CPH 931 — Fall 2009 — Dr. Charnigo

Lecture 5

Recalling logistic regression

Introduction. From CPH 930 you know that logistic regression is a standard technique for analyzing data when you have a dichotomous response variable. To briefly review, let Y denote the dichotomous response variable and $X_1, \ldots, X_k$ the explanatory variables. Typically $Y = 1$ represents an unfavorable outcome such as death or disease, while $Y = 0$ represents the complementary favorable outcome. We may assume that each explanatory variable is either continuous or dichotomous since a categorical explanatory variable with $m > 2$ categories can be expressed in terms of $(m - 1)$ dichotomous indicator variables.

The logistic regression model states that

$$\log \left[\frac{P(Y = 1|X_1 = x_1, \ldots, X_k = x_k)}{1 - P(Y = 1|X_1 = x_1, \ldots, X_k = x_k)} \right] = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k. \quad (1)$$

The fact that Y is dichotomous implies

$$P(Y = 0|X_1 = x_1, \ldots, X_k = x_k) = 1 - P(Y = 1|X_1 = x_1, \ldots, X_k = x_k),$$

so we can rewrite (1) as

$$\log \left[\frac{P(Y = 1|X_1 = x_1, \ldots, X_k = x_k)}{P(Y = 0|X_1 = x_1, \ldots, X_k = x_k)} \right] = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k. \quad (2)$$

Although recasting (1) as (2) may at first seem an unproductive exercise, we will see shortly that expression (2) helps us to relate logistic regression to the first major topic of today's lecture: polytomous regression.

Polytomous regression

Statistical framework. Suppose that Y is a categorical response variable but that there are more than two categories. For ease of presentation let us specify the number of categories as three and the possible values of Y as 0, 1, and 2. What follows can be straightforwardly adapted if there are more than three categories. Also, at this point we neither assume nor rule out the possibility that $Y = 2$ is "worse" than $Y = 1$ and $Y = 1$ is "worse" than $Y = 0$. For now, we simply regard $Y = 2$ as "different" from $Y = 1$, and we regard both $Y = 2$ and $Y = 1$ as "different" from $Y = 0$.

For $j = 0, 1, 2$ let $p_{j|\mathbf{x}}$ be shorthand for $P(Y = j|X_1 = x_1, \ldots, X_k = x_k)$. If we wish, we can still assert — as in expression (2) — that

$$\log\left[\frac{p_{1|\mathbf{x}}}{p_{0|\mathbf{x}}}\right] = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k. \tag{3}$$

However, there are two points to consider.

First, unless $p_{2|\mathbf{x}} = 0$, we cannot interpret $p_{1|\mathbf{x}}/p_{0|\mathbf{x}}$ as odds. For that we would need $1 - p_{1|\mathbf{x}}$ in the denominator instead of $p_{0|\mathbf{x}}$. Thus, some people refer to $p_{1|\mathbf{x}}/p_{0|\mathbf{x}}$ as an "odds-like" quantity, although I prefer to call it a "ratio of probabilities" or a "relative probability".

Second, expression (3) is an incomplete prescription for a statistical model since $p_{2|\mathbf{x}}$ is not determined by the ratio $p_{1|\mathbf{x}}/p_{0|\mathbf{x}}$. For instance, $p_{1|\mathbf{x}}/p_{0|\mathbf{x}}$ equals 1 if $p_{1|\mathbf{x}} = p_{0|\mathbf{x}} = 0.4$ (in which case $p_{2|\mathbf{x}} = 0.2$) but also equals 1 if $p_{1|\mathbf{x}} = p_{0|\mathbf{x}} = 0.1$ (in which case $p_{2|\mathbf{x}} = 0.8$).

However, if we assert that

$$\log\left[\frac{p_{2|\mathbf{x}}}{p_{0|\mathbf{x}}}\right] = \alpha^* + \beta_1^* x_1 + \cdots + \beta_k^* x_k, \tag{4}$$

then (3) and (4) together constitute a complete prescription for a statistical model. We refer to such a model as a polytomous regression model.

Example. The data set {BMIData.xls} contains information about caloric intake, fat intake, cholesterol intake, fiber intake, alcohol intake, and body mass index for 51 subjects. I have also created a variable "WEIGHT1" that equals 0, 1, or 2 according to the standard body mass index classifications: WEIGHT1 = 0 for people with BMI less than 25 (considered to be of healthy weight), WEIGHT1 = 1 for people with BMI between 25 and 30 (considered to be overweight), and WEIGHT1 = 2 for people with BMI greater than 30 (considered to be obese).

Suppose that we want to assess whether some of the intake variables predict body mass index classification. The usual approaches to multivariate data analysis from CPH 930 are not applicable. We cannot use ordinary logistic regression since WEIGHT1 has three categories instead of two. We cannot use linear regression since, with only three categories, WEIGHT1 is not discretized finely enough for us to treat it as approximately continuous. However, the method of polytomous regression may be applied. We will continue with this example later.

Interpreting coefficients in polytomous regression. Consider the polytomous regression model given by (3) and (4). The intercept coefficients are uninterpretable unless $X_1 = \cdots = X_k = 0$ can occur, which is not generally the case. However, the partial slope coefficients are amenable to interpretation.

The partial slope coefficient β_1 in (3) is the log ratio of relative probabilities corresponding to a one-unit increase in X_1 when $X_2, \ldots, X_k$ are fixed, where relative probability refers to the probability of $Y = 1$ divided by the probability of $Y = 0$. Or, if you prefer, $\exp[\beta_1]$ is the ratio of relative probabilities corresponding to a one-unit increase in X_1 when $X_2, \ldots, X_k$

are fixed. Symbolically,

$$\exp[\beta_1] = \frac{p_{1|\mathbf{x}_{new}}/p_{0|\mathbf{x}_{new}}}{p_{1|\mathbf{x}_{old}}/p_{0|\mathbf{x}_{old}}},$$

where the interpretation of $\mathbf{x}_{old}$ (resp., $\mathbf{x}_{new}$) is that the one-unit increase in X_1 has not yet taken place (resp., has taken place).

The partial slope coefficient β_1^* in (4) is the log ratio of relative probabilities corresponding to a one-unit increase in X_1 when $X_2, \ldots, X_k$ are fixed, where relative probability now refers to the probability of $Y = 2$ divided by the probability of $Y = 0$. Or, if you prefer, $\exp[\beta_1^*]$ is the ratio of relative probabilities corresponding to a one-unit increase in X_1 when $X_2, \ldots, X_k$ are fixed. That is,

$$\exp[\beta_1^*] = \frac{p_{2|\mathbf{x}_{new}}/p_{0|\mathbf{x}_{new}}}{p_{2|\mathbf{x}_{old}}/p_{0|\mathbf{x}_{old}}}.$$

Note that $Y = 0$ serves as a reference category: β_1 (resp., β_1^*) describes how the relative probability of $Y = 1$ (resp., $Y = 2$) with respect to $Y = 0$ changes with X_1. To see how the relative probability of $Y = 2$ with respect to $Y = 1$ changes with X_1, we must consider $\beta_1^* - \beta_1$. Specifically,

$$\exp[\beta_1^* - \beta_1] = \frac{\exp[\beta_1^*]}{\exp[\beta_1]} = \frac{p_{2|\mathbf{x}_{new}}/p_{1|\mathbf{x}_{new}}}{p_{2|\mathbf{x}_{old}}/p_{1|\mathbf{x}_{old}}}.$$

Of course β_j and β_j^* $(2 \leq j \leq k)$ are interpreted analogously but pertain to one-unit increases in X_j when $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k$ are fixed.

Inferences in polytomous regression. As in ordinary logistic regression, we estimate parameters by maximum likelihood and use Wald tests to assess hypotheses on individual partial slope coefficients.

The Wald test of $H_0 : \beta_1 = 0$ is performed by calculating

$$\chi^2 := \left[\hat{\beta}_1/se(\hat{\beta}_1)\right]^2,$$

where $\hat{\beta}_1$ is the maximum likelihood estimate of β_1 and $se(\hat{\beta}_1)$ is its standard error. If $\chi^2 > \chi^2_{1,1-\alpha}$, then H_0 is rejected at significance level α. A 95% confidence interval for β_1 is $\hat{\beta}_1 \pm 1.96 se(\hat{\beta}_1)$. Wald tests and confidence intervals for $\beta_1^*, \beta_2, \ldots, \beta_k^*$ are analogous.

In principle we can also perform a likelihood ratio test on multiple partial slope coefficients. For any $m \in \{1, \ldots, k\}$ the likelihood ratio test of $H_0 : \beta_1 = \beta_1^* = \cdots = \beta_m = \beta_m^* = 0$ entails fitting a "full" model with $X_1, \ldots, X_k$ and a "reduced" model with $X_{m+1}, \ldots, X_k$. Negative twice the log likelihood from the full model, denoted $-2 \log L$, is subtracted from negative twice the log likelihood from the reduced model, denoted $-2 \log L_{H_0}$. If the difference exceeds $\chi^2_{2m,1-\alpha}$, then H_0 is rejected at significance level α.

Unfortunately, PROC CATMOD in SAS does not provide the output necessary to perform a likelihood ratio test, except in the special case that every subject has a distinct combination of values on $X_{m+1}, \ldots, X_k$. However, PROC CATMOD does provide results for "multiparameter Wald tests" of $H_0 : \beta_1 = \beta_1^* = 0$, $H_0 : \beta_2 = \beta_2^* = 0$, and so forth. The null hypotheses for the multiparameter Wald tests state that X_1 is not needed in the model (given that $X_2, \ldots, X_k$ are in the model), that X_2 is not needed in the model, and so forth.

Example, continued. On page 4 of {BMIresults.rtf} are shown results for a polytomous regression model with WEIGHT1 as the response variable and caloric intake, fat intake as explanatory variables. For convenience let Y denote WEIGHT1 and X_1, X_2 denote caloric intake, fat intake. Caloric intake has been rescaled by a factor of 1/100 prior to fitting the polytomous regression model. That is, a one-unit increase in X_1 really corresponds to a 100-calorie increase.

The coefficients governing the relative probability of $Y = 1$ with respect to $Y = 0$ are estimated as $\hat{\alpha} = -4.6844$, $\hat{\beta}_1 = 0.0138$, and $\hat{\beta}_2 = 0.0455$ with respective standard errors 1.8427, 0.0814, and 0.0262.

With a 100-calorie increase but no change in fat intake, the relative probability of being overweight with respect to being at healthy weight is estimated to change by a multiplicative factor of $\exp[\hat{\beta}_1] = \exp[0.0138] = 1.014$. However, since $H_0 : \beta_1 = 0$ cannot be rejected (Wald test $\chi^2 = 0.03$, p-value $= 0.8650$), the factor of 1.014 is not significantly different from the 1 that would represent no change in the relative probability.

The coefficients governing the relative probability of $Y = 2$ with respect to $Y = 0$ are estimated as $\hat{\alpha}^* = -13.2674$, $\hat{\beta}_1^* = 0.1071$, and $\hat{\beta}_2^* = 0.0889$ with respective standard errors 4.1636, 0.1313, and 0.0410.

The multiparameter Wald test of $H_0 : \beta_1 = \beta_1^* = 0$ yields a p-value of 0.7063, while the multiparameter Wald test of $H_0 : \beta_2 = \beta_2^* = 0$ yields a p-value of 0.0672. Although PROC CATMOD in SAS does not have automated variable selection capabilities, we could perform backward elimination "manually" by discarding X_1 (since it generated the largest multiparameter Wald p-value), refitting the model with X_2 only, and deciding whether to keep X_2 according to its multiparameter Wald p-value in the refitted model.

Ordinal logistic regression

Statistical framework. Again suppose that Y is a categorical response variable with more than two categories. For ease of presentation we once more specify the number of categories as three and the possible values of Y as 0, 1, and 2. What follows can be straightforwardly adapted if there are more than three categories.

In polytomous regression we neither assumed nor ruled out the possibility that $Y = 2$ was "worse" than $Y = 1$ and $Y = 1$ was "worse" than $Y = 0$. If in fact $Y = 2$ is "worse" than $Y = 1$, which in turn is "worse" than $Y = 0$, then we refer to Y as an *ordinal* variable. Otherwise we refer to Y as a *nominal* variable. Polytomous regression is a valid method in either case. However, a simpler method is available if Y is ordinal.

The odds of having $Y = 2$ when $X_1 = x_1, \ldots, X_k = x_k$ are defined as

$$\frac{P(Y = 2|X_1 = x_1, \ldots, X_k = x_k)}{P(Y \leq 1|X_1 = x_1, \ldots, X_k = x_k)}. \tag{5}$$

Similarly, the odds of having $Y \geq 1$ when $X_1 = x_1, \ldots, X_k = x_k$ are defined as

$$\frac{P(Y \geq 1|X_1 = x_1, \ldots, X_k = x_k)}{P(Y = 0|X_1 = x_1, \ldots, X_k = x_k)}. \tag{6}$$

While we could define the odds of having $Y = 1$ rather than of having $Y \geq 1$, this would be awkward with an ordinal Y since the event complementary to $Y = 1$ is neither clearly "worse" than $Y = 1$ (because $Y = 0$ is included) nor clearly "better" (because $Y = 2$ is included).

Let us take $O_{2|\mathbf{x}}$ as shorthand for (5) and $O_{12|\mathbf{x}}$ as shorthand for (6). If we wanted, we could fit two separate ordinary logistic regression models,

$$\log O_{2|\mathbf{x}} = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k \tag{7}$$

and

$$\log O_{12|\mathbf{x}} = \alpha^* + \beta_1^* x_1 + \cdots + \beta_k^* x_k. \tag{8}$$

However, in many applications there is little reason to believe that β_j should differ much from β_j^* $(1 \leq j \leq k)$. If we impose the restrictions that $\beta_1 = \beta_1^*$, $\beta_2 = \beta_2^*$, and so forth (collectively referred to as "the proportional odds assumption"), then (8) becomes

$$\log O_{12|\mathbf{x}} = \alpha^* + \beta_1 x_1 + \cdots + \beta_k x_k. \tag{9}$$

We refer to (7) and (9) together as an ordinal logistic regression model or, more specifically, a proportional odds model.

Interpreting coefficients in ordinal logistic regression. Consider the ordinal logistic regression model given by (7) and (9). The intercept coefficients are uninterpretable unless $X_1 = \cdots = X_k = 0$ can occur, which is not generally the case. However, the partial slope coefficients are amenable to interpretation.

The partial slope coefficient β_1 is the log odds ratio corresponding to a one-unit increase in X_1 when $X_2, \ldots, X_k$ are fixed, where "odds ratio" can pertain either to the event that $Y = 2$ or to the event that $Y \geq 1$. Or, if you prefer, $\exp[\beta_1]$ is the odds ratio corresponding to a one-unit increase in X_1 when $X_2, \ldots, X_k$ are fixed. Symbolically,

$$\exp[\beta_1] = \frac{O_{2|\mathbf{x}_{new}}}{O_{2|\mathbf{x}_{old}}} = \frac{O_{12|\mathbf{x}_{new}}}{O_{12|\mathbf{x}_{old}}}.$$

Thus, the interpretations of partial slope coefficients in an ordinal logistic regression model are identical to those in an ordinary logistic regression model, except that any mention of "odds ratio" really applies simultaneously to two odds ratios, one pertaining to the event that $Y = 2$ and one pertaining to the event that $Y \geq 1$.

Inferences in ordinal logistic regression. As in ordinary logistic regression, we estimate parameters by maximum likelihood and use Wald tests to assess hypotheses on individual partial slope coefficients. In addition, PROC LOGISTIC in SAS does provide the output necessary to perform a likelihood ratio test on multiple partial slope coefficients. For completeness of this lecture I describe both testing procedures below.

The Wald test of $H_0 : \beta_1 = 0$ is performed by calculating

$$\chi^2 := \left[\hat{\beta}_1 / se(\hat{\beta}_1)\right]^2,$$

where $\hat{\beta}_1$ is the maximum likelihood estimate of β_1 and $se(\hat{\beta}_1)$ is its standard error. If $\chi^2 > \chi^2_{1,1-\alpha}$, then H_0 is rejected at significance level α. A 95% confidence interval for β_1 is $\hat{\beta}_1 \pm 1.96 se(\hat{\beta}_1)$.

For any $m \in \{1, \ldots, k\}$ the likelihood ratio test of $H_0 : \beta_1 = \cdots = \beta_m = 0$ entails fitting a "full" model with $X_1, \ldots, X_k$ and a "reduced" model with $X_{m+1}, \ldots, X_k$. Negative twice the log likelihood from the full model, denoted $-2 \log L$, is subtracted from negative twice the log likelihood from the reduced model, denoted $-2 \log L_{H_0}$. If the difference exceeds $\chi^2_{m,1-\alpha}$, then H_0 is rejected at significance level α.

Example, continued. Refer to {BMIresults2.rtf}, in which are shown results for a proportional odds model with WEIGHT1 as the response variable and caloric intake, fat intake as explanatory variables.

On page 1 the box labeled "Score Test for the Proportional Odds Assumption" provides the result for a test of the null hypothesis that (7) and (8) can be replaced by (7) and (9). We are looking for a *nonsignificant* result here; a significant result indicates a violation of the proportional odds assumption and, hence, reason to doubt the validity of the remaining results.

On page 1 the "Intercept and Covariates" column in the "Model Fit Statistics" box lists values of the AIC, the BIC, and negative twice the log likelihood for the ordinal logistic regression model with $X_1, \ldots, X_k$. In the "Intercept" column are values of the AIC, the BIC, and negative twice the log likelihood for a "null model" without $X_1, \ldots, X_k$.

On page 2 the box "Testing Global Null Hypothesis" presents the results of a likelihood ratio test (and of two other tests) for the "global null hypothesis" that all partial slope coefficients are zero. We see that χ^2 for the likelihood ratio test is 18.90 (which could also be deduced from the output on page 1 by subtraction of 82.07 from 100.97) with p-value less than 0.0001. We reject the global null hypothesis and conclude that caloric intake, fat intake are collectively useful in predicting body mass index classification.

On page 2 the box "Analysis of Maximum Likelihood Estimates" reveals that $\hat{\alpha} = -8.4711$, $\hat{\alpha}^* = -5.9528$, $\hat{\beta}_1 = 0.0390$, and $\hat{\beta}_2 = 0.0543$. If fat intake increases by one unit while caloric intake is unchanged, then the odds of being obese are multiplied by an estimated $\exp[\hat{\beta}_2] = 1.056$. However, the 1.056 is simultaneously the estimated multiplier for the odds of being overweight or obese.

We reject $H_0 : \beta_2 = 0$ at significance level 0.05 since the Wald test yields $\chi^2 = 5.7575$ and a p-value of 0.0164. We could calculate a 95% confidence interval for β_2 as $0.0543 \pm 1.96(0.0226)$, and exponentiating this would produce a 95% confidence interval for the odds ratio $\exp[\beta_2]$. The latter confidence interval is already reported in the SAS output as $[1.010, 1.104]$.

Discussion questions

1. Consider page 8 of {BMIresults.rtf}, on which are shown results for a polytomous regression model with WEIGHT1 as the response variable and caloric intake as the sole explanatory variable. Suppose that an individual typically consumes 3000 calories. Estimate the relative probability of being overweight versus of being at healthy weight for such an individual.

2. For an individual who typically consumes 3000 calories, estimate the relative probability of being obese versus of being at healthy weight.

3. For an individual who typically consumes 3000 calories, estimate the probabilities (not relative probabilities!) of being at healthy weight, being overweight, and being obese.