The LOGISTIC Procedure

Getting Started

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 use a MODEL statement similar to that used in the REG procedure:

```sas
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. The procedure also allows the input of binary response data that are grouped:

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

Here, n represents the number of trials and r represents the number of events.

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. The following invocation of PROC LOGISTIC fits the binary logit model to the grouped data:

```sas
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;

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

The results of this analysis are shown in the following tables.
Figure 42.1: Binary Logit Model

PROC LOGISTIC first lists background information in Figure 42.1 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.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>WORK.INGOTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Variable (Events)</td>
<td>r</td>
</tr>
<tr>
<td>Response Variable (Trials)</td>
<td>n</td>
</tr>
<tr>
<td>Model</td>
<td>binary logit</td>
</tr>
<tr>
<td>Optimization Technique</td>
<td>Fisher's scoring</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Observations Read</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations Used</td>
<td>19</td>
</tr>
<tr>
<td>Sum of Frequencies Read</td>
<td>387</td>
</tr>
<tr>
<td>Sum of Frequencies Used</td>
<td>387</td>
</tr>
</tbody>
</table>

Figure 42.2: Response Profile with Events/Trials Syntax

The “Response Profile” table (Figure 42.2) 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.

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Response Profile</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Event</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>Nonevent</td>
<td>375</td>
</tr>
</tbody>
</table>

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.
Figure 42.3: Fit Statistics and Hypothesis Tests

The "Model Fit Statistics" table (Figure 42.3) contains the Akaike 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 and Heat) are included in the "Testing Global Null Hypothesis: BETA=0" table (Figure 42.3).
Figure 42.4: Parameter Estimates and Odds Ratios

The "Analysis of Maximum Likelihood Estimates" table in Figure 42.4 lists the parameter estimates, their standard errors, and the results of the Wald test for individual parameters. The odds ratio for each effect parameter, estimated by exponentiating the corresponding parameter estimate, is shown in the "Odds Ratios Estimates" table (Figure 42.4), along with 95% Wald confidence intervals.

Using the parameter estimates, you can calculate the estimated logit of $\pi$ as

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

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

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

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.

Figure 42.5: Association Table

Finally, the "Association of Predicted Probabilities and Observed Responses" table (Figure 42.5) 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 "Rank Correlation of Observed Responses and Predicted Probabilities" section.
To illustrate the use of an alternative form of input data, the following program creates the INGOTS data set with 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  27 2.8 1  1  51 1.0 0 10
  7 2.2 0  7 14 2.2 1  2  27 1.7 1  4  27 4.0 1  1  51 2.2 0  1
  7 2.8 0 12  14 2.2 0 31  27 1.7 1  4  27 4.0 0 15  51 4.0 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 SAS statements invoke PROC LOGISTIC to fit the same model using the alternative form of the input data set.

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

Results of this analysis are the same as the previous one. 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 42.6.

```
The LOGISTIC Procedure

Response Profile

+----------------+------+-+-----+
| Ordered Value  | NotReady | Total Frequency |
+----------------+------+-+-----+
|                | 0     | 375           |
| 1               | 0     |               |
+----------------+------+-+-----+
| 2               | 1     | 12            |
+----------------+------+-+-----+

Probability modeled is NotReady=1.
```

**Figure 42.6:** Response Profile with Single-Trial Syntax

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 42.6. See the "Response Level Ordering" section for more details.