Performing Exact Logistic Regression with the SAS System — Revised 2009

Robert E. Derr, SAS Institute Inc., Cary, NC

ABSTRACT

Exact logistic regression has become an important analytical technique, especially in the pharmaceutical industry, since the usual asymptotic methods for analyzing small, skewed, or sparse data sets are unreliable. Inference based on enumerating the exact distributions of sufficient statistics for parameters of interest in a logistic regression model, conditional on the remaining parameters, is computationally infeasible for many problems. Efficient algorithms for generating the required conditional distributions were introduced in Hirji, Mehta, and Patel (1987) and Mehta, Patel, and Senchaudhuri (1992, 2000), thus making these methods computationally available. This paper discusses the theory and methods for exact logistic regression and illustrates their application with the LOGISTIC procedure in SAS/STAT® 9.2 software.

INTRODUCTION

Many clinical trials deal with the comparison of populations of subjects with categorical responses. Historically, statistical inference for such studies involved large-sample approximations, and fitting logistic regression models to such data was performed through the unconditional likelihood function.

However, asymptotic methods might be inadequate when sample sizes are small or the data are sparse, skewed, or heavily tied. Exact conditional inference remains valid in such situations.

The LOGISTIC, GENMOD, GLIMMIX, PROBIT, and CATMOD procedures perform unconditional likelihood inference for logit models, and the LOGISTIC and PHREG procedures can perform asymptotic conditional likelihood inference for logit models. SAS users have requested the ability to perform exact tests for logistic regression modeling. Many exact statistical tests have already been added to the FREQ and NPAR1WAY procedures, and as of SAS 8.1, SAS/STAT software includes exact logistic regression for binary (dichotomous) response variables in the LOGISTIC procedure. Exact methods for generalized logit (GLOGIT) models have been available in the LOGISTIC procedure since SAS 9.

The "METHODOLOGY" section in this paper presents the logistic regression model and the different likelihoods, and then explains how the exact analysis algorithm implemented in PROC LOGISTIC works. Details about the reported statistics are available in the appendix. The "SYNTAX" section describes the statements and options in the LOGISTIC procedure for the exact methods. The "EXAMPLES" section provides several examples to illustrate the syntax and the usefulness of the method.

Dose-Response Study

To demonstrate the usefulness of exact logistic regression, consider a small dose-response study. Researchers are interested in analyzing how mortality rates change with respect to dosage of a drug. The dose data set contains life and death outcomes for six levels of drug dosage (0 to 5). Three subjects are given each specific dose of the drug, and the number of deaths are recorded.

```
data dose;
   input Dose Deaths Total @@;
   datalines;
0 0 3   1 0 3   2 0 3   3 0 3
4 1 3   5 2 3
;
run;
```

All of the cells have counts that are less than 5, which makes the applicability of large-sample theory questionable. For each subject i who received dosage x_i, $i = 1, \ldots, 18$, let $Y_i = 1$ if the subject died, $Y_i = 0$ otherwise, and $\pi_i = \Pr(Y_i = 1 | x_i)$. Then the linear logistic model for this problem is $\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + x_i\beta$, which fits a common intercept and slope for the i subjects. In the following PROC LOGISTIC invocation, the EXACT statement requests an exact analysis and the ESTIMATE option produces exact parameter estimates.

```
proc logistic data=dose descending;
   model Deaths/Total = Dose;
   exact Dose / estimate=both;
run;
```

Figure 1 displays some of the unconditional asymptotic results that are produced by default. The likelihood ratio and score tests reject the null hypothesis that β is zero. However, the Wald test does not reject this null hypothesis. The seemingly conflicting conclusions of these tests are a telltale sign that the large-sample approximation is unreliable. The estimates for the intercept α and the slope β both have p-values greater than 0.05, indicating marginal influence. The confidence limits for the odds ratio of the dose parameter contain the value 1, from which you could conclude, if you accept the model, that there is no change in mortality with a change in dosage.

Figure 1 Output from Asymptotic Analysis

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.1478	1	0.0043
Score	5.7943	1	0.0161
Wald	2.7249	1	0.0988

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.4745	5.5677	2.8958	0.0888
Dose	1	2.0804	1.2603	2.7249	0.0988

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Dose	8.007	0.677	94.679

Figure 2 shows the results from the EXACT statement. The p-values in the "Conditional Exact Tests" table lead to rejecting the null hypothesis that β is zero (no conclusions can be made about α since it is "conditioned" away). Note that the p-values for the asymptotic estimates are larger than those for the exact estimates. However, Stokes, Davis, and Koch (1995) observe that, in general, the exact methods tend to produce more conservative results. The "Exact Parameter Estimates" table shows that the slope β is estimated to be $\widehat{\beta} = 1.8$. Since the 95% confidence interval for the odds ratio of $\widehat{\beta}$ does not contain 1, the odds of death increase significantly with dosage.

Figure 2 Output from Exact Analysis

The LOGISTIC Procedure

Exact Conditional Analysis

Conditional Exact Tests

Effect	Test	Statistic	--- p-Value --- Exact	Mid
Dose	Score	5.4724	0.0245	0.0190
	Probability	0.0110	0.0245	0.0190

Figure 2 *continued*

Exact Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		p-Value
Dose	1.8000	1.0784	0.1157	5.8665	0.0245

Exact Odds Ratios

Parameter	Estimate	95% Confidence Limits		p-Value
Dose	6.049	1.123	353.000	0.0245

The unconditional asymptotic and conditional exact results produce somewhat conflicting conclusions for this example. Stokes, Davis, and Koch (1995) recommend looking at the exact results when sample sizes are small and the approximate p-values are less than 0.10. For this example, the small sample size and the conflicting results for the asymptotic hypothesis tests indicate that an exact analysis would be more appropriate.

METHODOLOGY

The theory of exact conditional logistic regression analysis was originally laid out by Cox (1970), and the computational method employed in PROC LOGISTIC is described in Hirji, Mehta, and Patel (1987), Mehta, Patel, and Senchaudhuri (1992, 2000), and Hirji (1992). Other references that provide useful summaries of the derivations include Cox and Snell (1989), Agresti (1990), and Mehta and Patel (1995).

This section summarizes the methodology behind logistic regression and explains how the algorithm for exact computations works.

Logistic Regression

Consider n independent Bernoulli random variables $Y_1, \ldots, Y_n$ that have observed values $\boldsymbol{y}_0 = (y_{01}, \ldots, y_{0n})'$. For each observation $i = 1, \ldots, n$, let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip}, x_{i,p+1}, \ldots, x_{i,p+q})'$ be a $p+q$ vector of explanatory variables, and denote $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$. Let $\pi_i = \pi(\boldsymbol{x}_i) = \Pr(Y_i = 1 | \boldsymbol{x}_i)$ be the event probability for each $i = 1, \ldots, n$, and denote $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)'$. Then the logistic regression model is $\text{logit}(\boldsymbol{\pi}) = \boldsymbol{X\beta}$, or

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x}_i'\boldsymbol{\beta}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{p+q})'$ is the unknown parameter vector.

The joint probability of the observed $\boldsymbol{y}_0$ is a product of n Bernoulli functions:

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \pi_i^{y_{0,i}} (1-\pi_i)^{1-y_{0,i}} \\ &= \frac{\exp(\boldsymbol{y}_0'\boldsymbol{X\beta})}{\prod_{i=1}^{n}[1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})]} \end{aligned}$$

Unconditional likelihood inference is based on maximizing this likelihood function, and several asymptotic statistics (likelihood ratio, score, and Wald) can be used to perform hypothesis tests.

To perform conditional inference, first observe that the sufficient statistics for the β_j in the unconditional likelihood function are the corresponding $T_j = \sum_{i=1}^{n} y_i x_{ij}$, where y_i is a realization of Y_i. To create the probability density function (pdf) for $\boldsymbol{T} = (T_1, \ldots, T_{p+q})'$, sum over all binary sequences $\boldsymbol{y}$ that generate an observable $\boldsymbol{t}$

$$\Pr(\boldsymbol{T} = \boldsymbol{t}) = \frac{C(\boldsymbol{t})\exp(\boldsymbol{t}'\boldsymbol{\beta})}{\prod_{i=1}^{n}[1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]}$$

where $C(\boldsymbol{t}) = ||\{\boldsymbol{y} : \boldsymbol{y}'\boldsymbol{X} = \boldsymbol{t}'\}||$ is the number of sequences $\boldsymbol{y}$ that generate $\boldsymbol{t}$. Suppose the p parameters $\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_p)'$ are nuisance parameters; that is, the current analysis is geared toward the last q parameters $\boldsymbol{\beta}_1$. Denote

the sufficient statistics for the nuisance parameters as $\boldsymbol{T}_0 = (T_1, \ldots, T_p)$, the corresponding observed values as $\boldsymbol{t}_0$, and the corresponding columns of $\boldsymbol{X}$ as $\boldsymbol{X}_0$. Similarly, define $\boldsymbol{T}_1$, $\boldsymbol{t}_1$, and $\boldsymbol{X}_1$ for the parameters of interest. The nuisance parameters can be removed from the analysis by conditioning on their sufficient statistics to create the conditional likelihood

$$\begin{aligned} \Pr(\boldsymbol{T}_1 = \boldsymbol{t}_1 | \boldsymbol{T}_0 = \boldsymbol{t}_0) &= \frac{\Pr(\boldsymbol{T} = \boldsymbol{t})}{\Pr(\boldsymbol{T}_0 = \boldsymbol{t}_0)} \\ &= \frac{C(\boldsymbol{t}) \exp(\boldsymbol{t}_1' \boldsymbol{\beta}_1)}{\sum_{\boldsymbol{u}} C(\boldsymbol{u}, \boldsymbol{t}_0) \exp(\boldsymbol{u}' \boldsymbol{\beta}_1)} \end{aligned}$$

where $C(\boldsymbol{u}, \boldsymbol{t}_0)$ is the number of vectors $\boldsymbol{y}$ such that $\boldsymbol{y}'\boldsymbol{X}_1 = \boldsymbol{u}$ and $\boldsymbol{y}'\boldsymbol{X}_0 = \boldsymbol{t}_0$.

Conditional asymptotic inference is performed by maximizing the conditional likelihood and producing conditional statistics similar to the unconditional likelihood case.

Conditional exact inference is based on generating the conditional distribution for the parameters of interest. This distribution is called the *permutation distribution* or *exact conditional distribution*. The conditional pdf $\Pr(\boldsymbol{T}_1 = \boldsymbol{t}_1 | \boldsymbol{T}_0 = \boldsymbol{t}_0)$ is denoted as $f_{\boldsymbol{\beta}_1}(\boldsymbol{t}_1 | \boldsymbol{t}_0)$. The following section describes the generation of this distribution, and the appendix provides details about the tests and inferences.

Exact Conditional Distribution

The goal of the exact conditional analysis is to determine how likely the observed response $\boldsymbol{y}_0$ is with respect to all 2^n possible responses $\boldsymbol{y} = (y_1, \ldots, y_n)'$. One way to proceed is to generate every $\boldsymbol{y}$ vector for which $\boldsymbol{y}'\boldsymbol{X}_0 = \boldsymbol{t}_0$ and count the number of vectors $\boldsymbol{y}$ for which $\boldsymbol{y}'\boldsymbol{X}_1$ is equal to each unique $\boldsymbol{t}_1$.

Suppose you have the following data, and you want to find the permutation distribution of the sufficient statistics for $\boldsymbol{X}_1$ conditional on those for $\boldsymbol{X}_0$.

Table 1

Observation	y	x_0	x_1
1	0	1	1
2	1	1	1
3	0	1	2
4	1	1	0

Here, the observed data are $\boldsymbol{y}_0 = (0,1,0,1)'$, $\boldsymbol{X}_0 = (1,1,1,1)'$, and $\boldsymbol{X}_1 = (1,1,2,0)'$. The observed $\boldsymbol{t}$ is computed as $(t_0, t_1) = 0 \times (1,1) + 1 \times (1,1) + 0 \times (1,2) + 1 \times (1,0) = (2,1)$, so you are conditioning on $t_0 = 2$. Tabulate the 16 possible $\boldsymbol{y} = (y_1, y_2, y_3, y_4)'$ vectors and their resulting $\boldsymbol{t} = (t_0, t_1)$ vectors:

Table 2

Vector	y_1	y_2	y_3	y_4	t_0	t_1
1	0	0	0	0	0	0
2	0	0	0	1	1	0
3	0	0	1	0	1	2
4	0	0	1	1	2	2
5	0	1	0	0	1	1
6	0	1	0	1	2	1
7	0	1	1	0	2	3
8	0	1	1	1	3	3
9	1	0	0	0	1	1
10	1	0	0	1	2	1
11	1	0	1	0	2	3
12	1	0	1	1	3	3

Table 2 *continued*

Vector	y_1	y_2	y_3	y_4	t_0	t_1
13	1	1	0	0	2	2
14	1	1	0	1	3	2
15	1	1	1	0	3	4
16	1	1	1	1	4	4

The conditional distribution is derived from this joint distribution by extracting every vector with $t_0 = 2$:

Table 3

t_0	t_1	Frequency	Probability
2	1	2	2/6
2	2	2	2/6
2	3	2	2/6
Total		6	1

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 $\boldsymbol{y}$ vectors—more than a billion! You can reduce the number of vectors to look at if you are conditioning on the intercept by processing $\binom{30}{\sum_i y_{0,i}}$ vectors, but this does not improve the situation much.

The *multivariate shift algorithm* developed by Hirji, Mehta, and Patel (1987) is a faster method of generating and counting the $\boldsymbol{y}$ vectors for larger problems. The algorithm is based on the following observation. Given any $\boldsymbol{y} = (y_1, \ldots, y_n)'$ and a design $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, let $\boldsymbol{y}_{(i)} = (y_1, \ldots, y_i)'$ and

$$\boldsymbol{X}_{(i)} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i)' = \begin{pmatrix} x_{1,1} & \ldots & x_{1,p+q} \\ \vdots & & \vdots \\ x_{i,1} & \ldots & x_{i,p+q} \end{pmatrix}$$

be the first i rows of each matrix. Write the sufficient statistic based on these i rows as $\boldsymbol{t}'_{(i)} = \boldsymbol{y}'_{(i)}\boldsymbol{X}_{(i)}$. A recursion relation results: $\boldsymbol{t}_{(i+1)} = \boldsymbol{t}_{(i)} + y_{i+1}\boldsymbol{x}_{i+1}$.

The previous example is used to illustrate how this relation is exploited.

Figure 3 displays a tree diagram where each row after the first row corresponds to an observation i, and each node of the tree is denoted by a pair of digits that represent the value of $\boldsymbol{t}_{(i)}$. The top node in the tree is initially set to 00 and indicates that $t_{(0),0} = 0$ and $t_{(0),1} = 0$, or $\boldsymbol{t}_{(0)} = (0, 0)$. Each row of the tree is numbered; these numbers represent the *stages* of the algorithm. To move down the branches, add y times the next value of (x_0, x_1) to the current value of (t_0, t_1), for $y = 0$ and 1. For example, starting at stage 0 with $\boldsymbol{t}_{(0)} = (0, 0)$ =00, take $\boldsymbol{t}_{(0)} + y\boldsymbol{x}_1 = (0, 0) + 0(1, 1) = (0, 0)$ =00 as the value of the left branch of the first stage and $(0, 0) + 1(1, 1) = (1, 1)$ =11 for the right branch.

Figure 3 Stages of the Multivariate Shift Algorithm

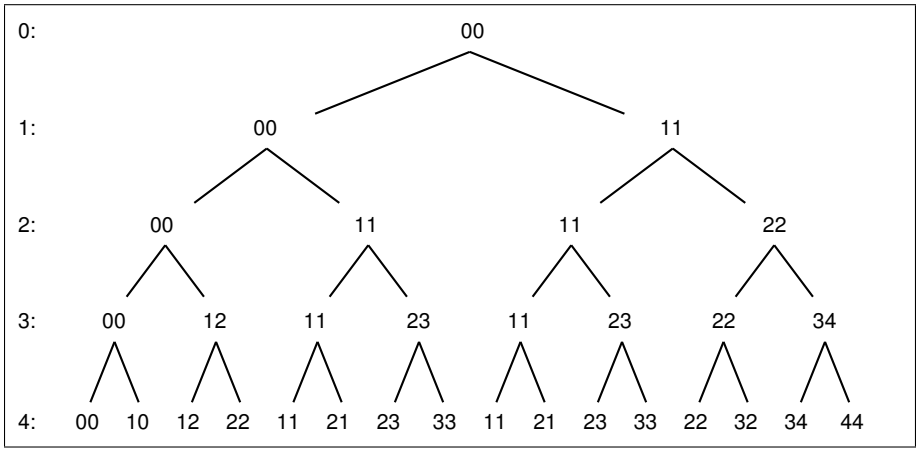


Table 4 displays the distribution created from the frequency table of the $2^4 = 16$ possible $\boldsymbol{t}$ vectors from the final stage of Figure 3.

Table 4

t_0	t_1	Frequency	Probability
0	0	1	1/16
1	0	1	1/16
1	1	2	2/16
1	2	1	1/16
2	1	2	2/16
2	2	2	2/16
2	3	2	2/16
3	2	1	1/16
3	3	2	2/16
3	4	1	1/16
4	4	1	1/16
Total		16	1

The conditional distribution obtained for the observed $t_0 = 2$ is the same as previously generated in Table 3.

You can observe five shortcuts from the example:

- There are two $(1, 1)$ nodes in the second stage of Figure 3, and the branches below those two values are identical. Computation time is significantly reduced if you process an entire stage and combine identical nodes; however, the trade-off is that a list of all valid nodes in a stage must be saved, increasing memory usage.

  In order to obtain the correct distribution, each node descended from this combined $(1, 1)$ node must count as two outcomes.

- In the third stage, there is no way to get from $(0, 0)$ to $(2, t_1)$ in one step by adding 0 or 1 times $(1, 0)$. Similarly, if the value of t_0 in the third stage is 3, it cannot be reduced to the necessary value of 2. These illustrate what Hirji, Mehta, and Patel (1987) call *infeasibility criteria*.

- The infeasibility criterion is more effective when the larger covariate values are processed first. For example, if the value of x0 for the fourth observation was 2 instead of 1, then you could obtain a $(2, t_1)$ from the $(0, 0)$ third stage node, and hence you would have to process the extra nodes.

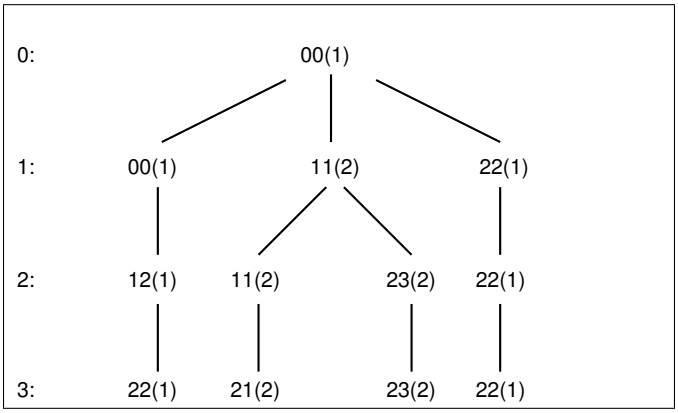
- Since the first two observations have the same covariate values, you can jump from stage 0 to stage 2 by combining the first two observations, incrementing the values in stage 0 along three branches with $i \times (1, 1)$ for

$i = 0, 1, 2$, and modifying the counts by $\binom{2}{i}$. This saves search time at the expense of computing binomial coefficients.

- After a distribution is computed for a set of effects, a distribution for any subset of these effects can be produced by scanning the larger distribution. In the example, the conditional distribution for $f_0(t_1|t_0 = 2)$ was produced from the joint distribution $f_0(t_0, t_1)$ by extracting members that have $t_0 = 2$.

To obtain the trimmed tree for the conditional distribution with the observed $t_0 = 2$, apply the five shortcuts to the tree in Figure 3 and obtain the tree shown in Figure 4. Attached to each node is a count, in parentheses, which enables you to compute the correct value for the conditional distribution. For example, the tree is initially set to 00 with a count of 1.

Figure 4 Stages of the Modified Multivariate Shift Algorithm

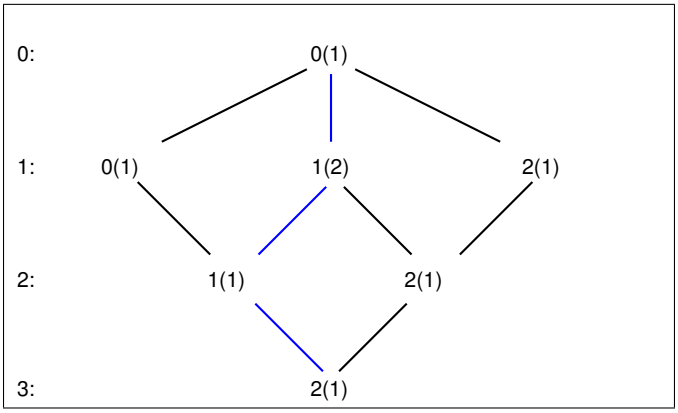


PROC LOGISTIC automatically uses these shortcuts in its implementation of the multivariate shift algorithm in order to improve performance.

PROC LOGISTIC also employs 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 $\boldsymbol{y}$ vector. These networks are combined and the set of feasible y_i is further reduced. Then the network algorithm uses this knowledge to build the exact distribution without adding as many intermediate $\boldsymbol{t}_{i+1}$ as the multivariate shift algorithm does.

The network generated from Figure 4 is shown in Figure 5. Note that the values displayed in Figure 5 are for t_0, not for t_1, and that the counts do not accumulate as you step through the stages.

Figure 5 Network Algorithm



There are four possible *paths* through this network, from the initial $t_{(0),0} = 0$ at stage 0 to the final $t_{(3),0} = 2$ in stage 3, and each step in the path carries a count. To compute how much each path contributes to the frequency distribution,

simply multiply the counts. For example, the path 0-1-1-2 accounts for $1 \times 2 \times 1 \times 1 = 2$ of the counts in the frequency table. To create the distribution for t_1 conditional on $t_0 = 2$, compute the sufficient statistics $\sum_i y_i x_{1i}$ and the product of the counts for each of the four paths as follows:

Table 5

Sufficient Statistics	Product of Counts
$(0 \times 1) + (1 \times 2) + (1 \times 0) = 2$	$1 \times 1 \times 1 \times 1 = 1$
$(1 \times 1) + (0 \times 2) + (1 \times 0) = 1$	$1 \times 2 \times 1 \times 1 = 2$
$(1 \times 1) + (1 \times 2) + (0 \times 0) = 3$	$1 \times 2 \times 1 \times 1 = 2$
$(2 \times 1) + (0 \times 2) + (0 \times 0) = 2$	$1 \times 1 \times 1 \times 1 = 1$

This results in the same conditional distribution as shown in Table 3.

The network algorithm is invoked by specifying the EXACTOPTIONS(METHOD=NETWORK) option of the PROC LOGISTIC statement. The NETWORK method can be faster and require less memory than the DIRECT method. The NETWORK method is invoked by default for most analyses.

A hybrid Monte Carlo and network algorithm is also available. It is described by Mehta, Patel, and Senchaudhuri (2000) and extends their 1992 algorithm by sampling from the combined network to build the exact distribution. This method creates a network and 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. For example, after producing the network in Figure 5, instead of stepping through every single path, the algorithm steps randomly through the network with probabilities that are proportional to the counts. So for the first step, you can move from 0 to 0 with probability $1/(1 + 2 + 1) = 0.25$, 0 to 1 with probability 0.5, and 0 to 2 with probability 0.25.

The hybrid method is invoked by specifying the EXACTOPTIONS(METHOD=NETWORKMC) option of the PROC LOGISTIC statement. 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.

The bulk of the computation time and memory is consumed by the creation of 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 trivial.

Exact Capabilities of PROC LOGISTIC

The exact conditional logistic regression analysis in PROC LOGISTIC provides the following:

- two tests for the null hypothesis that the parameters for the effects specified in the EXACT statement are zero: the exact probability test and the exact conditional scores test. For each test, the "Conditional Exact Tests" table displays the following:
  - a test statistic
  - an exact *p*-value, which is the probability of obtaining a more extreme statistic than the observed value, assuming the null hypothesis
  - a mid *p*-value, which adjusts for the discreteness of the distribution
- parameter estimates and odds ratios for each effect in the EXACT statement conditional on the values of all the other parameters in the model. For each estimate, the "Exact Parameter Estimates" and "Exact Odds Ratios" tables display the following:
  - the exact conditional maximum likelihood estimate (CMLE), or, in cases where the CMLE does not exist, the median unbiased estimate
  - the standard error of the estimate
  - one- or two-sided confidence limits
  - a one- or two-sided *p*-value for testing that the parameter estimate is zero or that the odds ratio is one
- optionally, output data sets that contain the derived distributions and summary statistics

Hypothesis tests can be generated for each individual effect in an EXACT statement or for all effects simultaneously. See the appendix for more detailed information about the reported tests and statistics.

SYNTAX

The following statements control the exact analyses in the LOGISTIC procedure. Items within the <> are optional.

PROC LOGISTIC < **EXACTONLY** > < **EXACTOPTIONS(options)** > **;**
EXACT < **'label'** > < **INTERCEPT** > < **effects** > < / **options** > **;**

Several EXACT statements can be specified in any program, but they must follow the MODEL statement. The EXACTOPTIONS option in the PROC LOGISTIC statement affects every exact analysis requested, whereas options in an EXACT statement are local to that statement.

PROC LOGISTIC Options

The EXACTONLY option suppresses the unconditional likelihood analyses that PROC LOGISTIC usually performs, and only the exact analyses are executed. Input data sets can be in single-trial or events/trials form, but the response variable must have at most two levels.

EXACTOPTIONS

Options specified in parentheses after the EXACTOPTIONS option apply to every EXACT statement in the program. The following *options* are available:

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. The ADDTOBS option was introduced in SAS/STAT 9.1.

BUILDSUBSETS builds every distribution for sampling. Some exact distributions are created by taking a subset of a previously generated exact distribution. The METHOD=NETWORKMC option has the effect of using fewer than the desired *n* samples; see the N= option for more details. The BUILDSUBSETS option suppresses this subsetting behavior. The BUILDSUBSETS option was introduced in SAS/STAT 9.2.

EPSILON=*value* controls how the partial sums $\sum_{i=1}^{j} y_i \boldsymbol{x}_i$ are compared. *value* must be between 0 and 1; by default, *value*=1E–8. The EPSILON= option was introduced in SAS/STAT 9.2.

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 the default when you are not conditioning out any covariates, is the default in releases prior to SAS/STAT 9.2 when you are conditioning out only the intercept, and is the default when LINK=GLOGIT.

NETWORK invokes a network algorithm described in Mehta, Patel, and Senchaudhuri (1992). This method builds a network for each parameter that you are conditioning out, combines the networks, and 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 and then samples from that network. This method does not reject any of the samples, but it uses 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.

The DIRECT method was the only method available until the NETWORK and NETWORKMC methods were introduced in SAS/STAT 9.0 along with the METHOD= option.

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 $f(\text{X1}, \text{X2})$, and let $f(\text{X1}|\text{X2} = x2)$ 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 $\hat{f}(\text{X1}, \text{X2})$ of $f(\text{X1}, \text{X2})$, from which the test is computed. However, the parameter estimate for X1 is computed from the subset of $\hat{f}(\text{X1}, \text{X2})$ that has $\text{X2} = x2$, 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 permit 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.

The N= option was introduced in SAS/STAT 9.0.

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. The ONDISK option was introduced in SAS/STAT 9.1.

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. The seed is displayed in the "Model Information" table. The SEED= option was introduced in SAS/STAT 9.0.

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. The STATUSN= option was introduced in SAS/STAT 9.0.

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.

EXACT Statement

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). The EXACT statement was introduced in SAS/STAT 8.1.

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 conditional 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 EXACTOPTIONS option METHOD=DIRECT is invoked 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 following options can be specified in each EXACT statement:

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. The ALPHA= option was introduced in SAS/STAT 8.1.

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. For classification variables, you must also specify the PARAM=REF option in the CLASS statement.

BOTH specifies that the parameters and odds ratios be estimated.

The standard error was added in SAS/STAT 9.2.

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 tests are indicated in the "Conditional Exact Tests" table by the label "Joint."

JOINTONLY
performs only the joint test that all of the parameters are simultaneously equal to zero. 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.

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. The CLTYPE= option was introduced in SAS/STAT 9.1.

MIDPFACTOR=δ_1 | (δ_1, δ_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 $\delta_2 = 1.0$. The MIDPFACTOR= option was introduced in SAS/STAT 9.1.

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.

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 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.

Use with Other Statements and Options

Several existing options can be used in conjunction with the EXACT statement. You can define classification effects with the CLASS statement, you can process the data with BY groups, you can include a frequency variable with the FREQ statement, and you can include stratification variables with the STRATA statement. The NOINT option in the MODEL statement suppresses the intercept term.

If you receive messages that indicate that the Newton-Raphson iterations for the parameter estimates or confidence intervals did not converge, specifying the ABSFCONV=, FCONV=, XCONV=, or MAXITER= options in the MODEL statement might help.

The CONTRAST, ODDSRATIO, OUTPUT, ROC, ROCCONTRAST, SCORE, TEST, and UNITS statements are not available with an exact analysis. 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.

Output

PROC LOGISTIC presents the exact conditional analysis results in several tables:

- The "Conditional Exact Tests" table displays the score and probability statistics for testing that all parameters for the specified effects are zero. By default, tests for a single-effect model are produced, but tests for multiple-effect models can also be requested. Exact and mid *p*-values are also generated.
- The "Exact Parameter Estimates" table displays the individual parameter estimates (conditional on all other parameters in the model), the standard error of the estimate, confidence limits, and a *p*-value for testing that the parameter is zero.
- The "Exact Odds Ratios" table displays odds ratios for individual parameters, confidence limits, and a *p*-value for testing that the odds ratio is 1.
- The "Sufficient Statistics" table displays the sufficient statistic for each parameter in the model. This table is generated only when you also specify the OUTDIST= option to output the distribution to a SAS data set. The information is useful for certain further analyses.

As with all SAS procedure output, you can use ODS (Output Delivery System) to create output data sets of the values included in these tables by specifying a statement such as the following:

```
ods output SuffStats=suff
           ExactTests=test
           ExactParmEst=est
           ExactOddsRatio=odds;
```

EXAMPLES

The following examples illustrate different types of exact analysis. The data in these examples were constructed solely for illustrative purposes. The "Sparse Data" example illustrates that the maximum likelihood estimate for the unconditional likelihood analysis might not exist, rendering the asymptotic inference impossible, while the exact conditional inference is still plausible. The "Stratified Analyses" example demonstrates how to use exact conditional analysis to adjust for within-strata correlation. The "Crossover Clinical Trial" example is a popular Phase II analysis for the pharmaceutical industry.

Sparse Data

There are several types of data for which unconditional maximum likelihood estimates fail to exist, or for which the theory is not applicable. For data with small cell counts, tests based on the asymptotic normality of the maximum likelihood estimates might not be valid. For other data, the maximum likelihood estimates might not exist and the estimated dispersion matrix might be unbounded. In this example, the data set separate contains variables which perfectly predict the response, yielding a complete separation of data points.

```
data separate;
   input A B Response count @@;
   datalines;
0 0 1 1  0 1 0 2  1 0 1 8  1 1 1 21
;
```

The following statements fit the logistic regression model:

$$\text{logit}(\pi_i) = \alpha + \text{A}\beta_1 + \text{B}\beta_2$$

The JOINT option tests the joint hypothesis that $\beta_1 = \beta_2 = 0$, and the ESTIMATE option produces the individual parameter estimates of β_1 and β_2. The OUTDIST= option creates a data set that contains all permutation distributions required for this analysis.

```
proc logistic data=separate;
   freq count;
   model Response=A B;
   exact A B / joint estimate outdist=dist;
run;
```

Figure 6 shows that the usual asymptotic analysis indicates that complete separation has occurred. You can see that the parameter estimates do not converge if you specify both the ITPRINT and NOCHECK options in the MODEL statement. However, exact tests and estimates for the conditional analysis can still be computed and are displayed in Figure 7.

Figure 6 Convergence Status

```
The LOGISTIC Procedure

Model Convergence Status

Complete separation of data points detected.
```

In Figure 7, the joint exact test of A and B is significant, but the B parameter appears insignificant. The median unbiased estimate is created instead of the conditional maximum likelihood estimate (CMLE) because the value of the observed sufficient statistic lies at an extreme of the derived distribution, implying that the CMLE does not exist. Even though the asymptotic results are unreliable, the exact analysis enables you to conclude that there is a significant effect due to A.

Figure 7 Output from EXACT Analysis

```
The LOGISTIC Procedure

Exact Conditional Analysis

Sufficient Statistics

Parameter     Value

Intercept         2
A                 0
B                 2

Conditional Exact Tests

                                        --- p-Value ---
Effect   Test          Statistic    Exact      Mid

Joint    Score           21.1153   0.0020   0.0010
         Probability      0.00202  0.0020   0.0010
A        Score           22.0000   0.0040   0.0020
         Probability      0.00395  0.0040   0.0020
B        Score            2.0000   0.3333   0.1667
         Probability      0.3333   0.3333   0.1667
```

Figure 7 *continued*

Exact Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		p-Value
A	-3.8398*	.	-Infinity	-1.0718	0.0079
B	0.6931*	.	-2.9704	Infinity	0.6667

NOTE: * indicates a median unbiased estimate.

Figure 8 displays the three permutation distributions created with the OUTDIST= option. The joint distribution of A and B conditional on the intercept is contained in observations 1 through 8, the distribution for A conditional on the intercept and B is in observations 9 through 11, and the distribution for B conditional on the intercept and A is in observations 12 and 13. The "Sufficient Statistics" table in Figure 7 enables you to identify the row that contains the observed values. You can see that it is $(\text{intercept}, \text{A}, \text{B}) = (2, 0, 2)$, which corresponds to the second, ninth, and thirteenth rows in Figure 8. Note that only the joint distribution for the A and B variables was computed from the multivariate shift algorithm; the univariate conditional distributions were extracted from the joint distribution to save CPU time. The OUTDIST= data set has three values in the distribution for the A variable and two for the B variable. If the permutation distribution is degenerate (has only one value), then the procedure does not produce any statistics and does not output the distribution. However, for small distributions, you have to decide whether there is enough information on which to base the estimates. In this simple example, there is probably too little information contained in the conditional distribution for the B variable.

Figure 8 OUTDIST= Data Set

Obs	A	B	Count	Score	Prob
1	0	1	2	20.2622	0.00403
2	0	2	1	21.1153	0.00202
3	1	0	8	8.9654	0.01613
4	1	1	37	4.4055	0.07460
5	1	2	42	4.9644	0.08468
6	2	0	28	5.5822	0.05645
7	2	1	168	0.7281	0.33871
8	2	2	210	0.9929	0.42339
9	0	.	1	22.0000	0.00395
10	1	.	42	4.5023	0.16601
11	2	.	210	0.1995	0.83004
12	.	1	2	0.5000	0.66667
13	.	2	1	2.0000	0.33333

Stratified Analyses

If your data are collected from different hospitals or different families, you can perform a stratified analysis to control for the within-group correlation. The STRATA statement names the *variables* that define strata or matched sets to use in a *stratified conditional logistic regression* of binary response data. Observations that have the same variable levels are in the same matched set. 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 in the STRATA statement to invoke the stratified analysis, and the usual unconditional asymptotic analysis is not performed. The stratified logistic model has the form

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

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

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

The stratified data set includes a response variable Y, two explanatory variables X1 and X2, and a stratification variable Stratum.

```
data stratified;
   input Stratum Y X1 X2 count @@;
   datalines;
1 0 1 1 1  2 0 1 2 3  3 0 1 0 2
1 0 2 1 1  2 0 2 2 3  3 0 2 1 1
```

```
1 1 1 0 1   2 1 2 0 1   3 1 1 0 1
1 1 2 0 1   2 1 3 1 2   3 1 2 2 2
1 1 3 0 2               3 1 3 2 1
;
```

The following statements produce both asymptotic conditional likelihood analysis and exact conditional analysis for comparison.

```
proc logistic descending;
   freq count;
   strata Stratum;
   model Y= X1 X2;
   exact X1 X2 / jointonly estimate;
run;
```

Figure 9 shows the results of the exact conditional analysis. The joint exact test for the X1 and X2 parameters rejects the null hypothesis. However, the X2 parameter appears insignificant.

Figure 9 Exact Conditional Analysis

The LOGISTIC Procedure

Exact Conditional Analysis

Conditional Exact Tests

Effect	Test	Statistic	--- p-Value --- Exact	Mid
Joint	Score	7.9291	0.0165	0.0162
	Probability	0.000612	0.0077	0.0074

Exact Parameter Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		p-Value
X1	1.9979	0.9283	0.3140	5.2012	0.0126
X2	-1.0097	0.6938	-2.9152	0.4142	0.1931

Figure 10 shows the results of the asymptotic conditional likelihood analysis.

Figure 10 Asymptotic Conditional Likelihood Analysis

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.6425	2	0.0081
Score	7.9291	2	0.0190
Wald	4.6510	2	0.0977

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
X1	1	2.3247	1.1159	4.3404	0.0372
X2	1	-1.1143	0.7292	2.3353	0.1265

Comparing Figure 9 with Figure 10, you can see that the value of the conditional score statistic for testing the overall null hypothesis $\beta_1 = \beta_2 = 0$ is 7.9291 for both the asymptotic conditional analysis and the exact conditional analysis. However, the asymptotic conditional analysis produces a *p*-value of 0.019 by comparing the value of the conditional score statistic to a chi-squared distribution with two degrees of freedom (since there are two parameters), while exact conditional analysis produces a *p*-value of 0.0165. Inference on individual parameters is often not the same between the exact conditional analysis and the asymptotic conditional likelihood results.

Crossover Clinical Trial

One common use of conditional logistic regression is in a crossover clinical trial. In this example, the subjects are given a sequence of drugs, and their response to each drug is recorded. Each subject is considered to be a separate stratum. The goal is to determine whether the drugs have the same effect, adjusting for period and carryover effects. In this example, researchers give 15 different subjects three different drugs (A, B, P=placebo) in three consecutive periods (P1, P2, P3), and their response in each period is 1 for improvement and 0 for no improvement. The carryover effect is a classification variable that indicates which drug was given in the preceding period.

```
data Crossover (drop=P1 P2 P3);
   input Subject P1$ P2$ P3$ Improve @@;
   Period=1; Drug=P1; Carry='0'; output;
   input Improve @@;
   Period=2; Drug=P2; Carry=P1;  output;
   input Improve @@;
   Period=3; Drug=P3; Carry=P2;  output;
   datalines;
1  A B P 0 0 0    6  B A P 0 0 0    11 P A B 0 1 0
2  A B P 1 1 0    7  B A P 1 1 0    12 P B A 1 0 1
3  A B P 0 1 1    8  B P A 0 0 1    13 P B A 0 0 1
4  A P B 1 0 1    9  B P A 1 0 1    14 P B A 0 1 0
5  A P B 1 0 0    10 B P A 0 1 0    15 P B A 0 1 1
;
```

The model to be fit is

$$\text{logit}(\pi_{hi}) = \alpha_h + I(\text{Drug} = \text{A})\beta_1 + I(\text{Drug} = \text{B})\beta_2 + I(i = 1)\beta_3 + I(i = 2)\beta_4$$

where h indexes the subject, α_h are the subject intercepts, i indexes the period, and the $I(\cdot)$ are indicator variables that take the value 1 when the condition is true. Note that this model ignores carryover effects.

```
proc logistic descending exactonly;
   class Drug Period/ param=ref;
   strata Subject;
   model Improve=Drug Period;
   exact 'one'   Drug Period / jointonly;
   exact 'two'   Drug;
   exact 'three' Period;
run;
```

Figure 11 Exact Conditional Analysis

```
                    Model Information

Data Set                          WORK.CROSSOVER
Response Variable                 Improve
Number of Response Levels         2
Number of Strata                  15
Number of Uninformative Strata    2
Frequency Uninformative           6
Model                             binary logit

     Number of Observations Read          45
     Number of Observations Used          45

                Response Profile

   Ordered                      Total
     Value      Improve      Frequency

         1            1             20
         2            0             25

   Probability modeled is Improve=1.
```

Figure 11 *continued*

Class Level Information

Class	Value	Design Variables	
Drug	A	1	0
	B	0	1
	P	0	0
Period	1	1	0
	2	0	1
	3	0	0

Strata Summary

Response Pattern	Improve 1	Improve 0	Number of Strata	Frequency
1	0	3	2	6
2	1	2	6	18
3	2	1	7	21

Conditional Exact Tests for 'one'

Effect	Test	Statistic	--- p-Value --- Exact	Mid
Joint	Score	6.1476	0.1835	0.1834
	Probability	0.000238	0.1973	0.1971

Conditional Exact Tests for 'two'

Effect	Test	Statistic	--- p-Value --- Exact	Mid
Drug	Score	5.6092	0.0583	0.0562
	Probability	0.00417	0.0583	0.0562

Conditional Exact Tests for 'three'

Effect	Test	Statistic	--- p-Value --- Exact	Mid
Period	Score	0.3501	0.8605	0.8296
	Probability	0.0618	0.8605	0.8296

Even though three EXACT statements are invoked in this example, PROC LOGISTIC computes only the permutation distribution for the joint test of the drug and period parameters; the other two distributions are derived from the joint distribution.

The exact conditional score *p*-value for the test of significance of all the parameters is 0.1835; hence, you cannot reject the null hypothesis. However, the exact conditional score *p*-value for the test of no drug effects, $\beta_1 = \beta_2 = 0$, is 0.0583, while the *p*-value for the test of no period effects, $\beta_3 = \beta_4 = 0$, is 0.8605, which suggests that the period term should be dropped from this model.

APPENDIX

Hypothesis Tests

Using the same notation as in the "METHODOLOGY" section, consider testing the null hypothesis $H_0: \boldsymbol{\beta}_1 = \mathbf{0}$ against the alternative $H_A: \boldsymbol{\beta}_1 \neq \mathbf{0}$, conditional on $\boldsymbol{T}_0 = \boldsymbol{t}_0$. Under the null hypothesis, the test statistic for the exact probability test is just $f_{\boldsymbol{\beta}_1=\mathbf{0}}(\boldsymbol{t}_1|\boldsymbol{t}_0)$, while the corresponding *p*-value is the probability of getting a less likely (more extreme) statistic,

$$p(\boldsymbol{t}_1|\boldsymbol{t}_0) = \sum_{\boldsymbol{u}\in\Omega_p} f_{\mathbf{0}}(\boldsymbol{u}|\boldsymbol{t}_0)$$

where $\Omega_p = \{\boldsymbol{u}$: there exist $\boldsymbol{y}$ with $\boldsymbol{y}'\boldsymbol{X}_1 = \boldsymbol{u}$, $\boldsymbol{y}'\boldsymbol{X}_0 = \boldsymbol{t}_0$, and $f_{\mathbf{0}}(\boldsymbol{u}|\boldsymbol{t}_0) \leq f_{\mathbf{0}}(\boldsymbol{t}_1|\boldsymbol{t}_0)\}$.

For the exact conditional scores test, the conditional mean $\boldsymbol{\mu}_1$ and variance matrix $\boldsymbol{\Sigma}_1$ of the $\boldsymbol{T}_1$ (conditional on $\boldsymbol{T}_0 = \boldsymbol{t}_0$) are calculated, and the score statistic for the observed value,

$$s = (\boldsymbol{t}_1 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{t}_1 - \boldsymbol{\mu}_1)$$

is compared to the score for each member of the distribution

$$S(\boldsymbol{T}_1) = (\boldsymbol{T}_1 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{T}_1 - \boldsymbol{\mu}_1)$$

The resulting *p*-value is

$$p(\boldsymbol{t}_1|\boldsymbol{t}_0) = Pr(S \geq s) = \sum_{\boldsymbol{u} \in \Omega_s} f_{\boldsymbol{0}}(\boldsymbol{\mu}|\boldsymbol{t}_0)$$

where $\Omega_s = \{\boldsymbol{\mu}$: there exist $\boldsymbol{y}$ with $\boldsymbol{y}'\boldsymbol{X}_1 = \boldsymbol{\mu}$, $\boldsymbol{y}'\boldsymbol{X}_0 = \boldsymbol{t}_0$, and $S(\boldsymbol{\mu}) \geq s\}$.

The mid-*p* statistic, defined as

$$p(\boldsymbol{t}_1|\boldsymbol{t}_0) - \frac{1}{2} f_{\boldsymbol{0}}(\boldsymbol{t}_1|\boldsymbol{t}_0)$$

was proposed by Lancaster (1961) to compensate for the discreteness of a distribution. See Agresti (1992) for more information.

Inference for a Single Parameter

Exact parameter estimates are derived for a single parameter β_i by regarding all the other parameters $\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_{p+q})'$ as nuisance parameters. The appropriate sufficient statistics are $\boldsymbol{T}_1 = T_i$ and $\boldsymbol{T}_0 = (T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_{p+q})'$, with their observed values denoted by the lowercase t. Hence, the conditional pdf used to create the parameter estimate for β_i is

$$f_{\beta_i}(t_i|\boldsymbol{t}_0) = \frac{C(t_i, \boldsymbol{t}_0) \exp(t_i \beta_i)}{\sum_{u \in \Omega} C(u, \boldsymbol{t}_0) \exp(u \beta_i)}$$

for $\Omega = \{u$: there exist $\boldsymbol{y}$ with $T_i = u$ and $\boldsymbol{T}_0 = \boldsymbol{t}_0\}$.

The maximum exact conditional likelihood estimate is the quantity $\widehat{\beta}_i$ which maximizes the conditional pdf, and the standard error of the parameter estimate is derived from the second derivative of the conditional likelihood (Agresti 2002). A Newton-Raphson algorithm is used to perform this search. However, if the observed t_i attains either its minimum or maximum value in the permutation 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; Hirji and Tang 1998) $\widehat{\beta}_i$ is produced that satisfies $f_{\widehat{\beta}_i}(t_i|\boldsymbol{t}_0) = \frac{1}{2}$, and a Newton-Raphson-type algorithm is used to perform the search.

Likelihood ratio tests based on the conditional pdf are used to test the null H_0: $\beta_i = 0$ against various alternatives. For testing against the alternative H_A: $\beta_i > 0$, the critical region for the uniformly most powerful test consists of the upper tail of values for T_i in the permutation distribution. Thus, the one-sided significance level $p_G(t_i; 0)$ is the probability of a more extreme (greater) value:

$$p_G(t_i; 0) = \sum_{u \geq t_i} f_0(u|\boldsymbol{t}_0)$$

The one-sided significance level $p_L(t_i; 0)$ against H_A: $\beta_i < 0$ is

$$p_L(t_i; 0) = \sum_{u \leq t_i} f_0(u|\boldsymbol{t}_0)$$

The minimum of these one-sided levels is reported when the ONESIDED option is specified. The two-sided significance level $p(t_i; 0)$ against H_A: $\beta_i \neq 0$ is calculated as

$$p(t_i; 0) = 2 \min[p_L(t_i; 0), p_G(t_i; 0)]$$

An upper $100(1 - 2\epsilon)$% confidence limit for $\widehat{\beta}_i$ that corresponds to the observed t_i is the solution $\beta_U(t_i)$ of $\epsilon = p_L(t_i, \beta_U(t_i))$, and the lower confidence limit is the solution $\beta_L(t_i)$ of $\epsilon = p_G(t_i, \beta_L(t_i))$. A Newton-Raphson procedure is used to search for the solutions.

ACKNOWLEDGMENTS

I am grateful to Dorothy Watson of the CALGB Statistical Center, and Virginia Clark, Greg Goodwin, Ying So, Maura Stokes, Randy Tobias, and Allen McDowell of the SAS Advanced Analytics Division for their valuable assistance in the preparation of this manuscript.

REFERENCES

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.

Cox, D. R. (1970), *Analysis of Binary Data*, London: Metheun.

Cox, D. R. and Snell, E. J. (1989), *The Analysis of Binary Data*, Second Edition, London: Chapman & Hall.

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. and Tang, M.-L. (1998), "A Comparison of Tests for Trend," *Communications in Statistics—Theory and Methods*, 7, 943–963.

Hirji, K. F., Tsiatis, A. A., and Mehta, C. R. (1989), "Median Unbiased Estimation for Binary Data," *American Statistician*, 43, 7–11.

Lancaster, H. O. (1961), "Significance Tests in Discrete Distributions," *Journal of the American Statistical Association*, 56, 223–234.

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.

Stokes, M. E., Davis, C. S., and Koch, G. (1995), *Categorical Data Analysis Using the SAS System*, Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Robert E. Derr
SAS Institute Inc.
SAS Campus Drive, S3046
Cary, NC 27513
Phone (919) 677-8000 ext 6137
FAX (919) 677-4444
E-mail Bob.Derr@sas.com

The correct bibliographic citation for this document is as follows:
Derr, R. E. 2009. "Performing Exact Logistic Regression with the SAS System-Revised 2009."
Available http://support.sas.com/rnd/app/papers/exactlogistic2009.pdf.

This document is a revised and enhanced version of the original paper:
Derr, R. E. 2000. "Performing Exact Logistic Regression with the SAS System." *Proceedings of the Twenty-fifth Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.