Recap: Nominal logistic regression allows us to predict probabilities for a response variable with more than 2 categories
- One category is the reference group and we estimate the odds ratio for 2 categories at a time (e.g., given that it is one of those 2 outcomes, the ratio of the odds for category $j$ to $J$).
- Model checking, statistical significance of individual or groups of predictors, and model utility are all analyzed in the same manner as before.
  - $df = (J - 1) \times (df \text{ for the variables in each logit)}$

Example 1: Hosmer and Lemeshow report on a study of low birth weight (less than 2500 grams) and associated risk factors. Data were collected as part of a larger study at Baystate Medical Center in Springfield, MA. The LowBirthWeight.jmp data set contains information on 189 births to women seen in the obstetrics clinic. The data in the birth weight category column have been coded as:
  - Low birth weight: birth weight $\leq 2500$
  - Below average birth weight: $2500 < \text{birth weight} \leq 3000$
  - Above average birth weight: $3000 < \text{birth weight} \leq 3500$
  - Heaviest: birth weight $> 3500$
I have also ordered the outcomes in this order as well.

(a) Create a two-way table of birth weight code by smoking status of mother (0 = No, 1 = Yes). Calculate the odds of lower birth weight categories vs. the heaviest category for smokers compared to nonsmokers. Keep in mind that we expect mother smoking to increase the odds of a lower weight for the child. How do these odds ratios compare to what we would get from logistic regression? $df$?

Above avg vs. Heaviest:
$$\frac{17}{11} \times \frac{35}{29} = 1.86 = e^{1.121}$$
$$\frac{16}{28} \times \frac{35}{22} = 2.31 = e^{1.393}$$
$$df = 6$$

(b) How do the odds ratios compare? What does this suggest?

Another way to look at these data is to create a new two-way table that examines the outcomes that are in the low birth weight category compared to the remaining categories.

<table>
<thead>
<tr>
<th></th>
<th>lowest birth weight</th>
<th>not lowest birth weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>non smoker</td>
<td>29</td>
<td>86</td>
</tr>
<tr>
<td>smoker</td>
<td>30</td>
<td>44</td>
</tr>
</tbody>
</table>

(c) Calculate and interpret the odds ratio for this table.
Non-smoker odds of lowest birth weight vs. not:
$$\frac{29}{186} = .157$$
Smoker odds of lower birth weight vs. not:
$$\frac{30}{44} = .682$$
Odds ratio:
$$\frac{30/44}{29/186} = 2.02$$
(d) Now create the table that groups the two lower birth weight categories vs. the two heavier birth weight categories.

<table>
<thead>
<tr>
<th></th>
<th>lower birth weight</th>
<th>heavier birth weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>non smoker</td>
<td>51</td>
<td>64</td>
</tr>
<tr>
<td>smoker</td>
<td>46</td>
<td>28</td>
</tr>
</tbody>
</table>

Calculate and interpret the odds ratio for this table.

\[
\frac{51/46}{64/28} = 2.06
\]

The odds of lower birth weight instead of heavier are twice as large for smokers.

(e) Finally create the table that three lower birth weight categories to the heaviest category.

<table>
<thead>
<tr>
<th></th>
<th>lower birth weights</th>
<th>heaviest birth weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>non smoker</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>smoker</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>

Calculate and interpret the odds ratio for this table.

\[
\frac{63/11}{80/35} = 2.51
\]

The odds of lower birth weight instead of heavier are about twice as large for smokers.

So for this approach we compare \(P(Y \leq k)\) to \(P(Y > k) = 1 - P(Y \leq k)\). \(P(Y \leq k)\) is the cumulative probability for \(Y, k = 1, \ldots, J\). So then we model the cumulative logits:

\[
\text{logit}[P(Y \leq k)] = \ln \frac{\sum_{j=k}^{J} P(Y = i)}{\sum_{j=1}^{k-1} P(Y = i)} = \frac{\pi_{k+1} + \cdots + \pi_{J}}{\pi_{k} + \cdots + \pi_{J}} = \beta_{0} + \beta_{1}x_{1} + \cdots
\]

**Definition:** The *proportional odds model* compares the odds of “cumulative probabilities” vs. their complements by assuming the odds ratio does not depend on the outcome category. That is, the effect of \(x_{i}\) is the same for all \(J - 1\) logits. Each logit defines a regular logistic regression curve for a binary response but shifted to the right or to the left.

The comparison at two different values of \(x\) will be of the form \(\beta(x - x^{*})\) so that the log odds ratio is proportional to the distance between the \(x\) values. In other words, the odds of making a response \(\leq j\) (rather than a “higher” response) at \(x = x_{1}\) are \(\exp(\beta(x_{1} - x_{2}))\) times the odds at \(x = x_{2}\) for any \(j\).

(f) Fit the ordinal logistic regression model using the *birth weight category* as the Response and *smoking status* as the predictor. (What is the “personality”?) What are the resulting 3 equations? What if we treat smoking as a (0,1) quantitative variable?

1. Low: \(-.736 + .380 \times \text{smoke} \quad e^{.380 \times 2} = 2.14\)
2. Below avg: \(.1327 + .380 \times \text{smoke} \quad e^{.380 \times 2} = 2.14\)
3. Above avg: \(1.247 + .380 \times \text{smoke} \quad e^{.380 \times 2} = 2.14\)

(g) Interpret the coefficient of the *smoke* variable.
Notice how a positive coefficient corresponds to being more likely to be in a lower birth weight category if the mother smokes.

(h) Does the proportional odds assumption seem reasonable for these data? How are you deciding?

(i) Is smoke a statistically significant predictor?

\[
\text{Wald: } 0.0053, \quad \text{Likelihood: } 0.0048
\]

Significance tests based on the ordinal model will be more powerful than tests that ignore ordering, partly due to the smaller degrees of freedom, when the model fits well. To test this constraint, we can compare this model to the nominal logistic regression model using a likelihood ratio test.

(j) Compare the log likelihood values for the two models using smoke as the predictor variable. What are the degrees of freedom for the chi-square test?

\[
-2 \left( \text{LL}_{\text{ordinal}} - \text{LL}_{\text{nominal}} \right) = 0.373
\]

\[\text{df} = 2, \quad p\text{-value} = .83\]

(not comparing full & reduced)

(k) Where does a similar test appear in the JMP output?

Note: SAS will perform a more appropriate “score test” for the proportional odds assumption (see p. 186-7). \( \text{df} = (J-1) \times p - 1 \).

(l) Now fit an ordinal logistic regression model associating the birth weight categories to mother’s weight. Interpret the slope coefficient, but consider a 10 lb increase in mother’s weight. Is this variable statistically significant?

\[
\text{For every 10 lb increase in mom’s wt, we predict } e^{\beta_{\text{wt}}} = .88 \text{-fold decrease}\]

in odds of being in lower birthweight categories rather than higher categories

To find the estimated probabilities, note that \( P(Y \leq k) = \frac{e^{\beta_0 + \beta_1 x_1 + ...}}{1 + e^{\beta_0 + \beta_1 x_1 + ...}} \)

(m) Estimate the probability of a birth weight of at most 3000g for a 182lb mother.
(n) In JMP, use the pull-down menu to select **Save > Save Probability Formula**. Where does the number you just calculated appear? What are the other *Cum[]* values? How do they compare? What would the *Cum[heaviest]* values be?

\[ \text{cum( below avg)} = \text{cum(heaviest)} = 1 \]

(o) If I know \( P(Y \leq k) \) and \( P(Y \leq k - 1) \), how do I find \( P(Y = k) \)?

\[ P(Y = k) = P(Y \leq k) - P(Y \leq k - 1) \]

(p) Using this relationship, find the estimated probability of below average birth weight (but not a low birth weight) when *mom's wt* = 182. Where does this value appear in the spreadsheet?

\[ .351 - .184 = .167 \]

\[ \uparrow \frac{1}{10} \text{(low)} \]

(q) What do the following graphs reveal?

Practice problem: Section 6.2.2 discusses the association between political party and political ideology. Fit the original logistic regression model in JMP and verify the odds ratio given in the text. Then, using the estimated probabilities from the model, create segmented bar graphs of the political ideologies for the two political parties. Interpret the results.

(I have created Table6.7.jmp and ordered the political ideology categories.)