 1
Female 51 0 Female 47 0
Male 54 1 Male 47 1
Female 35 0 Female 34 0
Female 48 0 Female 56 1
Male 46 1 Female 59 1
Female 46 1 Male 59 1
Male 38 1 Female 39 0
Male 49 1 Male 42 1
Male 50 1 Female 45 0
Female 47 0 Female 30 1
Female 39 0 Female 51 0
Female 45 0 Female 43 1
Male 39 1 Male 31 0
Female 39 0 Male 34 0
Female 52 1 Female 46 0
Male 58 1 Female 50 1
Female 32 0 Female 52 1
Female 35 0 Female 51 0;

proc format;
    value subscrib 1 = 'accept' 0 = 'reject';
run;

proc probit data=news;
    class sex;
    model subs(event="accept")=sex age / d=logistic itprint;
    format subs subscrib.;
    store out=LogitModel;
run;
Output 94.3.1 shows that there appears to be an effect due to both the variables sex and age. The positive coefficient for age indicates that older people are more likely to subscribe than younger people. The negative coefficient for sex indicates that females are less likely to subscribe than males.
You can use the SCORE statement in the PLM procedure to score new observations based on the fitted model saved by the STORE statement above. For example, to compute the probability of subscribing for one new observation with sex = ‘Female’ and age = 35 in the data set test, you can use the following statements:

```plaintext
data test;
  input sex $ age;
datalines;
Female 35;

proc plm restore=LogitModel;
  score data=test out=testout predicted / ilink;
run;

proc print data=testout;
run;
```

The ILINK option in the SCORE statement applies the inverse of the logit link to provide an estimate on the mean (probability) scale. **Output 94.3.2** shows the predicted probability for the new observation.

**Output 94.3.2** Predicted Probability for One New Observation

<table>
<thead>
<tr>
<th>Obs</th>
<th>sex</th>
<th>age</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>35</td>
<td>0.082205</td>
</tr>
</tbody>
</table>

**Example 94.4: An Epidemiology Study**

The data in this example, which are from an epidemiology study, consist of five variables: the number, r, of individuals surviving after an epidemic, out of n treated, for combinations of medicine dosage (dose), treatment (treat = A, B), and sex (sex = 0(Female), 1(Male)).

To see whether the two treatments have different effects on male and female individual survival rates, the interaction term between the two variables treat and sex is included in the model.

The following invocation of PROC PROBIT fits the binary probit model to the grouped data:

```plaintext
data epidemic;
  input treat$ dose n r sex @@;
  label dose = Dose;
datalines;
A 2.17 142 142 0 A .57 132 47 1
A 1.68 128 105 1 A 1.08 126 100 0
A 1.79 125 118 0 B 1.66 117 115 1
B 1.49 127 114 0 B 1.17 51 44 1
B 2.00 127 126 0 B .80 129 100 1;

data xval;
  input treat $ dose sex;
datalines;
B 2. 1;
```

The data in this example, which are from an epidemiology study, consist of five variables: the number, r, of individuals surviving after an epidemic, out of n treated, for combinations of medicine dosage (dose), treatment (treat = A, B), and sex (sex = 0(Female), 1(Male)).

To see whether the two treatments have different effects on male and female individual survival rates, the interaction term between the two variables treat and sex is included in the model.

The following invocation of PROC PROBIT fits the binary probit model to the grouped data:
Chapter 94: The PROBIT Procedure

proc probit optc lackfit covout data=epidemic
  outest = out1 xdata = xval
  Plots=(predpplot ippplot lpredplot);
class treat sex;
model r/n = dose treat sex sex*treat/corrb covb inversecl;
output out = out2 p =p;
run;

The results of this analysis are shown in the outputs that follow.

Output 94.4.1 displays the table of level information for all classification variables in the CLASS statement.

Output 94.4.1  Class Level Information

The Probit Procedure

<table>
<thead>
<tr>
<th>Class Level Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>treat</td>
</tr>
<tr>
<td>sex</td>
</tr>
</tbody>
</table>

Output 94.4.2 displays the table of parameter information for the effects in the MODEL statement.

Output 94.4.2  Parameter Information

<table>
<thead>
<tr>
<th>Parameter Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>intercept</td>
</tr>
<tr>
<td>dose</td>
</tr>
<tr>
<td>treatA</td>
</tr>
<tr>
<td>treatB</td>
</tr>
<tr>
<td>sex0</td>
</tr>
<tr>
<td>sex1</td>
</tr>
<tr>
<td>treatAsex0</td>
</tr>
<tr>
<td>treatAsex1</td>
</tr>
<tr>
<td>treatBsex0</td>
</tr>
<tr>
<td>treatBsex1</td>
</tr>
</tbody>
</table>

Output 94.4.3 displays background information about the model fit. Included are the name of the input data set, the response variables used, the numbers of observations, events, and trials, the type of distribution, and the final value of the log-likelihood function.
Output 94.4.3 Model Information

The Probit Procedure

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Events Variable</td>
</tr>
<tr>
<td>Trials Variable</td>
</tr>
<tr>
<td>Number of Observations</td>
</tr>
<tr>
<td>Number of Events</td>
</tr>
<tr>
<td>Number of Trials</td>
</tr>
<tr>
<td>Name of Distribution</td>
</tr>
<tr>
<td>Log Likelihood</td>
</tr>
</tbody>
</table>

Output 94.4.4 displays the table of goodness-of-fit tests requested with the LACKFIT option in the PROC PROBIT statement. Two goodness-of-fit statistics, the Pearson’s chi-square statistic and the likelihood ratio chi-square statistic, are computed. The grouping method for computing these statistics can be specified by the AGGREGATE= option. The details can be found in the AGGREGATE= option, and an example can be found in the second part of this example. By default, the PROBIT procedure uses the covariates in the MODEL statement to do grouping. Observations with the same values of the covariates in the MODEL statement are grouped into cells and the two statistics are computed according to these cells. The total number of cells and the number of levels for the response variable are reported next in the “Response-Covariate Profile.”

In this example, neither the Pearson’s chi-square nor the log-likelihood ratio chi-square tests are significant at the 0.1 level, which is the default test level used by the PROBIT procedure. That means that the model, which includes the interaction of treat and sex, is suitable for this epidemiology data set. (Further investigation shows that models without the interaction of treat and sex are not acceptable by either test.)

Output 94.4.4 Goodness-of-Fit Tests and Response-Covariate Profile

<table>
<thead>
<tr>
<th>Goodness-of-Fit Tests</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>4.9317</td>
<td>4</td>
<td>1.2329</td>
<td>0.2944</td>
</tr>
<tr>
<td>L.R. Chi-Square</td>
<td>5.7079</td>
<td>4</td>
<td>1.4270</td>
<td>0.2220</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response-Covariate Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Levels</td>
</tr>
<tr>
<td>Number of Covariate Values</td>
</tr>
</tbody>
</table>
Output 94.4.5 displays the Type III test results for all effects specified in the MODEL statement, which include the degrees of freedom for the effect, the Wald Chi-Square test statistic, and the \( p \)-value.

\begin{verbatim}
Output 94.4.5 Type III Tests

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>dose</td>
<td>1</td>
<td>42.1691</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>treat</td>
<td>1</td>
<td>16.1421</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>sex</td>
<td>1</td>
<td>1.7710</td>
<td>0.1833</td>
</tr>
<tr>
<td>treat*sex</td>
<td>1</td>
<td>13.9343</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
\end{verbatim}

Output 94.4.6 displays the table of parameter estimates for the model. The PROBIT procedure displays information for all the parameters of an effect. Degenerate parameters are indicated by 0 degree of freedom. Confidence intervals are computed for all parameters with nonzero degrees of freedom, including the natural threshold \( C \) if the OPTC option is specified in the PROC PROBIT statement. The confidence level can be specified by the ALPHA= option in the MODEL statement. The default confidence level is 95%.

\begin{verbatim}
Output 94.4.6 Analysis of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.8871</td>
<td>0.3632</td>
<td>-1.5991 -0.1752</td>
<td>5.96</td>
<td>0.0146</td>
</tr>
<tr>
<td>dose</td>
<td>1</td>
<td>1.6774</td>
<td>0.2583</td>
<td>1.1711 2.1837</td>
<td>42.17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>treat</td>
<td>A</td>
<td>-1.2537</td>
<td>0.2616</td>
<td>-1.7664 -0.7410</td>
<td>22.97</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>treat</td>
<td>B</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>sex</td>
<td>0</td>
<td>-0.4633</td>
<td>0.2289</td>
<td>-0.9119 -0.0147</td>
<td>4.10</td>
<td>0.0429</td>
</tr>
<tr>
<td>sex</td>
<td>1</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>treat*sex</td>
<td>A</td>
<td>1.2899</td>
<td>0.3456</td>
<td>0.6126 1.9672</td>
<td>13.93</td>
<td>0.0002</td>
</tr>
<tr>
<td>treat*sex</td>
<td>A</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>treat*sex</td>
<td>B</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>treat*sex</td>
<td>B</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0.2735</td>
<td>0.0946</td>
<td>0.0881 0.4589</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
\end{verbatim}

From Table 94.4.6, you can see the following results:

- The variable \textit{dose} has a significant positive effect on the survival rate.
- Individuals under treatment A have a lower survival rate.
- Male individuals have a higher survival rate.
- Female individuals under treatment A have a higher survival rate.
Output 94.4.7 and Output 94.4.8 display tables of estimated covariance matrix and estimated correlation matrix for estimated parameters with a nonzero degree of freedom, respectively. They are computed by the inverse of the Hessian matrix of the estimated parameters.

### Output 94.4.7 Estimated Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>dose</th>
<th>treatA</th>
<th>sex0</th>
<th>treatAsex0</th>
<th><em>C</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.13194</td>
<td>-0.087353</td>
<td>0.053551</td>
<td>0.030285</td>
<td>-0.067056</td>
<td>-0.028073</td>
</tr>
<tr>
<td>dose</td>
<td>-0.087353</td>
<td>0.066723</td>
<td>-0.047506</td>
<td>-0.034081</td>
<td>0.058620</td>
<td>0.018196</td>
</tr>
<tr>
<td>treatA</td>
<td>0.053551</td>
<td>-0.047506</td>
<td>0.068425</td>
<td>0.036063</td>
<td>-0.075323</td>
<td>-0.017084</td>
</tr>
<tr>
<td>sex0</td>
<td>0.030285</td>
<td>-0.034081</td>
<td>0.036063</td>
<td>0.052383</td>
<td>-0.063599</td>
<td>-0.008088</td>
</tr>
<tr>
<td>treatAsex0</td>
<td>-0.067056</td>
<td>0.058620</td>
<td>-0.075323</td>
<td>-0.063599</td>
<td>0.119408</td>
<td>0.019134</td>
</tr>
<tr>
<td><em>C</em></td>
<td>-0.028073</td>
<td>0.018196</td>
<td>-0.017084</td>
<td>-0.008088</td>
<td>0.019134</td>
<td>0.008948</td>
</tr>
</tbody>
</table>

### Output 94.4.8 Estimated Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>dose</th>
<th>treatA</th>
<th>sex0</th>
<th>treatAsex0</th>
<th><em>C</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.000000</td>
<td>-0.930998</td>
<td>0.563595</td>
<td>0.364284</td>
<td>-0.534227</td>
<td>-0.817027</td>
</tr>
<tr>
<td>dose</td>
<td>-0.930998</td>
<td>1.000000</td>
<td>-0.703083</td>
<td>-0.576477</td>
<td>0.656744</td>
<td>0.744699</td>
</tr>
<tr>
<td>treatA</td>
<td>0.563595</td>
<td>-0.703083</td>
<td>1.000000</td>
<td>0.602359</td>
<td>-0.833299</td>
<td>-0.690420</td>
</tr>
<tr>
<td>sex0</td>
<td>0.364284</td>
<td>-0.576477</td>
<td>0.602359</td>
<td>1.000000</td>
<td>-0.804154</td>
<td>-0.373565</td>
</tr>
<tr>
<td>treatAsex0</td>
<td>-0.534227</td>
<td>0.656744</td>
<td>-0.833299</td>
<td>-0.804154</td>
<td>1.000000</td>
<td>0.585364</td>
</tr>
<tr>
<td><em>C</em></td>
<td>-0.817027</td>
<td>0.744699</td>
<td>-0.690420</td>
<td>-0.373565</td>
<td>0.585364</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Output 94.4.9 displays the computed values and fiducial limits for the first single continuous variable `dose` in the MODEL statement, given the probability levels, without the effect of the natural threshold, and when the option INVERSECL in the MODEL statement is specified. If there is no single continuous variable in the MODEL specification but the INVERSECL option is specified, an error is reported.
If the XDATA= option is used to input a data set for the independent variables in the MODEL statement, the PROBIT procedure uses these values for the independent variables other than the single continuous variable. Missing values are not permitted in the XDATA= data set for the independent variables, although the value for the single continuous variable is not used in the computing of the fiducial limits. A suitable valid value should be given. In the data set xval created by the SAS statements on page 7663, dose = 2. Only one observation from the XDATA= data set is used to produce a probit analysis table for a combination of classification variable levels. If more than one observation is present in the XDATA= data set, only the last observation is used.
See the section “XDATA= SAS-data-set” on page 7639 for the default values for those effects other than the single continuous variable, for which the fiducial limits are computed.

In this example, there are two classification variables, `treat` and `sex`. Fiducial limits for the `dose` variable are computed for the highest level of the classification variables, `treat = B` and `sex = 1`, which is the default specification. Since these are the default values, you would get the same values and fiducial limits if you did not specify the `XDATA=` option in this example. The confidence level for the fiducial limits can be specified by the `ALPHA=` option in the `MODEL` statement. The default level is 95%.

If a LOG10 or LOG option is used in the `PROC PROBIT` statement, the values and the fiducial limits are computed for both the single continuous variable and its logarithm.

Output 94.4.10 displays the `OUTEST=` data set. All parameters for an effect are included. The name of a parameter is generated by combining the variable names and levels in the effect. The maximum length of a parameter name is 32.

**Output 94.4.10 Outest Data Set for Epidemiology Study**

```plaintext
Obs _MODEL_ _NAME_ _TYPE_ _DIST_ _STATUS_ _LNLIKE_ r Intercept dose
1 r PARMS Normal 0 Converged -387.247 -1.00000 -0.88714 1.67739
2 Intercept COV Normal 0 Converged -387.247 -0.88714 0.13194 -0.08735
3 dose COV Normal 0 Converged -387.247 1.67739 -0.08735 0.06672
4 treatA COV Normal 0 Converged -387.247 -1.25367 0.05355 -0.04751
5 treatB COV Normal 0 Converged -387.247 0.00000 0.00000 0.00000
6 sex0 COV Normal 0 Converged -387.247 -0.46329 0.03029 -0.03408
7 sex1 COV Normal 0 Converged -387.247 0.00000 0.00000 0.00000
8 treatAsex0 COV Normal 0 Converged -387.247 1.28991 -0.06706 0.05862
9 treatAsex1 COV Normal 0 Converged -387.247 0.00000 0.00000 0.00000
10 treatBsex0 COV Normal 0 Converged -387.247 0.00000 0.00000 0.00000
11 treatBsex1 COV Normal 0 Converged -387.247 0.00000 0.00000 0.00000
12 _C_ COV Normal 0 Converged -387.247 0.27347 -0.02807 0.01820
```

```plaintext
Obs treatA treatB sex0 sex1 treatAsex0 treatAsex1 treatBsex0 treatBsex1 _C_
1 -1.25367 0 -0.46329 0 1.28991 0 0 0 0.27347
2 0.05355 0 0.03029 0 -0.06706 0 0 0 -0.02807
3 -0.04751 0 -0.03408 0 0.05862 0 0 0 0.01820
4 0.06843 0 0.03606 0 -0.07532 0 0 0 -0.01708
5 0.00000 0 0.00000 0 0.00000 0 0 0 0.00000
6 0.03606 0 0.05238 0 -0.06360 0 0 0 -0.00809
7 0.00000 0 0.00000 0 0.00000 0 0 0 0.00000
8 -0.07532 0 -0.06360 0 0.11941 0 0 0 0.01913
9 0.00000 0 0.00000 0 0.00000 0 0 0 0.00000
10 0.00000 0 0.00000 0 0.00000 0 0 0 0.00000
11 0.00000 0 0.00000 0 0.00000 0 0 0 0.00000
12 -0.01708 0 -0.00809 0 0.01913 0 0 0 0.00895
```
The plots in the following three outputs, Output 94.4.11, Output 94.4.12, and Output 94.4.13, are generated by the PLOTS= option. The first plot, specified with the PREDPLOT option, is the plot of the predicted probability against the first single continuous variable dose in the MODEL statement. You can specify values of other independent variables in the MODEL statement by using an XDATA= data set or by using the default values.

The second plot, specified with the IPPPLOT option, is the inverse of the predicted probability plot with the fiducial limits. It should be pointed out that the fiducial limits are not just the inverse of the confidence limits in the predicted probability plot; see the section “Inverse Confidence Limits” on page 7638 for the computation of these limits. The third plot, specified with the LPREDPLOT option, is the plot of the linear predictor $\mathbf{x}'\beta$ against the first single continuous variable with the Wald confidence intervals.

Output 94.4.11 Predicted Probability Plot
Example 94.4: An Epidemiology Study

Output 94.4.12 Inverse Predicted Probability Plot

Output 94.4.13 Linear Predictor Plot
Example 94.5: Model Postfitting Analysis

Recall the previous example of an epidemic study, in which the `treat*sex` interaction is statistically significant. Suppose you want to know whether such an effect is the same at different levels of the two categorical variables.

The following SAS statements fit a probit model and use the SLICE statement to request analysis of the two-way interaction term `treat*sex`:

```sas
proc probit data=epidemic;
  class treat sex;
  model r/n = dose treat sex treat*sex;
  slice treat*sex / diff;
  effectplot;
run;
```

Output 94.5.1 displays the test results for the interaction effect. As you can see, the difference between the two treatments is not significant among females.

Output 94.5.1 Tests Conditional on treat*sex

<table>
<thead>
<tr>
<th>The Probit Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square Test for treat*sex Least Squares Means Slice</td>
</tr>
<tr>
<td>Slice</td>
</tr>
<tr>
<td>treat A</td>
</tr>
<tr>
<td>Chi-Square Test for treat*sex Least Squares Means Slice</td>
</tr>
<tr>
<td>Slice</td>
</tr>
<tr>
<td>treat B</td>
</tr>
<tr>
<td>Chi-Square Test for treat*sex Least Squares Means Slice</td>
</tr>
<tr>
<td>Slice</td>
</tr>
<tr>
<td>sex 0</td>
</tr>
<tr>
<td>Chi-Square Test for treat*sex Least Squares Means Slice</td>
</tr>
<tr>
<td>Slice</td>
</tr>
<tr>
<td>sex 1</td>
</tr>
</tbody>
</table>
The DIFF option computes effect differences between groups within the same slice. Results are displayed in Output 94.5.2.

**Output 94.5.2** Effect Differences Conditional on treat*sex

<table>
<thead>
<tr>
<th>Slice</th>
<th>sex</th>
<th>_sex</th>
<th>Estimate</th>
<th>Error</th>
<th>z Value</th>
<th>Pr &gt;</th>
<th>z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>treat A</td>
<td>0</td>
<td>1</td>
<td>0.5957</td>
<td>0.1384</td>
<td>4.30</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slice</th>
<th>sex</th>
<th>_sex</th>
<th>Estimate</th>
<th>Error</th>
<th>z Value</th>
<th>Pr &gt;</th>
<th>z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>treat B</td>
<td>0</td>
<td>1</td>
<td>-0.2956</td>
<td>0.1816</td>
<td>-1.63</td>
<td>0.1035</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slice</th>
<th>treat</th>
<th>_treat</th>
<th>Estimate</th>
<th>Error</th>
<th>z Value</th>
<th>Pr &gt;</th>
<th>z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sex 0</td>
<td>A</td>
<td>B</td>
<td>-0.00899</td>
<td>0.1702</td>
<td>-0.05</td>
<td>0.9579</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slice</th>
<th>treat</th>
<th>_treat</th>
<th>Estimate</th>
<th>Error</th>
<th>z Value</th>
<th>Pr &gt;</th>
<th>z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sex 1</td>
<td>A</td>
<td>B</td>
<td>-0.9003</td>
<td>0.1307</td>
<td>-6.89</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The EFFECTPLOT statement produces a predicted probability plot for dose by the four groups that are formed by the treat*sex interaction. The plot is displayed in Output 94.5.3. The two overlapping curves represent the two treatment groups for females, suggesting no treatment effect. It appears that males tend to respond to the two treatments differently: those on treatment B have a better survival rate, and those on treatment A have a worse chance of survival.
Output 94.5.3  Predicted Probability versus Dose Level by treat*sex

References


Subject Index

annotating
   CDF plots, 7586
   IPP plots, 7600
   LPRED plots, 7608
   predicted probability plots, 7626

biological assay data, 7572, 7638

CDF plots
   annotating, 7586
   axes, color, 7586
   font, specifying, 7587
   options summarized by function, 7584, 7605
   reference lines, options, 7587–7590
   threshold lines, options, 7590

CDFPLOT
   PROBIT procedure, 7583
   cumulative distribution function, 7572

deviance
   PROBIT procedure, 7621, 7637

dispersion parameter
   PROBIT procedure, 7637

distributions
   Gompertz, 7572
   logistic, 7572
   normal, 7572

extreme value distribution
   PROBIT procedure, 7634

fiducial limits, 7579, 7580, 7636

Gompertz distribution, 7572

graphics catalog, specifying
   PROBIT procedure, 7579

information matrix, 7633

INSET
   PROBIT procedure, 7595

inverse confidence limits
   PROBIT procedure, 7579, 7638

IPP plots
   annotating, 7600
   axes, color, 7600
   font, specifying, 7601
   options summarized by function, 7597
   reference lines, options, 7600–7602, 7604
   threshold lines, options, 7603

IPPPLOT
   PROBIT procedure, 7597

lack of fit tests, 7580, 7636

likelihood function, 7633

likelihood ratio chi-square test, 7634, 7636

log-likelihood function
   PROBIT procedure, 7633

logistic distribution, 7572
   PROBIT procedure, 7634

logistic regression, 7572

LPRED plots
   annotating, 7608
   axes, color, 7608
   font, specifying, 7609
   options summarized by function, 7608–7610, 7612
   reference lines, options, 7608–7610, 7612
   threshold lines, options, 7611

LPREDPLOT
   PROBIT procedure, 7605

missing values
   PROBIT procedure, 7632

multilevel response, 7635

natural response rate, 7572, 7580

Newton-Raphson algorithm
   PROBIT procedure, 7633

normal distribution, 7572
   PROBIT procedure, 7634

ODS Graphics names
   PROBIT procedure, 7644

options summary
   ESTIMATE statement, 7594

output table names
   PROBIT procedure, 7643

overdispersion
   PROBIT procedure, 7621

Pearson’s chi-square
   PROBIT procedure, 7618, 7621, 7637

Pearson’s chi-square test, 7579, 7636

predicted probability plots
   annotating, 7626
   axes, color, 7626
   font, specifying, 7627
   options summarized by function, 7623
   reference lines, options, 7626–7628, 7630
   threshold lines, options, 7629
PREDPPLOT
  PROBIT procedure, 7622
probit analysis
  insets, 7595
probit equation, 7572, 7635
PROBIT procedure
  Abbot's formula, 7632
  binary response data, 7572, 7573, 7635
  CDFPLOT, 7583
  deviance, 7621, 7637
  deviance statistic, 7636
  dispersion parameter, 7637
  extreme value distribution, 7634
  goodness-of-fit, 7618, 7621
  goodness-of-fit tests, 7579, 7580, 7618, 7636
  INSET, 7595
  inverse confidence limits, 7638
  IPPPLOT, 7597
  log-likelihood function, 7633
  logistic distribution, 7634
  LPREDPLOT, 7605
  maximum likelihood estimates, 7572
  missing values, 7632
  models, 7635
  multilevel response data, 7572, 7573, 7635
  natural response rate, 7573
  Newton-Raphson algorithm, 7633
  normal distribution, 7634
  ODS Graphics names, 7644
  ordering of effects, 7580
  output table names, 7643
  overdispersion, 7621
  Pearson's chi-square, 7618, 7621, 7636, 7637
  PREDPPLOT, 7622
  response level ordering, 7617
  subpopulation, 7572, 7621, 7637
  threshold response rate, 7573
  tolerance distribution, 7637

quantal response data, 7572

response level ordering
  PROBIT procedure, 7617

subpopulation
  PROBIT procedure, 7618, 7621, 7637

threshold response rate, 7572

Wald test
  PROBIT procedure, 7634
Syntax Index

AGGREGATE= option
  MODEL statement (PROBIT), 7618

BY statement
  PROBIT procedure, 7582

CDFPLOT statement, see PROBIT procedure, CDFPLOT statement, see PROBIT procedure, CDFPLOT statement options summarized by function, 7584 PROBIT procedure, 7583

CLASS statement
  PROBIT procedure, 7591

COVOUT option
  PROC PROBIT statement, 7578

CPREFIX= option
  CLASS statement (PROBIT), 7591

DATA= option
  PROC PROBIT statement, 7579

DESCENDING option
  CLASS statement (PROBIT), 7591
  MODEL statement, 7617

EFFECTPLOT statement
  PROBIT procedure, 7593

ESTIMATE statement
  PROBIT procedure, 7594

EVENT= option
  MODEL statement, 7617

GOUT= option
  PROC PROBIT statement, 7579

HPROB= option
  PROC PROBIT statement, 7579

INEST= option
  PROC PROBIT statement, 7579

INSET statement
  PROC PROBIT procedure, 7595

INVERSECL option
  PROC PROBIT statement, 7579

IPPLOT statement
  options summarized by function, 7597
  PROC PROBIT procedure, 7597

LACKFIT option
  PROC PROBIT statement, 7580

LOG option
  PROC PROBIT statement, 7580
  LOG10 option
  PROC PROBIT statement, 7580

LPREDPLOT statement
  options summarized by function, 7605
  PROC PROBIT procedure, 7605

LPREFIX= option
  CLASS statement (PROBIT), 7592

LSMEANS statement
  PROBIT procedure, 7613

LSMEANSTIMATE statement
  PROBIT procedure, 7614

MISSING option
  CLASS statement (PROBIT), 7592

NAMELEN= option
  PROC PROBIT statement, 7580

NOPRINT option
  PROC PROBIT statement, 7580

OPTC option
  PROC PROBIT statement, 7578, 7580

options
  CDFPLOT statement (PROBIT), 7584
  IPPPLOT statement (PROBIT), 7597
  LPREDPLOT statement (PROBIT), 7605
  PREDPLOT statement (PROBIT), 7623

ORDER= option
  CLASS statement (PROBIT), 7592
  MODEL statement, 7617
  PROC PROBIT statement, 7580

OUTEST= option
  PROC PROBIT statement, 7581

PARAM= option
  CLASS statement (PROBIT), 7592

PLOT= option
  PROC PROBIT statement, 7581

PREDPLOT statement
  options summarized by function, 7623
  PROC PROBIT procedure, 7622

PROBIT, 7572

PROBIT procedure, 7572 syntax, 7577

PROBIT procedure, BY statement, 7582

PROBIT procedure, CDFPLOT statement, 7583

ANNOTATE= option, 7586

CAXIS= option, 7586
CFIT= option, 7586
CFRAME= option, 7586
CGRID= option, 7587
CHREF= option, 7587
CLABBOX= option, 7587
CTEXT= option, 7587
CVREF= option, 7587
DESCRIPTION= option, 7587
FONT= option, 7587
HAXIS= option, 7587
HEIGHT= option, 7587
HLOWER= option, 7587
HOFFSET= option, 7588
HREF= option, 7588
HREFLABELS= option, 7588
HREFLABPOS= option, 7588
HUPPER= option, 7588
LEVEL option, 7588
LFIT option, 7588
LGRID option, 7589
LHREF= option, 7589
LVREF= option, 7589
NAME= option, 7589
NHTICK= option, 7589, 7602
NOFIT option, 7589
NOFRAME option, 7589
NOGRID option, 7589
NOHLABEL option, 7589
NOHTICK option, 7589
NOTHRESH option, 7589
NOVLABEL option, 7589
NOVTICK option, 7589
NAME= option, 7589
NHTICK= option, 7589
VAXIS= option, 7590
VAXISLABEL= option, 7590
VLOWER= option, 7590
VREF= option, 7590
VREFLABELS= option, 7590
VREFLABPOS= option, 7590
VUPPER= option, 7591
WAXIS= option, 7591
WFIT= option, 7591
WGRID= option, 7591
WREFL= option, 7591
APPROBIT procedure, CDFPPLOT statement
OFFSET= option, 7590
APPROBIT procedure, CLASS statement, 7591
CPREFIX= option, 7591
DESCENDING option, 7591
LPREFIX= option, 7591
MISSING option, 7592
ORDER= option, 7592
PARAM= option, 7592
RE= option, 7592
TRUNCATE option, 7593
APPROBIT procedure, EFFECTPLOT statement, 7593
APPROBIT procedure, ESTIMATE statement, 7594
APPROBIT procedure, INSET statement, 7595
keywords, 7595
APPROBIT procedure, IPPLOT statement
INHEIGHT= option, 7602
APPROBIT procedure, IPPPLOT statement, 7597
ANNOTATE= option, 7600
CAXIS= option, 7600
CFIT= option, 7600
CFRAME= option, 7600
CGRID= option, 7600
CHREF= option, 7600
CTEXT= option, 7600
CVREF= option, 7600
DESCRIPTION= option, 7601
FONT= option, 7601
HAXIS= option, 7601
HEIGHT= option, 7601
HLOWER= option, 7601
HOFFSET= option, 7601
MREF= option, 7601
NOHREFLABELS= option, 7601
NOCONF option, 7602
NOFIT option, 7603
NOFRAME option, 7603
NOGRID option, 7603
NOHLABEL option, 7603
NOHTICK option, 7603
NOTHRESH option, 7603
NOVLABEL option, 7603
NOVTICK option, 7603
options, 7597
THRESHLABPOS= option, 7603
TURNVLABELS option, 7603
VAR= option, 7583
NODATA option, 7603
NOCONF option, 7602
NOFIT option, 7603
NOFRAME option, 7603
NOGRID option, 7603
NOHLABEL option, 7603
NOHTICK option, 7603
NOTHRESH option, 7603
NOVLABEL option, 7603
NOVTICK option, 7603
options, 7597
THRESHLABPOS= option, 7603
TURNVLABELS option, 7603
VAR= option, 7597
VAXIS= option, 7603
VAXISLABEL= option, 7604
HEIGHT= option, 7627
HLOWER= option, 7627
HOFFSET= option, 7627
HREF= option, 7627
HREFLABELS= option, 7628
HREFLABPOS= option, 7628
HUPPER= option, 7627
INBORDER option, 7628
LFIT option, 7628
LGRID option, 7628
LHREF= option, 7628
LVREF= option, 7628
NAME= option, 7629
NOCONF option, 7629
NODATA option, 7629
NOFIT option, 7629
NOFRAME option, 7629
NOGRID option, 7629
NOHLABEL option, 7629
NOHTICK option, 7629
NOTHRESH option, 7629
NOVLABEL option, 7629
NOTICK option, 7629
options, 7623
THRESHLABPOS= option, 7629
TURNLABELS option, 7629
VAR= option, 7622
VAXIS= option, 7630
VAXISLABEL= option, 7630
VLOWER= option, 7630
VOFFSET= option, 7630
VREF= option, 7630
VREFLABELS= option, 7630
VREFLABPOS= option, 7630
VUPPER= option, 7630
WAXIS= option, 7631
WFIT= option, 7631
WGRID= option, 7631
WREFL= option, 7631
PROBIT procedure, PROC PROBIT statement, 7577
COVOUT option, 7578
DATA= option, 7579
GOUT= option, 7579
HPROB= option, 7579
INEST= option, 7579
INVERSECL option, 7579
LACKFIT option, 7580
LOG option, 7580
LOG10 option, 7580
NAMELEN= option, 7580
NOPRINT option, 7580
OPTC option, 7578, 7580
ORDER= option, 7580
OUTEST= option, 7581
PLOT= option, 7581
XDATA= option, 7582
PROBIT procedure, SLICE statement, 7631
PROBIT procedure, STORE statement, 7631
PROBIT procedure, TEST statement, 7631
PROBIT procedure, WEIGHT statement, 7631
PROC PROBIT statement, see PROBIT procedure
REF= option
CLASS statement (PROBIT), 7592
REFERENCE= option
MODEL statement, 7618
SCALE= option
MODEL statement (PROBIT), 7621
SLICE statement
PROBIT procedure, 7631
STORE statement
PROBIT procedure, 7631
TEST statement
LIFEREG procedure, 7631
TRUNCATE option
CLASS statement (PROBIT), 7593
VAR= option
CDFPLOT statement (PROBIT), 7583
IPPPLOT statement (PROBIT), 7597
LPREDPLOT statement (PROBIT), 7605
PREDPPlot statement (PROBIT), 7622
XDATA= option
PROC PROBIT statement, 7582