Hosmer-Lemeshow Statistic

The Hosmer-Lemeshow Statistic is another measure of lack of fit. Hosmer and Lemeshow recommend partitioning the observations into 10 equal sized groups according to their predicted probabilities. Then

\[ G_{HL}^2 = \sum_{j=1}^{10} \frac{(O_j - E_j)^2}{E_j(1 - E_j/n_j)} \sim \chi^2_8 \]

where

- \( n_j \) = Number of observations in the \( j^{th} \) group
- \( O_j = \sum_i y_{ij} \) = Observed number of cases in the \( j^{th} \) group
- \( E_j = \sum_i \hat{p}_{ij} \) = Expected number of cases in the \( j^{th} \) group

**Example:** Using the THS data described earlier (4/7/2005), I added the `lackfit` option to the model statement in PROC LOGISTIC as follows:

```plaintext
model cens28dy = trt age sbpcat aist hctbase / lackfit;
```

The output corresponding to the Hosmer-Lemeshow statistic is:

**Partition for the Hosmer and Lemeshow Test**

<table>
<thead>
<tr>
<th>Group</th>
<th>Total</th>
<th>CENS28DY = 0</th>
<th></th>
<th>CENS28DY = 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Observed</td>
<td>Expected</td>
<td>Observed</td>
<td>Expected</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0.14</td>
<td>9</td>
<td>8.86</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>0</td>
<td>1.60</td>
<td>9</td>
<td>7.40</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>4</td>
<td>3.68</td>
<td>5</td>
<td>5.32</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>9</td>
<td>5.28</td>
<td>0</td>
<td>3.72</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>7</td>
<td>6.52</td>
<td>2</td>
<td>2.48</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>5</td>
<td>7.47</td>
<td>4</td>
<td>1.53</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8.03</td>
<td>0</td>
<td>0.97</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8.50</td>
<td>2</td>
<td>0.50</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8.82</td>
<td>0</td>
<td>0.18</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>12</td>
<td>11.96</td>
<td>0</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**Hosmer and Lemeshow Goodness-of-Fit Test**

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.5799</td>
<td>8</td>
<td>0.0120</td>
</tr>
</tbody>
</table>
This output shows the observed and expected numbers of cases and controls within each group, and the final test statistic. The chi-square statistic suggests that there may be lack of fit, so I refit the model including a quadratic term for age:

\[
\text{model cens28dy = trt age|age sbpcat aist hctbase / lackfit;}
\]

<table>
<thead>
<tr>
<th>Group</th>
<th>Total</th>
<th>CENS28DY = 0</th>
<th>CENS28DY = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Observed</td>
<td>Expected</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>0</td>
<td>1.36</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>3</td>
<td>3.38</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>8</td>
<td>5.24</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>8</td>
<td>6.58</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>7</td>
<td>7.74</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>7</td>
<td>8.22</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>8</td>
<td>8.54</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8.85</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>12</td>
<td>11.97</td>
</tr>
</tbody>
</table>

Hosmer and Lemeshow Goodness-of-Fit Test

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.8690</td>
<td>8</td>
<td>0.2743</td>
</tr>
</tbody>
</table>

This model shows no evidence of lack of fit based on the H-L statistic, so apparently any lack of fit has been corrected.

Ideally, incorrect model specification such as non-linearity in the predictors or missing predictors should be detectable by this statistic. However, depending on the model, $G^2_{LH}$ may not be particularly sensitive to departures from the true model. The H-L test does not provide any information regarding the nature of the lack of fit observed. Other procedures may be more useful.

Other goodness-of-fit procedures

Tsiatis (A note on the goodness-of-fit test for the logistic regression model, *Biometrika*, 67:259-251, 1980) proposes a test for goodness-of-fit obtained by grouping observations into $k$ groups according to covariate values. The goodness-of-fit test uses the score test of $H_0$ that the fitted model is correct, versus the model which adds to the fitted model a categorical variable corresponding to the groupings. For example, for the model above, one could group observations by age categories (possibly equally spaced, or of equal size - the groups are somewhat arbitrary), and perform the score test for the addition of these groupings to the model (similar to what one would do in a forward selection procedure).
Lin, Wei and Ying (Model checking techniques based on cumulative residuals, *Biometrics*, 58:1-12, 2002) propose assessing goodness-of-fit using moving sums of residuals of the form:

\[ W(x) = \sum_i I(x - b < X_{ij} < x + b)(y_i - E(y_i)) \]

where \( I(\cdot) \) is one if the argument is true, and zero otherwise, \( x \) is an arbitrary value of covariate \( j \) and \( X_{ij} \) is the value of covariate \( j \) for subject \( i \). \( W(x) \) is simply the total observed - expected for all observations with the value covariate \( j \) within \( b \) units of \( x \). By plotting \( W(x) \) against \( x \) for values of \( x \) over the range of the covariate, patterns may be apparent which suggest lack of fit. LWY propose a formal test using simulation to assess whether the observed patterns are likely by chance alone. The score test of Tsiatis may also be used to provide a formal test.

**Polychotomous Responses** (*note that in the interest of time, I skipped over some of the material in this section during class*)

Now we consider the case in which we have more than two levels of outcome variable.

Examples

- Type of Disease
- Site of Onset (*e.g.* tumor)
- Disease Severity (ordered)
- Cause of Death

Suppose that we have \( k + 1 \) response categories, 0, 1, \ldots, \( k \). Under a multinomial model there are \( k + 1 \) event probabilities which sum to one. Hence there are \( k \) independent logits to model.

We consider three approaches to this modeling, all of which produce different results.

- Log-linear model
- Continuation Ratios
- Cumulative Logit

**Log-linear Model** (*Poisson Sampling model*)

Suppose we have a covariate with levels 0, 1, 2, \ldots. We may construct the following table:
For the $i,j$ cell, we have the log-linear model $\log \lambda_{ij} = \alpha_i + \beta_j + \gamma_{ij}$ where we impose the constraint that $\gamma_{0j} = \gamma_{i0} = 0 \ \forall \ i,j$.

We could, for example, choose a baseline category for the response, say $i = 0$. Then
\[
\log \left( \frac{\lambda_{ij}}{\lambda_{0j}} \right) = \log \left( \frac{p_{ij}}{p_{0j}} \right) = \alpha_i - \alpha_0 + \gamma_{ij}
\]
Then $\gamma_{ij}$ is the log odds ratio for response level $i$ versus level $0$ and exposure level $j$ versus level $0$. This may be called the baseline logit model.

Similarly, if the responses are naturally ordered, then it may be reasonable to consider adjacent categories:
\[
\log \left( \frac{\lambda_{i+1,j}}{\lambda_{i,j}} \right) = \log \left( \frac{p_{i+1,j}}{p_{i,j}} \right) = \alpha_{i+1} - \alpha_i + \gamma_{i+1,j} - \gamma_{i,j} = \alpha_i^* + \gamma_{i,j}^*
\]
Then $\gamma_{ij}^*$ is the log odds ratio for response level $i+1$ versus level $i$ and exposure level $j$ versus level $0$.

More generally, we might have
\[
\log \lambda_{ij} = \mathbf{x}_j^T \beta_i
\]
In which case
\[
\log \left( \frac{p_{ij}}{p_{0j}} \right) = \mathbf{x}_j^T (\beta_i - \beta_0) = \mathbf{x}_j^T \beta_i^*.
\]

We may compute the probabilities themselves via
\[
p_{ij} = \frac{e^{\mathbf{x}_j \beta_i^*}}{1 + \sum_{l \geq 1} e^{\mathbf{x}_j \beta_l^*}}
\]
For example, if $k = 2$, we have:
\[
p_{0j} = \frac{1}{1 + e^{\mathbf{x}_j \beta_1^*} + e^{\mathbf{x}_j \beta_2^*}}, \quad p_{1j} = \frac{e^{\mathbf{x}_j \beta_1^*}}{1 + e^{\mathbf{x}_j \beta_1^*} + e^{\mathbf{x}_j \beta_2^*}}, \quad p_{2j} = \frac{e^{\mathbf{x}_j \beta_2^*}}{1 + e^{\mathbf{x}_j \beta_1^*} + e^{\mathbf{x}_j \beta_2^*}}
\]

Hypothetical example: Suppose we consider death from Heart Disease or Cancer. There is no particular order to these outcomes, so this kind of model may make some sense. In the accompanying
plot, the probabilities of death from heart disease and cancer change as a function of age and this model allows them to change quite independently.

Hypothetical Example
Death from Heart Disease vs. Cancer

Note that if we consider a plot of logits for each of heart disease and cancer versus survival we have straight lines:
Continuation Ratios

Take \( k = 2 \) for a moment and consider two parallel schemes for constructing the 2 independent logits:

<table>
<thead>
<tr>
<th>Scheme A</th>
<th>Scheme B</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \log \left( \frac{p_{0j}}{p_{1j} + p_{2j}} \right) ]</td>
<td>[ \log \left( \frac{p_{2j}}{p_{0j} + p_{1j}} \right) ]</td>
</tr>
<tr>
<td>[ \log \left( \frac{p_{1j}}{p_{2j}} \right) ]</td>
<td>[ \log \left( \frac{p_{1j}}{p_{0j}} \right) ]</td>
</tr>
</tbody>
</table>

In General: \[ \log \left( \frac{p_{ij}}{\sum_{l>i} p_{lj}} \right) \quad \log \left( \frac{p_{ij}}{\sum_{l<i} p_{lj}} \right) \]

If the categories are 0, 1, 2, \ldots, \( k \), then under Scheme A, we first consider the lowest category (\( i = 0 \)) versus all remaining categories. For each category \( i, i < k \), we consider category \( i \) versus all higher categories. Under Scheme B, we consider for each category \( i, i < 0 \), we consider category \( i \) versus all lower categories.

For example, if the categories are “Alive”, “Heart Disease Death” and “Cancer Death”, we might consider logits for “Alive” versus “Dead” and among the deaths “Heart Disease Death” versus “Cancer Death”.

If we set the \( i \)th logit equal to \( x_j^T \beta_i \), we can express the multinomial probabilities of being in a particular category as follows:

take \( k = 3 \), Scheme A. First we have that

\[ p_{0j} = \Pr\{y_j = 0\} = \frac{e^{x_j^T \beta_0}}{1 + e^{x_j^T \beta_0}} \]

next,

\[ p_{1j} = \Pr\{y_j = 1\} = \Pr\{y_j = 1|y_j \geq 1\} \Pr\{y_j \geq 1\} = \frac{e^{x_j^T \beta_1}}{1 + e^{x_j^T \beta_1}} \frac{1}{1 + e^{x_j^T \beta_1}} \frac{1}{1 + e^{x_j^T \beta_0}}. \]

Similarly,

\[ p_{2j} = \Pr\{y_j = 2\} = \Pr\{y_j = 2|y_j \geq 1\} \Pr\{y_j \geq 1\} \Pr\{y_j \geq 1\} = \frac{e^{x_j^T \beta_2}}{1 + e^{x_j^T \beta_2}} \frac{1}{1 + e^{x_j^T \beta_2}} \frac{1}{1 + e^{x_j^T \beta_1}} \frac{1}{1 + e^{x_j^T \beta_0}} \]

Finally,

\[ p_{3j} = \Pr\{y_j = 3\} = \Pr\{y_j = 3|y_j \geq 1\} \Pr\{y_j \geq 1\} \Pr\{y_j \geq 1\} = \frac{1}{1 + e^{x_j^T \beta_2}} \frac{1}{1 + e^{x_j^T \beta_1}} \frac{1}{1 + e^{x_j^T \beta_0}}. \]
In all cases, the $p_{ij}$ are products of conditional probabilities each of which depends one only on one of the $\beta_j$. Since the likelihood is composed of products of the above $p_{ij}$, it will also factor into a product of conditional likelihoods, each of which can be independently maximized. Hence, this model can be fit as a series of binary logistic regression models.

Each binary model is fit by restricting the data set to only those observations with responses in categories $i$ through $k$ (under Scheme A) or $0$ through $i$ (under Scheme B) and considering response level $i$ versus the remaining categories.

Continuation Ratio models give different fitted values than log-linear models and Scheme A gives different fitted values than Scheme B.

**Cumulative Logits.**

Cumulative Logit models are typically used when we have ordinal responses. That is, we may consider the outcomes as a series of progressively worse or more extreme outcomes. In this case the relevant questions usually involve probabilities of subjects progressing from one category to the next, or probabilities of a subject being above of a given category versus at or lower than a given category.

**Example: Disease Severity**

![Disease Severity Chart]

In the above plot, the size of the ‘none’ category decreases slightly with exposure. This suggests that exposure increases the risk of disease without regard for severity. On the other hand, the ‘mild’ category also decreases with exposure, yet this fact by itself is not very meaningful since we don’t know whether or not subjects who otherwise would have been in the mild category were moved to the ‘none’ category or the ‘moderate’ or worse categories. The conclusion might be quite different depending on which of these happened. More meaningful is the observation that
the combined ‘none’ and ‘mild’ categories shrunk, suggesting that the rate of moderate or worse disease increased. This leads us to the cumulative logit model:

\[
\log \left( \frac{\Pr \{ y_j \leq i \}}{\Pr \{ y_j > i \}} \right) = \alpha_i + x_j^T \beta_i.
\]

If \( \beta_i = \beta \forall i \), this is the “Proportional Odds” model.

We also require that \( \alpha_0 \leq \alpha_1 \leq \ldots \) (although this will hold automatically in a proportional odds model).

We have

\[
\Pr \{ y_j \leq i \} = \frac{e^{\alpha_i + x_j^T \beta_i}}{1 + e^{\alpha_i + x_j^T \beta_i}}.
\]

Note that if \( \beta_i \neq \beta_i' \) for some \( i \neq i' \), then \( \alpha_i + \beta_i x = \alpha_{i'} + \beta_{i'} x \) for some \( x \) and the lines will cross. If \( \alpha_i + x_j^T \beta_i > \alpha_{i'} + x_j^T \beta_{i'} \) for some \( i < i' \) then \( \Pr \{ y_j \leq i \} > \Pr \{ y_j \leq i' \} \) which is nonsense. Hence if a cumulative logit model is NOT a proportional odds model, there may be particular values of \( x \) for which the predicted values are inconsistent.

Under the proportional odds model, the shapes of all curves are the same, they are shifted progressively to the left for increasing \( i \). The horizontal distance between adjacent curves is \( (\alpha_i - \alpha_{i-1})/\beta \).

Odds Ratios:

Let the odds of \( \{ y_j \leq i \} \) be

\[
\psi_{ij} = \frac{\Pr \{ y_j \leq i \}}{\Pr \{ y_j > i \}} = e^{\alpha_i + x_j^T \beta} \quad \text{so}
\]

\[
\frac{\psi_{ij}}{\psi_{ij'}} = e^{\alpha_i - \alpha_{i'}} \quad \text{is independent of } j \quad \text{(proportional odds)}
\]

and

\[
\frac{\psi_{ij}}{\psi_{ij'}} = e^{(x_j - x_{j'})^T \beta} \quad \text{is independent of } i
\]

With more than two levels of the outcome variable, SAS fits a proportional odds model and uses the score test for to test for proportionality.