

# **SAS/STAT<sup>®</sup> 14.2 User's Guide**

## **The PROBIT Procedure**

This document is an individual chapter from *SAS/STAT® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/STAT® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

### **SAS/STAT® 14.2 User's Guide**

Copyright © 2016, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government License Rights; Restricted Rights:** The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to <http://support.sas.com/thirdpartylicenses>.

# Chapter 94

## The PROBIT Procedure

### Contents

---

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

Displayed Output . . . . .	7641
ODS Table Names . . . . .	7643
ODS Graphics . . . . .	7643
Examples: PROBIT Procedure . . . . .	<b>7646</b>
Example 94.1: Dosage Levels . . . . .	7646
Example 94.2: Multilevel Response . . . . .	7655
Example 94.3: Logistic Regression and Scoring New Data . . . . .	7661
Example 94.4: An Epidemiology Study . . . . .	7663
Example 94.5: Model Postfitting Analysis . . . . .	7672
References . . . . .	<b>7674</b>

---

## 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

$$p = \Pr(Y = 0) = C + (1 - C)F(\mathbf{x}'\beta)$$

where

- $\beta$  is a vector of parameter estimates
- $F$  is a cumulative distribution function (normal, logistic, or extreme value)
- $\mathbf{x}$  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  $1, \dots, k, k + 1$ . For example, the severity of coronary disease can be classified into three response categories as 1=no disease, 2=angina pectoris, and 3=myocardial infraction. 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

$$\Pr(Y \leq 1 | \mathbf{x}) = F(\mathbf{x}'\boldsymbol{\beta})$$

$$\Pr(Y \leq i | \mathbf{x}) = F(\alpha_i + \mathbf{x}'\boldsymbol{\beta}), \quad 2 \leq i \leq k$$

where  $\alpha_2, \dots, \alpha_k$  are  $k - 1$  intercept parameters. By default, the covariate vector  $\mathbf{x}$  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:

```
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
;
```

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:

```
ods graphics on;

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

The `DATA=` option specifies that PROC PROBIT analyze 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

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

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**

Model Information	
Data Set	WORK.STUDY
Events Variable	Respond
Trials Variable	Number
Number of Observations	12
Number of Events	81
Number of Trials	180
Number of Events In Control Group	3
Number of Trials In Control Group	15
Name of Distribution	Normal
Log Likelihood	-104.3945783

**Figure 94.2** Model Parameter Estimates for the PROBIT Procedure

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-4.1438	1.3415	-6.7731	-1.5146	9.54	0.0020
Log10(Dose)	1	6.2308	1.8996	2.5076	9.9539	10.76	0.0010
_C_	1	0.2409	0.0523	0.1385	0.3433		

**Figure 94.3** Tolerance Distribution Estimates for the PROBIT Procedure

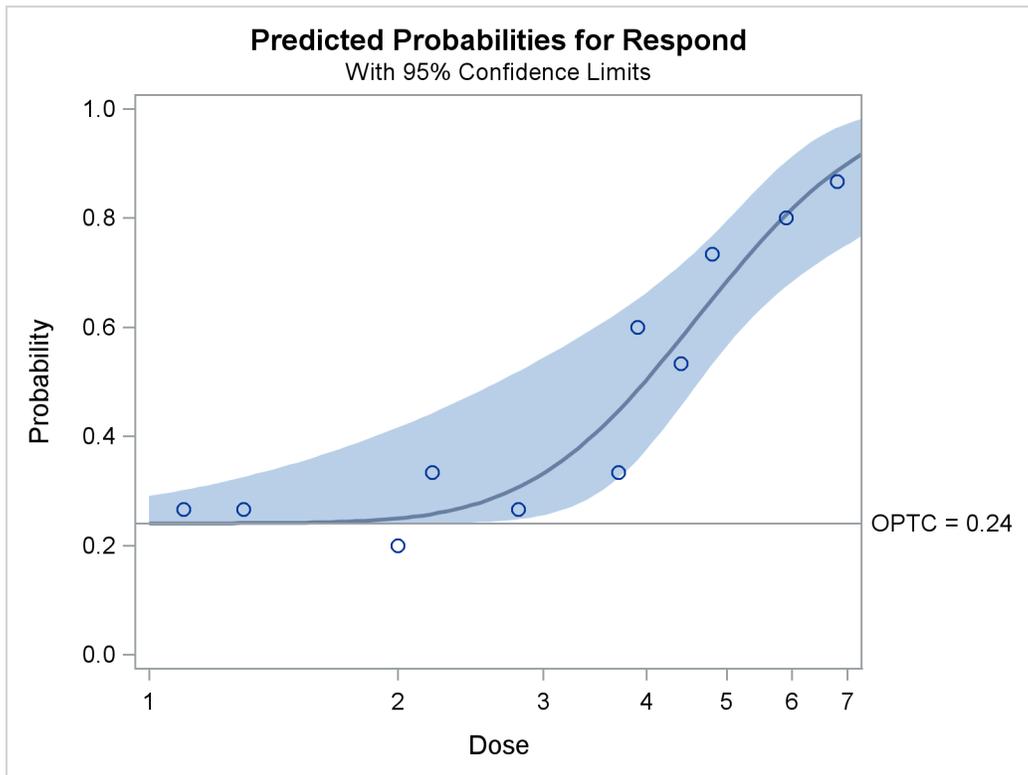
Estimated Covariance Matrix for Tolerance Parameters			
	MU	SIGMA	_C_
MU	0.001158	-0.000493	0.000954
SIGMA	-0.000493	0.002394	-0.000999
_C_	0.000954	-0.000999	0.002731

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.

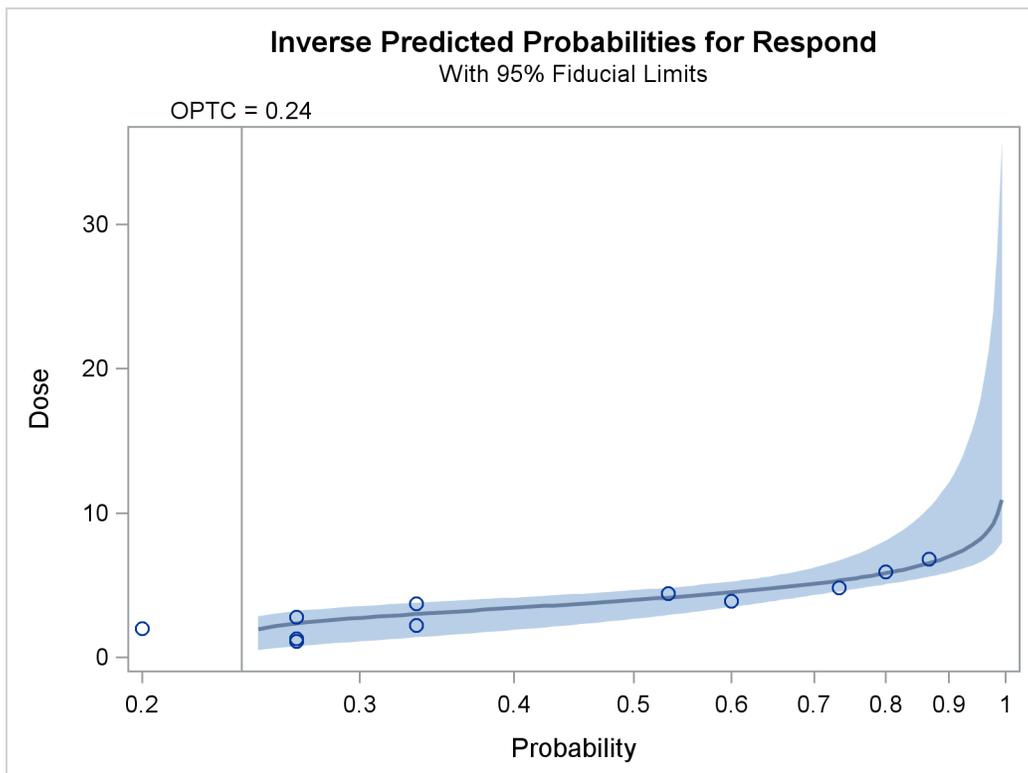
The PLOT=IPPPLOT option creates the plot in Figure 94.5, showing the inverse relationship between dosage level and observed response proportions/estimated probability values. The dashed lines represent pointwise fiducial limits for the predicted values of the dose variable, and a reference line is also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

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

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.

**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:

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

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 >**

#### **PLOT | PLOTS <=(plot-request <...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:

```
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

$$\hat{F}_j(\mathbf{x}) = C + (1 - C)F(\hat{a}_j + \mathbf{x}'\hat{\mathbf{b}})$$

where  $j = 1, \dots, k$  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,  $\hat{\mathbf{b}}$  is the vector of estimated parameters,  $\mathbf{x}$  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  $\mathbf{x}_{-1}$  be the vector of the rest of the covariates. Let the corresponding estimated parameters be  $\hat{b}_1$  and  $\hat{\mathbf{b}}_{-1}$ . Then

$$\hat{F}_j(\mathbf{x}) = C + (1 - C)F(\hat{a}_j + x_1\hat{b}_1 + \mathbf{x}'_{-1}\hat{\mathbf{b}}_{-1})$$

To plot  $\hat{F}_j$  as a function of  $x_1$ ,  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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

LEVEL= <i>character-list</i>	Specifies the names of the levels for which the CDF curves are requested
NOTHRESH	Suppresses the threshold line
THRESHLABPOS= <i>value</i>	Specifies the position for the label of the threshold line

### General Options

**Table 94.3** Color Options

CAXIS= <i>color</i>	Specifies color for axis
CFIT= <i>color</i>	Specifies color for fitted curves
CFRAME= <i>color</i>	Specifies color for frame
CGRID= <i>color</i>	Specifies color for grid lines
CHREF= <i>color</i>	Specifies color for HREF= lines
CLABBOX= <i>color</i>	Specifies color for label box
CTEXT= <i>color</i>	Specifies color for text
CVREF= <i>color</i>	Specifies color for VREF= lines

**Table 94.4** Options to Enhance Plots Produced on Graphics Devices

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

**Table 94.5** Axis Options

HAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for horizontal axis
HOFFSET= <i>value</i>	Specifies offset for horizontal axis
HLOWER= <i>value</i>	Specifies lower limit on horizontal axis scale
HUPPER= <i>value</i>	Specifies upper limit on horizontal axis scale
NHTICK= <i>n</i>	Specifies number of ticks for horizontal axis
NVTICK= <i>n</i>	Specifies number of ticks for vertical axis
VAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for vertical axis
VAXISLABEL= <i>'label'</i>	Specifies label for vertical axis
VOFFSET= <i>value</i>	Specifies offset for vertical axis
VLOWER= <i>value</i>	Specifies lower limit on vertical axis scale
VUPPER= <i>value</i>	Specifies upper limit on vertical axis scale
WAXIS= <i>n</i>	Specifies thickness for axis

**Table 94.6** Graphics Catalog Options

DESCRIPTION= <i>'string'</i>	Specifies description for graphics catalog member
NAME= <i>'string'</i>	Specifies name for plot in graphics catalog

**Table 94.7** Options for Text Enhancement

<b>FONT=font</b>	Specifies software font for text
<b>HEIGHT=value</b>	Specifies height of text outside framed areas
<b>INHEIGHT=value</b>	Specifies height of text inside framed areas

**Table 94.8** Options for Reference Lines

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

## Dictionary of Options

The following entries provide detailed descriptions of the *options* in the CDFPLOT 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 CDF plot. The ANNOTATE= data set you specify in the CDFPLOT 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 CDF 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.

**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.

**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.

<i>n</i>	Label Placement
1	Top
2	Staggered from top
3	Bottom
4	Staggered from bottom
5	Alternating from top
6	Alternating from bottom

**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.

**TURNVLABELS**

vertically strings out characters in vertical labels.

**NOVTICK**

suppresses vertical tick marks.

**THREHLABPOS=*n***

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

<i>n</i>	Label Placement
1	Left
2	Right

**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.

<i>n</i>	Label Placement
1	Left
2	Right

**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.

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

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*

Plot-Type and Description	Plot-Definition-Options
<p><b>BOX</b> Displays a box plot of continuous response data at each level of a CLASS effect, with predicted values superimposed and connected by a line. This is an alternative to the <b>INTERACTION</b> <i>plot-type</i>.</p>	<p><b>PLOTBY=</b> variable or CLASS effect <b>X=</b> CLASS variable or effect</p>
<p><b>CONTOUR</b> Displays a contour plot of predicted values against two continuous covariates.</p>	<p><b>PLOTBY=</b> variable or CLASS effect <b>X=</b> continuous variable <b>Y=</b> continuous variable</p>
<p><b>FIT</b> Displays a curve of predicted values versus a continuous variable.</p>	<p><b>PLOTBY=</b> variable or CLASS effect <b>X=</b> continuous variable</p>
<p><b>INTERACTION</b> Displays a plot of predicted values (possibly with error bars) versus the levels of a CLASS effect. The predicted values are connected with lines and can be grouped by the levels of another CLASS effect.</p>	<p><b>PLOTBY=</b> variable or CLASS effect <b>SLICEBY=</b> variable or CLASS effect <b>X=</b> CLASS variable or effect</p>
<p><b>MOSAIC</b> Displays a mosaic plot of predicted values using up to three CLASS effects.</p>	<p><b>PLOTBY=</b> variable or CLASS effect <b>X=</b> CLASS effects</p>
<p><b>SLICEFIT</b> Displays a curve of predicted values versus a continuous variable grouped by the levels of a CLASS effect.</p>	<p><b>PLOTBY=</b> variable or CLASS effect <b>SLICEBY=</b> variable or CLASS effect <b>X=</b> continuous variable</p>

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  $\mathbf{L}\boldsymbol{\beta}$ . 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 94.10** ESTIMATE Statement Options

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

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

```
INSET < keyword-list > < / options > ;
```

The box or table of summary information produced on plots made with the CDFPLOT, IPPLOT, LPREDPLOT, or PREDPLOT 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.

```
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.

```
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

NOBS	Number of observations
NTRIALS	Number of trials
NEVENTS	Number of events
C	User-input threshold
OPTC	Estimated natural threshold
NRESPLEV	Number of levels of the response variable

**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

CFILL= <i>color</i>	Specifies color for filling box
CFILLH= <i>color</i>	Specifies color for filling box header
CFRAME= <i>color</i>	Specifies color for frame
CHEADER= <i>color</i>	Specifies color for text in header
CTEXT= <i>color</i>	Specifies color for text

**Table 94.14** General Appearance Options

FONT= <i>font</i>	Specifies software font for text
HEIGHT= <i>value</i>	Specifies height of text
HEADER="quoted-string"	Specifies text for header or box title
NOFRAME	Omits frame around box
POS= <i>value</i> <DATA   PERCENT >	Determines the position of the inset. The <i>value</i> 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.
REFPOINT= <i>name</i>	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.

## IPPLOT Statement

**IPPLOT** < variable > < options > ;

The IPPLOT 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

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

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

To plot  $\hat{x}_1$  as a function of  $p$ ,  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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 IPPLOT 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 IPPLOT statement. The “Dictionary of Options” on page 7600 describes each option in detail.

**IPP Options****Table 94.15** Plot Layout Options for IPPLOT

NOCONF	Suppresses fiducial limits
NODATA	Suppresses observed data points on the plot
NOTHRESH	Suppresses the threshold line
THRESHLABPOS= <i>value</i>	Specifies the position for the label of the threshold line

**General Options****Table 94.16** Color Options

CAXIS= <i>color</i>	Specifies color for axis
CFIT= <i>color</i>	Specifies color for fitted curves
CFRAME= <i>color</i>	Specifies color for frame
CGRID= <i>color</i>	Specifies color for grid lines
CHREF= <i>color</i>	Specifies color for HREF= lines
CTEXT= <i>color</i>	Specifies color for text
CVREF= <i>color</i>	Specifies color for VREF= lines

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

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

**Table 94.18** Axis Options

HAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for horizontal axis
HOFFSET= <i>value</i>	Specifies offset for horizontal axis
HLOWER= <i>value</i>	Specifies lower limit on horizontal axis scale
HUPPER= <i>value</i>	Specifies upper limit on horizontal axis scale
NHTICK= <i>n</i>	Specifies number of ticks for horizontal axis
NVTICK= <i>n</i>	Specifies number of ticks for vertical axis
VAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for vertical axis
VAXISLABEL= <i>'label'</i>	Specifies label for vertical axis
VOFFSET= <i>value</i>	Specifies offset for vertical axis
VLOWER= <i>value</i>	Specifies lower limit on vertical axis scale
VUPPER= <i>value</i>	Specifies upper limit on vertical axis scale
WAXIS= <i>n</i>	Specifies thickness for axis

**Table 94.19** Options for Reference Lines

HREF<(INTERSECT)> = <i>value-list</i>	Requests horizontal reference line
HREFLABELS= ( <i>'label1',...,'labeln'</i> )	Specifies labels for HREF= lines
HREFLABPOS= <i>n</i>	Specifies vertical position of labels for HREF= lines
LHREF= <i>linetype</i>	Specifies line style for HREF= lines
LVREF= <i>linetype</i>	Specifies line style for VREF= lines
VREF<(INTERSECT)> = <i>value-list</i>	Requests vertical reference line
VREFLABELS= ( <i>'label1',...,'labeln'</i> )	Specifies labels for VREF= lines
VREFLABPOS= <i>n</i>	Specifies horizontal position of labels for VREF= lines

**Table 94.20** Graphics Catalog Options

DESCRIPTION= <i>'string'</i>	Specifies description for graphics catalog member
NAME= <i>'string'</i>	Specifies name for plot in graphics catalog

**Table 94.21** Options for Text Enhancement

<b>FONT=</b> <i>font</i>	Specifies software font for text
<b>HEIGHT=</b> <i>value</i>	Specifies height of text used outside framed areas
<b>INHEIGHT=</b> <i>value</i>	Specifies height of text inside framed areas

## Dictionary of Options

The following entries provide detailed descriptions of the *options* in the IPPLOT 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 IPP plot. The ANNOTATE= data set you specify in the IPPLOT 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 IPP 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. 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.

<i>n</i>	Label Placement
1	Top
2	Staggered from top
3	Bottom
4	Staggered from bottom
5	Alternating from top
6	Alternating from bottom

**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.

<i>n</i>	Label Placement
1	Top
2	Bottom

**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.

<i>n</i>	<b>Label Placement</b>
1	Left
2	Right

**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)  $\mathbf{x}'\mathbf{b}$  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,  $\mathbf{x}_{-1}$  be the vector of the rest of the covariates,  $\hat{\mathbf{b}}_{-1}$  be the vector of estimated parameters corresponding to  $\mathbf{x}_{-1}$ , and  $\hat{b}_1$  be the estimated parameter for the dose variable of interest.

To plot  $\hat{\mathbf{x}}'\mathbf{b}$  as a function of  $x_1$ ,  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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

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

**General Options****Table 94.23** Color Options

CAXIS= <i>color</i>	Specifies color for axis
CFIT= <i>color</i>	Specifies color for fitted curves
CFRAME= <i>color</i>	Specifies color for frame
CGRID= <i>color</i>	Specifies color for grid lines
CHREF= <i>color</i>	Specifies color for HREF= lines
CTEXT= <i>color</i>	Specifies color for text
CVREF= <i>color</i>	Specifies color for VREF= lines

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

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

**Table 94.25** Axis Options

HAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for horizontal axis
HOFFSET= <i>value</i>	Specifies offset for horizontal axis
HLOWER= <i>value</i>	Specifies lower limit on horizontal axis scale
HUPPER= <i>value</i>	Specifies upper limit on horizontal axis scale
NHTICK= <i>n</i>	Specifies number of ticks for horizontal axis
NVTICK= <i>n</i>	Specifies number of ticks for vertical axis
VAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for vertical axis
VAXISLABEL= <i>'label'</i>	Specifies label for vertical axis
VOFFSET= <i>value</i>	Specifies offset for vertical axis
VLOWER= <i>value</i>	Specifies lower limit on vertical axis scale
VUPPER= <i>value</i>	Specifies upper limit on vertical axis scale
WAXIS= <i>n</i>	Specifies thickness for axis

**Table 94.26** Graphics Catalog Options

DESCRIPTION= <i>'string'</i>	Specifies description for graphics catalog member
NAME= <i>'string'</i>	Specifies name for plot in graphics catalog

**Table 94.27** Options for Text Enhancement

FONT= <i>font</i>	Specifies software font for text
HEIGHT= <i>value</i>	Specifies height of text used outside framed areas
INHEIGHT= <i>value</i>	Specifies height of text inside framed areas

**Table 94.28** Options for Reference Lines

<b>HREF</b> <(INTERSECT)> = <i>value-list</i>	Requests horizontal reference line
<b>HREFLABELS</b> = ( <i>'label1'</i> , . . . , <i>'labeln'</i> )	Specifies labels for HREF= lines
<b>HREFLABPOS</b> = <i>n</i>	Specifies vertical position of labels for HREF= lines
<b>LHREF</b> = <i>linetype</i>	Specifies line style for HREF= lines
<b>LVREF</b> = <i>linetype</i>	Specifies line style for VREF= lines
<b>VREF</b> <(INTERSECT)> = <i>value-list</i>	Requests vertical reference line
<b>VREFLABELS</b> = ( <i>'label1'</i> , . . . , <i>'labeln'</i> )	Specifies labels for VREF= lines
<b>VREFLABPOS</b> = <i>n</i>	Specifies horizontal position of labels for VREF= lines

## 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'

**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.

<i>n</i>	Label Placement
1	Top
2	Staggered from top
3	Bottom
4	Staggered from bottom
5	Alternating from top
6	Alternating from bottom

**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*

**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 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.

<i>n</i>	<u>Label Placement</u>
1	Left
2	Right

**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.

<i>n</i>	Label Placement
1	Left
2	Right

**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.

**WAXIS=*n***

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

**WFIT=*n***

specifies line thickness for fitted lines. 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.

---

## LSMEANS Statement

**LSMEANS** < *model-effects* > < / *options* > ;

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 94.29** LSMEANS Statement Options

Option	Description
<b>Construction and Computation of LS-Means</b>	
AT	Modifies the covariate value in computing LS-means
BYLEVEL	Computes separate margins
DIFF	Requests differences of LS-means
OM=	Specifies the weighting scheme for LS-means computation as determined by the input data set
SINGULAR=	Tunes estimability checking
<b>Degrees of Freedom and <i>p</i>-values</b>	
ADJUST=	Determines the method for multiple-comparison adjustment of LS-means differences
ALPHA= $\alpha$	Determines the confidence level ( $1 - \alpha$ )
STEPDOWN	Adjusts multiple-comparison <i>p</i> -values further in a step-down fashion
<b>Statistical Output</b>	
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 <b>L</b> matrix
LINES	Produces a “Lines” display for pairwise LS-means differences
MEANS	Prints the LS-means
PLOTS=	Requests graphs of means and mean comparisons

**Table 94.29** *continued*

Option	Description
SEED=	Specifies the seed for computations that depend on random numbers
<b>Generalized Linear Modeling</b>	
EXP	Exponentiates and displays estimates of LS-means or LS-means differences
ILINK	Computes and displays estimates and standard errors of LS-means (but not differences) on the inverse linked scale
ODDSRATIO	Reports (simple) differences of least squares means in terms of odds ratios if permitted by the link function

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

## LSMESTIMATE Statement

```
LSMESTIMATE 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** LSMESTIMATE Statement Options

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

Table 94.30 *continued*

Option	Description
TESTVALUE= UPPER	Specifies values under the null hypothesis for tests Performs one-sided, upper-tailed inference
<b>Statistical Output</b>	
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 <b>L</b> matrix
ELSM	Prints the <b>K</b> matrix
JOINT	Produces a joint <i>F</i> 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
<b>Generalized Linear Modeling</b>	
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

```
<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

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

## 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:

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

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

**REF=**'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.

Type of List	Specification
List separated by blanks	<b>initial=3 4 5</b>
List separated by commas	<b>initial=3,4,5</b>

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:

```
PROB = .1 TO .9 by .1
PROB = .1 .2 .3 .4
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:

- |                     |  |
|---------------------|--|
| <b>D   DEVIANCE</b> | specifies that the dispersion parameter be estimated by the deviance divided by its degrees of freedom.  |
| <b>P   PEARSON</b>  | 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  $x'\beta$ , and the estimate of its standard error. Estimates of the probabilities,  $x'\beta$ , 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:

- |                     |  |
|---------------------|--|
| <i>keyword=name</i> | specifies the statistics to include in the output data set and assigns names to the new variables that contain the statistics. Specify a <i>keyword</i> for each desired statistic (see the following list of <i>keywords</i> ), an equal sign, and the variable to contain the statistic. The <i>keywords</i> allowed and the statistics they represent are as follows: |
|---------------------|--|

**PROB | P** cumulative probability estimates

$$p = C + (1 - C)F(a_j + \mathbf{x}'\boldsymbol{\beta})$$

**STD** standard error estimates of  $a_j + \mathbf{x}'\mathbf{b}$

**XBETA** estimates of  $a_j + \mathbf{x}'\boldsymbol{\beta}$

**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  $a_j + \mathbf{x}'\mathbf{b}$ , 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=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 probability is

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

for the binomial model and

$$\begin{aligned} \hat{p}_1 &= C + (1 - C)F(\mathbf{x}'\hat{\mathbf{b}}) \\ \hat{p}_j &= (1 - C)(F(\hat{a}_j + \mathbf{x}'\hat{\mathbf{b}}) - F(\hat{a}_{j-1} + \mathbf{x}'\hat{\mathbf{b}})) \quad j = 2, \dots, k - 1 \\ \hat{p}_k &= (1 - C)(1 - F(\hat{a}_{k-1} + \mathbf{x}'\hat{\mathbf{b}})) \end{aligned}$$

for the multinomial model with  $k$  response levels, where  $F$  is the cumulative distribution function used to model the probability,  $\mathbf{x}'$  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  $\hat{\mathbf{b}}'$  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  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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 PREDPLOT 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 all *options* by function. The “Dictionary of Options” on page 7626 describes each option in detail.

**PREDPLOT Options**

**Table 94.32** Plot Layout Options for PREDPLOT

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

**General Options****Table 94.33** Color Options

<i>CAXIS=</i> color	Specifies color for the axes
<i>CFIT=</i> color	Specifies color for fitted curves
<i>CFRAME=</i> color	Specifies color for frame
<i>CGRID=</i> color	Specifies color for grid lines
<i>CHREF=</i> color	Specifies color for HREF= lines
<i>CLABBOX=</i> color	Specifies color for label box
<i>CTEXT=</i> color	Specifies color for text
<i>CVREF=</i> color	Specifies color for VREF= lines

**Table 94.34** Options to Enhance Plots Produced on Graphics Devices

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

**Table 94.35** Axis Options

HAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for horizontal axis
HOFFSET= <i>value</i>	Specifies offset for horizontal axis
HLOWER= <i>value</i>	Specifies lower limit on horizontal axis scale
HUPPER= <i>value</i>	Specifies upper limit on horizontal axis scale
NHTICK= <i>n</i>	Specifies number of ticks for horizontal axis
NVTICK= <i>n</i>	Specifies number of ticks for vertical axis
VAXIS= <i>value1 to value2</i> < by <i>value3</i> >	Specifies tick mark values for vertical axis
VAXISLABEL= <i>'label'</i>	Specifies label for vertical axis
VOFFSET= <i>value</i>	Specifies offset for vertical axis
VLOWER= <i>value</i>	Specifies lower limit on vertical axis scale
VUPPER= <i>value</i>	Specifies upper limit on vertical axis scale
WAXIS= <i>n</i>	Specifies thickness for axis

**Table 94.36** Graphics Catalog Options

DESCRIPTION= <i>'string'</i>	Specifies description for graphics catalog member
NAME= <i>'string'</i>	Specifies name for plot in graphics catalog

**Table 94.37** Options for Text Enhancement

FONT= <i>font</i>	Specifies software font for text
HEIGHT= <i>value</i>	Specifies height of text used outside framed areas
INHEIGHT= <i>value</i>	Specifies height of text inside framed areas

**Table 94.38** Options for Reference Lines

<b>HREF</b> <(INTERSECT)> = <i>value-list</i>	Requests horizontal reference line
<b>HREFLABELS</b> = ( <i>'label1'</i> , . . . , <i>'labeln'</i> )	Specifies labels for HREF= lines
<b>HREFLABPOS</b> = <i>n</i>	Specifies vertical position of labels for HREF= lines
<b>LHREF</b> = <i>linetype</i>	Specifies line style for HREF= lines
<b>LVREF</b> = <i>linetype</i>	Specifies line style for VREF= lines
<b>VREF</b> <(INTERSECT)> = <i>value-list</i>	Requests vertical reference line
<b>VREFLABELS</b> = ( <i>'label1'</i> , . . . , <i>'labeln'</i> )	Specifies labels for VREF= lines
<b>VREFLABPOS</b> = <i>n</i>	Specifies horizontal position of labels for VREF= lines

## 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.

<i>n</i>	Label Placement
1	Top
2	Staggered from top
3	Bottom
4	Staggered from bottom
5	Alternating from top
6	Alternating from bottom

**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*

**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 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.

**THRESLABPOS=*n***

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

<i>n</i>	<u>Label Placement</u>
1	Left
2	Right

**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.

<i>n</i>	Label Placement
1	Left
2	Right

**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:

$$\Phi^{-1} \left( \frac{p - C}{1 - C} \right) = \mathbf{x}'\boldsymbol{\beta}$$

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  $\mathbf{x}$  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  $\boldsymbol{\beta}$  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:

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

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

```
model Y(ref='0') = 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,  $\mathbf{V}$ , 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

Cumulative Distribution Function	Distribution
$\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$	Normal
$\frac{1}{1+e^{-x}}$	Logistic
$1 - e^{-e^x}$	Extreme value or Gompertz

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.

Distribution	Mean	Variance
Normal	0	1
Logistic	0	$\pi^2/3$
Extreme value or Gompertz	$-\gamma$	$\pi^2/6$

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

$$p = C + (1 - C)F(\mathbf{x}'\mathbf{b})$$

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

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

$$\begin{aligned}
 p_1 &= C + (1 - C)F(\mathbf{x}'\mathbf{b}) \\
 p_2 &= (1 - C)(F(a_2 + \mathbf{x}'\mathbf{b}) - F(\mathbf{x}'\mathbf{b})) \\
 &\vdots \\
 p_j &= (1 - C)(F(a_j + \mathbf{x}'\mathbf{b}) - F(a_{j-1} + \mathbf{x}'\mathbf{b})) \\
 &\vdots \\
 p_k &= (1 - C)(1 - F(a_{k-1} + \mathbf{x}'\mathbf{b}))
 \end{aligned}$$

Thus, for a  $k$ -level response, there are  $k - 2$  additional parameters,  $a_2, a_3, \dots, a_{k-1}$ , estimated. These parameters are denoted by Intercept $j$ ,  $j = 2, 3, \dots, k - 1$ , 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 *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

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

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

$$\chi_D^2 = 2 \sum_{i=1}^m \sum_{j=1}^k r_{ij} \ln \left( \frac{r_{ij}}{n_i \hat{p}_{ij}} \right)$$

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  $\chi_P^2$  and the deviance  $\chi_D^2$  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

$$\hat{\sigma}^2 = \begin{cases} \chi_P^2 / (m(k - 1) - q) & \text{SCALE=PEARSON} \\ \chi_D^2 / (m(k - 1) - q) & \text{SCALE=DEVIANC} \\ (\text{constant})^2 & \text{SCALE=constant} \end{cases}$$

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

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

where  $F$  is the cumulative distribution function used to model the probability,  $\mathbf{x}^*$  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,  $\mathbf{b}^*$  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  $\mathbf{V}$  is the variance-covariance matrix of  $a$  and  $b$ :

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

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:

<code>_MODEL_</code>	a character variable containing the label of the MODEL statement, if present, or blank otherwise
<code>_NAME_</code>	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
<code>_TYPE_</code>	a character variable containing the type of the observation, either PARMS for parameter estimates or COV for covariance estimates
<code>_DIST_</code>	a character variable containing the name of the distribution modeled
<code>_LNLIKE_</code>	a numeric variable containing the last computed value of the log likelihood
<code>_C_</code>	a numeric variable containing the estimated threshold parameter
<code>INTERCEPT</code>	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, IPPLOT, 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 PREDPLOT. 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 94.39** Plot Statement Availability

Statement	Binomial	Multinomial
CDFPLOT	No	Yes
IPPPLOT	Yes	No
LPREDPLOT	Yes	Yes
PREDPLOT	Yes	Yes

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:

```
proc probit data=study log10 optc;
  model respond/number=dose;
  predpplot var=dose cfit=blue; inset;
  lpredplot var=dose cfit=blue; inset;
  ippplot   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 IPPLOT statement creates a similar plot. See the section “[IPPLOT 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 94.40** ODS Tables Produced by PROC PROBIT

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

## 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

Option	Plot
ALL	All appropriate plots
CDFPLOT	Estimated cumulative probability
IPPPLOT	Inverse predicted probability
LPREDPLOT	Linear predictor
NONE	No plot
PREDPPLOT	Predicted probability

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

ODS Graph Name	Plot Description	Statement	PLOTS= Option
CDFPlot	Estimated cumulative probability	PROC	CDFPLOT
IPPPlot	Inverse predicted probability	PROC	IPPPLOT
LPredPlot	Linear predictor	PROC	LPREDPLOT
PredPPlot	Predicted probability	PROC	PREDPPLOT

### CDF Plot

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

$$\hat{F}_j(\mathbf{x}) = C + (1 - C)F(\hat{a}_j + \mathbf{x}'\hat{\mathbf{b}})$$

where  $j = 1, \dots, k$  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,  $\hat{\mathbf{b}}$  is the vector of estimated parameters,  $\mathbf{x}$  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  $\mathbf{x}_{-1}$  be the vector of the rest of the covariates. Let the corresponding estimated parameters be  $\hat{b}_1$  and  $\hat{\mathbf{b}}_{-1}$ . Then

$$\hat{F}_j(\mathbf{x}) = C + (1 - C)F(\hat{a}_j + x_1\hat{b}_1 + \mathbf{x}'_{-1}\hat{\mathbf{b}}_{-1})$$

To plot  $\hat{F}_j$  as a function of  $x_1$ ,  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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

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

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

To plot  $\hat{x}_1$  as a function of  $p$ ,  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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  $\mathbf{x}'\mathbf{b}$  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,  $\mathbf{x}_{-1}$  be the vector of the rest of the covariates,  $\hat{\mathbf{b}}_{-1}$  be the vector of estimated parameters corresponding to  $\mathbf{x}_{-1}$ , and  $\hat{b}_1$  be the estimated parameter for the dose variable of interest.

To plot  $\hat{\mathbf{x}}'\mathbf{b}$  as a function of  $x_1$ ,  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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

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

for the binomial model and

$$\begin{aligned}\hat{p}_1 &= C + (1 - C)F(\mathbf{x}'\hat{\mathbf{b}}) \\ \hat{p}_j &= (1 - C)(F(\hat{a}_j + \mathbf{x}'\hat{\mathbf{b}}) - F(\hat{a}_{j-1} + \mathbf{x}'\hat{\mathbf{b}})), \quad j = 2, \dots, k - 1 \\ \hat{p}_k &= (1 - C)(1 - F(\hat{a}_{k-1} + \mathbf{x}'\hat{\mathbf{b}}))\end{aligned}$$

for the multinomial model with  $k$  response levels, where  $F$  is the cumulative distribution function used to model the probability,  $\mathbf{x}'$  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  $\hat{\mathbf{b}}'$  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  $\mathbf{x}_{-1}$  must be specified. You can use the XDATA= option to provide the values of  $\mathbf{x}_{-1}$  (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. Specifically, for a given level of Dose, 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](#):

```
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

Iteration History for Parameter Estimates				
Iter	Ridge	Loglikelihood	Intercept	Log10(Dose)
0	0	-51.292891	0	0
1	0	-37.881166	-1.355817008	2.635206083
2	0	-37.286169	-1.764939171	3.3408954936
3	0	-37.280389	-1.812147863	3.4172391614
4	0	-37.280388	-1.812704962	3.418117919
5	0	-37.280388	-1.812704962	3.418117919

Output 94.1.1 *continued*

Model Information	
Data Set	WORK.A
Events Variable	Response
Trials Variable	N
Number of Observations	7
Number of Events	38
Number of Trials	74
Name of Distribution	Normal
Log Likelihood	-37.28038802

Last Evaluation of the  
Negative of the Gradient

Intercept	Log10(Dose)
3.4349069E-7	-2.09809E-8

Last Evaluation of the Negative of the  
Hessian

	Intercept	Log10(Dose)
Intercept	36.005280383	20.152675982
Log10(Dose)	20.152675982	13.078826305

Goodness-of-Fit Tests

Statistic	Value	DF	Value/DF	Pr > ChiSq
Pearson Chi-Square	3.6497	5	0.7299	0.6009
L.R. Chi-Square	4.6381	5	0.9276	0.4616

Response-Covariate Profile

Response Levels	2
Number of Covariate Values	7

Analysis of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-1.8127	0.4493	-2.6934	-0.9320	16.27	<.0001
Log10(Dose)	1	3.4181	0.7455	1.9569	4.8794	21.02	<.0001

Probit Model in Terms  
of Tolerance  
Distribution

MU	SIGMA
0.53032254	0.29255866

Estimated Covariance  
Matrix for Tolerance  
Parameters

	MU	SIGMA
MU	0.002418	-0.000409
SIGMA	-0.000409	0.004072

The  $p$ -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.

### Output 94.1.2 Probit Analysis with Normal Distribution

#### The Probit Procedure

Probit Analysis on Log10(Dose)			
Probability	Log10(Dose)	95% Fiducial Limits	
0.01	-0.15027	-0.69518	0.07710
0.02	-0.07052	-0.55766	0.13475
0.03	-0.01992	-0.47064	0.17156
0.04	0.01814	-0.40534	0.19941
0.05	0.04911	-0.35233	0.22218
0.06	0.07546	-0.30731	0.24165
0.07	0.09857	-0.26793	0.25881
0.08	0.11926	-0.23273	0.27425
0.09	0.13807	-0.20080	0.28837
0.10	0.15539	-0.17147	0.30142
0.15	0.22710	-0.05086	0.35631
0.20	0.28410	0.04369	0.40124
0.25	0.33299	0.12343	0.44116
0.30	0.37690	0.19348	0.47857
0.35	0.41759	0.25658	0.51504
0.40	0.45620	0.31429	0.55182
0.45	0.49356	0.36754	0.58999
0.50	0.53032	0.41693	0.63057
0.55	0.56709	0.46296	0.67451
0.60	0.60444	0.50618	0.72271
0.65	0.64305	0.54734	0.77603
0.70	0.68374	0.58745	0.83550
0.75	0.72765	0.62776	0.90265
0.80	0.77655	0.66999	0.98008
0.85	0.83354	0.71675	1.07279
0.90	0.90525	0.77313	1.19191
0.91	0.92257	0.78646	1.22098
0.92	0.94139	0.80083	1.25265
0.93	0.96208	0.81653	1.28759
0.94	0.98519	0.83394	1.32672
0.95	1.01154	0.85367	1.37149
0.96	1.04250	0.87669	1.42424
0.97	1.08056	0.90480	1.48928
0.98	1.13116	0.94189	1.57602
0.99	1.21092	0.99987	1.71321

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

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

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	
Data Set	WORK.B
Events Variable	Response
Trials Variable	N
Number of Observations	7
Number of Events	38
Number of Trials	74
Name of Distribution	Logistic
Log Likelihood	-37.11065336

##### Analysis of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-3.2246	0.8861	-4.9613	-1.4880	13.24	0.0003
Log10(Dose)	1	5.9702	1.4492	3.1299	8.8105	16.97	<.0001

[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**

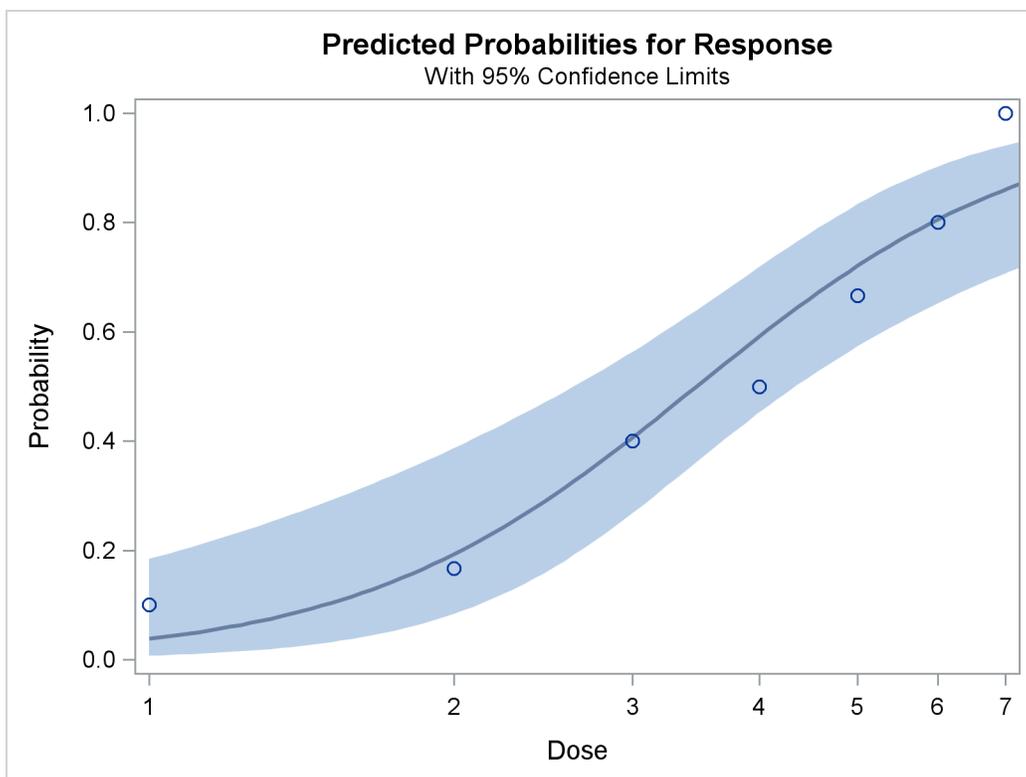
Probit Analysis on Log10(Dose)			
Probability	Log10(Dose)	95%	
		Fiducial Limits	
0.01	-0.22955	-0.97441	0.04234
0.02	-0.11175	-0.75158	0.12404
0.03	-0.04212	-0.62018	0.17265
0.04	0.00780	-0.52618	0.20771
0.05	0.04693	-0.45265	0.23533
0.06	0.07925	-0.39205	0.25826
0.07	0.10686	-0.34037	0.27796
0.08	0.13103	-0.29521	0.29530
0.09	0.15259	-0.25502	0.31085
0.10	0.17209	-0.21875	0.32498
0.15	0.24958	-0.07552	0.38207
0.20	0.30792	0.03092	0.42645
0.25	0.35611	0.11742	0.46451
0.30	0.39820	0.19143	0.49932
0.35	0.43644	0.25684	0.53275
0.40	0.47221	0.31588	0.56619
0.45	0.50651	0.36986	0.60089
0.50	0.54013	0.41957	0.63807
0.55	0.57374	0.46559	0.67894
0.60	0.60804	0.50846	0.72474
0.65	0.64381	0.54896	0.77673
0.70	0.68205	0.58815	0.83637
0.75	0.72414	0.62752	0.90582
0.80	0.77233	0.66915	0.98876
0.85	0.83067	0.71631	1.09242
0.90	0.90816	0.77562	1.23343
0.91	0.92766	0.79014	1.26931
0.92	0.94922	0.80607	1.30912
0.93	0.97339	0.82378	1.35391
0.94	1.00100	0.84384	1.40523
0.95	1.03332	0.86713	1.46546
0.96	1.07245	0.89511	1.53864
0.97	1.12237	0.93053	1.63228
0.98	1.19200	0.97952	1.76329
0.99	1.30980	1.06166	1.98569

**Output 94.1.6** Probit Analysis with Logistic Distribution**The Probit Procedure**

Probit Analysis on Dose			
Probability	Dose	95%	
		Fiducial Limits	
0.01	0.58945	0.10607	1.10241
0.02	0.77312	0.17718	1.33058
0.03	0.90757	0.23978	1.48817
0.04	1.01813	0.29773	1.61327
0.05	1.11413	0.35266	1.71922
0.06	1.20018	0.40546	1.81244
0.07	1.27896	0.45670	1.89654
0.08	1.35218	0.50675	1.97379
0.09	1.42100	0.55588	2.04572
0.10	1.48625	0.60430	2.11339
0.15	1.77656	0.84038	2.41030
0.20	2.03199	1.07379	2.66961
0.25	2.27043	1.31046	2.91416
0.30	2.50152	1.55393	3.15736
0.35	2.73172	1.80652	3.40996
0.40	2.96627	2.06957	3.68292
0.45	3.21006	2.34345	3.98927
0.50	3.46837	2.62768	4.34578
0.55	3.74746	2.92138	4.77466
0.60	4.05546	3.22451	5.30573
0.65	4.40366	3.53961	5.98041
0.70	4.80891	3.87391	6.86079
0.75	5.29836	4.24155	8.05044
0.80	5.92009	4.66820	9.74455
0.85	6.77126	5.20365	12.37149
0.90	8.09391	5.96508	17.11715
0.91	8.46559	6.16800	18.59129
0.92	8.89644	6.39837	20.37592
0.93	9.40575	6.66469	22.58957
0.94	10.02317	6.97977	25.42292
0.95	10.79732	7.36428	29.20549
0.96	11.81534	7.85438	34.56521
0.97	13.25466	8.52173	42.88232
0.98	15.55972	9.53941	57.98207
0.99	20.40815	11.52549	96.75820

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:

```

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

Model Information	
Data Set	WORK.MULTI
Dependent Variable	Symptoms
Weight Variable	N
Number of Observations	23
Name of Distribution	Normal
Log Likelihood	-345.9401767

Class Level Information		
Name	Levels	Values
Prep	2	stand test
Symptoms	3	None Mild Severe

#### Analysis of Maximum Likelihood Parameter Estimates

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

### Output 94.2.2 Multilevel Response: Parallel Analysis

#### The Probit Procedure

Model Information	
Data Set	WORK.MULTI
Dependent Variable	Symptoms
Weight Variable	N
Number of Observations	23
Name of Distribution	Normal
Log Likelihood	-346.306141

**Output 94.2.2** *continued*

Class Level Information		
Name	Levels	Values
Prep	2	stand test
Symptoms	3	None Mild Severe

Analysis of Maximum Likelihood Parameter Estimates								
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq	
				Intercept	1			3.4148
Intercept2	1	0.4678	0.0558	0.3584	0.5772	70.19	<.0001	
Prep	stand	1	-0.5675	0.1259	-0.8142	-0.3208	20.33	<.0001
Prep	test	0	0.0000	.	.	.	.	.
LDose	1	-2.3721	0.2949	-2.9502	-1.7940	64.68	<.0001	

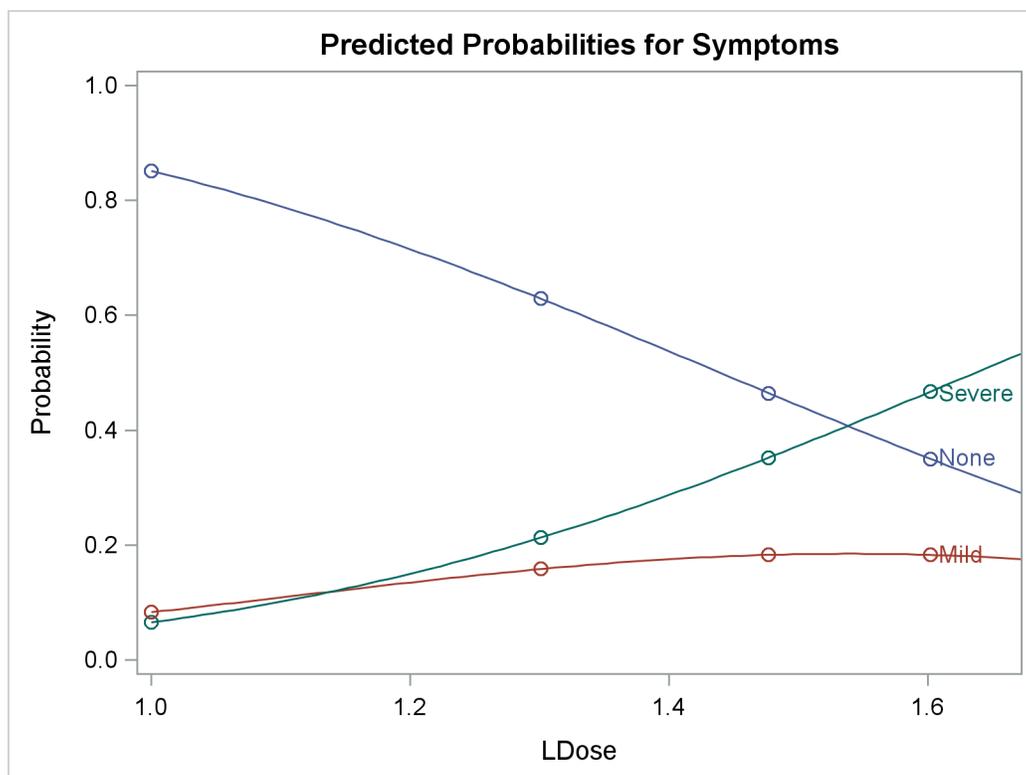
The negative coefficient associated with LDose indicates that the probability of having no symptoms (Symptoms='None') or no or mild symptoms (Symptoms='None' or Symptoms='Mild') decreases as LDose increases; that is, the probability of a severe symptom increases with LDose. This association is apparent for both treatment groups.

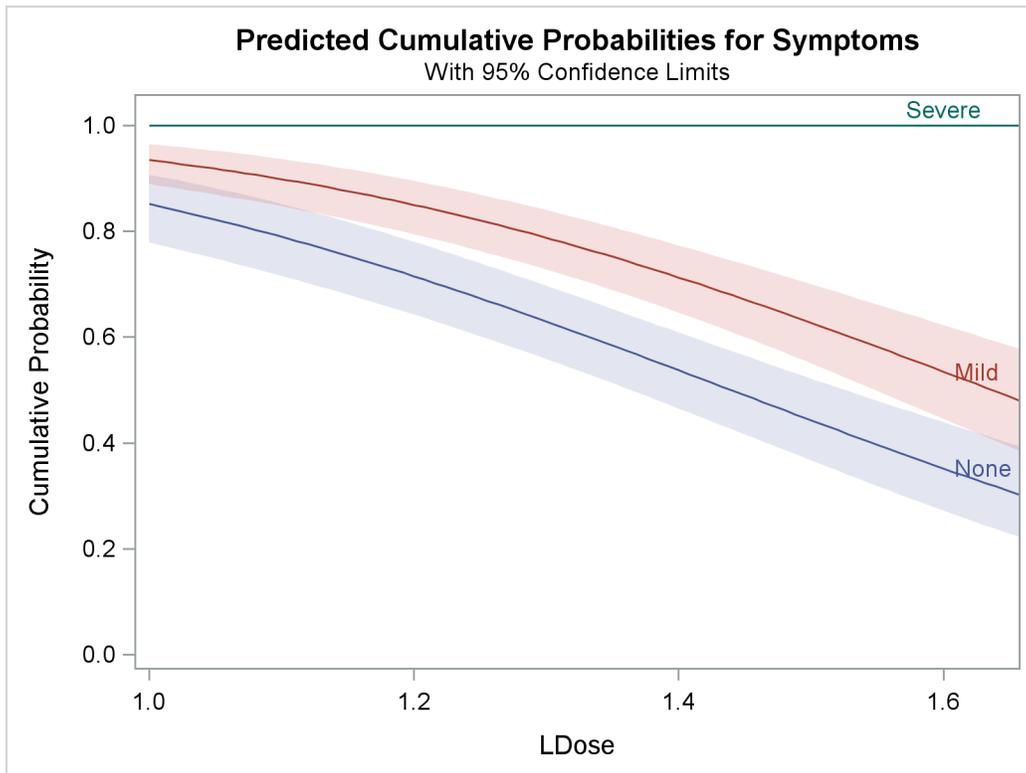
The negative coefficient associated with the standard treatment group (Prep = stand) indicates that the standard treatment is associated with more severe symptoms across all LDose values.

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](#).

```
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



**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](#).

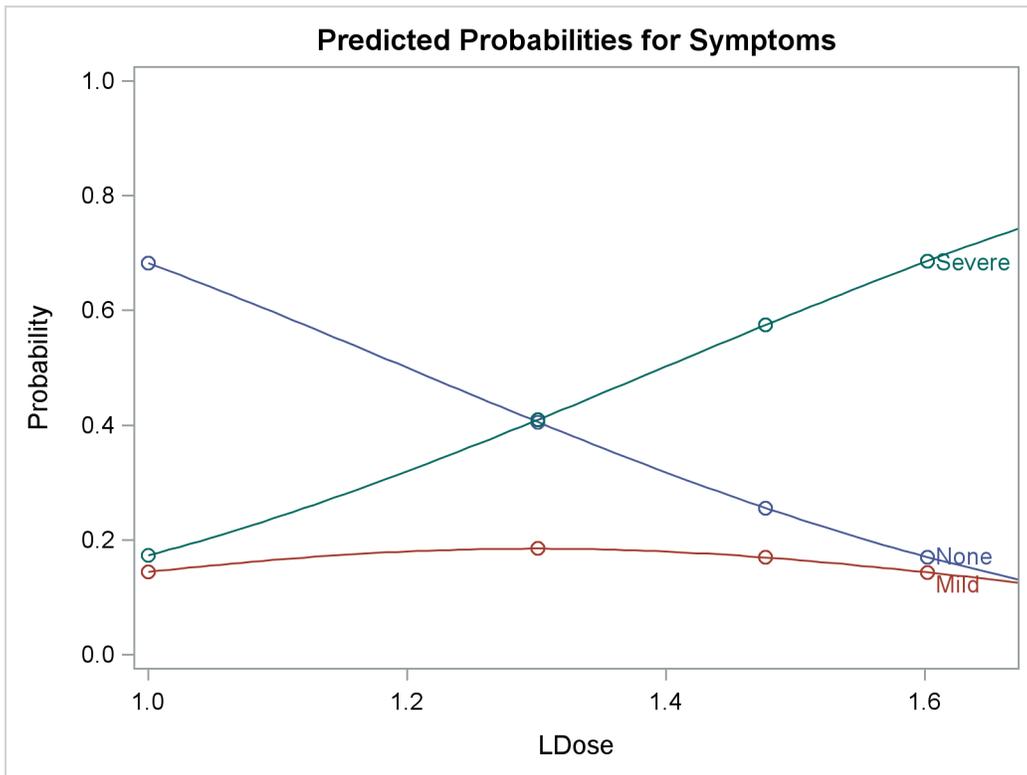
```

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

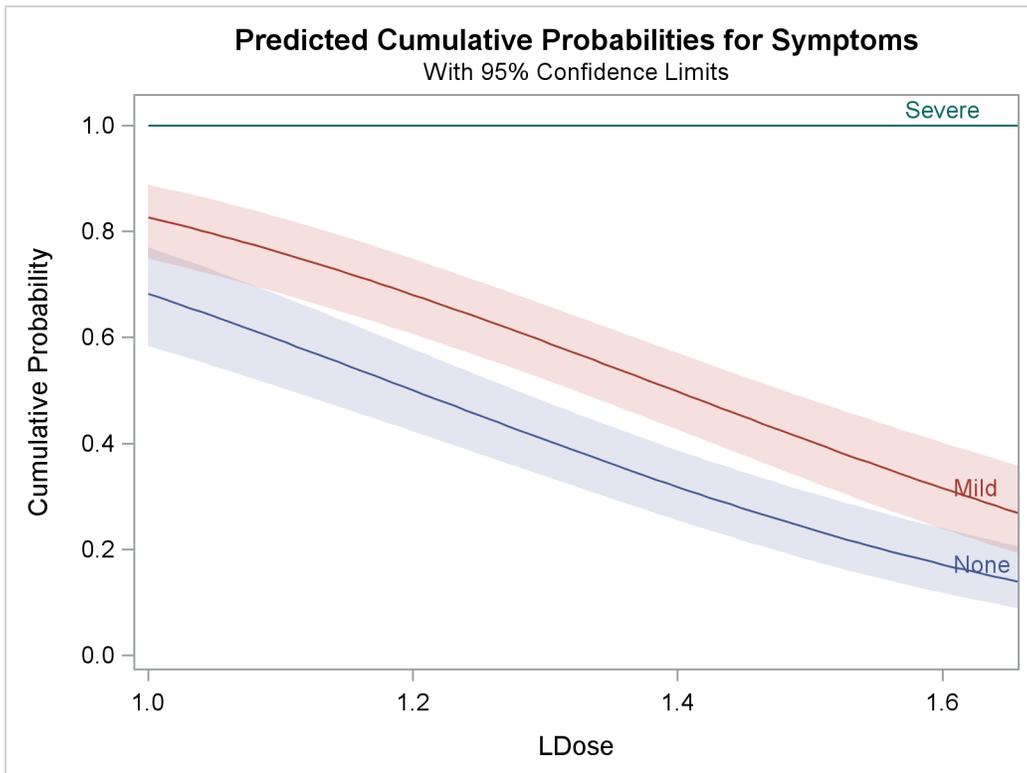
proc probit data=multi xdata=xrow
  plots=(predpplot(level=("None" "Mild" "Severe"))
         cdfplot(level=("None" "Mild" "Severe")));
  class Prep;
  parallel: model Symptoms(order=data)=Prep LDose / lackfit;
  weight N;
run;

```

**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** Logistic Regression of Subscription Status  
**The Probit Procedure**

Iteration History for Parameter Estimates					
Iter	Ridge	Loglikelihood	Intercept	sexFemale	age
0	0	-27.725887	0	0	0
1	0	-20.142659	-3.634567629	-1.648455751	0.1051634384
2	0	-19.52245	-5.254865196	-2.234724956	0.1506493473
3	0	-19.490439	-5.728485385	-2.409827238	0.1639621828
4	0	-19.490303	-5.76187293	-2.422349862	0.1649007124
5	0	-19.490303	-5.7620267	-2.422407743	0.1649050312
6	0	-19.490303	-5.7620267	-2.422407743	0.1649050312

Model Information	
Data Set	WORK.NEWS
Dependent Variable	subs
Number of Observations	40
Name of Distribution	Logistic
Log Likelihood	-19.49030281

Class Level Information		
Name	Levels	Values
sex	2	Female Male
subs	2	accept reject

**Last Evaluation of the Negative of the Gradient**

Intercept	sexFemale	age
-5.95557E-12	8.768324E-10	-1.6367E-8

**Last Evaluation of the Negative of the Hessian**

	Intercept	sexFemale	age
Intercept	6.4597397447	4.6042218284	292.04051848
sexFemale	4.6042218284	4.6042218284	216.20829515
age	292.04051848	216.20829515	13487.329973

**Analysis of Maximum Likelihood Parameter Estimates**

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq	
Intercept	1	-5.7620	2.7635	-11.1783	-0.3458	4.35	0.0371	
sex	Female	1	-2.4224	0.9559	-4.2959	-0.5489	6.42	0.0113
sex	Male	0	0.0000	.	.	.	.	.
age	1	0.1649	0.0652	0.0371	0.2927	6.40	0.0114	

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:

```
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

Obs	sex	age	Predicted
1	Female	35	0.082205

---

## 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:

```
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
;
```

```

proc probit optc lackfit covout data=epidemic
    outest = out1 xdata = xval
    Plots=(predpplot ipplot 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

Class Level Information		
Name	Levels	Values
treat	2	A B
sex	2	0 1

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

#### Output 94.4.2 Parameter Information

Parameter Information			
Parameter	Effect	treat	sex
Intercept	Intercept		
dose	dose		
treatA	treat	A	
treatB	treat	B	
sex0	sex		0
sex1	sex		1
treatAsex0	treat*sex	A	0
treatAsex1	treat*sex	A	1
treatBsex0	treat*sex	B	0
treatBsex1	treat*sex	B	1

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**

Model Information	
Data Set	WORK.EPIDEMIC
Events Variable	r
Trials Variable	n
Number of Observations	10
Number of Events	1011
Number of Trials	1204
Name of Distribution	Normal
Log Likelihood	-387.2467391

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

Goodness-of-Fit Tests				
Statistic	Value	DF	Value/DF	Pr > ChiSq
Pearson Chi-Square	4.9317	4	1.2329	0.2944
L.R. Chi-Square	5.7079	4	1.4270	0.2220

Response-Covariate Profile	
Response Levels	2
Number of Covariate Values	10

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.

**Output 94.4.5** Type III Tests

Type III Analysis of Effects			
Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
dose	1	42.1691	<.0001
treat	1	16.1421	<.0001
sex	1	1.7710	0.1833
treat*sex	1	13.9343	0.0002

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%.

**Output 94.4.6** Analysis of Parameter Estimates

Analysis of Maximum Likelihood Parameter Estimates								
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq	
				Intercept	1			-0.8871
dose	1	1.6774	0.2583	1.1711	2.1837	42.17	<.0001	
treat	A	1	-1.2537	0.2616	-1.7664	-0.7410	22.97	<.0001
treat	B	0	0.0000	.	.	.	.	.
sex	0	1	-0.4633	0.2289	-0.9119	-0.0147	4.10	0.0429
sex	1	0	0.0000	.	.	.	.	.
treat*sex	A 0	1	1.2899	0.3456	0.6126	1.9672	13.93	0.0002
treat*sex	A 1	0	0.0000	.	.	.	.	.
treat*sex	B 0	0	0.0000	.	.	.	.	.
treat*sex	B 1	0	0.0000	.	.	.	.	.
_C_	1	0.2735	0.0946	0.0881	0.4589			

From Table 94.4.6, you can see the following results:

- The variable 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

Estimated Covariance Matrix						
	Intercept	dose	treatA	sex0	treatAsex0	_C_
Intercept	0.131944	-0.087353	0.053551	0.030285	-0.067056	-0.028073
dose	-0.087353	0.066723	-0.047506	-0.034081	0.058620	0.018196
treatA	0.053551	-0.047506	0.068425	0.036063	-0.075323	-0.017084
sex0	0.030285	-0.034081	0.036063	0.052383	-0.063599	-0.008088
treatAsex0	-0.067056	0.058620	-0.075323	-0.063599	0.119408	0.019134
_C_	-0.028073	0.018196	-0.017084	-0.008088	0.019134	0.008948

#### Output 94.4.8 Estimated Correlation Matrix

Estimated Correlation Matrix						
	Intercept	dose	treatA	sex0	treatAsex0	_C_
Intercept	1.000000	-0.930998	0.563595	0.364284	-0.534227	-0.817027
dose	-0.930998	1.000000	-0.703083	-0.576477	0.656744	0.744699
treatA	0.563595	-0.703083	1.000000	0.602359	-0.833299	-0.690420
sex0	0.364284	-0.576477	0.602359	1.000000	-0.804154	-0.373565
treatAsex0	-0.534227	0.656744	-0.833299	-0.804154	1.000000	0.585364
_C_	-0.817027	0.744699	-0.690420	-0.373565	0.585364	1.000000

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.

**Output 94.4.9** Probit Analysis on Dose**The Probit Procedure**

Probit Analysis on dose			
Probability	dose	95% Fiducial Limits	
0.01	-0.85801	-1.81301	-0.33743
0.02	-0.69549	-1.58167	-0.21116
0.03	-0.59238	-1.43501	-0.13093
0.04	-0.51482	-1.32476	-0.07050
0.05	-0.45172	-1.23513	-0.02130
0.06	-0.39802	-1.15888	0.02063
0.07	-0.35093	-1.09206	0.05742
0.08	-0.30877	-1.03226	0.09039
0.09	-0.27043	-0.97790	0.12040
0.10	-0.23513	-0.92788	0.14805
0.15	-0.08900	-0.72107	0.26278
0.20	0.02714	-0.55706	0.35434
0.25	0.12678	-0.41669	0.43322
0.30	0.21625	-0.29095	0.50437
0.35	0.29917	-0.17477	0.57064
0.40	0.37785	-0.06487	0.63387
0.45	0.45397	0.04104	0.69546
0.50	0.52888	0.14481	0.75654
0.55	0.60380	0.24800	0.81819
0.60	0.67992	0.35213	0.88157
0.65	0.75860	0.45879	0.94803
0.70	0.84151	0.56985	1.01942
0.75	0.93099	0.68770	1.09847
0.80	1.03063	0.81571	1.18970
0.85	1.14677	0.95926	1.30171
0.90	1.29290	1.12867	1.45386
0.91	1.32819	1.16747	1.49273
0.92	1.36654	1.20867	1.53590
0.93	1.40870	1.25284	1.58450
0.94	1.45579	1.30084	1.64012
0.95	1.50949	1.35397	1.70515
0.96	1.57258	1.41443	1.78353
0.97	1.65015	1.48626	1.88238
0.98	1.75326	1.57833	2.01720
0.99	1.91577	1.71776	2.23537

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

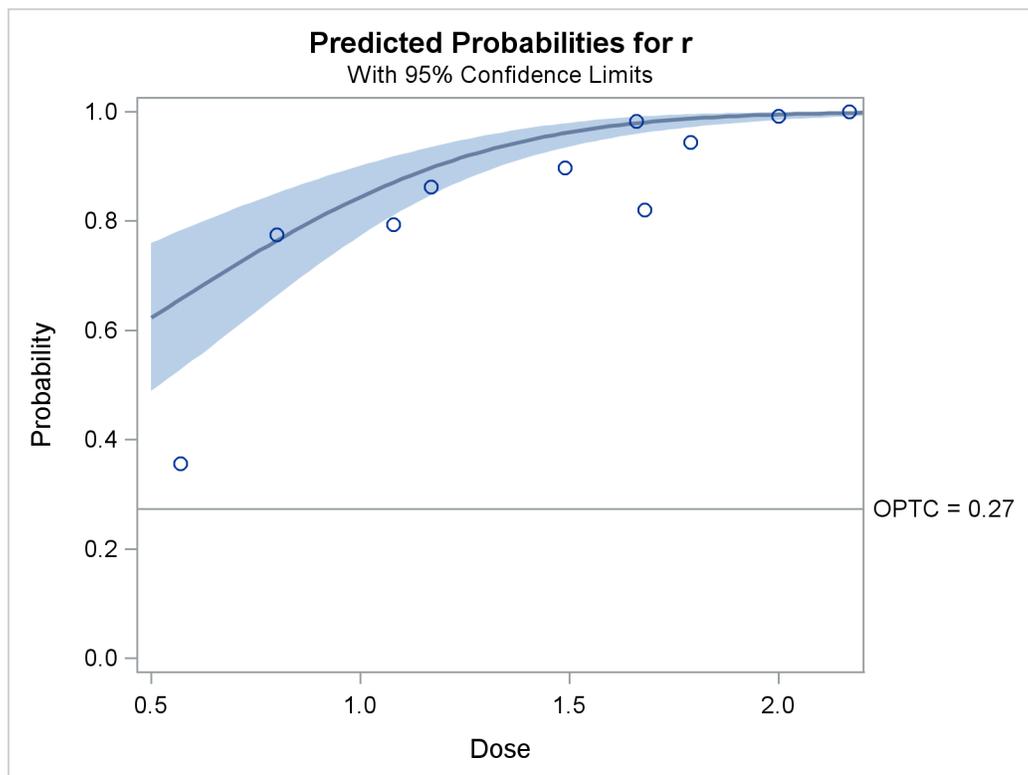
Obs	_MODEL_	_NAME_	_TYPE_	_DIST_	_STATUS_	_LNLIKE_	r	Intercept	dose
1	r	PARMS	Normal	0	Converged	-387.247	-1.00000	-0.88714	1.67739
2	Intercept	COV	Normal	0	Converged	-387.247	-0.88714	0.13194	-0.08735
3	dose	COV	Normal	0	Converged	-387.247	1.67739	-0.08735	0.06672
4	treatA	COV	Normal	0	Converged	-387.247	-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

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

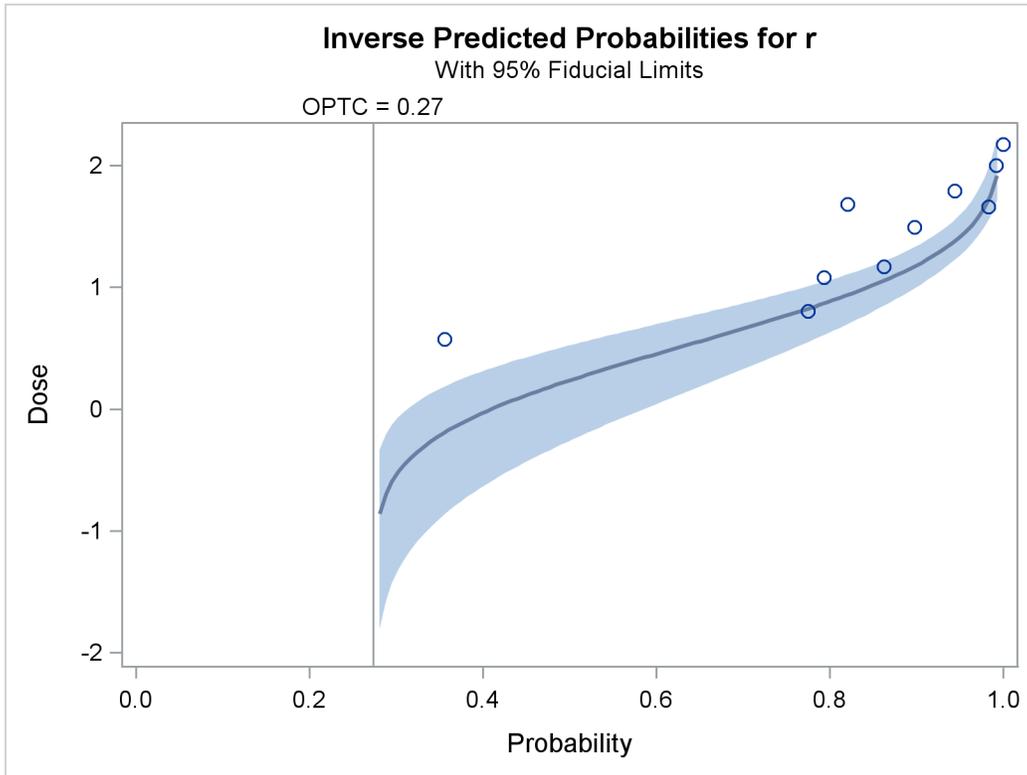
The plots in the following three outputs, [Output 94.4.11](#), [Output 94.4.12](#), and [Output 94.4.13](#), are generated by the PLOTS= option. The first plot, specified with the PREDPLOT option, is the plot of the predicted probability against the first single continuous variable dose in the MODEL 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  $x'$ , against the first single continuous variable with the Wald confidence intervals.

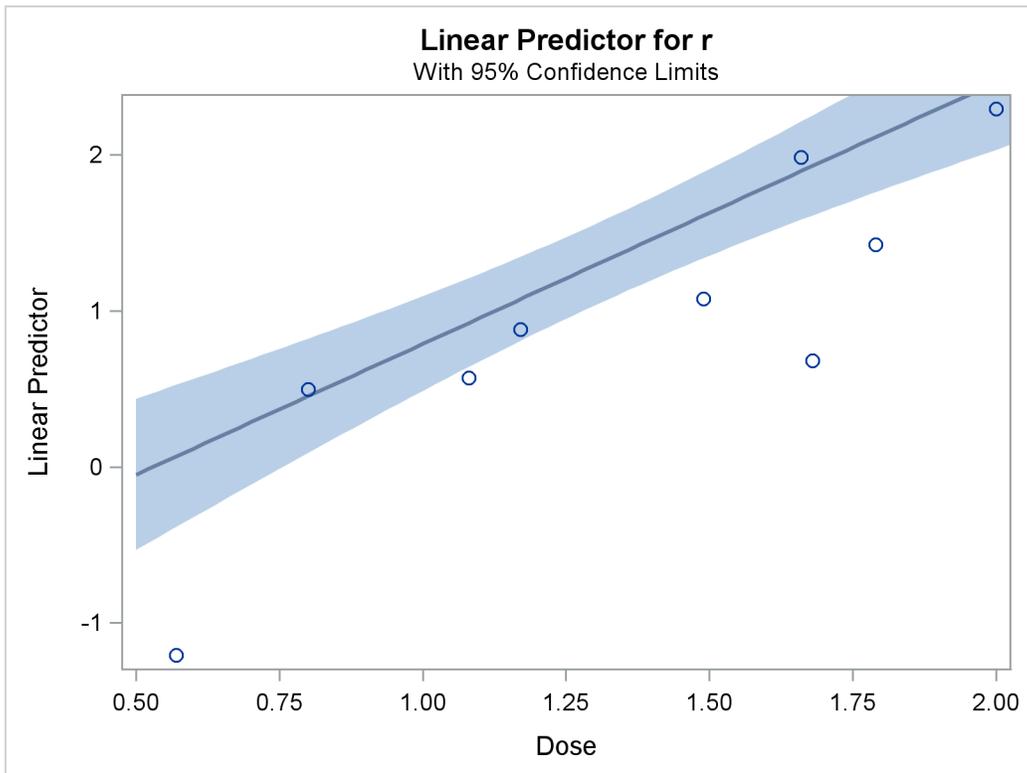
**Output 94.4.11** Predicted Probability Plot



**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`:

```
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`

The Probit Procedure			
Chi-Square Test for <code>treat*sex</code> Least Squares Means Slice			
	Num		
Slice	DF	Chi-Square	Pr > ChiSq
treat A	1	18.52	<.0001
Chi-Square Test for <code>treat*sex</code> Least Squares Means Slice			
	Num		
Slice	DF	Chi-Square	Pr > ChiSq
treat B	1	2.65	0.1035
Chi-Square Test for <code>treat*sex</code> Least Squares Means Slice			
	Num		
Slice	DF	Chi-Square	Pr > ChiSq
sex 0	1	0.00	0.9579
Chi-Square Test for <code>treat*sex</code> Least Squares Means Slice			
	Num		
Slice	DF	Chi-Square	Pr > ChiSq
sex 1	1	47.43	<.0001

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

Simple Differences of treat*sex Least Squares Means						
Slice	sex	_sex	Estimate	Standard Error	z Value	Pr >  z
treat A	0	1	0.5957	0.1384	4.30	<.0001

Simple Differences of treat*sex Least Squares Means						
Slice	sex	_sex	Estimate	Standard Error	z Value	Pr >  z
treat B	0	1	-0.2956	0.1816	-1.63	0.1035

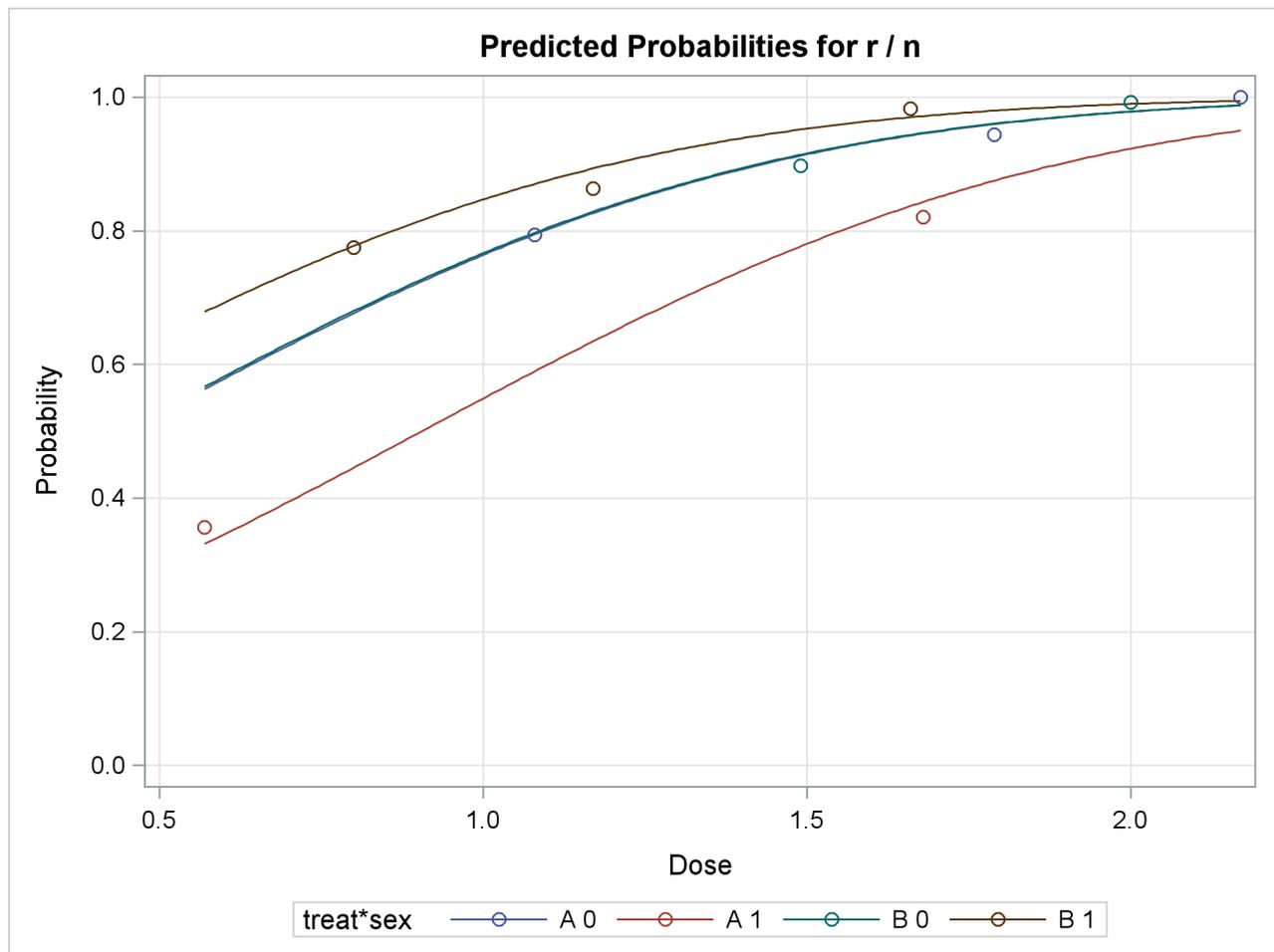
  

Simple Differences of treat*sex Least Squares Means						
Slice	treat	_treat	Estimate	Standard Error	z Value	Pr >  z
sex 0	A	B	-0.00899	0.1702	-0.05	0.9579

Simple Differences of treat*sex Least Squares Means						
Slice	treat	_treat	Estimate	Standard Error	z Value	Pr >  z
sex 1	A	B	-0.9003	0.1307	-6.89	<.0001

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

- Agresti, A. (2002). *Categorical Data Analysis*. 2nd ed. New York: John Wiley & Sons.
- Collett, D. (2003). *Modelling Binary Data*. 2nd ed. London: Chapman & Hall.
- Cox, D. R., and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman & Hall.
- Finney, D. J. (1971). *Probit Analysis*. 3rd ed. Cambridge: Cambridge University Press.
- Hubert, J. J., Bohidar, N. R., and Peace, K. E. (1988). "Assessment of Pharmacological Activity." In *Biopharmaceutical Statistics for Drug Development*, edited by K. E. Peace, 83–145. New York: Marcel Dekker.

# 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
  - 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

## PREDPLOT

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

PREDPLOT, 7622

response level ordering, 7617

subpopulation, 7618, 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
  - PROBIT procedure, 7595
- INVERSECL option
  - PROC PROBIT statement, 7579
- IPPPLOT statement
  - options summarized by function, 7597
  - 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
  - PROBIT procedure, 7605
- LPREFIX= option
  - CLASS statement (PROBIT), 7592
- LSMEANS statement
  - PROBIT procedure, 7613
- LSMESTIMATE 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
  - PREDPPLOT 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
- PREDPPLOT statement
  - options summarized by function, 7623
  - 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  
 INBORDER option, 7588  
 INHEIGHT= 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  
 NVTICK= option, 7589, 7602  
 options, 7584  
 THRESHLABPOS= option, 7590  
 TURNVLABELS option, 7589  
 VAR= option, 7583  
 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  
 PROBIT procedure, CDFPLOT statement  
 VOFFSET= option, 7590  
 PROBIT procedure, CLASS statement, 7591  
 CPREFIX= option, 7591  
 DESCENDING option, 7591  
 LPREFIX= option, 7592  
 MISSING option, 7592  
 ORDER= option, 7592  
 PARAM= option, 7592  
 REF= option, 7592  
 TRUNCATE option, 7593  
 PROBIT procedure, EFFECTPLOT statement, 7593  
 PROBIT procedure, ESTIMATE statement, 7594  
 PROBIT procedure, INSET statement, 7595  
 keywords, 7595  
 PROBIT procedure, IPLOT statement  
 INHEIGHT= option, 7602  
 PROBIT procedure, IPPLOT 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  
 HREF= option, 7601  
 HREFLABELS= option, 7601  
 HREFLABPOS= option, 7602  
 HUPPER= option, 7601  
 INBORDER option, 7602  
 LFIT option, 7602  
 LGRID option, 7602  
 LHREF= option, 7602  
 LVREF= option, 7602  
 NAME= option, 7602  
 NOCONF option, 7602  
 NODATA option, 7603  
 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

VLOWER= option, 7604  
 VREF= option, 7604  
 VREFLABELS= option, 7604  
 VREFLABPOS= option, 7604  
 VUPPER= option, 7604  
 WAXIS= option, 7604  
 WFIT= option, 7604  
 WGRID= option, 7604  
 WREFL= option, 7604  
 PROBIT procedure, IPPPLOT statement  
   VOFFSET= option, 7604  
 PROBIT procedure, LPLOT statement  
   INHEIGHT= option, 7610  
   NHTICK= option, 7611  
   NVTICK= option, 7611  
 PROBIT procedure, LPREDPLOT statement, 7605  
   ANNOTATE= option, 7608  
   CAXIS= option, 7608  
   CFIT= option, 7608  
   CFRAME= option, 7608  
   CGRID= option, 7608  
   CHREF= option, 7608  
   CTEXT= option, 7609  
   CVREF= option, 7609  
   DESCRIPTION= option, 7609  
   FONT= option, 7609  
   HAXIS= option, 7609  
   HEIGHT= option, 7609  
   HLOWER= option, 7609  
   HOFFSET= option, 7609  
   HREF= option, 7609  
   HREFLABELS= option, 7610  
   HREFLABPOS= option, 7610  
   HUPPER= option, 7609  
   INBORDER option, 7610  
   LEVEL option, 7610  
   LFIT option, 7610  
   LGRID option, 7610  
   LHREF= option, 7610  
   LVREF= option, 7610  
   NAME= option, 7611  
   NOCONF option, 7611  
   NODATA option, 7611  
   NOFIT option, 7611  
   NOFRAME option, 7611  
   NOGRID option, 7611  
   NOHLABEL option, 7611  
   NOHTICK option, 7611  
   NOTHRESH option, 7611  
   NOVLABEL option, 7611  
   NOVTICK option, 7611  
   options, 7605  
   THRESHLABPOS= option, 7611  
   TURNVLABELS option, 7611  
   VAR= option, 7605  
   VAXIS= option, 7612  
   VAXISLABEL= option, 7612  
   VLOWER= option, 7612  
   VOFFSET= option, 7612  
   VREF= option, 7612  
   VREFLABELS= option, 7612  
   VREFLABPOS= option, 7612  
   VUPPER= option, 7612  
   WAXIS= option, 7613  
   WFIT= option, 7613  
   WGRID= option, 7613  
   WREFL= option, 7613  
 PROBIT procedure, LSMEANS statement, 7613  
 PROBIT procedure, LSMESTIMATE statement, 7614  
 PROBIT procedure, MODEL statement, 7615  
   AGGREGATE= option, 7618  
   ALPHA= option, 7618  
   CONVERGE option, 7618  
   CORRB option, 7619  
   COVB option, 7619  
   DESCENDING option, 7617  
   DISTRIBUTION= option, 7619  
   EVENT= option, 7617  
   HPROB= option, 7619  
   INITIAL option, 7619  
   INTERCEPT= option, 7620  
   INVERSECL option, 7620  
   ITPRINT option, 7620  
   LACKFIT option, 7620  
   MAXITER= option, 7620  
   NOINT option, 7620  
   ORDER= option, 7617  
   REFERENCE= option, 7618  
   SCALE= option, 7621  
   SINGULAR= option, 7621  
 PROBIT procedure, OUTPUT statement, 7621  
 PROBIT procedure, PREDPLOT statement  
   INHEIGHT= option, 7628  
   LEVEL option, 7628  
   NHTICK= option, 7629  
   NVTICK= option, 7629  
 PROBIT procedure, PREDPPLOT statement, 7622  
   ANNOTATE= option, 7626  
   CAXIS= option, 7626  
   CFIT= option, 7626  
   CFRAME= option, 7626  
   CGRID= option, 7626  
   CHREF= option, 7626  
   CTEXT= option, 7627  
   CVREF= option, 7627  
   DESCRIPTION= option, 7627  
   FONT= option, 7627  
   HAXIS= option, 7627

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  
 NOVTICK option, 7629  
 options, 7623  
 THRESHLABPOS= option, 7629  
 TURNVLABELS 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  
 IPPLOT statement (PROBIT), 7597  
 LPREDPLOT statement (PROBIT), 7605  
 PREDPLOT statement (PROBIT), 7622  
 XDATA= option  
 PROC PROBIT statement, 7582