SAS/STAT® 12.3 User's Guide
The LOGISTIC Procedure
(Chapter)

This document is an individual chapter from *SAS/STAT® 12.3 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2013. *SAS/STAT® 12.3 User's Guide*. Cary, NC: SAS Institute Inc.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

July 2013

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at **support.sas.com/bookstore** or call 1-800-727-3228.

Chapter 54
The LOGISTIC Procedure

Contents

Overview: LOGISTIC Procedure

Binary responses (for example, success and failure), ordinal responses (for example, normal, mild, and severe), and nominal responses (for example, major TV networks viewed at a certain hour) arise in many fields of study. Logistic regression analysis is often used to investigate the relationship between these discrete responses and a set of explanatory variables. Texts that discuss logistic regression include Agresti (2002); Allison (1999); Collett (2003); Cox and Snell (1989); Hosmer and Lemeshow (2000); Stokes, Davis, and Koch (2012).

For binary response models, the response, Y, of an individual or an experimental unit can take on one of two possible values, denoted for convenience by 1 and 2 (for example, $Y = 1$ if a disease is present, otherwise $Y = 2$). Suppose $\mathbf{x}$ is a vector of explanatory variables and $\pi = \Pr(Y = 1 \mid \mathbf{x})$ is the response probability to be modeled. The linear logistic model has the form

$$\text{logit}(\pi) \equiv \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \boldsymbol{\beta}'\mathbf{x}$$

where α is the intercept parameter and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)'$ is the vector of s slope parameters. Notice that the LOGISTIC procedure, by default, models the probability of the *lower* response levels.

The logistic model shares a common feature with a more general class of linear models: a function $g = g(\mu)$ of the mean of the response variable is assumed to be linearly related to the explanatory variables. Since the mean μ implicitly depends on the stochastic behavior of the response, and the explanatory variables are assumed to be fixed, the function g provides the link between the random (stochastic) component and the systematic (deterministic) component of the response variable Y. For this reason, Nelder and Wedderburn (1972) refer to $g(\mu)$ as a link function. One advantage of the logit function over other link functions is that differences on the logistic scale are interpretable regardless of whether the data are sampled prospectively or retrospectively (McCullagh and Nelder 1989, Chapter 4). Other link functions that are widely used in practice are the probit function and the complementary log-log function. The LOGISTIC procedure enables you to choose one of these link functions, resulting in fitting a broader class of binary response models of the form

$$g(\pi) = \alpha + \boldsymbol{\beta}'\mathbf{x}$$

For ordinal response models, the response, Y, of an individual or an experimental unit might be restricted to one of a (usually small) number of ordinal values, denoted for convenience by $1, \ldots, k, k+1$. For example, the severity of coronary disease can be classified into three response categories as 1=no disease, 2=angina pectoris, and 3=myocardial infarction. The LOGISTIC 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

$$g(\Pr(Y \le i \mid \mathbf{x})) = \alpha_i + \boldsymbol{\beta}'\mathbf{x}, \quad i = 1, \ldots, k$$

where $\alpha_1, \ldots, \alpha_k$ are k intercept parameters, and $\boldsymbol{\beta}$ is the vector of slope parameters. This model has been considered by many researchers. Aitchison and Silvey (1957) and Ashford (1959) employ a probit scale and provide a maximum likelihood analysis; Walker and Duncan (1967) and Cox and Snell (1989) discuss the use of the log odds scale. For the log odds scale, the cumulative logit model is often referred to as the *proportional odds* model.

For nominal response logistic models, where the $k+1$ possible responses have no natural ordering, the logit model can also be extended to a *multinomial* model known as a *generalized* or *baseline-category* logit model, which has the form

$$\log\left(\frac{\Pr(Y=i \mid \mathbf{x})}{\Pr(Y=k+1 \mid \mathbf{x})}\right) = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}, \quad i = 1,\ldots,k$$

where the $\alpha_1,\ldots,\alpha_k$ are k intercept parameters, and the $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_k$ are k vectors of slope parameters. These models are a special case of the *discrete choice* or *conditional logit* models introduced by McFadden (1974).

The LOGISTIC procedure fits linear logistic regression models for discrete response data by the method of maximum likelihood. It can also perform conditional logistic regression for binary response data and exact logistic regression for binary and nominal response data. The maximum likelihood estimation is carried out with either the Fisher scoring algorithm or the Newton-Raphson algorithm, and you can perform the bias-reducing penalized likelihood optimization as discussed by Firth (1993) and Heinze and Schemper (2002). You can specify starting values for the parameter estimates. The logit link function in the logistic regression models can be replaced by the probit function, the complementary log-log function, or the generalized logit function.

Any term specified in the model is referred to as an *effect*. The LOGISTIC procedure enables you to specify categorical variables (also known as *classification* or *CLASS variables*) and continuous variables as explanatory effects. You can also specify more complex model terms such as interactions and nested terms in the same way as in the GLM procedure. You can create complex *constructed effects* with the EFFECT statement. An effect in the model that is not an interaction or a nested term or a constructed effect is referred to as a *main effect*.

The LOGISTIC procedure allows either a full-rank parameterization or a less-than-full-rank parameterization of the CLASS variables. The full-rank parameterization offers eight coding methods: effect, reference, ordinal, polynomial, and orthogonalizations of these. The effect coding is the same method that is used in the CATMOD procedure. The less-than-full-rank parameterization, often called *dummy coding*, is the same coding as that used in the GLM procedure.

The LOGISTIC procedure provides four effect selection methods: forward selection, backward elimination, stepwise selection, and best subset selection. The best subset selection is based on the likelihood score statistic. This method identifies a specified number of best models containing one, two, three effects, and so on, up to a single model containing effects for all the explanatory variables.

The LOGISTIC procedure has some additional options to control how to move effects in and out of a model with the forward selection, backward elimination, or stepwise selection model-building strategies. When there are no interaction terms, a main effect can enter or leave a model in a single step based on the p-value of the score or Wald statistic. When there are interaction terms, the selection process also depends on whether you want to preserve model hierarchy. These additional options enable you to specify whether model hierarchy is to be preserved, how model hierarchy is applied, and whether a single effect or multiple effects can be moved in a single step.

Odds ratio estimates are displayed along with parameter estimates. You can also specify the change in the continuous explanatory main effects for which odds ratio estimates are desired. Confidence intervals for the regression parameters and odds ratios can be computed based either on the profile-likelihood function or on the asymptotic normality of the parameter estimators. You can also produce odds ratios for effects that are involved in interactions or nestings, and for any type of parameterization of the CLASS variables.

Various methods to correct for overdispersion are provided, including Williams' method for grouped binary response data. The adequacy of the fitted model can be evaluated by various goodness-of-fit tests, including the Hosmer-Lemeshow test for binary response data.

Like many procedures in SAS/STAT software that enable the specification of CLASS variables, the LOGISTIC procedure provides a CONTRAST statement for specifying customized hypothesis tests concerning the model parameters. The CONTRAST statement also provides estimation of individual rows of contrasts, which is particularly useful for obtaining odds ratio estimates for various levels of the CLASS variables. The LOGISTIC procedure also provides testing capability through the ESTIMATE and TEST statements. Analyses of LS-means are enabled with the LSMEANS, LSMESTIMATE, and SLICE statements.

You can perform a conditional logistic regression on binary response data by specifying the STRATA statement. This enables you to perform matched-set and case-control analyses. The number of events and nonevents can vary across the strata. Many of the features available with the unconditional analysis are also available with a conditional analysis.

The LOGISTIC procedure enables you to perform exact logistic regression, also known as exact conditional logistic regression, by specifying one or more EXACT statements. You can test individual parameters or conduct a joint test for several parameters. The procedure computes two exact tests: the exact conditional score test and the exact conditional probability test. You can request exact estimation of specific parameters and corresponding odds ratios where appropriate. Point estimates, standard errors, and confidence intervals are provided. You can perform stratified exact logistic regression by specifying the STRATA statement.

Further features of the LOGISTIC procedure enable you to do the following:

- control the ordering of the response categories
- compute a generalized R Square measure for the fitted model
- reclassify binary response observations according to their predicted response probabilities
- test linear hypotheses about the regression parameters
- create a data set for producing a receiver operating characteristic curve for each fitted model
- specify contrasts to compare several receiver operating characteristic curves
- create a data set containing the estimated response probabilities, residuals, and influence diagnostics
- score a data set by using a previously fitted model

The LOGISTIC 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 more information about the plots implemented in PROC LOGISTIC, see the section "ODS Graphics" on page 4294.

The remaining sections of this chapter describe how to use PROC LOGISTIC and discuss the underlying statistical methodology. The section "Getting Started: LOGISTIC Procedure" on page 4166 introduces PROC LOGISTIC with an example for binary response data. The section "Syntax: LOGISTIC Procedure" on page 4174 describes the syntax of the procedure. The section "Details: LOGISTIC Procedure" on page 4237 summarizes the statistical technique employed by PROC LOGISTIC. The section "Examples: LOGISTIC Procedure" on page 4295 illustrates the use of the LOGISTIC procedure.

For more examples and discussion on the use of PROC LOGISTIC, see Stokes, Davis, and Koch (2012); Allison (1999); SAS Institute Inc. (1995).

Getting Started: LOGISTIC Procedure

The LOGISTIC procedure is similar in use to the other regression procedures in the SAS System. To demonstrate the similarity, suppose the response variable y is binary or ordinal, and x1 and x2 are two explanatory variables of interest. To fit a logistic regression model, you can specify a MODEL statement similar to that used in the REG procedure. For example:

```
proc logistic;
   model y=x1 x2;
run;
```

The response variable y can be either character or numeric. PROC LOGISTIC enumerates the total number of response categories and orders the response levels according to the response variable option ORDER= in the MODEL statement.

You can also input binary response data that are grouped. In the following statements, n represents the number of trials and r represents the number of events:

```
proc logistic;
   model r/n=x1 x2;
run;
```

The following example illustrates the use of PROC LOGISTIC. The data, taken from Cox and Snell (1989, pp. 10–11), consist of the number, r, of ingots not ready for rolling, out of n tested, for a number of combinations of heating time and soaking time.

```
data ingots;
   input Heat Soak r n @@;
   datalines;
7 1.0 0 10  14 1.0 0 31  27 1.0 1 56  51 1.0 3 13
7 1.7 0 17  14 1.7 0 43  27 1.7 4 44  51 1.7 0  1
7 2.2 0  7  14 2.2 2 33  27 2.2 0 21  51 2.2 0  1
7 2.8 0 12  14 2.8 0 31  27 2.8 1 22  51 4.0 0  1
7 4.0 0  9  14 4.0 0 19  27 4.0 1 16
;
```

The following invocation of PROC LOGISTIC fits the binary logit model to the grouped data. The continuous covariates Heat and Soak are specified as predictors, and the bar notation ("|") includes their interaction, Heat*Soak. The ODDSRATIO statement produces odds ratios in the presence of interactions, and a graphical display of the requested odds ratios is produced when ODS Graphics is enabled.

```
ods graphics on;
proc logistic data=ingots;
   model r/n = Heat | Soak;
   oddsratio Heat / at(Soak=1 2 3 4);
run;
ods graphics off;
```

The results of this analysis are shown in the following figures. PROC LOGISTIC first lists background information in Figure 54.1 about the fitting of the model. Included are the name of the input data set, the response variable(s) used, the number of observations used, and the link function used.

Figure 54.1 Binary Logit Model

```
                    The LOGISTIC Procedure

                      Model Information

Data Set                          WORK.INGOTS
Response Variable (Events)        r
Response Variable (Trials)        n
Model                             binary logit
Optimization Technique            Fisher's scoring


    Number of Observations Read          19
    Number of Observations Used          19
    Sum of Frequencies Read             387
    Sum of Frequencies Used             387
```

The "Response Profile" table (Figure 54.2) lists the response categories (which are Event and Nonevent when grouped data are input), their ordered values, and their total frequencies for the given data.

Figure 54.2 Response Profile with Events/Trials Syntax

```
                      Response Profile

          Ordered     Binary           Total
            Value     Outcome      Frequency

                1     Event               12
                2     Nonevent           375


                  Model Convergence Status

       Convergence criterion (GCONV=1E-8) satisfied.
```

The "Model Fit Statistics" table (Figure 54.3) contains Akaike's information criterion (AIC), the Schwarz criterion (SC), and the negative of twice the log likelihood (–2 Log L) for the intercept-only model and the fitted model. AIC and SC can be used to compare different models, and the ones with smaller values are preferred. Results of the likelihood ratio test and the efficient score test for testing the joint significance of the explanatory variables (Soak, Heat, and their interaction) are included in the "Testing Global Null Hypothesis: BETA=0" table (Figure 54.3); the small p-values reject the hypothesis that all slope parameters are equal to zero.

Figure 54.3 Fit Statistics and Hypothesis Tests

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates Log Likelihood	Full Log Likelihood
AIC	108.988	103.222	35.957
SC	112.947	119.056	51.791
-2 Log L	106.988	95.222	27.957

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11.7663	3	0.0082
Score	16.5417	3	0.0009
Wald	13.4588	3	0.0037

The "Analysis of Maximum Likelihood Estimates" table in Figure 54.4 lists the parameter estimates, their standard errors, and the results of the Wald test for individual parameters. Note that the Heat*Soak parameter is not significantly different from zero (p=0.727), nor is the Soak variable (p=0.6916).

Figure 54.4 Parameter Estimates

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.9901	1.6666	12.9182	0.0003
Heat	1	0.0963	0.0471	4.1895	0.0407
Soak	1	0.2996	0.7551	0.1574	0.6916
Heat*Soak	1	-0.00884	0.0253	0.1219	0.7270

The "Association of Predicted Probabilities and Observed Responses" table (Figure 54.5) contains four measures of association for assessing the predictive ability of a model. They are based on the number of pairs of observations with different response values, the number of concordant pairs, and the number of discordant pairs, which are also displayed. Formulas for these statistics are given in the section "Rank Correlation of Observed Responses and Predicted Probabilities" on page 4253.

Figure 54.5 Association Table

Association of Predicted Probabilities and Observed Responses

Percent Concordant	70.9	Somers' D	0.537
Percent Discordant	17.3	Gamma	0.608
Percent Tied	11.8	Tau-a	0.032
Pairs	4500	c	0.768

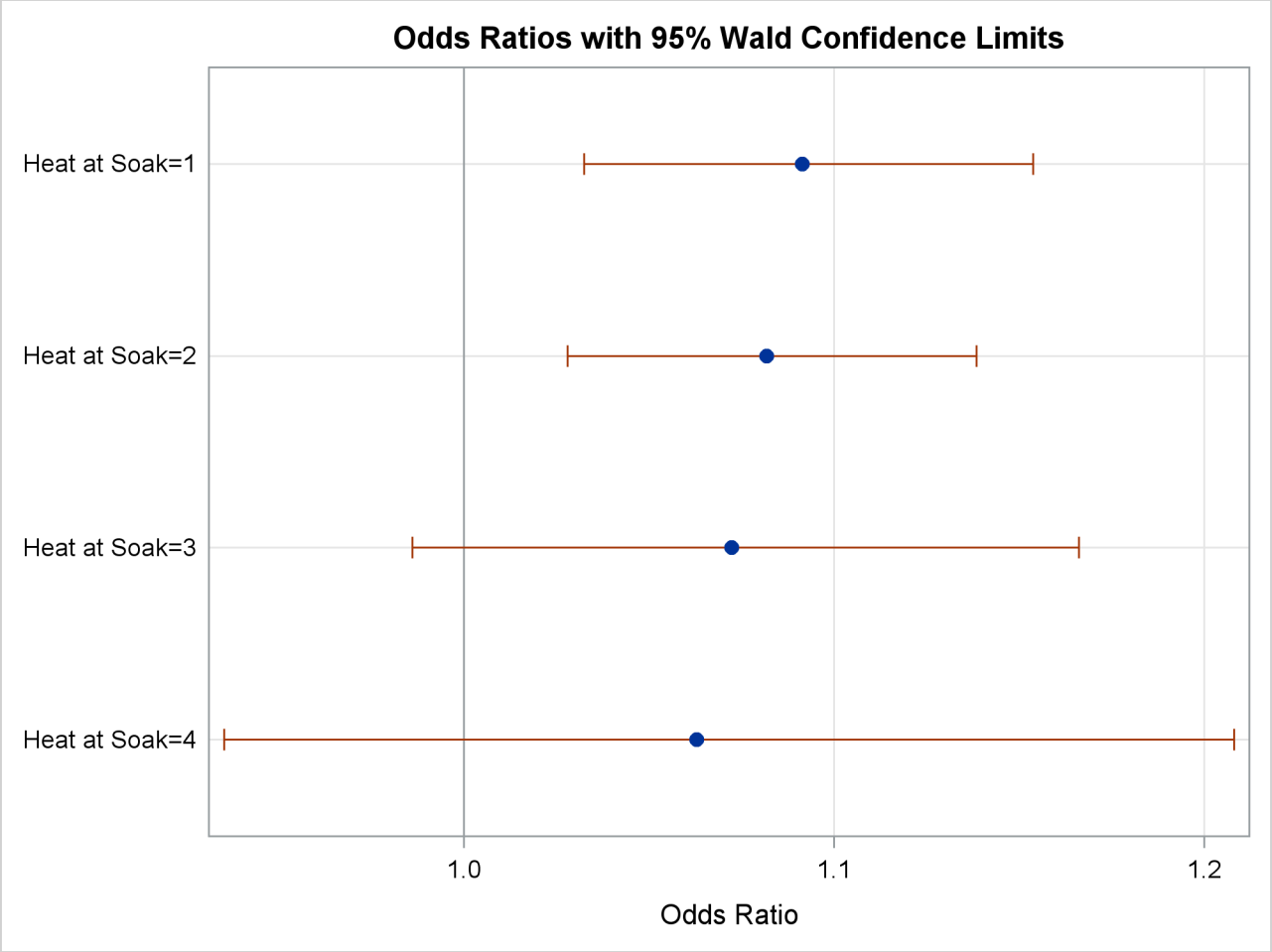
The ODDSRATIO statement produces the "Odds Ratio Estimates and Wald Confidence Intervals" table (Figure 54.6), and a graphical display of these estimates is shown in Figure 54.7. The differences between the odds ratios are small compared to the variability shown by their confidence intervals, which confirms the previous conclusion that the Heat*Soak parameter is not significantly different from zero.

Figure 54.6 Odds Ratios of Heat at Several Values of Soak

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
Heat at Soak=1	1.091	1.032	1.154
Heat at Soak=2	1.082	1.028	1.139
Heat at Soak=3	1.072	0.986	1.166
Heat at Soak=4	1.063	0.935	1.208

Figure 54.7 Plot of Odds Ratios of Heat at Several Values of Soak



Since the Heat*Soak interaction is nonsignificant, the following statements fit a main-effects model:

```
proc logistic data=ingots;
   model r/n = Heat Soak;
run;
```

The results of this analysis are shown in the following figures. The model information and response profiles are the same as those in Figure 54.1 and Figure 54.2 for the saturated model. The "Model Fit Statistics" table in Figure 54.8 shows that the AIC and SC for the main-effects model are smaller than for the saturated model, indicating that the main-effects model might be the preferred model. As in the preceding model, the "Testing Global Null Hypothesis: BETA=0" table indicates that the parameters are significantly different from zero.

Figure 54.8 Fit Statistics and Hypothesis Tests

The LOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates Log Likelihood	Intercept and Covariates Full Log Likelihood
AIC	108.988	101.346	34.080
SC	112.947	113.221	45.956
-2 Log L	106.988	95.346	28.080

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11.6428	2	0.0030
Score	15.1091	2	0.0005
Wald	13.0315	2	0.0015

The "Analysis of Maximum Likelihood Estimates" table in Figure 54.9 again shows that the Soak parameter is not significantly different from zero (p=0.8639). The odds ratio for each effect parameter, estimated by exponentiating the corresponding parameter estimate, is shown in the "Odds Ratios Estimates" table (Figure 54.9), along with 95% Wald confidence intervals. The confidence interval for the Soak parameter contains the value 1, which also indicates that this effect is not significant.

Figure 54.9 Parameter Estimates and Odds Ratios

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.5592	1.1197	24.6503	<.0001
Heat	1	0.0820	0.0237	11.9454	0.0005
Soak	1	0.0568	0.3312	0.0294	0.8639

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Heat	1.085	1.036	1.137
Soak	1.058	0.553	2.026

Figure 54.9 *continued*

```
      Association of Predicted Probabilities and Observed Responses

           Percent Concordant     64.4    Somers' D    0.460
           Percent Discordant     18.4    Gamma        0.555
           Percent Tied           17.2    Tau-a        0.028
           Pairs                  4500    c            0.730
```

Using these parameter estimates, you can calculate the estimated logit of π as

$$-5.5592 + 0.082 \times \text{Heat} + 0.0568 \times \text{Soak}$$

For example, if Heat=7 and Soak=1, then $\text{logit}(\widehat{\pi}) = -4.9284$. Using this logit estimate, you can calculate $\widehat{\pi}$ as follows:

$$\widehat{\pi} = 1/(1 + \mathrm{e}^{4.9284}) = 0.0072$$

This gives the predicted probability of the event (ingot not ready for rolling) for Heat=7 and Soak=1. Note that PROC LOGISTIC can calculate these statistics for you; use the OUTPUT statement with the PREDICTED= option, or use the SCORE statement.

To illustrate the use of an alternative form of input data, the following program creates the ingots data set with the new variables NotReady and Freq instead of n and r. The variable NotReady represents the response of individual units; it has a value of 1 for units not ready for rolling (event) and a value of 0 for units ready for rolling (nonevent). The variable Freq represents the frequency of occurrence of each combination of Heat, Soak, and NotReady. Note that, compared to the previous data set, NotReady=1 implies Freq=r, and NotReady=0 implies Freq=n–r.

```
data ingots;
   input Heat Soak NotReady Freq @@;
   datalines;
7 1.0 0 10  14 1.0 0 31  14 4.0 0 19  27 2.2 0 21  51 1.0 1  3
7 1.7 0 17  14 1.7 0 43  27 1.0 1  1  27 2.8 1  1  51 1.0 0 10
7 2.2 0  7  14 2.2 1  2  27 1.0 0 55  27 2.8 0 21  51 1.7 0  1
7 2.8 0 12  14 2.2 0 31  27 1.7 1  4  27 4.0 1  1  51 2.2 0  1
7 4.0 0  9  14 2.8 0 31  27 1.7 0 40  27 4.0 0 15  51 4.0 0  1
;
```

The following statements invoke PROC LOGISTIC to fit the main-effects model by using the alternative form of the input data set:

```
proc logistic data=ingots;
   model NotReady(event='1') = Heat Soak;
   freq Freq;
run;
```

Results of this analysis are the same as the preceding single-trial main-effects analysis. The displayed output for the two runs are identical except for the background information of the model fit and the "Response Profile" table shown in Figure 54.10.

Figure 54.10 Response Profile with Single-Trial Syntax

The LOGISTIC Procedure

Response Profile

Ordered Value	NotReady	Total Frequency
1	0	375
2	1	12

Probability modeled is NotReady=1.

By default, Ordered Values are assigned to the sorted response values in ascending order, and PROC LOGISTIC models the probability of the response level that corresponds to the Ordered Value 1. There are several methods to change these defaults; the preceding statements specify the response variable option EVENT= to model the probability of NotReady=1 as displayed in Figure 54.10. See the section "Response Level Ordering" on page 4237 for more details.

Syntax: LOGISTIC Procedure

The following statements are available in the LOGISTIC procedure:

PROC LOGISTIC < *options* > **;**
BY *variables* **;**
CLASS *variable* < **(***options***)** > < *variable* < **(***options***)** > . . . > < / *options* > **;**
CODE < *options* > **;**
CONTRAST *'label' effect values* < **,** *effect values***,** . . . > < / *options* > **;**
EFFECT *name***=***effect-type***(***variables* < / *options* >**)** **;**
EFFECTPLOT < *plot-type* < **(***plot-definition-options***)** > > < / *options* > **;**
ESTIMATE < *'label'* > *estimate-specification* < / *options* > **;**
EXACT < *'label'* > < **INTERCEPT** > < *effects* > < / *options* > **;**
EXACTOPTIONS *options* **;**
FREQ *variable* **;**
ID *variables* **;**
LSMEANS < *model-effects* > < / *options* > **;**
LSMESTIMATE *model-effect lsmestimate-specification* < / *options* > **;**
< *label***:** > **MODEL** *variable* < **(***variable_options***)** > **=** < *effects* > < / *options* > **;**
< *label***:** > **MODEL** *events/trials* **=** < *effects* > < / *options* > **;**
NLOPTIONS *options* **;**
ODDSRATIO < *'label'* > *variable* < / *options* > **;**
OUTPUT < **OUT=***SAS-data-set* > < *keyword***=***name* < *keyword***=***name* . . . > > < / *option* > **;**
ROC < *'label'* > < *specification* > < / *options* > **;**
ROCCONTRAST < *'label'* > < *contrast* > < / *options* > **;**
SCORE < *options* > **;**
SLICE *model-effect* < / *options* > **;**
STORE < **OUT=** >*item-store-name* < / **LABEL=***'label'* > **;**
STRATA *effects* < / *options* > **;**
< *label***:** > **TEST** *equation1* < **,***equation2***,** . . . > < / *option* > **;**
UNITS < *independent1***=***list1* < *independent2***=***list2* . . . > > < / *option* > **;**
WEIGHT *variable* < / *option* > **;**

The PROC LOGISTIC and MODEL statements are required. The CLASS and EFFECT statements (if specified) must precede the MODEL statement, and the CONTRAST, EXACT, and ROC statements (if specified) must follow the MODEL statement.

The PROC LOGISTIC, MODEL, and ROCCONTRAST statements can be specified at most once. If a FREQ or WEIGHT statement is specified more than once, the variable specified in the first instance is used. If a BY, OUTPUT, or UNITS statement is specified more than once, the last instance is used.

The rest of this section provides detailed syntax information for each of the preceding statements, beginning with the PROC LOGISTIC statement. The remaining statements are covered in alphabetical order. The CODE, EFFECT, EFFECTPLOT, ESTIMATE, LSMEANS, LSMESTIMATE, SLICE, and STORE statements are also available in many other procedures. Summary descriptions of functionality and syntax for these statements are provided, but you can find full documentation on them in the corresponding sections of Chapter 19, "Shared Concepts and Topics."

PROC LOGISTIC Statement

PROC LOGISTIC < *options* > **;**

The PROC LOGISTIC statement invokes the LOGISTIC procedure. Optionally, it identifies input and output data sets, suppresses the display of results, and controls the ordering of the response levels. Table 54.1 summarizes the options available in the PROC LOGISTIC statement.

Table 54.1 PROC LOGISTIC Statement Options

Option	Description
Input/Output Data Set Options	
COVOUT	Displays the estimated covariance matrix in the OUTEST= data set
DATA=	Names the input SAS data set
INEST=	Specifies the initial estimates SAS data set
INMODEL=	Specifies the model information SAS data set
NOCOV	Does not save covariance matrix in the OUTMODEL= data set
OUTDESIGN=	Specifies the design matrix output SAS data set
OUTDESIGNONLY	Outputs the design matrix only
OUTEST=	Specifies the parameter estimates output SAS data set
OUTMODEL=	Specifies the model output data set for scoring
Response and CLASS Variable Options	
DESCENDING	Reverses the sort order of the response variable
NAMELEN=	Specifies the maximum length of effect names
ORDER=	Specifies the sort order of the response variable
TRUNCATE	Truncates class level names
Displayed Output Options	
ALPHA=	Specifies the significance level for confidence intervals
NOPRINT	Suppresses all displayed output
PLOTS	Specifies options for plots
SIMPLE	Displays descriptive statistics
Large Data Set Option	
MULTIPASS	Does not copy the input SAS data set for internal computations
Control of Other Statement Options	
EXACTONLY	Performs exact analysis only
EXACTOPTIONS	Specifies global options for EXACT statements
ROCOPTIONS	Specifies global options for ROC statements

ALPHA=*number*

specifies the level of significance α for $100(1-\alpha)\%$ confidence intervals. The value *number* must be between 0 and 1; the default value is 0.05, which results in 95% intervals. This value is used as the default confidence level for limits computed by the following options:

Statement	Options
CONTRAST	ESTIMATE=
EXACT	ESTIMATE=
MODEL	CLODDS= CLPARM=
ODDSRATIO	CL=
OUTPUT	LOWER= UPPER=
PROC LOGISTIC	PLOTS=EFFECT(CLBAR CLBAND)
ROCCONTRAST	ESTIMATE=
SCORE	CLM

You can override the default in most of these cases by specifying the ALPHA= option in the separate statements.

COVOUT

adds the estimated covariance matrix to the OUTEST= data set. For the COVOUT option to have an effect, the OUTEST= option must be specified. See the section "OUTEST= Output Data Set" on page 4279 for more information.

DATA=*SAS-data-set*

names the SAS data set containing the data to be analyzed. If you omit the DATA= option, the procedure uses the most recently created SAS data set. The INMODEL= option cannot be specified with this option.

DESCENDING
DESC

reverses the sort order for the levels of the response variable. If both the DESCENDING and ORDER= options are specified, PROC LOGISTIC orders the levels according to the ORDER= option and then reverses that order. This option has the same effect as the response variable option DESCENDING in the MODEL statement. See the section "Response Level Ordering" on page 4237 for more detail.

EXACTONLY

requests only the exact analyses. The asymptotic analysis that PROC LOGISTIC usually performs is suppressed.

EXACTOPTIONS (*options***)**

specifies options that apply to every EXACT statement in the program. The available options are summarized here, and full descriptions are available in the EXACTOPTIONS statement.

Option	Description
ADDTOBS	Adds the observed sufficient statistic to the sampled exact distribution
BUILDSUBSETS	Builds every distribution for sampling
EPSILON=	Specifies the comparison fuzz for partial sums of sufficient statistics
MAXTIME=	Specifies the maximum time allowed in seconds
METHOD=	Specifies the DIRECT, NETWORK, or NETWORKMC algorithm
N=	Specifies the number of Monte Carlo samples
ONDISK	Uses disk space
SEED=	Specifies the initial seed for sampling
STATUSN=	Specifies the sampling interval for printing a status line
STATUSTIME=	Specifies the time interval for printing a status line

INEST=*SAS-data-set*
names the SAS data set that contains initial estimates for all the parameters in the model. If BY-group processing is used, it must be accommodated in setting up the INEST= data set. See the section "INEST= Input Data Set" on page 4280 for more information.

INMODEL=*SAS-data-set*
specifies the name of the SAS data set that contains the model information needed for scoring new data. This INMODEL= data set is the OUTMODEL= data set saved in a previous PROC LOGISTIC call. The OUTMODEL= data set should not be modified before its use as an INMODEL= data set.

The DATA= option cannot be specified with this option; instead, specify the data sets to be scored in the SCORE statements. FORMAT statements are not allowed when the INMODEL= data set is specified; variables in the DATA= and PRIOR= data sets in the SCORE statement should be formatted within the data sets.

You can specify the BY statement provided that the INMODEL= data set is created under the same BY-group processing.

The CLASS, EFFECT, EFFECTPLOT, ESTIMATE, EXACT, LSMEANS, LSMESTIMATE, MODEL, OUTPUT, ROC, ROCCONTRAST, SLICE, STORE, TEST, and UNIT statements are not available with the INMODEL= option.

MULTIPASS
forces the procedure to reread the DATA= data set as needed rather than require its storage in memory or in a temporary file on disk. By default, the data set is cleaned up and stored in memory or in a temporary file. This option can be useful for large data sets. All exact analyses are ignored in the presence of the MULTIPASS option. If a STRATA statement is specified, then the data set must first be grouped or sorted by the strata variables.

NAMELEN=*n*
specifies the maximum 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.

NOCOV
specifies that the covariance matrix not be saved in the OUTMODEL= data set. The covariance matrix is needed for computing the confidence intervals for the posterior probabilities in the OUT= data set in the SCORE statement. Specifying this option will reduce the size of the OUTMODEL= data set.

NOPRINT
suppresses all displayed output. Note that this option temporarily disables the Output Delivery System (ODS); see Chapter 20, "Using the Output Delivery System," for more information.

ORDER=DATA | FORMATTED | FREQ | INTERNAL
RORDER=DATA | FORMATTED | INTERNAL
specifies the sort order for the levels of the response variable. See the response variable option ORDER= in the MODEL statement for more information. For ordering of CLASS variable levels, see the ORDER= option in the CLASS statement.

OUTDESIGN=*SAS-data-set*
specifies the name of the data set that contains the design matrix for the model. The data set contains the same number of observations as the corresponding DATA= data set and includes the response

variable (with the same format as in the DATA= data set), the FREQ variable, the WEIGHT variable, the OFFSET= variable, and the design variables for the covariates, including the Intercept variable of constant value 1 unless the NOINT option in the MODEL statement is specified.

OUTDESIGNONLY
suppresses the model fitting and creates only the OUTDESIGN= data set. This option is ignored if the OUTDESIGN= option is not specified.

OUTEST=*SAS-data-set*
creates an output SAS data set that contains the final parameter estimates and, optionally, their estimated covariances (see the preceding COVOUT option). The output data set also includes a variable named _LNLIKE_, which contains the log likelihood. See the section "OUTEST= Output Data Set" on page 4279 for more information.

OUTMODEL=*SAS-data-set*
specifies the name of the SAS data set that contains the information about the fitted model. This data set contains sufficient information to score new data without having to refit the model. It is solely used as the input to the INMODEL= option in a subsequent PROC LOGISTIC call. The OUTMODEL= option is not available with the STRATA statement. Information in this data set is stored in a very compact form, so you should not modify it manually.

NOTE: The STORE statement can also be used to save your model. See the section "STORE Statement" on page 4233 for more information.

PLOTS <(*global-plot-options***)>< =***plot-request*< *(options)* >>

PLOTS <(*global-plot-options***)> =(***plot-request*< *(options)* >< *. . . plot-request*< *(options)* >>**)**
controls the plots produced through ODS Graphics. When you specify only one *plot-request*, you can omit the parentheses from around the *plot-request*. For example:

```
PLOTS = ALL
PLOTS = (ROC EFFECT INFLUENCE(UNPACK))
PLOTS(ONLY) = EFFECT(CLBAR SHOWOBS)
```

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

```
ods graphics on;
proc logistic plots=all;
   model y=x;
run;
ods graphics off;
```

For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" on page 600 in Chapter 21, "Statistical Graphics Using ODS."

If the PLOTS option is not specified or is specified with no *plot-requests*, then graphics are produced by default in the following situations:

- If the INFLUENCE or IPLOTS option is specified in the MODEL statement, then the line-printer plots are suppressed, and the INFLUENCE plots are produced unless the MAXPOINTS= cutoff is exceeded.

- If you specify the OUTROC= option in the MODEL statement, then ROC curves are produced. If you also specify a SELECTION= method, then an overlaid plot of all the ROC curves for each step of the selection process is displayed.
- If the OUTROC= option is specified in a SCORE statement, then the ROC curve for the scored data set is displayed.
- If you specify ROC statements, then an overlaid plot of the ROC curves for the model (or the selected model if a SELECTION= method is specified) and for all the ROC statement models is displayed.
- If an odds ratio table is produced, then a plot of the odds ratios and their confidence limits is displayed. These plots correspond to the default odds ratio table and the tables produced by the CLODDS= option in the MODEL statement and the tables produced by the ODDSRATIO statement.

For general information about ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS."

The following *global-plot-options* are available:

LABEL
displays a label on diagnostic plots to aid in identifying the outlying observations. This option enhances the plots produced by the DFBETAS, DPC, INFLUENCE, LEVERAGE, and PHAT options. If an ID statement is specified, then the plots are labeled with the ID variables. Otherwise, the observation number is displayed.

MAXPOINTS=NONE | *number*
suppresses the plots produced by the DFBETAS, DPC, INFLUENCE, LEVERAGE, and PHAT options if there are more than *number* observations. Also, observations are not displayed on the EFFECT plots when the cutoff is exceeded. The default is MAXPOINTS=5000. The cutoff is ignored if you specify MAXPOINTS=NONE.

ONLY
specifically requested *plot-requests* are displayed.

UNPACKPANELS | **UNPACK**
suppresses paneling. By default, multiple plots can appear in some output *panels*. Specify UNPACKPANEL to display each plot separately.

The following *plot-requests* are available:

ALL
produces all appropriate plots. You can specify other options with ALL. For example, to display all plots and unpack the DFBETAS plots you can specify `plots=(all dfbetas(unpack))`.

DFBETAS <(UNPACK)>
displays plots of DFBETAS versus the case (observation) number. This displays the statistics generated by the DFBETAS=_ALL_ option in the OUTPUT statement. The UNPACK option displays the plots separately. See Output 54.6.5 for an example of this plot.

DPC< (UNPACK) >
displays plots of DIFCHISQ and DIFDEV versus the predicted event probability, and colors the markers according to the value of the confidence interval displacement C. The UNPACK option displays the plots separately. See Output 54.6.8 for an example of this plot.

EFFECT< (*effect-options*) >
displays and enhances the effect plots for the model. For more information about effect plots and the available *effect-options*, see the section "PLOTS=EFFECT Plots" on page 4182.

NOTE: The EFFECTPLOT statement provides you with much of the same functionality and more options for creating effect plots. See Outputs 54.2.11, 54.3.5, 54.4.8, 54.7.4, and 54.15.4 for examples of effect plots.

INFLUENCE< (UNPACK | STDRES) >
displays index plots of RESCHI, RESDEV, leverage, confidence interval displacements C and CBar, DIFCHISQ, and DIFDEV. These plots are produced by default when any *plot-request* is specified and the MAXPOINTS= cutoff is not exceeded. The UNPACK option displays the plots separately. The STDRES option also displays index plots of STDRESCHI, STDRESDEV, and RESLIK. See Outputs 54.6.3 and 54.6.4 for examples of these plots.

LEVERAGE< (UNPACK) >
displays plots of DIFCHISQ, DIFDEV, confidence interval displacement C, and the predicted probability versus the leverage. The UNPACK option displays the plots separately. See Output 54.6.7 for an example of this plot.

NONE
suppresses all plots.

ODDSRATIO < (*oddsratio-options*) >
displays and enhances the odds ratio plots for the model. For more information about odds ratio plots and the available *oddsratio-options*, see the section "Odds Ratio Plots" on page 4185. See Outputs 54.7,54.2.9, 54.3.3, and 54.4.5 for examples of this plot.

PHAT< (UNPACK) >
displays plots of DIFCHISQ, DIFDEV, confidence interval displacement C, and leverage versus the predicted event probability. The UNPACK option displays the plots separately. See Output 54.6.6 for an example of this plot.

ROC< (ID< =*keyword* >) >
displays the ROC curve. If you also specify a SELECTION= method, then an overlaid plot of all the ROC curves for each step of the selection process is displayed. If you specify ROC statements, then an overlaid plot of the model (or the selected model if a SELECTION= method is specified) and the ROC statement models is displayed. If the OUTROC= option is specified in a SCORE statement, then the ROC curve for the scored data set is displayed.

The ID= option labels certain points on the ROC curve. Typically, the labeled points are closest to the upper-left corner of the plot, and points directly below or to the right of a labeled point are suppressed. This option is identical to, and has the same *keywords* as, the ID= suboption of the ROCOPTIONS option.

See Output 54.7.3 and Example 54.8 for examples of these ROC plots.

ROCOPTIONS (*options*)

specifies options that apply to every model specified in a ROC statement. Some of these options also apply to the SCORE statement. The following *options* are available:

ALPHA=*number*

sets the significance level for creating confidence limits of the areas and the pairwise differences. The ALPHA= value specified in the PROC LOGISTIC statement is the default. If neither ALPHA= value is specified, then ALPHA=0.05 by default.

EPS=*value*

is an alias for the ROCEPS= option in the MODEL statement. This value is used to determine which predicted probabilities are equal. The default value is the square root of the machine epsilon, which is about 1E–8.

ID< =*keyword* >

displays labels on certain points on the individual ROC curves and also on the SCORE statement's ROC curve. This option overrides the ID= suboption of the PLOTS=ROC option. If several observations lie at the same place on the ROC curve, the value for the last observation is displayed. If you specify the ID option with no *keyword*, any variables that are listed in the ID statement are used. If no ID statement is specified, the observation number is displayed. The following *keywords* are available:

PROB	displays the model predicted probability.
OBS	displays the (last) observation number.
SENSIT	displays the true positive fraction (sensitivity).
1MSPEC	displays the false positive fraction (1–specificity).
FALPOS	displays the fraction of nonevents that are predicted as events.
FALNEG	displays the fraction of events that are predicted as nonevents.
POSPRED	displays the positive predictive value (1–FALPOS).
NEGPRED	displays the negative predictive value (1–FALNEG).
MISCLASS	displays the misclassification rate.
ID	displays the ID variables.

The SENSIT, 1MSPEC, FALPOS, and FALNEG statistics are defined in the section "Receiver Operating Characteristic Curves" on page 4260. The misclassification rate is the number of events that are predicted as nonevents and the number of nonevents that are predicted as events as calculated by using the given cutpoint (predicted probability) divided by the number of observations. If the PEVENT= option is also specified, then FALPOS and FALNEG are computed using the first PEVENT= value and Bayes' theorem, as discussed in the section "Predicted Probability of an Event for Classification" on page 4255.

NODETAILS

suppresses the display of the model fitting information for the models specified in the ROC statements.

OUT=*SAS-data-set-name*
is an alias for the OUTROC= option in the MODEL statement.

WEIGHTED
uses frequency×weight in the ROC computations (Izrael et al. 2002) instead of just frequency. Typically, weights are considered in the fit of the model only, and hence are accounted for in the parameter estimates. The "Association of Predicted Probabilities and Observed Responses" table uses frequency (unless the BINWIDTH=0 option is also specified on the MODEL statement), and is suppressed when ROC comparisons are performed. This option also affects SCORE statement ROC and AUC computations.

SIMPLE
displays simple descriptive statistics (mean, standard deviation, minimum and maximum) for each continuous explanatory variable. For each CLASS variable involved in the modeling, the frequency counts of the classification levels are displayed. The SIMPLE option generates a breakdown of the simple descriptive statistics or frequency counts for the entire data set and also for individual response categories.

TRUNCATE
determines class levels by using no more than the first 16 characters of the formatted values of CLASS, response, and strata variables. When formatted values are longer than 16 characters, you can use this option to revert to the levels as determined in releases previous to SAS 9.0. This option invokes the same option in the CLASS statement.

PLOTS=EFFECT Plots

Only one PLOTS=EFFECT plot is produced by default; you must specify other *effect-options* to produce multiple plots. For binary response models, the following plots are produced when an EFFECT option is specified with no *effect-options*:

- If you only have continuous covariates in the model, then a plot of the predicted probability versus the first continuous covariate fixing all other continuous covariates at their means is displayed. See Output 54.7.4 for an example with one continuous covariate.
- If you only have classification covariates in the model, then a plot of the predicted probability versus the first CLASS covariate at each level of the second CLASS covariate, if any, holding all other CLASS covariates at their reference levels is displayed.
- If you have CLASS and continuous covariates, then a plot of the predicted probability versus the first continuous covariate at up to 10 cross-classifications of the CLASS covariate levels, while fixing all other continuous covariates at their means and all other CLASS covariates at their reference levels, is displayed. For example, if your model has four binary covariates, there are 16 cross-classifications of the CLASS covariate levels. The plot displays the 8 cross-classifications of the levels of the first three covariates while the fourth covariate is fixed at its reference level.

For polytomous response models, similar plots are produced by default, except that the response levels are used in place of the CLASS covariate levels. Plots for polytomous response models involving OFFSET= variables with multiple values are not available.

The following *effect-options* specify the type of graphic to produce:

AT(*variable=value-list* | **ALL<** *...variable=value-list* | **ALL >)**

specifies fixed values for a covariate. For continuous covariates, you can specify one or more numbers in the *value-list*. For classification covariates, you can specify one or more formatted levels of the covariate enclosed in single quotes (for example, **A='cat' 'dog'**), or you can specify the keyword ALL to select all levels of the classification variable. You can specify a variable at most once in the AT option. By default, continuous covariates are set to their means when they are not used on an axis, while classification covariates are set to their reference level when they are not used as an X=, SLICEBY=, or PLOTBY= effect. For example, for a model that includes a classification variable A={cat,dog} and a continuous covariate X, specifying **AT(A='cat' X=7 9)** will set A to 'cat' when A does not appear in the plot. When X does not define an axis it first produces plots setting $X = 7$ and then produces plots setting $X = 9$. Note in this example that specifying **AT(A=ALL)** is the same as specifying the PLOTBY=A option.

FITOBSONLY

computes the predicted values only at the observed data. If the FITOBSONLY option is omitted and the X-axis variable is continuous, the predicted values are computed at a grid of points extending slightly beyond the range of the data (see the EXTEND= option for more information). If the FITOBSONLY option is omitted and the X-axis effect is categorical, the predicted values are computed at all possible categories.

INDIVIDUAL

displays the individual probabilities instead of the cumulative probabilities. This option is available only with cumulative models, and it is not available with the LINK option.

LINK

displays the linear predictors instead of the probabilities on the Y axis. For example, for a binary logistic regression, the Y axis will be displayed on the logit scale. The INDIVIDUAL and POLYBAR options are not available with the LINK option.

PLOTBY=*effect*

displays an effect plot at each unique level of the PLOTBY= effect. You can specify *effect* as one CLASS variable or as an interaction of classification covariates. For polytomous-response models, you can also specify the response variable as the lone PLOTBY= effect. For nonsingular parameterizations, the complete cross-classification of the CLASS variables specified in the effect define the different PLOTBY= levels. When the GLM parameterization is used, the PLOTBY= levels can depend on the model and the data.

SLICEBY=*effect*

displays predicted probabilities at each unique level of the SLICEBY= effect. You can specify *effect* as one CLASS variable or as an interaction of classification covariates. For polytomous-response models, you can also specify the response variable as the lone SLICEBY= effect. For nonsingular parameterizations, the complete cross-classification of the CLASS variables specified in the effect define the different SLICEBY= levels. When the GLM parameterization is used, the SLICEBY= levels can depend on the model and the data.

X=*effect*

X=(*effect...effect***)**

specifies effects to be used on the X axis of the effect plots. You can specify several different X axes: continuous variables must be specified as main effects, while CLASS variables can be crossed. For nonsingular parameterizations, the complete cross-classification of the CLASS variables specified in

the effect define the axes. When the GLM parameterization is used, the X= levels can depend on the model and the data. The response variable is not allowed as an *effect*.

NOTE: Any variable not specified in a SLICEBY= or PLOTBY= option is available to be displayed on the X axis. A variable can be specified in at most one of the SLICEBY=, PLOTBY=, and X= options.

The following *effect-options* enhance the graphical output:

ALPHA=*number*
specifies the size of the confidence limits. The ALPHA= value specified in the PROC LOGISTIC statement is the default. If neither ALPHA= value is specified, then ALPHA=0.05 by default.

CLBAND< =YES | NO >
displays confidence limits on the plots. This option is not available with the INDIVIDUAL option. If you have CLASS covariates on the X axis, then error bars are displayed (see the CLBAR option) unless you also specify the CONNECT option.

CLBAR
displays the error bars on the plots when you have CLASS covariates on the X axis; if the X axis is continuous, then this invokes the CLBAND option. For polytomous-response models with CLASS covariates only and with the POLYBAR option specified, the stacked bar charts are replaced by side-by-side bar charts with error bars.

CLUSTER< =*percent* **>**
displays the levels of the SLICEBY= effect in a side-by-side fashion instead of stacking them. This option is available when you have CLASS covariates on the X axis. You can specify *percent* as a percentage of half the distance between X levels. The *percent* value must be between 0.1 and 1; the default *percent* depends on the number of X levels and the number of SLICEBY= levels. Default clustering can be removed by specifying the NOCLUSTER option.

CONNECT< =YES | NO >
JOIN< =YES | NO >
connects the predicted values with a line. This option is available when you have CLASS covariates on the X axis. Default connecting lines can be suppressed by specifying the NOCONNECT option.

EXTEND=*value*
extends continuous X axes by a factor of *value*/2 in each direction. By default, EXTEND=0.2.

MAXATLEN=*length*
specifies the maximum number of characters used to display the levels of all the fixed variables. If the text is too long, it is truncated and ellipses ("...") are appended. By default, *length* is equal to its maximum allowed value, 256.

NOCLUSTER
prevents clustering of the levels of the SLICEBY= effect. This option is available when you have CLASS covariates on the X axis.

NOCONNECT
removes the line that connects the predicted values. This option is available when you have CLASS covariates on the X axis.

POLYBAR
replaces scatter plots of polytomous response models with bar charts. This option has no effect on binary-response models, and it is overridden by the CONNECT option. By default, the X axis is chosen to be a crossing of available classification variables so that there are no more than 16 levels; if no such crossing is possible then the first available classification variable is used. You can override this default by specifying the X= option.

SHOWOBS<=YES | NO>
displays observations on the plot when the MAXPOINTS= cutoff is not exceeded. For events/trials notation, the observed proportions are displayed; for single-trial binary-response models, the observed events are displayed at $\hat{p} = 1$ and the observed nonevents are displayed at $\hat{p} = 0$. For polytomous response models the predicted probabilities at the observed values of the covariate are computed and displayed.

YRANGE=(<*min*><,*max*>)
displays the Y axis as [*min*,*max*]. Note that the axis might extend beyond your specified values. By default, the entire Y axis, [0,1], is displayed for the predicted probabilities. This option is useful if your predicted probabilities are all contained in some subset of this range.

Odds Ratio Plots

The odds ratios and confidence limits from the default Odds Ratio Estimates table and from the tables produced by the CLODDS= option or the ODDSRATIO statement can be displayed in a graphic. If you have many odds ratios, you can produce multiple graphics, or *panels*, by displaying subsets of the odds ratios. Odds ratios with duplicate labels are not displayed. See Outputs 54.2.9 and 54.3.3 for examples of odds ratio plots.

The following *oddsratio-options* modify the default odds ratio plot:

CLDISPLAY=SERIF | LINE | BAR<*width*>
controls the look of the confidence limit error bars. The default CLDISPLAY=SERIF displays the confidence limits as lines with serifs, CLDISPLAY=LINE removes the serifs from the error bars, and CLDISPLAY=BAR <*width*> displays the limits with a bar of width equal to the size of the marker. You can control the width of the bars and the size of the marker by specifying the *width* value as a percentage of the distance between the bars, $0 < \textit{width} \leq 1$.

NOTE: Your bar might disappear with small values of *width*.

DOTPLOT
displays dotted gridlines on the plot.

GROUP
displays the odds ratios in panels defined by the ODDSRATIO statements. The NPANELPOS= option is ignored when this option is specified.

LOGBASE=2 | E | 10
displays the odds ratio axis on the specified log scale.

NPANELPOS=*n*

breaks the plot into multiple graphics having at most |*n*| odds ratios per graphic. If *n* is positive, then the number of odds ratios per graphic is balanced; but if *n* is negative, then no balancing of the number of odds ratios takes place. By default, *n* = 0 and all odds ratios are displayed in a single plot. For example, suppose you want to display 21 odds ratios. Then specifying **NPANELPOS=20** displays two plots, the first with 11 odds ratios and the second with 10; but specifying **NPANELPOS=−20** displays 20 odds ratios in the first plot and only 1 odds ratio in the second.

ORDER=ASCENDING | DESCENDING

displays the odds ratios in sorted order. By default the odds ratios are displayed in the order in which they appear in the corresponding table.

RANGE=(< *min* **><** *,max* **>) | CLIP**

specifies the range of the displayed odds ratio axis. The RANGE=CLIP option has the same effect as specifying the minimum odds ratio as *min* and the maximum odds ratio as *max*. By default, all odds ratio confidence intervals are displayed.

TYPE=HORIZONTAL | HORIZONTALSTAT | VERTICAL | VERTICALBLOCK

controls the look of the graphic. The default TYPE=HORIZONTAL option places the odds ratio values on the X axis, while the TYPE=HORIZONTALSTAT option also displays the values of the odds ratios and their confidence limits on the right side of the graphic. The TYPE=VERTICAL option places the odds ratio values on the Y axis, while the TYPE=VERTICALBLOCK option (available only with the CLODDS= option) places the odds ratio values on the Y axis and puts boxes around the labels.

BY Statement

BY *variables* **;**

You can specify a BY statement with PROC LOGISTIC 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 LOGISTIC 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).

If a SCORE statement is specified, then define the *training data set* to be the DATA= data set or the INMODEL= data set in the PROC LOGISTIC statement, and define the *scoring data set* to be the DATA= data set and PRIOR= data set in the SCORE statement. The training data set contains all of the BY variables, and

the scoring data set must contain either all of them or none of them. If the scoring data set contains all the BY variables, matching is carried out between the training and scoring data sets. If the scoring data set does not contain any of the BY variables, the entire scoring data set is used for every BY group in the training data set and the BY variables are added to the output data sets that are specified in the SCORE statement.

CAUTION: The order of the levels in the response and classification variables is determined from all the data regardless of BY groups. However, different sets of levels might appear in different BY groups. This might affect the value of the reference level for these variables, and hence your interpretation of the model and the parameters.

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*.

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. Response variables do not need to be specified in the CLASS statement.

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 LOGISTIC 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 LOGISTIC interprets values of the ORDER= option.

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

specifies the parameterization method for the classification variable or variables. You can specify any of the *keywords* shown in the following table; the default is PARAM=EFFECT. Design matrix columns are created from CLASS variables according to the corresponding coding schemes:

Value of PARAM=	Coding
EFFECT	Effect coding
GLM	Less-than-full-rank reference cell coding (this *keyword* can be used only in a global option)
ORDINAL THERMOMETER	Cumulative parameterization for an ordinal CLASS variable
POLYNOMIAL POLY	Polynomial coding
REFERENCE REF	Reference cell coding
ORTHEFFECT	Orthogonalizes PARAM=EFFECT coding
ORTHORDINAL ORTHOTHERM	Orthogonalizes PARAM=ORDINAL coding
ORTHPOLY	Orthogonalizes PARAM=POLYNOMIAL coding
ORTHREF	Orthogonalizes PARAM=REFERENCE coding

All parameterizations are full rank, except for the GLM parameterization. The REF= option in the

CLASS statement determines the reference level for EFFECT and REFERENCE coding and for their orthogonal parameterizations. It also indirectly determines the reference level for a singular GLM parameterization through the order of levels.

If PARAM=ORTHPOLY or PARAM=POLY and the classification variable is numeric, then the ORDER= option in the CLASS statement is ignored, and the internal unformatted values are used. See the section "Other Parameterizations" on page 387 of Chapter 19, "Shared Concepts and Topics," for further details.

REF=*'level' | keyword*

specifies the reference level for PARAM=EFFECT, PARAM=REFERENCE, and their orthogonalizations. 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*. The default is REF=LAST.

FIRST designates the first ordered level as reference.

LAST designates the last ordered level as reference.

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.

Class Variable Naming Convention

Parameter names for a CLASS predictor variable are constructed by concatenating the CLASS variable name with the CLASS levels. However, for the POLYNOMIAL and orthogonal parameterizations, parameter names are formed by concatenating the CLASS variable name and keywords that reflect the parameterization. See the section "Other Parameterizations" on page 387 in Chapter 19, "Shared Concepts and Topics," for examples and further details.

Class Variable Parameterization with Unbalanced Designs

PROC LOGISTIC initially parameterizes the CLASS variables by looking at the levels of the variables across the complete data set. If you have an *unbalanced* replication of levels across variables or BY groups, then the design matrix and the parameter interpretation might be different from what you expect. For instance, suppose you have a model with one CLASS variable A with three levels (1, 2, and 3), and another CLASS variable B with two levels (1 and 2). If the third level of A occurs only with the first level of B, if you use the EFFECT parameterization, and if your model contains the effect A(B) and an intercept, then the design for A within the second level of B is not a differential effect. In particular, the design looks like the following:

		Design Matrix			
		A(B=1)		A(B=2)	
B	A	A1	A2	A1	A2
1	1	1	0	0	0
1	2	0	1	0	0
1	3	–1	–1	0	0
2	1	0	0	1	0
2	2	0	0	0	1

PROC LOGISTIC detects linear dependency among the last two design variables and sets the parameter for A2(B=2) to zero, resulting in an interpretation of these parameters as if they were reference- or dummy-coded. The REFERENCE or GLM parameterization might be more appropriate for such problems.

CODE Statement

CODE < *options* > **;**

The CODE statement enables you to write SAS DATA step code for computing predicted values of the fitted model either to a file or to a catalog entry. This code can then be included in a DATA step to score new data.

Table 54.2 summarizes the *options* available in the CODE statement.

Table 54.2 CODE Statement Options

Option	Description
CATALOG=	Names the catalog entry where the generated code is saved
DUMMIES	Retains the dummy variables in the data set
ERROR	Computes the error function
FILE=	Names the file where the generated code is saved
FORMAT=	Specifies the numeric format for the regression coefficients
GROUP=	Specifies the group identifier for array names and statement labels
IMPUTE	Imputes predicted values for observations with missing or invalid covariates
LINESIZE=	Specifies the line size of the generated code
LOOKUP=	Specifies the algorithm for looking up CLASS levels
RESIDUAL	Computes residuals

For details about the syntax of the CODE statement, see the section "CODE Statement" on page 390 in Chapter 19, "Shared Concepts and Topics."

CONTRAST Statement

CONTRAST *'label' row-description*< **,** . . . **,** *row-description* > < / *options* > **;**

where a *row-description* is defined as follows:

effect values< , . . . , effect values >

The CONTRAST statement provides a mechanism for obtaining customized hypothesis tests. It is similar to the CONTRAST and ESTIMATE statements in other modeling procedures.

The CONTRAST statement enables you to specify a matrix, $\mathbf{L}$, for testing the hypothesis $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta}$ is the vector of intercept and slope parameters. You must be familiar with the details of the model parameterization that PROC LOGISTIC uses (for more information, see the PARAM= option in the section "CLASS Statement" on page 4187). Optionally, the CONTRAST statement enables you to estimate each row, $\boldsymbol{l}_i'\boldsymbol{\beta}$, of $\mathbf{L}\boldsymbol{\beta}$ and test the hypothesis $\boldsymbol{l}_i'\boldsymbol{\beta} = 0$. Computed statistics are based on the asymptotic chi-square distribution of the Wald statistic.

There is no limit to the number of CONTRAST statements that you can specify, but they must appear after the MODEL statement.

The following parameters are specified in the CONTRAST statement:

label identifies the contrast in the displayed output. A label is required for every contrast specified, and it must be enclosed in quotes.

effect identifies an effect that appears in the MODEL statement. The name INTERCEPT can be used as an effect when one or more intercepts are included in the model. You do not need to include all effects that are included in the MODEL statement.

values are constants that are elements of the $\mathbf{L}$ matrix associated with the effect. To correctly specify your contrast, it is crucial to know the ordering of parameters within each effect and the variable levels associated with any parameter. The "Class Level Information" table shows the ordering of levels within variables. The E option, described later in this section, enables you to verify the proper correspondence of *values* to parameters. If too many values are specified for an effect, the extra ones are ignored. If too few values are specified, the remaining ones are set to 0.

Multiple degree-of-freedom hypotheses can be tested by specifying multiple *row-descriptions*; the rows of $\mathbf{L}$ are specified in order and are separated by commas. The degrees of freedom is the number of linearly independent constraints implied by the CONTRAST statement—that is, the rank of $\mathbf{L}$.

More details for specifying contrasts involving effects with full-rank parameterizations are given in the section "Full-Rank Parameterized Effects" on page 4192, while details for less-than-full-rank parameterized effects are given in the section "Less-Than-Full-Rank Parameterized Effects" on page 4193.

You can specify the following options after a slash (/):

ALPHA=*number*
specifies the level of significance α for the $100(1-\alpha)\%$ confidence interval for each contrast when the ESTIMATE option is specified. The value of *number* must be between 0 and 1. By default, *number* is equal to the value of the ALPHA= option in the PROC LOGISTIC statement, or 0.05 if that option is not specified.

E
displays the $\mathbf{L}$ matrix.

ESTIMATE=*keyword*

estimates and tests each individual contrast (that is, each row, $\boldsymbol{l}_i'\boldsymbol{\beta}$, of $\mathbf{L}\boldsymbol{\beta}$), exponentiated contrast ($e^{\boldsymbol{l}_i'\boldsymbol{\beta}}$), or predicted probability for the contrast ($g^{-1}(\boldsymbol{l}_i'\boldsymbol{\beta})$). PROC LOGISTIC displays the point estimate, its standard error, a Wald confidence interval, and a Wald chi-square test. The significance level of the confidence interval is controlled by the ALPHA= option. You can estimate the individual contrast, the exponentiated contrast, or the predicted probability for the contrast by specifying one of the following *keywords*:

PARM

estimates the individual contrast.

EXP

estimates the exponentiated contrast.

BOTH

estimates both the individual contrast and the exponentiated contrast.

PROB

estimates the predicted probability of the contrast.

ALL

estimates the individual contrast, the exponentiated contrast, and the predicted probability of the contrast.

For details about the computations of the standard errors and confidence limits, see the section "Linear Predictor, Predicted Probability, and Confidence Limits" on page 4253.

SINGULAR=*number*

tunes the estimability check. This option is ignored when a full-rank parameterization is specified. If $\mathbf{v}$ is a vector, define ABS($\mathbf{v}$) to be the largest absolute value of the elements of $\mathbf{v}$. For a row vector $\boldsymbol{l}'$ of the contrast matrix $\mathbf{L}$, define $c = \text{ABS}(\boldsymbol{l})$ if ABS($\boldsymbol{l}$) is greater than 0; otherwise, $c = 1$. If ABS($\boldsymbol{l}' - \boldsymbol{l}'\mathbf{T}$) is greater than $c*$*number*, then $\boldsymbol{l}$ is declared nonestimable. The $\mathbf{T}$ matrix is the Hermite form matrix $\mathbf{I}_0^-\mathbf{I}_0$, where $\mathbf{I}_0^-$ represents a generalized inverse of the (observed or expected) information matrix $\mathbf{I}_0$ of the null model. The value for *number* must be between 0 and 1; the default value is 1E–4.

Full-Rank Parameterized Effects

If an effect involving a CLASS variable with a full-rank parameterization does not appear in the CONTRAST statement, then all of its coefficients in the $\mathbf{L}$ matrix are set to 0.

If you use effect coding by default or by specifying PARAM=EFFECT in the CLASS statement, then all parameters are directly estimable and involve no other parameters. For example, suppose an effect-coded CLASS variable A has four levels. Then there are three parameters ($\beta_1, \beta_2, \beta_3$) representing the first three levels, and the fourth parameter is represented by

$$-\beta_1 - \beta_2 - \beta_3$$

To test the first versus the fourth level of A, you would test

$$\beta_1 = -\beta_1 - \beta_2 - \beta_3$$

or, equivalently,

$$2\beta_1 + \beta_2 + \beta_3 = 0$$

which, in the form $\mathbf{L}\boldsymbol{\beta} = 0$, is

$$\begin{bmatrix} 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 0$$

Therefore, you would use the following CONTRAST statement:

```
contrast '1 vs. 4' A 2 1 1;
```

To contrast the third level with the average of the first two levels, you would test

$$\frac{\beta_1 + \beta_2}{2} = \beta_3$$

or, equivalently,

$$\beta_1 + \beta_2 - 2\beta_3 = 0$$

Therefore, you would use the following CONTRAST statement:

```
contrast '1&2 vs. 3' A 1 1 -2;
```

Other CONTRAST statements are constructed similarly. For example:

```
contrast '1 vs. 2     '  A  1 -1  0;
contrast '1&2 vs. 4   '  A  3  3  2;
contrast '1&2 vs. 3&4'   A  2  2  0;
contrast 'Main Effect'   A  1  0  0,
                         A  0  1  0,
                         A  0  0  1;
```

Less-Than-Full-Rank Parameterized Effects

When you use the less-than-full-rank parameterization (by specifying PARAM=GLM in the CLASS statement), each row is checked for estimability; see the section "Estimable Functions" on page 59 in Chapter 3, "Introduction to Statistical Modeling with SAS/STAT Software," for more information. If PROC LOGISTIC finds a contrast to be nonestimable, it displays missing values in corresponding rows in the results. PROC LOGISTIC handles missing level combinations of classification variables in the same manner as PROC GLM: parameters corresponding to missing level combinations are not included in the model. This convention can affect the way in which you specify the **L** matrix in your CONTRAST statement. If the elements of **L** are not specified for an effect that contains a specified effect, then the elements of the specified effect are distributed over the levels of the higher-order effect just as the GLM procedure does for its CONTRAST and ESTIMATE statements. For example, suppose that the model contains effects A and B and their interaction A*B. If you specify a CONTRAST statement involving A alone, the **L** matrix contains nonzero terms for both A and A*B, since A*B contains A. See rule 4 in the section "Construction of Least Squares Means" on page 3363 in Chapter 42, "The GLM Procedure," for more details.

EFFECT Statement

EFFECT *name*=*effect-type* **(***variables* < / *options* >**) ;**

The EFFECT statement enables you to construct special collections of columns for design matrices. These collections are referred to as *constructed effects* to distinguish them from the usual model effects that are formed from continuous or classification variables, as discussed in the section "GLM Parameterization of Classification Variables and Effects" on page 383 in Chapter 19, "Shared Concepts and Topics."

You can specify the following *effect-types*:

COLLECTION is a collection effect that defines one or more variables as a single effect with multiple degrees of freedom. The variables in a collection are considered as a unit for estimation and inference.

LAG is a classification effect in which the level that is used for a given period corresponds to the level in the preceding period.

MULTIMEMBER | MM is a multimember classification effect whose levels are determined by one or more variables that appear in a CLASS statement.

POLYNOMIAL | POLY is a multivariate polynomial effect in the specified numeric variables.

SPLINE is a regression spline effect whose columns are univariate spline expansions of one or more variables. A spline expansion replaces the original variable with an expanded or larger set of new variables.

Table 54.3 summarizes the *options* available in the EFFECT statement.

Table 54.3 EFFECT Statement Options

Option	Description
Collection Effects Options	
DETAILS	Displays the constituents of the collection effect
Lag Effects Options	
DESIGNROLE=	Names a variable that controls to which lag design an observation is assigned
DETAILS	Displays the lag design of the lag effect
NLAG=	Specifies the number of periods in the lag
PERIOD=	Names the variable that defines the period
WITHIN=	Names the variable or variables that define the group within which each period is defined
Multimember Effects Options	
NOEFFECT	Specifies that observations with all missing levels for the multimember variables should have zero values in the corresponding design matrix columns
WEIGHT=	Specifies the weight variable for the contributions of each of the classification effects

Table 54.3 *continued*

Option	Description
Polynomial Effects Options	
DEGREE=	Specifies the degree of the polynomial
MDEGREE=	Specifies the maximum degree of any variable in a term of the polynomial
STANDARDIZE=	Specifies centering and scaling suboptions for the variables that define the polynomial
Spline Effects Options	
BASIS=	Specifies the type of basis (B-spline basis or truncated power function basis) for the spline expansion
DEGREE=	Specifies the degree of the spline transformation
KNOTMETHOD=	Specifies how to construct the knots for spline effects

For more information about the syntax of these *effect-types* and how columns of constructed effects are computed, see the section "EFFECT Statement" on page 393 in Chapter 19, "Shared Concepts and Topics."

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 54.4 describes the available *plot-types* and their *plot-definition-options*.

Table 54.4 *Plot-Types* and *Plot-Definition-Options*

Plot-Type and Description	Plot-Definition-Options
BOX 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 INTERACTION *plot-type*.	PLOTBY= variable or CLASS effect X= CLASS variable or effect
CONTOUR Displays a contour plot of predicted values against two continuous covariates.	PLOTBY= variable or CLASS effect X= continuous variable Y= continuous variable
FIT Displays a curve of predicted values versus a continuous variable.	PLOTBY= variable or CLASS effect X= continuous variable

Table 54.4 *continued*

Plot-Type and Description	Plot-Definition-Options
INTERACTION 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.	PLOTBY= variable or CLASS effect SLICEBY= variable or CLASS effect X= CLASS variable or effect
SLICEFIT Displays a curve of predicted values versus a continuous variable grouped by the levels of a CLASS effect.	PLOTBY= variable or CLASS effect SLICEBY= variable or CLASS effect X= continuous variable

For full details about the syntax and options of the EFFECTPLOT statement, see the section "EFFECTPLOT Statement" on page 411 in Chapter 19, "Shared Concepts and Topics."

See Outputs 54.2.11, 54.2.12, 54.3.5, 54.4.8, 54.7.4, and 54.15.4 for examples of plots produced by this statement.

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 54.5 summarizes the options available in the ESTIMATE statement.

Table 54.5 ESTIMATE Statement Options

Option	Description
Construction and Computation of Estimable Functions	
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

Table 54.5 *continued*

Option	Description
Degrees of Freedom and *p*-values	
ADJUST=	Determines the method for multiple comparison adjustment of estimates
ALPHA=α	Determines the confidence level ($1-\alpha$)
LOWER	Performs one-sided, lower-tailed inference
STEPDOWN	Adjusts multiplicity-corrected *p*-values further in a step-down fashion
TESTVALUE=	Specifies values under the null hypothesis for tests
UPPER	Performs one-sided, upper-tailed inference
Statistical Output	
CL	Constructs confidence limits
CORR	Displays the correlation matrix of estimates
COV	Displays the covariance matrix of estimates
E	Prints the **L** matrix
JOINT	Produces a joint *F* or chi-square test for the estimable functions
SEED=	Specifies the seed for computations that depend on random numbers
Generalized Linear Modeling	
CATEGORY=	Specifies how to construct estimable functions with multinomial data
EXP	Exponentiates and displays 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 437 in Chapter 19, "Shared Concepts and Topics."

EXACT Statement

EXACT < *'label'* > < **INTERCEPT** > < *effects* > < / *options* > **;**

The EXACT statement performs exact tests of the parameters for the specified *effects* and optionally estimates the parameters and outputs the exact conditional distributions. You can specify the keyword INTERCEPT and any effects in the MODEL statement. Inference on the parameters of the specified effects is performed by conditioning on the sufficient statistics of all the other model parameters (possibly including the intercept).

You can specify several EXACT statements, but they must follow the MODEL statement. Each statement can optionally include an identifying *label*. If several EXACT statements are specified, any statement without a label is assigned a label of the form "Exact*n*," where *n* indicates the *n*th EXACT statement. The label is included in the headers of the displayed exact analysis tables.

If a STRATA statement is also specified, then a stratified exact logistic regression is performed. The model contains a different intercept for each stratum, and these intercepts are conditioned out of the model along with any other nuisance parameters (parameters for effects specified in the MODEL statement that are not in the EXACT statement).

If the LINK=GLOGIT option is specified in the MODEL statement, then the METHOD=DIRECT option is invoked in the EXACTOPTIONS statement by default and a generalized logit model is fit. Since each effect specified in the MODEL statement adds k parameters to the model (where $k + 1$ is the number of response levels), exact analysis of the generalized logit model by using this method is limited to rather small problems.

The CONTRAST, ESTIMATE, LSMEANS, LSMESTIMATE, ODDSRATIO, OUTPUT, ROC, ROCCONTRAST, SCORE, SLICE, STORE, TEST, and UNITS statements are not available with an exact analysis; results from these statements are based on the asymptotic results. Exact analyses are not performed when you specify a WEIGHT statement, a link other than LINK=LOGIT or LINK=GLOGIT, an offset variable, the NOFIT option, or a model selection method. Exact estimation is not available for ordinal response models.

For classification variables, use of the reference parameterization is recommended.

The following options can be specified in each EXACT statement after a slash (/):

ALPHA=*number*

specifies the level of significance α for $100(1 - \alpha)\%$ confidence limits for the parameters or odds ratios. The value of *number* must be between 0 and 1. By default, *number* is equal to the value of the ALPHA= option in the PROC LOGISTIC statement, or 0.05 if that option is not specified.

CLTYPE=EXACT | MIDP

requests either the exact or mid-p confidence intervals for the parameter estimates. By default, the exact intervals are produced. The confidence coefficient can be specified with the ALPHA= option. The mid-p interval can be modified with the MIDPFACTOR= option. See the section "Exact Conditional Logistic Regression" on page 4274 for details.

ESTIMATE < =*keyword* >

estimates the individual parameters (conditioned on all other parameters) for the effects specified in the EXACT statement. For each parameter, a point estimate, a standard error, a confidence interval, and a p-value for a two-sided test that the parameter is zero are displayed. Note that the two-sided p-value is twice the one-sided p-value. You can optionally specify one of the following keywords:

PARM specifies that the parameters be estimated. This is the default.

ODDS specifies that the odds ratios be estimated. If you have classification variables, then you must also specify the PARAM=REF option in the CLASS statement.

BOTH specifies that both the parameters and odds ratios be estimated.

JOINT

performs the joint test that all of the parameters are simultaneously equal to zero, performs individual hypothesis tests for the parameter of each continuous variable, and performs joint tests for the parameters of each classification variable. The joint test is indicated in the "Conditional Exact Tests" table by the label "Joint."

JOINTONLY
performs only the joint test of the parameters. The test is indicated in the "Conditional Exact Tests" table by the label "Joint." When this option is specified, individual tests for the parameters of each continuous variable and joint tests for the parameters of the classification variables are not performed.

MIDPFACTOR=$\delta_1 \mid (\delta_1, \delta_2)$
sets the tie factors used to produce the mid-p hypothesis statistics and the mid-p confidence intervals. δ_1 modifies both the hypothesis tests and confidence intervals, while δ_2 affects only the hypothesis tests. By default, $\delta_1 = 0.5$ and $\delta_2 = 1.0$. See the section "Exact Conditional Logistic Regression" on page 4274 for details.

ONESIDED
requests one-sided confidence intervals and p-values for the individual parameter estimates and odds ratios. The one-sided p-value is the smaller of the left- and right-tail probabilities for the observed sufficient statistic of the parameter under the null hypothesis that the parameter is zero. The two-sided p-values (default) are twice the one-sided p-values. See the section "Exact Conditional Logistic Regression" on page 4274 for more details.

OUTDIST=*SAS-data-set*
names the SAS data set that contains the exact conditional distributions. This data set contains all of the exact conditional distributions that are required to process the corresponding EXACT statement. This data set contains the possible sufficient statistics for the parameters of the effects specified in the EXACT statement, the counts, and, when hypothesis tests are performed on the parameters, the probability of occurrence and the score value for each sufficient statistic. When you request an OUTDIST= data set, the observed sufficient statistics are displayed in the "Sufficient Statistics" table. See the section "OUTDIST= Output Data Set" on page 4281 for more information.

EXACT Statement Examples

In the following example, two exact tests are computed: one for x1 and the other for x2. The test for x1 is based on the exact conditional distribution of the sufficient statistic for the x1 parameter given the observed values of the sufficient statistics for the intercept, x2, and x3 parameters; likewise, the test for x2 is conditional on the observed sufficient statistics for the intercept, x1, and x3.

```
proc logistic;
   model y= x1 x2 x3;
   exact x1 x2;
run;
```

PROC LOGISTIC determines, from all the specified EXACT statements, the distinct conditional distributions that need to be evaluated. For example, there is only one exact conditional distribution for the following two EXACT statements:

```
exact 'One' x1 / estimate=parm;
exact 'Two' x1 / estimate=parm onesided;
```

For each EXACT statement, individual tests for the parameters of the specified effects are computed unless the JOINTONLY option is specified. Consider the following EXACT statements:

```
exact 'E12' x1 x2 / estimate;
exact 'E1'  x1    / estimate;
exact 'E2'  x2    / estimate;
exact 'J12' x1 x2 / joint;
```

In the E12 statement, the parameters for x1 and x2 are estimated and tested separately. Specifying the E12 statement is equivalent to specifying both the E1 and E2 statements. In the J12 statement, the joint test for the parameters of x1 and x2 is computed in addition to the individual tests for x1 and x2.

EXACTOPTIONS Statement

EXACTOPTIONS *options* **;**

The EXACTOPTIONS statement specifies options that apply to every EXACT statement in the program. The following *options* are available:

ABSFCONV=*value*

specifies the absolute function convergence criterion. Convergence requires a small change in the log-likelihood function in subsequent iterations,

$$|l_i - l_{i-1}| < value$$

where l_i is the value of the log-likelihood function at iteration i.

By default, ABSFCONV=1E–12. You can also specify the FCONV= and XCONV= criteria; optimizations are terminated as soon as one criterion is satisfied.

ADDTOBS

adds the observed sufficient statistic to the sampled exact distribution if the statistic was not sampled. This option has no effect unless the METHOD=NETWORKMC option is specified and the ESTIMATE option is specified in the EXACT statement. If the observed statistic has not been sampled, then the parameter estimate does not exist; by specifying this option, you can produce (biased) estimates.

BUILDSUBSETS

builds every distribution for sampling. By default, some exact distributions are created by taking a subset of a previously generated exact distribution. When the METHOD=NETWORKMC option is invoked, this subsetting behavior has the effect of using fewer than the desired n samples; see the N= option for more details. Use the BUILDSUBSETS option to suppress this subsetting.

EPSILON=*value*

controls how the partial sums $\sum_{i=1}^{j} y_i \mathbf{x}_i$ are compared. *value* must be between 0 and 1; by default, *value*=1E–8.

FCONV=*value*

specifies the relative function convergence criterion. Convergence requires a small relative change in the log-likelihood function in subsequent iterations,

$$\frac{|l_i - l_{i-1}|}{|l_{i-1}| + 1\text{E}{-}6} < value$$

where l_i is the value of the log likelihood at iteration i.

By default, FCONV=1E–8. You can also specify the ABSFCONV= and XCONV= criteria; if more than one criterion is specified, then optimizations are terminated as soon as one criterion is satisfied.

MAXTIME=*seconds*
specifies the maximum clock time (in seconds) that PROC LOGISTIC can use to calculate the exact distributions. If the limit is exceeded, the procedure halts all computations and prints a note to the LOG. The default maximum clock time is seven days.

METHOD=*keyword*
specifies which exact conditional algorithm to use for every EXACT statement specified. You can specify one of the following *keywords*:

DIRECT invokes the multivariate shift algorithm of Hirji, Mehta, and Patel (1987). This method directly builds the exact distribution, but it can require an excessive amount of memory in its intermediate stages. METHOD=DIRECT is invoked by default when you are conditioning out at most the intercept, or when the LINK=GLOGIT option is specified in the MODEL statement.

NETWORK invokes an algorithm described in Mehta, Patel, and Senchaudhuri (1992). This method builds a network for each parameter that you are conditioning out, combines the networks, then uses the multivariate shift algorithm to create the exact distribution. The NETWORK method can be faster and require less memory than the DIRECT method. The NETWORK method is invoked by default for most analyses.

NETWORKMC invokes the hybrid network and Monte Carlo algorithm of Mehta, Patel, and Senchaudhuri (1992). This method creates a network, then samples from that network; this method does not reject any of the samples at the cost of using a large amount of memory to create the network. METHOD=NETWORKMC is most useful for producing parameter estimates for problems that are too large for the DIRECT and NETWORK methods to handle and for which asymptotic methods are invalid—for example, for sparse data on a large grid.

N=*n*
specifies the number of Monte Carlo samples to take when the METHOD=NETWORKMC option is specified. By default, $n = 10{,}000$. If the procedure cannot obtain n samples due to a lack of memory, then a note is printed in the SAS log (the number of valid samples is also reported in the listing) and the analysis continues.

The number of samples used to produce any particular statistic might be smaller than n. For example, let *X1* and *X2* be continuous variables, denote their joint distribution by $f(X1,X2)$, and let $f(X1 \mid X2 = x2)$ denote the marginal distribution of *X1* conditioned on the observed value of *X2*. If you request the JOINT test of *X1* and *X2*, then n samples are used to generate the estimate $\hat{f}(X1,X2)$ of $f(X1,X2)$, from which the test is computed. However, the parameter estimate for *X1* is computed from the subset of $\hat{f}(X1,X2)$ that has $X2 = x2$, and this subset need not contain n samples. Similarly, the distribution for each level of a classification variable is created by extracting the appropriate subset from the joint distribution for the CLASS variable.

In some cases, the marginal sample size can be too small to admit accurate estimation of a particular statistic; a note is printed in the SAS log when a marginal sample size is less than 100. Increasing n

increases the number of samples used in a marginal distribution; however, if you want to control the sample size exactly, you can either specify the BUILDSUBSETS option or do both of the following:

- Remove the JOINT option from the EXACT statement.
- Create dummy variables in a DATA step to represent the levels of a CLASS variable, and specify them as independent variables in the MODEL statement.

NOLOGSCALE
specifies that computations for the exact conditional models be computed by using normal scaling. Log scaling can handle numerically larger problems than normal scaling; however, computations in the log scale are slower than computations in normal scale.

ONDISK
uses disk space instead of random access memory to build the exact conditional distribution. Use this option to handle larger problems at the cost of slower processing.

SEED=*seed*
specifies the initial seed for the random number generator used to take the Monte Carlo samples when the METHOD=NETWORKMC option is specified. The value of the SEED= option must be an integer. If you do not specify a seed, or if you specify a value less than or equal to zero, then PROC LOGISTIC uses the time of day from the computer's clock to generate an initial seed.

STATUSN=*number*
prints a status line in the SAS log after every *number* of Monte Carlo samples when the METHOD=NETWORKMC option is specified. The number of samples taken and the current exact p-value for testing the significance of the model are displayed. You can use this status line to track the progress of the computation of the exact conditional distributions.

STATUSTIME=*seconds*
specifies the time interval (in seconds) for printing a status line in the LOG. You can use this status line to track the progress of the computation of the exact conditional distributions. The time interval you specify is approximate; the actual time interval varies. By default, no status reports are produced.

XCONV=*value*
specifies the relative parameter convergence criterion. Convergence requires a small relative parameter change in subsequent iterations,

$$\max_j |\delta_j^{(i)}| < value$$

where

$$\delta_j^{(i)} = \begin{cases} \beta_j^{(i)} - \beta_j^{(i-1)} & |\beta_j^{(i-1)}| < 0.01 \\ \frac{\beta_j^{(i)} - \beta_j^{(i-1)}}{\beta_j^{(i-1)}} & \text{otherwise} \end{cases}$$

and $\beta_j^{(i)}$ is the estimate of the jth parameter at iteration i.

By default, XCONV=1E–4. You can also specify the ABSFCONV= and FCONV= criteria; if more than one criterion is specified, then optimizations are terminated as soon as one criterion is satisfied.

FREQ Statement

FREQ *variable* **;**

The FREQ statement identifies a *variable* that contains the frequency of occurrence of each observation. PROC LOGISTIC treats each observation as if it appears n times, where n is the value of the FREQ variable for the observation. If it is not an integer, the frequency value is truncated to an integer. If the frequency value is less than 1 or missing, the observation is not used in the model fitting. When the FREQ statement is not specified, each observation is assigned a frequency of 1. If you specify more than one FREQ statement, then the first statement is used.

If a SCORE statement is specified, then the FREQ variable is used for computing fit statistics and the ROC curve, but they are not required for scoring. If the DATA= data set in the SCORE statement does not contain the FREQ variable, the frequency values are assumed to be 1 and a warning message is issued in the LOG. If you fit a model and perform the scoring in the same run, the same FREQ variable is used for fitting and scoring. If you fit a model in a previous run and input it with the INMODEL= option in the current run, then the FREQ variable can be different from the one used in the previous run. However, if a FREQ variable was not specified in the previous run, you can still specify a FREQ variable in the current run.

ID Statement

ID *variable*< *variable,...* > **;**

The ID statement specifies variables in the DATA= data set that are used for labeling ROC curves and influence diagnostic plots. If more than one ID variable is specified, then the plots are labeled by concatenating the ID variable values together. See the PLOTS(LABEL) and ROCOPTIONS(ID) options in the PROC LOGISTIC statement for more details.

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 54.6 summarizes the options available in the LSMEANS statement.

Table 54.6 LSMEANS Statement Options

Option	Description
Construction and Computation of LS-Means	
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
Degrees of Freedom and *p*-values	
ADJUST=	Determines the method for multiple-comparison adjustment of LS-means differences
ALPHA=α	Determines the confidence level ($1-\alpha$)
STEPDOWN	Adjusts multiple-comparison *p*-values further in a step-down fashion
Statistical Output	
CL	Constructs confidence limits for means and mean differences
CORR	Displays the correlation matrix of LS-means
COV	Displays the covariance matrix of LS-means
E	Prints the **L** matrix
LINES	Produces a "Lines" display for pairwise LS-means differences
MEANS	Prints the LS-means
PLOTS=	Requests graphs of means and mean comparisons
SEED=	Specifies the seed for computations that depend on random numbers
Generalized Linear Modeling	
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 453 in Chapter 19, "Shared Concepts and Topics."

NOTE: If you have classification variables in your model, then the LSMEANS statement is allowed only if you also specify the PARAM=GLM option.

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 54.7 summarizes the options available in the LSMESTIMATE statement.

Table 54.7 LSMESTIMATE Statement Options

Option	Description
Construction and Computation of LS-Means	
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
Degrees of Freedom and *p*-values	
ADJUST=	Determines the method for multiple-comparison adjustment of LS-means differences
ALPHA=α	Determines the confidence level ($1-\alpha$)
LOWER	Performs one-sided, lower-tailed inference
STEPDOWN	Adjusts multiple-comparison *p*-values further in a step-down fashion
TESTVALUE=	Specifies values under the null hypothesis for tests
UPPER	Performs one-sided, upper-tailed inference
Statistical Output	
CL	Constructs confidence limits for means and mean differences
CORR	Displays the correlation matrix of LS-means
COV	Displays the covariance matrix of LS-means
E	Prints the **L** matrix
ELSM	Prints the **K** matrix
JOINT	Produces a joint *F* or chi-square test for the LS-means and LS-means differences
SEED=	Specifies the seed for computations that depend on random numbers

Table 54.7 *continued*

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

For details about the syntax of the LSMESTIMATE statement, see the section "LSMESTIMATE Statement" on page 470 in Chapter 19, "Shared Concepts and Topics."

NOTE: If you have classification variables in your model, then the LSMESTIMATE statement is allowed only if you also specify the PARAM=GLM option.

MODEL Statement

< *label* **:> MODEL** *variable* **<(** *variable_options* **)>** **=** **<** *effects* **>** **<** */ options* **>** **;**

< *label* **:> MODEL** *events/trials* **=** **<** *effects* **>** **<** */ options* **>** **;**

The MODEL statement names the response variable and the explanatory effects, including covariates, main effects, interactions, and nested effects; see the section "Specification of Effects" on page 3324 of Chapter 42, "The GLM Procedure," for more information. If you omit the explanatory effects, the procedure fits an intercept-only model. You must specify exactly one MODEL statement.

Two forms of the MODEL statement can be specified. The first form, referred to as *single-trial* syntax, is applicable to binary, ordinal, and nominal response data. The second form, referred to as *events/trials* syntax, is restricted to the case of binary response data. The single-trial syntax is used when each observation in the DATA= data set contains information about only a single trial, such as a single subject in an experiment. When each observation contains information about multiple binary-response trials, such as the counts of the number of subjects observed and the number responding, then events/trials syntax can be used.

In the events/trials syntax, you specify two variables that contain count data for a binomial experiment. These two variables are 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. The values of both *events* and (*trials–events*) must be nonnegative and the value of *trials* must be positive for the response to be valid.

In the single-trial syntax, you specify one variable (on the left side of the equal sign) as the response variable. This variable can be character or numeric. Variable_options specific to the response variable can be specified immediately after the response variable with parentheses around them.

For both forms of the MODEL statement, explanatory *effects* follow the equal sign. Variables can be either continuous or classification variables. Classification variables can be character or numeric, and they must be declared in the CLASS statement. When an effect is a classification variable, the procedure inserts a set of coded columns into the design matrix instead of directly entering a single column containing the values of the variable.

Response Variable Options

DESCENDING | DESC

reverses the order of the response categories. If both the DESCENDING and ORDER= options are specified, PROC LOGISTIC orders the response categories according to the ORDER= option and then reverses that order. See the section "Response Level Ordering" on page 4237 for more detail.

EVENT='*category***' | *keyword***

specifies the event category for the binary response model. PROC LOGISTIC 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 quotes, or you can specify one of the following keywords. The default is EVENT=FIRST.

FIRST

designates the first ordered category as the event.

LAST

designates the last ordered category as the event.

One of the most common sets of response levels is {0,1}, with 1 representing 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 generalized logit model and the binary response model. For the generalized logit model, each logit contrasts a nonreference category with 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 quotes, 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. This is the default.

Model Options

Table 54.8 summarizes the options available in the MODEL statement. These options can be specified after a slash (/).

Table 54.8 Model Statement Options

Option	Description
Model Specification Options	
LINK=	Specifies the link function
NOFIT	Suppresses model fitting
NOINT	Suppresses the intercept
OFFSET=	Specifies the offset variable
SELECTION=	Specifies the effect selection method
UNEQUALSLOPES	Specifies cumulative partial proportional odds models
Effect Selection Options	
BEST=	Controls the number of models displayed for SCORE selection
DETAILS	Requests detailed results at each step
FAST	Uses the fast elimination method
HIERARCHY=	Specifies whether and how hierarchy is maintained and whether a single effect or multiple effects are allowed to enter or leave the model per step
INCLUDE=	Specifies the number of effects included in every model
MAXSTEP=	Specifies the maximum number of steps for STEPWISE selection
SEQUENTIAL	Adds or deletes effects in sequential order
SLENTRY=	Specifies the significance level for entering effects
SLSTAY=	Specifies the significance level for removing effects
START=	Specifies the number of variables in the first model
STOP=	Specifies the number of variables in the final model
STOPRES	Adds or deletes variables by the residual chi-square criterion
Model-Fitting Specification Options	
ABSFCONV=	Specifies the absolute function convergence criterion
FCONV=	Specifies the relative function convergence criterion

Table 54.8 *continued*

Option	Description
FIRTH	Specifies Firth's penalized likelihood method
GCONV=	Specifies the relative gradient convergence criterion
MAXFUNCTION=	Specifies the maximum number of function calls for the conditional analysis
MAXITER=	Specifies the maximum number of iterations
NOCHECK	Suppresses checking for infinite parameters
RIDGING=	Specifies the technique used to improve the log-likelihood function when its value is worse than that of the previous step
SINGULAR=	Specifies the tolerance for testing singularity
TECHNIQUE=	Specifies the iterative algorithm for maximization
XCONV=	Specifies the relative parameter convergence criterion
Confidence Interval Options	
ALPHA=	Specifies α for the $100(1-\alpha)\%$ confidence intervals
CLODDS=	Computes confidence intervals for odds ratios
CLPARM=	Computes confidence intervals for parameters
PLCONV=	Specifies the profile-likelihood convergence criterion
Classification Options	
CTABLE	Displays the classification table
PEVENT=	Specifies prior event probabilities
PPROB=	Specifies probability cutpoints for classification
Overdispersion and Goodness-of-Fit Test Options	
AGGREGATE=	Determines subpopulations for Pearson chi-square and deviance
LACKFIT	Requests the Hosmer and Lemeshow goodness-of-fit test
SCALE=	Specifies the method to correct overdispersion
ROC Curve Options	
OUTROC=	Names the output ROC data set
ROCEPS=	Specifies the probability grouping criterion
Regression Diagnostics Options	
INFLUENCE	Displays influence statistics
IPLOTS	Requests index plots
Display Options	
CORRB	Displays the correlation matrix
COVB	Displays the covariance matrix
EXPB	Displays exponentiated values of the estimates
ITPRINT	Displays the iteration history
NODUMMYPRINT	Suppresses the "Class Level Information" table
PARMLABEL	Displays parameter labels
PCORR	Displays the partial correlation statistic
RSQUARE	Displays the generalized R Square
STB	Displays standardized estimates

Table 54.8 *continued*

Option	Description
Computational Options	
BINWIDTH=	Specifies the bin size for estimating association statistics
NOLOGSCALE	Performs calculations by using normal scaling

The following list describes these options.

ABSFCONV=*value*

specifies the absolute function convergence criterion. Convergence requires a small change in the log-likelihood function in subsequent iterations,

$$|l_i - l_{i-1}| < value$$

where l_i is the value of the log-likelihood function at iteration i. See the section "Convergence Criteria" on page 4242 for more information.

AGGREGATE< =*(variable-list)* **>**

specifies the subpopulations on which the Pearson chi-square test statistic and the likelihood ratio chi-square test statistic (deviance) are calculated. 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 explanatory variables in the MODEL statement. The deviance and Pearson goodness-of-fit statistics are calculated only when the SCALE= option is specified. Thus, the AGGREGATE (or AGGREGATE=) option has no effect if the SCALE= option is not specified.

See the section "Rescaling the Covariance Matrix" on page 4257 for more information.

ALPHA=*number*

sets the level of significance α for $100(1-\alpha)\%$ confidence intervals for regression parameters or odds ratios. The value of *number* must be between 0 and 1. By default, *number* is equal to the value of the ALPHA= option in the PROC LOGISTIC statement, or 0.05 if the option is not specified. This option has no effect unless confidence limits for the parameters (CLPARM= option) or odds ratios (CLODDS= option or ODDSRATIO statement) are requested.

BEST=*n*

specifies that *n* models with the highest score chi-square statistics are to be displayed for each model size. It is used exclusively with the SCORE model selection method. If the BEST= option is omitted and there are no more than 10 explanatory variables, then all possible models are listed for each model size. If the option is omitted and there are more than 10 explanatory variables, then the number of models selected for each model size is, at most, equal to the number of explanatory variables listed in the MODEL statement.

BINWIDTH=*width*

specifies the size of the bins used for estimating the association statistics. See the section "Rank Correlation of Observed Responses and Predicted Probabilities" on page 4253 for details. Valid values are $0 \leq width < 1$ (for polytomous response models, $0 < width < 1$). The default *width* is 0.002. If the *width* does not evenly divide the unit interval, it is reduced to a valid value and a message

is displayed in the SAS log. The width is also constrained by the amount of memory available on your machine; if you specify a *width* that is too small, it is adjusted to a value for which memory can be allocated and a note is displayed in the SAS log.

If you have a binary response and specify **BINWIDTH=0**, then no binning is performed and the exact values of the statistics are computed; this method is a bit slower and might require more memory than the binning approach.

The BINWIDTH= option is ignored and no binning is performed when a ROC statement is specified, when ROC graphics are produced, or when the SCORE statement computes an ROC area.

CLODDS=PL | WALD | BOTH

produces confidence intervals for odds ratios of main effects not involved in interactions or nestings. Computation of these confidence intervals is based on the profile likelihood (CLODDS=PL) or based on individual Wald tests (CLODDS=WALD). By specifying CLODDS=BOTH, the procedure computes two sets of confidence intervals for the odds ratios, one based on the profile likelihood and the other based on the Wald tests. The confidence coefficient can be specified with the ALPHA= option. The CLODDS=PL option is not available with the STRATA statement. Classification main effects that use parameterizations other than REF, EFFECT, or GLM are ignored. If you need to compute odds ratios for an effect involved in interactions or nestings, or using some other parameterization, then you should specify an ODDSRATIO statement for that effect.

CLPARM=PL | WALD | BOTH

requests confidence intervals for the parameters. Computation of these confidence intervals is based on the profile likelihood (CLPARM=PL) or individual Wald tests (CLPARM=WALD). If you specify CLPARM=BOTH, the procedure computes two sets of confidence intervals for the parameters, one based on the profile likelihood and the other based on individual Wald tests. The confidence coefficient can be specified with the ALPHA= option. The CLPARM=PL option is not available with the STRATA statement.

See the section "Confidence Intervals for Parameters" on page 4248 for more information.

CORRB

displays the correlation matrix of the parameter estimates.

COVB

displays the covariance matrix of the parameter estimates.

CTABLE

classifies the input binary response observations according to whether the predicted event probabilities are above or below some cutpoint value z in the range $(0, 1)$. An observation is predicted as an event if the predicted event probability exceeds or equals z. You can supply a list of cutpoints other than the default list by specifying the PPROB= option (page 4217). Also, false positive and negative rates can be computed as posterior probabilities by using Bayes' theorem. You can use the PEVENT= option to specify prior probabilities for computing these rates. The CTABLE option is ignored if the data have more than two response levels. The CTABLE option is not available with the STRATA statement.

For more information, see the section "Classification Table" on page 4255.

DETAILS

produces a summary of computational details for each step of the effect selection process. It produces the "Analysis of Effects Eligible for Entry" table before displaying the effect selected for entry for forward or stepwise selection. For each model fitted, it produces the "Type 3 Analysis of Effects" table if the fitted model involves CLASS variables, the "Analysis of Maximum Likelihood Estimates" table, and measures of association between predicted probabilities and observed responses. For the statistics included in these tables, see the section "Displayed Output" on page 4286. The DETAILS option has no effect when SELECTION=NONE.

EXPB

EXPEST

displays the exponentiated values ($e^{\widehat{\beta}_i}$) of the parameter estimates $\widehat{\beta}_i$ in the "Analysis of Maximum Likelihood Estimates" table for the logit model. These exponentiated values are the estimated odds ratios for parameters corresponding to the continuous explanatory variables, and for CLASS effects that use reference or GLM parameterizations.

FAST

uses a computational algorithm of Lawless and Singhal (1978) to compute a first-order approximation to the remaining slope estimates for each subsequent elimination of a variable from the model. Variables are removed from the model based on these approximate estimates. The FAST option is extremely efficient because the model is not refitted for every variable removed. The FAST option is used when SELECTION=BACKWARD and in the backward elimination steps when SELECTION=STEPWISE. The FAST option is ignored when SELECTION=FORWARD or SELECTION=NONE.

FCONV=*value*

specifies the relative function convergence criterion. Convergence requires a small relative change in the log-likelihood function in subsequent iterations,

$$\frac{|l_i - l_{i-1}|}{|l_{i-1}| + \text{1E–6}} < value$$

where l_i is the value of the log likelihood at iteration i. See the section "Convergence Criteria" on page 4242 for more information.

FIRTH

performs Firth's penalized maximum likelihood estimation to reduce bias in the parameter estimates (Heinze and Schemper 2002; Firth 1993). This method is useful in cases of separability, as often occurs when the event is rare, and is an alternative to performing an exact logistic regression. See the section "Firth's Bias-Reducing Penalized Likelihood" on page 4242 for more information.

NOTE: The intercept-only log likelihood is modified by using the full-model Hessian, computed with the slope parameters equal to zero. Therefore, in order to use the likelihood ratio test to compare models, you should use the log likelihoods from the "Model Fit Statistics" tables instead of the Likelihood Ratio statistic that is reported in the "Testing Global Null Hypothesis: BETA=0" table. When fitting a model and scoring a data set in the same PROC LOGISTIC step, the model is fit using Firth's penalty for parameter estimation purposes, but the penalty is not applied to the scored log likelihood.

GCONV=*value*

specifies the relative gradient convergence criterion. Convergence requires that the normalized prediction function reduction is small,

$$\frac{\mathbf{g}_i' \mathbf{I}_i^{-1} \mathbf{g}_i}{|l_i| + 1\text{E}{-}6} < value$$

where l_i is the value of the log-likelihood function, $\mathbf{g}_i$ is the gradient vector, and $\mathbf{I}_i$ is the negative (expected) Hessian matrix, all at iteration i. This is the default convergence criterion, and the default value is 1E–8. See the section "Convergence Criteria" on page 4242 for more information.

HIERARCHY=*keyword*

HIER=*keyword*

specifies whether and how the model hierarchy requirement is applied and whether a single effect or multiple effects are allowed to enter or leave the model in one step. You can specify that only CLASS effects, or both CLASS and interval effects, be subject to the hierarchy requirement. The HIERARCHY= option is ignored unless you also specify one of the following options: SELECTION=FORWARD, SELECTION=BACKWARD, or SELECTION=STEPWISE.

Model hierarchy refers to the requirement that, for any term to be in the model, all effects contained in the term must be present in the model. For example, in order for the interaction A*B to enter the model, the main effects A and B must be in the model. Likewise, neither effect A nor B can leave the model while the interaction A*B is in the model.

The keywords you can specify in the HIERARCHY= option are as follows:

NONE indicates that the model hierarchy is not maintained. Any single effect can enter or leave the model at any given step of the selection process.

SINGLE indicates that only one effect can enter or leave the model at one time, subject to the model hierarchy requirement. For example, suppose that you specify the main effects A and B and the interaction A*B in the model. In the first step of the selection process, either A or B can enter the model. In the second step, the other main effect can enter the model. The interaction effect can enter the model only when both main effects have already been entered. Also, before A or B can be removed from the model, the A*B interaction must first be removed. All effects (CLASS and interval) are subject to the hierarchy requirement.

SINGLECLASS is the same as HIERARCHY=SINGLE except that only CLASS effects are subject to the hierarchy requirement.

MULTIPLE indicates that more than one effect can enter or leave the model at one time, subject to the model hierarchy requirement. In a forward selection step, a single main effect can enter the model, or an interaction can enter the model together with all the effects that are contained in the interaction. In a backward elimination step, an interaction itself, or the interaction together with all the effects that the interaction contains, can be removed. All effects (CLASS and continuous) are subject to the hierarchy requirement.

MULTIPLECLASS is the same as HIERARCHY=MULTIPLE except that only CLASS effects are subject to the hierarchy requirement.

The default value is HIERARCHY=SINGLE, which means that model hierarchy is to be maintained for all effects (that is, both CLASS and continuous effects) and that only a single effect can enter or leave the model at each step.

INCLUDE=*n*

includes the first *n* effects in the MODEL statement in every model. By default, INCLUDE=0. The INCLUDE= option has no effect when SELECTION=NONE.

Note that the INCLUDE= and START= options perform different tasks: the INCLUDE= option includes the first *n* effects variables in every model, whereas the START= option requires only that the first *n* effects appear in the first model.

INFLUENCE< (STDRES) >

displays diagnostic measures for identifying influential observations in the case of a binary response model. For each observation, the INFLUENCE option displays the case number (which is the sequence number of the observation), the values of the explanatory variables included in the final model, and the regression diagnostic measures developed by Pregibon (1981). The STDRES option includes standardized and likelihood residuals in the display.

For a discussion of these diagnostic measures, see the section "Regression Diagnostics" on page 4263. When a STRATA statement is specified, the diagnostics are computed following Storer and Crowley (1985); see the section "Regression Diagnostic Details" on page 4272 for details.

IPLOTS

produces an index plot for the regression diagnostic statistics developed by Pregibon (1981). An index plot is a scatter plot with the regression diagnostic statistic represented on the Y axis and the case number on the X axis. See Example 54.6 for an illustration.

ITPRINT

displays the iteration history of the maximum-likelihood model fitting. The ITPRINT option also displays the last evaluation of the gradient vector and the final change in the –2 Log Likelihood.

LACKFIT< (*n*) >

performs the Hosmer and Lemeshow goodness-of-fit test (Hosmer and Lemeshow 2000) for the case of a binary response model. The subjects are divided into approximately 10 groups of roughly the same size based on the percentiles of the estimated probabilities. The discrepancies between the observed and expected number of observations in these groups are summarized by the Pearson chi-square statistic, which is then compared to a chi-square distribution with t degrees of freedom, where t is the number of groups minus n. By default, $n = 2$. A small p-value suggests that the fitted model is not an adequate model. The LACKFIT option is not available with the STRATA statement. See the section "The Hosmer-Lemeshow Goodness-of-Fit Test" on page 4259 for more information.

LINK=*keyword*

L=*keyword*

specifies the link function linking the response probabilities to the linear predictors. You can specify one of the following keywords. The default is LINK=LOGIT.

CLOGLOG is the complementary log-log function. PROC LOGISTIC fits the binary complementary log-log model when there are two response categories and fits the cumulative complementary log-log model when there are more than two response categories. The aliases are CCLOGLOG, CCLL, and CUMCLOGLOG.

GLOGIT is the generalized logit function. PROC LOGISTIC fits the generalized logit model where each nonreference category is contrasted with the reference category. You can use the response variable option REF= to specify the reference category.

LOGIT is the log odds function. PROC LOGISTIC fits the binary logit model when there are two response categories and fits the cumulative logit model when there are more than two response categories. The aliases are CLOGIT and CUMLOGIT.

PROBIT is the inverse standard normal distribution function. PROC LOGISTIC fits the binary probit model when there are two response categories and fits the cumulative probit model when there are more than two response categories. The aliases are NORMIT, CPROBIT, and CUMPROBIT.

The LINK= option is not available with the STRATA statement.

See the section "Link Functions and the Corresponding Distributions" on page 4238 for more details.

MAXFUNCTION=*number*

specifies the maximum number of function calls to perform when maximizing the conditional likelihood. This option is valid only when a STRATA statement or the UNEQUALSLOPES option is specified. The default values are as follows:

- 125 when the number of parameters $p < 40$
- 500 when $40 \leq p < 400$
- 1000 when $p \geq 400$

Since the optimization is terminated only after completing a full iteration, the number of function calls that are actually performed can exceed *number*. If convergence is not attained, the displayed output and all output data sets created by the procedure contain results based on the last maximum likelihood iteration.

MAXITER=*number*

specifies the maximum number of iterations to perform. By default, MAXITER=25. If convergence is not attained in *number* iterations, the displayed output and all output data sets created by the procedure contain results that are based on the last maximum likelihood iteration.

MAXSTEP=*n*

specifies the maximum number of times any explanatory variable is added to or removed from the model when SELECTION=STEPWISE. The default number is twice the number of explanatory variables in the MODEL statement. When the MAXSTEP= limit is reached, the stepwise selection process is terminated. All statistics displayed by the procedure (and included in output data sets) are based on the last model fitted. The MAXSTEP= option has no effect when SELECTION=NONE, FORWARD, or BACKWARD.

NOCHECK

disables the checking process to determine whether maximum likelihood estimates of the regression parameters exist. If you are sure that the estimates are finite, this option can reduce the execution time if the estimation takes more than eight iterations. For more information, see the section "Existence of Maximum Likelihood Estimates" on page 4242.

NODUMMYPRINT

NODESIGNPRINT

NODP

suppresses the "Class Level Information" table, which shows how the design matrix columns for the CLASS variables are coded.

NOINT

suppresses the intercept for the binary response model, the first intercept for the ordinal response model (which forces all intercepts to be nonnegative), or all intercepts for the generalized logit model. This can be particularly useful in conditional logistic analysis; see Example 54.11.

NOFIT

performs the global score test without fitting the model. The global score test evaluates the joint significance of the effects in the MODEL statement. No further analyses are performed. If the NOFIT option is specified along with other MODEL statement options, NOFIT takes effect and all other options except FIRTH, LINK=, NOINT, OFFSET=, ROC, and TECHNIQUE= are ignored. The NOFIT option is not available with the STRATA statement.

NOLOGSCALE

specifies that computations for the conditional and exact logistic regression models should be computed by using normal scaling. Log scaling can handle numerically larger problems than normal scaling; however, computations in the log scale are slower than computations in normal scale.

OFFSET=*name*

names the offset variable. The regression coefficient for this variable will be fixed at 1. For an example that uses this option, see Example 54.13. You can also use the OFFSET= option to restrict parameters to a fixed value. For example, if you want to restrict the parameter for variable X1 to 1 and the parameter for X2 to 2, compute Restrict= $X1 + 2 * X2$ in a DATA step, specify the option **`offset=Restrict`**, and leave X1 and X2 out of the model.

OUTROC=*SAS-data-set*

OUTR=*SAS-data-set*

creates, for binary response models, an output SAS data set that contains the data necessary to produce the receiver operating characteristic (ROC) curve. The OUTROC= option is not available with the STRATA statement. See the section "OUTROC= Output Data Set" on page 4283 for the list of variables in this data set.

PARMLABEL

displays the labels of the parameters in the "Analysis of Maximum Likelihood Estimates" table.

PCORR

computes the partial correlation statistic $sign(\beta_i)\sqrt{\frac{\chi_i^2-2}{-2\log L_0}}$ for each parameter i, where χ_i^2 is the Wald chi-square statistic for the parameter and $\log L_0$ is the log-likelihood of the intercept-only model (Hilbe 2009, p. 101). If $\chi_i^2 < 2$ then the partial correlation is set to 0. The partial correlation for the intercept terms is set to missing.

PEVENT=*value*

PEVENT=(*list***)**

specifies one prior probability or a list of prior probabilities for the event of interest. The false positive and false negative rates are then computed as posterior probabilities by Bayes' theorem. The prior probability is also used in computing the rate of correct prediction. For each prior probability in the given list, a classification table of all observations is computed. By default, the prior probability is the total sample proportion of events. The PEVENT= option is useful for stratified samples. It has no effect if the CTABLE option is not specified. For more information, see the section "False Positive, False Negative, and Correct Classification Rates Using Bayes' Theorem" on page 4256. Also see the PPROB= option for information about how the *list* is specified.

PLCL

is the same as specifying CLPARM=PL.

PLCONV=*value*

controls the convergence criterion for confidence intervals based on the profile-likelihood function. The quantity *value* must be a positive number, with a default value of 1E–4. The PLCONV= option has no effect if profile-likelihood confidence intervals (CLPARM=PL) are not requested.

PLRL

is the same as specifying CLODDS=PL.

PPROB=*value*

PPROB=(*list***)**

specifies one critical probability value (or cutpoint) or a list of critical probability values for classifying observations with the CTABLE option. Each *value* must be between 0 and 1. A response that has a cross validated predicted probability greater than or equal to the current PPROB= value is classified as an event response. The PPROB= option is ignored if the CTABLE option is not specified.

A classification table for each of several cutpoints can be requested by specifying a list. For example, the following statement requests a classification of the observations for each of the cutpoints 0.3, 0.5, 0.6, 0.7, and 0.8:

```
pprob= (0.3, 0.5 to 0.8 by 0.1)
```

If the PPROB= option is not specified, the default is to display the classification for a range of probabilities from the smallest estimated probability (rounded down to the nearest 0.02) to the highest estimated probability (rounded up to the nearest 0.02) with 0.02 increments.

RIDGING=ABSOLUTE | RELATIVE | NONE

specifies the technique used to improve the log-likelihood function when its value in the current iteration is less than that in the previous iteration. If you specify the RIDGING=ABSOLUTE option, the diagonal elements of the negative (expected) Hessian are inflated by adding the ridge value. If you specify the RIDGING=RELATIVE option, the diagonal elements are inflated by a factor of 1 plus the ridge value. If you specify the RIDGING=NONE option, the crude line search method of taking half a step is used instead of ridging. By default, RIDGING=RELATIVE.

RISKLIMITS

RL

WALDRL

is the same as specifying CLODDS=WALD.

ROCEPS=*number*

specifies a criterion for the ROC curve used for grouping estimated event probabilities that are close to each other. In each group, the difference between the largest and the smallest estimated event probabilities does not exceed the given value. The value for *number* must be between 0 and 1; the default value is the square root of the machine epsilon, which is about 1E–8 (in releases prior to 9.2, the default was 1E–4). The smallest estimated probability in each group serves as a cutpoint for predicting an event response. The ROCEPS= option has no effect unless the OUTROC= option, the BINWIDTH=0 option, or a ROC statement is specified.

RSQUARE

RSQ

requests a generalized R Square measure for the fitted model. For more information, see the section "Generalized Coefficient of Determination" on page 4246.

SCALE=*scale*

enables you to supply the value of the dispersion parameter or to specify the method for estimating the dispersion parameter. It also enables you to display the "Deviance and Pearson Goodness-of-Fit Statistics" table. To correct for overdispersion or underdispersion, the covariance matrix is multiplied by the estimate of the dispersion parameter. Valid values for *scale* are as follows:

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

P | PEARSON specifies that the dispersion parameter be estimated by the Pearson chi-square statistic divided by its degrees of freedom.

WILLIAMS <(*constant*)> specifies that Williams' method be used to model overdispersion. This option can be used only with the events/trials syntax. An optional *constant* can be specified as the scale parameter; otherwise, a scale parameter is estimated under the full model. A set of weights is created based on this scale parameter estimate. These weights can then be used in fitting subsequent models of fewer terms than the full model. When fitting these submodels, specify the computed scale parameter as *constant*. See Example 54.10 for an illustration.

N | NONE specifies that no correction is needed for the dispersion parameter; that is, the dispersion parameter remains as 1. This specification is used for requesting the deviance and the Pearson chi-square statistic without adjusting for overdispersion.

constant sets the estimate of the dispersion parameter to be the square of the given *constant*. For example, SCALE=2 sets the dispersion parameter to 4. The value *constant* must be a positive number.

You can use the AGGREGATE (or AGGREGATE=) option to define the subpopulations for calculating the Pearson chi-square statistic and the deviance. In the absence of the AGGREGATE (or AGGREGATE=) option, each observation is regarded as coming from a different subpopulation. For the events/trials syntax, each observation consists of n Bernoulli trials, where n is the value of the

trials variable. For single-trial syntax, each observation consists of a single response, and for this setting it is not appropriate to carry out the Pearson or deviance goodness-of-fit analysis. Thus, PROC LOGISTIC ignores specifications SCALE=P, SCALE=D, and SCALE=N when single-trial syntax is specified without the AGGREGATE (or AGGREGATE=) option.

The "Deviance and Pearson Goodness-of-Fit Statistics" table includes the Pearson chi-square statistic, the deviance, the degrees of freedom, the ratio of each statistic divided by its degrees of freedom, and the corresponding *p*-value. The SCALE= option is not available with the STRATA statement. For more information, see the section "Overdispersion" on page 4257.

SELECTION=BACKWARD | B
| FORWARD | F
| NONE | N
| STEPWISE | S
| SCORE

specifies the method used to select the variables in the model. BACKWARD requests backward elimination, FORWARD requests forward selection, NONE fits the complete model specified in the MODEL statement, and STEPWISE requests stepwise selection. SCORE requests best subset selection. By default, SELECTION=NONE.

For more information, see the section "Effect-Selection Methods" on page 4244.

SEQUENTIAL

SEQ

forces effects to be added to the model in the order specified in the MODEL statement or eliminated from the model in the reverse order of that specified in the MODEL statement. The model-building process continues until the next effect to be added has an insignificant adjusted chi-square statistic or until the next effect to be deleted has a significant Wald chi-square statistic. The SEQUENTIAL option has no effect when SELECTION=NONE.

SINGULAR=*value*

specifies the tolerance for testing the singularity of the Hessian matrix (Newton-Raphson algorithm) or the expected value of the Hessian matrix (Fisher scoring algorithm). The Hessian matrix is the matrix of second partial derivatives of the log-likelihood function. The test requires that a pivot for sweeping this matrix be at least this number times a norm of the matrix. Values of the SINGULAR= option must be numeric. By default, *value* is the machine epsilon times 1E7, which is approximately 1E–9.

SLENTRY=*value*

SLE=*value*

specifies the significance level of the score chi-square for entering an effect into the model in the FORWARD or STEPWISE method. Values of the SLENTRY= option should be between 0 and 1, inclusive. By default, SLENTRY=0.05. The SLENTRY= option has no effect when SELECTION=NONE, SELECTION=BACKWARD, or SELECTION=SCORE.

SLSTAY=*value*

SLS=*value*

specifies the significance level of the Wald chi-square for an effect to stay in the model in a backward elimination step. Values of the SLSTAY= option should be between 0 and 1, inclusive. By default, SLSTAY=0.05. The SLSTAY= option has no effect when SELECTION=NONE, SELECTION=FORWARD, or SELECTION=SCORE.

START=*n*

begins the FORWARD, BACKWARD, or STEPWISE effect selection process with the first *n* effects listed in the MODEL statement. The value of *n* ranges from 0 to *s*, where *s* is the total number of effects in the MODEL statement. The default value of *n* is *s* for the BACKWARD method and 0 for the FORWARD and STEPWISE methods. Note that START=*n* specifies only that the first *n* effects appear in the first model, while INCLUDE=*n* requires that the first *n* effects be included in every model. For the SCORE method, START=*n* specifies that the smallest models contain *n* effects, where *n* ranges from 1 to *s*; the default value is 1. The START= option has no effect when SELECTION=NONE.

STB

displays the standardized estimates for the parameters in the "Analysis of Maximum Likelihood Estimates" table. The standardized estimate of β_i is given by $\widehat{\beta}_i/(s/s_i)$, where s_i is the total sample standard deviation for the *i*th explanatory variable and

$$s = \begin{cases} \pi/\sqrt{3} & \text{Logistic} \\ 1 & \text{Normal} \\ \pi/\sqrt{6} & \text{Extreme-value} \end{cases}$$

The sample standard deviations for parameters associated with CLASS and EFFECT variables are computed using their codings. For the intercept parameters, the standardized estimates are set to missing.

STOP=*n*

specifies the maximum (SELECTION=FORWARD) or minimum (SELECTION=BACKWARD) number of effects to be included in the final model. The effect selection process is stopped when *n* effects are found. The value of *n* ranges from 0 to *s*, where *s* is the total number of effects in the MODEL statement. The default value of *n* is *s* for the FORWARD method and 0 for the BACKWARD method. For the SCORE method, STOP=*n* specifies that the largest models contain *n* effects, where *n* ranges from 1 to *s*; the default value of *n* is *s*. The STOP= option has no effect when SELECTION=NONE or STEPWISE.

STOPRES

SR

specifies that the removal or entry of effects be based on the value of the residual chi-square. If SELECTION=FORWARD, then the STOPRES option adds the effects into the model one at a time until the residual chi-square becomes insignificant (until the *p*-value of the residual chi-square exceeds the SLENTRY= *value*). If SELECTION=BACKWARD, then the STOPRES option removes effects from the model one at a time until the residual chi-square becomes significant (until the *p*-value of the residual chi-square becomes less than the SLSTAY= *value*). The STOPRES option has no effect when SELECTION=NONE or SELECTION=STEPWISE.

TECHNIQUE=FISHER | NEWTON

TECH=FISHER | NEWTON

specifies the optimization technique for estimating the regression parameters. NEWTON (or NR) is the Newton-Raphson algorithm and FISHER (or FS) is the Fisher scoring algorithm. Both techniques yield the same estimates, but the estimated covariance matrices are slightly different except for the case when the LOGIT link is specified for binary response data. The default is TECHNIQUE=FISHER. If the LINK=GLOGIT option is specified, then Newton-Raphson is the default and only available method. The TECHNIQUE= option is not applied to conditional and exact conditional analyses. This option is not available when the UNEQUALSLOPES option is specified. See the section "Iterative Algorithms for Model Fitting" on page 4240 for more details.

UNEQUALSLOPES< =*effect* **| (***effect-list***) >**

specifies one or more effects in a cumulative response model that have a different set of parameters for each response function. If you specify more than one effect, enclose the effects in parentheses. The effects must be explanatory effects that are specified in the MODEL statement.

If you do not specify this option, the cumulative response models make the parallel lines assumption, $F(\Pr(Y < i)) = \alpha_i + \mathbf{x}'\boldsymbol{\beta}$, where each response function has the same slope parameters $\boldsymbol{\beta}$. If you specify this option without an *effect* or *effect-list*, all slope parameters vary across the response functions, resulting in the model $F(\Pr(Y < i)) = \alpha_i + \mathbf{x}'\boldsymbol{\beta}_i$. Specifying an *effect* or *effect-list* enables you to choose which effects have different parameters across the response functions. If you select the first $\mathbf{x}_1$ parameters to have equal slopes and the remaining $\mathbf{x}_2$ parameters to have unequal slopes, the model can be written as $F(\Pr(Y < i)) = \alpha_i + \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2'\boldsymbol{\beta}_{2i}$. A model that uses the CLOGIT link is called a *partial proportional odds model* (Peterson and Harrell 1990).

For more information, see Example 54.17. The following statements are not available with this option: EFFECTPLOT, ESTIMATE, EXACT, LSMEANS, LSMESTIMATE, ROC, ROCCONTRAST, SLICE, STORE, and STRATA. The following options are not available with this option: FIRTH, RIDGING=, TECHNIQUE=, CTABLE, PEVENT=, PPROB=, OUTROC=, and ROCEPS=.

WALDCL

CL

is the same as specifying CLPARM=WALD.

XCONV=*value*

specifies the relative parameter convergence criterion. Convergence requires a small relative parameter change in subsequent iterations,

$$\max_j |\delta_j^{(i)}| < value$$

where

$$\delta_j^{(i)} = \begin{cases} \beta_j^{(i)} - \beta_j^{(i-1)} & |\beta_j^{(i-1)}| < 0.01 \\ \dfrac{\beta_j^{(i)} - \beta_j^{(i-1)}}{\beta_j^{(i-1)}} & \text{otherwise} \end{cases}$$

and $\beta_j^{(i)}$ is the estimate of the jth parameter at iteration i. See the section "Convergence Criteria" on page 4242 for more information.

NLOPTIONS Statement

NLOPTIONS < *options* > **;**

The NLOPTIONS statement controls the optimization process for conditional analyses (which result from specifying a STRATA statement) and for partial parallel slope models (which result from specifying the UNEQUALSLOPES option in the MODEL statement). An *option* specified in the NLOPTIONS statement takes precedence over the same option specified in the MODEL statement.

The default optimization techniques are chosen according to the number of parameters, p, as follows:

- Newton-Raphson with ridging when $p < 40$
- quasi-Newton when $40 \leq p < 400$
- conjugate gradient when $p \geq 400$

The available *options* are described in the section "NLOPTIONS Statement" on page 482 of Chapter 19, "Shared Concepts and Topics."

ODDSRATIO Statement

ODDSRATIO < *'label'* > *variable* < / *options* > **;**

The ODDSRATIO statement produces odds ratios for variable even when the variable is involved in interactions with other covariates, and for classification variables that use any parameterization. You can also specify variables on which constructed effects are based, in addition to the names of COLLECTION or MULTIMEMBER effects. You can specify several ODDSRATIO statements.

If variable is continuous, then the odds ratios honor any values specified in the UNITS statement. If variable is a classification variable, then odds ratios comparing each pairwise difference between the levels of variable are produced. If variable interacts with a continuous variable, then the odds ratios are produced at the mean of the interacting covariate by default. If variable interacts with a classification variable, then the odds ratios are produced at each level of the interacting covariate by default. The computed odds ratios are independent of the parameterization of any classification variable.

The odds ratios are uniquely labeled by concatenating the following terms to variable:

1. If this is a polytomous response model, then prefix the response variable and the level describing the logit followed by a colon; for example, "Y 0:".
2. If variable is continuous and the UNITS statement provides a value that is not equal to 1, then append "Units=value"; otherwise, if variable is a classification variable, then append the levels being contrasted; for example, "cat vs dog".
3. Append all interacting covariates preceded by "At"; for example, "At X=1.2 A=cat".

If you are also creating odds ratio plots, then this label is displayed on the plots (see the PLOTS option for more information). If you specify a 'label' in the ODDSRATIO statement, then the odds ratios produced by this statement are also labeled: 'label', 'label 2', 'label 3',..., and these are the labels used in the plots. If

there are any duplicated labels across all ODDSRATIO statements, then the corresponding odds ratios are not displayed on the plots.

The following *options* are available:

AT(*covariate=value-list* | **REF** | **ALL< ...***covariate=value-list* | **REF** | **ALL >)**
specifies fixed levels of the interacting covariates. If a specified *covariate* does not interact with the *variable*, then its AT list is ignored.

For continuous interacting covariates, you can specify one or more numbers in the *value-list*. For classification covariates, you can specify one or more formatted levels of the covariate enclosed in single quotes (for example, `A='cat' 'dog'`), you can specify the keyword REF to select the reference-level, or you can specify the keyword ALL to select all levels of the classification variable. By default, continuous covariates are set to their means, while CLASS covariates are set to ALL. For a model that includes a classification variable A={cat,dog} and a continuous covariate X, specifying `AT(A='cat' X=7 9)` will set A to 'cat', and X to 7 and then 9.

CL=WALD | **PL** | **BOTH**
specifies whether to create Wald or profile-likelihood confidence limits, or both. By default, Wald confidence limits are produced.

DIFF=REF | **ALL**
specifies whether the odds ratios for a classification variable are computed against the reference level, or all pairs of variable are compared. By default, DIFF=ALL. The DIFF= option is ignored when variable is continuous.

PLCONV=*value*
controls the convergence criterion for confidence intervals based on the profile-likelihood function. The quantity *value* must be a positive number, with a default value of 1E–4. The PLCONV= option has no effect if profile-likelihood confidence intervals (CL=PL) are not requested.

PLMAXITER=*n*
specifies the maximum number of iterations to perform. By default, PLMAXITER=25. If convergence is not attained in *n* iterations, the odds ratio or the confidence limits are set to missing. The PLMAXITER= option has no effect if profile-likelihood confidence intervals (CL=PL) are not requested.

PLSINGULAR=*value*
specifies the tolerance for testing the singularity of the Hessian matrix (Newton-Raphson algorithm) or the expected value of the Hessian matrix (Fisher scoring algorithm). The test requires that a pivot for sweeping this matrix be at least this number times a norm of the matrix. Values of the PLSINGULAR= option must be numeric. By default, *value* is the machine epsilon times 1E7, which is approximately 1E–9. The PLSINGULAR= option has no effect if profile-likelihood confidence intervals (CL=PL) are not requested.

OUTPUT Statement

OUTPUT < **OUT=***SAS-data-set* > < *options* > **;**

The OUTPUT statement creates a new SAS data set that contains all the variables in the input data set and, optionally, the estimated linear predictors and their standard error estimates, the estimates of the cumulative or individual response probabilities, and the confidence limits for the cumulative probabilities. Regression diagnostic statistics and estimates of cross validated response probabilities are also available for binary response models. If you specify more than one OUTPUT statement, only the last one is used. Formulas for the statistics are given in the sections "Linear Predictor, Predicted Probability, and Confidence Limits" on page 4253 and "Regression Diagnostics" on page 4263, and, for conditional logistic regression, in the section "Conditional Logistic Regression" on page 4271.

If you use the single-trial syntax, the data set also contains a variable named _LEVEL_, which indicates the level of the response that the given row of output is referring to. For instance, the value of the cumulative probability variable is the probability that the response variable is as large as the corresponding value of _LEVEL_. For details, see the section "OUT= Output Data Set in the OUTPUT Statement" on page 4280.

The estimated linear predictor, its standard error estimate, all predicted probabilities, and the confidence limits for the cumulative probabilities are computed for all observations in which the explanatory variables have no missing values, even if the response is missing. By adding observations with missing response values to the input data set, you can compute these statistics for new observations or for settings of the explanatory variables not present in the data without affecting the model fit. Alternatively, the SCORE statement can be used to compute predicted probabilities and confidence intervals for new observations.

Table 54.9 summarizes the options available in the OUTPUT statement. These options can be specified after a slash (/). The statistic and diagnostic options specify the statistics to be included in the output data set and name the new variables that contain the statistics. If a STRATA statement is specified, only the PREDICTED=, DFBETAS=, and H= options are available; see the section "Regression Diagnostic Details" on page 4272 for details.

Table 54.9 OUTPUT Statement Options

Option	Description
ALPHA=	Specifies α for the $100(1-\alpha)\%$ confidence intervals
OUT=	Names the output data set
Statistic Options	
LOWER=	Names the lower confidence limit
PREDICTED=	Names the predicted probabilities
PREDPROBS=	Requests the individual, cumulative, or cross validated predicted probabilities
STDXBETA=	Names the standard error estimate of the linear predictor
UPPER=	Names the upper confidence limit
XBETA=	Names the linear predictor
Diagnostic Options for Binary Response	
C=	Names the confidence interval displacement
CBAR=	Names the confidence interval displacement
DFBETAS=	Names the standardized deletion parameter differences
DIFCHISQ=	Names the deletion chi-square goodness-of-fit change
DIFDEV=	Names the deletion deviance change
H=	Names the leverage
RESCHI=	Names the Pearson chi-square residual
RESDEV=	Names the deviance residual

Table 54.9 *continued*

Option	Description
RESLIK=	Names the likelihood residual
STDRESCHI=	Names the standardized Pearson chi-square residual
STDRESDEV=	Names the standardized deviance residual

The following list describes these options.

ALPHA=*number*

sets the level of significance α for $100(1-\alpha)\%$ confidence limits for the appropriate response probabilities. The value of *number* must be between 0 and 1. By default, *number* is equal to the value of the ALPHA= option in the PROC LOGISTIC statement, or 0.05 if that option is not specified.

C=*name*

specifies the confidence interval displacement diagnostic that measures the influence of individual observations on the regression estimates.

CBAR=*name*

specifies the confidence interval displacement diagnostic that measures the overall change in the global regression estimates due to deleting an individual observation.

DFBETAS=_ALL_

DFBETAS=*var-list*

specifies the standardized differences in the regression estimates for assessing the effects of individual observations on the estimated regression parameters in the fitted model. You can specify a list of up to $s+1$ variable names, where s is the number of explanatory variables in the MODEL statement, or you can specify just the keyword _ALL_. In the former specification, the first variable contains the standardized differences in the intercept estimate, the second variable contains the standardized differences in the parameter estimate for the first explanatory variable in the MODEL statement, and so on. In the latter specification, the DFBETAS statistics are named DFBETA_*xxx*, where *xxx* is the name of the regression parameter. For example, if the model contains two variables X1 and X2, the specification DFBETAS=_ALL_ produces three DFBETAS statistics: DFBETA_Intercept, DFBETA_X1, and DFBETA_X2. If an explanatory variable is not included in the final model, the corresponding output variable named in DFBETAS=*var-list* contains missing values.

DIFCHISQ=*name*

specifies the change in the chi-square goodness-of-fit statistic attributable to deleting the individual observation.

DIFDEV=*name*

specifies the change in the deviance attributable to deleting the individual observation.

H=*name*

specifies the diagonal element of the hat matrix for detecting extreme points in the design space.

LOWER=*name*

L=*name*

names the variable containing the lower confidence limits for π, where π is the probability of the event response if events/trials syntax or single-trial syntax with binary response is specified; for a cumulative model, π is cumulative probability (that is, the probability that the response is less than

or equal to the value of _LEVEL_); for the generalized logit model, it is the individual probability (that is, the probability that the response category is represented by the value of _LEVEL_). See the ALPHA= option to set the confidence level.

OUT=*SAS-data-set*
names the output data set. If you omit the OUT= option, the output data set is created and given a default name by using the DATA*n* convention.

PREDICTED=*name*

PRED=*name*

PROB=*name*

P=*name*
names the variable containing the predicted probabilities. For the events/trials syntax or single-trial syntax with binary response, it is the predicted event probability. For a cumulative model, it is the predicted cumulative probability (that is, the probability that the response variable is less than or equal to the value of _LEVEL_); and for the generalized logit model, it is the predicted individual probability (that is, the probability of the response category represented by the value of _LEVEL_).

PREDPROBS=(*keywords***)**
requests individual, cumulative, or cross validated predicted probabilities. Descriptions of the *keywords* are as follows.

INDIVIDUAL | I requests the predicted probability of each response level. For a response variable Y with three levels, 1, 2, and 3, the individual probabilities are Pr(Y=1), Pr(Y=2), and Pr(Y=3).

CUMULATIVE | C requests the cumulative predicted probability of each response level. For a response variable Y with three levels, 1, 2, and 3, the cumulative probabilities are $\Pr(Y \leq 1)$, $\Pr(Y \leq 2)$, and $\Pr(Y \leq 3)$. The cumulative probability for the last response level always has the constant value of 1. For generalized logit models, the cumulative predicted probabilities are not computed and are set to missing.

CROSSVALIDATE | XVALIDATE | X requests the cross validated individual predicted probability of each response level. These probabilities are derived from the leave-one-out principle—that is, dropping the data of one subject and reestimating the parameter estimates. PROC LOGISTIC uses a less expensive one-step approximation to compute the parameter estimates. This option is valid only for binary response models; for nominal and ordinal models, the cross validated probabilities are not computed and are set to missing.

See the section "Details of the PREDPROBS= Option" on page 4227 at the end of this section for further details.

RESCHI=*name*
specifies the Pearson (chi-square) residual for identifying observations that are poorly accounted for by the model.

RESDEV=*name*
specifies the deviance residual for identifying poorly fitted observations.

RESLIK=*name*
specifies the likelihood residual for identifying poorly fitted observations.

STDRESCHI=*name*
specifies the standardized Pearson (chi-square) residual for identifying observations that are poorly accounted for by the model.

STDRESDEV=*name*
specifies the standardized deviance residual for identifying poorly fitted observations.

STDXBETA=*name*
names the variable containing the standard error estimates of XBETA. See the section "Linear Predictor, Predicted Probability, and Confidence Limits" on page 4253 for details.

UPPER=*name*
U=*name*
names the variable containing the upper confidence limits for π, where π is the probability of the event response if events/trials syntax or single-trial syntax with binary response is specified; for a cumulative model, π is cumulative probability (that is, the probability that the response is less than or equal to the value of _LEVEL_); for the generalized logit model, it is the individual probability (that is, the probability that the response category is represented by the value of _LEVEL_). See the ALPHA= option to set the confidence level.

XBETA=*name*
names the variable containing the estimates of the linear predictor $\alpha_i + \boldsymbol{\beta}'\mathbf{x}$, where i is the corresponding ordered value of _LEVEL_.

Details of the PREDPROBS= Option

You can request any of the three types of predicted probabilities. For example, you can request both the individual predicted probabilities and the cross validated probabilities by specifying PREDPROBS=(I X).

When you specify the PREDPROBS= option, two automatic variables, _FROM_ and _INTO_, are included for the single-trial syntax and only one variable, _INTO_, is included for the events/trials syntax. The variable _FROM_ contains the formatted value of the observed response. The variable _INTO_ contains the formatted value of the response level with the largest individual predicted probability.

If you specify PREDPROBS=INDIVIDUAL, the OUT= data set contains k additional variables representing the individual probabilities, one for each response level, where k is the maximum number of response levels across all BY groups. The names of these variables have the form IP_*xxx*, where *xxx* represents the particular level. The representation depends on the following situations:

- If you specify events/trials syntax, *xxx* is either 'Event' or 'Nonevent'. Thus, the variable containing the event probabilities is named IP_Event and the variable containing the nonevent probabilities is named IP_Nonevent.
- If you specify the single-trial syntax with more than one BY group, *xxx* is 1 for the first ordered level of the response, 2 for the second ordered level of the response, and so forth, as given in the "Response Profile" table. The variable containing the predicted probabilities Pr(Y=1) is named IP_1, where Y is the response variable. Similarly, IP_2 is the name of the variable containing the predicted probabilities Pr(Y=2), and so on.

- If you specify the single-trial syntax with no BY-group processing, *xxx* is the left-justified formatted value of the response level (the value might be truncated so that IP_*xxx* does not exceed 32 characters). For example, if Y is the response variable with response levels 'None', 'Mild', and 'Severe', the variables representing individual probabilities Pr(Y='None'), P(Y='Mild'), and P(Y='Severe') are named IP_None, IP_Mild, and IP_Severe, respectively.

If you specify PREDPROBS=CUMULATIVE, the OUT= data set contains *k* additional variables representing the cumulative probabilities, one for each response level, where *k* is the maximum number of response levels across all BY groups. The names of these variables have the form CP_*xxx*, where *xxx* represents the particular response level. The naming convention is similar to that given by PREDPROBS=INDIVIDUAL. The PREDPROBS=CUMULATIVE values are the same as those output by the PREDICT= option, but are arranged in variables on each output observation rather than in multiple output observations.

If you specify PREDPROBS=CROSSVALIDATE, the OUT= data set contains *k* additional variables representing the cross validated predicted probabilities of the *k* response levels, where *k* is the maximum number of response levels across all BY groups. The names of these variables have the form XP_*xxx*, where *xxx* represents the particular level. The representation is the same as that given by PREDPROBS=INDIVIDUAL except that for the events/trials syntax there are four variables for the cross validated predicted probabilities instead of two:

XP_EVENT_R1E is the cross validated predicted probability of an event when a single event is removed from the current observation.

XP_NONEVENT_R1E is the cross validated predicted probability of a nonevent when a single event is removed from the current observation.

XP_EVENT_R1N is the cross validated predicted probability of an event when a single nonevent is removed from the current observation.

XP_NONEVENT_R1N is the cross validated predicted probability of a nonevent when a single nonevent is removed from the current observation.

The cross validated predicted probabilities are precisely those used in the CTABLE option. See the section "Predicted Probability of an Event for Classification" on page 4255 for details of the computation.

ROC Statement

ROC < *'label'* > < *specification* > < / *options* > **;**

The ROC statements specify models to be used in the ROC comparisons. You can specify more than one ROC statement. ROC statements are identified by their *label*—if you do not specify a *label*, the *i*th ROC statement is labeled "ROC*i*". Additionally, the specified or selected model is labeled with the MODEL statement label or "Model" if the MODEL label is not present. The *specification* can be either a list of effects that have previously been specified in the MODEL statement, or PRED=*variable*, where the *variable* does not have to be specified in the MODEL statement. The PRED= option enables you to input a criterion produced outside PROC LOGISTIC; for example, you can fit a random-intercept model by using PROC GLIMMIX or use survey weights in PROC SURVEYLOGISTIC, then use the predicted values from those models to produce an ROC curve for the comparisons. If you do not make a *specification*, then an intercept-only model is fit to the data, resulting in a noninformative ROC curve that can be used for comparing the area under another ROC curve to 0.5.

You can specify a ROCCONTRAST statement and a ROCOPTIONS option in the PROC LOGISTIC statement to control how the models are compared, while the PLOTS=ROC option controls the ODS Graphics displays. See Example 54.8 for an example that uses the ROC statement.

If you specify any *options*, then a "ROC Model Information" table summarizing the new ROC model is displayed. The *options* are ignored for the PRED= specification. The following *options* are available:

NOOFFSET
does not include an offset variable if the OFFSET= option is specified in the MODEL statement. A constant offset has no effect on the ROC curve, although the cutpoints might be different, but a nonconstant offset can affect the parameter estimates and hence the ROC curve.

LINK=*keyword*
specifies the link function to be used in the model. The available keywords are LOGIT, NORMIT, and CLOGLOG. The logit link is the default. Note that the LINK= option in the MODEL statement is ignored.

ROCCONTRAST Statement

ROCCONTRAST < *'label'* > < *contrast* > < / *options* > **;**

The ROCCONTRAST statement compares the different ROC models. You can specify only one ROCCONTRAST statement. The ROCOPTIONS options in the PROC LOGISTIC statement control how the models are compared. You can specify one of the following *contrast* specifications:

REFERENCE< (MODEL | *'roc-label'***) >**
produces a contrast matrix of differences between each ROC curve and a reference curve. The MODEL keyword specifies that the reference curve is that produced from the MODEL statement; the *roc-label* specifies the label of the ROC curve that is to be used as the reference curve. If neither the MODEL keyword nor the *roc-label* label is specified, then the reference ROC curve is either the curve produced from the MODEL statement, the selected model if a selection method is specified, or the model from the first ROC statement if the NOFIT option is specified.

ADJACENTPAIRS
produces a contrast matrix of each ROC curve minus the succeeding curve.

matrix
specifies the contrast in the form `row1,row2,...`, where each *row* contains the coefficients used to compare the ROC curves. Each *row* must contain the same number of entries as there are ROC curves being compared. The elements of each *row* refer to the ROC statements in the order in which they are specified. However, the first element of each *row* refers either to the fitted model, the selected model if a SELECTION= method is specified, or the first specified ROC statement if the NOFIT option is specified.

If no *contrast* is specified, then the REFERENCE contrast with the default reference curve is used. See the section "Comparing ROC Curves" on page 4261 for more information about comparing ROC curves, and see Example 54.8 for an example.

The following *options* are available:

E

displays the contrast.

ESTIMATE < = ROWS | ALLPAIRS >

produces estimates of each row of the contrast when ESTIMATE or ESTIMATE=ROWS is specified. If the ESTIMATE=ALLPAIRS option is specified, then estimates of every pairwise difference of ROC curves are produced.

The row contrasts are labeled "ModelLabel1 – ModelLabel2", where the model labels are as described in the ROC statement; in particular, for the REFERENCE contrast, ModelLabel2 is the reference model label. If you specify your own contrast matrix, then the *i*th contrast row estimate is labeled "Row*i*".

COV

displays covariance matrices used in the computations.

SCORE Statement

SCORE < *options* > **;**

The SCORE statement creates a data set that contains all the data in the DATA= data set together with posterior probabilities and, optionally, prediction confidence intervals. Fit statistics are displayed on request. If you have binary response data, the SCORE statement can be used to create a data set containing data for the ROC curve. You can specify several SCORE statements. FREQ, WEIGHT, and BY statements can be used with the SCORE statements. The SCORE statement is not available with the STRATA statement.

If a SCORE statement is specified in the same run as fitting the model, FORMAT statements should be specified after the SCORE statement in order for the formats to apply to all the DATA= and PRIOR= data sets in the SCORE statement.

See the section "Scoring Data Sets" on page 4266 for more information, and see Example 54.15 for an illustration of how to use this statement.

Table 54.10 summarizes the options available in the SCORE statement.

Table 54.10 SCORE Statement Options

Option	Description
ALPHA=	Specifies the significance level
CLM	Outputs the Wald-test-based confidence limits
CUMULATIVE	Outputs the cumulative predicted probabilities
DATA=	Names the SAS data that you want to score
FITSTAT	Displays fit statistics
OUT=	Names the SAS data set that contains the predicted information
OUTROC=	Names the SAS data set that contains the ROC curve
PRIOR=	Names the SAS data set that contains the priors of the response categories
PRIOREVENT=	Specifies the prior event probability
ROCEPS=	Specifies the criterion for grouping estimated event probabilities

You can specify the following options:

ALPHA=*number*
specifies the significance level α for $100(1-\alpha)\%$ confidence intervals. By default, the value of *number* is equal to the ALPHA= option in the PROC LOGISTIC statement, or 0.05 if that option is not specified. This option has no effect unless the CLM option in the SCORE statement is requested.

CLM
outputs the Wald-test-based confidence limits for the predicted probabilities. This option is not available when the INMODEL= data set is created with the NOCOV option.

CUMULATIVE
outputs the cumulative predicted probabilities $\Pr(Y \leq i), i = 1, \ldots, k+1$, to the OUT= data set. This option is valid only when you have more than two response levels; otherwise, the option is ignored and a note is printed in the SAS log. These probabilities are named CP_*level_i*, where *level_i* is the *i*th response level.

If the CLM option is also specified in the SCORE statement, then the Wald-based confidence limits for the cumulative predicted probabilities are also output. The confidence limits are named CLCL_*level_i* and CUCL_*level_i*. In particular, for the lowest response level, the cumulative values (CP, CLCL, CUCL) should be identical to the individual values (P, LCL, UCL), and for the highest response level CP=CLCL=CUCL=1.

DATA=*SAS-data-set*
names the SAS data set that you want to score. If you omit the DATA= option in the SCORE statement, then scoring is performed on the DATA= input data set in the PROC LOGISTIC statement, if specified; otherwise, the DATA=_LAST_ data set is used.

It is not necessary for the DATA= data set in the SCORE statement to contain the response variable unless you are specifying the FITSTAT or OUTROC= option.

Only those variables involved in the fitted model effects are required in the DATA= data set in the SCORE statement. For example, the following statements use forward selection to select effects:

```
proc logistic data=Neuralgia outmodel=sasuser.Model;
   class Treatment Sex;
   model Pain(event='Yes')= Treatment|Sex Age
         / selection=forward sle=.01;
run;
```

Suppose Treatment and Age are the effects selected for the final model. You can score a data set that does not contain the variable Sex since the effect Sex is not in the model that the scoring is based on. For example, the following statements score the Neuralgia data set after dropping the Sex variable:

```
proc logistic inmodel=sasuser.Model;
   score data=Neuralgia(drop=Sex);
run;
```

FITSTAT
displays fit statistics for the data set you are scoring. The data set must contain the response variable. See the section "Fit Statistics for Scored Data Sets" on page 4266 for details.

OUT=*SAS-data-set*
names the SAS data set that contains the predicted information. If you omit the OUT= option, the output data set is created and given a default name by using the DATA*n* convention.

OUTROC=*SAS-data-set*
names the SAS data set that contains the ROC curve for the DATA= data set. The ROC curve is computed only for binary response data. See the section "OUTROC= Output Data Set" on page 4283 for the list of variables in this data set.

PRIOR=*SAS-data-set*
names the SAS data set that contains the priors of the response categories. The priors can be values proportional to the prior probabilities; thus, they do not necessarily sum to one. This data set should include a variable named _PRIOR_ that contains the prior probabilities. For events/trials MODEL statement syntax, this data set should also include an _OUTCOME_ variable that contains the values EVENT and NONEVENT; for single-trial syntax, this data set should include the response variable that contains the unformatted response categories. See Example 54.15 for an example.

PRIOREVENT=*value*
specifies the prior event probability for a binary response model. If both PRIOR= and PRIOREVENT= options are specified, the PRIOR= option takes precedence.

ROCEPS=*value*
specifies the criterion for grouping estimated event probabilities that are close to each other for the ROC curve. In each group, the difference between the largest and the smallest estimated event probability does not exceed the given value. The *value* must be between 0 and 1; the default value is the square root of the machine epsilon, which is about 1E–8 (in releases prior to 9.2, the default was 1E–4). The smallest estimated probability in each group serves as a cutpoint for predicting an event response. The ROCEPS= option has no effect if the OUTROC= option is not specified in the SCORE statement.

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 498 in Chapter 19, "Shared Concepts and Topics."

NOTE: If you have classification variables in your model, then the SLICE statement is allowed only if you also specify the PARAM=GLM option.

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 501 in Chapter 19, "Shared Concepts and Topics."

STRATA Statement

STRATA *variable* < (*option*) > . . . < *variable* < (*option*) > > < / *options* > **;**

The STRATA statement names the *variables* that define *strata* or *matched sets* to use in *stratified logistic regression* of binary response data.

Observations that have the same *variable* values are in the same matched set. For a stratified logistic model, you can analyze 1: 1, 1: n, m: n, and general m_i: n_i matched sets where the number of cases and controls varies across strata. At least one variable must be specified to invoke the stratified analysis, and the usual unconditional asymptotic analysis is not performed. The stratified logistic model has the form

$$\text{logit}(\pi_{hi}) = \alpha_h + \mathbf{x}'_{hi}\boldsymbol{\beta}$$

where π_{hi} is the event probability for the ith observation in stratum h with covariates $\mathbf{x}_{hi}$ and where the stratum-specific intercepts α_h are the nuisance parameters that are to be conditioned out.

STRATA variables can also be specified in the MODEL statement as classification or continuous covariates; however, the effects are nondegenerate only when crossed with a nonstratification variable. Specifying several STRATA statements is the same as specifying one STRATA statement that contains all the strata variables. The STRATA variables can be either character or numeric, and the formatted values of the STRATA variables determine the levels. Thus, you can also use formats to group values into levels; see the discussion of the FORMAT procedure in the *Base SAS Procedures Guide*.

The "Strata Summary" table is displayed by default. For an exact logistic regression, it displays the number of strata that have a specific number of events and non-events. For example, if you are analyzing a 1: 5 matched study, this table enables you to verify that every stratum in the analysis has exactly one event and five non-events. Strata that contain only events or only non-events are reported in this table, but such strata are uninformative and are not used in the analysis.

If an EXACT statement is also specified, then a stratified *exact* logistic regression is performed.

The EFFECTPLOT, SCORE, and WEIGHT statements are not available with a STRATA statement. The following MODEL options are also not supported with a STRATA statement: CLPARM=PL, CLODDS=PL, CTABLE, FIRTH, LACKFIT, LINK=, NOFIT, OUTMODEL=, OUTROC=, ROC, and SCALE=.

The following *option* can be specified for a stratification variable by enclosing the option in parentheses after the variable name, or it can be specified globally for all STRATA variables after a slash (/).

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

The following strata *options* are also available after the slash:

CHECKDEPENDENCY | CHECK=*keyword*
specifies which variables are to be tested for dependency before the analysis is performed. The available *keywords* are as follows:

NONE performs no dependence checking. Typically, a message about a singular information matrix is displayed if you have dependent variables. Dependent variables can be identified after the analysis by noting any missing parameter estimates.

COVARIATES checks dependence between covariates and an added intercept. Dependent covariates are removed from the analysis. However, covariates that are linear functions of the strata variable might not be removed, which results in a singular information matrix message being displayed in the SAS log. This is the default.

ALL checks dependence between all the strata and covariates. This option can adversely affect performance if you have a large number of strata.

NOSUMMARY
suppresses the display of the "Strata Summary" table.

INFO
displays the "Strata Information" table, which includes the stratum number, levels of the STRATA variables that define the stratum, the number of events, the number of non-events, and the total frequency for each stratum. Since the number of strata can be very large, this table is displayed only by request.

TEST Statement

< *label***:>** **TEST** *equation1* **<,** *equation2***,** *...* **>** **<** */ option* **>** **;**

The TEST statement tests linear hypotheses about the regression coefficients. The Wald test is used to perform a joint test of the null hypotheses $H_0: \mathbf{L}\boldsymbol{\beta} = \mathbf{c}$ specified in a single TEST statement, where $\boldsymbol{\beta}$ is the vector of intercept and slope parameters. When $\mathbf{c} = \mathbf{0}$ you should specify a CONTRAST statement instead.

Each *equation* specifies a linear hypothesis (a row of the $\mathbf{L}$ matrix and the corresponding element of the $\mathbf{c}$ vector). Multiple *equations* are separated by commas. The *label*, which must be a valid SAS name, is used to identify the resulting output and should always be included. You can submit multiple TEST statements.

The form of an *equation* is as follows:

*term***<** ± term **...>** **<= ±***term* **<±***term***...>>**

where *term* is a parameter of the model, or a constant, or a constant times a parameter. Intercept and CLASS variable parameter names should be specified as described in the section "Parameter Names in the OUTEST= Data Set" on page 4279. Note for generalized logit models that this enables you to construct tests of parameters from specific logits. When no equal sign appears, the expression is set to 0. The following statements illustrate possible uses of the TEST statement:

```
proc logistic;
   model y= a1 a2 a3 a4;
   test1: test intercept + .5 * a2 = 0;
   test2: test intercept + .5 * a2;
   test3: test a1=a2=a3;
   test4: test a1=a2, a2=a3;
run;
```

Note that the first and second TEST statements are equivalent, as are the third and fourth TEST statements.

You can specify the following option in the TEST statement after a slash(/):

PRINT

displays intermediate calculations in the testing of the null hypothesis H_0: $\mathbf{L}\boldsymbol{\beta} = \mathbf{c}$. This includes $\mathbf{L}\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\mathbf{L}'$ bordered by $(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{c})$ and $[\mathbf{L}\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\mathbf{L}']^{-1}$ bordered by $[\mathbf{L}\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\mathbf{L}']^{-1}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{c})$, where $\widehat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ and $\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$ is the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$.

For more information, see the section "Testing Linear Hypotheses about the Regression Coefficients" on page 4262.

UNITS Statement

UNITS < *independent1*=*list1* < *independent2*=*list2*. . . >> < / *option*> **;**

The UNITS statement enables you to specify units of change for the continuous explanatory variables so that customized odds ratios can be estimated. If you specify more than one UNITS statement, only the last one is used. An estimate of the corresponding odds ratio is produced for each unit of change specified for an explanatory variable. The UNITS statement is ignored for CLASS variables. Odds ratios are computed only for main effects that are not involved in interactions or nestings, unless an ODDSRATIO statement is also specified. If the CLODDS= option is specified in the MODEL statement, the corresponding confidence limits for the odds ratios are also displayed, as are odds ratios and confidence limits for any CLASS main effects that are not involved in interactions or nestings. The CLASS effects must use the GLM, reference, or effect coding.

The UNITS statement also enables you to customize the odds ratios for effects specified in ODDSRATIO statements, in which case interactions and nestings are allowed, and CLASS variables can be specified with any parameterization.

The term *independent* is the name of an explanatory variable and *list* represents a list of units of change, separated by spaces, that are of interest for that variable. Each unit of change in a list has one of the following forms:

- *number*
- SD or –SD
- *number* * SD

where *number* is any nonzero number, and SD is the sample standard deviation of the corresponding independent variable. For example, $X = -2$ requests an odds ratio that represents the change in the odds when

the variable X is decreased by two units. $X = 2*\text{SD}$ requests an estimate of the change in the odds when X is increased by two sample standard deviations.

You can specify the following option in the UNITS statement after a slash(/):

DEFAULT=*list*

gives a list of units of change for all explanatory variables that are not specified in the UNITS statement. Each unit of change can be in any of the forms described previously. If the DEFAULT= option is not specified, PROC LOGISTIC does not produce customized odds ratio estimates for any continuous explanatory variable that is not listed in the UNITS statement.

For more information, see the section "Odds Ratio Estimation" on page 4250.

WEIGHT Statement

WEIGHT *variable* < / *option* > **;**

When a WEIGHT statement appears, each observation in the input data set is weighted by the value of the WEIGHT variable. Unlike a FREQ variable, the values of the WEIGHT variable can be nonintegral and are not truncated. Observations with negative, zero, or missing values for the WEIGHT variable are not used in the model fitting. When the WEIGHT statement is not specified, each observation is assigned a weight of 1. The WEIGHT statement is not available with the STRATA statement. If you specify more than one WEIGHT statement, then the first WEIGHT variable is used.

If a SCORE statement is specified, then the WEIGHT variable is used for computing fit statistics and the ROC curve, but it is not required for scoring. If the DATA= data set in the SCORE statement does not contain the WEIGHT variable, the weights are assumed to be 1 and a warning message is issued in the SAS log. If you fit a model and perform the scoring in the same run, the same WEIGHT variable is used for fitting and scoring. If you fit a model in a previous run and input it with the INMODEL= option in the current run, then the WEIGHT variable can be different from the one used in the previous run; however, if a WEIGHT variable was not specified in the previous run, you can still specify a WEIGHT variable in the current run.

CAUTION: PROC LOGISTIC does not compute the proper variance estimators if you are analyzing survey data and specifying the sampling weights through the WEIGHT statement. The SURVEYLOGISTIC procedure is designed to perform the necessary, and correct, computations.

The following option can be added to the WEIGHT statement after a slash (/):

NORMALIZE

NORM

causes the weights specified by the WEIGHT variable to be normalized so that they add up to the actual sample size. Weights w_i are normalized by multiplying them by $\frac{n}{\sum_{i=1}^{n} w_i}$, where n is the sample size. With this option, the estimated covariance matrix of the parameter estimators is invariant to the scale of the WEIGHT variable.

Details: LOGISTIC Procedure

Missing Values

Any observation with missing values for the response, offset, strata, or explanatory variables is excluded from the analysis; however, missing values are valid for variables specified with the MISSING option in the CLASS or STRATA statement. Observations with a nonpositive or missing weight or with a frequency less than 1 are also excluded. The estimated linear predictor and its standard error estimate, the fitted probabilities and confidence limits, and the regression diagnostic statistics are not computed for any observation with missing offset or explanatory variable values. However, if only the response value is missing, the linear predictor, its standard error, the fitted individual and cumulative probabilities, and confidence limits for the cumulative probabilities can be computed and output to a data set by using the OUTPUT statement.

Response Level Ordering

Response level ordering is important because, by default, PROC LOGISTIC models the probability of response levels with *lower Ordered Value*. Ordered Values are assigned to response levels in ascending sorted order (that is, the lowest response level is assigned Ordered Value 1, the next lowest is assigned Ordered Value 2, and so on) and are displayed in the "Response Profiles" table. If your response variable Y takes values in $\{1, \ldots, k+1\}$, then, by default, the functions modeled with the binary or cumulative model are

$$\text{logit}(\Pr(Y \le i|\mathbf{x})), \quad i = 1, \ldots, k$$

and for the generalized logit model the functions modeled are

$$\log\left(\frac{\Pr(Y = i|\mathbf{x})}{\Pr(Y = k+1|\mathbf{x})}\right), \quad i = 1, \ldots, k$$

where the highest Ordered Value $Y = k+1$ is the reference level. You can change which probabilities are modeled by specifying the EVENT=, REF=, DESCENDING, or ORDER= response variable options in the MODEL statement.

For binary response data with event and nonevent categories, if your event category has a higher Ordered Value, then by default the nonevent is modeled. Since the default response function modeled is

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

where π is the probability of the response level assigned Ordered Value 1, and since

$$\text{logit}(\pi) = -\text{logit}(1-\pi)$$

the effect of modeling the nonevent is to change the signs of α and $\boldsymbol{\beta}$ in the model for the event, $\text{logit}(\pi) = \alpha + \boldsymbol{\beta}'\mathbf{x}$.

For example, suppose the binary response variable Y takes the values 1 and 0 for event and nonevent, respectively, and Exposure is the explanatory variable. By default, PROC LOGISTIC assigns Ordered

Value 1 to response level Y=0, and Ordered Value 2 to response level Y=1. As a result, PROC LOGISTIC models the probability of the nonevent (Ordered Value=1) category, and your parameter estimates have the opposite sign from those in the model for the event. To model the event without using a DATA step to change the values of the variable Y, you can control the ordering of the response levels or select the event or reference level, as shown in the following list:

- 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. This option is most useful for generalized logit models where the EVENT= option cannot be used.

```
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. In the following example, Y=1 is assigned the formatted value 'event' and Y=0 is assigned the formatted value 'nonevent'. Since ORDER=FORMATTED by default, Ordered Value 1 is assigned to response level Y=1, so the procedure models the event.

```
proc format;
   value Disease 1='event' 0='nonevent';
run;
proc logistic;
   format Y Disease.;
   model Y=Exposure;
run;
```

Link Functions and the Corresponding Distributions

Four link functions are available in the LOGISTIC procedure. The logit function is the default. To specify a different link function, use the LINK= option in the MODEL statement. The link functions and the corresponding distributions are as follows:

- The logit function

  $$g(p) = \log(p/(1-p))$$

  is the inverse of the cumulative logistic distribution function, which is

  $$F(x) = 1/(1+\exp(-x)) = \exp(x)/(1+\exp(x))$$

- The probit (or normit) function

  $$g(p) = \Phi^{-1}(p)$$

  is the inverse of the cumulative standard normal distribution function, which is

  $$F(x) = \Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-z^2/2)dz$$

  Traditionally, the probit function contains the additive constant 5, but throughout PROC LOGISTIC, the terms probit and normit are used interchangeably.

- The complementary log-log function

  $$g(p) = \log(-\log(1-p))$$

  is the inverse of the cumulative extreme-value function (also called the Gompertz distribution), which is

  $$F(x) = 1 - \exp(-\exp(x))$$

- The generalized logit function extends the binary logit link to a vector of levels $(p_1, \ldots, p_{k+1})$ by contrasting each level with a fixed level

  $$g(p_i) = \log(p_i/p_{k+1}) \quad i = 1, \ldots, k$$

The variances of the normal, logistic, and extreme-value distributions are not the same. Their respective means and variances are shown in the following table:

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

Here γ is the Euler constant. In comparing parameter estimates from different link functions, you need to take into account the different scalings of the corresponding distributions and, for the complementary log-log function, 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 logit link function should be about $\pi/\sqrt{3}$ larger than the estimates from the probit link function.

Determining Observations for Likelihood Contributions

If you use events/trials MODEL statement syntax, split each observation into two observations. One has response value 1 with a frequency equal to the frequency of the original observation (which is 1 if the FREQ statement is not used) times the value of the *events* variable. The other observation has response value 2 and a frequency equal to the frequency of the original observation times the value of (*trials–events*). These two observations will have the same explanatory variable values and the same FREQ and WEIGHT values as the original observation.

For either single-trial or events/trials syntax, let j index all observations. In other words, for single-trial syntax, j indexes the actual observations. And, for events/trials syntax, j indexes the observations after splitting (as described in the preceding paragraph). If your data set has 30 observations and you use single-trial syntax, j has values from 1 to 30; if you use events/trials syntax, j has values from 1 to 60.

Suppose the response variable in a cumulative response model can take on the ordered values $1, \ldots, k, k+1$, where k is an integer ≥ 1. The likelihood for the jth observation with ordered response value y_j and explanatory variables vector $\mathbf{x}_j$ is given by

$$L_j = \begin{cases} F(\alpha_1 + \boldsymbol{\beta}'\mathbf{x}_j) & y_j = 1 \\ F(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_j) - F(\alpha_{i-1} + \boldsymbol{\beta}'\mathbf{x}_j) & 1 < y_j = i \leq k \\ 1 - F(\alpha_k + \boldsymbol{\beta}'\mathbf{x}_j) & y_j = k+1 \end{cases}$$

where $F(\cdot)$ is the logistic, normal, or extreme-value distribution function, $\alpha_1, \ldots, \alpha_k$ are ordered intercept parameters, and $\boldsymbol{\beta}$ is the common slope parameter vector.

For the generalized logit model, letting the $k+1$st level be the reference level, the intercepts $\alpha_1, \ldots, \alpha_k$ are unordered and the slope vector $\boldsymbol{\beta}_i$ varies with each logit. The likelihood for the jth observation with response value y_j and explanatory variables vector $\mathbf{x}_j$ is given by

$$L_j = \Pr(Y = y_j | \mathbf{x}_j) = \begin{cases} \dfrac{e^{\alpha_i + \mathbf{x}_j'\boldsymbol{\beta}_i}}{1 + \sum_{m=1}^{k} e^{\alpha_m + \mathbf{x}_j'\boldsymbol{\beta}_m}} & 1 \leq y_j = i \leq k \\ \dfrac{1}{1 + \sum_{m=1}^{k} e^{\alpha_m + \mathbf{x}_j'\boldsymbol{\beta}_m}} & y_j = k+1 \end{cases}$$

Iterative Algorithms for Model Fitting

Two iterative maximum likelihood algorithms are available in PROC LOGISTIC for fitting an unconditional logistic regression, and these two methods are discussed in this section. For conditional logistic regression and models with the UNEQUALSLOPES specification, see the section "NLOPTIONS Statement" on page 4222 for details about available optimization techniques. Exact logistic regression uses a special algorithm described in the section "Exact Conditional Logistic Regression" on page 4274.

The default maximum likelihood algorithm is the Fisher scoring method, which is equivalent to fitting by iteratively reweighted least squares. The alternative algorithm is the Newton-Raphson method. Both algorithms give the same parameter estimates; however, the estimated covariance matrix of the parameter estimators can differ slightly. This is due to the fact that Fisher scoring is based on the expected information matrix while the Newton-Raphson method is based on the observed information matrix. In the case of a binary logit model, the observed and expected information matrices are identical, resulting in identical estimated covariance matrices for both algorithms. You can specify the TECHNIQUE= option to select a fitting algorithm, and specify the FIRTH option to perform a bias-reducing penalized maximum likelihood fit. Note for generalized logit models that only the Newton-Raphson technique is available.

Iteratively Reweighted Least Squares Algorithm (Fisher Scoring)

Consider the multinomial variable $\mathbf{Z}_j = (Z_{1j}, \ldots, Z_{k+1,j})'$ such that

$$Z_{ij} = \begin{cases} 1 & \text{if } Y_j = i \\ 0 & \text{otherwise} \end{cases}$$

With π_{ij} denoting the probability that the jth observation has response value i, the expected value of $\mathbf{Z}_j$ is $\boldsymbol{\pi}_j = (\pi_{1j}, \ldots, \pi_{k+1,j})'$ where $\pi_{k+1,j} = 1 - \sum_{i=1}^{k} \pi_{ij}$. The covariance matrix of $\mathbf{Z}_j$ is $\mathbf{V}_j$, which is the covariance matrix of a multinomial random variable for one trial with parameter vector $\boldsymbol{\pi}_j$. Let $\boldsymbol{\beta}$ be the vector of regression parameters; in other words, $\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_s)'$. Let $\mathbf{D}_j$ be the matrix of partial derivatives of $\boldsymbol{\pi}_j$ with respect to $\boldsymbol{\beta}$. The estimating equation for the regression parameters is

$$\sum_j \mathbf{D}_j' \mathbf{W}_j (\mathbf{Z}_j - \boldsymbol{\pi}_j) = \mathbf{0}$$

where $\mathbf{W}_j = w_j f_j \mathbf{V}_j^-$, w_j and f_j are the weight and frequency of the jth observation, and $\mathbf{V}_j^-$ is a generalized inverse of $\mathbf{V}_j$. PROC LOGISTIC chooses $\mathbf{V}_j^-$ as the inverse of the diagonal matrix with $\boldsymbol{\pi}_j$ as the diagonal.

With a starting value of $\boldsymbol{\beta}^{(0)}$, the maximum likelihood estimate of $\boldsymbol{\beta}$ is obtained iteratively as

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + (\sum_j \mathbf{D}_j' \mathbf{W}_j \mathbf{D}_j)^{-1} \sum_j \mathbf{D}_j' \mathbf{W}_j (\mathbf{Z}_j - \boldsymbol{\pi}_j)$$

where $\mathbf{D}_j$, $\mathbf{W}_j$, and $\boldsymbol{\pi}_j$ are evaluated at $\boldsymbol{\beta}^{(m)}$. The expression after the plus sign is the step size. If the likelihood evaluated at $\boldsymbol{\beta}^{(m+1)}$ is less than that evaluated at $\boldsymbol{\beta}^{(m)}$, then $\boldsymbol{\beta}^{(m+1)}$ is recomputed by step-halving or ridging as determined by the value of the RIDGING= option. The iterative scheme continues until convergence is obtained—that is, until $\boldsymbol{\beta}^{(m+1)}$ is sufficiently close to $\boldsymbol{\beta}^{(m)}$. Then the maximum likelihood estimate of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(m+1)}$.

The covariance matrix of $\widehat{\boldsymbol{\beta}}$ is estimated by

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}) = (\sum_j \widehat{\mathbf{D}}_j' \widehat{\mathbf{W}}_j \widehat{\mathbf{D}}_j)^{-1} = \widehat{\mathbf{I}}^{-1}$$

where $\widehat{\mathbf{D}}_j$ and $\widehat{\mathbf{W}}_j$ are, respectively, $\mathbf{D}_j$ and $\mathbf{W}_j$ evaluated at $\widehat{\boldsymbol{\beta}}$. $\widehat{\mathbf{I}}$ is the information matrix, or the negative expected Hessian matrix, evaluated at $\widehat{\boldsymbol{\beta}}$.

By default, starting values are zero for the slope parameters, and for the intercept parameters, starting values are the observed cumulative logits (that is, logits of the observed cumulative proportions of response). Alternatively, the starting values can be specified with the INEST= option.

Newton-Raphson Algorithm

For cumulative models, let the parameter vector be $\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_s)'$, and for the generalized logit model let $\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_k, \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_k')'$. The gradient vector and the Hessian matrix are given, respectively, by

$$\mathbf{g} = \sum_j w_j f_j \frac{\partial l_j}{\partial \boldsymbol{\beta}}$$

$$\mathbf{H} = \sum_j w_j f_j \frac{\partial^2 l_j}{\partial \boldsymbol{\beta}^2}$$

where $l_j = \log L_j$ is the log likelihood for the jth observation. With a starting value of $\boldsymbol{\beta}^{(0)}$, the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained iteratively until convergence is obtained:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} - \mathbf{H}^{-1}\mathbf{g}$$

where $\mathbf{H}$ and $\mathbf{g}$ are evaluated at $\boldsymbol{\beta}^{(m)}$. If the likelihood evaluated at $\boldsymbol{\beta}^{(m+1)}$ is less than that evaluated at $\boldsymbol{\beta}^{(m)}$, then $\boldsymbol{\beta}^{(m+1)}$ is recomputed by step-halving or ridging.

The covariance matrix of $\widehat{\boldsymbol{\beta}}$ is estimated by

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}) = \widehat{\mathbf{I}}^{-1}$$

where the observed information matrix $\widehat{\mathbf{I}} = -\widehat{\mathbf{H}}$ is computed by evaluating $\mathbf{H}$ at $\widehat{\boldsymbol{\beta}}$.

Firth's Bias-Reducing Penalized Likelihood

Firth's method is currently available only for binary logistic models. It replaces the usual score (gradient) equation

$$g(\beta_j) = \sum_{i=1}^{n}(y_i - \pi_i)x_{ij} = 0 \quad (j = 1, \ldots, p)$$

where p is the number of parameters in the model, with the modified score equation

$$g(\beta_j)^* = \sum_{i=1}^{n}\{y_i - \pi_i + h_i(0.5 - \pi_i)\}x_{ij} = 0 \quad (j = 1, \ldots, p)$$

where the h_is are the ith diagonal elements of the hat matrix $\mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}$ and $\mathbf{W} = \mathrm{diag}\{\pi_i(1 - \pi_i)\}$. The Hessian matrix is not modified by this penalty, and the optimization method is performed in the usual manner.

Convergence Criteria

Four convergence criteria are available: ABSFCONV=, FCONV=, GCONV=, and XCONV=. If you specify more than one convergence criterion, the optimization is terminated as soon as one of the criteria is satisfied. If none of the criteria is specified, the default is GCONV=1E–8.

If you specify a STRATA statement or the UNEQUALSLOPES option in the MODEL statement, all unspecified (or nondefault) criteria are also compared to zero. For example, specifying only the criterion XCONV=1E–8 but attaining FCONV=0 terminates the optimization even if the XCONV= criterion is not satisfied, because the log likelihood has reached its maximum. More convergence criteria are also available; see the section "NLOPTIONS Statement" on page 4222 for details.

Existence of Maximum Likelihood Estimates

The likelihood equation for a logistic regression model does not always have a finite solution. Sometimes there is a nonunique maximum on the boundary of the parameter space, at infinity. The existence, finiteness, and uniqueness of maximum likelihood estimates for the logistic regression model depend on the patterns of data points in the observation space (Albert and Anderson 1984; Santner and Duffy 1986). Existence checks are not performed for conditional logistic regression.

Consider a binary response model. Let Y_j be the response of the jth subject, and let $\mathbf{x}_j$ be the vector of explanatory variables (including the constant 1 associated with the intercept). There are three mutually exclusive and exhaustive types of data configurations: complete separation, quasi-complete separation, and overlap.

Complete Separation There is a complete separation of data points if there exists a vector $\mathbf{b}$ that correctly allocates all observations to their response groups; that is,

$$\begin{cases} \mathbf{b}'\mathbf{x}_j > 0 & Y_j = 1 \\ \mathbf{b}'\mathbf{x}_j < 0 & Y_j = 2 \end{cases}$$

This configuration gives nonunique infinite estimates. If the iterative process of maximizing the likelihood function is allowed to continue, the log likelihood diminishes to zero, and the dispersion matrix becomes unbounded.

Quasi-complete Separation The data are not completely separable, but there is a vector $\mathbf{b}$ such that

$$\begin{cases} \mathbf{b}'\mathbf{x}_j \geq 0 & Y_j = 1 \\ \mathbf{b}'\mathbf{x}_j \leq 0 & Y_j = 2 \end{cases}$$

and equality holds for at least one subject in each response group. This configuration also yields nonunique infinite estimates. If the iterative process of maximizing the likelihood function is allowed to continue, the dispersion matrix becomes unbounded and the log likelihood diminishes to a nonzero constant.

Overlap If neither complete nor quasi-complete separation exists in the sample points, there is an overlap of sample points. In this configuration, the maximum likelihood estimates exist and are unique.

Complete separation and quasi-complete separation are problems typically encountered with small data sets. Although complete separation can occur with any type of data, quasi-complete separation is not likely with truly continuous explanatory variables.

The LOGISTIC procedure uses a simple empirical approach to recognize the data configurations that lead to infinite parameter estimates. The basis of this approach is that any convergence method of maximizing the log likelihood must yield a solution giving complete separation, if such a solution exists. In maximizing the log likelihood, there is no checking for complete or quasi-complete separation if convergence is attained in eight or fewer iterations. Subsequent to the eighth iteration, the probability of the observed response is computed for each observation. If the predicted response equals the observed response for every observation, there is a complete separation of data points and the iteration process is stopped. If the complete separation of data has not been determined and an observation is identified to have an extremely large probability ($\geq$0.95) of predicting the observed response, there are two possible situations. First, there is overlap in the data set, and the observation is an atypical observation of its own group. The iterative process, if allowed to continue, will stop when a maximum is reached. Second, there is quasi-complete separation in the data set, and the asymptotic dispersion matrix is unbounded. If any of the diagonal elements of the dispersion matrix for the standardized observations vectors (all explanatory variables standardized to zero mean and unit variance) exceeds 5000, quasi-complete separation is declared and the iterative process is stopped. If either complete separation or quasi-complete separation is detected, a warning message is displayed in the procedure output.

Checking for quasi-complete separation is less foolproof than checking for complete separation. The NOCHECK option in the MODEL statement turns off the process of checking for infinite parameter estimates. In cases of complete or quasi-complete separation, turning off the checking process typically results

in the procedure failing to converge. The presence of a WEIGHT statement also turns off the checking process.

To address the separation issue, you can change your model, specify the FIRTH option to use Firth's penalized likelihood method, or for small data sets specify an EXACT statement to perform an exact logistic regression.

Effect-Selection Methods

Five effect-selection methods are available by specifying the SELECTION= option in the MODEL statement. The simplest method (and the default) is SELECTION=NONE, for which PROC LOGISTIC fits the complete model as specified in the MODEL statement. The other four methods are FORWARD for forward selection, BACKWARD for backward elimination, STEPWISE for stepwise selection, and SCORE for best subsets selection. Intercept parameters are forced to stay in the model unless the NOINT option is specified.

When SELECTION=FORWARD, PROC LOGISTIC first estimates parameters for effects forced into the model. These effects are the intercepts and the first *n* explanatory effects in the MODEL statement, where *n* is the number specified by the START= or INCLUDE= option in the MODEL statement (*n* is zero by default). Next, the procedure computes the score chi-square statistic for each effect not in the model and examines the largest of these statistics. If it is significant at the SLENTRY= level, the corresponding effect is added to the model. Once an effect is entered in the model, it is never removed from the model. The process is repeated until none of the remaining effects meet the specified level for entry or until the STOP= value is reached.

When SELECTION=BACKWARD, parameters for the complete model as specified in the MODEL statement are estimated unless the START= option is specified. In that case, only the parameters for the intercepts and the first *n* explanatory effects in the MODEL statement are estimated, where *n* is the number specified by the START= option. Results of the Wald test for individual parameters are examined. The least significant effect that does not meet the SLSTAY= level for staying in the model is removed. Once an effect is removed from the model, it remains excluded. The process is repeated until no other effect in the model meets the specified level for removal or until the STOP= value is reached. Backward selection is often less successful than forward or stepwise selection because the full model fit in the first step is the model most likely to result in a complete or quasi-complete separation of response values as described in the section "Existence of Maximum Likelihood Estimates" on page 4242.

The SELECTION=STEPWISE option is similar to the SELECTION=FORWARD option except that effects already in the model do not necessarily remain. Effects are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further effect can be added to the model or if the current model is identical to a previously visited model.

For SELECTION=SCORE, PROC LOGISTIC uses the branch-and-bound algorithm of Furnival and Wilson (1974) to find a specified number of models with the highest likelihood score (chi-square) statistic for all possible model sizes, from 1, 2, 3 effect models, and so on, up to the single model containing all of the explanatory effects. The number of models displayed for each model size is controlled by the BEST= option. You can use the START= option to impose a minimum model size, and you can use the STOP= option to impose a maximum model size. For instance, with BEST=3, START=2, and STOP=5, the SCORE selection method displays the best three models (that is, the three models with the highest score chi-squares)

containing 2, 3, 4, and 5 effects. The SELECTION=SCORE option is not available for models with CLASS variables.

The options FAST, SEQUENTIAL, and STOPRES can alter the default criteria for entering or removing effects from the model when they are used with the FORWARD, BACKWARD, or STEPWISE selection method.

Model Fitting Information

For the jth observation, let $\hat{\pi}_j$ be the estimated probability of the observed response. The three criteria displayed by the LOGISTIC procedure are calculated as follows:

- –2 log likelihood:

$$-2 \text{ Log L} = -2 \sum_j \frac{w_j}{\sigma^2} f_j \log(\hat{\pi}_j)$$

  where w_j and f_j are the weight and frequency values of the jth observation, and σ^2 is the dispersion parameter, which equals 1 unless the SCALE= option is specified. For binary response models that use events/trials MODEL statement syntax, this is

$$-2 \text{ Log L} = -2 \sum_j \frac{w_j}{\sigma^2} f_j [\log \binom{n_j}{r_j} + r_j \log(\hat{\pi}_j) + (n_j - r_j) \log(1 - \hat{\pi}_j)]$$

  where r_j is the number of events, n_j is the number of trials, $\hat{\pi}_j$ is the estimated event probability, and the statistic is reported both with and without the constant term.

- Akaike's information criterion:

$$\text{AIC} = -2 \text{ Log L} + 2p$$

  where p is the number of parameters in the model. For cumulative response models, $p = k + s$, where k is the total number of response levels minus one and s is the number of explanatory effects. For the generalized logit model, $p = k(s + 1)$.

- Schwarz (Bayesian information) criterion:

$$\text{SC} = -2 \text{ Log L} + p \log(\sum_j f_j n_j)$$

  where p is the number of parameters in the model, n_j is the number of trials when events/trials syntax is specified, and $n_j = 1$ with single-trial syntax.

The AIC and SC statistics give two different ways of adjusting the –2 Log L statistic for the number of terms in the model and the number of observations used. These statistics can be used when comparing different models for the same data (for example, when you use the SELECTION=STEPWISE option in the MODEL statement). The models being compared do not have to be nested; lower values of the statistics indicate a more desirable model.

The difference in the –2 Log L statistics between the intercepts-only model and the specified model has a $p - k$ degree-of-freedom chi-square distribution under the null hypothesis that all the explanatory effects in the model are zero, where p is the number of parameters in the specified model and k is the number of intercepts. The likelihood ratio test in the "Testing Global Null Hypothesis: BETA=0" table displays this difference and the associated p-value for this statistic. The score and Wald tests in that table test the same hypothesis and are asymptotically equivalent; see the sections "Residual Chi-Square" on page 4247 and "Testing Linear Hypotheses about the Regression Coefficients" on page 4262 for details.

Generalized Coefficient of Determination

Cox and Snell (1989, pp. 208–209) propose the following generalization of the coefficient of determination to a more general linear model:

$$R^2 = 1 - \left\{ \frac{L(\mathbf{0})}{L(\widehat{\boldsymbol{\beta}})} \right\}^{\frac{2}{n}}$$

where $L(\mathbf{0})$ is the likelihood of the intercept-only model, $L(\widehat{\boldsymbol{\beta}})$ is the likelihood of the specified model, $n = \sum_j f_j n_j$ is the sample size, f_j is the frequency of the jth observation, and n_j is the number of trials when events/trials syntax is specified or $n_j = 1$ with single-trial syntax.

The quantity R^2 achieves a maximum of less than one for discrete models, where the maximum is given by

$$R^2_{\max} = 1 - \{L(\mathbf{0})\}^{\frac{2}{n}}$$

Nagelkerke (1991) proposes the following adjusted coefficient, which can achieve a maximum value of one:

$$\tilde{R}^2 = \frac{R^2}{R^2_{\max}}$$

Specifying the NORMALIZE option in the WEIGHT statement makes these coefficients invariant to the scale of the weights.

Like the AIC and SC statistics described in the section "Model Fitting Information" on page 4245, R^2 and $\tilde{R}^2$ are most useful for comparing competing models that are not necessarily nested—larger values indicate better models. More properties and interpretation of R^2 and $\tilde{R}^2$ are provided in Nagelkerke (1991). In the "Testing Global Null Hypothesis: BETA=0" table, R^2 is labeled as "RSquare" and $\tilde{R}^2$ is labeled as "Max-rescaled RSquare." Use the RSQUARE option to request R^2 and $\tilde{R}^2$.

Score Statistics and Tests

To understand the general form of the score statistics, let $\mathbf{g}(\boldsymbol{\beta})$ be the vector of first partial derivatives of the log likelihood with respect to the parameter vector $\boldsymbol{\beta}$, and let $\mathbf{H}(\boldsymbol{\beta})$ be the matrix of second partial derivatives of the log likelihood with respect to $\boldsymbol{\beta}$. That is, $\mathbf{g}(\boldsymbol{\beta})$ is the gradient vector, and $\mathbf{H}(\boldsymbol{\beta})$ is the Hessian matrix. Let $\mathbf{I}(\boldsymbol{\beta})$ be either $-\mathbf{H}(\boldsymbol{\beta})$ or the expected value of $-\mathbf{H}(\boldsymbol{\beta})$. Consider a null hypothesis H_0. Let $\widehat{\boldsymbol{\beta}}_{H_0}$ be the MLE of $\boldsymbol{\beta}$ under H_0. The chi-square score statistic for testing H_0 is defined by

$$\mathbf{g}'(\widehat{\boldsymbol{\beta}}_{H_0})\mathbf{I}^{-1}(\widehat{\boldsymbol{\beta}}_{H_0})\mathbf{g}(\widehat{\boldsymbol{\beta}}_{H_0})$$

and it has an asymptotic χ^2 distribution with r degrees of freedom under H_0, where r is the number of restrictions imposed on $\boldsymbol{\beta}$ by H_0.

Residual Chi-Square

When you use SELECTION=FORWARD, BACKWARD, or STEPWISE, the procedure calculates a residual chi-square score statistic and reports the statistic, its degrees of freedom, and the p-value. This section describes how the statistic is calculated.

Suppose there are s explanatory effects of interest. The full cumulative response model has a parameter vector

$$\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_s)'$$

where $\alpha_1, \ldots, \alpha_k$ are intercept parameters, and $\beta_1, \ldots, \beta_s$ are the common slope parameters for the s explanatory effects. The full generalized logit model has a parameter vector

$$\begin{aligned} \boldsymbol{\beta} &= (\alpha_1, \ldots, \alpha_k, \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_k')' \quad \text{with} \\ \boldsymbol{\beta}_i' &= (\beta_{i1}, \ldots, \beta_{is}), \quad i = 1, \ldots, k \end{aligned}$$

where β_{ij} is the slope parameter for the jth effect in the ith logit.

Consider the null hypothesis $H_0: \beta_{t+1} = \cdots = \beta_s = 0$, where $t < s$ for the cumulative response model, and $H_0: \beta_{i,t+1} = \cdots = \beta_{is} = 0, t < s, i = 1, \ldots, k$, for the generalized logit model. For the reduced model with t explanatory effects, let $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_k$ be the MLEs of the unknown intercept parameters, let $\widehat{\beta}_1, \ldots, \widehat{\beta}_t$ be the MLEs of the unknown slope parameters, and let $\widehat{\boldsymbol{\beta}}_{i(t)}' = (\widehat{\beta}_{i1}, \ldots, \widehat{\beta}_{it}), i = 1, \ldots, k$, be those for the generalized logit model. The residual chi-square is the chi-square score statistic testing the null hypothesis H_0; that is, the residual chi-square is

$$\mathrm{g}'(\widehat{\boldsymbol{\beta}}_{H_0})\mathbf{I}^{-1}(\widehat{\boldsymbol{\beta}}_{H_0})\mathrm{g}(\widehat{\boldsymbol{\beta}}_{H_0})$$

where for the cumulative response model $\widehat{\boldsymbol{\beta}}_{H_0} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_k, \widehat{\beta}_1, \ldots, \widehat{\beta}_t, 0, \ldots, 0)'$, and for the generalized logit model $\widehat{\boldsymbol{\beta}}_{H_0} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_k, \widehat{\boldsymbol{\beta}}_{1(t)}', \mathbf{0}_{(s-t)}', \ldots \widehat{\boldsymbol{\beta}}_{k(t)}', \mathbf{0}_{(s-t)}')'$, where $\mathbf{0}_{(s-t)}$ denotes a vector of $s - t$ zeros.

The residual chi-square has an asymptotic chi-square distribution with $s - t$ degrees of freedom ($k(s - t)$ for the generalized logit model). A special case is the global score chi-square, where the reduced model consists of the k intercepts and no explanatory effects. The global score statistic is displayed in the "Testing Global Null Hypothesis: BETA=0" table. The table is not produced when the NOFIT option is used, but the global score statistic is displayed.

Testing Individual Effects Not in the Model

These tests are performed when you specify SELECTION=FORWARD or STEPWISE, and are displayed when the DETAILS option is specified. In the displayed output, the tests are labeled "Score Chi-Square" in the "Analysis of Effects Not in the Model" table and in the "Summary of Stepwise (Forward) Selection" table. This section describes how the tests are calculated.

Suppose that k intercepts and t explanatory variables (say $v_1, \ldots, v_t$) have been fit to a model and that v_{t+1} is another explanatory variable of interest. Consider a full model with the k intercepts and $t + 1$ explanatory variables $(v_1, \ldots, v_t, v_{t+1})$ and a reduced model with v_{t+1} excluded. The significance of v_{t+1} adjusted for $v_1, \ldots, v_t$ can be determined by comparing the corresponding residual chi-square with a chi-square distribution with one degree of freedom (k degrees of freedom for the generalized logit model).

Testing the Parallel Lines Assumption

For an ordinal response, PROC LOGISTIC performs a test of the parallel lines assumption. In the displayed output, this test is labeled "Score Test for the Equal Slopes Assumption" when the LINK= option is NORMIT or CLOGLOG. When LINK=LOGIT, the test is labeled as "Score Test for the Proportional Odds Assumption" in the output. For small sample sizes, this test might be too liberal (Stokes, Davis, and Koch 2000, p. 249). This section describes the methods used to calculate the test.

For this test the number of response levels, $k+1$, is assumed to be strictly greater than 2. Let Y be the response variable taking values $1,\ldots,k,k+1$. Suppose there are s explanatory variables. Consider the general cumulative model without making the parallel lines assumption

$$g(\Pr(Y \le i \mid \mathbf{x})) = (1,\mathbf{x}')\boldsymbol{\beta}_i, \quad 1 \le i \le k$$

where $g(\cdot)$ is the link function, and $\boldsymbol{\beta}_i = (\alpha_i,\beta_{i1},\ldots,\beta_{is})'$ is a vector of unknown parameters consisting of an intercept α_i and s slope parameters $\beta_{i1},\ldots,\beta_{is}$. The parameter vector for this general cumulative model is

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_k')'$$

Under the null hypothesis of parallelism H_0: $\beta_{1m} = \beta_{2m} = \cdots = \beta_{km}, 1 \le m \le s$, there is a single common slope parameter for each of the s explanatory variables. Let $\beta_1,\ldots,\beta_s$ be the common slope parameters. Let $\widehat{\alpha}_1,\ldots,\widehat{\alpha}_k$ and $\widehat{\beta}_1,\ldots,\widehat{\beta}_s$ be the MLEs of the intercept parameters and the common slope parameters. Then, under H_0, the MLE of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{H_0} = (\widehat{\boldsymbol{\beta}}_1',\ldots,\widehat{\boldsymbol{\beta}}_k')' \quad \text{with} \quad \widehat{\boldsymbol{\beta}}_i = (\widehat{\alpha}_i,\widehat{\beta}_1,\ldots,\widehat{\beta}_s)' \quad 1 \le i \le k$$

and the chi-square score statistic $\mathrm{g}'(\widehat{\boldsymbol{\beta}}_{H_0})\mathbf{I}^{-1}(\widehat{\boldsymbol{\beta}}_{H_0})\mathrm{g}(\widehat{\boldsymbol{\beta}}_{H_0})$ has an asymptotic chi-square distribution with $s(k-1)$ degrees of freedom. This tests the parallel lines assumption by testing the equality of separate slope parameters simultaneously for all explanatory variables.

Confidence Intervals for Parameters

There are two methods of computing confidence intervals for the regression parameters. One is based on the profile-likelihood function, and the other is based on the asymptotic normality of the parameter estimators. The latter is not as time-consuming as the former, since it does not involve an iterative scheme; however, it is not thought to be as accurate as the former, especially with small sample size. You use the CLPARM= option to request confidence intervals for the parameters.

Likelihood Ratio-Based Confidence Intervals

The likelihood ratio-based confidence interval is also known as the profile-likelihood confidence interval. The construction of this interval is derived from the asymptotic χ^2 distribution of the generalized likelihood ratio test (Venzon and Moolgavkar 1988). Suppose that the parameter vector is $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_s)'$ and you want to compute a confidence interval for β_j. The profile-likelihood function for $\beta_j = \gamma$ is defined as

$$l_j^*(\gamma) = \max_{\boldsymbol{\beta} \in \mathcal{B}_j(\gamma)} l(\boldsymbol{\beta})$$

where $\mathcal{B}_j(\gamma)$ is the set of all $\boldsymbol{\beta}$ with the jth element fixed at γ, and $l(\boldsymbol{\beta})$ is the log-likelihood function for $\boldsymbol{\beta}$. If $l_{\max} = l(\widehat{\boldsymbol{\beta}})$ is the log likelihood evaluated at the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$, then $2(l_{\max} - l_j^*(\beta_j))$ has a limiting chi-square distribution with one degree of freedom if β_j is the true parameter value. Let $l_0 = l_{\max} - 0.5\chi_1^2(1-\alpha)$, where $\chi_1^2(1-\alpha)$ is the $100(1-\alpha)$ percentile of the chi-square distribution with one degree of freedom. A $100(1-\alpha)\%$ confidence interval for β_j is

$$\{\gamma : l_j^*(\gamma) \geq l_0\}$$

The endpoints of the confidence interval are found by solving numerically for values of β_j that satisfy equality in the preceding relation. To obtain an iterative algorithm for computing the confidence limits, the log-likelihood function in a neighborhood of $\boldsymbol{\beta}$ is approximated by the quadratic function

$$\tilde{l}(\boldsymbol{\beta} + \boldsymbol{\delta}) = l(\boldsymbol{\beta}) + \boldsymbol{\delta}'\mathbf{g} + \frac{1}{2}\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta}$$

where $\mathbf{g} = \mathbf{g}(\boldsymbol{\beta})$ is the gradient vector and $\mathbf{V} = \mathbf{V}(\boldsymbol{\beta})$ is the Hessian matrix. The increment $\boldsymbol{\delta}$ for the next iteration is obtained by solving the likelihood equations

$$\frac{d}{d\boldsymbol{\delta}}\{\tilde{l}(\boldsymbol{\beta} + \boldsymbol{\delta}) + \lambda(\mathbf{e}_j'\boldsymbol{\delta} - \gamma)\} = \mathbf{0}$$

where λ is the Lagrange multiplier, $\mathbf{e}_j$ is the jth unit vector, and γ is an unknown constant. The solution is

$$\boldsymbol{\delta} = -\mathbf{V}^{-1}(\mathbf{g} + \lambda\mathbf{e}_j)$$

By substituting this $\boldsymbol{\delta}$ into the equation $\tilde{l}(\boldsymbol{\beta} + \boldsymbol{\delta}) = l_0$, you can estimate λ as

$$\lambda = \pm\left(\frac{2(l_0 - l(\boldsymbol{\beta}) + \frac{1}{2}\mathbf{g}'\mathbf{V}^{-1}\mathbf{g})}{\mathbf{e}_j'\mathbf{V}^{-1}\mathbf{e}_j}\right)^{\frac{1}{2}}$$

The upper confidence limit for β_j is computed by starting at the maximum likelihood estimate of $\boldsymbol{\beta}$ and iterating with positive values of λ until convergence is attained. The process is repeated for the lower confidence limit by using negative values of λ.

Convergence is controlled by the value ϵ specified with the PLCONV= option in the MODEL statement (the default value of ϵ is 1E–4). Convergence is declared on the current iteration if the following two conditions are satisfied:

$$|l(\boldsymbol{\beta}) - l_0| \leq \epsilon$$

and

$$(\mathbf{g} + \lambda\mathbf{e}_j)'\mathbf{V}^{-1}(\mathbf{g} + \lambda\mathbf{e}_j) \leq \epsilon$$

Wald Confidence Intervals

Wald confidence intervals are sometimes called the normal confidence intervals. They are based on the asymptotic normality of the parameter estimators. The $100(1-\alpha)\%$ Wald confidence interval for β_j is given by

$$\widehat{\beta}_j \pm z_{1-\alpha/2}\widehat{\sigma}_j$$

where z_p is the $100p$ percentile of the standard normal distribution, $\widehat{\beta}_j$ is the maximum likelihood estimate of β_j, and $\widehat{\sigma}_j$ is the standard error estimate of $\widehat{\beta}_j$.

Odds Ratio Estimation

Consider a dichotomous response variable with outcomes *event* and *nonevent*. Consider a dichotomous risk factor variable X that takes the value 1 if the risk factor is present and 0 if the risk factor is absent. According to the logistic model, the log odds function, $\text{logit}(X)$, is given by

$$\text{logit}(X) \equiv \log\left(\frac{\Pr(event \mid X)}{\Pr(nonevent \mid X)}\right) = \alpha + X\beta$$

The odds ratio ψ is defined as the ratio of the odds for those with the risk factor ($X = 1$) to the odds for those without the risk factor ($X = 0$). The log of the odds ratio is given by

$$\log(\psi) \equiv \log(\psi(X=1, X=0)) = \text{logit}(X=1) - \text{logit}(X=0) = (\alpha + 1\times\beta) - (\alpha + 0\times\beta) = \beta$$

In general, the odds ratio can be computed by exponentiating the difference of the logits between any two population profiles. This is the approach taken by the ODDSRATIO statement, so the computations are available regardless of parameterization, interactions, and nestings. However, as shown in the preceding equation for $\log(\psi)$, odds ratios of main effects can be computed as functions of the parameter estimates, and the remainder of this section is concerned with this methodology.

The parameter, β, associated with X represents the change in the log odds from $X = 0$ to $X = 1$. So the odds ratio is obtained by simply exponentiating the value of the parameter associated with the risk factor. The odds ratio indicates how the odds of the event change as you change X from 0 to 1. For instance, $\psi = 2$ means that the odds of an event when $X = 1$ are twice the odds of an event when $X = 0$. You can also express this as follows: the percent change in the odds of an event from $X = 0$ to $X = 1$ is $(\psi - 1)100\% = 100\%$.

Suppose the values of the dichotomous risk factor are coded as constants a and b instead of 0 and 1. The odds when $X = a$ become $\exp(\alpha + a\beta)$, and the odds when $X = b$ become $\exp(\alpha + b\beta)$. The odds ratio corresponding to an increase in X from a to b is

$$\psi = \exp[(b-a)\beta] = [\exp(\beta)]^{b-a} \equiv [\exp(\beta)]^c$$

Note that for any a and b such that $c = b - a = 1, \psi = \exp(\beta)$. So the odds ratio can be interpreted as the change in the odds for any increase of one unit in the corresponding risk factor. However, the change in odds for some amount other than one unit is often of greater interest. For example, a change of one pound in body weight might be too small to be considered important, while a change of 10 pounds might be more meaningful. The odds ratio for a change in X from a to b is estimated by raising the odds ratio estimate for a unit change in X to the power of $c = b - a$ as shown previously.

For a polytomous risk factor, the computation of odds ratios depends on how the risk factor is parameterized. For illustration, suppose that Race is a risk factor with four categories: White, Black, Hispanic, and Other.

For the effect parameterization scheme (PARAM=EFFECT) with White as the reference group (REF='White'), the design variables for Race are as follows:

	Design Variables		
Race	X_1	X_2	X_3
Black	1	0	0
Hispanic	0	1	0
Other	0	0	1
White	–1	–1	–1

The log odds for Black is

$$\begin{aligned} \text{logit(Black)} &= \alpha + (X_1 = 1)\beta_1 + (X_2 = 0)\beta_2 + (X_3 = 0)\beta_3 \\ &= \alpha + \beta_1 \end{aligned}$$

The log odds for White is

$$\begin{aligned} \text{logit(White)} &= \alpha + (X_1 = -1)\beta_1 + (X_2 = -1)\beta_2 + (X_3 = -1)\beta_3 \\ &= \alpha - \beta_1 - \beta_2 - \beta_3 \end{aligned}$$

Therefore, the log odds ratio of Black versus White becomes

$$\begin{aligned} \log(\psi(\text{Black, White})) &= \text{logit(Black)} - \text{logit(White)} \\ &= 2\beta_1 + \beta_2 + \beta_3 \end{aligned}$$

For the reference cell parameterization scheme (PARAM=REF) with White as the reference cell, the design variables for race are as follows:

	Design Variables		
Race	X_1	X_2	X_3
Black	1	0	0
Hispanic	0	1	0
Other	0	0	1
White	0	0	0

The log odds ratio of Black versus White is given by

$$\begin{aligned} \log(\psi(\text{Black, White})) &= \text{logit(Black)} - \text{logit(White)} \\ &= (\alpha + (X_1 = 1)\beta_1 + (X_2 = 0)\beta_2 + (X_3 = 0)\beta_3) - \\ &\quad (\alpha + (X_1 = 0)\beta_1 + (X_2 = 0)\beta_2 + (X_3 = 0)\beta_3) \\ &= \beta_1 \end{aligned}$$

For the GLM parameterization scheme (PARAM=GLM), the design variables are as follows:

	Design Variables			
Race	X_1	X_2	X_3	X_4
Black	1	0	0	0
Hispanic	0	1	0	0
Other	0	0	1	0
White	0	0	0	1

The log odds ratio of Black versus White is

$$\begin{aligned}
\log(\psi(\text{Black, White})) &= \text{logit(Black)} - \text{logit(White)} \\
&= (\alpha + (X_1 = 1)\beta_1 + (X_2 = 0)\beta_2 + (X_3 = 0)\beta_3 + (X_4 = 0)\beta_4) - \\
&\quad (\alpha + (X_1 = 0)\beta_1 + (X_2 = 0)\beta_2 + (X_3 = 0)\beta_3 + (X_4 = 1)\beta_4) \\
&= \beta_1 - \beta_4
\end{aligned}$$

Consider the hypothetical example of heart disease among race in Hosmer and Lemeshow (2000, p. 56). The entries in the following contingency table represent counts:

	Race			
Disease Status	White	Black	Hispanic	Other
Present	5	20	15	10
Absent	20	10	10	10

The computation of odds ratio of Black versus White for various parameterization schemes is tabulated in Table 54.11.

Table 54.11 Odds Ratio of Heart Disease Comparing Black to White

	Parameter Estimates				
PARAM=	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	**Odds Ratio Estimates**
EFFECT	0.7651	0.4774	0.0719		$\exp(2 \times 0.7651 + 0.4774 + 0.0719) = 8$
REF	2.0794	1.7917	1.3863		$\exp(2.0794) = 8$
GLM	2.0794	1.7917	1.3863	0.0000	$\exp(2.0794) = 8$

Since the log odds ratio ($\log(\psi)$) is a linear function of the parameters, the Wald confidence interval for $\log(\psi)$ can be derived from the parameter estimates and the estimated covariance matrix. Confidence intervals for the odds ratios are obtained by exponentiating the corresponding confidence limits for the log odd ratios. In the displayed output of PROC LOGISTIC, the "Odds Ratio Estimates" table contains the odds ratio estimates and the corresponding 95% Wald confidence intervals. For continuous explanatory variables, these odds ratios correspond to a unit increase in the risk factors.

To customize odds ratios for specific units of change for a continuous risk factor, you can use the UNITS statement to specify a list of relevant units for each explanatory variable in the model. Estimates of these customized odds ratios are given in a separate table. Let (L_j, U_j) be a confidence interval for $\log(\psi)$. The corresponding lower and upper confidence limits for the customized odds ratio $\exp(c\beta_j)$ are $\exp(cL_j)$ and $\exp(cU_j)$, respectively (for $c > 0$), or $\exp(cU_j)$ and $\exp(cL_j)$, respectively (for $c < 0$). You use the CLODDS= option or ODDSRATIO statement to request the confidence intervals for the odds ratios.

For a generalized logit model, odds ratios are computed similarly, except k odds ratios are computed for each effect, corresponding to the k logits in the model.

Rank Correlation of Observed Responses and Predicted Probabilities

The predicted mean score of an observation is the sum of the Ordered Values (shown in the "Response Profile" table) minus one, weighted by the corresponding predicted probabilities for that observation; that is, the predicted means score $= \sum_{i=1}^{k+1}(i-1)\widehat{\pi}_i$, where $k+1$ is the number of response levels and $\widehat{\pi}_i$ is the predicted probability of the ith (ordered) response.

A pair of observations with different observed responses is said to be *concordant* if the observation with the lower ordered response value has a lower predicted mean score than the observation with the higher ordered response value. If the observation with the lower ordered response value has a higher predicted mean score than the observation with the higher ordered response value, then the pair is *discordant*. If the pair is neither concordant nor discordant, it is a *tie*. Enumeration of the total numbers of concordant and discordant pairs is carried out by categorizing the predicted mean score into intervals of length $k/500$ and accumulating the corresponding frequencies of observations. Note that the length of these intervals can be modified by specification of the BINWIDTH= option in the MODEL statement.

Let N be the sum of observation frequencies in the data. Suppose there are a total of t pairs with different responses: n_c of them are concordant, n_d of them are discordant, and $t - n_c - n_d$ of them are tied. PROC LOGISTIC computes the following four indices of rank correlation for assessing the predictive ability of a model:

c	$=$	$(n_c + 0.5(t - n_c - n_d))/t$
Somers' D (Gini coefficient)	$=$	$(n_c - n_d)/t$
Goodman-Kruskal Gamma	$=$	$(n_c - n_d)/(n_c + n_d)$
Kendall's Tau-a	$=$	$(n_c - n_d)/(0.5N(N-1))$

If there are no ties, then Somers' D (Gini's coefficient) $= 2c - 1$. Note that the concordance index, c, also gives an estimate of the area under the receiver operating characteristic (ROC) curve when the response is binary (Hanley and McNeil 1982). See the section "ROC Computations" on page 4261 for more information about this area.

For binary responses, the predicted mean score is equal to the predicted probability for Ordered Value 2. As such, the preceding definition of concordance is consistent with the definition used in previous releases for the binary response model.

These statistics are not available when the STRATA statement is specified.

Linear Predictor, Predicted Probability, and Confidence Limits

This section describes how predicted probabilities and confidence limits are calculated by using the maximum likelihood estimates (MLEs) obtained from PROC LOGISTIC. For a specific example, see the section "Getting Started: LOGISTIC Procedure" on page 4166. Predicted probabilities and confidence limits can be output to a data set with the OUTPUT statement.

Binary and Cumulative Response Models

For a vector of explanatory variables $\mathbf{x}$, the linear predictor

$$\eta_i = g(\Pr(Y \le i \mid \mathbf{x})) = \alpha_i + \mathbf{x}'\boldsymbol{\beta} \quad 1 \le i \le k$$

is estimated by

$$\hat{\eta}_i = \widehat{\alpha}_i + \mathbf{x}'\widehat{\boldsymbol{\beta}}$$

where $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}$ are the MLEs of α_i and $\boldsymbol{\beta}$. The estimated standard error of η_i is $\hat{\sigma}(\hat{\eta}_i)$, which can be computed as the square root of the quadratic form $(1, \mathbf{x}')\widehat{\mathbf{V}}_{\mathbf{b}}(1, \mathbf{x}')'$, where $\widehat{\mathbf{V}}_{\mathbf{b}}$ is the estimated covariance matrix of the parameter estimates. The asymptotic $100(1-\alpha)\%$ confidence interval for η_i is given by

$$\hat{\eta}_i \pm z_{\alpha/2}\hat{\sigma}(\hat{\eta}_i)$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile point of a standard normal distribution.

The predicted probability and the $100(1-\alpha)\%$ confidence limits for $\pi_i = \Pr(Y \le i \mid \mathbf{x})$ are obtained by back-transforming the corresponding measures for the linear predictor, as shown in the following table:

Link	Predicted Probability	100(1–α)% Confidence Limits
LOGIT	$1/(1+\exp(-\hat{\eta}_i))$	$1/(1+\exp(-\hat{\eta}_i \pm z_{\alpha/2}\hat{\sigma}(\hat{\eta}_i)))$
PROBIT	$\Phi(\hat{\eta}_i)$	$\Phi(\hat{\eta}_i \pm z_{\alpha/2}\hat{\sigma}(\hat{\eta}_i))$
CLOGLOG	$1-\exp(-\exp(\hat{\eta}_i))$	$1-\exp(-\exp(\hat{\eta}_i \pm z_{\alpha/2}\hat{\sigma}(\hat{\eta}_i)))$

The CONTRAST statement also enables you to estimate the exponentiated contrast, $e^{\hat{\eta}_i}$. The corresponding standard error is $e^{\hat{\eta}_i}\hat{\sigma}(\hat{\eta}_i)$, and the confidence limits are computed by exponentiating those for the linear predictor: $\exp\{\hat{\eta}_i \pm z_{\alpha/2}\hat{\sigma}(\hat{\eta}_i)\}$.

Generalized Logit Model

For a vector of explanatory variables $\mathbf{x}$, define the linear predictors $\eta_i = \alpha_i + \mathbf{x}'\boldsymbol{\beta}_i$, and let π_i denote the probability of obtaining the response value i:

$$\pi_i = \begin{cases} \pi_{k+1}e^{\eta_i} & 1 \le i \le k \\ \dfrac{1}{1+\sum_{j=1}^{k} e^{\eta_j}} & i = k+1 \end{cases}$$

By the *delta method*,

$$\sigma^2(\pi_i) = \left(\frac{\partial \pi_i}{\partial \boldsymbol{\beta}}\right)' \mathbf{V}(\boldsymbol{\beta}) \frac{\partial \pi_i}{\partial \boldsymbol{\beta}}$$

A $100(1-\alpha)\%$ confidence level for π_i is given by

$$\widehat{\pi}_i \pm z_{\alpha/2}\hat{\sigma}(\widehat{\pi}_i)$$

where $\widehat{\pi}_i$ is the estimated expected probability of response i, and $\hat{\sigma}(\widehat{\pi}_i)$ is obtained by evaluating $\sigma(\pi_i)$ at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$.

Note that the contrast $\hat{\eta}_i$ and exponentiated contrast $e^{\hat{\eta}_i}$, their standard errors, and their confidence intervals are computed in the same fashion as for the cumulative response models, replacing $\boldsymbol{\beta}$ with $\boldsymbol{\beta}_i$.

Classification Table

For binary response data, the response is either an *event* or a *nonevent*. In PROC LOGISTIC, the response with Ordered Value 1 is regarded as the event, and the response with Ordered Value 2 is the nonevent. PROC LOGISTIC models the probability of the event. From the fitted model, a predicted event probability can be computed for each observation. A method to compute a reduced-bias estimate of the predicted probability is given in the section "Predicted Probability of an Event for Classification" on page 4255. If the predicted event probability exceeds or equals some cutpoint value $z \in [0, 1]$, the observation is predicted to be an event observation; otherwise, it is predicted as a nonevent. A 2×2 frequency table can be obtained by cross-classifying the observed and predicted responses. The CTABLE option produces this table, and the PPROB= option selects one or more cutpoints. Each cutpoint generates a classification table. If the PEVENT= option is also specified, a classification table is produced for each combination of PEVENT= and PPROB= values.

The accuracy of the classification is measured by its *sensitivity* (the ability to predict an event correctly) and *specificity* (the ability to predict a nonevent correctly). *Sensitivity* is the proportion of event responses that were predicted to be events. *Specificity* is the proportion of nonevent responses that were predicted to be nonevents. PROC LOGISTIC also computes three other conditional probabilities: *false positive rate*, *false negative rate*, and *rate of correct classification*. The *false positive rate* is the proportion of predicted event responses that were observed as nonevents. The *false negative rate* is the proportion of predicted nonevent responses that were observed as events. Given prior probabilities specified with the PEVENT= option, these conditional probabilities can be computed as posterior probabilities by using Bayes' theorem.

Predicted Probability of an Event for Classification

When you classify a set of binary data, if the same observations used to fit the model are also used to estimate the classification error, the resulting error-count estimate is biased. One way of reducing the bias is to remove the binary observation to be classified from the data, reestimate the parameters of the model, and then classify the observation based on the new parameter estimates. However, it would be costly to fit the model by leaving out each observation one at a time. The LOGISTIC procedure provides a less expensive one-step approximation to the preceding parameter estimates. Let $\widehat{\boldsymbol{\beta}}$ be the MLE of the parameter vector $(\alpha, \beta_1, \ldots, \beta_s)'$ based on all observations. Let $\widehat{\boldsymbol{\beta}}_{(j)}$ denote the MLE computed without the jth observation. The one-step estimate of $\widehat{\boldsymbol{\beta}}_{(j)}$ is given by

$$\widehat{\boldsymbol{\beta}}^1_{(j)} = \widehat{\boldsymbol{\beta}} - \frac{w_j(y_j - \widehat{\pi}_j)}{1 - h_j}\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}) \begin{pmatrix} 1 \\ \mathbf{x}_j \end{pmatrix}$$

where

y_j is 1 for an observed event response and 0 otherwise

w_j is the weight of the observation

$\widehat{\pi}_j$ is the predicted event probability based on $\widehat{\boldsymbol{\beta}}$

h_j is the hat diagonal element (defined on page 4263) with $n_j = 1$ and $r_j = y_j$

$\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$ is the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$

False Positive, False Negative, and Correct Classification Rates Using Bayes' Theorem

Suppose n_1 of n individuals experience an event, such as a disease. Let this group be denoted by $\mathcal{C}_1$, and let the group of the remaining $n_2 = n - n_1$ individuals who do not have the disease be denoted by $\mathcal{C}_2$. The jth individual is classified as giving a positive response if the predicted probability of disease ($\widehat{\pi}^*_{(j)}$) is large. The probability $\widehat{\pi}^*_{(j)}$ is the reduced-bias estimate based on the one-step approximation given in the preceding section. For a given cutpoint z, the jth individual is predicted to give a positive response if $\widehat{\pi}^*_{(j)} \geq z$.

Let B denote the event that a subject has the disease, and let $\bar{B}$ denote the event of not having the disease. Let A denote the event that the subject responds positively, and let $\bar{A}$ denote the event of responding negatively. Results of the classification are represented by two conditional probabilities, $\Pr(A|B)$ and $\Pr(A|\bar{B})$, where $\Pr(A|B)$ is the sensitivity and $\Pr(A|\bar{B})$ is one minus the specificity.

These probabilities are given by

$$\Pr(A|B) = \frac{\sum_{j\in\mathcal{C}_1} I(\widehat{\pi}^*_{(j)} \geq z)}{n_1}$$
$$\Pr(A|\bar{B}) = \frac{\sum_{j\in\mathcal{C}_2} I(\widehat{\pi}^*_{(j)} \geq z)}{n_2}$$

where $I(\cdot)$ is the indicator function.

Bayes' theorem is used to compute several rates of the classification. For a given prior probability $\Pr(B)$ of the disease, the false positive rate P_{F+}, the false negative rate P_{F-}, and the correct classification rate P_C are given by Fleiss (1981, pp. 4–5) as follows:

$$\begin{aligned} P_{F+} = \Pr(\bar{B}|A) &= \frac{\Pr(A|\bar{B})[1-\Pr(B)]}{\Pr(A|\bar{B}) + \Pr(B)[\Pr(A|B) - \Pr(A|\bar{B})]} \\ P_{F-} = \Pr(B|\bar{A}) &= \frac{[1-\Pr(A|B)]\Pr(B)}{1 - \Pr(A|\bar{B}) - \Pr(B)[\Pr(A|B) - \Pr(A|\bar{B})]} \\ P_C = \Pr(B|A) + \Pr(\bar{B}|\bar{A}) &= \Pr(A|B)\Pr(B) + \Pr(\bar{A}|\bar{B})[1-\Pr(B)] \end{aligned}$$

The prior probability $\Pr(B)$ can be specified by the PEVENT= option. If the PEVENT= option is not specified, the sample proportion of diseased individuals is used; that is, $\Pr(B) = n_1/n$. In such a case, the false positive rate and the false negative rate reduce to

$$\begin{aligned} P_{F+} &= \frac{\sum_{j\in\mathcal{C}_2} I(\widehat{\pi}^*_{(j)} \geq z)}{\sum_{j\in\mathcal{C}_1} I(\widehat{\pi}^*_{(j)} \geq z) + \sum_{j\in\mathcal{C}_2} I(\widehat{\pi}^*_{(j)} \geq z)} \\ P_{F-} &= \frac{\sum_{j\in\mathcal{C}_1} I(\widehat{\pi}^*_{(j)} < z)}{\sum_{j\in\mathcal{C}_1} I(\widehat{\pi}^*_{(j)} < z) + \sum_{j\in\mathcal{C}_2} I(\widehat{\pi}^*_{(j)} < z)} \\ P_C &= \frac{\sum_{j\in\mathcal{C}_1} I(\widehat{\pi}^*_{(j)} \geq z) + \sum_{j\in\mathcal{C}_2} I(\widehat{\pi}^*_{(j)} < z)}{n} \end{aligned}$$

Note that for a stratified sampling situation in which n_1 and n_2 are chosen a priori, n_1/n is not a desirable estimate of $\Pr(B)$. For such situations, the PEVENT= option should be specified.

Overdispersion

For a correctly specified model, the Pearson chi-square statistic and the deviance, divided by their degrees of freedom, should be approximately equal to one. When their values are much larger than one, the assumption of binomial variability might not be valid and the data are said to exhibit overdispersion. Underdispersion, which results in the ratios being less than one, occurs less often in practice.

When fitting a model, there are several problems that can cause the goodness-of-fit statistics to exceed their degrees of freedom. Among these are such problems as outliers in the data, using the wrong link function, omitting important terms from the model, and needing to transform some predictors. These problems should be eliminated before proceeding to use the following methods to correct for overdispersion.

Rescaling the Covariance Matrix

One way of correcting overdispersion is to multiply the covariance matrix by a dispersion parameter. This method assumes that the sample sizes in each subpopulation are approximately equal. You can supply the value of the dispersion parameter directly, or you can estimate the dispersion parameter based on either the Pearson chi-square statistic or the deviance for the fitted model.

The Pearson chi-square statistic χ^2_P and the deviance χ^2_D are given by

$$\begin{aligned}
\chi^2_P &= \sum_{i=1}^{m}\sum_{j=1}^{k+1}\frac{(r_{ij}-n_i\widehat{\pi}_{ij})^2}{n_i\widehat{\pi}_{ij}} \\
\chi^2_D &= 2\sum_{i=1}^{m}\sum_{j=1}^{k+1} r_{ij}\log\left(\frac{r_{ij}}{n_i\widehat{\pi}_{ij}}\right)
\end{aligned}$$

where m is the number of subpopulation profiles, $k+1$ is the number of response levels, r_{ij} is the total weight (sum of the product of the frequencies and the weights) associated with jth level responses in the ith profile, $n_i = \sum_{j=1}^{k+1} r_{ij}$, and $\widehat{\pi}_{ij}$ is the fitted probability for the jth level at the ith profile. Each of these chi-square statistics has $mk-p$ degrees of freedom, where p is the number of parameters estimated. The dispersion parameter is estimated by

$$\widehat{\sigma^2} = \begin{cases} \chi^2_P/(mk-p) & \text{SCALE=PEARSON} \\ \chi^2_D/(mk-p) & \text{SCALE=DEVIANCE} \\ (constant)^2 & \text{SCALE=}constant \end{cases}$$

In order for the Pearson 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 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 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.

Williams' Method

Suppose that the data consist of n binomial observations. For the ith observation, let r_i/n_i be the observed proportion and let $\mathbf{x}_i$ be the associated vector of explanatory variables. Suppose that the response probability for the ith observation is a random variable P_i with mean and variance

$$E(P_i) = \pi_i \quad \text{and} \quad V(P_i) = \phi\pi_i(1-\pi_i)$$

where p_i is the probability of the event, and ϕ is a nonnegative but otherwise unknown scale parameter. Then the mean and variance of r_i are

$$E(r_i) = n_i\pi_i \quad \text{and} \quad V(r_i) = n_i\pi_i(1-\pi_i)[1+(n_i-1)\phi]$$

Williams (1982) estimates the unknown parameter ϕ by equating the value of Pearson's chi-square statistic for the full model to its approximate expected value. Suppose w_i^* is the weight associated with the ith observation. The Pearson chi-square statistic is given by

$$\chi^2 = \sum_{i=1}^{n} \frac{w_i^*(r_i - n_i\widehat{\pi}_i)^2}{n_i\widehat{\pi}_i(1-\widehat{\pi}_i)}$$

Let $g'(\cdot)$ be the first derivative of the link function $g(\cdot)$. The approximate expected value of χ^2 is

$$E_{\chi^2} = \sum_{i=1}^{n} w_i^*(1-w_i^* v_i d_i)[1+\phi(n_i-1)]$$

where $v_i = n_i/(\pi_i(1-\pi_i)[g'(\pi_i)]^2)$ and d_i is the variance of the linear predictor $\widehat{\alpha}_i + \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$. The scale parameter ϕ is estimated by the following iterative procedure.

At the start, let $w_i^* = 1$ and let π_i be approximated by r_i/n_i, $i = 1, 2, \ldots, n$. If you apply these weights and approximated probabilities to χ^2 and E_{χ^2} and then equate them, an initial estimate of ϕ is

$$\hat{\phi}_0 = \frac{\chi^2 - (n-p)}{\sum_i (n_i-1)(1-v_i d_i)}$$

where p is the total number of parameters. The initial estimates of the weights become $\hat{w}_{i0}^* = [1+(n_i-1)\hat{\phi}_0]^{-1}$. After a weighted fit of the model, the $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}$ are recalculated, and so is χ^2. Then a revised estimate of ϕ is given by

$$\hat{\phi}_1 = \frac{\chi^2 - \sum_i w_i^*(1-w_i^* v_i d_i)}{w_i^*(n_i-1)(1-w_i^* v_i d_i)}$$

The iterative procedure is repeated until χ^2 is very close to its degrees of freedom.

Once ϕ has been estimated by $\hat{\phi}$ under the full model, weights of $(1+(n_i-1)\hat{\phi})^{-1}$ can be used to fit models that have fewer terms than the full model. See Example 54.10 for an illustration.

NOTE: If the WEIGHT statement is specified with the NORMALIZE option, then the initial w_i^* values are set to the normalized weights, and the weights resulting from Williams' method will not add up to the actual sample size. However, the estimated covariance matrix of the parameter estimates remains invariant to the scale of the WEIGHT variable.

The Hosmer-Lemeshow Goodness-of-Fit Test

Sufficient replication within subpopulations is required to make the Pearson and deviance goodness-of-fit tests valid. When there are one or more continuous predictors in the model, the data are often too sparse to use these statistics. Hosmer and Lemeshow (2000) proposed a statistic that they show, through simulation, is distributed as chi-square when there is no replication in any of the subpopulations. This test is available only for binary response models.

First, the observations are sorted in increasing order of their estimated event probability. The event is the response level specified in the response variable option EVENT=, or the response level that is not specified in the REF= option, or, if neither of these options was specified, then the event is the response level identified in the "Response Profiles" table as "Ordered Value 1". The observations are then divided into approximately 10 groups according to the following scheme. Let N be the total number of subjects. Let M be the target number of subjects for each group given by

$$M = [0.1 \times N + 0.5]$$

where $[x]$ represents the integral value of x. If the single-trial syntax is used, blocks of subjects are formed of observations with identical values of the explanatory variables. Blocks of subjects are not divided when being placed into groups.

Suppose there are n_1 subjects in the first block and n_2 subjects in the second block. The first block of subjects is placed in the first group. Subjects in the second block are added to the first group if

$$n_1 < M \quad \text{and} \quad n_1 + [0.5 \times n_2] \leq M$$

Otherwise, they are placed in the second group. In general, suppose subjects of the $(j-1)$ block have been placed in the kth group. Let c be the total number of subjects currently in the kth group. Subjects for the jth block (containing n_j subjects) are also placed in the kth group if

$$c < M \quad \text{and} \quad c + [0.5 \times n_j] \leq M$$

Otherwise, the n_j subjects are put into the next group. In addition, if the number of subjects in the last group does not exceed $[0.05 \times N]$ (half the target group size), the last two groups are collapsed to form only one group.

Note that the number of groups, g, can be smaller than 10 if there are fewer than 10 patterns of explanatory variables. There must be at least three groups in order for the Hosmer-Lemeshow statistic to be computed.

The Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson chi-square statistic from the $2 \times g$ table of observed and expected frequencies, where g is the number of groups. The statistic is written

$$\chi^2_{HL} = \sum_{i=1}^{g} \frac{(O_i - N_i \bar{\pi}_i)^2}{N_i \bar{\pi}_i (1 - \bar{\pi}_i)}$$

where N_i is the total frequency of subjects in the ith group, O_i is the total frequency of event outcomes in the ith group, and $\bar{\pi}_i$ is the average estimated predicted probability of an event outcome for the ith group. (Note that the predicted probabilities are computed as shown in the section "Linear Predictor, Predicted Probability, and Confidence Limits" on page 4253 and are not the cross validated estimates discussed in the section "Classification Table" on page 4255.) The Hosmer-Lemeshow statistic is then compared to a chi-square distribution with $(g-n)$ degrees of freedom, where the value of n can be specified in the LACKFIT option in the MODEL statement. The default is $n = 2$. Large values of χ^2_{HL} (and small p-values) indicate a lack of fit of the model.

Receiver Operating Characteristic Curves

ROC curves are used to evaluate and compare the performance of diagnostic tests; they can also be used to evaluate model fit. An ROC curve is just a plot of the proportion of true positives (events predicted to be events) versus the proportion of false positives (nonevents predicted to be events).

In a sample of n individuals, suppose n_1 individuals are observed to have a certain condition or event. Let this group be denoted by $\mathcal{C}_1$, and let the group of the remaining $n_2 = n - n_1$ individuals who do not have the condition be denoted by $\mathcal{C}_2$. Risk factors are identified for the sample, and a logistic regression model is fitted to the data. For the jth individual, an estimated probability $\widehat{\pi}_j$ of the event of interest is calculated. Note that the $\widehat{\pi}_j$ are computed as shown in the section "Linear Predictor, Predicted Probability, and Confidence Limits" on page 4253 and are not the cross validated estimates discussed in the section "Classification Table" on page 4255.

Suppose the n individuals undergo a test for predicting the event and the test is based on the estimated probability of the event. Higher values of this estimated probability are assumed to be associated with the event. A receiver operating characteristic (ROC) curve can be constructed by varying the cutpoint that determines which estimated event probabilities are considered to predict the event. For each cutpoint z, the following measures can be output to a data set by specifying the OUTROC= option in the MODEL statement or the OUTROC= option in the SCORE statement:

$$
\begin{aligned}
\text{\_POS\_}(z) &= \sum_{i \in \mathcal{C}_1} I(\widehat{\pi}_i \geq z) \\
\text{\_NEG\_}(z) &= \sum_{i \in \mathcal{C}_2} I(\widehat{\pi}_i < z) \\
\text{\_FALPOS\_}(z) &= \sum_{i \in \mathcal{C}_2} I(\widehat{\pi}_i \geq z) \\
\text{\_FALNEG\_}(z) &= \sum_{i \in \mathcal{C}_1} I(\widehat{\pi}_i < z) \\
\text{\_SENSIT\_}(z) &= \frac{\text{\_POS\_}(z)}{n_1} \\
\text{\_1MSPEC\_}(z) &= \frac{\text{\_FALPOS\_}(z)}{n_2}
\end{aligned}
$$

where $I(\cdot)$ is the indicator function.

Note that _POS_(z) is the number of correctly predicted event responses, _NEG_(z) is the number of correctly predicted nonevent responses, _FALPOS_(z) is the number of falsely predicted event responses, _FALNEG_(z) is the number of falsely predicted nonevent responses, _SENSIT_(z) is the sensitivity of the test, and _1MSPEC_(z) is one minus the specificity of the test.

The ROC curve is a plot of sensitivity (_SENSIT_) against 1–specificity (_1MSPEC_). The plot can be produced by using the PLOTS option or by using the GPLOT or SGPLOT procedure with the OUTROC= data set. See Example 54.7 for an illustration. The area under the ROC curve, as determined by the trapezoidal rule, is estimated by the concordance index, c, in the "Association of Predicted Probabilities and Observed Responses" table.

Comparing ROC Curves

ROC curves can be created from each model fit in a selection routine, from the specified model in the MODEL statement, from specified models in ROC statements, or from input variables which act as $\widehat{\pi}$ in the preceding discussion. Association statistics are computed for these models, and the models are compared when the ROCCONTRAST statement is specified. The ROC comparisons are performed by using a contrast matrix to take differences of the areas under the empirical ROC curves (DeLong, DeLong, and Clarke-Pearson 1988). For example, if you have three curves and the second curve is the reference, the contrast used for the overall test is

$$\mathbf{L}_1 = \begin{pmatrix} \boldsymbol{l}_1' \\ \boldsymbol{l}_2' \end{pmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

and you can optionally estimate and test each row of this contrast, in order to test the difference between the reference curve and each of the other curves. If you do not want to use a reference curve, the global test optionally uses the following contrast:

$$\mathbf{L}_2 = \begin{pmatrix} \boldsymbol{l}_1' \\ \boldsymbol{l}_2' \end{pmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

You can also specify your own contrast matrix. Instead of estimating the rows of these contrasts, you can request that the difference between every pair of ROC curves be estimated and tested.

By default for the reference contrast, the specified or selected model is used as the reference unless the NOFIT option is specified in the MODEL statement, in which case the first ROC model is the reference.

In order to label the contrasts, a name is attached to every model. The name for the specified or selected model is the MODEL statement label, or "Model" if the MODEL label is not present. The ROC statement models are named with their labels, or as "ROC*i*" for the *i*th ROC statement if a label is not specified. The contrast $\mathbf{L}_1$ is labeled as "Reference = ModelName", where ModelName is the reference model name, while $\mathbf{L}_2$ is labeled "Adjacent Pairwise Differences". The estimated rows of the contrast matrix are labeled "ModelName1 – ModelName2". In particular, for the rows of $\mathbf{L}_1$, ModelName2 is the reference model name. If you specify your own contrast matrix, then the contrast is labeled "Specified" and the *i*th contrast row estimates are labeled "Row*i*".

If ODS Graphics is enabled, then all ROC curves are displayed individually and are also overlaid in a final display. If a selection method is specified, then the curves produced in each step of the model selection process are overlaid onto a single plot and are labeled "Step*i*", and the selected model is displayed on a separate plot and on a plot with curves from specified ROC statements. See Example 54.8 for an example.

ROC Computations

The trapezoidal area under an empirical ROC curve is equal to the Mann-Whitney two-sample rank measure of association statistic (a generalized U-statistic) applied to two samples, $\{X_i\}, i = 1, \ldots, n_1$, in $\mathcal{C}_1$ and $\{Y_i\}, i = 1, \ldots, n_2$, in $\mathcal{C}_2$. PROC LOGISTIC uses the predicted probabilities in place of $\mathbf{X}$ and $\mathbf{Y}$; however, in general any criterion could be used. Denote the frequency of observation i in $\mathcal{C}_k$ as f_{ki}, and denote the total frequency in $\mathcal{C}_k$ as F_k. The WEIGHTED option replaces f_{ki} with $f_{ki}w_{ki}$, where w_{ki} is the weight of

observation i in group $\mathcal{C}_k$. The trapezoidal area under the curve is computed as

$$\begin{aligned} \hat{c} &= \frac{1}{F_1 F_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \psi(X_i, Y_j) f_{1i} f_{2j} \\ \psi(X, Y) &= \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases} \end{aligned}$$

so that $E(\hat{c}) = \Pr(Y < X) + \frac{1}{2}\Pr(Y = X)$. Note that the concordance index, c, in the "Association of Predicted Probabilities and Observed Responses" table does not use weights unless both the WEIGHTED and BINWIDTH=0 options are specified. Also, in this table, c is computed by creating 500 bins and binning the X_i and Y_j; this results in more ties than the preceding method (unless the BINWIDTH=0 or ROCEPS=0 option is specified), so c is not necessarily equal to $E(\hat{c})$.

To compare K empirical ROC curves, first compute the trapezoidal areas. Asymptotic normality of the estimated area follows from U-statistic theory, and a covariance matrix $\mathbf{S}$ can be computed; see DeLong, DeLong, and Clarke-Pearson (1988) for details. A Wald confidence interval for the rth area, $1 \leq r \leq K$, can be constructed as

$$\hat{c}_r \pm z_{1-\frac{\alpha}{2}} s_{r,r}$$

where $s_{r,r}$ is the rth diagonal of $\mathbf{S}$.

For a contrast of ROC curve areas, $\mathbf{Lc}$, the statistic

$$(\hat{\mathbf{c}} - \mathbf{c})'\mathbf{L}' \left[\mathbf{LSL}'\right]^{-1} \mathbf{L}(\hat{\mathbf{c}} - \mathbf{c})$$

has a chi-square distribution with df=rank($\mathbf{LSL}'$). For a row of the contrast, $\boldsymbol{l}'\mathbf{c}$,

$$\frac{\boldsymbol{l}'\hat{\mathbf{c}} - \boldsymbol{l}'\mathbf{c}}{[\boldsymbol{l}'\mathbf{S}\boldsymbol{l}]^{1/2}}$$

has a standard normal distribution. The corresponding confidence interval is

$$\boldsymbol{l}'\hat{\mathbf{c}} \pm z_{1-\frac{\alpha}{2}} \left[\boldsymbol{l}'\mathbf{S}\boldsymbol{l}\right]^{1/2}$$

Testing Linear Hypotheses about the Regression Coefficients

Linear hypotheses for $\boldsymbol{\beta}$ are expressed in matrix form as

$$H_0: \mathbf{L}\boldsymbol{\beta} = \mathbf{c}$$

where $\mathbf{L}$ is a matrix of coefficients for the linear hypotheses, and $\mathbf{c}$ is a vector of constants. The vector of regression coefficients $\boldsymbol{\beta}$ includes slope parameters as well as intercept parameters. The Wald chi-square statistic for testing H_0 is computed as

$$\chi^2_W = (\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{c})'[\mathbf{L}\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\mathbf{L}']^{-1}(\mathbf{L}\widehat{\boldsymbol{\beta}} - \mathbf{c})$$

where $\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$ is the estimated covariance matrix. Under H_0, χ^2_W has an asymptotic chi-square distribution with r degrees of freedom, where r is the rank of $\mathbf{L}$.

Regression Diagnostics

For binary response data, regression diagnostics developed by Pregibon (1981) can be requested by specifying the INFLUENCE option. For diagnostics available with conditional logistic regression, see the section "Regression Diagnostic Details" on page 4272. These diagnostics can also be obtained from the OUTPUT statement.

This section uses the following notation:

r_j, n_j	r_j is the number of event responses out of n_j trials for the jth observation. If events/trials syntax is used, r_j is the value of *events* and n_j is the value of *trials*. For single-trial syntax, $n_j = 1$, and $r_j = 1$ if the ordered response is 1, and $r_j = 0$ if the ordered response is 2.
w_j	is the weight of the jth observation.
π_j	is the probability of an event response for the jth observation given by $\pi_j = F(\alpha + \boldsymbol{\beta}'\mathbf{x}_j)$, where $F(\cdot)$ is the inverse link function defined on page 4238.
$\widehat{\boldsymbol{\beta}}$	is the maximum likelihood estimate (MLE) of $(\alpha, \beta_1, \ldots, \beta_s)'$.
$\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})$	is the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$.
$\hat{p}_j, \hat{q}_j$	$\hat{p}_j$ is the estimate of π_j evaluated at $\widehat{\boldsymbol{\beta}}$, and $\hat{q}_j = 1 - \hat{p}_j$.

Pregibon (1981) suggests using the index plots of several diagnostic statistics to identify influential observations and to quantify the effects on various aspects of the maximum likelihood fit. In an index plot, the diagnostic statistic is plotted against the observation number. In general, the distributions of these diagnostic statistics are not known, so cutoff values cannot be given for determining when the values are large. However, the IPLOTS and INFLUENCE options in the MODEL statement and the PLOTS option in the PROC LOGISTIC statement provide displays of the diagnostic values, allowing visual inspection and comparison of the values across observations. In these plots, if the model is correctly specified and fits all observations well, then no extreme points should appear.

The next five sections give formulas for these diagnostic statistics.

Hat Matrix Diagonal (Leverage)

The diagonal elements of the hat matrix are useful in detecting extreme points in the design space where they tend to have larger values. The jth diagonal element is

$$h_j = \begin{cases} \widetilde{w}_j (1, \mathbf{x}_j') \widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}) (1, \mathbf{x}_j')' & \text{Fisher scoring} \\ \widehat{w}_j (1, \mathbf{x}_j') \widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}) (1, \mathbf{x}_j')' & \text{Newton-Raphson} \end{cases}$$

where

$$\begin{aligned} \widetilde{w}_j &= \frac{w_j n_j}{\hat{p}_j \hat{q}_j [g'(\hat{p}_j)]^2} \\ \widehat{w}_j &= \widetilde{w}_j + \frac{w_j (r_j - n_j \hat{p}_j)[\hat{p}_j \hat{q}_j g''(\hat{p}_j) + (\hat{q}_j - \hat{p}_j) g'(\hat{p}_j)]}{(\hat{p}_j \hat{q}_j)^2 [g'(\hat{p}_j)]^3} \end{aligned}$$

and $g'(\cdot)$ and $g''(\cdot)$ are the first and second derivatives of the link function $g(\cdot)$, respectively.

For a binary response logit model, the hat matrix diagonal elements are

$$h_j = w_j n_j \hat{p}_j \hat{q}_j (1, \mathbf{x}_j') \widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}) \begin{pmatrix} 1 \\ \mathbf{x}_j \end{pmatrix}$$

If the estimated probability is extreme (less than 0.1 and greater than 0.9, approximately), then the hat diagonal might be greatly reduced in value. Consequently, when an observation has a very large or very small estimated probability, its hat diagonal value is not a good indicator of the observation's distance from the design space (Hosmer and Lemeshow 2000, p. 171).

Residuals

Residuals are useful in identifying observations that are not explained well by the model. Pearson residuals are components of the Pearson chi-square statistic and deviance residuals are components of the deviance. The Pearson residual for the *j*th observation is

$$\chi_j = \frac{\sqrt{w_j}(r_j - n_j \hat{p}_j)}{\sqrt{n_j \hat{p}_j \hat{q}_j}}$$

The Pearson chi-square statistic is the sum of squares of the Pearson residuals.

The deviance residual for the *j*th observation is

$$d_j = \begin{cases} -\sqrt{-2w_j n_j \log(\hat{q}_j)} & \text{if } r_j = 0 \\ \pm\sqrt{2w_j[r_j \log(\frac{r_j}{n_j \hat{p}_j}) + (n_j - r_j)\log(\frac{n_j - r_j}{n_j \hat{q}_j})]} & \text{if } 0 < r_j < n_j \\ \sqrt{-2w_j n_j \log(\hat{p}_j)} & \text{if } r_j = n_j \end{cases}$$

where the plus (minus) in $\pm$ is used if r_j/n_j is greater (less) than $\hat{p}_j$. The deviance is the sum of squares of the deviance residuals.

The STDRES option in the INFLUENCE and PLOTS=INFLUENCE options computes three more residuals (Collett 2003). The Pearson and deviance residuals are standardized to have approximately unit variance:

$$e_{p_j} = \frac{\chi_j}{\sqrt{1 - h_j}}$$

$$e_{d_j} = \frac{d_j}{\sqrt{1 - h_j}}$$

The likelihood residuals, which estimate components of a likelihood ratio test of deleting an individual observation, are a weighted combination of the standardized Pearson and deviance residuals

$$e_{l_j} = \text{sign}(r_j - n_j \hat{p}_j)\sqrt{h_j e_{p_j}^2 + (1 - h_j) e_{d_j}^2}$$

DFBETAS

For each parameter estimate, the procedure calculates a DFBETAS diagnostic for each observation. The DFBETAS diagnostic for an observation is the standardized difference in the parameter estimate due to deleting the observation, and it can be used to assess the effect of an individual observation on each estimated parameter of the fitted model. Instead of reestimating the parameter every time an observation is deleted, PROC LOGISTIC uses the one-step estimate. See the section "Predicted Probability of an Event for Classification" on page 4255. For the *j*th observation, the DFBETAS are given by

$$\text{DFBETAS}i_j = \boldsymbol{\Delta}_i \widehat{\boldsymbol{\beta}}^1_j / \hat{\sigma}_i$$

where $i = 0, 1, \ldots, s$, $\hat{\sigma}_i$ is the standard error of the *i*th component of $\widehat{\boldsymbol{\beta}}$, and $\boldsymbol{\Delta}_i \widehat{\boldsymbol{\beta}}^1_j$ is the *i*th component of the one-step difference

$$\boldsymbol{\Delta}\widehat{\boldsymbol{\beta}}^1_j = \frac{w_j(r_j - n_j\hat{p}_j)}{1-h_j}\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\begin{pmatrix} 1 \\ \mathbf{x}_j \end{pmatrix}$$

$\boldsymbol{\Delta}\widehat{\boldsymbol{\beta}}^1_j$ is the approximate change $(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^1_j)$ in the vector of parameter estimates due to the omission of the *j*th observation. The DFBETAS are useful in detecting observations that are causing instability in the selected coefficients.

C and CBAR

C and CBAR are confidence interval displacement diagnostics that provide scalar measures of the influence of individual observations on $\widehat{\boldsymbol{\beta}}$. These diagnostics are based on the same idea as the Cook distance in linear regression theory (Cook and Weisberg 1982), but use the one-step estimate. C and CBAR for the *j*th observation are computed as

$$C_j = \chi^2_j h_j/(1-h_j)^2$$

and

$$\overline{C}_j = \chi^2_j h_j/(1-h_j)$$

respectively.

Typically, to use these statistics, you plot them against an index and look for outliers.

DIFDEV and DIFCHISQ

DIFDEV and DIFCHISQ are diagnostics for detecting ill-fitted observations; in other words, observations that contribute heavily to the disagreement between the data and the predicted values of the fitted model. DIFDEV is the change in the deviance due to deleting an individual observation while DIFCHISQ is the change in the Pearson chi-square statistic for the same deletion. By using the one-step estimate, DIFDEV and DIFCHISQ for the *j*th observation are computed as

$$\text{DIFDEV} = d^2_j + \overline{C}_j$$

and

$$\text{DIFCHISQ} = \overline{C}_j/h_j$$

Scoring Data Sets

Scoring a data set, which is especially important for predictive modeling, means applying a previously fitted model to a new data set in order to compute the conditional, or *posterior*, probabilities of each response category given the values of the explanatory variables in each observation.

The SCORE statement enables you to score new data sets and output the scored values and, optionally, the corresponding confidence limits into a SAS data set. If the response variable is included in the new data set, then you can request fit statistics for the data, which is especially useful for test or validation data. If the response is binary, you can also create a SAS data set containing the *receiver operating characteristic* (ROC) curve. You can specify multiple SCORE statements in the same invocation of PROC LOGISTIC.

By default, the posterior probabilities are based on implicit prior probabilities that are proportional to the frequencies of the response categories in the *training data* (the data used to fit the model). Explicit prior probabilities should be specified with the PRIOR= or PRIOREVENT= option when the sample proportions of the response categories in the training data differ substantially from the operational data to be scored. For example, to detect a rare category, it is common practice to use a training set in which the rare categories are overrepresented; without prior probabilities that reflect the true incidence rate, the predicted posterior probabilities for the rare category will be too high. By specifying the correct priors, the posterior probabilities are adjusted appropriately.

The model fit to the DATA= data set in the PROC LOGISTIC statement is the default model used for the scoring. Alternatively, you can save a model fit in one run of PROC LOGISTIC and use it to score new data in a subsequent run. The OUTMODEL= option in the PROC LOGISTIC statement saves the model information in a SAS data set. Specifying this data set in the INMODEL= option of a new PROC LOGISTIC run will score the DATA= data set in the SCORE statement without refitting the model.

The STORE statement can also be used to save your model. The PLM procedure can use this model to score new data sets; see Chapter 69, "The PLM Procedure," for more information. You cannot specify priors in PROC PLM.

Fit Statistics for Scored Data Sets

Specifying the FITSTAT option displays the following fit statistics when the data set being scored includes the response variable:

Statistic	**Description**
Total frequency	$F = \sum_i f_i n_i$
Total weight	$W = \sum_i f_i w_i n_i$
Log likelihood	$\log L = \sum_i f_i w_i \log(\widehat{\pi}_i)$
Full log likelihood	$\log L_f = \text{constant} + \log L$
Misclassification (error) rate	$\frac{\sum_i 1\{\text{F\_}Y_i \neq \text{I\_}Y_i\} f_i n_i}{F}$
AIC	$-2 \log L_f + 2p$
AICC	$-2 \log L_f + \frac{2pn}{n-p-1}$
BIC	$-2 \log L_f + p \log(n)$
SC	$-2 \log L_f + p \log(F)$
R-square	$R^2 = 1 - \left(\frac{L_0}{L}\right)^{2/F}$
Maximum-rescaled R-square	$\frac{R^2}{1 - L_0^{2/F}}$
AUC	Area under the ROC curve
Brier score (polytomous response)	$\frac{1}{W} \sum_i f_i w_i \sum_j (y_{ij} - \widehat{\pi}_{ij})^2$
Brier score (binary response)	$\frac{1}{W} \sum_i f_i w_i (r_i (1 - \widehat{\pi}_i)^2 + (n_i - r_i) \widehat{\pi}_i^2)$
Brier reliability (events/trials syntax)	$\frac{1}{W} \sum_i f_i w_i (r_i / n_i - \widehat{\pi}_i)^2$

In the preceding table, f_i is the frequency of the ith observation in the data set being scored, w_i is the weight of the observation, and $n = \sum_i f_i$. The number of trials when events/trials syntax is specified is n_i, and with single-trial syntax $n_i = 1$. The values F_Y_i and I_Y_i are described in the section "OUT= Output Data Set in a SCORE Statement" on page 4281. The indicator function $1\{A\}$ is 1 if A is true and 0 otherwise. The likelihood of the model is L, and L_0 denotes the likelihood of the intercept-only model. For polytomous response models, y_i is the observed polytomous response level, $\widehat{\pi}_{ij}$ is the predicted probability of the jth response level for observation i, and $y_{ij} = 1\{y_i = j\}$. For binary response models, $\widehat{\pi}_i$ is the predicted probability of the observation, r_i is the number of events when you specify events/trials syntax, and $r_i = y_i$ when you specify single-trial syntax.

The log likelihood, Akaike's information criterion (AIC), and Schwarz criterion (SC) are described in the section "Model Fitting Information" on page 4245. The full log likelihood is displayed for models specified with events/trials syntax, and the constant term is described in the section "Model Fitting Information" on page 4245. The AICC is a small-sample bias-corrected version of the AIC (Hurvich and Tsai 1993; Burnham and Anderson 1998). The Bayesian information criterion (BIC) is the same as the SC except when events/trials syntax is specified. The area under the ROC curve for binary response models is defined in the section "ROC Computations" on page 4261. The R-square and maximum-rescaled R-square statistics, defined in "Generalized Coefficient of Determination" on page 4246, are not computed when you specify both an OFFSET= variable and the INMODEL= data set. The Brier score (Brier 1950) is the weighted squared difference between the predicted probabilities and their observed response levels. For events/trials syntax, the Brier reliability is the weighted squared difference between the predicted probabilities and the observed proportions (Murphy 1973).

Posterior Probabilities and Confidence Limits

Let F be the inverse link function. That is,

$$F(t) = \begin{cases} \frac{1}{1+\exp(-t)} & \text{logistic} \\ \Phi(t) & \text{normal} \\ 1-\exp(-\exp(t)) & \text{complementary log-log} \end{cases}$$

The first derivative of F is given by

$$F'(t) = \begin{cases} \frac{\exp(-t)}{(1+\exp(-t))^2} & \text{logistic} \\ \phi(t) & \text{normal} \\ \exp(t)\exp(-\exp(t)) & \text{complementary log-log} \end{cases}$$

Suppose there are $k+1$ response categories. Let Y be the response variable with levels $1,\ldots,k+1$. Let $\mathbf{x} = (x_0, x_1, \ldots, x_s)'$ be a $(s+1)$-vector of covariates, with $x_0 \equiv 1$. Let $\boldsymbol{\beta}$ be the vector of intercept and slope regression parameters.

Posterior probabilities are given by

$$p(Y=i|\mathbf{x}) = \frac{p_o(Y=i|\mathbf{x})\frac{\widetilde{p}(Y=i)}{p_o(Y=i)}}{\sum_j p_o(Y=j|\mathbf{x})\frac{\widetilde{p}(Y=j)}{p_o(Y=j)}} \quad i = 1,\ldots,k+1$$

where the old posterior probabilities ($p_o(Y=i|\mathbf{x}), i=1,\ldots,k+1$) are the conditional probabilities of the response categories given $\mathbf{x}$, the old priors ($p_o(Y=i), i=1,\ldots,k+1$) are the sample proportions of response categories of the training data, and the new priors ($\widetilde{p}(Y=i), i=1,\ldots,k+1$) are specified in the PRIOR= or PRIOREVENT= option. To simplify notation, absorb the old priors into the new priors; that is

$$p(Y=i) = \frac{\widetilde{p}(Y=i)}{p_o(Y=i)} \quad i = 1,\ldots,k+1$$

Note if the PRIOR= and PRIOREVENT= options are not specified, then $p(Y=i)=1$.

The posterior probabilities are functions of $\boldsymbol{\beta}$ and their estimates are obtained by substituting $\boldsymbol{\beta}$ by its MLE $\widehat{\boldsymbol{\beta}}$. The variances of the estimated posterior probabilities are given by the *delta method* as follows:

$$\mathrm{Var}(\widehat{p}(Y=i|\mathbf{x})) = \left[\frac{\partial p(Y=i|\mathbf{x})}{\partial \boldsymbol{\beta}}\right]' \mathrm{Var}(\widehat{\boldsymbol{\beta}}) \left[\frac{\partial p(Y=i|\mathbf{x})}{\partial \boldsymbol{\beta}}\right]$$

where

$$\frac{\partial p(Y=i|\mathbf{x})}{\partial \boldsymbol{\beta}} = \frac{\frac{\partial p_o(Y=i|\mathbf{x})}{\partial \boldsymbol{\beta}} p(Y=i)}{\sum_j p_o(Y=j|\mathbf{x})p(Y=j)} - \frac{p_o(Y=i|\mathbf{x})p(Y=i)\sum_j \frac{\partial p_o(Y=j|\mathbf{x})}{\partial \boldsymbol{\beta}} p(Y=j)}{[\sum_j p_o(Y=j|\mathbf{x})p(Y=j)]^2}$$

and the old posterior probabilities $p_o(Y=i|\mathbf{x})$ are described in the following sections.

A $100(1-\alpha)\%$ confidence interval for $p(Y=i|\mathbf{x})$ is

$$\widehat{p}(Y=i|\mathbf{x}) \pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\widehat{p}(Y=i|\mathbf{x}))}$$

where z_τ is the upper 100τ percentile of the standard normal distribution.

Binary and Cumulative Response Models

Let $\alpha_1, \ldots, \alpha_k$ be the intercept parameters and let $\boldsymbol{\beta}_s$ be the vector of slope parameters. Denote $\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_k, \boldsymbol{\beta}_s')'$. Let

$$\eta_i = \eta_i(\boldsymbol{\beta}) = \alpha_i + \mathbf{x}'\boldsymbol{\beta}_s, i = 1, \ldots, k$$

Estimates of $\eta_1, \ldots, \eta_k$ are obtained by substituting the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$.

The predicted probabilities of the responses are

$$\widehat{p_o}(Y = i\,|\mathbf{x}) = \widehat{\Pr}(Y = i) = \begin{cases} F(\hat{\eta}_1) & i = 1 \\ F(\hat{\eta}_i) - F(\hat{\eta}_{i-1}) & i = 2, \ldots, k \\ 1 - F(\hat{\eta}_k) & i = k+1 \end{cases}$$

For $i = 1, \ldots, k$, let $\boldsymbol{\delta}_i(\mathbf{x})$ be a $(k+1)$ column vector with ith entry equal to 1, $k+1$ entry equal to $\mathbf{x}$, and all other entries 0. The derivative of $p_o(Y = i\,|\mathbf{x})$ with respect to $\boldsymbol{\beta}$ are

$$\frac{\partial p_o(Y = i\,|\mathbf{x})}{\partial \boldsymbol{\beta}} = \begin{cases} F'(\alpha_1 + \mathbf{x}'\boldsymbol{\beta}_s)\boldsymbol{\delta}_1(\mathbf{x}) & i = 1 \\ F'(\alpha_i + \mathbf{x}'\boldsymbol{\beta}_s)\boldsymbol{\delta}_i(\mathbf{x}) - F'(\alpha_{i-1} + \mathbf{x}'\boldsymbol{\beta}_s)\boldsymbol{\delta}_{i-1}(\mathbf{x}) & i = 2, \ldots, k \\ -F'(\alpha_k + \mathbf{x}'\boldsymbol{\beta}_s)\boldsymbol{\delta}_k(\mathbf{x}) & i = k+1 \end{cases}$$

The cumulative posterior probabilities are

$$p(Y \le i\,|\mathbf{x}) = \frac{\sum_{j=1}^{i} p_o(Y = j\,|\mathbf{x})p(Y = j)}{\sum_{j=1}^{k+1} p_o(Y = j\,|\mathbf{x})p(Y = j)} = \sum_{j=1}^{i} p(Y = j\,|\mathbf{x}) \quad i = 1, \ldots, k+1$$

Their derivatives are

$$\frac{\partial p(Y \le i\,|\mathbf{x})}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{i} \frac{\partial p(Y = j\,|\mathbf{x})}{\partial \boldsymbol{\beta}} \quad i = 1, \ldots, k+1$$

In the delta-method equation for the variance, replace $p(Y = \cdot|\mathbf{x})$ with $p(Y \le \cdot|\mathbf{x})$.

Finally, for the cumulative response model, use

$$\begin{aligned} \widehat{p_o}(Y \le i\,|\mathbf{x}) &= F(\hat{\eta}_i) \quad i = 1, \ldots, k \\ \widehat{p_o}(Y \le k+1|\mathbf{x}) &= 1 \\ \frac{\partial p_o(Y \le i\,|\mathbf{x})}{\partial \boldsymbol{\beta}} &= F'(\alpha_i + \mathbf{x}'\boldsymbol{\beta}_s)\delta_i(\mathbf{x}) \quad i = 1, \ldots, k \\ \frac{\partial p_o(Y \le k+1|\mathbf{x})}{\partial \boldsymbol{\beta}} &= 0 \end{aligned}$$

Generalized Logit Model

Consider the last response level (Y=k+1) as the reference. Let $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ be the (intercept and slope) parameter vectors for the first k logits, respectively. Denote $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_k')'$. Let $\eta = (\eta_1, \ldots, \eta_k)'$ with

$$\eta_i = \eta_i(\boldsymbol{\beta}) = \mathbf{x}'\boldsymbol{\beta}_i \quad i = 1, \ldots, k$$

Estimates of $\eta_1, \ldots, \eta_k$ are obtained by substituting the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$.

The predicted probabilities are

$$\begin{aligned} \widehat{p_o}(Y = k+1|\mathbf{x}) \equiv \Pr(Y = k+1|\mathbf{x}) &= \frac{1}{1 + \sum_{l=1}^{k} \exp(\hat{\eta}_l)} \\ \widehat{p_o}(Y = i|\mathbf{x}) \equiv \Pr(Y = i|\mathbf{x}) &= \widehat{p_o}(Y = k+1|\mathbf{x}) \exp(\eta_i), i = 1, \ldots, k \end{aligned}$$

The derivative of $p_o(Y = i|\mathbf{x})$ with respect to $\boldsymbol{\beta}$ are

$$\begin{aligned} \frac{\partial p_o(Y = i|\mathbf{x})}{\partial \boldsymbol{\beta}} &= \frac{\partial \eta}{\partial \boldsymbol{\beta}} \frac{\partial p_o(Y = i|\mathbf{x})}{\partial \eta} \\ &= (I_k \otimes \mathbf{x}) \left(\frac{\partial p_o(Y = i|\mathbf{x})}{\partial \eta_1}, \cdots, \frac{\partial p_o(Y = i|\mathbf{x})}{\partial \eta_k} \right)' \end{aligned}$$

where

$$\frac{\partial p_o(Y = i|\mathbf{x})}{\partial \eta_j} = \begin{cases} p_o(Y = i|\mathbf{x})(1 - p_o(Y = i|\mathbf{x})) & j = i \\ -p_o(Y = i|\mathbf{x}) p_o(Y = j|\mathbf{x}) & \text{otherwise} \end{cases}$$

Special Case of Binary Response Model with No Priors

Let $\boldsymbol{\beta}$ be the vector of regression parameters. Let

$$\eta = \eta(\boldsymbol{\beta}) = \mathbf{x}'\boldsymbol{\beta}$$

The variance of $\hat{\eta}$ is given by

$$\mathrm{Var}(\hat{\eta}) = \mathbf{x}' \mathrm{Var}(\widehat{\boldsymbol{\beta}}) \mathbf{x}$$

A $100(1-\alpha)$ percent confidence interval for η is

$$\hat{\eta} \pm z_{1-\alpha/2} \sqrt{\widehat{\mathrm{Var}}(\hat{\eta})}$$

Estimates of $p_o(Y = 1|\mathbf{x})$ and confidence intervals for the $p_o(Y = 1|\mathbf{x})$ are obtained by back-transforming $\hat{\eta}$ and the confidence intervals for η, respectively. That is,

$$\widehat{p_o}(Y = 1|\mathbf{x}) = F(\hat{\eta})$$

and the confidence intervals are

$$F\left(\hat{\eta} \pm z_{1-\alpha/2} \sqrt{\widehat{\mathrm{Var}}(\hat{\eta})} \right)$$

Conditional Logistic Regression

The method of maximum likelihood described in the preceding sections relies on large-sample asymptotic normality for the validity of estimates and especially of their standard errors. When you do not have a large sample size compared to the number of parameters, this approach might be inappropriate and might result in biased inferences. This situation typically arises when your data are stratified and you fit intercepts to each stratum so that the number of parameters is of the same order as the sample size. For example, in a 1:1 matched pairs study with n pairs and p covariates, you would estimate $n-1$ intercept parameters and p slope parameters. Taking the stratification into account by "conditioning out" (and not estimating) the stratum-specific intercepts gives consistent and asymptotically normal MLEs for the slope coefficients. See Breslow and Day (1980) and Stokes, Davis, and Koch (2012) for more information. If your nuisance parameters are not just stratum-specific intercepts, you can perform an exact conditional logistic regression.

Computational Details

For each stratum h, $h = 1,\ldots,H$, number the observations as $i = 1,\ldots,n_h$ so that hi indexes the ith observation in stratum h. Denote the p covariates for the hith observation as $\mathbf{x}_{hi}$ and its binary response as y_{hi}, and let $\mathbf{y} = (y_{11},\ldots,y_{1n_1},\ldots,y_{H1},\ldots,y_{Hn_H})'$, $\mathbf{X}_h = (\mathbf{x}_{h1}\ldots\mathbf{x}_{hn_h})'$, and $\mathbf{X} = (\mathbf{X}_1'\ldots\mathbf{X}_H')'$. Let the dummy variables $z_h, h = 1,\ldots,H$, be indicator functions for the strata ($z_h = 1$ if the observation is in stratum h), and denote $\mathbf{z}_{hi} = (z_1,\ldots,z_H)$ for the hith observation, $\mathbf{Z}_h = (\mathbf{z}_{h1}\ldots\mathbf{z}_{hn_h})'$, and $\mathbf{Z} = (\mathbf{Z}_1'\ldots\mathbf{Z}_H')'$. Denote $\mathbf{X}^* = (\mathbf{Z}|\mathbf{X})$ and $\mathbf{x}^*_{hi} = (\mathbf{z}'_{hi}|\mathbf{x}'_{hi})'$. Arrange the observations in each stratum h so that $y_{hi} = 1$ for $i = 1,\ldots,m_h$, and $y_{hi} = 0$ for $i = m_{h+1},\ldots,n_h$. Suppose all observations have unit frequency.

Consider the binary logistic regression model on page 4163 written as

$$\text{logit}(\boldsymbol{\pi}) = \mathbf{X}^*\boldsymbol{\theta}$$

where the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}',\boldsymbol{\beta}')'$ consists of $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_H)'$, α_h is the intercept for stratum $h, h = 1,\ldots,H$, and $\boldsymbol{\beta}$ is the parameter vector for the p covariates.

From the section "Determining Observations for Likelihood Contributions" on page 4239, you can write the likelihood contribution of observation $hi, i = 1,\ldots,n_h, h = 1,\ldots,H$, as

$$L_{hi}(\boldsymbol{\theta}) = \frac{e^{y_{hi}\mathbf{x}^{*\prime}_{hi}\boldsymbol{\theta}}}{1+e^{\mathbf{x}^{*\prime}_{hi}\boldsymbol{\theta}}}$$

where $y_{hi} = 1$ when the response takes Ordered Value 1, and $y_{hi} = 0$ otherwise.

The full likelihood is

$$L(\boldsymbol{\theta}) = \prod_{h=1}^{H}\prod_{i=1}^{n_h} L_{hi}(\boldsymbol{\theta}) = \frac{e^{\mathbf{y}'\mathbf{X}^*\boldsymbol{\theta}}}{\prod_{h=1}^{H}\prod_{i=1}^{n_h}\left(1+e^{\mathbf{x}^{*\prime}_{hi}\boldsymbol{\theta}}\right)}$$

Unconditional likelihood inference is based on maximizing this likelihood function.

When your nuisance parameters are the stratum-specific intercepts $(\alpha_1,\ldots,\alpha_H)'$, and the slopes $\boldsymbol{\beta}$ are your parameters of interest, "conditioning out" the nuisance parameters produces the conditional likelihood (Lachin 2000)

$$L(\boldsymbol{\beta}) = \prod_{h=1}^{H} L_h(\boldsymbol{\beta}) = \prod_{h=1}^{H}\frac{\prod_{i=1}^{m_h}\exp(\mathbf{x}'_{hi}\boldsymbol{\beta})}{\sum\prod_{j=j_1}^{j_{m_h}}\exp(\mathbf{x}'_{hj}\boldsymbol{\beta})}$$

where the summation is over all $\binom{n_h}{m_h}$ subsets $\{j_1, \ldots, j_{m_h}\}$ of m_h observations chosen from the n_h observations in stratum h. Note that the nuisance parameters have been factored out of this equation.

For conditional asymptotic inference, maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ of the regression parameters are obtained by maximizing the conditional likelihood, and asymptotic results are applied to the conditional likelihood function and the maximum likelihood estimators. A relatively fast method of computing this conditional likelihood and its derivatives is given by Gail, Lubin, and Rubinstein (1981) and Howard (1972). The optimization techniques can be controlled by specifying the NLOPTIONS statement.

Sometimes the log likelihood converges but the estimates diverge. This condition is flagged by having inordinately large standard errors for some of your parameter estimates, and can be monitored by specifying the ITPRINT option. Unfortunately, broad existence criteria such as those discussed in the section "Existence of Maximum Likelihood Estimates" on page 4242 do not exist for this model. It might be possible to circumvent such a problem by standardizing your independent variables before fitting the model.

Regression Diagnostic Details

Diagnostics are used to indicate observations that might have undue influence on the model fit or that might be outliers. Further investigation should be performed before removing such an observation from the data set.

The derivations in this section use an augmentation method described by Storer and Crowley (1985), which provides an estimate of the "one-step" DFBETAS estimates advocated by Pregibon (1984). The method also provides estimates of conditional stratum-specific predicted values, residuals, and leverage for each observation. The augmentation method can take a lot of time and memory.

Following Storer and Crowley (1985), the log-likelihood contribution can be written as

$$\begin{aligned} l_h &= \log(L_h) = \mathbf{y}_h'\boldsymbol{\gamma}_h - a(\boldsymbol{\gamma}_h) \quad \text{where} \\ a(\boldsymbol{\gamma}_h) &= \log\left[\sum \prod_{j=j_1}^{j_{m_h}} \exp(\boldsymbol{\gamma}_{hj})\right] \end{aligned}$$

and the h subscript on matrices indicates the submatrix for the stratum, $\boldsymbol{\gamma}_h = (\gamma_{h1}, \ldots, \gamma_{hn_h})'$, and $\gamma_{hi} = \mathbf{x}_{hi}'\boldsymbol{\beta}$. Then the gradient and information matrix are

$$\begin{aligned} \mathbf{g}(\boldsymbol{\beta}) &= \left\{\frac{\partial l_h}{\partial \boldsymbol{\beta}}\right\}_{h=1}^{H} = \mathbf{X}'(\mathbf{y} - \boldsymbol{\pi}) \\ \boldsymbol{\Lambda}(\boldsymbol{\beta}) &= \left\{\frac{\partial^2 l_h}{\partial \boldsymbol{\beta}^2}\right\}_{h=1}^{H} = \mathbf{X}'\mathrm{diag}(\mathbf{U}_1, \ldots, \mathbf{U}_H)\mathbf{X} \end{aligned}$$

where

$$\begin{aligned}
\pi_{hi} &= \frac{\partial a(\boldsymbol{\gamma}_h)}{\partial \gamma_{hi}} = \frac{\sum_{j(i)} \prod_{j=j_1}^{j_{m_h}} \exp(\gamma_{hj})}{\sum \prod_{j=j_1}^{j_{m_h}} \exp(\gamma_{hj})} \\
\boldsymbol{\pi}_h &= (\pi_{h1}, \ldots, \pi_{hn_h}) \\
\mathbf{U}_h &= \frac{\partial^2 a(\boldsymbol{\gamma}_h)}{\partial \boldsymbol{\gamma}_h^2} = \left\{ \frac{\partial^2 \boldsymbol{a}(\boldsymbol{\gamma}_h)}{\partial \gamma_{hi} \partial \gamma_{hj}} \right\} = \{a_{ij}\} \\
a_{ij} &= \frac{\sum_{k(i,j)} \prod_{k=k_1}^{k_{m_h}} \exp(\gamma_{hk})}{\sum \prod_{k=k_1}^{k_{m_h}} \exp(\gamma_{hk})} - \frac{\partial a(\boldsymbol{\gamma}_h)}{\partial \gamma_{hi}} \frac{\partial a(\boldsymbol{\gamma}_h)}{\partial \gamma_{hj}} = \pi_{hij} - \pi_{hi}\pi_{hj}
\end{aligned}$$

and where π_{hi} is the conditional stratum-specific probability that subject i in stratum h is a case, the summation on $j(i)$ is over all subsets from $\{1, \ldots, n_h\}$ of size m_h that contain the index i, and the summation on $k(i, j)$ is over all subsets from $\{1, \ldots, n_h\}$ of size m_h that contain the indices i and j.

To produce the true one-step estimate $\boldsymbol{\beta}_{hi}^1$, start at the MLE $\widehat{\boldsymbol{\beta}}$, delete the hith observation, and use this reduced data set to compute the next Newton-Raphson step. Note that if there is only one event or one nonevent in a stratum, deletion of that single observation is equivalent to deletion of the entire stratum. The augmentation method does not take this into account.

The augmented model is

$$\text{logit}(\Pr(y_{hi} = 1 | \mathbf{x}_{hi})) = \mathbf{x}_{hi}'\boldsymbol{\beta} + \mathbf{z}_{hi}'\gamma$$

where $\mathbf{z}_{hi} = (0, \ldots, 0, 1, 0, \ldots, 0)'$ has a 1 in the hith coordinate, and use $\boldsymbol{\beta}^0 = (\widehat{\boldsymbol{\beta}}', 0)'$ as the initial estimate for $(\boldsymbol{\beta}', \gamma)'$. The gradient and information matrix before the step are

$$\begin{aligned}
\mathbf{g}(\boldsymbol{\beta}^0) &= \begin{bmatrix} \mathbf{X}' \\ \mathbf{z}_{hi}' \end{bmatrix} (\mathbf{y} - \boldsymbol{\pi}) = \begin{bmatrix} \mathbf{0} \\ y_{hi} - \pi_{hi} \end{bmatrix} \\
\boldsymbol{\Lambda}(\boldsymbol{\beta}^0) &= \begin{bmatrix} \mathbf{X}' \\ \mathbf{z}_{hi}' \end{bmatrix} \mathbf{U} [\mathbf{X} \quad \mathbf{z}_{hi}] = \begin{bmatrix} \boldsymbol{\Lambda}(\boldsymbol{\beta}) & \mathbf{X}'\mathbf{U}\mathbf{z}_{hi} \\ \mathbf{z}_{hi}'\mathbf{U}\mathbf{X} & \mathbf{z}_{hi}'\mathbf{U}\mathbf{z}_{hi} \end{bmatrix}
\end{aligned}$$

Inserting the $\boldsymbol{\beta}^0$ and $(\mathbf{X}', \mathbf{z}_{hi}')'$ into the Gail, Lubin, and Rubinstein (1981) algorithm provides the appropriate estimates of $\mathbf{g}(\boldsymbol{\beta}^0)$ and $\boldsymbol{\Lambda}(\boldsymbol{\beta}^0)$. Indicate these estimates with $\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\widehat{\boldsymbol{\beta}})$, $\widehat{\mathbf{U}} = \mathbf{U}(\widehat{\boldsymbol{\beta}})$, $\widehat{\mathbf{g}}$, and $\widehat{\boldsymbol{\Lambda}}$.

DFBETA is computed from the information matrix as

$$\begin{aligned}
\Delta_{hi}\boldsymbol{\beta} &= \boldsymbol{\beta}^0 - \boldsymbol{\beta}_{hi}^1 \\
&= -\widehat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{\beta}^0)\widehat{\mathbf{g}}(\boldsymbol{\beta}^0) \\
&= -\widehat{\boldsymbol{\Lambda}}^{-1}(\widehat{\boldsymbol{\beta}})(\mathbf{X}'\widehat{\mathbf{U}}\mathbf{z}_{hi})\mathbf{M}^{-1}\mathbf{z}_{hi}'(\mathbf{y} - \widehat{\boldsymbol{\pi}})
\end{aligned}$$

where

$$\mathbf{M} = (\mathbf{z}_{hi}'\widehat{\mathbf{U}}\mathbf{z}_{hi}) - (\mathbf{z}_{hi}'\widehat{\mathbf{U}}\mathbf{X})\widehat{\boldsymbol{\Lambda}}^{-1}(\widehat{\boldsymbol{\beta}})(\mathbf{X}'\widehat{\mathbf{U}}\mathbf{z}_{hi})$$

For each observation in the data set, a DFBETA statistic is computed for each parameter β_j, $1 \le j \le p$, and standardized by the standard error of β_j from the full data set to produce the estimate of DFBETAS.

The estimated leverage is defined as

$$h_{hi} = \frac{\text{trace}\{(\mathbf{z}'_{hi}\widehat{\mathbf{U}}\mathbf{X})\widehat{\mathbf{\Lambda}}^{-1}(\widehat{\boldsymbol{\beta}})(\mathbf{X}'\widehat{\mathbf{U}}\mathbf{z}_{hi})\}}{\text{trace}\{\mathbf{z}'_{hi}\widehat{\mathbf{U}}\mathbf{z}_{hi}\}}$$

This definition of leverage produces different values from those defined by Pregibon (1984); Moolgavkar, Lustbader, and Venzon (1985); Hosmer and Lemeshow (2000); however, it has the advantage that no extra computations beyond those for the DFBETAS are required.

The estimated residuals $e_{hi} = y_{hi} - \widehat{\pi}_{hi}$ are obtained from $\widehat{\mathbf{g}}(\boldsymbol{\beta}^0)$, and the weights, or predicted probabilities, are then $\widehat{\pi}_{hi} = y_{hi} - e_{hi}$. The residuals are standardized and reported as (estimated) Pearson residuals:

$$\frac{r_{hi} - n_{hi}\widehat{\pi}_{hi}}{\sqrt{n_{hi}\widehat{\pi}_{hi}(1-\widehat{\pi}_{hi})}}$$

where r_{hi} is the number of events in the observation and n_{hi} is the number of trials.

The STDRES option in the INFLUENCE and PLOTS=INFLUENCE options computes the standardized Pearson residual:

$$e_{s,hi} = \frac{e_{hi}}{\sqrt{1-h_{hi}}}$$

For events/trials MODEL statement syntax, treat each observation as two observations (the first for the nonevents and the second for the events) with frequencies $f_{h,2i-1} = n_{hi} - r_{hi}$ and $f_{h,2i} = r_{hi}$, and augment the model with a matrix $\mathbf{Z}_{hi} = [\mathbf{z}_{h,2i-1}\mathbf{z}_{h,2i}]$ instead of a single $\mathbf{z}_{hi}$ vector. Writing $\gamma_{hi} = \mathbf{x}'_{hi}\boldsymbol{\beta}\, f_{hi}$ in the preceding section results in the following gradient and information matrix:

$$\mathbf{g}(\boldsymbol{\beta}^0) = \begin{bmatrix} \mathbf{0} \\ f_{h,2i-1}(y_{h,2i-1} - \pi_{h,2i-1}) \\ f_{h,2i}(y_{h,2i} - \pi_{h,2i}) \end{bmatrix}$$

$$\mathbf{\Lambda}(\boldsymbol{\beta}^0) = \begin{bmatrix} \mathbf{\Lambda}(\boldsymbol{\beta}) & \mathbf{X}'\text{diag}(\mathbf{f})\mathbf{U}\text{diag}(\mathbf{f})\mathbf{Z}_{hi} \\ \mathbf{Z}'_{hi}\text{diag}(\mathbf{f})\mathbf{U}\text{diag}(\mathbf{f})\mathbf{X} & \mathbf{Z}'_{hi}\text{diag}(\mathbf{f})\mathbf{U}\text{diag}(\mathbf{f})\mathbf{Z}_{hi} \end{bmatrix}$$

The predicted probabilities are then $\widehat{\pi}_{hi} = y_{h,2i} - e_{h,2i}/r_{h,2i}$, while the leverage and the DFBETAS are produced from $\mathbf{\Lambda}(\boldsymbol{\beta}^0)$ in a fashion similar to that for the preceding single-trial equations.

Exact Conditional Logistic Regression

The theory of exact logistic regression, also known as exact conditional logistic regression, was originally laid out by Cox (1970), and the computational methods employed in PROC LOGISTIC are described in Hirji, Mehta, and Patel (1987); Hirji (1992); Mehta, Patel, and Senchaudhuri (1992). Other useful references for the derivations include Cox and Snell (1989); Agresti (1990); Mehta and Patel (1995).

Exact conditional inference is based on generating the conditional distribution for the sufficient statistics of the parameters of interest. This distribution is called the *permutation* or *exact conditional* distribution. Using the notation in the section "Computational Details" on page 4271, follow Mehta and Patel (1995) and

first note that the sufficient statistics $\mathbf{T} = (T_1, \ldots, T_p)$ for the parameter vector of intercepts and slopes, $\boldsymbol{\beta}$, are

$$T_j = \sum_{i=1}^{n} y_i x_{ij}, \quad j = 1, \ldots, p$$

Denote a vector of observable sufficient statistics as $\mathbf{t} = (t_1, \ldots, t_p)'$.

The probability density function (PDF) for $\mathbf{T}$ can be created by summing over all binary sequences $\mathbf{y}$ that generate an observable $\mathbf{t}$ and letting $C(\mathbf{t}) = ||\{\mathbf{y} : \mathbf{y}'\mathbf{X} = \mathbf{t}'\}||$ denote the number of sequences $\mathbf{y}$ that generate $\mathbf{t}$

$$\Pr(\mathbf{T} = \mathbf{t}) = \frac{C(\mathbf{t}) \exp(\mathbf{t}'\boldsymbol{\beta})}{\prod_{i=1}^{n}[1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})]}$$

In order to condition out the nuisance parameters, partition the parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\mathrm{N}}', \boldsymbol{\beta}_{\mathrm{I}}')'$, where $\boldsymbol{\beta}_{\mathrm{N}}$ is a $p_{\mathrm{N}} \times 1$ vector of the nuisance parameters, and $\boldsymbol{\beta}_{\mathrm{I}}$ is the parameter vector for the remaining $p_{\mathrm{I}} = p - p_{\mathrm{N}}$ parameters of interest. Likewise, partition $\mathbf{X}$ into $\mathbf{X}_{\mathrm{N}}$ and $\mathbf{X}_{\mathrm{I}}$, $\mathbf{T}$ into $\mathbf{T}_{\mathrm{N}}$ and $\mathbf{T}_{\mathrm{I}}$, and $\mathbf{t}$ into $\mathbf{t}_{\mathrm{N}}$ and $\mathbf{t}_{\mathrm{I}}$. The nuisance parameters can be removed from the analysis by conditioning on their sufficient statistics to create the conditional likelihood of $\mathbf{T}_{\mathrm{I}}$ given $\mathbf{T}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}$,

$$\begin{aligned}\Pr(\mathbf{T}_{\mathrm{I}} = \mathbf{t}_{\mathrm{I}} | \mathbf{T}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}) &= \frac{\Pr(\mathbf{T} = \mathbf{t})}{\Pr(\mathbf{T}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}})} \\ = f_{\boldsymbol{\beta}_{\mathrm{I}}}(\mathbf{t}_{\mathrm{I}} | \mathbf{t}_{\mathrm{N}}) &= \frac{C(\mathbf{t}_{\mathrm{N}}, \mathbf{t}_{\mathrm{I}}) \exp(\mathbf{t}_{\mathrm{I}}'\boldsymbol{\beta}_{\mathrm{I}})}{\sum_u C(\mathbf{t}_{\mathrm{N}}, \mathbf{u}) \exp(\mathbf{u}'\boldsymbol{\beta}_{\mathrm{I}})}\end{aligned}$$

where $C(\mathbf{t}_{\mathrm{N}}, \mathbf{u})$ is the number of vectors $\mathbf{y}$ such that $\mathbf{y}'\mathbf{X}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}$ and $\mathbf{y}'\mathbf{X}_{\mathrm{I}} = \mathbf{u}$. Note that the nuisance parameters have factored out of this equation, and that $C(\mathbf{t}_{\mathrm{N}}, \mathbf{t}_{\mathrm{I}})$ is a constant.

The goal of the exact conditional analysis is to determine how likely the observed response $\mathbf{y}_0$ is with respect to all 2^n possible responses $\mathbf{y} = (y_1, \ldots, y_n)'$. One way to proceed is to generate every $\mathbf{y}$ vector for which $\mathbf{y}'\mathbf{X}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}$, and count the number of vectors $\mathbf{y}$ for which $\mathbf{y}'\mathbf{X}_{\mathrm{I}}$ is equal to each unique $\mathbf{t}_{\mathrm{I}}$. Generating the conditional distribution from complete enumeration of the joint distribution is conceptually simple; however, this method becomes computationally infeasible very quickly. For example, if you had only 30 observations, you would have to scan through 2^{30} different $\mathbf{y}$ vectors.

Several algorithms are available in PROC LOGISTIC to generate the exact distribution. All of the algorithms are based on the following observation. Given any $\mathbf{y} = (y_1, \ldots, y_n)'$ and a design $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, let $\mathbf{y}_{(i)} = (y_1, \ldots, y_i)'$ and $\mathbf{X}_{(i)} = (\mathbf{x}_1, \ldots, \mathbf{x}_i)'$ be the first i rows of each matrix. Write the sufficient statistic based on these i rows as $\mathbf{t}_{(i)}' = \mathbf{y}_{(i)}'\mathbf{X}_{(i)}$. A recursion relation results: $\mathbf{t}_{(i+1)} = \mathbf{t}_{(i)} + y_{i+1}\mathbf{x}_{i+1}$.

The following methods are available:

- The *multivariate shift algorithm* developed by Hirji, Mehta, and Patel (1987), which steps through the recursion relation by adding one observation at a time and building an intermediate distribution at each step. If it determines that $\mathbf{t}_{(i)}$ for the nuisance parameters could eventually equal $\mathbf{t}$, then $\mathbf{t}_{(i)}$ is added to the intermediate distribution.

- An extension of the multivariate shift algorithm to generalized logit models by Hirji (1992). Since the generalized logit model fits a new set of parameters to each logit, the number of parameters in the

model can easily get too large for this algorithm to handle. Note for these models that the hypothesis tests for each effect are computed across the logit functions, while individual parameters are estimated for each logit function.

- A network algorithm described in Mehta, Patel, and Senchaudhuri (1992), which builds a network for each parameter that you are conditioning out in order to identify feasible y_i for the $\mathbf{y}$ vector. These networks are combined and the set of feasible y_i is further reduced, and then the multivariate shift algorithm uses this knowledge to build the exact distribution without adding as many intermediate $\mathbf{t}_{(i+1)}$ as the multivariate shift algorithm does.
- A hybrid Monte Carlo and network algorithm described by Mehta, Patel, and Senchaudhuri (2000), which extends their 1992 algorithm by sampling from the combined network to build the exact distribution.

The bulk of the computation time and memory for these algorithms is consumed by the creation of the networks and the exact joint distribution. After the joint distribution for a set of effects is created, the computational effort required to produce hypothesis tests and parameter estimates for any subset of the effects is (relatively) trivial. See the section "Computational Resources for Exact Logistic Regression" on page 4284 for more computational notes about exact analyses.

NOTE: An alternative to using these exact conditional methods is to perform Firth's bias-reducing penalized likelihood method (see the FIRTH option in the MODEL statement); this method has the advantage of being much faster and less memory intensive than exact algorithms, but it might not converge to a solution.

Hypothesis Tests

Consider testing the null hypothesis $H_0: \boldsymbol{\beta}_{\mathrm{I}} = \mathbf{0}$ against the alternative $H_A: \boldsymbol{\beta}_{\mathrm{I}} \neq \mathbf{0}$, conditional on $\mathbf{T}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}$. Under the null hypothesis, the test statistic for the *exact probability test* is just $f_{\boldsymbol{\beta}_{\mathrm{I}}=\mathbf{0}}(\mathbf{t}_{\mathrm{I}}|\mathbf{t}_{\mathrm{N}})$, while the corresponding p-value is the probability of getting a less likely (more extreme) statistic,

$$p(\mathbf{t}_{\mathrm{I}}|\mathbf{t}_{\mathrm{N}}) = \sum_{u \in \Omega_p} f_0(\mathbf{u}|\mathbf{t}_{\mathrm{N}})$$

where $\Omega_p = \{\mathbf{u}$: there exist $\mathbf{y}$ with $\mathbf{y}'\mathbf{X}_{\mathrm{I}} = \mathbf{u}$, $\mathbf{y}'\mathbf{X}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}$, and $f_0(\mathbf{u}|\mathbf{t}_{\mathrm{N}}) \leq f_0(\mathbf{t}_{\mathrm{I}}|\mathbf{t}_{\mathrm{N}})\}$.

For the *exact conditional scores test*, the conditional mean $\boldsymbol{\mu}_{\mathrm{I}}$ and variance matrix $\boldsymbol{\Sigma}_{\mathrm{I}}$ of the $\mathbf{T}_{\mathrm{I}}$ (conditional on $\mathbf{T}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}$) are calculated, and the score statistic for the observed value,

$$s = (\mathbf{t}_{\mathrm{I}} - \boldsymbol{\mu}_{\mathrm{I}})' \boldsymbol{\Sigma}_{\mathrm{I}}^{-1} (\mathbf{t}_{\mathrm{I}} - \boldsymbol{\mu}_{\mathrm{I}})$$

is compared to the score for each member of the distribution

$$S(\mathbf{T}_{\mathrm{I}}) = (\mathbf{T}_{\mathrm{I}} - \boldsymbol{\mu}_{\mathrm{I}})' \boldsymbol{\Sigma}_{\mathrm{I}}^{-1} (\mathbf{T}_{\mathrm{I}} - \boldsymbol{\mu}_{\mathrm{I}})$$

The resulting p-value is

$$p(\mathbf{t}_{\mathrm{I}}|\mathbf{t}_{\mathrm{N}}) = \Pr(S \geq s) = \sum_{u \in \Omega_s} f_0(\mathbf{u}|\mathbf{t}_{\mathrm{N}})$$

where $\Omega_s = \{\mathbf{u}$: there exist $\mathbf{y}$ with $\mathbf{y}'\mathbf{X}_{\mathrm{I}} = \mathbf{u}$, $\mathbf{y}'\mathbf{X}_{\mathrm{N}} = \mathbf{t}_{\mathrm{N}}$, and $S(\mathbf{u}) \geq s\}$.

The mid-p statistic, defined as

$$p(\mathbf{t}_\mathrm{I}|\mathbf{t}_\mathrm{N}) - \frac{1}{2} f_0(\mathbf{t}_\mathrm{I}|\mathbf{t}_\mathrm{N})$$

was proposed by Lancaster (1961) to compensate for the discreteness of a distribution. See Agresti (1992) for more information. However, to allow for more flexibility in handling ties, you can write the mid-p statistic as (based on a suggestion by Lamotte (2002) and generalizing Vollset, Hirji, and Afifi (1991))

$$\sum_{u \in \Omega_<} f_0(\mathbf{u}|\mathbf{t}_\mathrm{N}) + \delta_1 f_0(\mathbf{t}_\mathrm{I}|\mathbf{t}_\mathrm{N}) + \delta_2 \sum_{u \in \Omega_=} f_0(\mathbf{u}|\mathbf{t}_\mathrm{N})$$

where, for $i \in \{p, s\}$, $\Omega_<$ is Ω_i using strict inequalities, and $\Omega_=$ is Ω_i using equalities with the added restriction that $\mathbf{u} \neq \mathbf{t}_\mathrm{I}$. Letting $(\delta_1, \delta_2) = (0.5, 1.0)$ yields Lancaster's mid-p.

CAUTION: When the exact distribution has ties and METHOD=NETWORKMC is specified, the Monte Carlo algorithm estimates $p(\mathbf{t}|\mathbf{t}_\mathrm{N})$ with error, and hence it cannot determine precisely which values contribute to the reported p-values. For example, if the exact distribution has densities $\{0.2, 0.2, 0.2, 0.4\}$ and if the observed statistic has probability 0.2, then the exact probability p-value is exactly 0.6. Under Monte Carlo sampling, if the densities after N samples are $\{0.18, 0.21, 0.23, 0.38\}$ and the observed probability is 0.21, then the resulting p-value is 0.39. Therefore, the exact probability test p-value for this example fluctuates between 0.2, 0.4, and 0.6, and the reported p-values are actually lower bounds for the true p-values. If you need more precise values, you can specify the OUTDIST= option, determine appropriate cutoff values for the observed probability and score, and then construct the true p-value estimates from the OUTDIST= data set and display them in the SAS log by using the following statements:

```
data _null_;
   set outdist end=end;
   retain pvalueProb 0 pvalueScore 0;
   if prob < ProbCutOff then pvalueProb+prob;
   if score > ScoreCutOff then pvalueScore+prob;
   if end then put pvalueProb= pvalueScore=;
run;
```

Inference for a Single Parameter

Exact parameter estimates are derived for a single parameter β_i by regarding all the other parameters $\boldsymbol{\beta}_\mathrm{N} = (\beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_{p_\mathrm{N}+p_\mathrm{I}})'$ as nuisance parameters. The appropriate sufficient statistics are $\mathbf{T}_\mathrm{I} = T_i$ and $\mathbf{T}_\mathrm{N} = (T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_{p_\mathrm{N}+p_\mathrm{I}})'$, with their observed values denoted by the lowercase t. Hence, the conditional PDF used to create the parameter estimate for β_i is

$$f_{\beta_i}(t_i|\mathbf{t}_\mathrm{N}) = \frac{C(\mathbf{t}_\mathrm{N}, t_i)\exp(t_i\beta_i)}{\sum_{u \in \Omega} C(\mathbf{t}_\mathrm{N}, u)\exp(u\beta_i)}$$

for $\Omega = \{u\colon$ there exist $\mathbf{y}$ with $T_i = u$ and $\mathbf{T}_\mathrm{N} = \mathbf{t}_\mathrm{N}\}$.

The maximum exact conditional likelihood estimate is the quantity $\widehat{\beta}_i$, which maximizes the conditional PDF. A Newton-Raphson algorithm is used to perform this search. However, if the observed t_i attains either its maximum or minimum value in the exact distribution (that is, either $t_i = \min\{u : u \in \Omega\}$ or $t_i = \max\{u : u \in \Omega\}$), then the conditional PDF is monotonically increasing in β_i and cannot be maximized. In this case, a median unbiased estimate (Hirji, Tsiatis, and Mehta 1989) $\widehat{\beta}_i$ is produced that satisfies $f_{\widehat{\beta}_i}(t_i|\mathbf{t}_\mathrm{N}) = 0.5$, and a Newton-Raphson algorithm is used to perform the search.

The standard error of the exact conditional likelihood estimate is just the negative of the inverse of the second derivative of the exact conditional log likelihood (Agresti 2002).

Likelihood ratio tests based on the conditional PDF are used to test the null $H_0: \beta_i = 0$ against the alternative $H_A: \beta_i > 0$. The critical region for this UMP test consists of the upper tail of values for T_i in the exact distribution. Thus, the one-sided significance level $p_+(t_i; 0)$ is

$$p_+(t_i; 0) = \sum_{u \geq t_i} f_0(u|\mathbf{t}_N)$$

Similarly, the one-sided significance level $p_-(t_i; 0)$ against $H_A: \beta_i < 0$ is

$$p_-(t_i; 0) = \sum_{u \leq t_i} f_0(u|\mathbf{t}_N)$$

The two-sided significance level $p(t_i; 0)$ against $H_A: \beta_i \neq 0$ is calculated as

$$p(t_i; 0) = 2 \min[p_-(t_i; 0), p_+(t_i; 0)]$$

An upper $100(1-2\epsilon)\%$ exact confidence limit for $\widehat{\beta}_i$ corresponding to the observed t_i is the solution $\beta_U(t_i)$ of $\epsilon = p_-(t_i, \beta_U(t_i))$, while the lower exact confidence limit is the solution $\beta_L(t_i)$ of $\epsilon = p_+(t_i, \beta_L(t_i))$. Again, a Newton-Raphson procedure is used to search for the solutions. Note that one of the confidence limits for a median unbiased estimate is set to infinity, but the other is still computed at ϵ. This results in the display of a one-sided $100(1-\epsilon)\%$ confidence interval; if you want the 2ϵ limit instead, you can specify the ONESIDED option.

Specifying the ONESIDED option displays only one p-value and one confidence interval, because small values of $p_+(t_i; 0)$ and $p_-(t_i; 0)$ support different alternative hypotheses and only one of these p-values can be less than 0.50.

The mid-p confidence limits are the solutions to $\min\{p_-(t_i, \beta(t_i)), p_+(t_i, \beta(t_i))\} - (1-\delta_1) f_{\beta(t_i)}(t_i|\mathbf{t}_N) = \epsilon$ for $\epsilon = \alpha/2, 1-\alpha/2$ (Vollset, Hirji, and Afifi 1991). $\delta_1 = 1$ produces the usual exact (or *max-p*) confidence interval, $\delta_1 = 0.5$ yields the mid-p interval, and $\delta_1 = 0$ gives the *min-p* interval. The mean of the endpoints of the max-p and min-p intervals provides the *mean-p* interval as defined by Hirji, Mehta, and Patel (1988).

Estimates and confidence intervals for the odds ratios are produced by exponentiating the estimates and interval endpoints for the parameters.

Notes about Exact *p*-Values

In the "Conditional Exact Tests" table, the exact probability test is not necessarily a sum of tail areas and can be inflated if the distribution is skewed. The more robust exact conditional scores test is a sum of tail areas and is generally preferred over the exact probability test.

The p-value reported for a single parameter in the "Exact Parameter Estimates" table is twice the one-sided tail area of a likelihood ratio test against the null hypothesis of the parameter equaling zero.

Input and Output Data Sets

OUTEST= Output Data Set

The OUTEST= data set contains one observation for each BY group containing the maximum likelihood estimates of the regression coefficients. If you also use the COVOUT option in the PROC LOGISTIC statement, there are additional observations containing the rows of the estimated covariance matrix. If you specify SELECTION=FORWARD, BACKWARD, or STEPWISE, only the estimates of the parameters and covariance matrix for the final model are output to the OUTEST= data set.

Variables in the OUTEST= Data Set

The OUTEST= data set contains the following variables:

- any BY variables specified
- _LINK_, a character variable of length 8 with four possible values: CLOGLOG for the complementary log-log function, LOGIT for the logit function, NORMIT for the probit (alias normit) function, and GLOGIT for the generalized logit function
- _TYPE_, a character variable of length 8 with two possible values: PARMS for parameter estimates or COV for covariance estimates. If an EXACT statement is also specified, then two other values are possible: EPARMMLE for the exact maximum likelihood estimates and EPARMMUE for the exact median unbiased estimates.
- _NAME_, a character variable containing the name of the response variable when _TYPE_=PARMS, EPARMMLE, and EPARMMUE, or the name of a model parameter when _TYPE_=COV
- _STATUS_, a character variable that indicates whether the estimates have converged
- one variable for each intercept parameter
- one variable for each slope parameter and one variable for the offset variable if the OFFSET= option if specified. If an effect is not included in the final model in a model building process, the corresponding parameter estimates and covariances are set to missing values.
- _LNLIKE_, the log likelihood

Parameter Names in the OUTEST= Data Set

If there are only two response categories in the entire data set, the intercept parameter is named Intercept. If there are more than two response categories in the entire data set, the intercept parameters are named Intercept_xxx, where xxx is the value (formatted if a format is applied) of the corresponding response category.

For continuous explanatory variables, the names of the parameters are the same as the corresponding variables. For CLASS variables, the parameter names are obtained by concatenating the corresponding CLASS variable name with the CLASS category; see the section "Class Variable Naming Convention" on page 4189 for more details. For interaction and nested effects, the parameter names are created by concatenating the names of each effect.

For the generalized logit model, names of parameters corresponding to each nonreference category contain _xxx as the suffix, where xxx is the value (formatted if a format is applied) of the corresponding nonreference

category. For example, suppose the variable Net3 represents the television network (ABC, CBS, and NBC) viewed at a certain time. The following statements fit a generalized logit model with Age and Gender (a CLASS variable with values Female and Male) as explanatory variables:

```
proc logistic;
   class Gender;
   model Net3 = Age Gender / link=glogit;
run;
```

There are two logit functions, one contrasting ABC with NBC and the other contrasting CBS with NBC. For each logit, there are three parameters: an intercept parameter, a slope parameter for Age, and a slope parameter for Gender (since there are only two gender levels and the EFFECT parameterization is used by default). The names of the parameters and their descriptions are as follows:

Intercept_ABC	intercept parameter for the logit contrasting ABC with NBC
Intercept_CBS	intercept parameter for the logit contrasting CBS with NBC
Age_ABC	Age slope parameter for the logit contrasting ABC with NBC
Age_CBS	Age slope parameter for the logit contrasting CBS with NBC
GenderFemale_ABC	Gender=Female slope parameter for the logit contrasting ABC with NBC
GenderFemale_CBS	Gender=Female slope parameter for the logit contrasting CBS with NBC

INEST= Input Data Set

You can specify starting values for the iterative algorithm in the INEST= data set. The INEST= data set has the same structure as the OUTEST= data set but is not required to have all the variables or observations that appear in the OUTEST= data set. A previous OUTEST= data set can be used as, or modified for use as, an INEST= data set.

The INEST= data set must contain the intercept variables (named Intercept for binary response models and Intercept, Intercept_2, Intercept_3, and so forth, for ordinal and nominal response models) and all explanatory variables in the MODEL statement. If BY processing is used, the INEST= data set should also include the BY variables, and there must be one observation for each BY group. If the INEST= data set also contains the _TYPE_ variable, only observations with _TYPE_ value 'PARMS' are used as starting values.

OUT= Output Data Set in the OUTPUT Statement

The OUT= data set in the OUTPUT statement contains all the variables in the input data set along with statistics you request by specifying *keyword=name* options or the PREDPROBS= option in the OUTPUT statement. In addition, if you use the single-trial syntax and you request any of the XBETA=, STDXBETA=, PREDICTED=, LCL=, and UCL= options, the OUT= data set contains the automatic variable _LEVEL_. The value of _LEVEL_ identifies the response category upon which the computed values of XBETA=, STDXBETA=, PREDICTED=, LCL=, and UCL= are based.

When there are more than two response levels, only variables named by the XBETA=, STDXBETA=, PREDICTED=, LOWER=, and UPPER= options and the variables given by PREDPROBS=(INDIVIDUAL CUMULATIVE) have their values computed; the other variables have missing values. If you fit a generalized logit model, the cumulative predicted probabilities are not computed.

When there are only two response categories, each input observation produces one observation in the OUT= data set.

If there are more than two response categories and you specify only the PREDPROBS= option, then each input observation produces one observation in the OUT= data set. However, if you fit an ordinal (cumulative) model and specify options other than the PREDPROBS= options, each input observation generates as many output observations as one fewer than the number of response levels, and the predicted probabilities and their confidence limits correspond to the cumulative predicted probabilities. If you fit a generalized logit model and specify options other than the PREDPROBS= options, each input observation generates as many output observations as the number of response categories; the predicted probabilities and their confidence limits correspond to the probabilities of individual response categories.

For observations in which only the response variable is missing, values of the XBETA=, STDXBETA=, PREDICTED=, UPPER=, LOWER=, and the PREDPROBS= options are computed even though these observations do not affect the model fit. This enables, for instance, predicted probabilities to be computed for new observations.

OUT= Output Data Set in a SCORE Statement

The OUT= data set in a SCORE statement contains all the variables in the data set being scored. The data set being scored can be either the input DATA= data set in the PROC LOGISTIC statement or the DATA= data set in the SCORE statement. The DATA= data set in the SCORE statement does not need to contain the response variable.

If the data set being scored contains the response variable, then denote the *normalized* levels (left-justified, formatted values of 16 characters or less) of your response variable Y by $Y_1, \ldots, Y_{k+1}$. For each response level, the OUT= data set also contains the following:

- F_*Y*, the normalized levels of the response variable Y in the data set being scored. If the events/trials syntax is used, the F_*Y* variable is not created.
- I_*Y*, the normalized levels that the observations are classified into. Note that an observation is classified into the level with the largest probability. If the events/trials syntax is used, the _INTO_ variable is created instead, and it contains the values EVENT and NONEVENT.
- P_Y_i, the posterior probabilities of the normalized response level Y_i
- If the CLM option is specified in the SCORE statement, the OUT= data set also includes the following:
  - LCL_Y_i, the lower $100(1-\alpha)\%$ confidence limits for P_Y_i
  - UCL_Y_i, the upper $100(1-\alpha)\%$ confidence limits for P_Y_i

OUTDIST= Output Data Set

The OUTDIST= data set contains every exact conditional distribution necessary to process the corresponding EXACT statement. For example, the following statements create one distribution for the x1 parameter and another for the x2 parameters, and produce the data set dist shown in Table 54.12:

```
data test;
   input y x1 x2 count;
   datalines;
0 0 0 1
1 0 0 1
0 1 1 2
1 1 1 1
1 0 2 3
1 1 2 1
1 2 0 3
1 2 1 2
1 2 2 1
;

proc logistic data=test exactonly;
   class x2 / param=ref;
   model y=x1 x2;
   exact x1 x2/ outdist=dist;
run;
proc print data=dist;
run;
```

Table 54.12 OUTDIST= Data Set

Obs	x1	x20	x21	Count	Score	Prob
1	.	0	0	3	5.81151	0.03333
2	.	0	1	15	1.66031	0.16667
3	.	0	2	9	3.12728	0.10000
4	.	1	0	15	1.46523	0.16667
5	.	1	1	18	0.21675	0.20000
6	.	1	2	6	4.58644	0.06667
7	.	2	0	19	1.61869	0.21111
8	.	2	1	2	3.27293	0.02222
9	.	3	0	3	6.27189	0.03333
10	2	.	.	6	3.03030	0.12000
11	3	.	.	12	0.75758	0.24000
12	4	.	.	11	0.00000	0.22000
13	5	.	.	18	0.75758	0.36000
14	6	.	.	3	3.03030	0.06000

The first nine observations in the dist data set contain an exact distribution for the parameters of the x2 effect (hence the values for the x1 parameter are missing), and the remaining five observations are for the x1 parameter. If a joint distribution was created, there would be observations with values for both the x1 and x2 parameters. For CLASS variables, the corresponding parameters in the dist data set are identified by concatenating the variable name with the appropriate classification level.

The data set contains the possible sufficient statistics of the parameters for the effects specified in the EXACT statement, and the Count variable contains the number of different responses that yield these statistics. In particular, there are six possible response vectors $\mathbf{y}$ for which the dot product $\mathbf{y}'\mathbf{x1}$ was equal to 2, and

for which $\mathbf{y'x20}$, $\mathbf{y'x21}$, and $\mathbf{y'1}$ were equal to their actual observed values (displayed in the "Sufficient Statistics" table).

When hypothesis tests are performed on the parameters, the Prob variable contains the probability of obtaining that statistic (which is just the count divided by the total count), and the Score variable contains the score for that statistic.

The OUTDIST= data set can contain a different exact conditional distribution for each specified EXACT statement. For example, consider the following EXACT statements:

```
exact 'O1'   x1    /           outdist=o1;
exact 'OJ12' x1 x2 / jointonly outdist=oj12;
exact 'OA12' x1 x2 / joint     outdist=oa12;
exact 'OE12' x1 x2 / estimate  outdist=oe12;
```

The O1 statement outputs a single exact conditional distribution. The OJ12 statement outputs only the joint distribution for x1 and x2. The OA12 statement outputs three conditional distributions: one for x1, one for x2, and one jointly for x1 and x2. The OE12 statement outputs two conditional distributions: one for x1 and the other for x2. Data set oe12 contains both the x1 and x2 variables; the distribution for x1 has missing values in the x2 column while the distribution for x2 has missing values in the x1 column.

OUTROC= Output Data Set

The OUTROC= data set contains data necessary for producing the ROC curve, and can be created by specifying the OUTROC= option in the MODEL statement or the OUTROC= option in the SCORE statement: It has the following variables:

- any BY variables specified
- _STEP_, the model step number. This variable is not included if model selection is not requested.
- _PROB_, the estimated probability of an event. These estimated probabilities serve as cutpoints for predicting the response. Any observation with an estimated event probability that exceeds or equals _PROB_ is predicted to be an event; otherwise, it is predicted to be a nonevent. Predicted probabilities that are close to each other are grouped together, with the maximum allowable difference between the largest and smallest values less than a constant that is specified by the ROCEPS= option. The smallest estimated probability is used to represent the group.
- _POS_, the number of correctly predicted event responses
- _NEG_, the number of correctly predicted nonevent responses
- _FALPOS_, the number of falsely predicted event responses
- _FALNEG_, the number of falsely predicted nonevent responses
- _SENSIT_, the sensitivity, which is the proportion of event observations that were predicted to have an event response
- _1MSPEC_, one minus specificity, which is the proportion of nonevent observations that were predicted to have an event response

Note that none of these statistics are affected by the bias-correction method discussed in the section "Classification Table" on page 4255. An ROC curve is obtained by plotting _SENSIT_ against _1MSPEC_.

For more information, see the section "Receiver Operating Characteristic Curves" on page 4260.

Computational Resources

The memory needed to fit an unconditional model is approximately $8n(p+2)+24(p+2)^2$ bytes, where p is the number of parameters estimated and n is the number of observations in the data set. For cumulative response models with more than two response levels, a test of the parallel lines assumption requires an additional memory of approximately $4k^2(m+1)^2+24(m+2)^2$ bytes, where k is the number of response levels and m is the number of slope parameters. However, if this additional memory is not available, the procedure skips the test and finishes the other computations. You might need more memory if you use the SELECTION= option for model building.

The data that consist of relevant variables (including the design variables for model effects) and observations for fitting the model are stored in a temporary utility file. If sufficient memory is available, such data will also be kept in memory; otherwise, the data are reread from the utility file for each evaluation of the likelihood function and its derivatives, with the resulting execution time of the procedure substantially increased. Specifying the MULTIPASS option in the MODEL statement avoids creating this utility file and also does not store the data in memory; instead, the DATA= data set is reread when needed. This saves approximately $8n(p+2)$ bytes of memory but increases the execution time.

If a conditional logistic regression is performed, then approximately $4(m^2+m+4)\max_h(m_h)+(8s_H+36)H+12s_H$ additional bytes of memory are needed, where m_h is the number of events in stratum h, H is the total number of strata, and s_H is the number of variables used to define the strata. If the CHECKDEPENDENCY=ALL option is specified in the STRATA statement, then an extra $4(m+H)(m+H+1)$ bytes are required, and the resulting execution time of the procedure might be substantially increased.

Computational Resources for Exact Logistic Regression

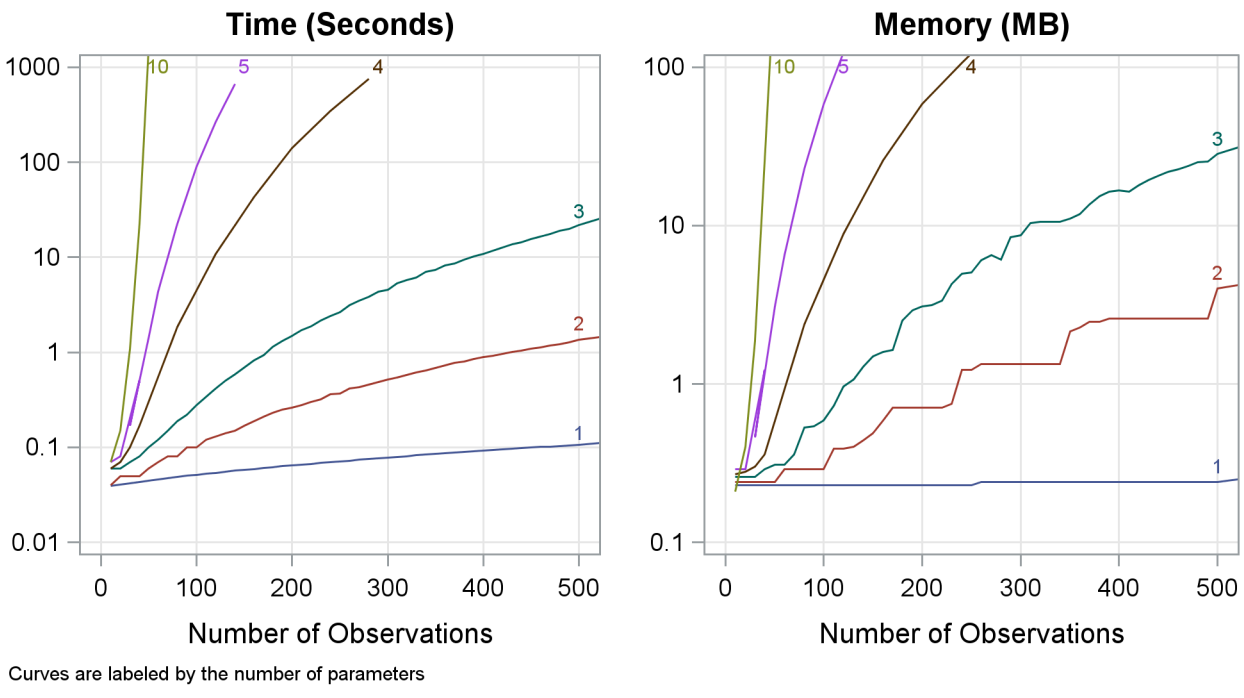
Many problems require a prohibitive amount of time and memory for exact computations, depending on the speed and memory available on your computer. For such problems, consider whether exact methods are really necessary. Stokes, Davis, and Koch (2012) suggest looking at exact p-values when the sample size is small and the approximate p-values from the unconditional analysis are less than 0.10, and they provide *rules of thumb* for determining when various models are valid.

A formula does not exist that can predict the amount of time and memory necessary to generate the exact conditional distributions for a particular problem. The time and memory required depends on several factors, including the total sample size, the number of parameters of interest, the number of nuisance parameters, and the order in which the parameters are processed. To provide a feel for how these factors affect performance, 19 data sets containing Nobs $\in \{10,\ldots,500\}$ observations consisting of up to 10 independent uniform binary covariates (X1,...,*XN*) and a binary response variable (Y), are generated, and the following statements create exact conditional distributions for X1 conditional on the other covariates by using the default METHOD=NETWORK. Figure 54.11 displays results obtained on a 400Mhz PC with 768MB RAM running Microsoft Windows NT.

```
data one;
   do obs=1 to HalfNobs;
```

```
      do Y=0 to 1;
         X1=round(ranuni(0));
         ...
         XN=round(ranuni(0));
         output;
      end;
   end;
options fullstimer;
proc logistic exactonly;
   exactoptions method=network maxtime=1200;
   class X1...XN / param=ref;
   model Y=X1...XN;
   exact X1 / outdist=dist;
run;
```

Figure 54.11 Mean Time and Memory Required



At any time while PROC LOGISTIC is deriving the distributions, you can terminate the computations by pressing the system interrupt key sequence (see the SAS Companion for your system) and choosing to stop computations. If you run out of memory, see the SAS Companion for your system to see how to allocate more.

You can use the EXACTOPTIONS option MAXTIME= to limit the total amount of time PROC LOGISTIC uses to derive all of the exact distributions. If PROC LOGISTIC does not finish within that time, the procedure terminates.

Calculation of frequencies are performed in the log scale by default. This reduces the need to check for excessively large frequencies but can be slower than not scaling. You can turn off the log scaling by specifying the NOLOGSCALE option in the EXACTOPTIONS statement. If a frequency in the exact distribution is larger than the largest integer that can be held in double precision, a warning is printed to the SAS log. But

since inaccuracies due to adding small numbers to these large frequencies might have little or no effect on the statistics, the exact computations continue.

You can monitor the progress of the procedure by submitting your program with the EXACTOPTIONS option STATUSTIME=. If the procedure is too slow, you can try another method by specifying the EXACTOPTIONS option METHOD=, you can try reordering the variables in the MODEL statement (note that CLASS variables are always processed before continuous covariates), or you can try reparameterizing your classification variables as in the following statement:

```
class class-variables / param=ref ref=first order=freq;
```

Displayed Output

If you use the NOPRINT option in the PROC LOGISTIC statement, the procedure does not display any output. Otherwise, the tables displayed by the LOGISTIC procedure are discussed in the following section in the order in which they appear in the output. Some of the tables appear only in conjunction with certain options or statements; see the section "ODS Table Names" on page 4291 for details.

NOTE: The EFFECT, ESTIMATE, LSMEANS, LSMESTIMATE, and SLICE statements also create tables, which are not listed in this section. For information about these tables, see the corresponding sections of Chapter 19, "Shared Concepts and Topics."

Table Summary

Model Information and the Number of Observations

See the section "Missing Values" on page 4237 for information about missing-value handling, and the sections "FREQ Statement" on page 4203 and "WEIGHT Statement" on page 4236 for information about valid frequencies and weights.

Response Profile

Displays the Ordered Value assigned to each response level. See the section "Response Level Ordering" on page 4237 for details.

Class Level Information

Displays the design values for each CLASS explanatory variable. See the section "Other Parameterizations" on page 387 in Chapter 19, "Shared Concepts and Topics," for details.

Simple Statistics Tables

The following tables are displayed if you specify the SIMPLE option in the PROC LOGISTIC statement:

- **Descriptive Statistics for Continuous Explanatory Variables**
- **Frequency Distribution of Class Variables**
- **Weight Distribution of Class Variables**
  Displays if you also specify a WEIGHT statement.

Strata Tables for (Exact) Conditional Logistic Regression

The following tables are displayed if you specify a STRATA statement:

- **Strata Summary**
  Shows the pattern of the number of events and the number of nonevents in a stratum. See the section "STRATA Statement" on page 4233 for more information.
- **Strata Information**
  Displays if you specify the INFO option in a STRATA statement.

Maximum Likelihood Iteration History

Displays if you specify the ITPRINT option in the MODEL statement. See the sections "Iterative Algorithms for Model Fitting" on page 4240, "Convergence Criteria" on page 4242, and "Existence of Maximum Likelihood Estimates" on page 4242 for details.

Deviance and Pearson Goodness-of-Fit Statistics

Displays if you specify the SCALE= option in the MODEL statement. Small p-values reject the null hypothesis that the fitted model is adequate. See the section "Overdispersion" on page 4257 for details.

Score Test for the Equal Slopes (Proportional Odds) Assumption

Tests the parallel lines assumption if you fit an ordinal response model with the LINK=CLOGLOG or LINK=PROBIT options. If you specify LINK=LOGIT, this is called the "Proportional Odds" assumption. The table is not displayed if you specify the UNEQUALSLOPES option in the MODEL statement. Small p-values reject the null hypothesis that the slope parameters for each explanatory variable are constant across all the response functions. See the section "Testing the Parallel Lines Assumption" on page 4248 for details.

Model Fit Statistics

Computes various fit criteria based on a model with intercepts only and a model with intercepts and explanatory variables. If you specify the NOINT option in the MODEL statement, these statistics are calculated without considering the intercept parameters. See the section "Model Fitting Information" on page 4245 for details.

Testing Global Null Hypothesis: BETA=0

Tests the joint effect of the explanatory variables included in the model. Small p-values reject the null hypothesis that all slope parameters are equal to zero, $H_0: \boldsymbol{\beta} = \mathbf{0}$. See the sections "Model Fitting Information" on page 4245, "Residual Chi-Square" on page 4247, and "Testing Linear Hypotheses about the Regression Coefficients" on page 4262 for details. If you also specify the RSQUARE option in the MODEL statement, two generalized R Square measures are included; see the section "Generalized Coefficient of Determination" on page 4246 for details.

Score Test for Global Null Hypothesis

Displays instead of the "Testing Global Null Hypothesis: BETA=0" table if the NOFIT option is specified in the MODEL statement. The global score test evaluates the joint significance of the effects in the MODEL statement. Small p-values reject the null hypothesis that all slope parameters are equal to zero, $H_0: \boldsymbol{\beta} = \mathbf{0}$. See the section "Residual Chi-Square" on page 4247 for details.

Model Selection Tables

The tables in this section are produced when the SELECTION= option is specified in the MODEL statement. See the section "Effect-Selection Methods" on page 4244 for more information.

- **Residual Chi-Square Test**
  Displays if you specify SELECTION=FORWARD, BACKWARD, or STEPWISE in the MODEL statement. Small p-values reject the null hypothesis that the reduced model is adequate. See the section "Residual Chi-Square" on page 4247 for details.

- **Analysis of Effects Eligible for Entry**
  Displays if you specify the DETAILS option and the SELECTION=FORWARD or STEPWISE option in the MODEL statement. Small p-values reject $H_0: \beta_i \neq 0$. The score chi-square is used to determine entry; see the section "Testing Individual Effects Not in the Model" on page 4247 for details.

- **Analysis of Effects Eligible for Removal**
  Displays if you specify the SELECTION=BACKWARD or STEPWISE option in the MODEL statement. Small p-values reject $H_0: \beta_i = 0$. The Wald chi-square is used to determine removal; see the section "Testing Linear Hypotheses about the Regression Coefficients" on page 4262 for details.

- **Analysis of Effects Removed by Fast Backward Elimination**
  Displays if you specify the FAST option and the SELECTION=BACKWARD or STEPWISE option in the MODEL statement. This table gives the approximate chi-square statistic for the variable removed, the corresponding p-value with respect to a chi-square distribution with one degree of freedom, the residual chi-square statistic for testing the joint significance of the variable and the preceding ones, the degrees of freedom, and the p-value of the residual chi-square with respect to a chi-square distribution with the corresponding degrees of freedom.

- **Summary of Forward, Backward, and Stepwise Selection**
  Displays if you specify SELECTION=FORWARD, BACKWARD, or STEPWISE in the MODEL statement. The score chi-square is used to determine entry; see the section "Testing Individual Effects Not in the Model" on page 4247 for details. The Wald chi-square is used to determine removal; see the section "Testing Linear Hypotheses about the Regression Coefficients" on page 4262 for details.

- **Regression Models Selected by Score Criterion**
  Displays the score chi-square for all models if you specify the SELECTION=SCORE option in the MODEL statement. Small p-values reject the null hypothesis that the fitted model is adequate. See the section "Effect-Selection Methods" on page 4244 for details.

Type 3 Analysis of Effect

Displays if the model contains a CLASS variable. Performs Wald chi-square tests of the joint effect of the parameters for each CLASS variable in the model. Small p-values reject $H_0: \boldsymbol{\beta}_i = 0$. See the section "Testing Linear Hypotheses about the Regression Coefficients" on page 4262 for details.

Analysis of Maximum Likelihood Estimates

CLASS effects are identified by their (nonreference) level. For generalized logit models, a response variable column displays the nonreference level of the logit. The table includes the following:

- the estimated standard error of the parameter estimate, computed as the square root of the corresponding diagonal element of the estimated covariance matrix
- the Wald chi-square statistic, computed by squaring the ratio of the parameter estimate divided by its standard error estimate. See the section "Testing Linear Hypotheses about the Regression Coefficients" on page 4262 for details.
- the p-value tests the null hypothesis $H_0: \beta_i = 0$; small values reject the null.
- the standardized estimate for the slope parameter, if you specify the STB option in the MODEL statement. See the STB option on page 4220 for details.
- exponentiated values of the estimates of the slope parameters, if you specify the EXPB option in the MODEL statement. See the EXPB option on page 4212 for details.
- the label of the variable, if you specify the PARMLABEL option in the MODEL statement and if space permits. Due to constraints on the line size, the variable label might be suppressed in order to display the table in one panel. Use the SAS system option LINESIZE= to specify a larger line size to accommodate variable labels. A shorter line size can break the table into two panels allowing labels to be displayed.

Odds Ratio Estimates

Displays the odds ratio estimates and the corresponding 95% Wald confidence intervals for variables that are not involved in nestings or interactions. For continuous explanatory variables, these odds ratios correspond to a unit increase in the risk factors. See the section "Odds Ratio Estimation" on page 4250 for details.

Association of Predicted Probabilities and Observed Responses

See the section "Rank Correlation of Observed Responses and Predicted Probabilities" on page 4253 for details.

Parameter Estimates and Profile-Likelihood or Wald Confidence Intervals

Displays if you specify the CLPARM= option in the MODEL statement. See the section "Confidence Intervals for Parameters" on page 4248 for details.

Odds Ratio Estimates and Profile-Likelihood or Wald Confidence Intervals

Displays if you specify the ODDSRATIO statement for any effects with any class parameterizations. Also displays if you specify the CLODDS= option in the MODEL statement, except odds ratios are computed only for main effects not involved in interactions or nestings, and if the main effect is a CLASS variable, the parameterization must be EFFECT, REFERENCE, or GLM. See the section "Odds Ratio Estimation" on page 4250 for details.

Estimated Covariance or Correlation Matrix

Displays if you specify the COVB or CORRB option in the MODEL statement. See the section "Iterative Algorithms for Model Fitting" on page 4240 for details.

Contrast Test Results

Displays the Wald test for each specified CONTRAST statement. Small p-values reject $H_0: \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$. The "Coefficients of Contrast" table displays the contrast matrix if you specify the E option, and the "Contrast Estimation and Testing Results by Row" table displays estimates and Wald tests for each row of the contrast matrix if you specify the ESTIMATE= option. See the sections "CONTRAST Statement" on page 4190, "Testing Linear Hypotheses about the Regression Coefficients" on page 4262, and "Linear Predictor, Predicted Probability, and Confidence Limits" on page 4253 for details.

Linear Hypotheses Testing Results

Displays the Wald test for each specified TEST statement. See the sections "Testing Linear Hypotheses about the Regression Coefficients" on page 4262 and "TEST Statement" on page 4234 for details.

Hosmer and Lemeshow Goodness-of-Fit Test

Displays if you specify the LACKFIT option in the MODEL statement. Small p-values reject the null hypothesis that the fitted model is adequate. The "Partition for the Hosmer and Lemeshow Test" table displays the grouping used in the test. See the section "The Hosmer-Lemeshow Goodness-of-Fit Test" on page 4259 for details.

Classification Table

Displays if you use the CTABLE option in the MODEL statement. If you specify a list of cutpoints with the PPROB= option, then the cutpoints are displayed in the Prob Level column. If you specify the prior event probabilities with the PEVENT= option, then the probabilities are displayed in the Prob Event column. The Correct column displays the number of correctly classified events and nonevents, the Incorrect Event column displays the number of nonevents incorrectly classified as events, and the Incorrect Nonevent column gives the number of nonevents incorrectly classified as events. See the section "Classification Table" on page 4255 for more details.

Regression Diagnostics

Displays if you specify the INFLUENCE option in the MODEL statement. See the section "Regression Diagnostics" on page 4263 for more information about diagnostics from an unconditional analysis, and the section "Regression Diagnostic Details" on page 4272 for information about diagnostics from a conditional analysis.

Fit Statistics for SCORE Data

Displays if you specify the FITSTAT option in the SCORE statement. See the section "Scoring Data Sets" on page 4266 for details.

ROC Association Statistic and Contrast Tables

Displayed if a ROC statement or a ROCCONTRAST statement is specified. See the section "ROC Computations" on page 4261 for details about the Mann-Whitney statistics and the test and estimation computations, and see the section "Rank Correlation of Observed Responses and Predicted Probabilities" on page 4253 for details about the other statistics.

Exact Conditional Logistic Regression Tables

The tables in this section are produced when the EXACT statement is specified. If the METHOD=NETWORKMC option is specified, the test and estimate tables are renamed "Monte Carlo" tables and a Monte Carlo standard error column ($\sqrt{p(1-p)/n}$) is displayed.

- **Sufficient Statistics**
  Displays if you request an OUTDIST= data set in an EXACT statement. The table lists the parameters and their observed sufficient statistics.

- **(Monte Carlo) Conditional Exact Tests**
  See the section "Hypothesis Tests" on page 4276 for details.

- **(Monte Carlo) Exact Parameter Estimates**
  Displays if you specify the ESTIMATE option in the EXACT statement. This table gives individual parameter estimates for each variable (conditional on the values of all the other parameters in the model), confidence limits, and a two-sided *p*-value (twice the one-sided *p*-value) for testing that the parameter is zero. See the section "Inference for a Single Parameter" on page 4277 for details.

- **(Monte Carlo) Exact Odds Ratios**
  Displays if you specify the ESTIMATE=ODDS or ESTIMATE=BOTH option in the EXACT statement. See the section "Inference for a Single Parameter" on page 4277 for details.

ODS Table Names

PROC LOGISTIC 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 Table 54.13. For more information about ODS, see Chapter 20, "Using the Output Delivery System."

The EFFECT, ESTIMATE, LSMEANS, LSMESTIMATE, and SLICE statements also create tables, which are not listed in Table 54.13. For information about these tables, see the corresponding sections of Chapter 19, "Shared Concepts and Topics."

Table 54.13 ODS Tables Produced by PROC LOGISTIC

ODS Table Name	Description	Statement	Option
Association	Association of predicted probabilities and observed responses	MODEL (without STRATA)	Default
BestSubsets	Best subset selection	MODEL	SELECTION=SCORE
ClassFreq	Frequency breakdown of CLASS variables	PROC	Simple (with CLASS vars)
ClassLevelInfo	CLASS variable levels and design variables	MODEL	Default (with CLASS vars)
Classification	Classification table	MODEL	CTABLE
ClassWgt	Weight breakdown of CLASS variables	PROC, WEIGHT	Simple (with CLASS vars)
CLOddsPL	Odds ratio estimates and profile-likelihood confidence intervals	MODEL	CLODDS=PL
CLOddsWald	Odds ratio estimates and Wald confidence intervals	MODEL	CLODDS=WALD

Table 54.13 *continued*

ODS Table Name	Description	Statement	Option
CLParmPL	Parameter estimates and profile-likelihood confidence intervals	MODEL	CLPARM=PL
CLParmWald	Parameter estimates and Wald confidence intervals	MODEL	CLPARM=WALD
ContrastCoeff	L matrix from CONTRAST	CONTRAST	E
ContrastEstimate	Estimates from CONTRAST	CONTRAST	ESTIMATE=
ContrastTest	Wald test for CONTRAST	CONTRAST	Default
ConvergenceStatus	Convergence status	MODEL	Default
CorrB	Estimated correlation matrix of parameter estimators	MODEL	CORRB
CovB	Estimated covariance matrix of parameter estimators	MODEL	COVB
CumulativeModelTest	Test of the cumulative model assumption	MODEL	(Ordinal response)
EffectNotInModel	Test for effects not in model	MODEL	SELECTION=S|F
ExactOddsRatio	Exact odds ratios	EXACT	ESTIMATE=ODDS, ESTIMATE=BOTH
ExactParmEst	Parameter estimates	EXACT	ESTIMATE, ESTIMATE=PARM, ESTIMATE=BOTH
ExactTests	Conditional exact tests	EXACT	Default
FastElimination	Fast backward elimination	MODEL	SELECTION=B,FAST
FitStatistics	Model fit statistics	MODEL	Default
GlobalScore	Global score test	MODEL	NOFIT
GlobalTests	Test for global null hypothesis	MODEL	Default
GoodnessOfFit	Pearson and deviance goodness-of-fit tests	MODEL	SCALE
IndexPlots	Batch capture of the index plots	MODEL	IPLOTS
Influence	Regression diagnostics	MODEL	INFLUENCE
IterHistory	Iteration history	MODEL	ITPRINT
LackFitChiSq	Hosmer-Lemeshow chi-square test results	MODEL	LACKFIT
LackFitPartition	Partition for the Hosmer-Lemeshow test	MODEL	LACKFIT
LastGradient	Last evaluation of gradient	MODEL	ITPRINT
Linear	Linear combination	PROC	Default
LogLikeChange	Final change in the log likelihood	MODEL	ITPRINT
ModelBuildingSummary	Summary of model building	MODEL	SELECTION=B|F|S
ModelInfo	Model information	PROC	Default
NObs	Number of observations	PROC	Default

Table 54.13 *continued*

ODS Table Name	Description	Statement	Option
OddsEst	Adjusted odds ratios	UNITS	Default
OddsRatios	Odds ratio estimates	MODEL	Default
OddsRatiosWald	Odds ratio estimates and Wald confidence intervals	ODDSRATIOS	CL=WALD
OddsRatiosPL	Odds ratio estimates and PL confidence intervals	ODDSRATIOS	CL=PL
ParameterEstimates	Maximum likelihood estimates of model parameters	MODEL	Default
RSquare	R-square	MODEL	RSQUARE
ResidualChiSq	Residual chi-square	MODEL	SELECTION=F|B
ResponseProfile	Response profile	PROC	Default
ROCAssociation	Association table for ROC models	ROC	Default
ROCContrastCoeff	L matrix from ROCCONTRAST	ROCCONTRAST	E
ROCContrastCov	Covariance of ROCCONTRAST rows	ROCCONTRAST	COV
ROCContrastEstimate	Estimates from ROCCONTRAST	ROCCONTRAST	ESTIMATE=
ROCContrastTest	Wald test from ROCCONTRAST	ROCCONTRAST	Default
ROCCov	Covariance between ROC curves	ROCCONTRAST	COV
ScoreFitStat	Fit statistics for Scored data	SCORE	FITSTAT
SimpleStatistics	Summary statistics for explanatory variables	PROC	SIMPLE
StrataSummary	Number of strata with specific response frequencies	STRATA	Default
StrataInfo	Event and nonevent frequencies for each stratum	STRATA	INFO
SuffStats	Sufficient statistics	EXACT	OUTDIST=
TestPrint1	**L**[Cov(**b**)]**L**' and **Lb**-**c**	TEST	PRINT
TestPrint2	Ginv(**L**[Cov(**b**)]**L**') and Ginv(**L**[Cov(**b**)]**L**')(**Lb**-**c**)	TEST	PRINT
TestStmts	Linear hypotheses testing results	TEST	Default
Type3	Type 3 tests of effects	MODEL	Default (with CLASS variables)

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 600 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 599 in Chapter 21, "Statistical Graphics Using ODS."

You must also specify the options in the PROC LOGISTIC statement that are indicated in Table 54.14.

When ODS Graphics is enabled, then the EFFECT, EFFECTPLOT, ESTIMATE, LSMEANS, LSMESTIMATE, and SLICE statements can produce plots that are associated with their analyses. For information about these plots, see the corresponding sections of Chapter 19, "Shared Concepts and Topics."

PROC LOGISTIC 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 54.14.

Table 54.14 Graphs Produced by PROC LOGISTIC

ODS Graph Name	Plot Description	Statement or Option
DfBetasPlot	Panel of dfbetas by case number	PLOTS=DFBETAS or MODEL / INFLUENCE or IPLOTS
	Effect dfbetas by case number	PLOTS=DFBETAS(UNPACK)
DPCPlot	Difchisq and/or difdev by predicted probability by CI displacement C	PLOTS=DPC
EffectPlot	Predicted probability	PLOTS=EFFECT
InfluencePlots	Panel of influence statistics by case number	PLOTS=INFLUENCE or MODEL / INFLUENCE or IPLOTS
CBarPlot	CI displacement Cbar by case number	PLOTS=INFLUENCE(UNPACK)
CPlot	CI displacement C by case number	PLOTS=INFLUENCE(UNPACK)
DevianceResidualPlot	Deviance residual by case number	PLOTS=INFLUENCE(UNPACK)
DifChisqPlot	Difchisq by case number	PLOTS=INFLUENCE(UNPACK)
DifDeviancePlot	Difdev by case number	PLOTS=INFLUENCE(UNPACK)
LeveragePlot	Hat diagonal by case number	PLOTS=INFLUENCE(UNPACK)
LikelihoodResidualPlot	Likelihood residual by case number	PLOTS=INFLUENCE(UNPACK STDRES)
PearsonResidualPlot	Pearson chi-square residual by case number	PLOTS=INFLUENCE(UNPACK)
StdDevianceResidualPlot	Standardized deviance residual by case number	PLOTS=INFLUENCE(UNPACK STDRES)
StdPearsonResidualPlot	Standardized Pearson chi-square residual by case number	PLOTS=INFLUENCE(UNPACK STDRES)

Table 54.14 *continued*

ODS Graph Name	Plot Description	Statement or Option
LeveragePlots	Panel of influence statistics by leverage	PLOTS=LEVERAGE
LeverageCPlot	CI displacement C by leverage	PLOTS=LEVERAGE(UNPACK)
LeverageDifChisqPlot	Difchisq by leverage	PLOTS=LEVERAGE(UNPACK)
LeverageDifDevPlot	Difdev by leverage	PLOTS=LEVERAGE(UNPACK)
LeveragePhatPlot	Predicted probability by leverage	PLOTS=LEVERAGE(UNPACK)
ORPlot	Odds ratios	Default or PLOTS=ODDSRATIO and MODEL / CLODDS= or ODDSRATIO
PhatPlots	Panel of influence by predicted probability	PLOTS=PHAT
PhatCPlot	CI displacement C by predicted probability	PLOTS=PHAT(UNPACK)
PhatDifChisqPlot	Difchisq by predicted probability	PLOTS=PHAT(UNPACK)
PhatDifDevPlot	Difdev by predicted probability	PLOTS=PHAT(UNPACK)
PhatLeveragePlot	Leverage by predicted probability	PLOTS=PHAT(UNPACK)
ROCCurve	Receiver operating characteristics curve	PLOTS=ROC or MODEL / OUTROC= or SCORE OUTROC= or ROC
ROCOverlay	ROC curves for comparisons	PLOTS=ROC and MODEL / SELECTION= or ROC

Examples: LOGISTIC Procedure

Example 54.1: Stepwise Logistic Regression and Predicted Values

Consider a study on cancer remission (Lee 1974). The data consist of patient characteristics and whether or not cancer remission occurred. The following DATA step creates the data set Remission containing seven variables. The variable remiss is the cancer remission indicator variable with a value of 1 for remission and a value of 0 for nonremission. The other six variables are the risk factors thought to be related to cancer remission.

```
data Remission;
   input remiss cell smear infil li blast temp;
   label remiss='Complete Remission';
   datalines;
1   .8   .83  .66  1.9  1.1     .996
1   .9   .36  .32  1.4   .74    .992
```

```
0   .8   .88  .7    .8   .176   .982
0  1     .87  .87   .7  1.053   .986
1   .9   .75  .68  1.3   .519   .98
0  1     .65  .65   .6   .519   .982
1   .95  .97  .92  1    1.23    .992
0   .95  .87  .83  1.9  1.354  1.02
0  1     .45  .45   .8   .322   .999
0   .95  .36  .34   .5  0      1.038
0   .85  .39  .33   .7   .279   .988
0   .7   .76  .53  1.2   .146   .982
0   .8   .46  .37   .4   .38   1.006
0   .2   .39  .08   .8   .114   .99
0  1     .9   .9   1.1  1.037   .99
1  1     .84  .84  1.9  2.064  1.02
0   .65  .42  .27   .5   .114  1.014
0  1     .75  .75  1    1.322  1.004
0   .5   .44  .22   .6   .114   .99
1  1     .63  .63  1.1  1.072   .986
0  1     .33  .33   .4   .176  1.01
0   .9   .93  .84   .6  1.591  1.02
1  1     .58  .58  1     .531  1.002
0   .95  .32  .3   1.6   .886   .988
1  1     .6   .6   1.7   .964   .99
1  1     .69  .69   .9   .398   .986
0  1     .73  .73   .7   .398   .986
;
```

The following invocation of PROC LOGISTIC illustrates the use of stepwise selection to identify the prognostic factors for cancer remission. A significance level of 0.3 is required to allow a variable into the model (SLENTRY=0.3), and a significance level of 0.35 is required for a variable to stay in the model (SLSTAY=0.35). A detailed account of the variable selection process is requested by specifying the DETAILS option. The Hosmer and Lemeshow goodness-of-fit test for the final selected model is requested by specifying the LACKFIT option. The OUTEST= and COVOUT options in the PROC LOGISTIC statement create a data set that contains parameter estimates and their covariances for the final selected model. The response variable option EVENT= chooses remiss=1 (remission) as the event so that the probability of remission is modeled. The OUTPUT statement creates a data set that contains the cumulative predicted probabilities and the corresponding confidence limits, and the individual and cross validated predicted probabilities for each observation.

```
title 'Stepwise Regression on Cancer Remission Data';
proc logistic data=Remission outest=betas covout;
   model remiss(event='1')=cell smear infil li blast temp
                / selection=stepwise
                  slentry=0.3
                  slstay=0.35
                  details
                  lackfit;
   output out=pred p=phat lower=lcl upper=ucl
          predprob=(individual crossvalidate);
run;
```

```
proc print data=betas;
   title2 'Parameter Estimates and Covariance Matrix';
run;

proc print data=pred;
   title2 'Predicted Probabilities and 95% Confidence Limits';
run;
```

In stepwise selection, an attempt is made to remove any insignificant variables from the model before adding a significant variable to the model. Each addition or deletion of a variable to or from a model is listed as a separate step in the displayed output, and at each step a new model is fitted. Details of the model selection steps are shown in Outputs 54.1.1 through 54.1.5.

Prior to the first step, the intercept-only model is fit and individual score statistics for the potential variables are evaluated (Output 54.1.1).

Output 54.1.1 Startup Model

```
                Stepwise Regression on Cancer Remission Data

                           The LOGISTIC Procedure

Step  0. Intercept entered:


                          Model Convergence Status

                Convergence criterion (GCONV=1E-8) satisfied.


                             -2 Log L = 34.372


                  Analysis of Maximum Likelihood Estimates

                                      Standard          Wald
      Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq

      Intercept     1     -0.6931      0.4082        2.8827        0.0895


                          Residual Chi-Square Test

                     Chi-Square       DF     Pr > ChiSq

                         9.4609        6         0.1493
```

Output 54.1.1 *continued*

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
cell	1	1.8893	0.1693
smear	1	1.0745	0.2999
infil	1	1.8817	0.1701
li	1	7.9311	0.0049
blast	1	3.5258	0.0604
temp	1	0.6591	0.4169

In Step 1 (Output 54.1.2), the variable li is selected into the model since it is the most significant variable among those to be chosen ($p = 0.0049 < 0.3$). The intermediate model that contains an intercept and li is then fitted. li remains significant ($p = 0.0146 < 0.35$) and is not removed.

Output 54.1.2 Step 1 of the Stepwise Analysis

Step 1. Effect li entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	36.372	30.073
SC	37.668	32.665
-2 Log L	34.372	26.073

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.2988	1	0.0040
Score	7.9311	1	0.0049
Wald	5.9594	1	0.0146

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.7771	1.3786	7.5064	0.0061
li	1	2.8973	1.1868	5.9594	0.0146

Output 54.1.2 *continued*

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
li	18.124	1.770	185.563

Association of Predicted Probabilities and Observed Responses

Percent Concordant	84.0	Somers' D	0.710
Percent Discordant	13.0	Gamma	0.732
Percent Tied	3.1	Tau-a	0.328
Pairs	162	c	0.855

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
3.1174	5	0.6819

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
li	1	5.9594	0.0146

NOTE: No effects for the model in Step 1 are removed.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
cell	1	1.1183	0.2903
smear	1	0.1369	0.7114
infil	1	0.5715	0.4497
blast	1	0.0932	0.7601
temp	1	1.2591	0.2618

In Step 2 (Output 54.1.3), the variable temp is added to the model. The model then contains an intercept and the variables li and temp. Both li and temp remain significant at 0.35 level; therefore, neither li nor temp is removed from the model.

Output 54.1.3 Step 2 of the Stepwise Analysis

Step 2. Effect temp entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	36.372	30.648
SC	37.668	34.535
-2 Log L	34.372	24.648

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.7239	2	0.0077
Score	8.3648	2	0.0153
Wald	5.9052	2	0.0522

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	47.8448	46.4381	1.0615	0.3029
li	1	3.3017	1.3593	5.9002	0.0151
temp	1	-52.4214	47.4897	1.2185	0.2697

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
li	27.158	1.892	389.856
temp	<0.001	<0.001	>999.999

Association of Predicted Probabilities and Observed Responses

Percent Concordant	87.0	Somers' D	0.747
Percent Discordant	12.3	Gamma	0.752
Percent Tied	0.6	Tau-a	0.345
Pairs	162	c	0.873

Output 54.1.3 *continued*

```
                    Residual Chi-Square Test

                 Chi-Square       DF     Pr > ChiSq

                     2.1429        4         0.7095


               Analysis of Effects Eligible for Removal

                                        Wald
              Effect      DF    Chi-Square    Pr > ChiSq

              li           1        5.9002        0.0151
              temp         1        1.2185        0.2697


NOTE: No effects for the model in Step 2 are removed.


                 Analysis of Effects Eligible for Entry

                                       Score
              Effect      DF    Chi-Square    Pr > ChiSq

              cell         1        1.4700        0.2254
              smear        1        0.1730        0.6775
              infil        1        0.8274        0.3630
              blast        1        1.1013        0.2940
```

In Step 3 (Output 54.1.4), the variable cell is added to the model. The model then contains an intercept and the variables li, temp, and cell. None of these variables are removed from the model since all are significant at the 0.35 level.

Output 54.1.4 Step 3 of the Stepwise Analysis

```
Step  3. Effect cell entered:


                        Model Convergence Status

             Convergence criterion (GCONV=1E-8) satisfied.


                          Model Fit Statistics

                                                 Intercept
                                    Intercept          and
                Criterion                Only    Covariates

                AIC                    36.372        29.953
                SC                     37.668        35.137
                -2 Log L               34.372        21.953
```

Output 54.1.4 *continued*

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12.4184	3	0.0061
Score	9.2502	3	0.0261
Wald	4.8281	3	0.1848

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	67.6339	56.8875	1.4135	0.2345
cell	1	9.6521	7.7511	1.5507	0.2130
li	1	3.8671	1.7783	4.7290	0.0297
temp	1	-82.0737	61.7124	1.7687	0.1835

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
cell	>999.999	0.004	>999.999
li	47.804	1.465	>999.999
temp	<0.001	<0.001	>999.999

Association of Predicted Probabilities and Observed Responses

Percent Concordant	88.9	Somers' D	0.778
Percent Discordant	11.1	Gamma	0.778
Percent Tied	0.0	Tau-a	0.359
Pairs	162	c	0.889

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.1831	3	0.9803

Analysis of Effects Eligible for Removal

Effect	DF	Wald Chi-Square	Pr > ChiSq
cell	1	1.5507	0.2130
li	1	4.7290	0.0297
temp	1	1.7687	0.1835

Output 54.1.4 *continued*

NOTE: No effects for the model in Step 3 are removed.

Analysis of Effects Eligible for Entry

Effect	DF	Score Chi-Square	Pr > ChiSq
smear	1	0.0956	0.7572
infil	1	0.0844	0.7714
blast	1	0.0208	0.8852

Finally, none of the remaining variables outside the model meet the entry criterion, and the stepwise selection is terminated. A summary of the stepwise selection is displayed in Output 54.1.5.

Output 54.1.5 Summary of the Stepwise Selection

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	li		1	1	7.9311		0.0049
2	temp		1	2	1.2591		0.2618
3	cell		1	3	1.4700		0.2254

Results of the Hosmer and Lemeshow test are shown in Output 54.1.6. There is no evidence of a lack of fit in the selected model ($p = 0.5054$).

Output 54.1.6 Display of the LACKFIT Option

Partition for the Hosmer and Lemeshow Test

Group	Total	remiss = 1 Observed	remiss = 1 Expected	remiss = 0 Observed	remiss = 0 Expected
1	3	0	0.00	3	3.00
2	3	0	0.01	3	2.99
3	3	0	0.19	3	2.81
4	3	0	0.56	3	2.44
5	4	1	1.09	3	2.91
6	3	2	1.35	1	1.65
7	3	2	1.84	1	1.16
8	3	3	2.15	0	0.85
9	2	1	1.80	1	0.20

Output 54.1.6 *continued*

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
6.2983	7	0.5054

The data set betas created by the OUTEST= and COVOUT options is displayed in Output 54.1.7. The data set contains parameter estimates and the covariance matrix for the final selected model. Note that all explanatory variables listed in the MODEL statement are included in this data set; however, variables that are not included in the final model have all missing values.

Output 54.1.7 Data Set of Estimates and Covariances

Stepwise Regression on Cancer Remission Data
Parameter Estimates and Covariance Matrix

Obs	_LINK_	_TYPE_	_STATUS_	_NAME_	Intercept	cell
1	LOGIT	PARMS	0 Converged	remiss	67.63	9.652
2	LOGIT	COV	0 Converged	Intercept	3236.19	157.097
3	LOGIT	COV	0 Converged	cell	157.10	60.079
4	LOGIT	COV	0 Converged	smear	.	.
5	LOGIT	COV	0 Converged	infil	.	.
6	LOGIT	COV	0 Converged	li	64.57	6.945
7	LOGIT	COV	0 Converged	blast	.	.
8	LOGIT	COV	0 Converged	temp	-3483.23	-223.669

Obs	smear	infil	li	blast	temp	_LNLIKE_	_ESTTYPE_
1	.	.	3.8671	.	-82.07	-10.9767	MLE
2	.	.	64.5726	.	-3483.23	-10.9767	MLE
3	.	.	6.9454	.	-223.67	-10.9767	MLE
4	.	.	.	.	.	-10.9767	MLE
5	.	.	.	.	.	-10.9767	MLE
6	.	.	3.1623	.	-75.35	-10.9767	MLE
7	.	.	.	.	.	-10.9767	MLE
8	.	.	-75.3513	.	3808.42	-10.9767	MLE

The data set pred created by the OUTPUT statement is displayed in Output 54.1.8. It contains all the variables in the input data set, the variable phat for the (cumulative) predicted probability, the variables lcl and ucl for the lower and upper confidence limits for the probability, and four other variables (IP_1, IP_0, XP_1, and XP_0) for the PREDPROBS= option. The data set also contains the variable _LEVEL_, indicating the response value to which phat, lcl, and ucl refer. For instance, for the first row of the OUTPUT data set, the values of _LEVEL_ and phat, lcl, and ucl are 1, 0.72265, 0.16892, and 0.97093, respectively; this means that the estimated probability that remiss=1 is 0.723 for the given explanatory variable values, and the corresponding 95% confidence interval is (0.16892, 0.97093). The variables IP_1 and IP_0 contain the predicted probabilities that remiss=1 and remiss=0, respectively. Note that values of phat and IP_1 are identical since they both contain the probabilities that remiss=1. The variables XP_1 and XP_0 contain the cross validated predicted probabilities that remiss=1 and remiss=0, respectively.

Output 54.1.8 Predicted Probabilities and Confidence Intervals

Stepwise Regression on Cancer Remission Data
Predicted Probabilities and 95% Confidence Limits

Obs	remiss	cell	smear	infil	li	blast	temp	_FROM_	_INTO_	IP_0
1	1	0.80	0.83	0.66	1.9	1.100	0.996	1	1	0.27735
2	1	0.90	0.36	0.32	1.4	0.740	0.992	1	1	0.42126
3	0	0.80	0.88	0.70	0.8	0.176	0.982	0	0	0.89540
4	0	1.00	0.87	0.87	0.7	1.053	0.986	0	0	0.71742
5	1	0.90	0.75	0.68	1.3	0.519	0.980	1	1	0.28582
6	0	1.00	0.65	0.65	0.6	0.519	0.982	0	0	0.72911
7	1	0.95	0.97	0.92	1.0	1.230	0.992	1	0	0.67844
8	0	0.95	0.87	0.83	1.9	1.354	1.020	0	1	0.39277
9	0	1.00	0.45	0.45	0.8	0.322	0.999	0	0	0.83368
10	0	0.95	0.36	0.34	0.5	0.000	1.038	0	0	0.99843
11	0	0.85	0.39	0.33	0.7	0.279	0.988	0	0	0.92715
12	0	0.70	0.76	0.53	1.2	0.146	0.982	0	0	0.82714

Obs	IP_1	XP_0	XP_1	_LEVEL_	phat	lcl	ucl
1	0.72265	0.43873	0.56127	1	0.72265	0.16892	0.97093
2	0.57874	0.47461	0.52539	1	0.57874	0.26788	0.83762
3	0.10460	0.87060	0.12940	1	0.10460	0.00781	0.63419
4	0.28258	0.67259	0.32741	1	0.28258	0.07498	0.65683
5	0.71418	0.36901	0.63099	1	0.71418	0.25218	0.94876
6	0.27089	0.67269	0.32731	1	0.27089	0.05852	0.68951
7	0.32156	0.72923	0.27077	1	0.32156	0.13255	0.59516
8	0.60723	0.09906	0.90094	1	0.60723	0.10572	0.95287
9	0.16632	0.80864	0.19136	1	0.16632	0.03018	0.56123
10	0.00157	0.99840	0.00160	1	0.00157	0.00000	0.68962
11	0.07285	0.91723	0.08277	1	0.07285	0.00614	0.49982
12	0.17286	0.63838	0.36162	1	0.17286	0.00637	0.87206

Output 54.1.8 *continued*

Stepwise Regression on Cancer Remission Data
Predicted Probabilities and 95% Confidence Limits

Obs	remiss	cell	smear	infil	li	blast	temp	_FROM_	_INTO_	IP_0
13	0	0.80	0.46	0.37	0.4	0.380	1.006	0	0	0.99654
14	0	0.20	0.39	0.08	0.8	0.114	0.990	0	0	0.99982
15	0	1.00	0.90	0.90	1.1	1.037	0.990	0	1	0.42878
16	1	1.00	0.84	0.84	1.9	2.064	1.020	1	1	0.28530
17	0	0.65	0.42	0.27	0.5	0.114	1.014	0	0	0.99938
18	0	1.00	0.75	0.75	1.0	1.322	1.004	0	0	0.77711
19	0	0.50	0.44	0.22	0.6	0.114	0.990	0	0	0.99846
20	1	1.00	0.63	0.63	1.1	1.072	0.986	1	1	0.35089
21	0	1.00	0.33	0.33	0.4	0.176	1.010	0	0	0.98307
22	0	0.90	0.93	0.84	0.6	1.591	1.020	0	0	0.99378
23	1	1.00	0.58	0.58	1.0	0.531	1.002	1	0	0.74739
24	0	0.95	0.32	0.30	1.6	0.886	0.988	0	1	0.12989

Obs	IP_1	XP_0	XP_1	_LEVEL_	phat	lcl	ucl
13	0.00346	0.99644	0.00356	1	0.00346	0.00001	0.46530
14	0.00018	0.99981	0.00019	1	0.00018	0.00000	0.96482
15	0.57122	0.35354	0.64646	1	0.57122	0.25303	0.83973
16	0.71470	0.47213	0.52787	1	0.71470	0.15362	0.97189
17	0.00062	0.99937	0.00063	1	0.00062	0.00000	0.62665
18	0.22289	0.73612	0.26388	1	0.22289	0.04483	0.63670
19	0.00154	0.99842	0.00158	1	0.00154	0.00000	0.79644
20	0.64911	0.42053	0.57947	1	0.64911	0.26305	0.90555
21	0.01693	0.98170	0.01830	1	0.01693	0.00029	0.50475
22	0.00622	0.99348	0.00652	1	0.00622	0.00003	0.56062
23	0.25261	0.84423	0.15577	1	0.25261	0.06137	0.63597
24	0.87011	0.03637	0.96363	1	0.87011	0.40910	0.98481

Stepwise Regression on Cancer Remission Data
Predicted Probabilities and 95% Confidence Limits

Obs	remiss	cell	smear	infil	li	blast	temp	_FROM_	_INTO_	IP_0
25	1	1.00	0.60	0.60	1.7	0.964	0.990	1	1	0.06868
26	1	1.00	0.69	0.69	0.9	0.398	0.986	1	0	0.53949
27	0	1.00	0.73	0.73	0.7	0.398	0.986	0	0	0.71742

Obs	IP_1	XP_0	XP_1	_LEVEL_	phat	lcl	ucl
25	0.93132	0.08017	0.91983	1	0.93132	0.44114	0.99573
26	0.46051	0.62312	0.37688	1	0.46051	0.16612	0.78529
27	0.28258	0.67259	0.32741	1	0.28258	0.07498	0.65683

Next, a different variable selection method is used to select prognostic factors for cancer remission, and an efficient algorithm is employed to eliminate insignificant variables from a model. The following statements invoke PROC LOGISTIC to perform the backward elimination analysis:

```
title 'Backward Elimination on Cancer Remission Data';
proc logistic data=Remission;
   model remiss(event='1')=temp cell li smear blast
         / selection=backward fast slstay=0.2 ctable;
run;
```

The backward elimination analysis (SELECTION=BACKWARD) starts with a model that contains all explanatory variables given in the MODEL statement. By specifying the FAST option, PROC LOGISTIC eliminates insignificant variables without refitting the model repeatedly. This analysis uses a significance level of 0.2 to retain variables in the model (SLSTAY=0.2), which is different from the previous stepwise analysis where SLSTAY=.35. The CTABLE option is specified to produce classifications of input observations based on the final selected model.

Results of the fast elimination analysis are shown in Output 54.1.9 and Output 54.1.10. Initially, a full model containing all six risk factors is fit to the data (Output 54.1.9). In the next step (Output 54.1.10), PROC LOGISTIC removes blast, smear, cell, and temp from the model all at once. This leaves li and the intercept as the only variables in the final model. Note that in this analysis, only parameter estimates for the final model are displayed because the DETAILS option has not been specified.

Output 54.1.9 Initial Step in Backward Elimination

```
            Backward Elimination on Cancer Remission Data

                        The LOGISTIC Procedure

                          Model Information

Data Set                        WORK.REMISSION
Response Variable               remiss                Complete Remission
Number of Response Levels       2
Model                           binary logit
Optimization Technique          Fisher's scoring


               Number of Observations Read          27
               Number of Observations Used          27


                        Response Profile

                 Ordered                        Total
                   Value       remiss       Frequency

                       1            0              18
                       2            1               9

                  Probability modeled is remiss=1.
```

Output 54.1.9 *continued*

Backward Elimination Procedure

Step 0. The following effects were entered:

Intercept temp cell li smear blast

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	36.372	33.857
SC	37.668	41.632
-2 Log L	34.372	21.857

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12.5146	5	0.0284
Score	9.3295	5	0.0966
Wald	4.7284	5	0.4499

Output 54.1.10 Fast Elimination Step

Step 1. Fast Backward Elimination:

Analysis of Effects Removed by Fast Backward Elimination

Effect Removed	Chi-Square	DF	Pr > ChiSq	Residual Chi-Square	DF	Pr > Residual ChiSq
blast	0.0008	1	0.9768	0.0008	1	0.9768
smear	0.0951	1	0.7578	0.0959	2	0.9532
cell	1.5134	1	0.2186	1.6094	3	0.6573
temp	0.6535	1	0.4189	2.2628	4	0.6875

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Output 54.1.10 *continued*

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	36.372	30.073
SC	37.668	32.665
-2 Log L	34.372	26.073

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.2988	1	0.0040
Score	7.9311	1	0.0049
Wald	5.9594	1	0.0146

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
2.8530	4	0.5827

Summary of Backward Elimination

Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	blast	1	4	0.0008	0.9768
1	smear	1	3	0.0951	0.7578
1	cell	1	2	1.5134	0.2186
1	temp	1	1	0.6535	0.4189

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.7771	1.3786	7.5064	0.0061
li	1	2.8973	1.1868	5.9594	0.0146

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
li	18.124	1.770	185.563

Output 54.1.10 *continued*

Association of Predicted Probabilities and Observed Responses

Percent Concordant	84.0	Somers' D	0.710
Percent Discordant	13.0	Gamma	0.732
Percent Tied	3.1	Tau-a	0.328
Pairs	162	c	0.855

Note that you can also use the FAST option when SELECTION=STEPWISE. However, the FAST option operates only on backward elimination steps. In this example, the stepwise process only adds variables, so the FAST option would not be useful.

Results of the CTABLE option are shown in Output 54.1.11.

Output 54.1.11 Classifying Input Observations

Classification Table

Prob Level	Correct Event	Correct Non-Event	Incorrect Event	Incorrect Non-Event	Percentages Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.060	9	0	18	0	33.3	100.0	0.0	66.7	.
0.080	9	2	16	0	40.7	100.0	11.1	64.0	0.0
0.100	9	4	14	0	48.1	100.0	22.2	60.9	0.0
0.120	9	4	14	0	48.1	100.0	22.2	60.9	0.0
0.140	9	7	11	0	59.3	100.0	38.9	55.0	0.0
0.160	9	10	8	0	70.4	100.0	55.6	47.1	0.0
0.180	9	10	8	0	70.4	100.0	55.6	47.1	0.0
0.200	8	13	5	1	77.8	88.9	72.2	38.5	7.1
0.220	8	13	5	1	77.8	88.9	72.2	38.5	7.1
0.240	8	13	5	1	77.8	88.9	72.2	38.5	7.1
0.260	6	13	5	3	70.4	66.7	72.2	45.5	18.8
0.280	6	13	5	3	70.4	66.7	72.2	45.5	18.8
0.300	6	13	5	3	70.4	66.7	72.2	45.5	18.8
0.320	6	14	4	3	74.1	66.7	77.8	40.0	17.6
0.340	5	14	4	4	70.4	55.6	77.8	44.4	22.2
0.360	5	14	4	4	70.4	55.6	77.8	44.4	22.2
0.380	5	15	3	4	74.1	55.6	83.3	37.5	21.1
0.400	5	15	3	4	74.1	55.6	83.3	37.5	21.1
0.420	5	15	3	4	74.1	55.6	83.3	37.5	21.1
0.440	5	15	3	4	74.1	55.6	83.3	37.5	21.1
0.460	4	16	2	5	74.1	44.4	88.9	33.3	23.8
0.480	4	16	2	5	74.1	44.4	88.9	33.3	23.8
0.500	4	16	2	5	74.1	44.4	88.9	33.3	23.8
0.520	4	16	2	5	74.1	44.4	88.9	33.3	23.8
0.540	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.560	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.580	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.600	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.620	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.640	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.660	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.680	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.700	3	16	2	6	70.4	33.3	88.9	40.0	27.3
0.720	2	16	2	7	66.7	22.2	88.9	50.0	30.4
0.740	2	16	2	7	66.7	22.2	88.9	50.0	30.4
0.760	2	16	2	7	66.7	22.2	88.9	50.0	30.4
0.780	2	16	2	7	66.7	22.2	88.9	50.0	30.4
0.800	2	17	1	7	70.4	22.2	94.4	33.3	29.2
0.820	2	17	1	7	70.4	22.2	94.4	33.3	29.2
0.840	0	17	1	9	63.0	0.0	94.4	100.0	34.6
0.860	0	17	1	9	63.0	0.0	94.4	100.0	34.6
0.880	0	17	1	9	63.0	0.0	94.4	100.0	34.6
0.900	0	17	1	9	63.0	0.0	94.4	100.0	34.6
0.920	0	17	1	9	63.0	0.0	94.4	100.0	34.6
0.940	0	17	1	9	63.0	0.0	94.4	100.0	34.6
0.960	0	18	0	9	66.7	0.0	100.0	.	33.3

Each row of the "Classification Table" corresponds to a cutpoint applied to the predicted probabilities, which is given in the Prob Level column. The 2×2 frequency tables of observed and predicted responses are given by the next four columns. For example, with a cutpoint of 0.5, 4 events and 16 nonevents were classified correctly. On the other hand, 2 nonevents were incorrectly classified as events and 5 events were incorrectly classified as nonevents. For this cutpoint, the correct classification rate is 20/27 (=74.1%), which is given in the sixth column. Accuracy of the classification is summarized by the sensitivity, specificity, and false positive and negative rates, which are displayed in the last four columns. You can control the number of cutpoints used, and their values, by using the PPROB= option.

Example 54.2: Logistic Modeling with Categorical Predictors

Consider a study of the analgesic effects of treatments on elderly patients with neuralgia. Two test treatments and a placebo are compared. The response variable is whether the patient reported pain or not. Researchers recorded the age and gender of 60 patients and the duration of complaint before the treatment began. The following DATA step creates the data set Neuralgia:

```
data Neuralgia;
   input Treatment $ Sex $ Age Duration Pain $ @@;
   datalines;
P  F  68   1  No   B  M  74  16  No  P  F  67  30  No
P  M  66  26  Yes  B  F  67  28  No  B  F  77  16  No
A  F  71  12  No   B  F  72  50  No  B  F  76   9  Yes
A  M  71  17  Yes  A  F  63  27  No  A  F  69  18  Yes
B  F  66  12  No   A  M  62  42  No  P  F  64   1  Yes
A  F  64  17  No   P  M  74   4  No  A  F  72  25  No
P  M  70   1  Yes  B  M  66  19  No  B  M  59  29  No
A  F  64  30  No   A  M  70  28  No  A  M  69   1  No
B  F  78   1  No   P  M  83   1  Yes B  F  69  42  No
B  M  75  30  Yes  P  M  77  29  Yes P  F  79  20  Yes
A  M  70  12  No   A  F  69  12  No  B  F  65  14  No
B  M  70   1  No   B  M  67  23  No  A  M  76  25  Yes
P  M  78  12  Yes  B  M  77   1  Yes B  F  69  24  No
P  M  66   4  Yes  P  F  65  29  No  P  M  60  26  Yes
A  M  78  15  Yes  B  M  75  21  Yes A  F  67  11  No
P  F  72  27  No   P  F  70  13  Yes A  M  75   6  Yes
B  F  65   7  No   P  F  68  27  Yes P  M  68  11  Yes
P  M  67  17  Yes  B  M  70  22  No  A  M  65  15  No
P  F  67   1  Yes  A  M  67  10  No  P  F  72  11  Yes
A  F  74   1  No   B  M  80  21  Yes A  F  69   3  No
;
```

The data set Neuralgia contains five variables: Treatment, Sex, Age, Duration, and Pain. The last variable, Pain, is the response variable. A specification of Pain=Yes indicates there was pain, and Pain=No indicates no pain. The variable Treatment is a categorical variable with three levels: A and B represent the two test treatments, and P represents the placebo treatment. The gender of the patients is given by the categorical variable Sex. The variable Age is the age of the patients, in years, when treatment began. The duration of complaint, in months, before the treatment began is given by the variable Duration.

The following statements use the LOGISTIC procedure to fit a two-way logit with interaction model for the effect of Treatment and Sex, with Age and Duration as covariates. The categorical variables Treatment and

Sex are declared in the CLASS statement.

```
proc logistic data=Neuralgia;
   class Treatment Sex;
   model Pain= Treatment Sex Treatment*Sex Age Duration / expb;
run;
```

In this analysis, PROC LOGISTIC models the probability of no pain (Pain=No). By default, effect coding is used to represent the CLASS variables. Two design variables are created for Treatment and one for Sex, as shown in Output 54.2.1.

Output 54.2.1 Effect Coding of CLASS Variables

The LOGISTIC Procedure

Class Level Information

Class	Value	Design Variables	
Treatment	A	1	0
	B	0	1
	P	-1	-1
Sex	F	1	
	M	-1	

PROC LOGISTIC displays a table of the Type 3 analysis of effects based on the Wald test (Output 54.2.2). Note that the Treatment*Sex interaction and the duration of complaint are not statistically significant (p = 0.9318 and p = 0.8752, respectively). This indicates that there is no evidence that the treatments affect pain differently in men and women, and no evidence that the pain outcome is related to the duration of pain.

Output 54.2.2 Wald Tests of Individual Effects

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Treatment	2	11.9886	0.0025
Sex	1	5.3104	0.0212
Treatment*Sex	2	0.1412	0.9318
Age	1	7.2744	0.0070
Duration	1	0.0247	0.8752

Parameter estimates are displayed in Output 54.2.3. The Exp(Est) column contains the exponentiated parameter estimates requested with the EXPB option. These values can, but do not necessarily, represent odds ratios for the corresponding variables. For continuous explanatory variables, the Exp(Est) value corresponds to the odds ratio for a unit increase of the corresponding variable. For CLASS variables that use effect coding, the Exp(Est) values have no direct interpretation as a comparison of levels. However, when

the reference coding is used, the Exp(Est) values represent the odds ratio between the corresponding level and the reference level. Following the parameter estimates table, PROC LOGISTIC displays the odds ratio estimates for those variables that are not involved in any interaction terms. If the variable is a CLASS variable, the odds ratio estimate comparing each level with the reference level is computed regardless of the coding scheme. In this analysis, since the model contains the Treatment*Sex interaction term, the odds ratios for Treatment and Sex were not computed. The odds ratio estimates for Age and Duration are precisely the values given in the Exp(Est) column in the parameter estimates table.

Output 54.2.3 Parameter Estimates with Effect Coding

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	19.2236	7.1315	7.2661	0.0070	2.232E8
Treatment	A	1	0.8483	0.5502	2.3773	0.1231	2.336
Treatment	B	1	1.4949	0.6622	5.0956	0.0240	4.459
Sex	F	1	0.9173	0.3981	5.3104	0.0212	2.503
Treatment*Sex	A F	1	-0.2010	0.5568	0.1304	0.7180	0.818
Treatment*Sex	B F	1	0.0487	0.5563	0.0077	0.9302	1.050
Age		1	-0.2688	0.0996	7.2744	0.0070	0.764
Duration		1	0.00523	0.0333	0.0247	0.8752	1.005

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Age	0.764	0.629	0.929
Duration	1.005	0.942	1.073

The following PROC LOGISTIC statements illustrate the use of forward selection on the data set Neuralgia to identify the effects that differentiate the two Pain responses. The option SELECTION=FORWARD is specified to carry out the forward selection. The term Treatment|Sex@2 illustrates another way to specify main effects and two-way interactions. (Note that, in this case, the "@2" is unnecessary because no interactions besides the two-way interaction are possible).

```
proc logistic data=Neuralgia;
   class Treatment Sex;
   model Pain=Treatment|Sex@2 Age Duration
         /selection=forward expb;
run;
```

Results of the forward selection process are summarized in Output 54.2.4. The variable Treatment is selected first, followed by Age and then Sex. The results are consistent with the previous analysis (Output 54.2.2) in which the Treatment*Sex interaction and Duration are not statistically significant.

Output 54.2.4 Effects Selected into the Model

The LOGISTIC Procedure

Summary of Forward Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq
1	Treatment	2	1	13.7143	0.0011
2	Age	1	2	10.6038	0.0011
3	Sex	1	3	5.9959	0.0143

Output 54.2.5 shows the Type 3 analysis of effects, the parameter estimates, and the odds ratio estimates for the selected model. All three variables, Treatment, Age, and Sex, are statistically significant at the 0.05 level (p=0.0018, p=0.0213, and p=0.0057, respectively). Since the selected model does not contain the Treatment*Sex interaction, odds ratios for Treatment and Sex are computed. The estimated odds ratio is 24.022 for treatment A versus placebo, 41.528 for Treatment B versus placebo, and 6.194 for female patients versus male patients. Note that these odds ratio estimates are not the same as the corresponding values in the Exp(Est) column in the parameter estimates table because effect coding was used. From Output 54.2.5, it is evident that both Treatment A and Treatment B are better than the placebo in reducing pain; females tend to have better improvement than males; and younger patients are faring better than older patients.

Output 54.2.5 Type 3 Effects and Parameter Estimates with Effect Coding

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Treatment	2	12.6928	0.0018
Sex	1	5.3013	0.0213
Age	1	7.6314	0.0057

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	19.0804	6.7882	7.9007	0.0049	1.9343E8
Treatment	A	1	0.8772	0.5274	2.7662	0.0963	2.404
Treatment	B	1	1.4246	0.6036	5.5711	0.0183	4.156
Sex	F	1	0.9118	0.3960	5.3013	0.0213	2.489
Age		1	-0.2650	0.0959	7.6314	0.0057	0.767

Output 54.2.5 *continued*

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Treatment A vs P	24.022	3.295	175.121
Treatment B vs P	41.528	4.500	383.262
Sex F vs M	6.194	1.312	29.248
Age	0.767	0.636	0.926

Finally, the following statements refit the previously selected model, except that reference coding is used for the CLASS variables instead of effect coding:

```
ods graphics on;
proc logistic data=Neuralgia plots(only)=(oddsratio(range=clip));
   class Treatment Sex /param=ref;
   model Pain= Treatment Sex Age;
   oddsratio Treatment;
   oddsratio Sex;
   oddsratio Age;
   contrast 'Pairwise A vs P' Treatment 1  0 / estimate=exp;
   contrast 'Pairwise B vs P' Treatment 0  1 / estimate=exp;
   contrast 'Pairwise A vs B' Treatment 1 -1 / estimate=exp;
   contrast 'Female vs Male' Sex 1 / estimate=exp;
   effectplot / at(Sex=all) noobs;
   effectplot slicefit(sliceby=Sex plotby=Treatment) / noobs;
run;
ods graphics off;
```

The ODDSRATIO statements compute the odds ratios for the covariates. Four CONTRAST statements are specified; they provide another method of producing the odds ratios. The three contrasts labeled 'Pairwise' specify a contrast vector, L, for each of the pairwise comparisons between the three levels of Treatment. The contrast labeled 'Female vs Male' compares female to male patients. The option ESTIMATE=EXP is specified in all CONTRAST statements to exponentiate the estimates of $\mathbf{L}'\boldsymbol{\beta}$. With the given specification of contrast coefficients, the first of the 'Pairwise' CONTRAST statements corresponds to the odds ratio of A versus P, the second corresponds to B versus P, and the third corresponds to A versus B. You can also specify the 'Pairwise' contrasts in a single contrast statement with three rows. The 'Female vs Male' CONTRAST statement corresponds to the odds ratio that compares female to male patients.

The PLOTS(ONLY)= option displays only the requested odds ratio plot when ODS Graphics is enabled. The EFFECTPLOT statements do not honor the ONLY option, and display the fitted model. The first EFFECTPLOT statement by default produces a plot of the predicted values against the continuous Age variable, grouped by the Treatment levels. The AT option produces one plot for males and another for females; the NOOBS option suppresses the display of the observations. In the second EFFECTPLOT statement, a SLICEFIT plot is specified to display the Age variable on the X axis, the fits are grouped by the Sex levels, and the PLOTBY= option produces a panel of plots that displays each level of the Treatment variable.

The reference coding is shown in Output 54.2.6. The Type 3 analysis of effects, the parameter estimates for the reference coding, and the odds ratio estimates are displayed in Output 54.2.7. Although the parameter

estimates are different because of the different parameterizations, the "Type 3 Analysis of Effects" table and the "Odds Ratio" table remain the same as in Output 54.2.5. With effect coding, the treatment A parameter estimate (0.8772) estimates the effect of treatment A compared to the average effect of treatments A, B, and placebo. The treatment A estimate (3.1790) under the reference coding estimates the difference in effect of treatment A and the placebo treatment.

Output 54.2.6 Reference Coding of CLASS Variables

The LOGISTIC Procedure

Class Level Information

Class	Value	Design Variables	
Treatment	A	1	0
	B	0	1
	P	0	0
Sex	F	1	
	M	0	

Output 54.2.7 Type 3 Effects and Parameter Estimates with Reference Coding

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Treatment	2	12.6928	0.0018
Sex	1	5.3013	0.0213
Age	1	7.6314	0.0057

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	15.8669	6.4056	6.1357	0.0132
Treatment A	1	3.1790	1.0135	9.8375	0.0017
Treatment B	1	3.7264	1.1339	10.8006	0.0010
Sex F	1	1.8235	0.7920	5.3013	0.0213
Age	1	-0.2650	0.0959	7.6314	0.0057

The ODDSRATIO statement results are shown in Output 54.2.8, and the resulting plot is displayed in Output 54.2.9. Note in Output 54.2.9 that the odds ratio confidence limits are truncated due to specifying the RANGE=CLIP option; this enables you to see which intervals contain "1" more clearly. The odds ratios are identical to those shown in the "Odds Ratio Estimates" table in Output 54.2.7 with the addition of the odds ratio for "Treatment A vs B". Both treatments A and B are highly effective over placebo in reducing pain, as can be seen from the odds ratios comparing treatment A against P and treatment B against P (the second

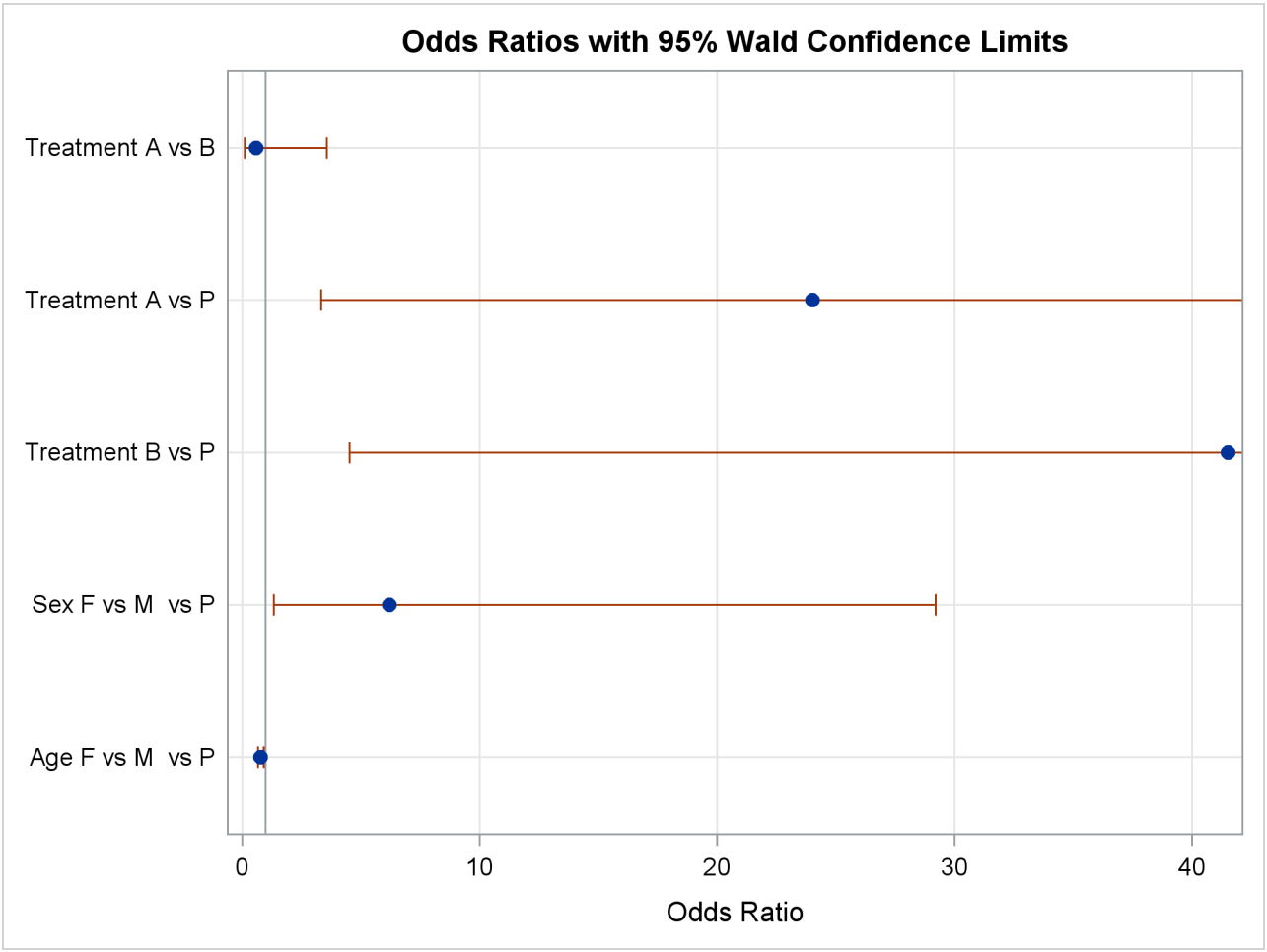
and third rows in the table). However, the 95% confidence interval for the odds ratio comparing treatment A to B is (0.0932, 3.5889), indicating that the pain reduction effects of these two test treatments are not very different. Again, the 'Sex F vs M' odds ratio shows that female patients fared better in obtaining relief from pain than male patients. The odds ratio for Age shows that a patient one year older is 0.77 times as likely to show no pain; that is, younger patients have more improvement than older patients.

Output 54.2.8 Results from the ODDSRATIO Statements

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
Treatment A vs B	0.578	0.093	3.589
Treatment A vs P	24.022	3.295	175.121
Treatment B vs P	41.528	4.500	383.262
Sex F vs M	6.194	1.312	29.248
Age	0.767	0.636	0.926

Output 54.2.9 Plot of the ODDSRATIO Statement Results



Output 54.2.10 contains two tables: the "Contrast Test Results" table and the "Contrast Estimation and Testing Results by Row" table. The former contains the overall Wald test for each CONTRAST statement. The latter table contains estimates and tests of individual contrast rows. The estimates for the first two rows of the 'Pairwise' CONTRAST statements are the same as those given in the two preceding odds ratio tables (Output 54.2.7 and Output 54.2.8). The third row estimates the odds ratio comparing A to B, agreeing with Output 54.2.8, and the last row computes the odds ratio comparing pain relief for females to that for males.

Output 54.2.10 Results of CONTRAST Statements

Contrast Test Results

Contrast	DF	Wald Chi-Square	Pr > ChiSq
Pairwise A vs P	1	9.8375	0.0017
Pairwise B vs P	1	10.8006	0.0010
Pairwise A vs B	1	0.3455	0.5567
Female vs Male	1	5.3013	0.0213

Contrast Estimation and Testing Results by Row

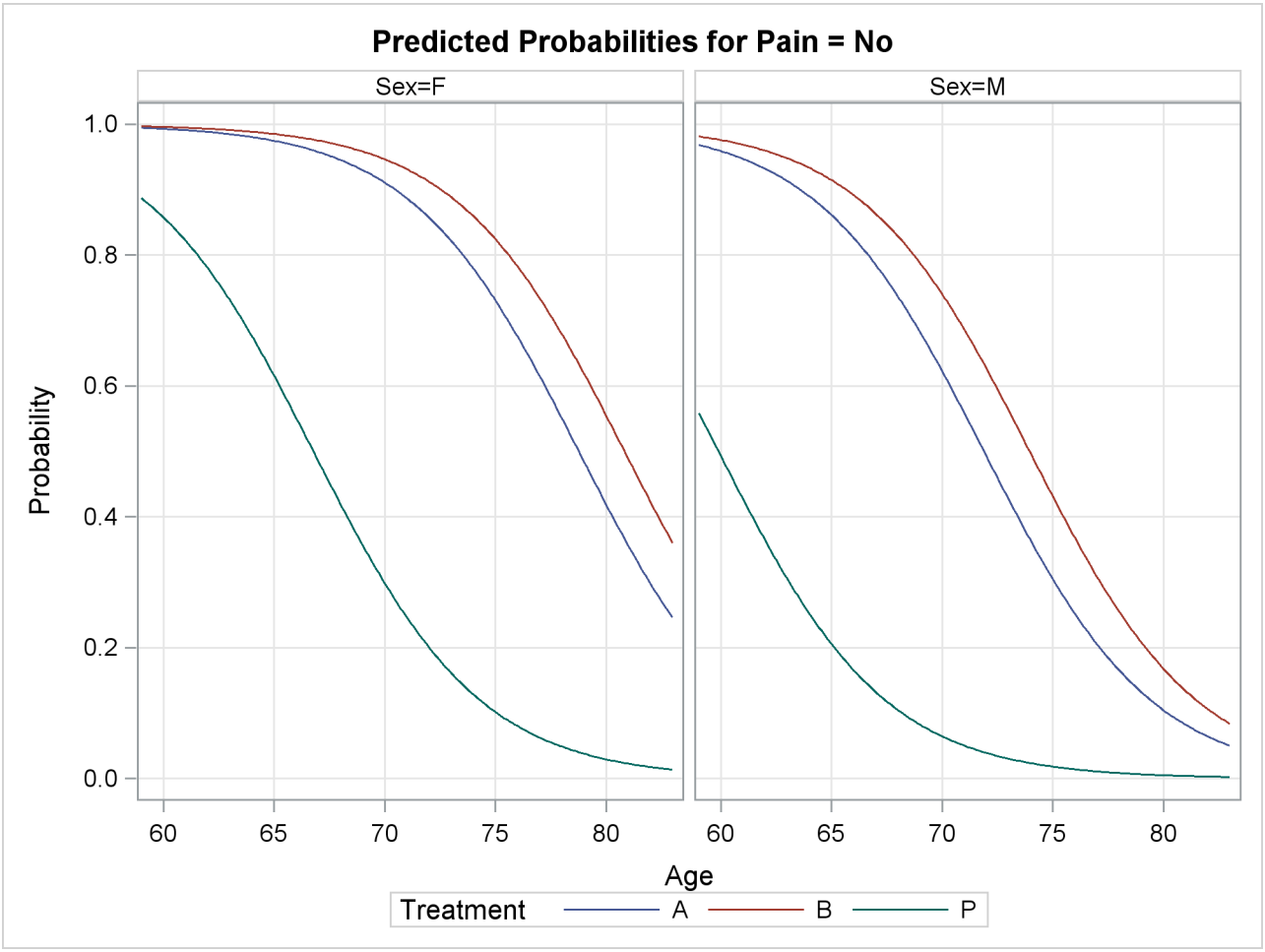
Contrast	Type	Row	Estimate	Standard Error	Alpha	Confidence Limits	
Pairwise A vs P	EXP	1	24.0218	24.3473	0.05	3.2951	175.1
Pairwise B vs P	EXP	1	41.5284	47.0877	0.05	4.4998	383.3
Pairwise A vs B	EXP	1	0.5784	0.5387	0.05	0.0932	3.5889
Female vs Male	EXP	1	6.1937	4.9053	0.05	1.3116	29.2476

Contrast Estimation and Testing Results by Row

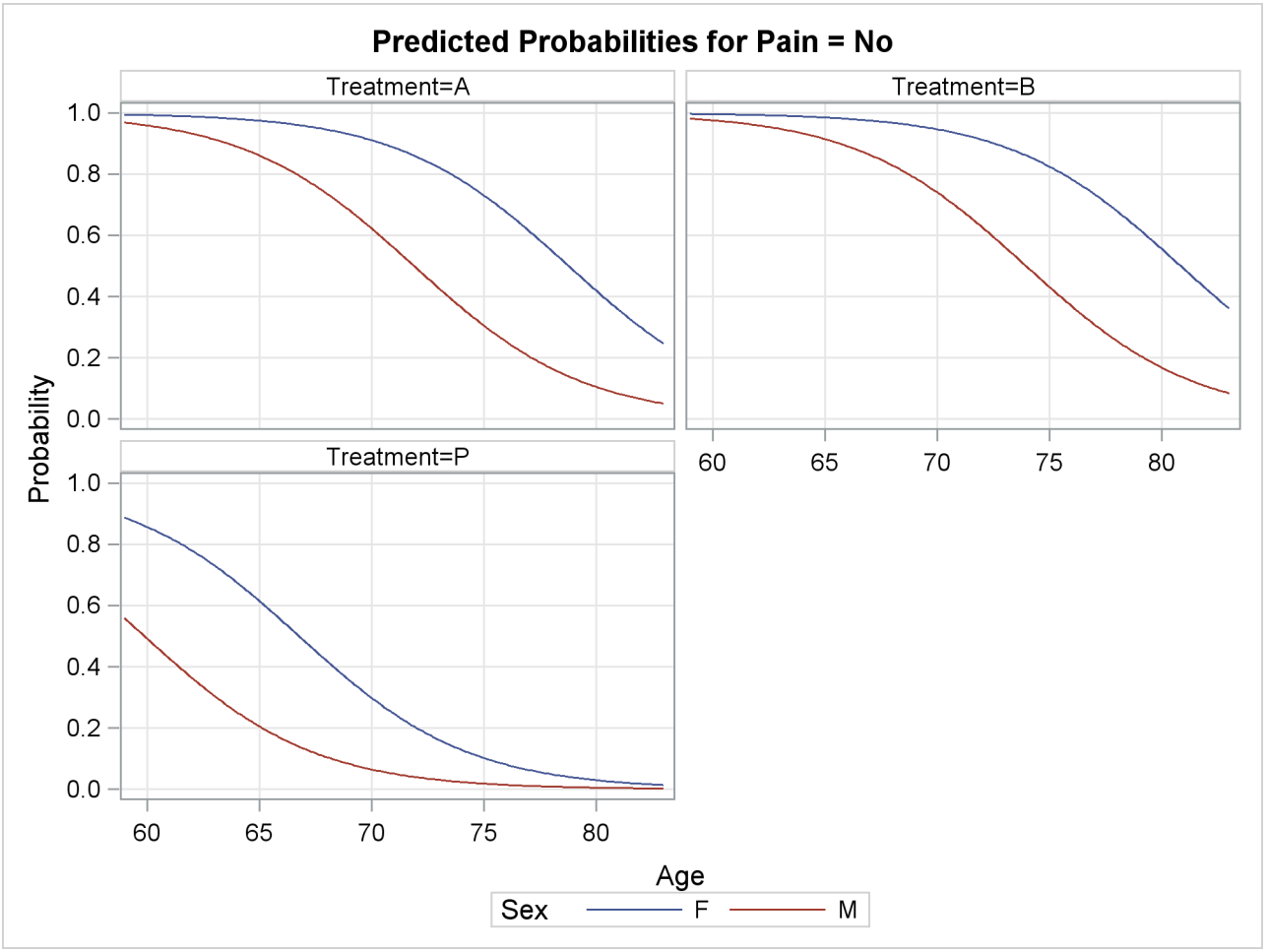
Contrast	Type	Row	Wald Chi-Square	Pr > ChiSq
Pairwise A vs P	EXP	1	9.8375	0.0017
Pairwise B vs P	EXP	1	10.8006	0.0010
Pairwise A vs B	EXP	1	0.3455	0.5567
Female vs Male	EXP	1	5.3013	0.0213

ANCOVA-style plots of the model-predicted probabilities against the Age variable for each combination of Treatment and Sex are displayed in Output 54.2.11 and Output 54.2.12. These plots confirm that females always have a higher probability of pain reduction in each treatment group, the placebo treatment has a lower probability of success than the other treatments, and younger patients respond to treatment better than older patients.

Output 54.2.11 Model-Predicted Probabilities by Sex



Output 54.2.12 Model-Predicted Probabilities by Treatment



Example 54.3: Ordinal Logistic Regression

Consider a study of the effects on taste of various cheese additives. Researchers tested four cheese additives and obtained 52 response ratings for each additive. Each response was measured on a scale of nine categories ranging from strong dislike (1) to excellent taste (9). The data, given in McCullagh and Nelder (1989, p. 175) in the form of a two-way frequency table of additive by rating, are saved in the data set Cheese by using the following program. The variable y contains the response rating. The variable Additive specifies the cheese additive (1, 2, 3, or 4). The variable freq gives the frequency with which each additive received each rating.

```
data Cheese;
   do Additive = 1 to 4;
      do y = 1 to 9;
         input freq @@;
         output;
      end;
   end;
   label y='Taste Rating';
   datalines;
```

```
0  0  1  7  8  8 19  8  1
6  9 12 11  7  6  1  0  0
1  1  6  8 23  7  5  1  0
0  0  0  1  3  7 14 16 11
;
```

The response variable y is ordinally scaled. A cumulative logit model is used to investigate the effects of the cheese additives on taste. The following statements invoke PROC LOGISTIC to fit this model with y as the response variable and three indicator variables as explanatory variables, with the fourth additive as the reference level. With this parameterization, each Additive parameter compares an additive to the fourth additive. The COVB option displays the estimated covariance matrix. The ODDSRATIO statement computes odds ratios for all combinations of the Additive levels. The PLOTS option produces a graphical display of the odds ratios, and the EFFECTPLOT statement displays the predicted probabilities.

```
ods graphics on;
proc logistic data=Cheese plots(only)=oddsratio(range=clip);
   freq freq;
   class Additive (param=ref ref='4');
   model y=Additive / covb;
   oddsratio Additive;
   effectplot / polybar;
   title 'Multiple Response Cheese Tasting Experiment';
run;
ods graphics off;
```

The "Response Profile" table in Output 54.3.1 shows that the strong dislike (y=1) end of the rating scale is associated with lower Ordered Values in the "Response Profile" table; hence the probability of disliking the additives is modeled.

The score chi-square for testing the proportional odds assumption is 17.287, which is not significant with respect to a chi-square distribution with 21 degrees of freedom ($p = 0.694$). This indicates that the proportional odds assumption is reasonable. The positive value (1.6128) for the parameter estimate for Additive1 indicates a tendency toward the lower-numbered categories of the first cheese additive relative to the fourth. In other words, the fourth additive tastes better than the first additive. The second and third additives are both less favorable than the fourth additive. The relative magnitudes of these slope estimates imply the preference ordering: fourth, first, third, second.

Output 54.3.1 Proportional Odds Model Regression Analysis

```
            Multiple Response Cheese Tasting Experiment

                       The LOGISTIC Procedure

                         Model Information

      Data Set                        WORK.CHEESE
      Response Variable               y                    Taste Rating
      Number of Response Levels       9
      Frequency Variable              freq
      Model                           cumulative logit
      Optimization Technique          Fisher's scoring
```

Output 54.3.1 *continued*

Number of Observations Read	36
Number of Observations Used	28
Sum of Frequencies Read	208
Sum of Frequencies Used	208

Response Profile

Ordered Value	y	Total Frequency
1	1	7
2	2	10
3	3	19
4	4	27
5	5	41
6	6	28
7	7	39
8	8	25
9	9	12

Probabilities modeled are cumulated over the lower Ordered Values.

NOTE: 8 observations having nonpositive frequencies or weights were excluded since they do not contribute to the analysis.

Class Level Information

Class	Value	Design Variables		
Additive	1	1	0	0
	2	0	1	0
	3	0	0	1
	4	0	0	0

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
17.2866	21	0.6936

Output 54.3.1 *continued*

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	875.802	733.348
SC	902.502	770.061
-2 Log L	859.802	711.348

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	148.4539	3	<.0001
Score	111.2670	3	<.0001
Wald	115.1504	3	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Additive	3	115.1504	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 1	1	-7.0801	0.5624	158.4851	<.0001
Intercept 2	1	-6.0249	0.4755	160.5500	<.0001
Intercept 3	1	-4.9254	0.4272	132.9484	<.0001
Intercept 4	1	-3.8568	0.3902	97.7087	<.0001
Intercept 5	1	-2.5205	0.3431	53.9704	<.0001
Intercept 6	1	-1.5685	0.3086	25.8374	<.0001
Intercept 7	1	-0.0669	0.2658	0.0633	0.8013
Intercept 8	1	1.4930	0.3310	20.3439	<.0001
Additive 1	1	1.6128	0.3778	18.2265	<.0001
Additive 2	1	4.9645	0.4741	109.6427	<.0001
Additive 3	1	3.3227	0.4251	61.0931	<.0001

Association of Predicted Probabilities and Observed Responses

Percent Concordant	67.6	Somers' D	0.578
Percent Discordant	9.8	Gamma	0.746
Percent Tied	22.6	Tau-a	0.500
Pairs	18635	c	0.789

The odds ratio results in Output 54.3.2 show the preferences more clearly. For example, the "Additive 1 vs 4" odds ratio says that the first additive has 5.017 times the odds of receiving a lower score than the fourth

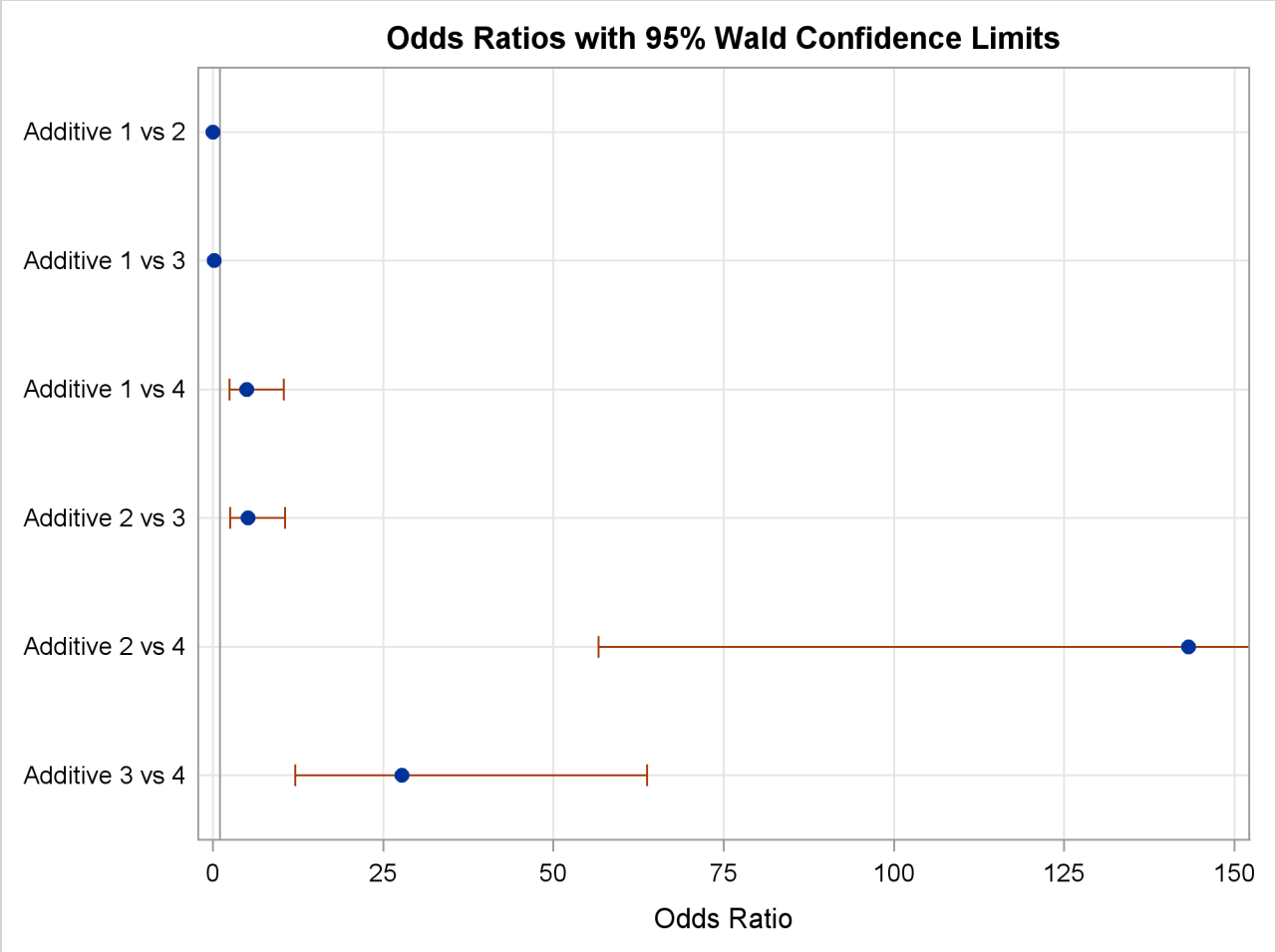
additive; that is, the first additive is 5.017 times more likely than the fourth additive to receive a lower score. Output 54.3.3 displays the odds ratios graphically; the range of the confidence limits is truncated by the RANGE=CLIP option, so you can see that "1" is not contained in any of the intervals.

Output 54.3.2 Odds Ratios of All Pairs of Additive Levels

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
Additive 1 vs 2	0.035	0.015	0.080
Additive 1 vs 3	0.181	0.087	0.376
Additive 1 vs 4	5.017	2.393	10.520
Additive 2 vs 3	5.165	2.482	10.746
Additive 2 vs 4	143.241	56.558	362.777
Additive 3 vs 4	27.734	12.055	63.805

Output 54.3.3 Plot of Odds Ratios for Additive



The estimated covariance matrix of the parameters is displayed in Output 54.3.4.

Output 54.3.4 Estimated Covariance Matrix

Estimated Covariance Matrix

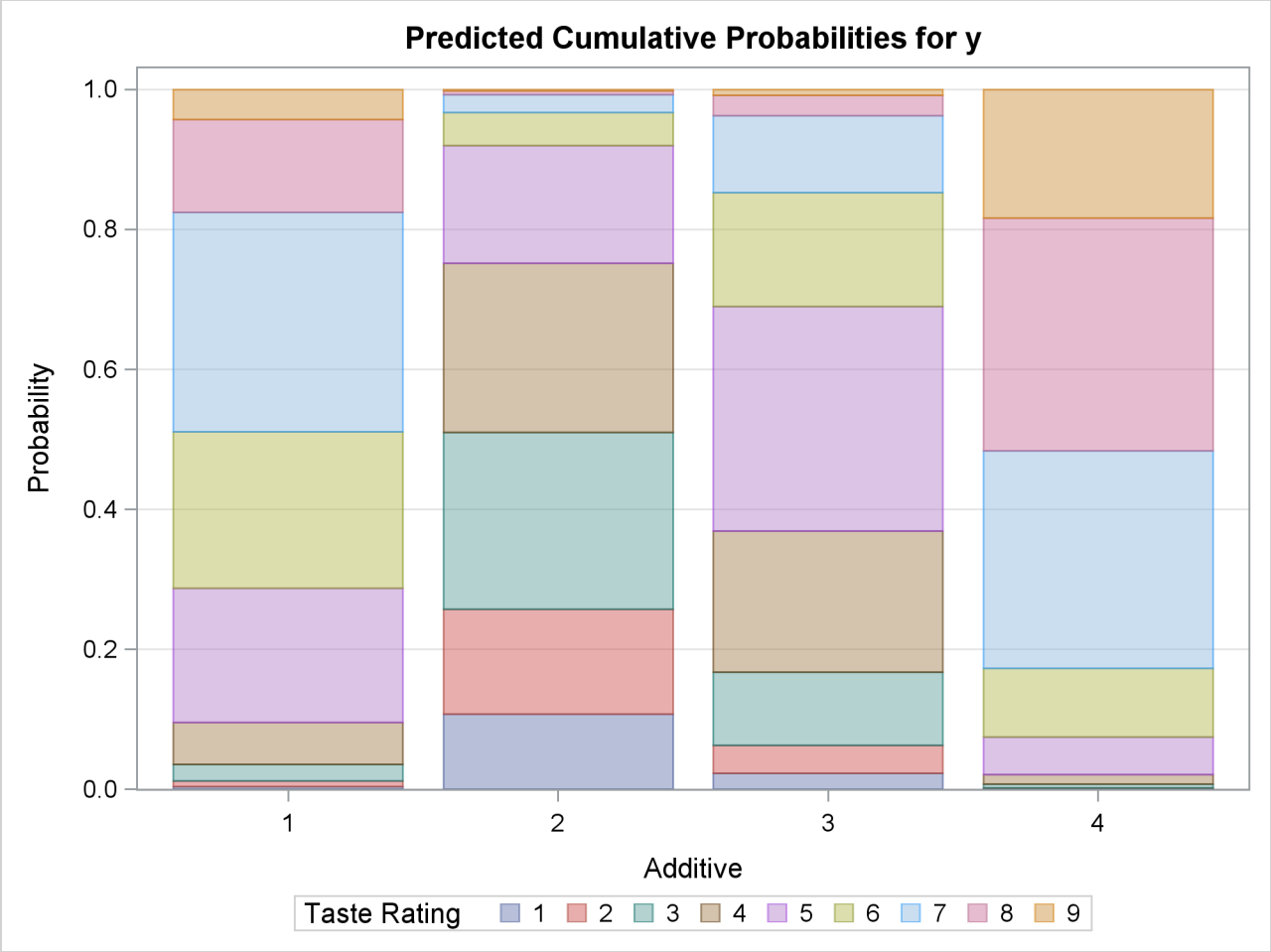
Parameter	Intercept_ 1	Intercept_ 2	Intercept_ 3	Intercept_ 4	Intercept_ 5
Intercept_1	0.316291	0.219581	0.176278	0.147694	0.114024
Intercept_2	0.219581	0.226095	0.177806	0.147933	0.11403
Intercept_3	0.176278	0.177806	0.182473	0.148844	0.114092
Intercept_4	0.147694	0.147933	0.148844	0.152235	0.114512
Intercept_5	0.114024	0.11403	0.114092	0.114512	0.117713
Intercept_6	0.091085	0.091081	0.091074	0.091109	0.091821
Intercept_7	0.057814	0.057813	0.057807	0.05778	0.057721
Intercept_8	0.041304	0.041304	0.0413	0.041277	0.041162
Additive1	-0.09419	-0.09421	-0.09427	-0.09428	-0.09246
Additive2	-0.18686	-0.18161	-0.1687	-0.14717	-0.11415
Additive3	-0.13565	-0.13569	-0.1352	-0.13118	-0.11207

Estimated Covariance Matrix

Parameter	Intercept_ 6	Intercept_ 7	Intercept_ 8	Additive1	Additive2	Additive3
Intercept_1	0.091085	0.057814	0.041304	-0.09419	-0.18686	-0.13565
Intercept_2	0.091081	0.057813	0.041304	-0.09421	-0.18161	-0.13569
Intercept_3	0.091074	0.057807	0.0413	-0.09427	-0.1687	-0.1352
Intercept_4	0.091109	0.05778	0.041277	-0.09428	-0.14717	-0.13118
Intercept_5	0.091821	0.057721	0.041162	-0.09246	-0.11415	-0.11207
Intercept_6	0.09522	0.058312	0.041324	-0.08521	-0.09113	-0.09122
Intercept_7	0.058312	0.07064	0.04878	-0.06041	-0.05781	-0.05802
Intercept_8	0.041324	0.04878	0.109562	-0.04436	-0.0413	-0.04143
Additive1	-0.08521	-0.06041	-0.04436	0.142715	0.094072	0.092128
Additive2	-0.09113	-0.05781	-0.0413	0.094072	0.22479	0.132877
Additive3	-0.09122	-0.05802	-0.04143	0.092128	0.132877	0.180709

Output 54.3.5 displays the probability of each taste rating y within each additive. You can see that Additive=1 mostly receives ratings of 5 to 7, Additive=2 mostly receives ratings of 2 to 5, Additive=3 mostly receives ratings of 4 to 6, and Additive=4 mostly receives ratings of 7 to 9, which also confirms the previously discussed preference orderings.

Output 54.3.5 Model-Predicted Probabilities



Example 54.4: Nominal Response Data: Generalized Logits Model

Over the course of one school year, third graders from three different schools are exposed to three different styles of mathematics instruction: a self-paced computer-learning style, a team approach, and a traditional class approach. The students are asked which style they prefer and their responses, classified by the type of program they are in (a regular school day versus a regular day supplemented with an afternoon school program), are displayed in Table 54.15. The data set is from Stokes, Davis, and Koch (2012), and is also analyzed in the section "Generalized Logits Model" on page 1784 of Chapter 30, "The CATMOD Procedure."

Table 54.15 School Program Data

School	Program	Learning Style Preference		
		Self	Team	Class
1	Regular	10	17	26
1	Afternoon	5	12	50
2	Regular	21	17	26
2	Afternoon	16	12	36
3	Regular	15	15	16
3	Afternoon	12	12	20

The levels of the response variable (self, team, and class) have no essential ordering, so a logistic regression is performed on the generalized logits. The model to be fit is

$$\log\left(\frac{\pi_{hij}}{\pi_{hir}}\right) = \alpha_j + \mathbf{x}'_{hi}\boldsymbol{\beta}_j$$

where π_{hij} is the probability that a student in school h and program i prefers teaching style j, $j \neq r$, and style r is the baseline style (in this case, class). There are separate sets of intercept parameters α_j and regression parameters $\boldsymbol{\beta}_j$ for each logit, and the vector $\mathbf{x}_{hi}$ is the set of explanatory variables for the hith population. Thus, two logits are modeled for each school and program combination: the logit comparing self to class and the logit comparing team to class.

The following statements create the data set school and request the analysis. The LINK=GLOGIT option forms the generalized logits. The response variable option ORDER=DATA means that the response variable levels are ordered as they exist in the data set: self, team, and class; thus, the logits are formed by comparing self to class and by comparing team to class. The ODDSRATIO statement produces odds ratios in the presence of interactions, and a graphical display of the requested odds ratios is produced when ODS Graphics is enabled.

```
data school;
   length Program $ 9;
   input School Program $ Style $ Count @@;
   datalines;
1 regular   self 10  1 regular   team 17  1 regular   class 26
1 afternoon self  5  1 afternoon team 12  1 afternoon class 50
2 regular   self 21  2 regular   team 17  2 regular   class 26
2 afternoon self 16  2 afternoon team 12  2 afternoon class 36
3 regular   self 15  3 regular   team 15  3 regular   class 16
3 afternoon self 12  3 afternoon team 12  3 afternoon class 20
;

ods graphics on;
proc logistic data=school;
   freq Count;
   class School Program(ref=first);
   model Style(order=data)=School Program School*Program / link=glogit;
   oddsratio program;
run;
ods graphics off;
```

Summary information about the model, the response variable, and the classification variables are displayed in Output 54.4.1.

Output 54.4.1 Analysis of Saturated Model

```
                    The LOGISTIC Procedure

                      Model Information

Data Set                          WORK.SCHOOL
Response Variable                 Style
Number of Response Levels         3
Frequency Variable                Count
Model                             generalized logit
Optimization Technique            Newton-Raphson


    Number of Observations Read          18
    Number of Observations Used          18
    Sum of Frequencies Read             338
    Sum of Frequencies Used             338


                  Response Profile

          Ordered                      Total
            Value     Style        Frequency

                1     self                79
                2     team                85
                3     class              174

Logits modeled use Style='class' as the reference category.


                Class Level Information

                                      Design
Class       Value              Variables

School      1                   1      0
            2                   0      1
            3                  -1     -1

Program     afternoon          -1
            regular             1


                Model Convergence Status

        Convergence criterion (GCONV=1E-8) satisfied.
```

The "Testing Global Null Hypothesis: BETA=0" table in Output 54.4.2 shows that the parameters are significantly different from zero.

Output 54.4.2 Analysis of Saturated Model

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	699.404	689.156
SC	707.050	735.033
-2 Log L	695.404	665.156

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	30.2480	10	0.0008
Score	28.3738	10	0.0016
Wald	25.6828	10	0.0042

However, the "Type 3 Analysis of Effects" table in Output 54.4.3 shows that the interaction effect is clearly nonsignificant.

Output 54.4.3 Analysis of Saturated Model

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
School	4	14.5522	0.0057
Program	2	10.4815	0.0053
School*Program	4	1.7439	0.7827

Analysis of Maximum Likelihood Estimates

Parameter			Style	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			self	1	-0.8097	0.1488	29.5989	<.0001
Intercept			team	1	-0.6585	0.1366	23.2449	<.0001
School	1		self	1	-0.8194	0.2281	12.9066	0.0003
School	1		team	1	-0.2675	0.1881	2.0233	0.1549
School	2		self	1	0.2974	0.1919	2.4007	0.1213
School	2		team	1	-0.1033	0.1898	0.2961	0.5863
Program	regular		self	1	0.3985	0.1488	7.1684	0.0074
Program	regular		team	1	0.3537	0.1366	6.7071	0.0096
School*Program	1	regular	self	1	0.2751	0.2281	1.4547	0.2278
School*Program	1	regular	team	1	0.1474	0.1881	0.6143	0.4332
School*Program	2	regular	self	1	-0.0998	0.1919	0.2702	0.6032
School*Program	2	regular	team	1	-0.0168	0.1898	0.0079	0.9293

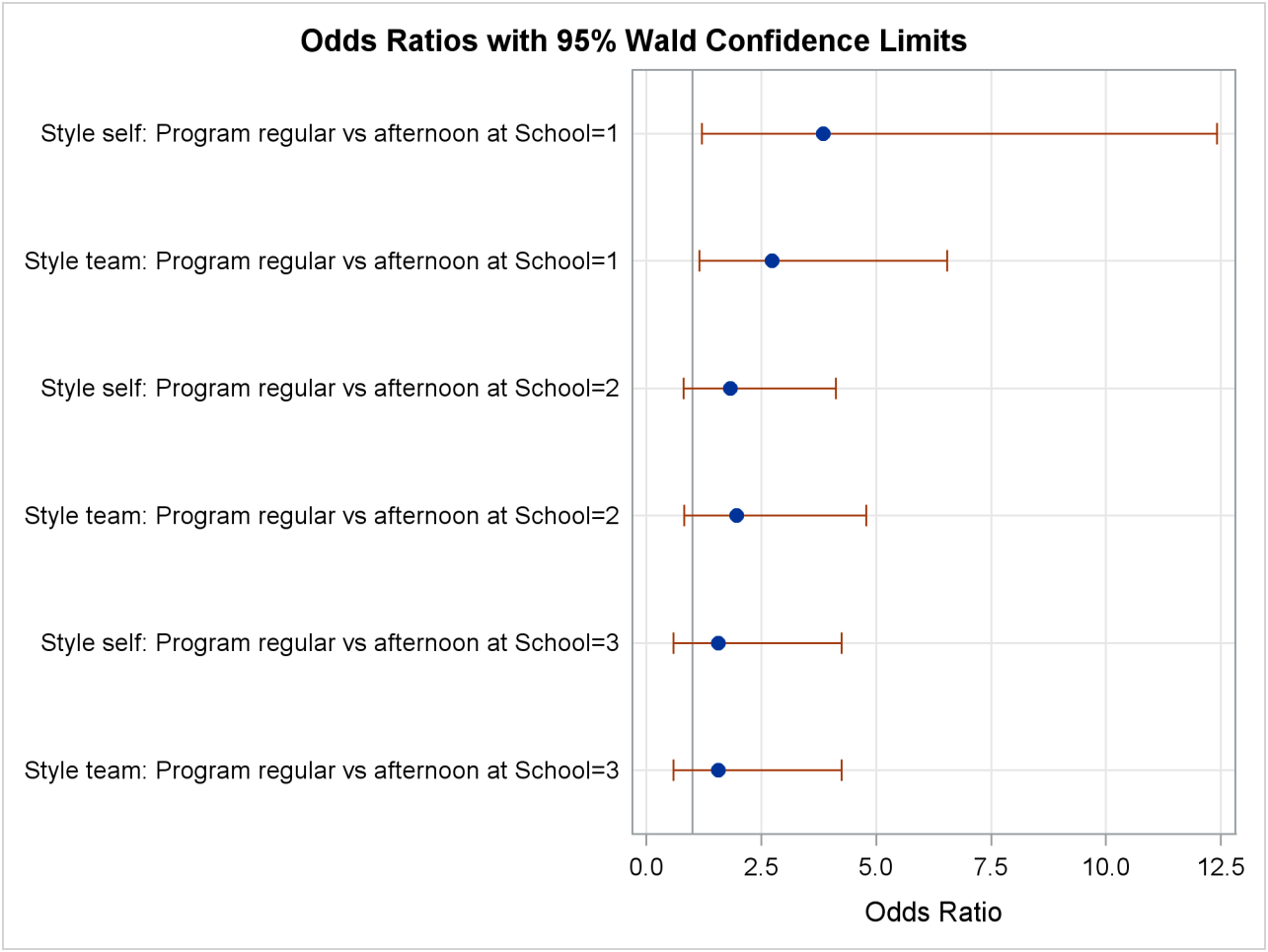
The table produced by the ODDSRATIO statement is displayed in Output 54.4.4. The differences between the program preferences are small across all the styles (logits) compared to their variability as displayed by the confidence limits in Output 54.4.5, confirming that the interaction effect is nonsignificant.

Output 54.4.4 Odds Ratios for Style

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
Style self: Program regular vs afternoon at School=1	3.846	1.190	12.435
Style team: Program regular vs afternoon at School=1	2.724	1.132	6.554
Style self: Program regular vs afternoon at School=2	1.817	0.798	4.139
Style team: Program regular vs afternoon at School=2	1.962	0.802	4.799
Style self: Program regular vs afternoon at School=3	1.562	0.572	4.265
Style team: Program regular vs afternoon at School=3	1.562	0.572	4.265

Output 54.4.5 Plot of Odds Ratios for Style



Since the interaction effect is clearly nonsignificant, a main-effects model is fit with the following statements. The EFFECTPLOT statement creates a plot of the predicted values versus the levels of the School

variable at each level of the Program variables. The CLM option adds confidence bars, and the NOOBS option suppresses the display of the observations.

```
ods graphics on;
proc logistic data=school;
   freq Count;
   class School Program(ref=first);
   model Style(order=data)=School Program / link=glogit;
   effectplot interaction(plotby=Program) / clm noobs;
run;
ods graphics off;
```

All of the global fit tests in Output 54.4.6 suggest the model is significant, and the Type 3 tests show that the school and program effects are also significant.

Output 54.4.6 Analysis of Main-Effects Model

The LOGISTIC Procedure

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	699.404	682.934
SC	707.050	713.518
-2 Log L	695.404	666.934

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	28.4704	6	<.0001
Score	27.1190	6	0.0001
Wald	25.5881	6	0.0003

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
School	4	14.8424	0.0050
Program	2	10.9160	0.0043

The parameter estimates, tests for individual parameters, and odds ratios are displayed in Output 54.4.7. The Program variable has nearly the same effect on both logits, while School=1 has the largest effect of the schools.

Output 54.4.7 Estimates

Analysis of Maximum Likelihood Estimates

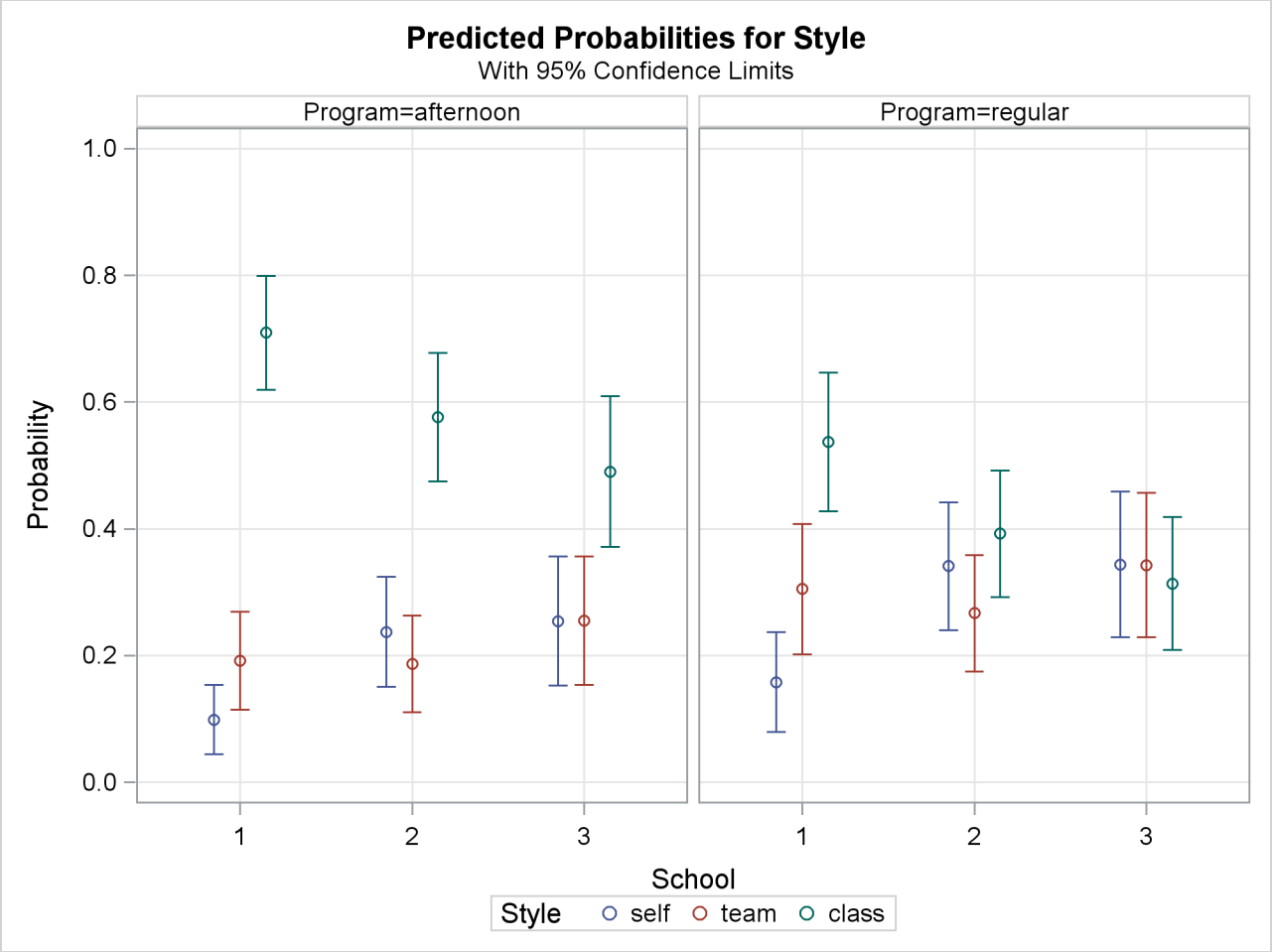
Parameter		Style	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		self	1	−0.7978	0.1465	29.6502	<.0001
Intercept		team	1	−0.6589	0.1367	23.2300	<.0001
School	1	self	1	−0.7992	0.2198	13.2241	0.0003
School	1	team	1	−0.2786	0.1867	2.2269	0.1356
School	2	self	1	0.2836	0.1899	2.2316	0.1352
School	2	team	1	−0.0985	0.1892	0.2708	0.6028
Program	regular	self	1	0.3737	0.1410	7.0272	0.0080
Program	regular	team	1	0.3713	0.1353	7.5332	0.0061

Odds Ratio Estimates

Effect			Style	Point Estimate	95% Wald Confidence Limits	
School	1 vs 3		self	0.269	0.127	0.570
School	1 vs 3		team	0.519	0.267	1.010
School	2 vs 3		self	0.793	0.413	1.522
School	2 vs 3		team	0.622	0.317	1.219
Program	regular	vs afternoon	self	2.112	1.215	3.670
Program	regular	vs afternoon	team	2.101	1.237	3.571

The interaction plots in Output 54.4.8 show that School=1 and Program=afternoon have a preference for the traditional classroom style. Of course, since these are not simultaneous confidence intervals, the nonoverlapping 95% confidence limits do not take the place of an actual test.

Output 54.4.8 Model-Predicted Probabilities



Example 54.5: Stratified Sampling

Consider the hypothetical example in Fleiss (1981, pp. 6–7), in which a test is applied to a sample of 1,000 people known to have a disease and to another sample of 1,000 people known not to have the same disease. In the diseased sample, 950 test positive; in the nondiseased sample, only 10 test positive. If the true disease rate in the population is 1 in 100, specifying PEVENT=0.01 results in the correct false positive and negative rates for the stratified sampling scheme. Omitting the PEVENT= option is equivalent to using the overall sample disease rate (1000/2000 = 0.5) as the value of the PEVENT= option, which would ignore the stratified sampling.

The statements to create the data set and perform the analysis are as follows:

```
data Screen;
   do Disease='Present','Absent';
      do Test=1,0;
         input Count @@;
         output;
      end;
   end;
```

```
   datalines;
950  50
 10 990
;

proc logistic data=Screen;
   freq Count;
   model Disease(event='Present')=Test
         / pevent=.5 .01 ctable pprob=.5;
run;
```

The response variable option EVENT= indicates that Disease='Present' is the event. The CTABLE option is specified to produce a classification table. Specifying PPROB=0.5 indicates a cutoff probability of 0.5. A list of two probabilities, 0.5 and 0.01, is specified for the PEVENT= option; 0.5 corresponds to the overall sample disease rate, and 0.01 corresponds to a true disease rate of 1 in 100.

The classification table is shown in Output 54.5.1.

Output 54.5.1 False Positive and False Negative Rates

The LOGISTIC Procedure

Classification Table

		Correct		Incorrect		Percentages				
Prob Event	Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	0.500	950	990	10	50	97.0	95.0	99.0	1.0	4.8
0.010	0.500	950	990	10	50	99.0	95.0	99.0	51.0	0.1

In the classification table, the column "Prob Level" represents the cutoff values (the settings of the PPROB= option) for predicting whether an observation is an event. The "Correct" columns list the numbers of subjects that are correctly predicted as events and nonevents, respectively, and the "Incorrect" columns list the number of nonevents incorrectly predicted as events and the number of events incorrectly predicted as nonevents, respectively. For PEVENT=0.5, the false positive rate is 1% and the false negative rate is 4.8%. These results ignore the fact that the samples were stratified and incorrectly assume that the overall sample proportion of disease (which is 0.5) estimates the true disease rate. For a true disease rate of 0.01, the false positive rate and the false negative rate are 51% and 0.1%, respectively, as shown in the second line of the classification table.

Example 54.6: Logistic Regression Diagnostics

In a controlled experiment to study the effect of the rate and volume of air intake on a transient reflex vasoconstriction in the skin of the digits, 39 tests under various combinations of rate and volume of air intake were obtained (Finney 1947). The endpoint of each test is whether or not vasoconstriction occurred. Pregibon (1981) uses this set of data to illustrate the diagnostic measures he proposes for detecting influential observations and to quantify their effects on various aspects of the maximum likelihood fit.

The vasoconstriction data are saved in the data set vaso:

```
data vaso;
   length Response $12;
   input Volume Rate Response @@;
   LogVolume=log(Volume);
   LogRate=log(Rate);
   datalines;
3.70  0.825  constrict       3.50  1.09   constrict
1.25  2.50   constrict       0.75  1.50   constrict
0.80  3.20   constrict       0.70  3.50   constrict
0.60  0.75   no_constrict    1.10  1.70   no_constrict
0.90  0.75   no_constrict    0.90  0.45   no_constrict
0.80  0.57   no_constrict    0.55  2.75   no_constrict
0.60  3.00   no_constrict    1.40  2.33   constrict
0.75  3.75   constrict       2.30  1.64   constrict
3.20  1.60   constrict       0.85  1.415  constrict
1.70  1.06   no_constrict    1.80  1.80   constrict
0.40  2.00   no_constrict    0.95  1.36   no_constrict
1.35  1.35   no_constrict    1.50  1.36   no_constrict
1.60  1.78   constrict       0.60  1.50   no_constrict
1.80  1.50   constrict       0.95  1.90   no_constrict
1.90  0.95   constrict       1.60  0.40   no_constrict
2.70  0.75   constrict       2.35  0.03   no_constrict
1.10  1.83   no_constrict    1.10  2.20   constrict
1.20  2.00   constrict       0.80  3.33   constrict
0.95  1.90   no_constrict    0.75  1.90   no_constrict
1.30  1.625  constrict
;
```

In the data set vaso, the variable Response represents the outcome of a test. The variable LogVolume represents the log of the volume of air intake, and the variable LogRate represents the log of the rate of air intake.

The following statements invoke PROC LOGISTIC to fit a logistic regression model to the vasoconstriction data, where Response is the response variable, and LogRate and LogVolume are the explanatory variables. Regression diagnostics are displayed when ODS Graphics is enabled, and the INFLUENCE option is specified to display a table of the regression diagnostics.

```
ods graphics on;
title 'Occurrence of Vasoconstriction';
proc logistic data=vaso;
   model Response=LogRate LogVolume/influence iplots;
run;
ods graphics off;
```

Results of the model fit are shown in Output 54.6.1. Both LogRate and LogVolume are statistically significant to the occurrence of vasoconstriction ($p = 0.0131$ and $p = 0.0055$, respectively). Their positive parameter estimates indicate that a higher inspiration rate or a larger volume of air intake is likely to increase the probability of vasoconstriction.

Output 54.6.1 Logistic Regression Analysis for Vasoconstriction Data

Occurrence of Vasoconstriction

The LOGISTIC Procedure

Model Information

Data Set	WORK.VASO
Response Variable	Response
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	39
Number of Observations Used	39

Response Profile

Ordered Value	Response	Total Frequency
1	constrict	20
2	no_constrict	19

Probability modeled is Response='constrict'.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	56.040	35.227
SC	57.703	40.218
-2 Log L	54.040	29.227

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	24.8125	2	<.0001
Score	16.6324	2	0.0002
Wald	7.8876	2	0.0194

Output 54.6.1 *continued*

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	−2.8754	1.3208	4.7395	0.0295
LogRate	1	4.5617	1.8380	6.1597	0.0131
LogVolume	1	5.1793	1.8648	7.7136	0.0055

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
LogRate	95.744	2.610	>999.999
LogVolume	177.562	4.592	>999.999

Association of Predicted Probabilities and Observed Responses

Percent Concordant	93.7	Somers' D	0.874
Percent Discordant	6.3	Gamma	0.874
Percent Tied	0.0	Tau-a	0.448
Pairs	380	c	0.937

The INFLUENCE option displays the values of the explanatory variables (LogRate and LogVolume) for each observation, a column for each diagnostic produced, and the *case number* that represents the sequence number of the observation (Output 54.6.2).

Output 54.6.2 Regression Diagnostics from the INFLUENCE Option

Regression Diagnostics

Case Number	Covariates LogRate	Log Volume	Pearson Residual	Deviance Residual	Hat Matrix Diagonal	Intercept DfBeta	LogRate DfBeta
1	-0.1924	1.3083	0.2205	0.3082	0.0927	-0.0165	0.0193
2	0.0862	1.2528	0.1349	0.1899	0.0429	-0.0134	0.0151
3	0.9163	0.2231	0.2923	0.4049	0.0612	-0.0492	0.0660
4	0.4055	-0.2877	3.5181	2.2775	0.0867	1.0734	-0.9302
5	1.1632	-0.2231	0.5287	0.7021	0.1158	-0.0832	0.1411
6	1.2528	-0.3567	0.6090	0.7943	0.1524	-0.0922	0.1710
7	-0.2877	-0.5108	-0.0328	-0.0464	0.00761	-0.00280	0.00274
8	0.5306	0.0953	-1.0196	-1.1939	0.0559	-0.1444	0.0613
9	-0.2877	-0.1054	-0.0938	-0.1323	0.0342	-0.0178	0.0173
10	-0.7985	-0.1054	-0.0293	-0.0414	0.00721	-0.00245	0.00246
11	-0.5621	-0.2231	-0.0370	-0.0523	0.00969	-0.00361	0.00358
12	1.0116	-0.5978	-0.5073	-0.6768	0.1481	-0.1173	0.0647
13	1.0986	-0.5108	-0.7751	-0.9700	0.1628	-0.0931	-0.00946
14	0.8459	0.3365	0.2559	0.3562	0.0551	-0.0414	0.0538
15	1.3218	-0.2877	0.4352	0.5890	0.1336	-0.0940	0.1408
16	0.4947	0.8329	0.1576	0.2215	0.0402	-0.0198	0.0234
17	0.4700	1.1632	0.0709	0.1001	0.0172	-0.00630	0.00701
18	0.3471	-0.1625	2.9062	2.1192	0.0954	0.9595	-0.8279
19	0.0583	0.5306	-1.0718	-1.2368	0.1315	-0.2591	0.2024
20	0.5878	0.5878	0.2405	0.3353	0.0525	-0.0331	0.0421
21	0.6931	-0.9163	-0.1076	-0.1517	0.0373	-0.0180	0.0158
22	0.3075	-0.0513	-0.4193	-0.5691	0.1015	-0.1449	0.1237
23	0.3001	0.3001	-1.0242	-1.1978	0.0761	-0.1961	0.1275
24	0.3075	0.4055	-1.3684	-1.4527	0.0717	-0.1281	0.0410
25	0.5766	0.4700	0.3347	0.4608	0.0587	-0.0403	0.0570
26	0.4055	-0.5108	-0.1595	-0.2241	0.0548	-0.0366	0.0329
27	0.4055	0.5878	0.3645	0.4995	0.0661	-0.0327	0.0496
28	0.6419	-0.0513	-0.8989	-1.0883	0.0647	-0.1423	0.0617
29	-0.0513	0.6419	0.8981	1.0876	0.1682	0.2367	-0.1950
30	-0.9163	0.4700	-0.0992	-0.1400	0.0507	-0.0224	0.0227
31	-0.2877	0.9933	0.6198	0.8064	0.2459	0.1165	-0.0996
32	-3.5066	0.8544	-0.00073	-0.00103	0.000022	-3.22E-6	3.405E-6
33	0.6043	0.0953	-1.2062	-1.3402	0.0510	-0.0882	-0.0137
34	0.7885	0.0953	0.5447	0.7209	0.0601	-0.0425	0.0877
35	0.6931	0.1823	0.5404	0.7159	0.0552	-0.0340	0.0755
36	1.2030	-0.2231	0.4828	0.6473	0.1177	-0.0867	0.1381
37	0.6419	-0.0513	-0.8989	-1.0883	0.0647	-0.1423	0.0617
38	0.6419	-0.2877	-0.4874	-0.6529	0.1000	-0.1395	0.1032
39	0.4855	0.2624	0.7053	0.8987	0.0531	0.0326	0.0190

Output 54.6.2 *continued*

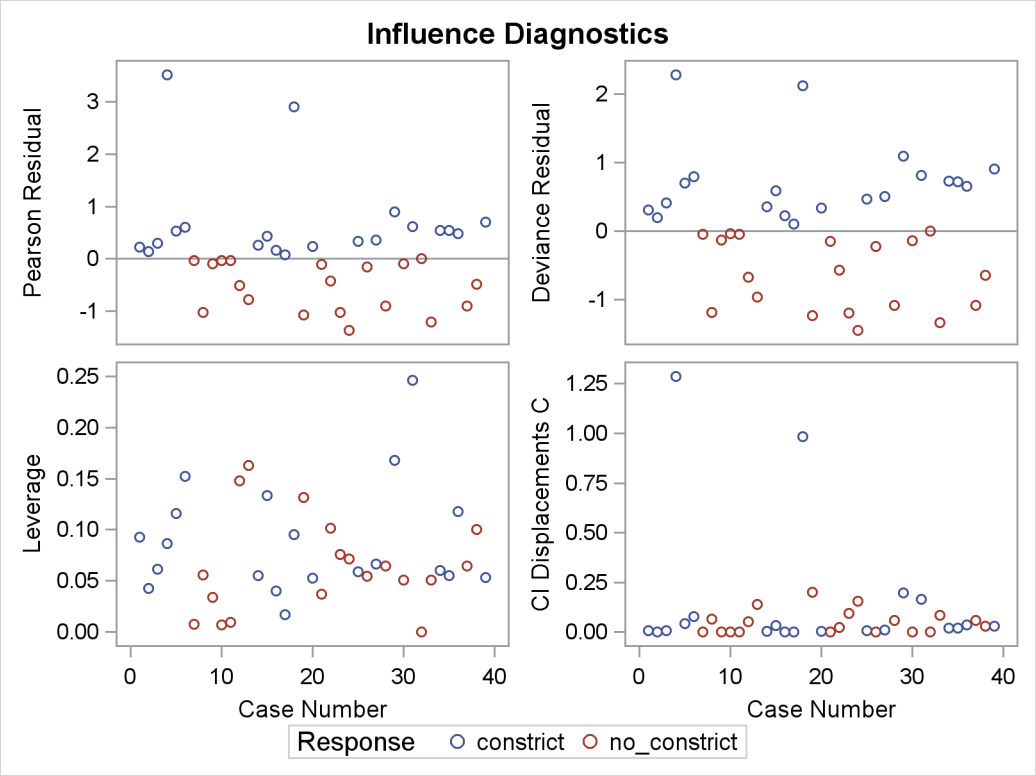
Regression Diagnostics

Case Number	Log Volume DfBeta	Confidence Interval Displacement C	Confidence Interval Displacement CBar	Delta Deviance	Delta Chi-Square
1	0.0556	0.00548	0.00497	0.1000	0.0536
2	0.0261	0.000853	0.000816	0.0369	0.0190
3	0.0589	0.00593	0.00557	0.1695	0.0910
4	-1.0180	1.2873	1.1756	6.3626	13.5523
5	0.0583	0.0414	0.0366	0.5296	0.3161
6	0.0381	0.0787	0.0667	0.6976	0.4376
7	0.00265	8.321E-6	8.258E-6	0.00216	0.00109
8	0.0570	0.0652	0.0616	1.4870	1.1011
9	0.0153	0.000322	0.000311	0.0178	0.00911
10	0.00211	6.256E-6	6.211E-6	0.00172	0.000862
11	0.00319	0.000014	0.000013	0.00274	0.00138
12	0.1651	0.0525	0.0447	0.5028	0.3021
13	0.1775	0.1395	0.1168	1.0577	0.7175
14	0.0527	0.00404	0.00382	0.1307	0.0693
15	0.0643	0.0337	0.0292	0.3761	0.2186
16	0.0307	0.00108	0.00104	0.0501	0.0259
17	0.00914	0.000089	0.000088	0.0101	0.00511
18	-0.8477	0.9845	0.8906	5.3817	9.3363
19	-0.00488	0.2003	0.1740	1.7037	1.3227
20	0.0518	0.00338	0.00320	0.1156	0.0610
21	0.0208	0.000465	0.000448	0.0235	0.0120
22	0.1179	0.0221	0.0199	0.3437	0.1956
23	0.0357	0.0935	0.0864	1.5212	1.1355
24	-0.1004	0.1558	0.1447	2.2550	2.0171
25	0.0708	0.00741	0.00698	0.2193	0.1190
26	0.0373	0.00156	0.00147	0.0517	0.0269
27	0.0788	0.0101	0.00941	0.2589	0.1423
28	0.1025	0.0597	0.0559	1.2404	0.8639
29	0.0286	0.1961	0.1631	1.3460	0.9697
30	0.0159	0.000554	0.000526	0.0201	0.0104
31	0.1322	0.1661	0.1253	0.7755	0.5095
32	2.48E-6	1.18E-11	1.18E-11	1.065E-6	5.324E-7
33	-0.00216	0.0824	0.0782	1.8744	1.5331
34	0.0671	0.0202	0.0190	0.5387	0.3157
35	0.0711	0.0180	0.0170	0.5295	0.3091
36	0.0631	0.0352	0.0311	0.4501	0.2641
37	0.1025	0.0597	0.0559	1.2404	0.8639
38	0.1397	0.0293	0.0264	0.4526	0.2639
39	0.0489	0.0295	0.0279	0.8355	0.5254

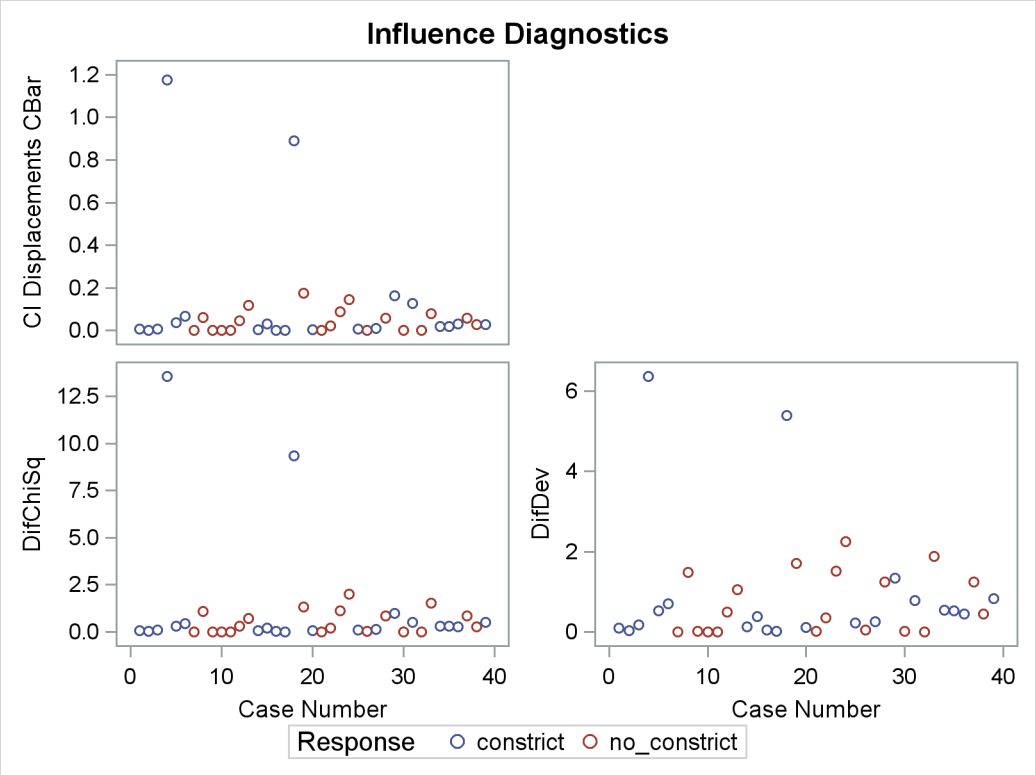
The index plots produced by the IPLOTS option are essentially the same line-printer plots as those produced by the INFLUENCE option, but with a 90-degree rotation and perhaps on a more refined scale. Since ODS Graphics is enabled, the line-printer plots from the INFLUENCE and IPLOTS options are suppressed and ODS Graphics versions of the plots are displayed in Outputs 54.6.3 through 54.6.5. For general information about ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS." For specific information about the graphics available in the LOGISTIC procedure, see the section "ODS Graphics" on page 4294. The vertical axis of an index plot represents the value of the diagnostic, and the horizontal axis represents the sequence (case number) of the observation. The index plots are useful for identification of extreme values.

The index plots of the Pearson residuals and the deviance residuals (Output 54.6.3) indicate that case 4 and case 18 are poorly accounted for by the model. The index plot of the diagonal elements of the hat matrix (Output 54.6.3) suggests that case 31 is an extreme point in the design space. The index plots of DFBETAS (Output 54.6.5) indicate that case 4 and case 18 are causing instability in all three parameter estimates. The other four index plots in Outputs 54.6.3 and 54.6.4 also point to these two cases as having a large impact on the coefficients and goodness of fit.

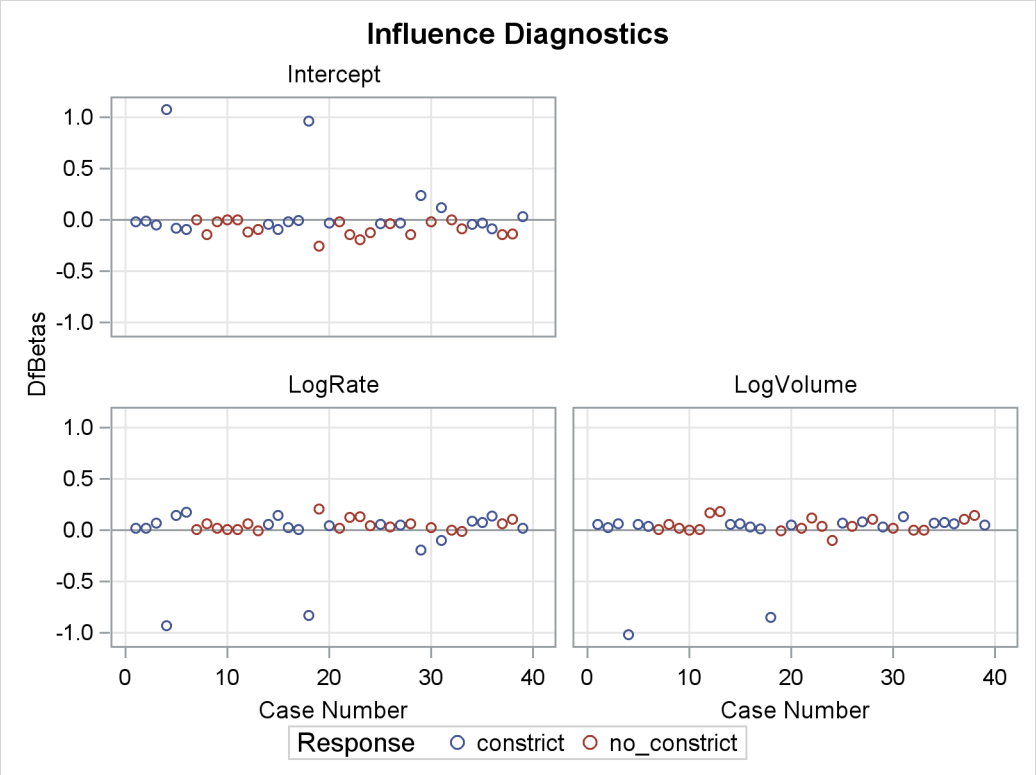
Output 54.6.3 Residuals, Hat Matrix, and CI Displacement C



Output 54.6.4 CI Displacement CBar, Change in Deviance and Pearson Chi-Square



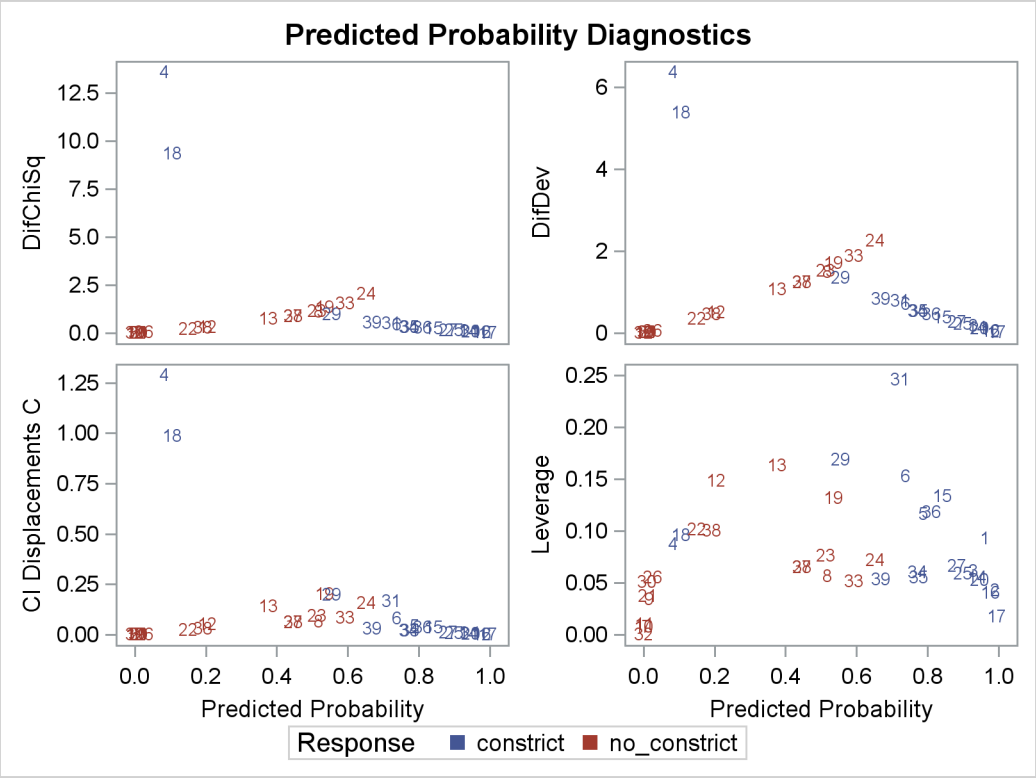
Output 54.6.5 DFBETAS Plots



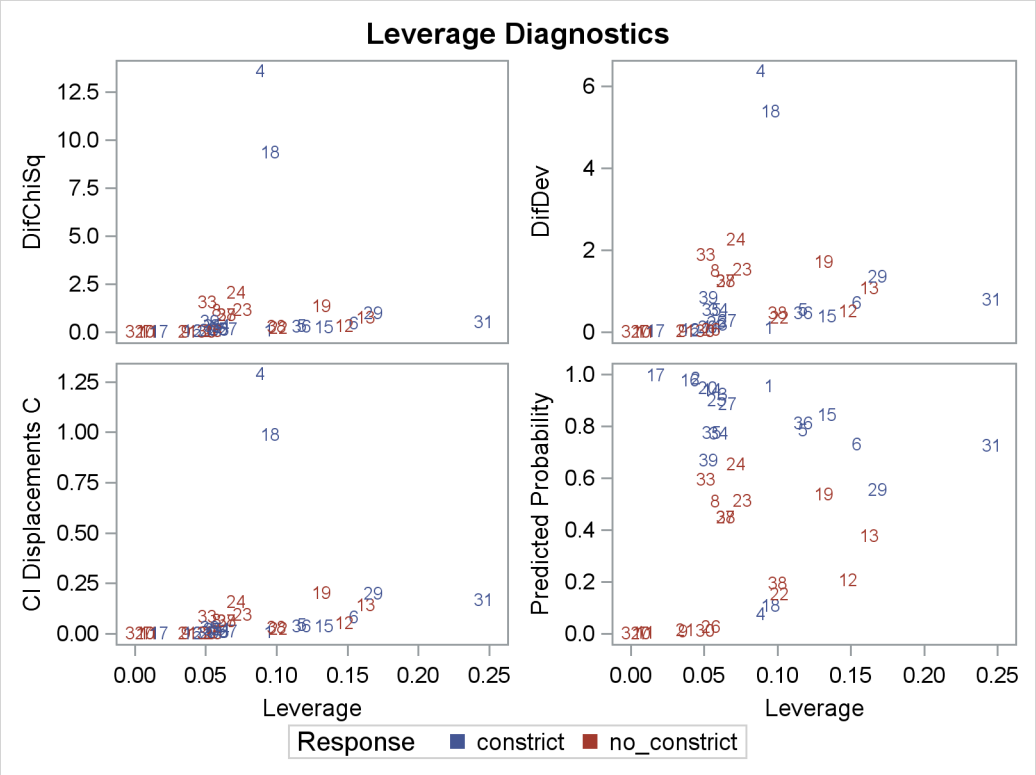
Other versions of diagnostic plots can be requested by specifying the appropriate options in the PLOTS= option. For example, the following statements produce three other sets of influence diagnostic plots: the PHAT option plots several diagnostics against the predicted probabilities (Output 54.6.6), the LEVERAGE option plots several diagnostics against the leverage (Output 54.6.7), and the DPC option plots the deletion diagnostics against the predicted probabilities and colors the observations according to the confidence interval displacement diagnostic (Output 54.6.8). The LABEL option displays the observation numbers on the plots. In all plots, you are looking for the outlying observations, and again cases 4 and 18 are noted.

```
ods graphics on;
proc logistic data=vaso plots(only label)=(phat leverage dpc);
   model Response=LogRate LogVolume;
run;
ods graphics off;
```

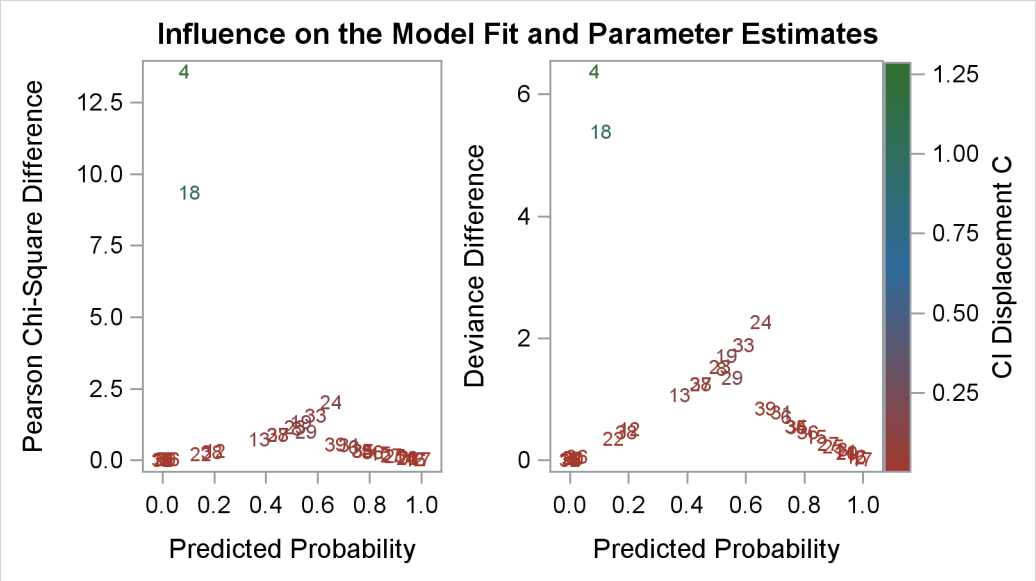
Output 54.6.6 Diagnostics versus Predicted Probability



Output 54.6.7 Diagnostics versus Leverage



Output 54.6.8 Three Diagnostics



Example 54.7: ROC Curve, Customized Odds Ratios, Goodness-of-Fit Statistics, R-Square, and Confidence Limits

This example plots an ROC curve, estimates a customized odds ratio, produces the traditional goodness-of-fit analysis, displays the generalized R Square measures for the fitted model, calculates the normal confidence intervals for the regression parameters, and produces a display of the probability function and prediction curves for the fitted model. The data consist of three variables: n (number of subjects in the sample), disease (number of diseased subjects in the sample), and age (age for the sample). A linear logistic regression model is used to study the effect of age on the probability of contracting the disease. The statements to produce the data set and perform the analysis are as follows:

```
data Data1;
   input disease n age;
   datalines;
 0 14 25
 0 20 35
 0 19 45
 7 18 55
 6 12 65
17 17 75
;

ods graphics on;
proc logistic data=Data1 plots(only)=roc(id=obs);
   model disease/n=age / scale=none
                         clparm=wald
                         clodds=pl
                         rsquare;
   units age=10;
   effectplot;
run;
ods graphics off;
```

The option SCALE=NONE is specified to produce the deviance and Pearson goodness-of-fit analysis without adjusting for overdispersion. The RSQUARE option is specified to produce generalized R Square measures of the fitted model. The CLPARM=WALD option is specified to produce the Wald confidence intervals for the regression parameters. The UNITS statement is specified to produce customized odds ratio estimates for a change of 10 years in the age variable, and the CLODDS=PL option is specified to produce profile-likelihood confidence limits for the odds ratio. The PLOTS= option with ODS Graphics enabled produces a graphical display of the ROC curve, and the EFFECTPLOT statement displays the model fit.

The results in Output 54.7.1 show that the deviance and Pearson statistics indicate no lack of fit in the model.

Output 54.7.1 Deviance and Pearson Goodness-of-Fit Analysis

The LOGISTIC Procedure

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	7.7756	4	1.9439	0.1002
Pearson	6.6020	4	1.6505	0.1585

Number of events/trials observations: 6

Output 54.7.2 shows that the R-square for the model is 0.74. The odds of an event increases by a factor of 7.9 for each 10-year increase in age.

Output 54.7.2 R-Square, Confidence Intervals, and Customized Odds Ratio

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates Log Likelihood	Full Log Likelihood
AIC	124.173	52.468	18.075
SC	126.778	57.678	23.285
-2 Log L	122.173	48.468	14.075

R-Square 0.5215 Max-rescaled R-Square 0.8925

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	73.7048	1	<.0001
Score	55.3274	1	<.0001
Wald	23.3475	1	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-12.5016	2.5555	23.9317	<.0001
age	1	0.2066	0.0428	23.3475	<.0001

Output 54.7.2 *continued*

Association of Predicted Probabilities and Observed Responses

Percent Concordant	92.6	Somers' D	0.906
Percent Discordant	2.0	Gamma	0.958
Percent Tied	5.4	Tau-a	0.384
Pairs	2100	c	0.953

Parameter Estimates and Wald Confidence Intervals

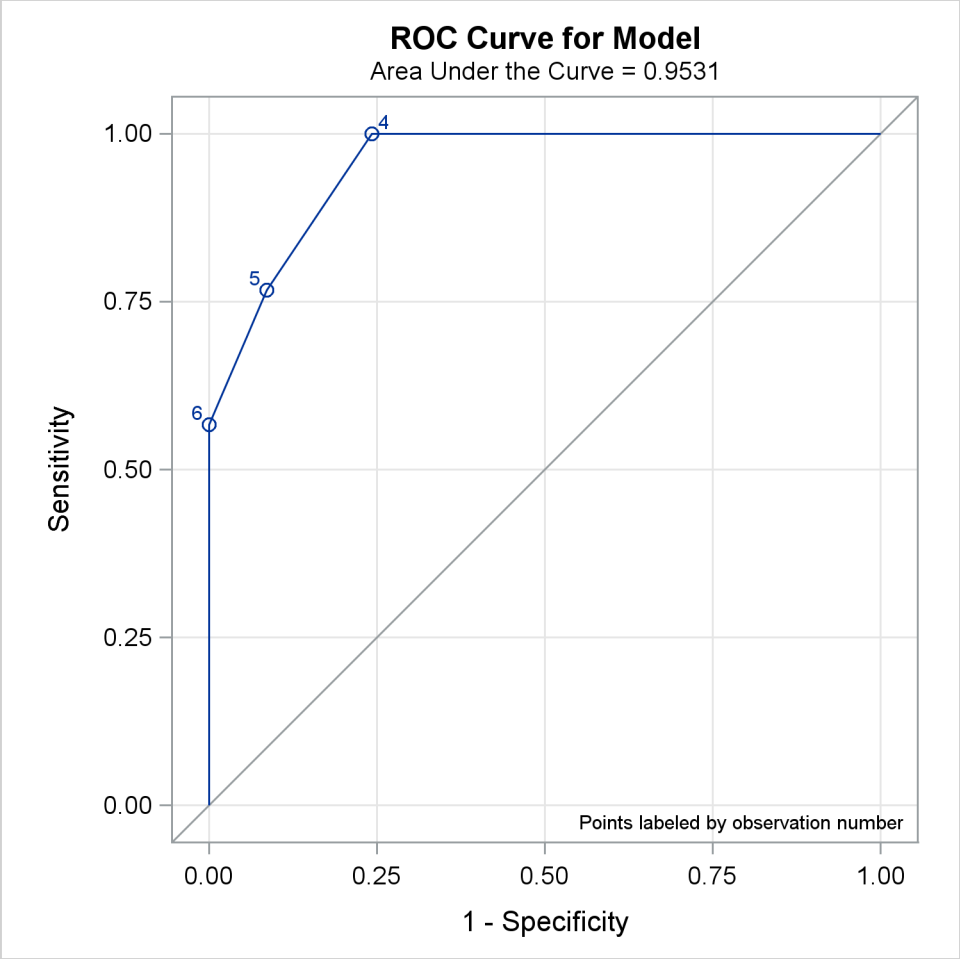
Parameter	Estimate	95% Confidence Limits	
Intercept	-12.5016	-17.5104	-7.4929
age	0.2066	0.1228	0.2904

Odds Ratio Estimates and Profile-Likelihood Confidence Intervals

Effect	Unit	Estimate	95% Confidence Limits	
age	10.0000	7.892	3.881	21.406

Since ODS Graphics is enabled, a graphical display of the ROC curve is produced as shown in Output 54.7.3.

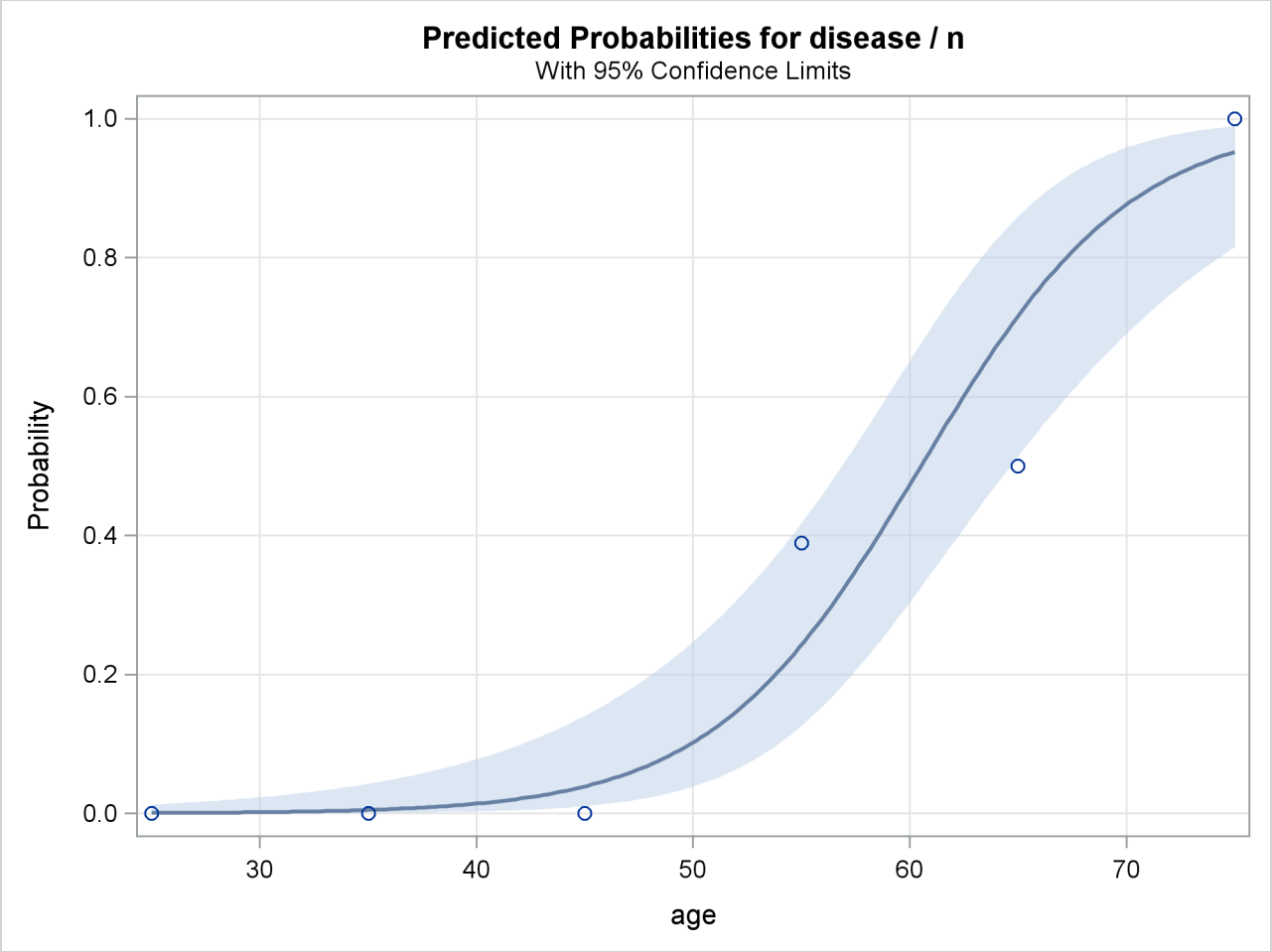
Output 54.7.3 Receiver Operating Characteristic Curve



Note that the area under the ROC curve is estimated by the statistic *c* in the "Association of Predicted Probabilities and Observed Responses" table. In this example, the area under the ROC curve is 0.953.

Since there is only one continuous covariate and since ODS Graphics is enabled, the EFFECTPLOT statement produces a graphical display of the predicted probability curve with bounding 95% confidence limits as shown in Output 54.7.4.

Output 54.7.4 Predicted Probability and 95% Prediction Limits



Example 54.8: Comparing Receiver Operating Characteristic Curves

DeLong, DeLong, and Clarke-Pearson (1988) report on 49 patients with ovarian cancer who also suffer from an intestinal obstruction. Three (correlated) screening tests are measured to determine whether a patient will benefit from surgery. The three tests are the K-G score and two measures of nutritional status: total protein and albumin. The data are as follows:

```
data roc;
   input alb tp totscore popind @@;
   totscore = 10 - totscore;
   datalines;
3.0 5.8 10 0   3.2 6.3  5 1   3.9 6.8  3 1   2.8 4.8  6 0
3.2 5.8  3 1   0.9 4.0  5 0   2.5 5.7  8 0   1.6 5.6  5 1
3.8 5.7  5 1   3.7 6.7  6 1   3.2 5.4  4 1   3.8 6.6  6 1
4.1 6.6  5 1   3.6 5.7  5 1   4.3 7.0  4 1   3.6 6.7  4 0
2.3 4.4  6 1   4.2 7.6  4 0   4.0 6.6  6 0   3.5 5.8  6 1
3.8 6.8  7 1   3.0 4.7  8 0   4.5 7.4  5 1   3.7 7.4  5 1
3.1 6.6  6 1   4.1 8.2  6 1   4.3 7.0  5 1   4.3 6.5  4 1
3.2 5.1  5 1   2.6 4.7  6 1   3.3 6.8  6 0   1.7 4.0  7 0
```

```
3.7 6.1  5 1   3.3 6.3  7 1   4.2 7.7  6 1   3.5 6.2  5 1
2.9 5.7  9 0   2.1 4.8  7 1   2.8 6.2  8 0   4.0 7.0  7 1
3.3 5.7  6 1   3.7 6.9  5 1   3.6 6.6  5 1
;
```

In the following statements, the NOFIT option is specified in the MODEL statement to prevent PROC LOGISTIC from fitting the model with three covariates. Each ROC statement lists one of the covariates, and PROC LOGISTIC then fits the model with that single covariate. Note that the original data set contains six more records with missing values for one of the tests, but PROC LOGISTIC ignores all records with missing values; hence there is a common sample size for each of the three models. The ROCCONTRAST statement implements the nonparameteric approach of DeLong, DeLong, and Clarke-Pearson (1988) to compare the three ROC curves, the REFERENCE option specifies that the K-G Score curve is used as the reference curve in the contrast, the E option displays the contrast coefficients, and the ESTIMATE option computes and tests each comparison. With ODS Graphics enabled, the **plots=roc(id=prob)** specification in the PROC LOGISTIC statement displays several plots, and the plots of individual ROC curves have certain points labeled with their predicted probabilities.

```
ods graphics on;
proc logistic data=roc plots=roc(id=prob);
   model popind(event='0') = alb tp totscore / nofit;
   roc 'Albumin' alb;
   roc 'K-G Score' totscore;
   roc 'Total Protein' tp;
   roccontrast reference('K-G Score') / estimate e;
run;
ods graphics off;
```

The initial model information is displayed in Output 54.8.1.

Output 54.8.1 Initial LOGISTIC Output

```
                    The LOGISTIC Procedure

                      Model Information

Data Set                          WORK.ROC
Response Variable                 popind
Number of Response Levels         2
Model                             binary logit
Optimization Technique            Fisher's scoring


   Number of Observations Read          43
   Number of Observations Used          43


                  Response Profile

   Ordered                        Total
     Value       popind     Frequency

         1            0            12
         2            1            31

    Probability modeled is popind=0.
```

Output 54.8.1 *continued*

Score Test for Global Null Hypothesis		
Chi-Square	DF	Pr > ChiSq
10.7939	3	0.0129

For each ROC model, the model fitting details in Outputs 54.8.2, 54.8.4, and 54.8.6 can be suppressed with the ROCOPTIONS(NODETAILS) option; however, the convergence status is always displayed.

The ROC curves for the three models are displayed in Outputs 54.8.3, 54.8.5, and 54.8.7. Note that the labels on the ROC curve are produced by specifying the ID=PROB option, and are the predicted probabilities for the cutpoints.

Output 54.8.2 Fit Tables for Popind=Alb

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	52.918	49.384
SC	54.679	52.907
-2 Log L	50.918	45.384

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.5339	1	0.0187
Score	5.6893	1	0.0171
Wald	4.6869	1	0.0304

Analysis of Maximum Likelihood Estimates

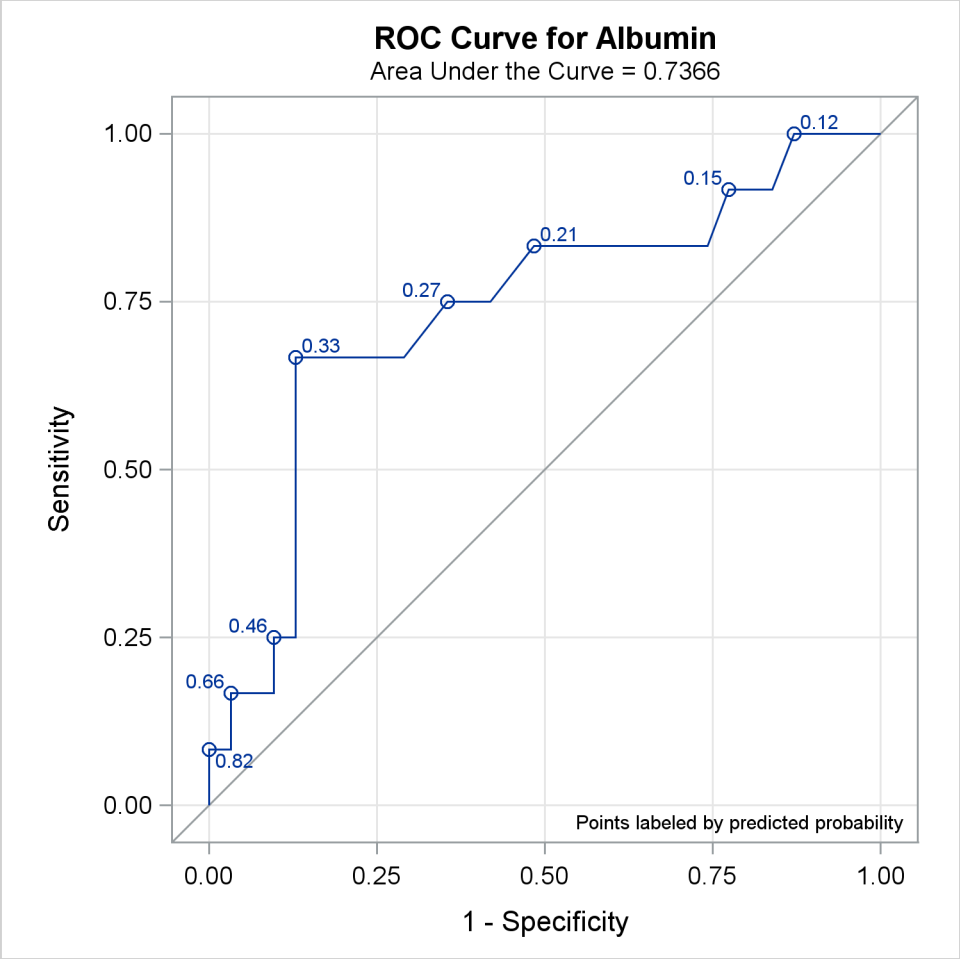
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.4646	1.5913	2.3988	0.1214
alb	1	-1.0520	0.4859	4.6869	0.0304

Output 54.8.2 *continued*

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
alb	0.349	0.135	0.905

Output 54.8.3 ROC Curve for Popind=Alb



Output 54.8.4 Fit Tables for Popind=Totscore

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	52.918	46.262
SC	54.679	49.784
-2 Log L	50.918	42.262

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.6567	1	0.0033
Score	8.3613	1	0.0038
Wald	6.3845	1	0.0115

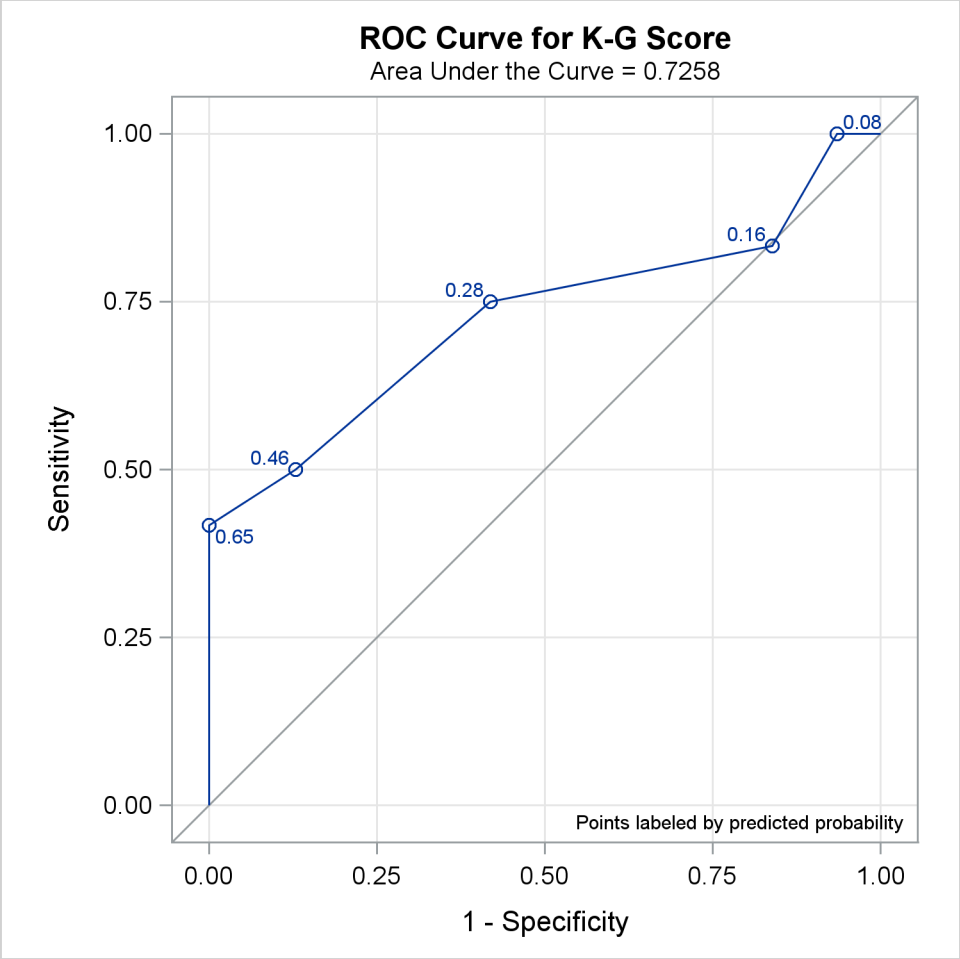
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.1542	1.2477	2.9808	0.0843
totscore	1	-0.7696	0.3046	6.3845	0.0115

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
totscore	0.463	0.255	0.841

Output 54.8.5 ROC Curve for Popind=Totscore



Output 54.8.6 Fit Tables for Popind=Tp

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	52.918	51.794
SC	54.679	55.316
-2 Log L	50.918	47.794

Output 54.8.6 *continued*

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3.1244	1	0.0771
Score	3.1123	1	0.0777
Wald	2.9059	1	0.0883

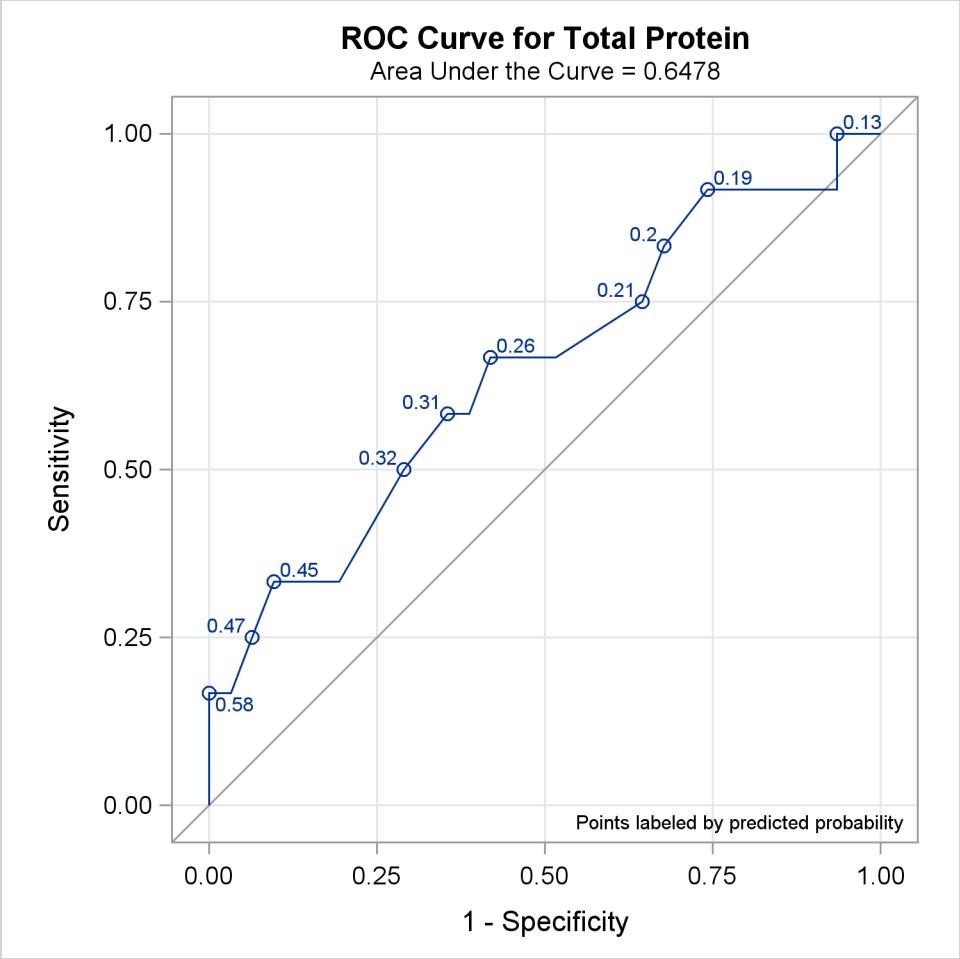
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.8295	2.2065	1.6445	0.1997
tp	1	-0.6279	0.3683	2.9059	0.0883

Odds Ratio Estimates

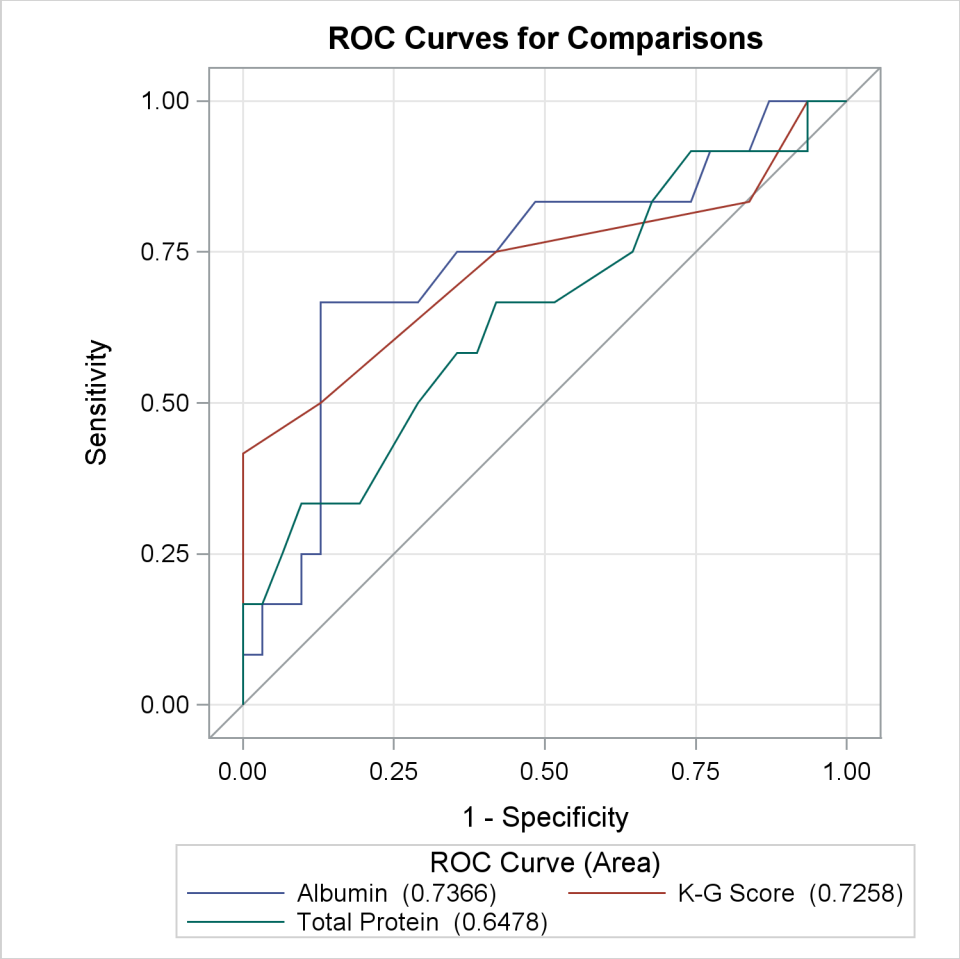
Effect	Point Estimate	95% Wald Confidence Limits	
tp	0.534	0.259	1.099

Output 54.8.7 ROC Curve for Popind=Tp



All ROC curves being compared are also overlaid on the same plot, as shown in Output 54.8.8.

Output 54.8.8 Overlay of All Models Being Compared



Output 54.8.9 displays the association statistics, and displays the area under the ROC curve along with its standard error and a confidence interval for each model in the comparison. The confidence interval for Total Protein contains 0.50; hence it is not significantly different from random guessing, which is represented by the diagonal line in the preceding ROC plots.

Output 54.8.9 ROC Association Table

ROC Association Statistics

	Mann-Whitney						
		Standard	95% Wald		Somers' D		
ROC Model	Area	Error	Confidence Limits		(Gini)	Gamma	Tau-a
Albumin	0.7366	0.0927	0.5549	0.9182	0.4731	0.4809	0.1949
K-G Score	0.7258	0.1028	0.5243	0.9273	0.4516	0.5217	0.1860
Total Protein	0.6478	0.1000	0.4518	0.8439	0.2957	0.3107	0.1218

Output 54.8.10 shows that the contrast used 'K-G Score' as the reference level. This table is produced by specifying the E option in the ROCCONTRAST statement.

Output 54.8.10 ROC Contrast Coefficients

ROC Contrast Coefficients

ROC Model	Row1	Row2
Albumin	1	0
K-G Score	-1	-1
Total Protein	0	1

Output 54.8.11 shows that the 2-degrees-of-freedom test that the 'K-G Score' is different from at least one other test is not significant at the 0.05 level.

Output 54.8.11 ROC Test Results (2 Degrees of Freedom)

ROC Contrast Test Results

Contrast	DF	Chi-Square	Pr > ChiSq
Reference = K-G Score	2	2.5340	0.2817

Output 54.8.12 is produced by specifying the ESTIMATE option in the ROCCONTRAST statement. Each row shows that the curves are not significantly different.

Output 54.8.12 ROC Contrast Row Estimates (1-Degree-of-Freedom Tests)

ROC Contrast Estimation and Testing Results by Row

Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
Albumin - K-G Score	0.0108	0.0953	-0.1761	0.1976	0.0127	0.9102
Total Protein - K-G Score	-0.0780	0.1046	-0.2830	0.1271	0.5554	0.4561

Example 54.9: Goodness-of-Fit Tests and Subpopulations

A study is done to investigate the effects of two binary factors, A and B, on a binary response, Y. Subjects are randomly selected from subpopulations defined by the four possible combinations of levels of A and B. The number of subjects responding with each level of Y is recorded, and the following DATA step creates the data set One:

```
data One;
   do A=0,1;
      do B=0,1;
         do Y=1,2;
            input F @@;
            output;
         end;
      end;
   end;
   datalines;
23 63 31 70 67 100 70 104
;
```

The following statements fit a full model to examine the main effects of A and B as well as the interaction effect of A and B:

```
proc logistic data=One;
   freq F;
   model Y=A B A*B;
run;
```

Results of the model fit are shown in Output 54.9.1. Notice that neither the A*B interaction nor the B main effect is significant.

Output 54.9.1 Full Model Fit

```
                    The LOGISTIC Procedure

                      Model Information

Data Set                          WORK.ONE
Response Variable                 Y
Number of Response Levels         2
Frequency Variable                F
Model                             binary logit
Optimization Technique            Fisher's scoring


   Number of Observations Read           8
   Number of Observations Used           8
   Sum of Frequencies Read             528
   Sum of Frequencies Used             528


                Response Profile

      Ordered                        Total
        Value             Y     Frequency

            1             1           191
            2             2           337

         Probability modeled is Y=1.
```

Output 54.9.1 *continued*

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	693.061	691.914
SC	697.330	708.990
-2 Log L	691.061	683.914

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7.1478	3	0.0673
Score	6.9921	3	0.0721
Wald	6.9118	3	0.0748

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.0074	0.2436	17.1015	<.0001
A	1	0.6069	0.2903	4.3714	0.0365
B	1	0.1929	0.3254	0.3515	0.5533
A*B	1	-0.1883	0.3933	0.2293	0.6321

Pearson and deviance goodness-of-fit tests cannot be obtained for this model since a full model containing four parameters is fit, leaving no residual degrees of freedom. For a binary response model, the goodness-of-fit tests have $m - q$ degrees of freedom, where m is the number of subpopulations and q is the number of model parameters. In the preceding model, $m = q = 4$, resulting in zero degrees of freedom for the tests.

The following statements fit a reduced model containing only the A effect, so two degrees of freedom become available for testing goodness of fit. Specifying the SCALE=NONE option requests the Pearson and deviance statistics. With single-trial syntax, the AGGREGATE= option is needed to define the subpopulations in the study. Specifying AGGREGATE=(A B) creates subpopulations of the four combinations of levels of A and B. Although the B effect is being dropped from the model, it is still needed to define the original subpopulations in the study. If AGGREGATE=(A) were specified, only two subpopulations would be created from the levels of A, resulting in $m = q = 2$ and zero degrees of freedom for the tests.

```
proc logistic data=One;
   freq F;
   model Y=A / scale=none aggregate=(A B);
run;
```

The goodness-of-fit tests in Output 54.9.2 show that dropping the B main effect and the A*B interaction simultaneously does not result in significant lack of fit of the model. The tests' large *p*-values indicate insufficient evidence for rejecting the null hypothesis that the model fits.

Output 54.9.2 Reduced Model Fit

The LOGISTIC Procedure

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	0.3541	2	0.1770	0.8377
Pearson	0.3531	2	0.1765	0.8382

Number of unique profiles: 4

Example 54.10: Overdispersion

In a seed germination test, seeds of two cultivars were planted in pots of two soil conditions. The following statements create the data set seeds, which contains the observed proportion of seeds that germinated for various combinations of cultivar and soil condition. The variable n represents the number of seeds planted in a pot, and the variable r represents the number germinated. The indicator variables cult and soil represent the cultivar and soil condition, respectively.

```
data seeds;
   input pot n r cult soil;
   datalines;
 1 16     8       0       0
 2 51    26       0       0
 3 45    23       0       0
 4 39    10       0       0
 5 36     9       0       0
 6 81    23       1       0
 7 30    10       1       0
 8 39    17       1       0
 9 28     8       1       0
10 62    23       1       0
11 51    32       0       1
12 72    55       0       1
13 41    22       0       1
14 12     3       0       1
15 13    10       0       1
16 79    46       1       1
17 30    15       1       1
18 51    32       1       1
19 74    53       1       1
20 56    12       1       1
;
```

PROC LOGISTIC is used as follows to fit a logit model to the data, with cult, soil, and cult × soil interaction as explanatory variables. The option SCALE=NONE is specified to display goodness-of-fit statistics.

```
proc logistic data=seeds;
   model r/n=cult soil cult*soil/scale=none;
   title 'Full Model With SCALE=NONE';
run;
```

Results of fitting the full factorial model are shown in Output 54.10.1. Both Pearson χ^2 and deviance are highly significant ($p < 0.0001$), suggesting that the model does not fit well.

Output 54.10.1 Results of the Model Fit for the Two-Way Layout

Full Model With SCALE=NONE

The LOGISTIC Procedure

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	68.3465	16	4.2717	<.0001
Pearson	66.7617	16	4.1726	<.0001

Number of events/trials observations: 20

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates Log Likelihood	Full Log Likelihood
AIC	1256.852	1213.003	156.533
SC	1261.661	1232.240	175.769
-2 Log L	1254.852	1205.003	148.533

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	49.8488	3	<.0001
Score	49.1682	3	<.0001
Wald	47.7623	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.3788	0.1489	6.4730	0.0110
cult	1	-0.2956	0.2020	2.1412	0.1434
soil	1	0.9781	0.2128	21.1234	<.0001
cult*soil	1	-0.1239	0.2790	0.1973	0.6569

If the link function and the model specification are correct and if there are no outliers, then the lack of fit might be due to overdispersion. Without adjusting for the overdispersion, the standard errors are likely to be underestimated, causing the Wald tests to be too sensitive. In PROC LOGISTIC, there are three SCALE= options to accommodate overdispersion. With unequal sample sizes for the observations, SCALE=WILLIAMS is preferred. The Williams model estimates a scale parameter ϕ by equating the value of Pearson χ^2 for the full model to its approximate expected value. The full model considered in the following statements is the model with cultivar, soil condition, and their interaction. Using a full model reduces the risk of contaminating ϕ with lack of fit due to incorrect model specification.

```
proc logistic data=seeds;
   model r/n=cult soil cult*soil / scale=williams;
   title 'Full Model With SCALE=WILLIAMS';
run;
```

Results of using Williams' method are shown in Output 54.10.2. The estimate of ϕ is 0.075941 and is given in the formula for the Weight Variable at the beginning of the displayed output.

Output 54.10.2 Williams' Model for Overdispersion

```
                  Full Model With SCALE=WILLIAMS

                      The LOGISTIC Procedure

                        Model Information

Data Set                          WORK.SEEDS
Response Variable (Events)        r
Response Variable (Trials)        n
Weight Variable                   1 / ( 1 + 0.075941 * (n - 1) )
Model                             binary logit
Optimization Technique            Fisher's scoring


           Number of Observations Read          20
           Number of Observations Used          20
           Sum of Frequencies Read             906
           Sum of Frequencies Used             906
           Sum of Weights Read            198.3216
           Sum of Weights Used            198.3216


                     Response Profile

      Ordered     Binary           Total            Total
        Value     Outcome      Frequency           Weight

            1     Event              437         92.95346
            2     Nonevent           469        105.36819


                Model Convergence Status

        Convergence criterion (GCONV=1E-8) satisfied.
```

Output 54.10.2 *continued*

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	16.4402	16	1.0275	0.4227
Pearson	16.0000	16	1.0000	0.4530

Number of events/trials observations: 20

NOTE: Since the Williams method was used to accommodate overdispersion, the Pearson chi-squared statistic and the deviance can no longer be used to assess the goodness of fit of the model.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates Log Likelihood	Full Log Likelihood
AIC	276.155	273.586	44.579
SC	280.964	292.822	63.815
-2 Log L	274.155	265.586	36.579

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.5687	3	0.0356
Score	8.4856	3	0.0370
Wald	8.3069	3	0.0401

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.3926	0.2932	1.7932	0.1805
cult	1	-0.2618	0.4160	0.3963	0.5290
soil	1	0.8309	0.4223	3.8704	0.0491
cult*soil	1	-0.0532	0.5835	0.0083	0.9274

Since neither cult nor cult × soil is statistically significant ($p = 0.5290$ and $p = 0.9274$, respectively), a reduced model that contains only the soil condition factor is fitted, with the observations weighted by $1/(1 + 0.075941(N - 1))$. This can be done conveniently in PROC LOGISTIC by including the scale estimate in the SCALE=WILLIAMS option as follows:

```
proc logistic data=seeds;
   model r/n=soil / scale=williams(0.075941);
   title 'Reduced Model With SCALE=WILLIAMS(0.075941)';
run;
```

Results of the reduced model fit are shown in Output 54.10.3. Soil condition remains a significant factor (p = 0.0064) for the seed germination.

Output 54.10.3 Reduced Model with Overdispersion Controlled

Reduced Model With SCALE=WILLIAMS(0.075941)

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.5249	0.2076	6.3949	0.0114
soil	1	0.7910	0.2902	7.4284	0.0064

Example 54.11: Conditional Logistic Regression for Matched Pairs Data

In matched pairs, or *case-control*, studies, conditional logistic regression is used to investigate the relationship between an outcome of being an event (case) or a nonevent (control) and a set of prognostic factors.

The following data are a subset of the data from the Los Angeles Study of the Endometrial Cancer Data in Breslow and Day (1980). There are 63 matched pairs, each consisting of a case of endometrial cancer (Outcome=1) and a control (Outcome=0). The case and corresponding control have the same ID. Two prognostic factors are included: Gall (an indicator variable for gall bladder disease) and Hyper (an indicator variable for hypertension). The goal of the case-control analysis is to determine the relative risk for gall bladder disease, controlling for the effect of hypertension.

```
data Data1;
   do ID=1 to 63;
      do Outcome = 1 to 0 by -1;
         input Gall Hyper @@;
         output;
      end;
   end;
   datalines;
0 0  0 0    0 0  0 0    0 1  0 1    0 0  1 0    1 0  0 1
0 1  0 0    1 0  0 0    1 1  0 1    0 0  0 0    0 0  0 0
1 0  0 0    0 0  0 1    1 0  0 1    1 0  1 0    1 0  0 1
0 1  0 0    0 0  1 1    0 0  1 1    0 0  0 1    0 1  0 0
0 0  1 1    0 1  0 1    0 1  0 0    0 0  0 0    0 0  0 0
0 0  0 1    1 0  0 1    0 0  0 1    1 0  0 0    0 1  0 0
0 1  0 0    0 1  0 0    0 1  0 0    0 0  0 0    1 1  1 1
0 0  0 1    0 1  0 0    0 1  0 1    0 1  0 1    0 1  0 0
0 0  0 0    0 1  1 0    0 0  0 1    0 0  0 0    1 0  0 0
0 0  0 0    1 1  0 0    0 1  0 0    0 0  0 0    0 1  0 1
0 0  0 0    0 1  0 1    0 1  0 0    0 1  0 0    1 0  0 0
0 0  0 0    1 1  1 0    0 0  0 0    0 0  0 0    1 1  0 0
1 0  1 0    0 1  0 0    1 0  0 0
;
```

There are several ways to approach this problem with PROC LOGISTIC:

- Specify the STRATA statement to perform a conditional logistic regression.
- Specify EXACT and STRATA statements to perform an exact logistic regression on the original data set, if you believe the data set is too small or too sparse for the usual asymptotics to hold.
- Transform each matched pair into a single observation, and then specify a PROC LOGISTIC statement on this transformed data without a STRATA statement; this also performs a conditional logistic regression and produces essentially the same results.
- Specify an EXACT statement on the transformed data.

SAS statements and selected results for these four approaches are given in the remainder of this example.

Conditional Analysis Using the STRATA Statement

In the following statements, PROC LOGISTIC is invoked with the ID variable declared in the STRATA statement to obtain the conditional logistic model estimates for a model containing Gall as the only predictor variable:

```
proc logistic data=Data1;
   strata ID;
   model outcome(event='1')=Gall;
run;
```

Results from the conditional logistic analysis are shown in Output 54.11.1. Note that there is no intercept term in the "Analysis of Maximum Likelihood Estimates" tables.

The odds ratio estimate for Gall is 2.60, which is marginally significant ($p = 0.0694$) and which is an estimate of the relative risk for gall bladder disease. A 95% confidence interval for this relative risk is (0.927, 7.293).

Output 54.11.1 Conditional Logistic Regression (Gall as Risk Factor)

```
                    The LOGISTIC Procedure

                     Conditional Analysis

                      Model Information

Data Set                          WORK.DATA1
Response Variable                 Outcome
Number of Response Levels         2
Number of Strata                  63
Model                             binary logit
Optimization Technique            Newton-Raphson ridge


     Number of Observations Read         126
     Number of Observations Used         126
```

Output 54.11.1 *continued*

Response Profile

Ordered Value	Outcome	Total Frequency
1	0	63
2	1	63

Probability modeled is Outcome=1.

Strata Summary

Response Pattern	Outcome 0	Outcome 1	Number of Strata	Frequency
1	1	1	63	126

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
AIC	87.337	85.654
SC	87.337	88.490
-2 Log L	87.337	83.654

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3.6830	1	0.0550
Score	3.5556	1	0.0593
Wald	3.2970	1	0.0694

Analysis of Conditional Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Gall	1	0.9555	0.5262	3.2970	0.0694

Output 54.11.1 *continued*

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Gall	2.600	0.927	7.293

Exact Analysis Using the STRATA Statement

When you believe there are not enough data or that the data are too sparse, you can perform a stratified exact logistic regression. The following statements perform stratified exact logistic regressions on the original data set by specifying both the STRATA and EXACT statements:

```
proc logistic data=Data1 exactonly;
   strata ID;
   model outcome(event='1')=Gall;
   exact Gall / estimate=both;
run;
```

Output 54.11.2 Exact Logistic Regression (Gall as Risk Factor)

The LOGISTIC Procedure

Exact Conditional Analysis

Exact Conditional Tests

Effect	Test	Statistic	--- p-Value --- Exact	Mid
Gall	Score	3.5556	0.0963	0.0799
	Probability	0.0327	0.0963	0.0799

Exact Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Two-sided p-Value
Gall	0.9555	0.5262	-0.1394	2.2316	0.0963

Exact Odds Ratios

Parameter	Estimate	95% Confidence Limits		Two-sided p-Value
Gall	2.600	0.870	9.315	0.0963

Note that the score statistic in the "Conditional Exact Tests" table in Output 54.11.2 is identical to the score statistic in Output 54.11.1 from the conditional analysis. The exact odds ratio confidence interval is much wider than its conditional analysis counterpart, but the parameter estimates are similar. The exact analysis confirms the marginal significance of Gall as a predictor variable.

Conditional Analysis Using Transformed Data

When each matched set consists of one event and one nonevent, the conditional likelihood is given by

$$\prod_i (1 + \exp(-\boldsymbol{\beta}'(\mathbf{x}_{i1} - \mathbf{x}_{i0}))^{-1}$$

where $\mathbf{x}_{i1}$ and $\mathbf{x}_{i0}$ are vectors representing the prognostic factors for the event and nonevent, respectively, of the ith matched set. This likelihood is identical to the likelihood of fitting a logistic regression model to a set of data with constant response, where the model contains no intercept term and has explanatory variables given by $\mathbf{d}_i = \mathbf{x}_{i1} - \mathbf{x}_{i0}$ (Breslow 1982).

To apply this method, the following DATA step transforms each matched pair into a single observation, where the variables Gall and Hyper contain the differences between the corresponding values for the case and the control (case–control). The variable Outcome, which will be used as the response variable in the logistic regression model, is given a constant value of 0 (which is the Outcome value for the control, although any constant, numeric or character, will suffice).

```
data Data2;
   set Data1;
   drop id1 gall1 hyper1;
   retain id1 gall1 hyper1 0;
   if (ID = id1) then do;
      Gall=gall1-Gall; Hyper=hyper1-Hyper;
      output;
   end;
   else do;
      id1=ID; gall1=Gall; hyper1=Hyper;
   end;
run;
```

Note that there are 63 observations in the data set, one for each matched pair. Since the number of observations n is halved, statistics that depend on n such as R Square (see the "Generalized Coefficient of Determination" on page 4246 section) will be incorrect. The variable Outcome has a constant value of 0.

In the following statements, PROC LOGISTIC is invoked with the NOINT option to obtain the conditional logistic model estimates. Because the option CLODDS=PL is specified, PROC LOGISTIC computes a 95% profile-likelihood confidence interval for the odds ratio for each predictor variable; note that profile-likelihood confidence intervals are not currently available when a STRATA statement is specified.

```
proc logistic data=Data2;
   model outcome=Gall / noint clodds=PL;
run;
```

The results are not displayed here.

Exact Analysis Using Transformed Data

Sometimes the original data set in a matched-pairs study is too large for the exact methods to handle. In such cases it might be possible to use the transformed data set. The following statements perform exact logistic regressions on the transformed data set. The results are not displayed here.

```
proc logistic data=Data2 exactonly;
   model outcome=Gall / noint;
   exact Gall / estimate=both;
run;
```

Example 54.12: Firth's Penalized Likelihood Compared with Other Approaches

Firth's penalized likelihood approach is a method of addressing issues of separability, small sample sizes, and bias of the parameter estimates. This example performs some comparisons between results from using the FIRTH option to results from the usual unconditional, conditional, and exact logistic regression analyses. When the sample size is large enough, the unconditional estimates and the Firth penalized-likelihood estimates should be nearly the same. These examples show that Firth's penalized likelihood approach compares favorably with unconditional, conditional, and exact logistic regression; however, this is not an exhaustive analysis of Firth's method. For more detailed analyses with separable data sets, see Heinze (2006, 1999) and Heinze and Schemper (2002).

Comparison on 2x2 Tables with One Zero Cell

A 2×2 table with one cell having zero frequency, where the rows of the table are the levels of a covariate while the columns are the levels of the response variable, is an example of a quasi-completely separated data set. The parameter estimate for the covariate under unconditional logistic regression will move off to infinity, although PROC LOGISTIC will stop the iterations at an earlier point in the process. An exact logistic regression is sometimes performed to determine the importance of the covariate in describing the variation in the data, but the median-unbiased parameter estimate, while finite, might not be near the true value, and one confidence limit (for this example, the upper) is always infinite.

The following DATA step produces 1000 different 2×2 tables, all following an underlying probability structure, with one cell having a near zero probability of being observed:

```
%let beta0=-15;
%let beta1=16;
data one;
   keep sample X y pry;
   do sample=1 to 1000;
      do i=1 to 100;
         X=rantbl(987987,.4,.6)-1;
         xb= &beta0 + X*&beta1;
         exb=exp(xb);
         pry= exb/(1+exb);
         cut= ranuni(393993);
         if (pry < cut) then y=1; else y=0;
         output;
      end;
```

```
   end;
run;
```

The following statements perform the bias-corrected and exact logistic regression on each of the 1000 different data sets, output the odds ratio tables by using the ODS OUTPUT statement, and compute various statistics across the data sets by using the MEANS procedure:

```
ods exclude all;
proc logistic data=one;
   by sample;
   class X(param=ref);
   model y(event='1')=X / firth clodds=pl;
   ods output cloddspl=firth;
run;
proc logistic data=one exactonly;
   by sample;
   class X(param=ref);
   model y(event='1')=X;
   exact X / estimate=odds;
   ods output exactoddsratio=exact;
run;
ods select all;
proc means data=firth;
   var LowerCL OddsRatioEst UpperCL;
run;
proc means data=exact;
   var LowerCL Estimate UpperCL;
run;
```

The results of the PROC MEANS statements are summarized in Table 54.16. You can see that the odds ratios are all quite large; the confidence limits on every table suggest that the covariate X is a significant factor in explaining the variability in the data.

Table 54.16 Odds Ratio Results

Method	Mean Estimate	Standard Error	Minimum Lower CL	Maximum Upper CL
Firth	231.59	83.57	10.40	111317
Exact	152.02	52.30	11.16	∞

Comparison on Case-Control Data

Case-control models contain an intercept term for every case-control pair in the data set. This means that there are a large number of parameters compared to the number of observations. Breslow and Day (1980) note that the estimates from unconditional logistic regression are biased with the corresponding odds ratios off by a power of 2 from the true value; conditional logistic regression was developed to remedy this.

The following DATA step produces 1000 case-control data sets, with pair indicating the strata:

```
%let beta0=1;
%let beta1=2;
data one;
   do sample=1 to 1000;
      do pair=1 to 20;
         ran=ranuni(939393);
         a=3*ranuni(9384984)-1;
         pdf0= pdf('NORMAL',a,.4,1);
         pdf1= pdf('NORMAL',a,1,1);
         pry0= pdf0/(pdf0+pdf1);
         pry1= 1-pry0;
         xb= log(pry0/pry1);
         x= (xb-&beta0*pair/100) / &beta1;
         y=0;
         output;
         x= (-xb-&beta0*pair/100) / &beta1;
         y=1;
         output;
      end;
   end;
run;
```

Unconditional, conditional, exact, and Firth-adjusted analyses are performed on the data sets, and the mean, minimum, and maximum odds ratios and the mean upper and lower limits for the odds ratios are displayed in Table 54.17. **CAUTION:** Due to the exact analyses, this program takes a long time and a lot of resources to run. You might want to reduce the number of samples generated.

```
ods exclude all;
proc logistic data=one;
   by sample;
   class pair / param=ref;
   model y=x pair / clodds=pl;
   ods output cloddspl=oru;
run;
data oru;
   set oru;
   if Effect='x';
   rename lowercl=lclu uppercl=uclu oddsratioest=orestu;
run;
proc logistic data=one;
   by sample;
   strata pair;
   model y=x / clodds=wald;
   ods output cloddswald=orc;
run;
data orc;
   set orc;
   if Effect='x';
   rename lowercl=lclc uppercl=uclc oddsratioest=orestc;
run;
proc logistic data=one exactonly;
   by sample;
   strata pair;
```

```
   model y=x;
   exact x / estimate=both;
   ods output ExactOddsRatio=ore;
run;
proc logistic data=one;
   by sample;
   class pair / param=ref;
   model y=x pair / firth clodds=pl;
   ods output cloddspl=orf;
run;
data orf;
   set orf;
   if Effect='x';
   rename lowercl=lclf uppercl=uclf oddsratioest=orestf;
run;
data all;
   merge oru orc ore orf;
run;
ods select all;
proc means data=all;
run;
```

You can see from Table 54.17 that the conditional, exact, and Firth-adjusted results are all comparable, while the unconditional results are several orders of magnitude different.

Table 54.17 Odds Ratio Estimates

Method	N	Minimum	Mean	Maximum
Unconditional	1000	0.00045	112.09	38038
Conditional	1000	0.021	4.20	195
Exact	1000	0.021	4.20	195
Firth	1000	0.018	4.89	71

Further examination of the data set all shows that the differences between the square root of the unconditional odds ratio estimates and the conditional estimates have mean –0.00019 and standard deviation 0.0008, verifying that the unconditional odds ratio is about the square of the conditional odds ratio. The conditional and exact conditional odds ratios are also nearly equal, with their differences having mean 3E–7 and standard deviation 6E–6. The differences between the Firth and the conditional odds ratios can be large (mean 0.69, standard deviation 5.40), but their relative differences, $\frac{Firth - Conditional}{Conditional}$, have mean 0.20 with standard deviation 0.19, so the largest differences occur with the larger estimates.

Example 54.13: Complementary Log-Log Model for Infection Rates

Antibodies produced in response to an infectious disease like malaria remain in the body after the individual has recovered from the disease. A serological test detects the presence or absence of such antibodies. An individual with such antibodies is called seropositive. In geographic areas where the disease is endemic, the inhabitants are at fairly constant risk of infection. The probability of an individual never having been

infected in Y years is $\exp(-\mu Y)$, where μ is the mean number of infections per year (see the appendix of Draper, Voller, and Carpenter 1972). Rather than estimating the unknown μ, epidemiologists want to estimate the probability of a person living in the area being infected in one year. This infection rate γ is given by

$$\gamma = 1 - e^{-\mu}$$

The following statements create the data set sero, which contains the results of a serological survey of malarial infection. Individuals of nine age groups (Group) were tested. The variable A represents the midpoint of the age range for each age group. The variable N represents the number of individuals tested in each age group, and the variable R represents the number of individuals that are seropositive.

```
data sero;
   input Group A N R;
   X=log(A);
   label X='Log of Midpoint of Age Range';
   datalines;
1  1.5  123  8
2  4.0  132  6
3  7.5  182 18
4 12.5  140 14
5 17.5  138 20
6 25.0  161 39
7 35.0  133 19
8 47.0   92 25
9 60.0   74 44
;
```

For the ith group with the age midpoint A_i, the probability of being seropositive is $p_i = 1 - \exp(-\mu A_i)$. It follows that

$$\log(-\log(1 - p_i)) = \log(\mu) + \log(A_i)$$

By fitting a binomial model with a complementary log-log link function and by using X=log(A) as an offset term, you can estimate $\alpha = \log(\mu)$ as an intercept parameter. The following statements invoke PROC LOGISTIC to compute the maximum likelihood estimate of α. The LINK=CLOGLOG option is specified to request the complementary log-log link function. Also specified is the CLPARM=PL option, which requests the profile-likelihood confidence limits for α.

```
proc logistic data=sero;
   model R/N= / offset=X
                link=cloglog
                clparm=pl
                scale=none;
   title 'Constant Risk of Infection';
run;
```

Results of fitting this constant risk model are shown in Output 54.13.1.

Output 54.13.1 Modeling Constant Risk of Infection

```
                          Constant Risk of Infection

                            The LOGISTIC Procedure

                              Model Information

Data Set                          WORK.SERO
Response Variable (Events)        R
Response Variable (Trials)        N
Offset Variable                   X                    Log of Midpoint of Age Range
Model                             binary cloglog
Optimization Technique            Fisher's scoring


                     Number of Observations Read           9
                     Number of Observations Used           9
                     Sum of Frequencies Read            1175
                     Sum of Frequencies Used            1175


                               Response Profile

                        Ordered     Binary           Total
                          Value     Outcome      Frequency

                              1     Event              193
                              2     Nonevent           982


                    Intercept-Only Model Convergence Status

                  Convergence criterion (GCONV=1E-8) satisfied.


                               -2 Log L = 967.1158


                Deviance and Pearson Goodness-of-Fit Statistics

          Criterion          Value       DF     Value/DF     Pr > ChiSq

          Deviance         41.5032        8       5.1879         <.0001
          Pearson          50.6883        8       6.3360         <.0001

                     Number of events/trials observations: 9
```

Output 54.13.1 *continued*

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.6605	0.0725	4133.5626	<.0001
X	0	1.0000	0	.	.

Parameter Estimates and Profile-Likelihood Confidence Intervals

Parameter	Estimate	95% Confidence Limits	
Intercept	-4.6605	-4.8057	-4.5219

Output 54.13.1 shows that the maximum likelihood estimate of $\alpha = \log(\mu)$ and its estimated standard error are $\widehat{\alpha} = -4.6605$ and $\widehat{\sigma}_{\widehat{\alpha}} = 0.0725$, respectively. The infection rate is estimated as

$$\widehat{\gamma} = 1 - e^{-\widehat{\mu}} = 1 - e^{-e^{\widehat{\beta}_0}} = 1 - e^{-e^{-4.6605}} = 0.00942$$

The 95% confidence interval for γ, obtained by back-transforming the 95% confidence interval for α, is (0.0082, 0.0108); that is, there is a 95% chance that, in repeated sampling, the interval of 8 to 11 infections per thousand individuals contains the true infection rate.

The goodness-of-fit statistics for the constant risk model are statistically significant ($p < 0.0001$), indicating that the assumption of constant risk of infection is not correct. You can fit a more extensive model by allowing a separate risk of infection for each age group. Suppose μ_i is the mean number of infections per year for the ith age group. The probability of seropositive for the ith group with the age midpoint A_i is $p_i = 1 - \exp(-\mu_i A_i)$, so that

$$\log(-\log(1 - p_i)) = \log(\mu_i) + \log(A_i)$$

In the following statements, a complementary log-log model is fit containing Group as an explanatory classification variable with the GLM coding (so that a dummy variable is created for each age group), no intercept term, and X=log(A) as an offset term. The ODS OUTPUT statement saves the estimates and their 95% profile-likelihood confidence limits to the ClparmPL data set. Note that $\log(\mu_i)$ is the regression parameter associated with Group$= i$.

```
proc logistic data=sero;
   ods output ClparmPL=ClparmPL;
   class Group / param=glm;
   model R/N=Group / noint
                     offset=X
                     link=cloglog
                     clparm=pl;
   title 'Infectious Rates and 95% Confidence Intervals';
run;
```

Results of fitting the model with a separate risk of infection are shown in Output 54.13.2.

Output 54.13.2 Modeling Separate Risk of Infection

Infectious Rates and 95% Confidence Intervals

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Group	1	1	-3.1048	0.3536	77.0877	<.0001
Group	2	1	-4.4542	0.4083	119.0164	<.0001
Group	3	1	-4.2769	0.2358	328.9593	<.0001
Group	4	1	-4.7761	0.2674	319.0600	<.0001
Group	5	1	-4.7165	0.2238	443.9920	<.0001
Group	6	1	-4.5012	0.1606	785.1350	<.0001
Group	7	1	-5.4252	0.2296	558.1114	<.0001
Group	8	1	-4.9987	0.2008	619.4666	<.0001
Group	9	1	-4.1965	0.1559	724.3157	<.0001
X		0	1.0000	0	.	.

Parameter Estimates and Profile-Likelihood Confidence Intervals

Parameter		Estimate	95% Confidence Limits	
Group	1	-3.1048	-3.8880	-2.4833
Group	2	-4.4542	-5.3769	-3.7478
Group	3	-4.2769	-4.7775	-3.8477
Group	4	-4.7761	-5.3501	-4.2940
Group	5	-4.7165	-5.1896	-4.3075
Group	6	-4.5012	-4.8333	-4.2019
Group	7	-5.4252	-5.9116	-5.0063
Group	8	-4.9987	-5.4195	-4.6289
Group	9	-4.1965	-4.5164	-3.9037

For the first age group (Group=1), the point estimate of $\log(\mu_1)$ is –3.1048, which transforms into an infection rate of $1 - \exp(-\exp(-3.1048)) = 0.0438$. A 95% confidence interval for this infection rate is obtained by transforming the 95% confidence interval for $\log(\mu_1)$. For the first age group, the lower and upper confidence limits are $1 - \exp(-\exp(-3.8880) = 0.0203$ and $1 - \exp(-\exp(-2.4833)) = 0.0801$, respectively; that is, there is a 95% chance that, in repeated sampling, the interval of 20 to 80 infections per thousand individuals contains the true infection rate. The following statements perform this transformation on the estimates and confidence limits saved in the ClparmPL data set; the resulting estimated infection rates in one year's time for each age group are displayed in Table 54.18. Note that the infection rate for the first age group is high compared to that of the other age groups.

```
data ClparmPL;
   set ClparmPL;
   Estimate=round( 1000*( 1-exp(-exp(Estimate)) ) );
   LowerCL =round( 1000*( 1-exp(-exp(LowerCL )) ) );
   UpperCL =round( 1000*( 1-exp(-exp(UpperCL )) ) );
run;
```

Table 54.18 Infection Rate in One Year

Age Group	Number Infected per 1,000 People		
	Point Estimate	95% Confidence Limits	
		Lower	Upper
1	44	20	80
2	12	5	23
3	14	8	21
4	8	5	14
5	9	6	13
6	11	8	15
7	4	3	7
8	7	4	10
9	15	11	20

Example 54.14: Complementary Log-Log Model for Interval-Censored Survival Times

Often survival times are not observed more precisely than the interval (for instance, a day) within which the event occurred. Survival data of this form are known as grouped or interval-censored data. A discrete analog of the continuous proportional hazards model (Prentice and Gloeckler 1978; Allison 1982) is used to investigate the relationship between these survival times and a set of explanatory variables.

Suppose T_i is the discrete survival time variable of the ith subject with covariates $\mathbf{x}_i$. The discrete-time hazard rate λ_{it} is defined as

$$\lambda_{it} = \Pr(T_i = t \mid T_i \geq t, \mathbf{x}_i), \quad t = 1, 2, \ldots$$

Using elementary properties of conditional probabilities, it can be shown that

$$\Pr(T_i = t) = \lambda_{it} \prod_{j=1}^{t-1} (1 - \lambda_{ij}) \quad \text{and} \quad \Pr(T_i > t) = \prod_{j=1}^{t} (1 - \lambda_{ij})$$

Suppose t_i is the observed survival time of the ith subject. Suppose $\delta_i = 1$ if $T_i = t_i$ is an event time and 0 otherwise. The likelihood for the grouped survival data is given by

$$\begin{aligned} L &= \prod_i [\Pr(T_i = t_i)]^{\delta_i} [\Pr(T_i > t_i)]^{1-\delta_i} \\ &= \prod_i \left(\frac{\lambda_{it_i}}{1 - \lambda_{it_i}} \right)^{\delta_i} \prod_{j=1}^{t_i} (1 - \lambda_{ij}) \\ &= \prod_i \prod_{j=1}^{t_i} \left(\frac{\lambda_{ij}}{1 - \lambda_{ij}} \right)^{y_{ij}} (1 - \lambda_{ij}) \end{aligned}$$

where $y_{ij} = 1$ if the ith subject experienced an event at time $T_i = j$ and 0 otherwise.

Note that the likelihood L for the grouped survival data is the same as the likelihood of a binary response model with event probabilities λ_{ij}. If the data are generated by a continuous-time proportional hazards model, Prentice and Gloeckler (1978) have shown that

$$\lambda_{ij} = 1 - \exp(-\exp(\alpha_j + \boldsymbol{\beta}'\mathbf{x}_i))$$

which can be rewritten as

$$\log(-\log(1 - \lambda_{ij})) = \alpha_j + \boldsymbol{\beta}'\mathbf{x}_i$$

where the coefficient vector $\boldsymbol{\beta}$ is identical to that of the continuous-time proportional hazards model, and α_j is a constant related to the conditional survival probability in the interval defined by $T_i = j$ at $\mathbf{x}_i = \mathbf{0}$. The grouped data survival model is therefore equivalent to the binary response model with complementary log-log link function. To fit the grouped survival model by using PROC LOGISTIC, you must treat each discrete time unit for each subject as a separate observation. For each of these observations, the response is dichotomous, corresponding to whether or not the subject died in the time unit.

Consider a study of the effect of insecticide on flour beetles. Four different concentrations of an insecticide were sprayed on separate groups of flour beetles. The following DATA step saves the number of male and female flour beetles dying in successive intervals in the data set Beetles:

```
data Beetles(keep=time sex conc freq);
   input time m20 f20 m32 f32 m50 f50 m80 f80;
   conc=.20; freq= m20; sex=1; output;
             freq= f20; sex=2; output;
   conc=.32; freq= m32; sex=1; output;
             freq= f32; sex=2; output;
   conc=.50; freq= m50; sex=1; output;
             freq= f50; sex=2; output;
   conc=.80; freq= m80; sex=1; output;
             freq= f80; sex=2; output;
   datalines;
 1   3   0  7  1  5  0  4  2
 2  11   2 10  5  8  4 10  7
 3  10   4 11 11 11  6  8 15
 4   7   8 16 10 15  6 14  9
 5   4   9  3  5  4  3  8  3
 6   3   3  2  1  2  1  2  4
 7   2   0  1  0  1  1  1  1
 8   1   0  0  1  1  4  0  1
 9   0   0  1  1  0  0  0  0
10   0   0  0  0  0  0  1  1
11   0   0  0  0  1  1  0  0
12   1   0  0  0  0  1  0  0
13   1   0  0  0  0  1  0  0
14 101 126 19 47  7 17  2  4
;
```

The data set Beetles contains four variables: time, sex, conc, and freq. The variable time represents the interval death time; for example, time=2 is the interval between day 1 and day 2. Insects surviving the duration (13 days) of the experiment are given a time value of 14. The variable sex represents the sex of the insects (1=male, 2=female), conc represents the concentration of the insecticide (mg/cm^2), and freq represents the frequency of the observations.

To use PROC LOGISTIC with the grouped survival data, you must expand the data so that each beetle has a separate record for each day of survival. A beetle that died in the third day (time=3) would contribute three observations to the analysis, one for each day it was alive at the beginning of the day. A beetle that survives the 13-day duration of the experiment (time=14) would contribute 13 observations.

The following DATA step creates a new data set named Days containing the beetle-day observations from the data set Beetles. In addition to the variables sex, conc, and freq, the data set contains an outcome variable y and a classification variable day. The variable y has a value of 1 if the observation corresponds to the day that the beetle died, and it has a value of 0 otherwise. An observation for the first day will have a value of 1 for day; an observation for the second day will have a value of 2 for day, and so on. For instance, Output 54.14.1 shows an observation in the Beetles data set with time=3, and Output 54.14.2 shows the corresponding beetle-day observations in the data set Days.

```
data Days;
   set Beetles;
   do day=1 to time;
      if (day < 14) then do;
         y= (day=time);
         output;
      end;
   end;
run;
```

Output 54.14.1 An Observation with Time=3 in Beetles Data Set

Obs	time	conc	freq	sex
17	3	0.2	10	1

Output 54.14.2 Corresponding Beetle-Day Observations in Days

Obs	time	conc	freq	sex	day	y
25	3	0.2	10	1	1	0
26	3	0.2	10	1	2	0
27	3	0.2	10	1	3	1

The following statements invoke PROC LOGISTIC to fit a complementary log-log model for binary data with the response variable Y and the explanatory variables day, sex, and Variableconc. Specifying the EVENT= option ensures that the event (y=1) probability is modeled. The GLM coding in the CLASS statement creates an indicator column in the design matrix for each level of day. The coefficients of the indicator effects for day can be used to estimate the baseline survival function. The NOINT option is specified to prevent any redundancy in estimating the coefficients of day. The Newton-Raphson algorithm is used for the maximum likelihood estimation of the parameters.

```
proc logistic data=Days outest=est1;
   class day / param=glm;
   model y(event='1')= day sex conc
         / noint link=cloglog technique=newton;
   freq freq;
run;
```

Results of the model fit are given in Output 54.14.3. Both sex and conc are statistically significant for the survival of beetles sprayed by the insecticide. Female beetles are more resilient to the chemical than male beetles, and increased concentration of the insecticide increases its effectiveness.

Output 54.14.3 Parameter Estimates for the Grouped Proportional Hazards Model

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
day	1	1	−3.9314	0.2934	179.5602	<.0001
day	2	1	−2.8751	0.2412	142.0596	<.0001
day	3	1	−2.3985	0.2299	108.8833	<.0001
day	4	1	−1.9953	0.2239	79.3960	<.0001
day	5	1	−2.4920	0.2515	98.1470	<.0001
day	6	1	−3.1060	0.3037	104.5799	<.0001
day	7	1	−3.9704	0.4230	88.1107	<.0001
day	8	1	−3.7917	0.4007	89.5233	<.0001
day	9	1	−5.1540	0.7316	49.6329	<.0001
day	10	1	−5.1350	0.7315	49.2805	<.0001
day	11	1	−5.1131	0.7313	48.8834	<.0001
day	12	1	−5.1029	0.7313	48.6920	<.0001
day	13	1	−5.0951	0.7313	48.5467	<.0001
sex		1	−0.5651	0.1141	24.5477	<.0001
conc		1	3.0918	0.2288	182.5665	<.0001

The coefficients of parameters for the day variable are the maximum likelihood estimates of $\alpha_1, \ldots, \alpha_{13}$, respectively. The baseline survivor function $S_0(t)$ is estimated by

$$\hat{S}_0(t) = \widehat{\Pr}(T > t) = \prod_{j \le t} \exp(-\exp(\widehat{\alpha}_j))$$

and the survivor function for a given covariate pattern (sex=x_1 and conc=x_2) is estimated by

$$\hat{S}(t) = [\hat{S}_0(t)]^{\exp(-0.5651x_1 + 3.0918x_2)}$$

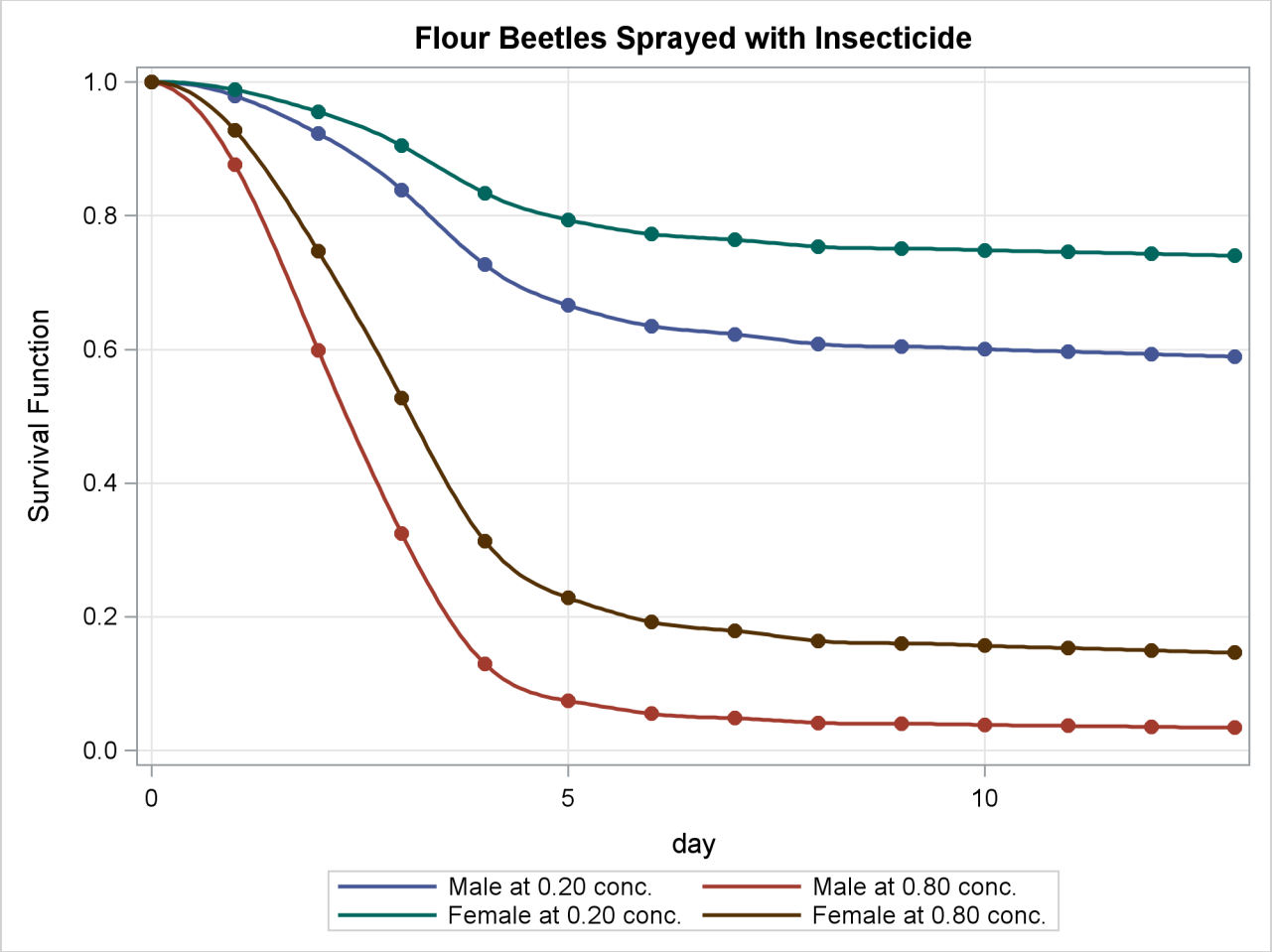
The following statements compute the survival curves for male and female flour beetles exposed to the insecticide in concentrations of 0.20 mg/cm^2 and 0.80 mg/cm^2:

```
data one (keep=day survival element s_m20 s_f20 s_m80 s_f80);
   array dd day1-day13;
   array sc[4] m20 f20 m80 f80;
   array s_sc[4] s_m20 s_f20 s_m80 s_f80 (1 1 1 1);
   set est1;
   m20= exp(sex + .20 * conc);
   f20= exp(2 * sex + .20 * conc);
   m80= exp(sex + .80 * conc);
   f80= exp(2 * sex + .80 * conc);
   survival=1;
   day=0;
   output;
   do over dd;
      element= exp(-exp(dd));
      survival= survival * element;
      do i=1 to 4;
         s_sc[i] = survival ** sc[i];
      end;
      day + 1;
      output;
   end;
run;
```

Instead of plotting the curves as step functions, the following statements use the PBSPLINE statement in the SGPLOT procedure to smooth the curves with a penalized B-spline. See Chapter 97, "The TRANSREG Procedure," for details about the implementation of the penalized B-spline method. The SAS autocall macro %MODSTYLE is specified to change the marker symbols for the plot. For more information about the %MODSTYLE macro, see the section "Style Template Modification Macro" on page 667 in Chapter 21, "Statistical Graphics Using ODS." The smoothed survival curves are displayed in Output 54.14.4.

```
%modstyle(name=LogiStyle,parent=htmlblue,markers=circlefilled);
ods listing style=LogiStyle;
proc sgplot data=one;
   title 'Flour Beetles Sprayed with Insecticide';
   xaxis grid integer;
   yaxis grid label='Survival Function';
   pbspline y=s_m20 x=day /
      legendlabel = "Male at 0.20 conc." name="pred1";
   pbspline y=s_m80 x=day /
      legendlabel = "Male at 0.80 conc." name="pred2";
   pbspline y=s_f20 x=day /
      legendlabel = "Female at 0.20 conc." name="pred3";
   pbspline y=s_f80 x=day /
      legendlabel = "Female at 0.80 conc." name="pred4";
   discretelegend "pred1" "pred2" "pred3" "pred4" / across=2;
run;
```

Output 54.14.4 Predicted Survival at Insecticide Concentrations of 0.20 and 0.80 mg/cm^2



The probability of survival is displayed on the vertical axis. Notice that most of the insecticide effect occurs by day 6 for both the high and low concentrations.

Example 54.15: Scoring Data Sets

This example first illustrates the syntax used for scoring data sets, then uses a previously scored data set to score a new data set. A generalized logit model is fit to the remote-sensing data set used in the section "Example 33.4: Linear Discriminant Analysis of Remote-Sensing Data on Crops" on page 2146 of Chapter 33, "The DISCRIM Procedure," to illustrate discrimination and classification methods. In the following DATA step, the response variable is Crop and the prognostic factors are x1 through x4:

```
data Crops;
   length Crop $ 10;
   infile datalines truncover;
   input Crop $ @@;
   do i=1 to 3;
      input x1-x4 @@;
      if (x1 ^= .) then output;
```

```
   end;
   input;
   datalines;
Corn       16 27 31 33  15 23 30 30  16 27 27 26
Corn       18 20 25 23  15 15 31 32  15 32 32 15
Corn       12 15 16 73
Soybeans   20 23 23 25  24 24 25 32  21 25 23 24
Soybeans   27 45 24 12  12 13 15 42  22 32 31 43
Cotton     31 32 33 34  29 24 26 28  34 32 28 45
Cotton     26 25 23 24  53 48 75 26  34 35 25 78
Sugarbeets 22 23 25 42  25 25 24 26  34 25 16 52
Sugarbeets 54 23 21 54  25 43 32 15  26 54  2 54
Clover     12 45 32 54  24 58 25 34  87 54 61 21
Clover     51 31 31 16  96 48 54 62  31 31 11 11
Clover     56 13 13 71  32 13 27 32  36 26 54 32
Clover     53 08 06 54  32 32 62 16
;
```

In the following statements, you specify a SCORE statement to use the fitted model to score the Crops data. The data together with the predicted values are saved in the data set Score1. The output from the EFFECTPLOT statement is discussed at the end of this section.

```
ods graphics on;
proc logistic data=Crops;
   model Crop=x1-x4 / link=glogit;
   score out=Score1;
   effectplot slicefit(x=x3);
run;
ods graphics off;
```

In the following statements, the model is fit again, and the data and the predicted values are saved into the data set Score2. The OUTMODEL= option saves the fitted model information in the permanent SAS data set sasuser.CropModel, and the STORE statement saves the fitted model information into the SAS data set CropModel2. Both the OUTMODEL= option and the STORE statement are specified to illustrate their use; you would usually specify only one of these model-storing methods.

```
proc logistic data=Crops outmodel=sasuser.CropModel;
   model Crop=x1-x4 / link=glogit;
   score data=Crops out=Score2;
   store CropModel2;
run;
```

To score data without refitting the model, specify the INMODEL= option to identify a previously saved SAS data set of model information. In the following statements, the model is read from the sasuser.CropModel data set, and the data and the predicted values are saved in the data set Score3. Note that the data set being scored does not have to include the response variable.

```
proc logistic inmodel=sasuser.CropModel;
   score data=Crops out=Score3;
run;
```

Another method available to score the data without refitting the model is to invoke the PLM procedure. In the following statements, the stored model is named in the SOURCE= option. The PREDICTED= option computes the linear predictors, and the ILINK option transforms the linear predictors to the probability

scale. The SCORE statement scores the Crops data set, and the predicted probabilities are saved in the data set ScorePLM. See Chapter 69, "The PLM Procedure," for more information.

```
proc plm source=CropModel2;
   score data=Crops out=ScorePLM predicted=p / ilink;
run;
```

For each observation in the Crops data set, the ScorePLM data set contains 5 observations—one for each level of the response variable. The following statements transform this data set into a form that is similar to the other scored data sets in this example:

```
proc transpose data=ScorePLM out=Score4 prefix=P_ let;
   id _LEVEL_;
   var p;
   by x1-x4  notsorted;
run;
data Score4(drop=_NAME_ _LABEL_);
   merge Score4 Crops(keep=Crop x1-x4);
   F_Crop=Crop;
run;
proc summary data=ScorePLM nway;
   by x1-x4 notsorted;
   var p;
   output out=into maxid(p(_LEVEL_))=I_Crop;
run;
data Score4;
   merge Score4 into(keep=I_Crop);
run;
```

To set prior probabilities on the responses, specify the PRIOR= option to identify a SAS data set containing the response levels and their priors. In the following statements, the Prior data set contains the values of the response variable (because this example uses single-trial MODEL statement syntax) and a _PRIOR_ variable containing values proportional to the default priors. The data and the predicted values are saved in the data set Score5.

```
data Prior;
   length Crop $10.;
   input Crop _PRIOR_;
   datalines;
Clover     11
Corn        7
Cotton      6
Soybeans    6
Sugarbeets  6
;

proc logistic inmodel=sasuser.CropModel;
   score data=Crops prior=prior out=Score5 fitstat;
run;
```

The "Fit Statistics for SCORE Data" table displayed in Output 54.15.1 shows that 47.22% of the observations are misclassified.

Output 54.15.1 Fit Statistics for Data Set Prior

Fit Statistics for SCORE Data

Data Set	Total Frequency	Log Likelihood	Error Rate	AIC	AICC	BIC
WORK.CROPS	36	-32.2247	0.4722	104.4493	160.4493	136.1197

Data Set	SC	R-Square	Max-Rescaled R-Square	AUC	Brier Score
WORK.CROPS	136.1197	0.744081	0.777285	.	0.492712

The data sets Score1, Score2, Score3, Score4, and Score5 are identical. The following statements display the scoring results in Output 54.15.2:

```
proc freq data=Score1;
   table F_Crop*I_Crop / nocol nocum nopercent;
run;
```

Output 54.15.2 Classification of Data Used for Scoring

Table of F_Crop by I_Crop

F_Crop(From: Crop) I_Crop(Into: Crop)

Frequency / Row Pct	Clover	Corn	Cotton	Soybeans	Sugarbeets	Total
Clover	6	0	2	2	1	11
	54.55	0.00	18.18	18.18	9.09	
Corn	0	7	0	0	0	7
	0.00	100.00	0.00	0.00	0.00	
Cotton	4	0	1	1	0	6
	66.67	0.00	16.67	16.67	0.00	
Soybeans	1	1	1	3	0	6
	16.67	16.67	16.67	50.00	0.00	
Sugarbeets	2	0	0	2	2	6
	33.33	0.00	0.00	33.33	33.33	
Total	13	8	4	8	3	36

The following statements use the previously fitted and saved model in the sasuser.CropModel data set to score the observations in a new data set, Test. The results of scoring the test data are saved in the ScoredTest data set and displayed in Output 54.15.3.

```
data Test;
   input Crop $ 1-10 x1-x4;
   datalines;
Corn       16 27 31 33
Soybeans   21 25 23 24
Cotton     29 24 26 28
Sugarbeets 54 23 21 54
Clover     32 32 62 16
;

proc logistic noprint inmodel=sasuser.CropModel;
   score data=Test out=ScoredTest;
run;

proc print data=ScoredTest label noobs;
   var F_Crop I_Crop P_Clover P_Corn P_Cotton P_Soybeans P_Sugarbeets;
run;
```

Output 54.15.3 Classification of Test Data

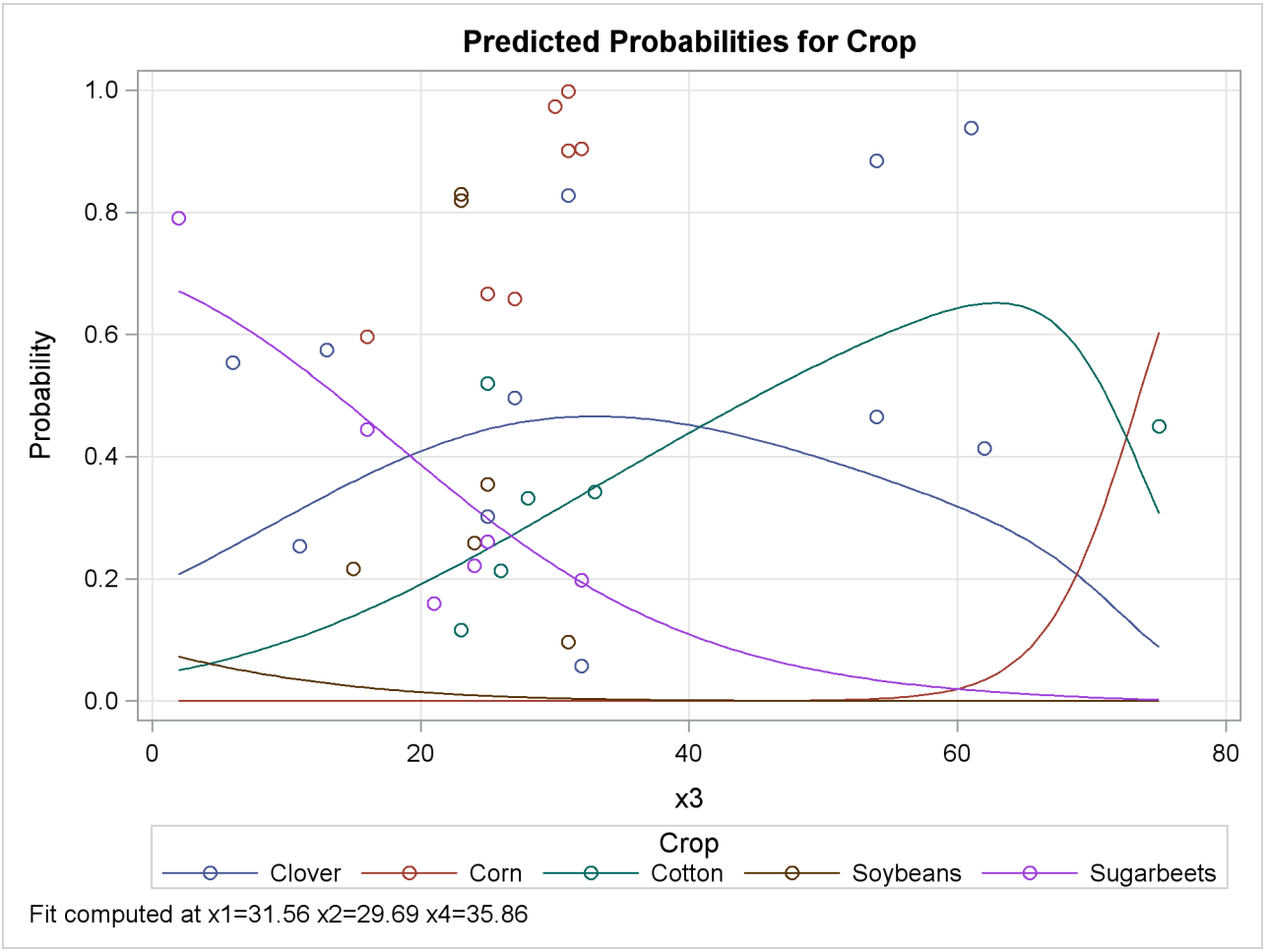
From: Crop	Into: Crop	Predicted Probability: Crop=Clover	Predicted Probability: Crop=Corn
Corn	Corn	0.00342	0.90067
Soybeans	Soybeans	0.04801	0.03157
Cotton	Clover	0.43180	0.00015
Sugarbeets	Clover	0.66681	0.00000
Clover	Cotton	0.41301	0.13386

Predicted Probability: Crop=Cotton	Predicted Probability: Crop=Soybeans	Predicted Probability: Crop=Sugarbeets
0.00500	0.08675	0.00416
0.02865	0.82933	0.06243
0.21267	0.07623	0.27914
0.17364	0.00000	0.15955
0.43649	0.00033	0.01631

The EFFECTPLOT statement that is specified in the first PROC LOGISTIC invocation produces a plot of the model-predicted probabilities versus X3 while holding the other three covariates at their means (Output 54.15.4). This plot shows how the value of X3 affects the probabilities of the various crops when the other prognostic factors are fixed at their means. If you are interested in the effect of X3 when the other covariates are fixed at a certain level—say, 10—specify the following EFFECTPLOT statement.

```
effectplot slicefit(x=x3) / at(x1=10 x2=10 x4=10)
```

Output 54.15.4 Model-Predicted Probabilities



Example 54.16: Using the LSMEANS Statement

Recall the main-effects model fit to the Neuralgia data set in Example 54.2. The Treatment*Sex interaction, which was previously shown to be nonsignificant, is added back into the model for this discussion.

In the following statements, the ODDSRATIO statement is specified to produce odds ratios of pairwise differences of the Treatment parameters in the presence of the Sex interaction. The LSMEANS statement is specified with several options: the E option displays the coefficients that are used to compute the LS-means for each Treatment level, the DIFF option takes all pairwise differences of the LS-means for the levels of the Treatment variable, the ODDSRATIO option computes odds ratios of these differences, the CL option produces confidence intervals for the differences and odds ratios, and the ADJUST=BON option performs a very conservative adjustment of the *p*-values and confidence intervals.

```
proc logistic data=Neuralgia;
   class Treatment Sex / param=glm;
   model Pain= Treatment|Sex Age;
   oddsratio Treatment;
   lsmeans Treatment / e diff oddsratio cl adjust=bon;
run;
```

The results from the ODDSRATIO statement are displayed in Output 54.16.1. All pairwise differences of levels of the Treatment effect are compared. However, because of the interaction between the Treatment and Sex variables, each difference is computed at each of the two levels of the Sex variable. These results show that the difference between Treatment levels A and B is insignificant for both genders.

To compute these odds ratios, you must first construct a linear combination of the parameters, $\boldsymbol{l}'\boldsymbol{\beta}$, for each level that is compared with all other levels fixed at some value. For example, to compare Treatment=A with B for Sex=F, you fix the Age variable at its mean, 70.05, and construct the following $\boldsymbol{l}$ vectors:

		Treatment			Sex		Treatment*Sex						
	Intercept	A	B	P	F	M	AF	AM	BF	BM	PF	PM	Age
$\boldsymbol{l}'_A$	1	1	0	0	1	0	1	0	0	0	0	0	70.05
$\boldsymbol{l}'_B$	1	0	1	0	1	0	0	0	1	0	0	0	70.05
$\boldsymbol{l}'_A - \boldsymbol{l}'_B$	0	1	–1	0	0	0	1	0	–1	0	0	0	0

Then the odds ratio for Treatment A versus B at Sex=F is computed as $\exp((\boldsymbol{l}'_A - \boldsymbol{l}'_B)\boldsymbol{\beta})$. Different $\boldsymbol{l}$ vectors must be similarly constructed when Sex=M because the resulting odds ratio will be different due to the interaction.

Output 54.16.1 Odds Ratios from the ODDSRATIO Statement

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
Treatment A vs B at Sex=F	0.398	0.016	9.722
Treatment A vs P at Sex=F	16.892	1.269	224.838
Treatment B vs P at Sex=F	42.492	2.276	793.254
Treatment A vs B at Sex=M	0.663	0.078	5.623
Treatment A vs P at Sex=M	34.766	1.807	668.724
Treatment B vs P at Sex=M	52.458	2.258	>999.999

The results from the LSMEANS statement are displayed in Output 54.16.2 through Output 54.16.4.

The LS-means are computed by constructing each of the $\boldsymbol{l}$ coefficient vectors shown in Output 54.16.2, and then computing $\boldsymbol{l}'\boldsymbol{\beta}$. The LS-means are not estimates of the event probabilities; they are estimates of the linear predictors on the logit scale. In order to obtain event probabilities, you need to apply the inverse-link transformation by specifying the ILINK option in the LSMEANS statement. Notice in Output 54.16.2 that the Sex rows do not indicate either Sex=F or Sex=M. Instead, the LS-means are computed at an average of these two levels, so only one result needs to be reported. For more information about the construction of LS-means, see the section "Construction of Least Squares Means" on page 3363 of Chapter 42, "The GLM Procedure."

Output 54.16.2 Treatment LS-Means Coefficients

Coefficients for Treatment Least Squares Means

Parameter	Treatment	Sex	Row1	Row2	Row3
Intercept: Pain=No			1	1	1
Treatment A	A		1		
Treatment B	B			1	
Treatment P	P				1
Sex F		F	0.5	0.5	0.5
Sex M		M	0.5	0.5	0.5
Treatment A * Sex F	A	F	0.5		
Treatment A * Sex M	A	M	0.5		
Treatment B * Sex F	B	F		0.5	
Treatment B * Sex M	B	M		0.5	
Treatment P * Sex F	P	F			0.5
Treatment P * Sex M	P	M			0.5
Age			70.05	70.05	70.05

The Treatment LS-means shown in Output 54.16.3 are all significantly nonzero at the 0.05 level. These LS-means are *predicted population margins* of the logits; that is, they estimate the marginal means over a balanced population, and they are effectively the within-Treatment means appropriately adjusted for the other effects in the model. The LS-means are not event probabilities; in order to obtain event probabilities, you need to apply the inverse-link transformation by specifying the ILINK option in the LSMEANS statement. For more information about LS-means, see the section "LSMEANS Statement" on page 453 of Chapter 19, "Shared Concepts and Topics."

Output 54.16.3 Treatment LS-Means

Treatment Least Squares Means

Treatment	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper
A	1.3195	0.6664	1.98	0.0477	0.05	0.01331	2.6257
B	1.9864	0.7874	2.52	0.0116	0.05	0.4431	3.5297
P	−1.8682	0.7620	−2.45	0.0142	0.05	−3.3618	−0.3747

Pairwise differences between the Treatment LS-means, requested with the DIFF option, are displayed in Output 54.16.4. The LS-mean for the level that is displayed in the _Treatment column is subtracted from the LS-mean for the level in the Treatment column, so the first row displays the LS-mean for Treatment level A minus the LS-mean for Treatment level B. The Pr > |z| column indicates that the A and B levels are not significantly different; however, both of these levels are different from level P. If the inverse-link transformation is specified with the ILINK option, then these differences do not transform back to differences in probabilities.

There are two odds ratios for Treatment level A versus B in Output 54.16.1; these are constructed at each level of the interacting covariate Sex. In contrast, there is only one LS-means odds ratio for Treatment level A versus B in Output 54.16.4. This odds ratio is computed at an average of the interacting effects by creating the $\boldsymbol{l}$ vectors shown in Output 54.16.2 (the Row1 column corresponds to l_A and the Row2 column corresponds to l_B) and computing $\exp(\boldsymbol{l}_A'\boldsymbol{\beta} - \boldsymbol{l}_B'\boldsymbol{\beta})$.

Since multiple tests are performed, you can protect yourself from falsely significant results by adjusting your *p*-values for multiplicity. The ADJUST=BON option performs the very conservative Bonferroni adjustment, and adds the columns labeled with 'Adj' to Output 54.16.4. Comparing the Pr > |z| column to the Adj P column, you can see that the *p*-values are adjusted upwards; in this case, there is no change in your conclusions. The confidence intervals are also adjusted for multiplicity—all adjusted intervals are wider than the unadjusted intervals, but again your conclusions in this example are unchanged.

Output 54.16.4 Differences and Odds Ratios for the Treatment LS-Means

Differences of Treatment Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Treatment	_Treatment	Estimate	Standard Error	z Value	Pr > |z|	Adj P	Alpha
A	B	-0.6669	1.0026	-0.67	0.5059	1.0000	0.05
A	P	3.1877	1.0376	3.07	0.0021	0.0064	0.05
B	P	3.8547	1.2126	3.18	0.0015	0.0044	0.05

Differences of Treatment Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Treatment	_Treatment	Lower	Upper	Adj Lower	Adj Upper	Odds Ratio
A	B	-2.6321	1.2982	-3.0672	1.7334	0.513
A	P	1.1541	5.2214	0.7037	5.6717	24.234
B	P	1.4780	6.2313	0.9517	6.7576	47.213

Differences of Treatment Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Treatment	_Treatment	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio	Adj Lower Odds Ratio	Adj Upper Odds Ratio
A	B	0.072	3.663	0.047	5.660
A	P	3.171	185.195	2.021	290.542
B	P	4.384	508.441	2.590	860.612

If you want to jointly test whether the active treatments are different from the placebo, you can specify a custom hypothesis test with the LSMESTIMATE statement. In the following statements, the LS-means for the two treatments are contrasted against the LS-mean of the placebo, and the JOINT option performs a joint test that the two treatments are not different from placebo.

```
proc logistic data=Neuralgia;
   class Treatment Sex / param=glm;
   model Pain= Treatment|Sex Age;
   lsmestimate treatment 1 0 -1, 0 1 -1 / joint;
run;
```

Output 54.16.5 displays the results from the LSMESTIMATE statement. The "Least Squares Means Es-

timate" table displays the differences of the two active treatments against the placebo, and the results are identical to the second and third rows of Output 54.16.3. The "Chi-Square Test for Least Squares Means Estimates" table displays the joint test. In all of these tests, you reject the null hypothesis that the treatment has the same effect as the placebo.

Output 54.16.5 Custom LS-Mean Tests

Least Squares Means Estimates

Effect	Label	Estimate	Standard Error	z Value	Pr > |z|
Treatment	Row 1	3.1877	1.0376	3.07	0.0021
Treatment	Row 2	3.8547	1.2126	3.18	0.0015

Chi-Square Test for Least Squares Means Estimates

Effect	Num DF	Chi-Square	Pr > ChiSq
Treatment	2	12.13	0.0023

If you want to work with LS-means but you prefer to compute the Treatment odds ratios within the Sex levels in the same fashion as the ODDSRATIO statement does, you can specify the SLICE statement. In the following statements, you specify the same options in the SLICE statement as you do in the LSMEANS statement, except that you also specify the SLICEBY= option to perform an LS-means analysis partitioned into sets that are defined by the Sex variable:

```
proc logistic data=Neuralgia;
   class Treatment Sex / param=glm;
   model Pain= Treatment|Sex Age;
   slice Treatment*Sex / sliceby=Sex diff oddsratio cl adjust=bon;
run;
```

The results for Sex=F are displayed in Output 54.16.6 and Output 54.16.7. The joint test in Output 54.16.6 tests the equality of the LS-means of the levels of Treatment for Sex=F, and rejects equality at level 0.05. In Output 54.16.7, the odds ratios and confidence intervals match those reported for Sex=F in Output 54.16.1, and multiplicity adjustments are performed.

Output 54.16.6 Joint Test of Treatment Equality for Females

Chi-Square Test for Treatment*Sex Least Squares Means Slice

Slice	Num DF	Chi-Square	Pr > ChiSq
Sex F	2	8.22	0.0164

Output 54.16.7 Differences of the Treatment LS-Means for Females

Simple Differences of Treatment*Sex Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Estimate	Standard Error	z Value	Pr > |z|	Adj P
Sex F	A	B	-0.9224	1.6311	-0.57	0.5717	1.0000
Sex F	A	P	2.8269	1.3207	2.14	0.0323	0.0970
Sex F	B	P	3.7493	1.4933	2.51	0.0120	0.0361

Simple Differences of Treatment*Sex Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Alpha	Lower	Upper	Adj Lower	Adj Upper
Sex F	A	B	0.05	-4.1193	2.2744	-4.8272	2.9824
Sex F	A	P	0.05	0.2384	5.4154	-0.3348	5.9886
Sex F	B	P	0.05	0.8225	6.6761	0.1744	7.3243

Simple Differences of Treatment*Sex Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
Sex F	A	B	0.398	0.016	9.722
Sex F	A	P	16.892	1.269	224.838
Sex F	B	P	42.492	2.276	793.254

Simple Differences of Treatment*Sex
Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Adj Lower Odds Ratio	Adj Upper Odds Ratio
Sex F	A	B	0.008	19.734
Sex F	A	P	0.715	398.848
Sex F	B	P	1.190	>999.999

Similarly, the results for Sex=M are shown in Output 54.16.8 and Output 54.16.9.

Output 54.16.8 Joint Test of Treatment Equality for Males

Chi-Square Test for Treatment*Sex
Least Squares Means Slice

Slice	Num DF	Chi-Square	Pr > ChiSq
Sex M	2	6.64	0.0361

Output 54.16.9 Differences of the Treatment LS-Means for Males

Simple Differences of Treatment*Sex Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Estimate	Standard Error	z Value	Pr > |z|	Adj P
Sex M	A	B	-0.4114	1.0910	-0.38	0.7061	1.0000
Sex M	A	P	3.5486	1.5086	2.35	0.0187	0.0560
Sex M	B	P	3.9600	1.6049	2.47	0.0136	0.0408

Simple Differences of Treatment*Sex Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Alpha	Lower	Upper	Adj Lower	Adj Upper
Sex M	A	B	0.05	-2.5496	1.7268	-3.0231	2.2003
Sex M	A	P	0.05	0.5919	6.5054	-0.06286	7.1601
Sex M	B	P	0.05	0.8145	7.1055	0.1180	7.8021

Simple Differences of Treatment*Sex Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Odds Ratio	Lower Confidence Limit for Odds Ratio	Upper Confidence Limit for Odds Ratio
Sex M	A	B	0.663	0.078	5.623
Sex M	A	P	34.766	1.807	668.724
Sex M	B	P	52.458	2.258	>999.999

Simple Differences of Treatment*Sex
Least Squares Means
Adjustment for Multiple Comparisons: Bonferroni

Slice	Treatment	_Treatment	Adj Lower Odds Ratio	Adj Upper Odds Ratio
Sex M	A	B	0.049	9.028
Sex M	A	P	0.939	>999.999
Sex M	B	P	1.125	>999.999

Example 54.17: Partial Proportional Odds Model

Cameron and Trivedi (1998, p. 68) studied the number of doctor visits from the Australian Health Survey 1977–78. The data set contains a dependent variable, dvisits, which contains the number of doctor visits in the past two weeks (0, 1, or 2, where 2 represents two or more visits) and the following explanatory variables: sex, which indicates whether the patient is female; age, which contains the patient's age in years divided by 100; income, which contains the patient's annual income (in units of $10,000); levyplus, which indicates whether the patient has private health insurance; freepoor, which indicates that the patient has free government health insurance due to low income; freerepa, which indicates that the patient has free government health insurance for other reasons; illness, which contains the number of illnesses in the past two weeks; actdays, which contains the number of days the illness caused reduced activity; hscore, which is a questionnaire score; chcond1, which indicates a chronic condition that does not limit activity; and chcond2, which indicates a chronic condition that limits activity.

```
data docvisit;
   input sex age agesq income levyplus freepoor freerepa
         illness actdays hscore chcond1 chcond2 dvisits;
   if ( dvisits > 2) then dvisits = 2;
datalines;
1 0.19 0.0361 0.55  1  0  0  1  4  1  0  0  1
1 0.19 0.0361 0.45  1  0  0  1  2  1  0  0  1
0 0.19 0.0361 0.90  0  0  0  3  0  0  0  0  1

   ... more lines ...

1 0.37 0.1369 0.25  0  0  1  1  0  1  0  0  0
1 0.52 0.2704 0.65  0  0  0  0  0  0  0  0  0
0 0.72 0.5184 0.25  0  0  1  0  0  0  0  0  0
;
```

Because the response variable dvisits has three levels, the proportional odds model constructs two response functions. There is an intercept parameter for each of the two response functions, $\alpha_1 < \alpha_2$, and common slope parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{12})$ across the functions. The model can be written as

$$\text{logit}(\Pr(Y \leq i \mid \mathbf{x})) = \alpha_i + \boldsymbol{\beta}'\mathbf{x}, \quad i = 1, 2$$

The following statements fit a proportional odds model to this data:

```
proc logistic data=docvisit;
   model dvisits = sex age agesq income levyplus
                   freepoor freerepa illness actdays hscore
                   chcond1 chcond2;
run;
```

Selected results are displayed in Output 54.17.1.

Output 54.17.1 Test of Proportional Odds Assumption

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
27.4256	12	0.0067

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	734.2971	12	<.0001
Score	811.8964	12	<.0001
Wald	690.7156	12	<.0001

The test of the proportional odds assumption in Output 54.17.1 rejects the null hypothesis that all the slopes are equal across the two response functions. This test is very anticonservative; that is, it tends to reject the null hypothesis even when the proportional odds assumption is reasonable.

The proportional odds assumption for ordinal response models can be relaxed by specifying the UNEQUALSLOPES option in the MODEL statement. A fully nonproportional odds model has different slope parameters $\boldsymbol{\beta}_i = (\beta_{1,i}, \ldots, \beta_{12,i})$ for every logit i:

$$\text{logit}(\Pr(Y \leq i \mid \mathbf{x})) = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}, \quad i = 1, 2$$

The nonproportional odds model is fit with the following statements. The TEST statements test the proportional odds assumption for each of the covariates in the model.

```
proc logistic data=docvisit;
   model dvisits = sex age agesq income levyplus
                   freepoor freerepa illness actdays hscore
                   chcond1 chcond2 / unequalslopes;
   sex:      test sex_0     =sex_1;
   age:      test age_0     =age_1;
   agesq:    test agesq_0   =agesq_1;
   income:   test income_0  =income_1;
   levyplus: test levyplus_0=levyplus_1;
   freepoor: test freepoor_0=freepoor_1;
   freerepa: test freerepa_0=freerepa_1;
   illness:  test illness_0 =illness_1;
   actdays:  test actdays_0 =actdays_1;
   hscore:   test hscore_0  =hscore_1;
   chcond1:  test chcond1_0 =chcond1_1;
   chcond2:  test chcond2_0 =chcond2_1;
run;
```

Selected results from fitting the nonproportional odds model to the data are displayed in Output 54.17.2.

Output 54.17.2 Results for Nonproportional Odds Model

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	761.4797	24	<.0001
Score	957.6793	24	<.0001
Wald	688.2306	24	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	dvisits	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	0	1	2.3238	0.2754	71.2018	<.0001
Intercept	1	1	4.2862	0.4890	76.8368	<.0001
sex	0	1	−0.2637	0.0818	10.3909	0.0013
sex	1	1	−0.1232	0.1451	0.7210	0.3958
age	0	1	1.7489	1.5115	1.3389	0.2472
age	1	1	−2.0974	2.6003	0.6506	0.4199
agesq	0	1	−2.4718	1.6636	2.2076	0.1373
agesq	1	1	2.6883	2.8398	0.8961	0.3438
income	0	1	−0.00857	0.1266	0.0046	0.9460
income	1	1	0.6464	0.2375	7.4075	0.0065
levyplus	0	1	−0.2658	0.0997	7.0999	0.0077
levyplus	1	1	−0.2869	0.1820	2.4848	0.1150
freepoor	0	1	0.6773	0.2601	6.7811	0.0092
freepoor	1	1	0.9020	0.4911	3.3730	0.0663
freerepa	0	1	−0.4044	0.1382	8.5637	0.0034
freerepa	1	1	−0.0958	0.2361	0.1648	0.6848
illness	0	1	−0.2645	0.0287	84.6792	<.0001
illness	1	1	−0.3083	0.0499	38.1652	<.0001
actdays	0	1	−0.1521	0.0116	172.2764	<.0001
actdays	1	1	−0.1863	0.0134	193.7700	<.0001
hscore	0	1	−0.0620	0.0172	12.9996	0.0003
hscore	1	1	−0.0568	0.0252	5.0940	0.0240
chcond1	0	1	−0.1140	0.0909	1.5721	0.2099
chcond1	1	1	−0.2478	0.1743	2.0201	0.1552
chcond2	0	1	−0.2660	0.1255	4.4918	0.0341
chcond2	1	1	−0.3146	0.2116	2.2106	0.1371

Output 54.17.2 *continued*

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
sex	1.0981	1	0.2947
age	2.5658	1	0.1092
agesq	3.8309	1	0.0503
income	8.8006	1	0.0030
levyplus	0.0162	1	0.8989
freepoor	0.2569	1	0.6122
freerepa	2.0099	1	0.1563
illness	0.8630	1	0.3529
actdays	6.9407	1	0.0084
hscore	0.0476	1	0.8273
chcond1	0.6906	1	0.4060
chcond2	0.0615	1	0.8042

The preceding nonproportional odds model fits $12 \times 2 = 24$ slope parameters, and the model seems to overfit the data. You can obtain a more parsimonious model by specifying a subset of the parameters to have nonproportional odds. The following statements allow the parameters for the variables in the "Linear Hypotheses Testing Results" table that have p-values less than 0.1 (actdays, agesq, and income) to vary across the response functions:

```
proc logistic data=docvisit;
   model dvisits= sex age agesq income levyplus freepoor
         freerepa illness actdays hscore chcond1 chcond2
   / unequalslopes=(actdays agesq income);
run;
```

Selected results from fitting this partial proportional odds model are displayed in Output 54.17.3.

Output 54.17.3 Results for Partial Proportional Odds Model

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	752.5512	15	<.0001
Score	947.3269	15	<.0001
Wald	683.4719	15	<.0001

Output 54.17.3 *continued*

Analysis of Maximum Likelihood Estimates

Parameter	dvisits	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	0	1	2.3882	0.2716	77.2988	<.0001
Intercept	1	1	3.7597	0.3138	143.5386	<.0001
sex		1	−0.2485	0.0807	9.4789	0.0021
age		1	1.3000	1.4864	0.7649	0.3818
agesq	0	1	−2.0110	1.6345	1.5139	0.2186
agesq	1	1	−0.8789	1.6512	0.2833	0.5945
income	0	1	0.0209	0.1261	0.0275	0.8683
income	1	1	0.4283	0.2221	3.7190	0.0538
levyplus		1	−0.2703	0.0989	7.4735	0.0063
freepoor		1	0.6936	0.2589	7.1785	0.0074
freerepa		1	−0.3648	0.1358	7.2155	0.0072
illness		1	−0.2707	0.0281	92.7123	<.0001
actdays	0	1	−0.1522	0.0115	173.5696	<.0001
actdays	1	1	−0.1868	0.0129	209.7134	<.0001
hscore		1	−0.0609	0.0166	13.5137	0.0002
chcond1		1	−0.1200	0.0901	1.7756	0.1827
chcond2		1	−0.2628	0.1227	4.5849	0.0323

The partial proportional odds model can be written in the same form as the nonproportional odds model by letting $\mathbf{x} = (x_1, \ldots, x_q, x_{q+1}, \ldots, x_{12})$ and $\boldsymbol{\beta}_i = (\beta_1, \ldots, \beta_q, \beta_{q+1,i}, \ldots, \beta_{12,i})$, so the first q parameters have proportional odds and the remaining parameters do not. The last 12–q parameters can be rewritten to have a common slope: $\beta_{q+j} + \gamma_{q+j,i}$, $j = 1, \ldots, 12-q$, where the new parameters $\boldsymbol{\gamma}_i$ contain the increments from the common slopes. The model in this form makes it obvious that the proportional odds model is a submodel of the partial proportional odds models, and both of these are submodels of the nonproportional odds model. This means that you can use likelihood ratio tests to compare models.

You can use the following statements to compute the likelihood ratio tests from the Likelihood Ratio row of the "Testing Global Null hypothesis: BETA=0" tables in the preceding outputs:

```
data a;
   label p='Pr>ChiSq';
   format p 8.6;
   input Test $10. ChiSq1 DF1 ChiSq2 DF2;
   ChiSq= ChiSq1-ChiSq2;
   DF= DF1-DF2;
   p=1-probchi(ChiSq,DF);
   keep Test Chisq DF p;
   datalines;
Non vs PO     761.4797 24    734.2971 12
PPO vs PO     752.5512 15    734.2971 12
Non vs PPO    761.4797 24    752.5512 15
;

proc print data=a label noobs;
   var Test ChiSq DF p;
run;
```

Output 54.17.4 Likelihood Ratio Tests

Test	ChiSq	DF	Pr>ChiSq
Non vs PO	27.1826	12	0.007273
PPO vs PO	18.2541	3	0.000390
Non vs PPO	8.9285	9	0.443900

Therefore, you reject the proportional odds model in favor of both the nonproportional odds model and the partial proportional odds model, and the partial proportional odds model fits as well as the nonproportional odds model. The likelihood ratio test of the nonproportional odds model versus the proportional odds model is very similar to the score test of the proportional odds assumption in Output 54.17.1 because of the large sample size (Stokes, Davis, and Koch 2000, p. 249).

NOTE: The proportional odds model has increasing intercepts, which ensures the increasing nature of the cumulative response functions. However, none of the parameters in the partial or nonproportional odds models are constrained. Because of this, sometimes during the optimization process a predicted individual probability can be negative; the optimization continues because it might recover from this situation. Sometimes your final model will predict negative individual probabilities for some of the observations; in this case a message is displayed, and you should check your data for outliers and possibly redefine your model. Other times the model fits your data well, but if you try to score new data you can get negative individual probabilities. This means the model is not appropriate for the data you are trying to score, a message is displayed, and the estimates are set to missing.

References

Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, New York: John Wiley & Sons.

Agresti, A. (1990), *Categorical Data Analysis*, New York: John Wiley & Sons.

Agresti, A. (1992), "A Survey of Exact Inference for Contingency Tables," *Statistical Science*, 7, 131–177.

Agresti, A. (2002), *Categorical Data Analysis*, Second Edition, New York: John Wiley & Sons.

Aitchison, J. and Silvey, S. (1957), "The Generalization of Probit Analysis to the Case of Multiple Responses," *Biometrika*, 44, 131–140.

Albert, A. and Anderson, J. A. (1984), "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika*, 71, 1–10.

Allison, P. D. (1982), "Discrete-Time Methods for the Analysis of Event Histories," in S. Leinhardt, ed., *Sociological Methods and Research*, volume 15, 61–98, San Francisco: Jossey-Bass.

Allison, P. D. (1999), *Logistic Regression Using the SAS System: Theory and Application*, Cary, NC: SAS Institute Inc.

Ashford, J. R. (1959), "An Approach to the Analysis of Data for Semi-quantal Responses in Biology Response," *Biometrics*, 15, 573–581.

Bartolucci, A. A. and Fraser, M. D. (1977), "Comparative Step-Up and Composite Test for Selecting Prognostic Indicator Associated with Survival," *Biometrical Journal*, 19, 437–448.

Breslow, N. E. (1982), "Covariance Adjustment of Relative-Risk Estimates in Matched Studies," *Biometrics*, 38, 661–672.

Breslow, N. E. and Day, N. E. (1980), *Statistical Methods in Cancer Research, Volume I: The Analysis of Case-Control Studies*, IARC Scientific Publications, No. 32, Lyon, France: International Agency for Research on Cancer.

Brier, G. W. (1950), "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review*, 78(1), 1–3.

Burnham, K. P. and Anderson, D. R. (1998), *Model Selection and Inference: A Practical Information-Theoretic Approach*, New York: Springer-Verlag.

Cameron, A. C. and Trivedi, P. K. (1998), *Regression Analysis of Count Data*, Cambridge: Cambridge University Press.

Collett, D. (2003), *Modelling Binary Data*, Second Edition, London: Chapman & Hall.

Cook, R. D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman & Hall.

Cox, D. R. (1970), *The Analysis of Binary Data*, New York: Chapman & Hall.

Cox, D. R. (1972), "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Series B*, 20, 187–220, with discussion.

Cox, D. R. and Snell, E. J. (1989), *The Analysis of Binary Data*, Second Edition, London: Chapman & Hall.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988), "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics*, 44, 837–845.

Draper, C. C., Voller, A., and Carpenter, R. G. (1972), "The Epidemiologic Interpretation of Serologic Data in Malaria," *American Journal of Tropical Medicine and Hygiene*, 21, 696–703.

Finney, D. J. (1947), "The Estimation from Individual Records of the Relationship between Dose and Quantal Response," *Biometrika*, 34, 320–334.

Firth, D. (1993), "Bias Reduction of Maximum Likelihood Estimates," *Biometrika*, 80, 27–38.

Fleiss, J. L. (1981), *Statistical Methods for Rates and Proportions*, Second Edition, New York: John Wiley & Sons.

Freeman, D. H., Jr. (1987), *Applied Categorical Data Analysis*, New York: Marcel Dekker.

Furnival, G. M. and Wilson, R. W. (1974), "Regression by Leaps and Bounds," *Technometrics*, 16, 499–511.

Gail, M. H., Lubin, J. H., and Rubinstein, L. V. (1981), "Likelihood Calculations for Matched Case-Control Studies and Survival Studies with Tied Death Times," *Biometrika*, 68, 703–707.

Hanley, J. A. and McNeil, B. J. (1982), "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve," *Radiology*, 143, 29–36.

Harrell, F. E. (1986), "The LOGIST Procedure," *SUGI Supplemental Library Guide, Version 5 Edition*.

Heinze, G. (1999), *The Application of Firth's Procedure to Cox and Logistic Regression*, Technical Report 10/1999, update in January 2001, Section of Clinical Biometrics, Department of Medical Computer Sciences, University of Vienna.

Heinze, G. (2006), "A Comparative Investigation of Methods for Logistic Regression with Separated or Nearly Separated Data," *Statistics in Medicine*, 25, 4216–4226.

Heinze, G. and Schemper, M. (2002), "A Solution to the Problem of Separation in Logistic Regression," *Statistics in Medicine*, 21, 2409–2419.

Hilbe, J. M. (2009), *Logistic Regression Models*, London: Chapman & Hall/CRC.

Hirji, K. F. (1992), "Computing Exact Distributions for Polytomous Response Data," *Journal of the American Statistical Association*, 87, 487–492.

Hirji, K. F., Mehta, C. R., and Patel, N. R. (1987), "Computing Distributions for Exact Logistic Regression," *Journal of the American Statistical Association*, 82, 1110–1117.

Hirji, K. F., Mehta, C. R., and Patel, N. R. (1988), "Exact Inference for Matched Case-Control Studies," *Biometrics*, 44, 803–814.

Hirji, K. F., Tsiatis, A. A., and Mehta, C. R. (1989), "Median Unbiased Estimation for Binary Data," *The American Statistician*, 43, 7–11.

Hosmer, D. W., Jr. and Lemeshow, S. (2000), *Applied Logistic Regression*, Second Edition, New York: John Wiley & Sons.

Howard, S. (1972), "Discussion on the Paper by Cox," in *Regression Models and Life Tables*, volume 34 of *Journal of the Royal Statistical Society, Series B*, 187–220, with discussion.

Hurvich, C. M. and Tsai, C. (1993), "A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection," *Journal of Time Series Analysis*.

Izrael, D., Battaglia, A. A., Hoaglin, D. C., and Battaglia, M. P. (2002), "Use of the ROC Curve and the Bootstrap in Comparing Weighted Logistic Regression Models," in *Proceedings of the Twenty-seventh Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc., available at www2.sas.com/proceedings/sugi27/p248-27.pdf.

Lachin, J. M. (2000), *Biostatistical Methods: The Assessment of Relative Risks*, New York: John Wiley & Sons.

Lamotte, L. R. (2002), personal communication, June 2002.

Lancaster, H. O. (1961), "Significance Tests in Discrete Distributions," *Journal of the American Statistical Association*, 56, 223–234.

Lawless, J. F. and Singhal, K. (1978), "Efficient Screening of Nonnormal Regression Models," *Biometrics*, 34, 318–327.

Lee, E. T. (1974), "A Computer Program for Linear Logistic Regression Analysis," *Computer Programs in Biomedicine*, 80–92.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, Second Edition, London: Chapman & Hall.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behaviour," in P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press.

Mehta, C. R., Patel, N., and Senchaudhuri, P. (1992), "Exact Stratified Linear Rank Tests for Ordered Categorical and Binary Data," *Journal of Computational and Graphical Statistics*, 1, 21–40.

Mehta, C. R., Patel, N., and Senchaudhuri, P. (2000), "Efficient Monte Carlo Methods for Conditional Logistic Regression," *Journal of the American Statistical Association*, 95, 99–108.

Mehta, C. R. and Patel, N. R. (1995), "Exact Logistic Regression: Theory and Examples," *Statistics in Medicine*, 14, 2143–2160.

Moolgavkar, S. H., Lustbader, E. D., and Venzon, D. J. (1985), "Assessing the Adequacy of the Logistic Regression Model for Matched Case-Control Studies," *Statistics in Medicine*, 4, 425–435.

Murphy, A. H. (1973), "A New Vector Partition of the Probability Score," *Journal of Applied Meterology*, 12, 595–600.

Naessens, J. M., Offord, K. P., Scott, W. F., and Daood, S. L. (1986), "The MCSTRAT Procedure," in *SUGI Supplemental Library User's Guide, Version 5 Edition*, 307–328, Cary, NC: SAS Institute Inc.

Nagelkerke, N. J. D. (1991), "A Note on a General Definition of the Coefficient of Determination," *Biometrika*, 78, 691–692.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society, Series A*, 135, 370–384.

Peterson, B. and Harrell, F. E., Jr. (1990), "Partial Proportional Odds Models for Ordinal Response Variables," *Journal of the Royal Statistical Society, Series B*, 39, 205–217.

Pregibon, D. (1981), "Logistic Regression Diagnostics," *Annals of Statistics*, 9, 705–724.

Pregibon, D. (1984), "Data Analytic Methods for Matched Case-Control Studies," *Biometrics*, 40, 639–651.

Prentice, P. L. and Gloeckler, L. A. (1978), "Regression Analysis of Grouped Survival Data with Applications to Breast Cancer Data," *Biometrics*, 34, 57–67.

Press, S. J. and Wilson, S. (1978), "Choosing between Logistic Regression and Discriminant Analysis," *Journal of the American Statistical Association*, 73, 699–705.

Santner, T. J. and Duffy, E. D. (1986), "A Note on A. Albert and J. A. Anderson's Conditions for the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika*, 73, 755–758.

SAS Institute Inc. (1995), *Logistic Regression Examples Using the SAS System*, Cary, NC: SAS Institute Inc.

Stokes, M. E., Davis, C. S., and Koch, G. G. (2000), *Categorical Data Analysis Using the SAS System*, Second Edition, Cary, NC: SAS Institute Inc.

Stokes, M. E., Davis, C. S., and Koch, G. G. (2012), *Categorical Data Analysis Using SAS*, Third edition Edition, Cary, NC: SAS Institute Inc.

Storer, B. E. and Crowley, J. (1985), "A Diagnostic for Cox Regression and General Conditional Likelihoods," *Journal of the American Statistical Association*, 80, 139–147.

Venzon, D. J. and Moolgavkar, S. H. (1988), "A Method for Computing Profile-Likelihood Based Confidence Intervals," *Applied Statistics*, 37, 87–94.

Vollset, S. E., Hirji, K. F., and Afifi, A. A. (1991), "Evaluation of Exact and Asymptotic Interval Estimators in Logistic Analysis of Matched Case-Control Studies," *Biometrics*, 47, 1311–1325.

Walker, S. H. and Duncan, D. B. (1967), "Estimation of the Probability of an Event as a Function of Several Independent Variables," *Biometrika*, 54, 167–179.

Williams, D. A. (1982), "Extra-binomial Variation in Logistic Linear Models," *Applied Statistics*, 31, 144–148.

Subject Index

Syntax Index

Your Turn

We welcome your feedback.

- If you have comments about this book, please send them to **yourturn@sas.com**. Include the full title and page numbers (if applicable).
- If you have comments about the software, please send them to **suggest@sas.com**.