12 Polytomous Logistic Regression

Contents

D.G. Kleinbaum and M. Klein, *Logistic Regression*, Statistics for Biology and Health,
DOI 10.1007/978-1-4419-1742-3_12,

Introduction

In this chapter, the standard logistic model is extended to handle outcome variables that have more than two categories. Polytomous logistic regression is used when the categories of the outcome variable are nominal, that is, they do not have any natural order. When the categories of the outcome variable do have a natural order, ordinal logistic regression may also be appropriate.

The focus of this chapter is on polytomous logistic regression. The mathematical form of the polytomous model and its interpretation are developed. The formulas for the odds ratio and confidence intervals are derived, and techniques for testing hypotheses and assessing the statistical significance of independent variables are shown.

Abbreviated Outline

The outline below gives the user a preview of the material to be covered by the presentation. A detailed outline for review purposes follows the presentation.

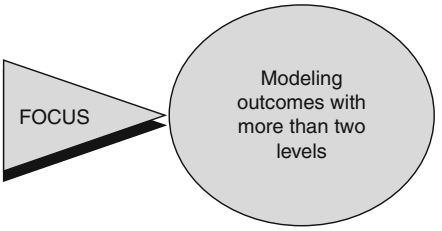
Objectives

Upon completing this chapter, the learner should be able to:

1. State or recognize the difference between nominal and ordinal variables.
2. State or recognize when the use of polytomous logistic regression may be appropriate.
3. State or recognize the polytomous regression model.
4. Given a printout of the results of a polytomous logistic regression:
   a. State the formula and compute the odds ratio
   b. State the formula and compute a confidence interval for the odds ratio
   c. Test hypotheses about the model parameters using the likelihood ratio test or the Wald test, stating the null hypothesis and the distribution of the test statistic with the corresponding degrees of freedom under the null hypothesis
5. Recognize how running a polytomous logistic regression differs from running multiple standard logistic regressions.

Presentation

I. Overview



This presentation and the presentation that follows describe approaches for extending the standard logistic regression model to accommodate a disease, or outcome, variable that has more than two categories. Up to this point, our focus has been on models that involve a dichotomous outcome variable, such as disease present/absent. However, there may be situations in which the investigator has collected data on multiple levels of a single outcome. We describe the *form* and key *characteristics* of one model for such multilevel outcome variables: the polytomous logistic regression model.

Examples of multilevel outcomes:

1. Absent, mild, moderate, severe
2. In situ, locally invasive, metastatic
3. Choice of treatment regimen

Examples of outcome variables with more than two levels might include (1) disease symptoms that have been classified by subjects as being absent, mild, moderate, or severe, (2) invasiveness of a tumor classified as in situ, locally invasive, or metastatic, or (3) patients' preferred treatment regimen, selected from among three or more options.

One approach: dichotomize outcome

Change

0	1	2

to

0	1	2

One possible approach to the analysis of data with a polytomous outcome would be to choose an appropriate cut-point, dichotomize the multilevel outcome variable, and then simply utilize the logistic modeling techniques discussed in previous chapters.

EXAMPLE

Change

None	Mild	Moderate	Severe

to

None or mild	Moderate or severe

For example, if the outcome symptom severity has four categories of severity, one might compare subjects with none or only mild symptoms to those with either moderate or severe symptoms.

Disadvantage of dichotomizing: Loss of detail (e.g., mild vs. none? moderate vs. mild?)

The disadvantage of dichotomizing a polytomous outcome is loss of detail in describing the outcome of interest. For example, in the scenario given above, we can no longer compare mild vs. none or moderate vs. mild. This loss of detail may, in turn, affect the conclusions made about the exposure–disease relationship.

Alternate approach: Use model for a polytomous outcome

Nominal or ordinal outcome?

The detail of the original data coding can be retained through the use of models developed specifically for polytomous outcomes. The specific form that the model takes depends, in part, on whether the multilevel outcome variable is measured on a nominal or an ordinal scale.

Nominal: Different categories; no ordering

EXAMPLE

Endometrial cancer subtypes:

- Adenosquamous
- Adenocarcinoma
- Other

Nominal variables simply indicate different categories. An example is histological subtypes of cancer. For endometrial cancer, three possible subtypes are adenosquamous, adenocarcinoma, and other.

Ordinal: Levels have natural ordering

EXAMPLE

Tumor grade:

- Well differentiated
- Moderately differentiated
- Poorly differentiated

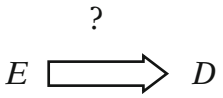
Ordinal variables have a natural ordering among the levels. An example is cancer tumor grade, ranging from well differentiated to moderately differentiated to poorly differentiated tumors.

Nominal outcome $\Rightarrow$ Polytomous model

Ordinal outcome $\Rightarrow$ Ordinal model or polytomous model

An outcome variable that has three or more nominal categories can be modeled using polytomous logistic regression. An outcome variable with three or more ordered categories can also be modeled using polytomous regression, but can also be modeled with ordinal logistic regression, provided that certain assumptions are met. Ordinal logistic regression is discussed in detail in Chap. 13.

II. Polytomous Logistic Regression: An Example with Three Categories



When modeling a multilevel outcome variable, the epidemiological question remains the same: What is the relationship of one or more exposure or study variables (E) to a disease or illness outcome (D)?

EXAMPLE

Simplest case of polytomous model:

- Outcome with three categories
- One dichotomous exposure variable

In this section, we present an example of a polytomous logistic regression model with one dichotomous exposure variable and an outcome (D) that has three categories. This is the simplest case of a polytomous model. Later in the presentation, we discuss extending the polytomous model to more than one predictor variable and then to outcomes with more than three categories.

Data source:
Black/White Cancer Survival Study

$$E = \text{AGEGP}\begin{cases} 0 & \text{if 50–64} \\ 1 & \text{if 65–79} \end{cases}$$

$$D = \text{SUBTYPE}\begin{cases} 0 & \text{if Adenocarcinoma} \\ 1 & \text{if Adenosquamous} \\ 2 & \text{if Other} \end{cases}$$

The example uses data from the National Cancer Institute's Black/White Cancer Survival Study (Hill et al., 1995). Suppose we are interested in assessing the effect of age group on histological subtype among women with primary endometrial cancer. AGEGP, the exposure variable, is coded as 0 for aged 50–64 or 1 for aged 65–79. The disease variable, histological subtype, is coded 0 for adenocarcinoma, 1 for adenosquamous, and 2 for other.

There is no inherent order in the outcome variable. The 0, 1, and 2 coding of the disease categories is arbitrary.

SUBTYPE (0, 1, 2) uses arbitrary coding.

The 3×2 table of the data is presented on the left.

	AGEGP 50–64 $E = 0$	AGEGP 65–79 $E = 1$
Adenocarcinoma $D = 0$	77	109
Adenosquamous $D = 1$	11	34
Other $D = 2$	18	39

Outcome categories:

A B C D

Reference (arbitrary choice)

Then compare:

A vs. C, B vs. C, and D vs. C

EXAMPLE (*continued*)

Reference group = Adenocarcinoma

Two comparisons:

1. Adenosquamous ($D = 1$) vs. Adenocarcinoma ($D = 0$)
2. Other ($D = 2$) vs. Adenocarcinoma ($D = 0$)

Using data from table:

$$\widehat{\text{OR}}_{1 \text{ vs. } 0} = \frac{77 \times 34}{109 \times 11} = 2.18$$

$$\widehat{\text{OR}}_{2 \text{ vs. } 0} = \frac{77 \times 39}{109 \times 18} = 1.53$$

With polytomous logistic regression, one of the categories of the outcome variable is designated as the reference category and each of the other levels is compared with this reference. The choice of reference category can be arbitrary and is at the discretion of the researcher. See example at left. Changing the reference category does not change the form of the model, but it does change the interpretation of the parameter estimates in the model.

In our three-outcome example, the Adenocarcinoma group has been designated as the reference category. We are therefore interested in modeling two main comparisons. We want to compare subjects with an Adenosquamous outcome (category 1) to those subjects with an Adenocarcinoma outcome (category 0) and we also want to compare subjects with an Other outcome (category 2) to those subjects with an Adenocarcinoma outcome (category 0).

If we consider these two comparisons separately, the crude odds ratios can be calculated using data from the preceding table. The crude odds ratio comparing Adenosquamous (category 1) to Adenocarcinoma (category 0) is the product of 77 and 34 divided by the product of 109 and 11, which equals 2.18. Similarly, the crude odds ratio comparing Other (category 2) to Adenocarcinoma (category 0) is the product of 77 and 39 divided by the product of 109 and 18, which equals 1.53.

Dichotomous vs. polytomous model: Odds vs. "odds-like" expressions

$$\text{logit } \text{P}(\mathbf{X}) = \ln\left[\frac{\text{P}(D = 1 \mid \mathbf{X})}{\text{P}(D = 0 \mid \mathbf{X})}\right] = \alpha + \sum_{i=1}^{k} \beta_i X_i$$

Recall that for a dichotomous outcome variable coded as 0 or 1, the logit form of the logistic model, logit P($\mathbf{X}$), is defined as the natural log of the odds for developing a disease for a person with a set of independent variables specified by $\mathbf{X}$. This logit form can be written as the linear function shown on the left.

Odds of disease: a ratio of probabilities

The odds for developing disease can be viewed as a ratio of probabilities. For a dichotomous outcome variable coded 0 and 1, the odds of disease equal the probability that disease equals 1 divided by 1 minus the probability that disease equals 1, or the probability that disease equals 1 divided by the probability that disease equals 0.

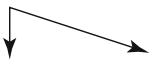
Dichotomous outcome:

$$\text{odds} = \frac{\text{P}(D=1)}{1-\text{P}(D=1)} = \frac{\text{P}(D=1)}{\text{P}(D=0)}$$

Polytomous outcome (three categories):

For polytomous logistic regression with a three-level variable coded 0, 1, and 2, there are two analogous expressions, one for each of the two comparisons we are making. These expressions are also in the form of a ratio of probabilities.

Use "odds-like" expressions for two comparisons



(1) $\frac{\text{P}(D=1)}{\text{P}(D=0)}$ (2) $\frac{\text{P}(D=2)}{\text{P}(D=0)}$

In polytomous logistic regression with three levels, we therefore define our model using two expressions for the natural log of these "odds-like" quantities. The first is the natural log of the probability that the outcome is in category 1 divided by the probability that the outcome is in category 0; the second is the natural log of the probability that the outcome is in category 2 divided by the probability that the outcome is in category 0.

The logit form of model uses ln of "odds-like" expressions

(1) $\ln\left[\frac{\text{P}(D=1)}{\text{P}(D=0)}\right]$ (2) $\ln\left[\frac{\text{P}(D=2)}{\text{P}(D=0)}\right]$

$$\text{P}(D=0)+\text{P}(D=1)+\text{P}(D=2)=1$$

BUT

$$\text{P}(D=1)+\text{P}(D=0)\neq 1$$
$$\text{P}(D=2)+\text{P}(D=0)\neq 1$$

When there are three categories of the outcome, the sum of the probabilities for the three outcome categories must be equal to 1, the total probability. Because each comparison considers only two probabilities, the probabilities in the ratio do not sum to 1. Thus, the two "odds-like" expressions are not true odds. However, if we restrict our interest to just the two categories being considered in a given ratio, we may still conceptualize the expression as an odds. In other words, each expression is an odds *only* if we condition on the outcome being in one of the two categories of interest. For ease of the subsequent discussion, we will use the term "odds" rather than "odds-like" for these expressions.

Therefore:

$$\frac{\text{P}(D=1)}{\text{P}(D=0)} \text{ and } \frac{\text{P}(D=2)}{\text{P}(D=0)}$$

"odds-like" but not true odds (unless analysis restricted to two categories)

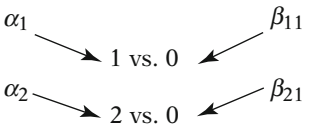
Model for three categories, one predictor (X_1 = AGEGP):

$$\ln\left[\frac{P(D=1\,|\,X_1)}{P(D=0\,|\,X_1)}\right] = \alpha_1 + \beta_{11}X_1$$

Because our example has three outcome categories and one predictor (i.e., AGEGP), our polytomous model requires two regression expressions. One expression gives the log of the probability that the outcome is in category 1 divided by the probability that the outcome is in category 0, which equals α_1 plus β_{11} times X_1.

$$\ln\left[\frac{P(D=2\,|\,X_1)}{P(D=0\,|\,X_1)}\right] = \alpha_2 + \beta_{21}X_1$$

We are also *simultaneously* modeling the log of the probability that the outcome is in category 2 divided by the probability that the outcome is in category 0, which equals α_2 plus β_{21} times X_1.



Both the alpha and beta terms have a subscript to indicate which comparison is being made (i.e., category **1** vs. 0 or category **2** vs. 0).

III. Odds Ratio with Three Categories

$\hat{\alpha}_1$ $\hat{\alpha}_2$ } Estimates obtained
$\hat{\beta}_{11}$ $\hat{\beta}_{21}$ } as in SLR

Once a polytomous logistic regression model has been fit and the parameters (intercepts and beta coefficients) have been estimated, we can then calculate estimates of the disease–exposure association in a similar manner to the methods used in standard logistic regression (SLR).

Special case for one predictor
where $X_1 = 1$ or $X_1 = 0$

Consider the special case in which the only independent variable is the exposure variable and the exposure is coded 0 and 1. To assess the effect of the exposure on the outcome, we compare $X_1 = 1$ to $X_1 = 0$.

Two odds ratios:

OR_1 (category 1 vs. category 0) (Adenosquamous vs. Adenocarcinoma)

OR_2 (category 2 vs. category 0) (Other vs. Adenocarcinoma)

We need to calculate two odds ratios, one that compares category 1 (Adenosquamous) to category 0 (Adenocarcinoma) and one that compares category 2 (Other) to category 0 (Adenocarcinoma).

Recall that we are actually calculating a ratio of two "odds-like" expressions. However, we continue the conventional use of the term odds ratio for our discussion.

$$OR_1 = \frac{[P(D=1|X=1)/P(D=0|X=1)]}{[P(D=1|X=0)/P(D=0|X=0)]}$$

$$OR_2 = \frac{[P(D=2|X=1)/P(D=0|X=1)]}{[P(D=2|X=0)/P(D=0|X=0)]}$$

Each odds ratio is calculated in a manner similar to that used in standard logistic regression. The two OR formulas are shown on the left.

Adenosquamous vs. Adenocarcinoma:

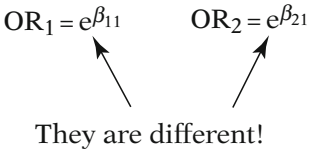
$$OR_1 = \frac{\exp[\alpha_1 + \beta_{11}(1)]}{\exp[\alpha_1 + \beta_{11}(0)]} = e^{\beta_{11}}$$

Other vs. Adenocarcinoma:

$$OR_2 = \frac{\exp[\alpha_2 + \beta_{21}(1)]}{\exp[\alpha_2 + \beta_{21}(0)]} = e^{\beta_{21}}$$

Using our previously defined probabilities of the log odds, we substitute the two values of X_1 for the exposure (i.e., 0 and 1) into those expressions. After dividing, we see that the odds ratio for the first comparison (Adenosquamous vs. Adenocarcinoma) is e to the β_{11}.

The odds ratio for the second comparison (Other vs. Adenocarcinoma) is e to the β_{21}.



We obtain two different odds ratio expressions, one utilizing β_{11} and the other utilizing β_{21}. Thus, quantifying the association between the exposure and outcome depends on which levels of the outcome are being compared.

General case for one predictor

$$\mathrm{OR}_g = \exp\left[\beta_{g1}\left(X_1^{**} - X_1^{*}\right)\right], \text{ where } g = 1, 2$$

The special case of a dichotomous predictor can be generalized to include categorical or continuous predictors. To compare any two levels ($X_1 = X_1^{**}$ vs. $X_1 = X_1^{*}$) of a predictor, the odds ratio formula is e to the β_{g1} times $(X_1^{**} - X_1^{*})$, where g defines the category of the disease variable (1 or 2) being compared with the reference category (0).

Computer output for polytomous model:

Is output listed in ascending or descending order?

The output generated by a computer package for polytomous logistic regression includes alphas and betas for the log odds terms being modeled. Packages vary in the presentation of output, and the coding of the variables must be considered to correctly read and interpret the computer output for a given package. For example, in SAS, if $D = 0$ is designated as the reference category, the output is listed in descending order (see Appendix). This means that the listing of parameters pertaining to the comparison with category $D = 2$ precedes the listing of parameters pertaining to the comparison with category $D = 1$, as shown on the left.

EXAMPLE

SAS

Reference category: $D = 0$
Parameters for $D = 2$ comparison precede $D = 1$ comparison.

Variable	Estimate symbol
Intercept 1	$\hat{\alpha}_2$
Intercept 2	$\hat{\alpha}_1$
X_1	$\hat{\beta}_{21}$
X_1	$\hat{\beta}_{11}$

EXAMPLE

Variable	Estimate	S.E.	Symbol
Intercept 1	−1.4534	0.2618	$\hat{\alpha}_2$
Intercept 2	−1.9459	0.3223	$\hat{\alpha}_1$
AGEGP	0.4256	0.3215	$\hat{\beta}_{21}$
AGEGP	0.7809	0.3775	$\hat{\beta}_{11}$

The results for the polytomous model examining histological subtype and age are presented on the left. The results were obtained from running PROC LOGISTIC in SAS. See the Computer Appendix for computer coding.

There are two sets of parameter estimates. The output is listed in descending order, with α_2 labeled as Intercept 1 and α_1 labeled as intercept 2. If $D = 2$ had been designated as the reference category, the output would have been in ascending order.

EXAMPLE (*continued*)

Other vs. Adenocarcinoma:

$$\ln\left[\frac{\hat{P}(D=2\,|\,X_1)}{\hat{P}(D=0\,|\,X_1)}\right] = -1.4534 + (0.4256)\text{AGEGP}$$

$$\widehat{\text{OR}}_2 = \exp[\hat{\beta}_{21}] = \exp(0.4256) = 1.53$$

Adenosquamous vs. Adenocarcinoma:

$$\ln\left[\frac{\hat{P}(D=1\,|\,X_1)}{\hat{P}(D=0\,|\,X_1)}\right] = -1.9459 + (0.7809)\text{AGEGP}$$

$$\text{OR}_1 = \exp[\hat{\beta}_{11}] = \exp(0.7809) = 2.18$$

Special case

One dichotomous exposure ⇒ polytomous model ORs = crude ORs

Interpretation of ORs

For older vs. younger subjects:

- Other tumor category more likely than Adenocarcinoma ($\widehat{\text{OR}}_2 = 1.53$)
- Adenosquamous even more likely than Adenocarcinoma ($\widehat{\text{OR}}_1 = 2.18$)

The equation for the estimated log odds of Other (category 2) vs. Adenocarcinoma (category 0) is negative 1.4534 plus 0.4256 times age group.

Exponentiating the beta estimate for age in this model yields an estimated odds ratio of 1.53.

The equation for the estimated log odds of Adenosquamous (category 1) vs. Adenocarcinoma (category 0) is negative 1.9459 plus 0.7809 times age group.

Exponentiating the beta estimate for AGEGP in this model yields an estimated odds ratio of 2.18.

The odds ratios from the polytomous model (i.e., 1.53 and 2.18) are the same as those we obtained earlier when calculating the crude odds ratios from the data table before modeling. In the special case, where there is one *dichotomous* exposure variable, the crude estimate of the odds ratio will match the estimate of the odds ratio obtained from a polytomous model (or from a standard logistic regression model).

We can interpret the odds ratios by saying that, for women diagnosed with primary endometrial cancer, older subjects (aged 65–79) relative to younger subjects (aged 50–64) were more likely to have their tumors categorized as Other than as Adenocarcinoma ($\widehat{\text{OR}}_2 = 1.53$) and were even more likely to have their tumors classified as Adenosquamous than as Adenocarcinoma ($\widehat{\text{OR}}_1 = 2.18$).

Interpretation of alphas

Log odds where all Xs set to 0.

Not informative if sampling done by outcome (i.e., "disease") status.

What is the interpretation of the alpha coefficients? They represent the log of the odds where all independent variables are set to zero (i.e., $X_i = 0$ for $i = 1$ to k). The intercepts are not informative, however, if sampling is done by outcome (i.e., disease status). For example, suppose the subjects in the endometrial cancer example had been selected based on tumor type, with age group (i.e., exposure status) determined after selection. This would be analogous to a case-control study design. Although the intercepts are not informative in this setting, the odds ratio is still a valid measure with this sampling method.

IV. Statistical Inference with Three Categories

Two types of inferences:

1. Hypothesis testing about parameters
2. Interval estimation around parameters

Procedures for polytomous outcomes or generalizations of SLR

In polytomous logistic regression, as with standard logistic regression (i.e., a dichotomous outcome), two types of statistical inferences are often of interest: (1) testing hypotheses and (2) deriving interval estimates around parameters. Procedures for both of these are straightforward generalizations of those that apply to logistic regression modeling with a dichotomous outcome variable (i.e., SLR).

95% CI for OR (one predictor)

$$\exp\left\{\hat{\beta}_{g1}(X_1^{**} - X_1^{*}) \pm 1.96(X_1^{**} - X_1^{*})s_{\hat{\beta}_{g1}}\right\}$$

The confidence interval estimation is analogous to the standard logistic regression situation. For one predictor variable, with any levels (X_1^{**} and X_1^{*}) of that variable, the large-sample formula for a 95% confidence interval is of the general form shown at left.

EXAMPLE

Estimated standard errors: (X_1 = AGEGP)

$s_{\hat{\beta}_{21}} = 0.3215, \quad s_{\hat{\beta}_{11}} = 0.3775$

Continuing with the endometrial cancer example, the estimated standard errors for the parameter estimates for AGEGP are 0.3215 for $\hat{\beta}_{21}$ and 0.3775 for $\hat{\beta}_{11}$.

EXAMPLE (*continued*)

95% CI for OR_2

$$= \exp[0.4256 \pm 1.96(0.3215)]$$
$$= (0.82, 2.87)$$

95% CI for OR_1

$$= \exp[0.7809 \pm 1.96(0.3775)]$$
$$= (1.04, 4.58)$$

The 95% confidence interval for OR_2 is calculated as 0.82 to 2.87, as shown on the left. The 95% confidence interval for OR_1 is calculated as 1.04 to 4.58.

Likelihood ratio test

Assess significance of X_1

2 βs tested at the same time

$\Downarrow$

2 degrees of freedom

As with a standard logistic regression, we can use a likelihood ratio test to assess the significance of the independent variable in our model. We must keep in mind, however, that rather than testing one beta coefficient for an independent variable, we are now testing two at the same time. There is a coefficient for each comparison being made (i.e., $D = 2$ vs. $D = 0$ and $D = 1$ vs. $D = 0$). This affects the number of parameters tested and, therefore, the degrees of freedom associated with the test.

EXAMPLE

3 levels of D and 1 predictor

$\Downarrow$

2 αs and 2 βs

Full model:

$$\ln\left[\frac{P(D = g \mid X_1)}{P(D = 0 \mid X_1)}\right] = \alpha_g + \beta_{g1}X_1, \quad g = 1, 2$$

In our example, we have a three-level outcome variable and a single predictor variable, the exposure. As the model indicates, we have two intercepts and two beta coefficients.

Reduced model:

$$\ln\left[\frac{P(D = g)}{P(D = 0)}\right] = \alpha_g, \quad g = 1, 2$$

H_0: $\beta_{11} = \beta_{21} = 0$

If we are interested in testing for the significance of the beta coefficient corresponding to the exposure, we begin by fitting a full model (with the exposure variable in it) and then comparing that to a reduced model containing only the intercepts.

The null hypothesis is that the beta coefficients corresponding to the exposure variable are both equal to zero.

Likelihood ratio test statistic:

$$-2\ln L_{\text{reduced}} - (-2\ln L_{\text{full}}) \sim \chi^2$$

with df = number of parameters set to zero under H_0

The likelihood ratio test is calculated as negative two times the log likelihood ($\ln L$) from the reduced model minus negative two times the log likelihood from the full model. The resulting statistic is distributed approximately chi-square, with degrees of freedom (df) equal to the number of parameters set equal to zero under the null hypothesis.

EXAMPLE

	$-2\ln L$
Reduced:	514.4
Full:	508.9

Difference = 5.5
df = 2
P-value = 0.06

In the endometrial cancer example, negative two times the log likelihood for the reduced model is 514.4, and for the full model is 508.9. The difference is 5.5. The chi-square *P*-value for this test statistic, with two degrees of freedom, is 0.06. The two degrees of freedom are for the two beta coefficients being tested, one for each comparison. We conclude that AGEGP is statistically significant at the 0.10 level but not at the 0.05 level.

Wald test

β for single outcome level tested

Whereas the likelihood ratio test allows for the assessment of the effect of an independent variable across all levels of the outcome simultaneously, it is possible that one might be interested in evaluating the effect of the independent variable at a single outcome level. A Wald test can be performed in this situation.

For two levels:

H_0: $\beta_{11} = 0$ H_0: $\beta_{21} = 0$

$$Z = \frac{\hat{\beta}_{g1}}{s_{\hat{\beta}_{g1}}} \sim N(0,1)$$

The null hypothesis, for each level of interest, is that the beta coefficient is equal to zero. The Wald test statistics are computed as described earlier, by dividing the estimated coefficient by its standard error. This test statistic has an approximate normal distribution.

EXAMPLE

H_0: $\beta_{11} = 0$ (category 1 vs. 0)

$$Z = \frac{0.7809}{0.3775} = 2.07, \quad P = 0.04$$

H_0: $\beta_{21} = 0$ (category 2 vs. 0)

$$Z = \frac{0.4256}{0.3215} = 1.32, \quad P = 0.19$$

Continuing with our example, the null hypothesis for the Adenosquamous vs. Adenocarcinoma comparison (i.e., category 1 vs. 0) is that β_{11} equals zero. The Wald statistic for β_{11} is equal to 2.07, with a *P*-value of 0.04. The null hypothesis for the Other vs. Adenocarcinoma comparison (i.e., category 2 vs. 0) is that β_{21} equals zero. The Wald statistic for β_{21} is equal to 1.32, with a *P*-value of 0.19.

Conclusion: Is AGEGP significant?

⇒ Yes: Adenocarcinoma vs. Adenosquamous

⇒ No: Other vs. Adenosquamous.

At the 0.05 level of significance, we reject the null hypothesis for β_{11} but not for β_{21}. We conclude that AGEGP is statistically significant for the Adenosquamous vs. Adenocarcinoma comparison (category 1 vs. 0), but not for the Other vs. Adenocarcinoma comparison (category 2 vs. 0).

Decision: Retain or drop *both* β_{11} and β_{21} from model

We must either keep both betas (β_{11} and β_{21}) for an independent variable or drop both betas when modeling in polytomous regression. Even if only one beta is significant, both betas must be retained if the independent variable is to remain in the model.

V. Extending the Polytomous Model to *G* Outcomes and *k* Predictors

Adding more independent variables

Expanding the model to add more independent variables is straightforward. We can add k independent variables for each of the outcome comparisons.

$$\ln\left[\frac{\mathrm{P}(D=1\,|\,\mathbf{X})}{\mathrm{P}(D=0\,|\,\mathbf{X})}\right] = \alpha_1 + \sum_{i=1}^{k} \beta_{1i}X_i$$

$$\ln\left[\frac{\mathrm{P}(D=2\,|\,\mathbf{X})}{\mathrm{P}(D=0\,|\,\mathbf{X})}\right] = \alpha_2 + \sum_{i=1}^{k} \beta_{2i}X_i$$

The log odds comparing category 1 to category 0 is equal to α_1 plus the summation of the k independent variables times their β_1 coefficients. The log odds comparing category 2 to category 0 is equal to α_2 plus the summation of the k independent variables times their β_2 coefficients.

Same procedures for OR, CI, and hypothesis testing

The procedures for calculation of the odds ratios, confidence intervals, and for hypothesis testing remain the same.

EXAMPLE

$$D = \text{SUBTYPE}\begin{cases} 0 & \text{if Adenocarcinoma} \\ 1 & \text{if Adenosquamous} \\ 2 & \text{if Other} \end{cases}$$

Predictors

X_1 = AGEGP
X_2 = ESTROGEN
X_3 = SMOKING

To illustrate, we return to our endometrial cancer example. Suppose we wish to consider the effects of estrogen use and smoking status as well as AGEGP on histological subtype ($D = 0, 1, 2$). The model now contains three predictor variables: X_1 = AGEGP, X_2 = ESTROGEN, and X_3 = SMOKING.

EXAMPLE (*continued*)

$$X_1 = \text{AGEGP} \begin{cases} 0 & \text{if 50–64} \\ 1 & \text{if 65–79} \end{cases}$$

$$X_2 = \text{ESTROGEN} \begin{cases} 0 & \text{if never user} \\ 1 & \text{if ever user} \end{cases}$$

$$X_3 = \text{SMOKING} \begin{cases} 0 & \text{if former or never smoker} \\ 1 & \text{if current smoker} \end{cases}$$

Recall that AGEGP is coded as 0 for aged 50–64 or 1 for aged 65–79. Both estrogen use and smoking status are also coded as dichotomous variables. ESTROGEN is coded as 1 for ever user and 0 for never user. SMOKING is coded as 1 for current smoker and 0 for former or never smoker.

Adenosquamous vs. Adenocarcinoma:

$$\ln\left[\frac{\mathrm{P}(D=1\,|\,\mathbf{X})}{\mathrm{P}(D=0\,|\,\mathbf{X})}\right] = \alpha_1 + \beta_{11}X_1 + \beta_{12}X_2 + \beta_{13}X_3$$

The log odds comparing Adenosquamous ($D = 1$) to Adenocarcinoma ($D = 0$) is equal to α_1 plus β_{11} times X_1 plus β_{12} times X_2 plus β_{13} times X_3.

Other vs. Adenocarcinoma:

$$\ln\left[\frac{\mathrm{P}(D=2\,|\,\mathbf{X})}{\mathrm{P}(D=0\,|\,\mathbf{X})}\right] = \alpha_2 + \beta_{21}X_1 + \beta_{22}X_2 + \beta_{23}X_3$$

Similarly, the log odds comparing Other type ($D = 2$) to Adenocarcinoma ($D = 0$) is equal to α_2 plus β_{21} times X_1 plus β_{22} times X_2 plus β_{23} times X_3.

Variable	Estimate	S.E.	Symbol
Intercept 1	−1.2032	0.3190	$\hat{\alpha}_2$
Intercept 2	−1.8822	0.4025	$\hat{\alpha}_1$
AGEGP	0.2823	0.3280	$\hat{\beta}_{21}$
AGEGP	0.9871	0.4118	$\hat{\beta}_{11}$
ESTROGEN	−0.1071	0.3067	$\hat{\beta}_{22}$
ESTROGEN	−0.6439	0.3436	$\hat{\beta}_{12}$
SMOKING	−1.7913	1.0460	$\hat{\beta}_{23}$
SMOKING	0.8895	0.5254	$\hat{\beta}_{13}$

The output for the analysis is shown on the left. There are two beta estimates for each of the three predictor variables in the model. Thus, there are a total of eight parameters in the model, including the intercepts.

EXAMPLE (*continued*)

Adenosquamous vs. Adenocarcinoma:

$$\widehat{\text{OR}}_1 = \frac{\exp[\hat{\alpha}_1 + \hat{\beta}_{11}(1) + \hat{\beta}_{12}(X_2) + \hat{\beta}_{13}(X_3)]}{\exp[\hat{\alpha}_1 + \hat{\beta}_{11}(0) + \hat{\beta}_{12}(X_2) + \hat{\beta}_{13}(X_3)]} = \exp\hat{\beta}_{11} = \exp(0.9871) = 2.68$$

Other vs. Adenocarcinoma:

$$\widehat{\text{OR}}_2 = \frac{\exp[\hat{\alpha}_2 + \beta_{21}(1) + \hat{\beta}_{22}(X_2) + \hat{\beta}_{23}(X_3)]}{\exp[\hat{\alpha}_2 + \hat{\beta}_{21}(0) + \hat{\beta}_{22}(X_2) + \hat{\beta}_{23}(X_3)]} = \exp\hat{\beta}_{21} = \exp(0.2823) = 1.33$$

Interpretation of ORs

Three-variable vs. one-variable model

Three-variable model:

$\Rightarrow$ AGEGP | ESTROGEN, SMOKING

One-variable model:

$\Rightarrow$ AGEGP | no control variables

Odds ratios for effect of AGEGP:

	Model	
Comparison	AGEGP ESTROGEN SMOKING	AGEGP
1 vs. 0	2.68	2.18
2 vs. 0	1.33	1.53

Results suggest bias for single-predictor model:

- Toward null for comparison of category 1 vs. 0
- Away from null for comparison of category 2 vs. 0.

Suppose we are interested in the effect of AGEGP, controlling for the effects of ESTROGEN and SMOKING. The odds ratio for the effect of AGEGP in the comparison of Adenosquamous ($D = 1$) to Adenocarcinoma ($D = 0$) is equal to e to the $\hat{\beta}_{11}$ or exp(0.9871) equals 2.68.

The odds ratio for the effect of AGEGP in the comparison of Other type ($D = 2$) to Adenocarcinoma ($D = 0$) is equal to e to the $\hat{\beta}_{21}$ or exp(0.2823) equals 1.33.

Our interpretation of the results for the three-variable model differs from that of the one-variable model. The effect of AGEGP on the outcome is now estimated while controlling for the effects of ESTROGEN and SMOKING.

If we compare the model with three predictor variables with the model with only AGEGP included, the effect of AGEGP in the reduced model is weaker for the comparison of Adenosquamous to Adenocarcinoma ($\widehat{\text{OR}} = 2.18$ vs. 2.68), but is stronger for the comparison of Other to Adenocarcinoma ($\widehat{\text{OR}} = 1.53$ vs. 1.33).

These results suggest that estrogen use and smoking status act as confounders of the relationship between age group and the tumor category outcome. The results of the single-predictor model suggest a bias toward the null value (i.e., 1) for the comparison of Adenosquamous to Adenocarcinoma, whereas the results suggest a bias away from the null for the comparison of Other to Adenocarcinoma. These results illustrate that assessment of confounding can have added complexity in the case of multilevel outcomes.

EXAMPLE (*continued*)

95% confidence intervals

Use standard errors from three-variable model:

$$s_{\hat{\beta}_{11}} = 0.4118, \quad s_{\hat{\beta}_{21}} = 0.3280$$

95% CI for OR_1

$= \exp[0.9871 \pm 1.96(0.4118)$

$= (1.20, 6.01)$

95% CI for OR_2

$= \exp[0.2832 \pm 1.96(0.3280)$

$= (0.70, 2.52)$

Likelihood ratio test, Wald tests } same procedures as with one predictor

Likelihood ratio test

	$-2 \ln L$
Reduced:	500.97
Full:	494.41

Difference: 6.56
($\sim \chi^2$, with 2 df)
P-value = 0.04

Wald tests

$H_0: \beta_{11} = 0$ (category 1 vs. 0)

$$Z = \frac{0.9871}{0.4118} = 2.40, \quad P = 0.02$$

$H_0: \beta_{21} = 0$ (category 2 vs. 0)

$$Z = \frac{0.2832}{0.3280} = 0.86, \quad P = 0.39$$

The 95% confidence intervals are calculated using the standard errors of the parameter estimates from the three-variable model, which are 0.4118 and 0.3280 for $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$, respectively.

These confidence intervals are calculated with the usual large-sample formula as shown on the left. For OR_1, this yields a confidence interval of 1.20 to 6.01, whereas for OR_2, this yields a confidence interval of 0.70 to 2.52. The confidence interval for OR_2 contains the null value (i.e., 1.0), whereas the interval for OR_1 does not.

The procedures for the likelihood ratio test and for the Wald tests follow the same format as described earlier for the polytomous model with one independent variable.

The likelihood ratio test compares the reduced model without the age group variable to the full model with the age group variable. This test is distributed approximately chi-square with two degrees of freedom. Minus two times the log likelihood for the reduced model is 500.97, and for the full model, it is 494.41. The difference of 6.56 is statistically significant at the 0.05 level ($P = 0.04$).

The Wald tests are carried out as before, with the same null hypotheses. The Wald statistic for β_{11} is equal to 2.40 and for β_{21} is equal to 0.86. The *P*-value for β_{11} is 0.02, while the *P*-value for β_{21} is 0.39. We therefore reject the null hypothesis for β_{11} but not for β_{21}.

EXAMPLE (*continued*)

Conclusion: Is AGEGP significant?*
⇒ Yes: Adenocarcinoma vs. Adenosquamous
⇒ No: Other vs. Adenosquamous.

*Controlling for ESTROGEN and SMOKING

Decision: Retain or drop AGEGP from model.

We conclude that AGEGP is statistically significant for the Adenosquamous vs. Adenocarcinoma comparison (category 1 vs. 0), but not for the Other vs. Adenocarcinoma comparison (category 2 vs. 0), controlling for ESTROGEN and SMOKING.

The researcher must make a decision about whether to retain AGEGP in the model. If we are interested in both comparisons, then both betas must be retained, even though only one is statistically significant.

We can also consider interaction terms in a polytomous logistic model.

Adding interaction terms

$D = (0, 1, 2)$

Two independent variables (X_1, X_2)

$$\text{log odds} = \alpha_g + \beta_{g1}X_1 + \beta_{g2}X_2 + \beta_{g3}X_1X_2,$$

where $g = 1, 2$

Consider a disease variable that has three categories ($D = 0, 1, 2$) as in our previous example. Suppose our model includes two independent variables, X_1 and X_2, and that we are interested in the potential interaction between these two variables. The log odds could be modeled as α_1 plus $\beta_{g1}X_1$ plus $\beta_{g2}X_2$ plus $\beta_{g3}X_1X_2$. The subscript g ($g = 1, 2$) indicates which comparison is being made (i.e., category 2 vs. 0, or category 1 vs. 0).

Likelihood ratio test

To test significance of interaction terms

H_0: $\beta_{13} = \beta_{23} = 0$

To test for the significance of the interaction term, a likelihood ratio test with two degrees of freedom can be done. The null hypothesis is that β_{13} equals β_{23} equals zero.

Full model: $\alpha_g + \beta_{g1}X_1 + \beta_{g2}X_2 + \beta_{g3}X_1X_2$

Reduced model: $\alpha_g + \beta_{g1}X_1 + \beta_{g2}X_2$,

where $g = 1, 2$

A full model with the interaction term would be fit and its likelihood compared against a reduced model without the interaction term.

Wald test

To test significance of interaction term at each level

H_0: $\beta_{13} = 0$

H_0: $\beta_{23} = 0$

It is also possible to test the significance of the interaction term at each level with Wald tests. The null hypotheses would be that β_{13} equals zero and that β_{23} equals zero. Recall that both terms must either be retained or dropped.

Extending model to G outcomes

The model also easily extends for outcomes with more than three levels.

Outcome variable has G levels: $(0, 1, 2, \ldots, G-1)$

Assume that the outcome has G levels (0, 1, 2, . . . , $G - 1$). There are now $G - 1$ possible comparisons with the reference category.

$$\ln\left[\frac{\mathrm{P}(D=g\,|\,\mathbf{X})}{\mathrm{P}(D=0\,|\,\mathbf{X})}\right] = \alpha_g + \sum_{i=1}^{k} \beta_{gi} X_i,$$

where $g = 1, 2, \ldots, G-1$

If the reference category is 0, we can define the model in terms of $G - 1$ expressions of the following form: the log odds of the probability that the outcome is in category g divided by the probability the outcome is in category 0 equals α_g plus the summation of the k independent variables times their β_g coefficients.

Calculation of ORs and CIs as before

The odds ratios and corresponding confidence intervals for the $G - 1$ comparisons of category g to category 0 are calculated in the manner previously described. There are now $G - 1$ estimated odds ratios and corresponding confidence intervals, for the effect of each independent variable in the model.

Likelihood ratio test } same
Wald tests } procedures

The likelihood ratio test and Wald test are also calculated as before.

Likelihood ratio test

$$-2\ln L_{\text{reduced}} - (-2\ln L_{\text{full}}) \sim \chi^2$$

with df = number of parameters set to zero under H_0 ($= G - 1$ if $k = 1$)

For the likelihood ratio test, we test $G - 1$ parameter estimates simultaneously for each independent variable. Thus, for testing one independent variable, we have $G - 1$ degrees of freedom for the chi-square test statistic comparing the reduced and full models.

Wald test

$$Z = \frac{\hat{\beta}_{g1}}{s_{\hat{\beta}_{g1}}} \sim N(0,1),$$

where $g = 1, 2, \ldots, G-1$

We can also perform a Wald test to examine the significance of individual betas. We have $G - 1$ coefficients that can be tested for each independent variable. As before, the set of coefficients must either be retained or dropped.

VI. Likelihood Function for Polytomous Model

(Section may be omitted.)

We now present the likelihood function for polytomous logistic regression. This section may be omitted without loss of continuity.

Outcome with three levels

We will write the function for an outcome variable with three categories. Once the likelihood is defined for three outcome categories, it can easily be extended to G outcome categories.

Consider probabilities of three outcomes:

$$\mathrm{P}(D=0), \mathrm{P}(D=1), \mathrm{P}(D=2)$$

We begin by examining the individual probabilities for the three outcomes discussed in our earlier example, that is, the probabilities of the tumor being classified as Adenocarcinoma ($D=0$), Adenosquamous ($D=1$), or Other ($D=2$).

Logistic regression: dichotomous outcome:

$$\mathrm{P}(D=1\,|\,\mathbf{X}) = \frac{1}{1+\exp\left[-\left(\alpha+\sum_{i=1}^{k}\beta_i X_i\right)\right]}$$

$$\mathrm{P}(D=0\,|\,\mathbf{X}) = 1-\mathrm{P}(D=1\,|\,\mathbf{X})$$

Recall that in logistic regression with a dichotomous outcome variable, we were able to write an expression for the probability that the outcome variable was in category 1, as shown on the left, and for the probability the outcome was in category 0, which is 1 minus the first probability.

Polytomous regression: three-level outcome:

$$\mathrm{P}(D=0\,|\,\mathbf{X}) + \mathrm{P}(D=1\,|\,\mathbf{X}) + \mathrm{P}(D=2\,|\,\mathbf{X}) = 1$$

Similar expressions can be written for a three-level outcome. As noted earlier, the sum of the probabilities for the three outcomes must be equal to 1, the total probability.

$$h_1(\mathbf{X}) = \alpha_1 + \sum_{i=1}^{k}\beta_{1i}X_i$$

$$h_2(\mathbf{X}) = \alpha_2 + \sum_{i=1}^{k}\beta_{2i}X_i$$

To simplify notation, we can let $h_1(\mathbf{X})$ be equal to α_1 plus the summation of the k independent variables times their β_1 coefficients and $h_2(\mathbf{X})$ be equal to α_2 plus the summation of the k independent variables times their β_2 coefficients.

$$\frac{\mathrm{P}(D=1\,|\,\mathbf{X})}{\mathrm{P}(D=0\,|\,\mathbf{X})} = \exp[h_1(\mathbf{X})]$$

$$\frac{\mathrm{P}(D=2\,|\,\mathbf{X})}{\mathrm{P}(D=0\,|\,\mathbf{X})} = \exp[h_2(\mathbf{X})]$$

The probability for the outcome being in category 1 divided by the probability for the outcome being in category 0 is modeled as e to the $h_1(\mathbf{X})$ and the ratio of probabilities for category 2 and category 0 is modeled as e to the $h_2(\mathbf{X})$.

Solve for $P(D = 1 \mid \mathbf{X})$ and $P(D = 2 \mid \mathbf{X})$ in terms of $P(D = 0 \mid \mathbf{X})$.

Rearranging these equations allows us to solve for the probability that the outcome is in category 1, and for the probability that the outcome is in category 2, in terms of the probability that the outcome is in category 0.

$$P(D = 1 \mid \mathbf{X}) = P(D = 0 \mid \mathbf{X}) \exp[h_1(\mathbf{X})]$$
$$P(D = 2 \mid \mathbf{X}) = P(D = 0 \mid \mathbf{X}) \exp[h_2(\mathbf{X})]$$

The probability that the outcome is in category 1 is equal to the probability that the outcome is in category 0 times e to the $h_1(\mathbf{X})$. Similarly, the probability that the outcome is in category 2 is equal to the probability that the outcome is in category 0 times e to the $h_2(\mathbf{X})$.

$$P(D = 0 \mid \mathbf{X}) + P(D = 0 \mid \mathbf{X}) \exp[h_1(\mathbf{X})] + P(D = 0 \mid \mathbf{X}) \exp[h_2(\mathbf{X})] = 1$$

These quantities can be substituted into the total probability equation and summed to 1.

Factoring out $P(D = 0|\mathbf{X})$:

$$P(D = 0 \mid \mathbf{X})[1 + \exp h_1(\mathbf{X}) + \exp h_2(\mathbf{X})] = 1$$

With some algebra, we find that

$$P(D = 0 \mid \mathbf{X}) = \frac{1}{1 + \exp[h_1(\mathbf{X})] + \exp[h_2(\mathbf{X})]}$$

With some simple algebra, we can see that the probability that the outcome is in category 0 is 1 divided by the quantity 1 plus e to the $h_1(\mathbf{X})$ plus e to the $h_2(\mathbf{X})$.

and that

$$P(D = 1 \mid \mathbf{X}) = \frac{\exp[h_1(\mathbf{X})]}{1 + \exp[h_1(\mathbf{X})] + \exp[h_2(\mathbf{X})]}$$

Substituting this value into our earlier equation for the probability that the outcome is in category 1, we obtain the probability that the outcome is in category 1 as e to the $h_1(\mathbf{X})$ divided by one plus e to the $h_1(\mathbf{X})$ plus e to the $h_2(\mathbf{X})$.

and that

$$P(D = 2 \mid \mathbf{X}) = \frac{\exp[h_2(\mathbf{X})]}{1 + \exp[h_1(\mathbf{X})] + \exp[h_2(\mathbf{X})]}$$

The probability that the outcome is in category 2 can be found in a similar way, as shown on the left.

$L \Leftrightarrow$ joint probability of observed data.
The ML method chooses parameter estimates that maximize L

Recall that the likelihood function (L) represents the joint probability of observing the data that have been collected and that the method of maximum likelihood (ML) chooses that estimator of the set of unknown parameters that maximizes the likelihood.

Subjects: $j = 1, 2, 3, \ldots, n$

$$y_{j0} = \begin{cases} 1 & \text{if outcome} = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$y_{j1} = \begin{cases} 1 & \text{if outcome} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$y_{j2} = \begin{cases} 1 & \text{if outcome} = 2 \\ 0 & \text{otherwise} \end{cases}$$

Assume that there are n subjects in the dataset, numbered from $j = 1$ to n. If the outcome for subject j is in category 0, then we let an indicator variable, y_{j0}, be equal to 1, otherwise y_{j0} is equal to 0. We similarly create indicator variables y_{j1} and y_{j2} to indicate whether the subject's outcome is in category 1 or category 2.

$$\mathrm{P}(D = 0 \mid \mathbf{X})^{y_{j0}}\ \mathrm{P}(D = 1 \mid \mathbf{X})^{y_{j1}} \times \mathrm{P}(D = 2 \mid \mathbf{X})^{y_{j2}}$$

The contribution of each subject to the likelihood is the probability that the outcome is in category 0, raised to the y_{j0} power, times the probability that the outcome is in category 1, raised to the y_{j1}, times the probability that the outcome is in category 2, raised to the y_{j2}.

$$y_{j0} + y_{j1} + y_{j2} = 1$$

since each subject has one outcome

Note that each individual subject contributes to only one of the category probabilities, since only one of the indicator variables will be nonzero.

$$\prod_{j=1}^{n} \mathrm{P}(D=0 \mid \mathbf{X})^{y_{j0}} \mathrm{P}(D=1 \mid \mathbf{X})^{y_{j1}} \mathrm{P}(D=2 \mid \mathbf{X})^{y_{j2}}$$

The joint probability for the likelihood is the product of all the individual subject probabilities, assuming subject outcomes are independent.

Likelihood for G outcome categories:

$$\prod_{j=1}^{n} \prod_{g=0}^{G-1} \mathrm{P}(D = g \mid \mathbf{X})^{y_{jg}},$$

The likelihood can be generalized to include G outcome categories by taking the product of each individual's contribution across the G outcome categories.

where

$$y_{jg} = \begin{cases} 1 \text{ if the } j\text{th subject has } D = g \\ \quad (g = 0, 1, \ldots, G-1) \\ 0 \text{ if otherwise} \end{cases}$$

Estimated αs and βs are those which maximize L

The unknown parameters that will be estimated by maximizing the likelihood are the alphas and betas in the probability that the disease outcome is in category g, where g equals $0, 1, \ldots, G-1$.

VII. Polytomous vs. Multiple Standard Logistic Regressions

Polytomous vs. separate logistic models

One may wonder how using a polytomous model compares with using two or more separate dichotomous logistic models.

Polytomous model uses data on all outcome categories in *L*.

Separate standard logistic model uses data ononly two outcome categories at a time.

⇓

Parameter and variance estimates may differ.

The likelihood function for the polytomous model utilizes the data involving all categories of the outcome variable in a single structure. In contrast, the likelihood function for a dichotomous logistic model utilizes the data involving only two categories of the outcome variable. In other words, different likelihood functions are used when fitting each dichotomous model separately than when fitting a polytomous model that considers all levels simultaneously. Consequently, both the estimation of the parameters and the estimation of the variances of the parameter estimates may differ when comparing the results from fitting separate dichotomous models to the results from the polytomous model.

Special case: One dichotomous predictor Polytomous and standard logistic models ⇒ same estimates

In the special case of a polytomous model with one dichotomous predictor, fitting separate logistic models yields the same parameter estimates and variance estimates as fitting the polytomous model.

VIII. SUMMARY

✓ Chapter 9: Polytomous Logistic Regression

This presentation is now complete. We have described a method of analysis, polytomous regression, for the situation where the outcome variable has more than two categories.

We suggest that you review the material covered here by reading the detailed outline that follows. Then, do the practice exercises and test.

Chapter 10: Ordinal Logistic Regression	If there is no inherent ordering of the outcome categories, a polytomous regression model is appropriate. If there is an inherent ordering of the outcome categories, then an ordinal logistic regression model may also be appropriate. The proportional odds model is one such ordinal model, which may be used if the proportional odds assumption is met. This model is discussed in Chap. 10.

Detailed Outline

I. Overview (pages 432–433)

A. Focus: modeling outcomes with more than two levels.

B. Using previously described techniques by combining outcome categories.

C. Nominal vs. ordinal outcomes.

II. Polytomous logistic regression: An example with three categories (pages 434–437)

A. Nominal outcome: variable has no inherent order.

B. Consider "odds-like" expressions, which are ratios of probabilities.

C. Example with three categories and one predictor (X_1):

$$\ln\left[\frac{\mathrm{P}(D=1\,|\,X_1)}{\mathrm{P}(D=0\,|\,X_1)}\right] = \alpha_1 + \beta_{11}X_1,$$

$$\ln\left[\frac{\mathrm{P}(D=2\,|\,X_1)}{\mathrm{P}(D=0\,|\,X_1)}\right] = \alpha_2 + \beta_{21}X_1.$$

III. Odds ratio with three categories (pages 437–441)

A. Computation of OR in polytomous regression is analogous to standard logistic regression, except that there is a separate odds ratio for each comparison.

B. The general formula for the odds ratio for any two levels of the exposure variable (X_1^{**} and X_1^{*}) in a no-interaction model is

$$\mathrm{OR}_g = \exp\left[(\beta_{g1}(X_1^{**} - X_1^{*})\right], \quad \text{where } g = 1, 2.$$

IV. Statistical inference with three categories (pages 441–444)

A. Two types of statistical inferences are often of interest in polytomous regression:

i. Testing hypotheses

ii. Deriving interval estimates

B. Confidence interval estimation is analogous to standard logistic regression.

C. The general large-sample formula (no-interaction model) for a 95% confidence interval for comparison of outcome level g vs. the reference category, for any two levels of the independent variable (X_1^{**} and X_1^{*}), is

$$\exp\left\{\hat{\beta}_{g1}(X_1^{**} - X_1^{*}) \pm 1.96(X_1^{**} - X_1^{*})s_{\hat{\beta}_{g1}}\right\}.$$

D. The likelihood ratio test is used to test hypotheses about the significance of the predictor variable(s).
   i. With three levels of the outcome variable, there are two comparisons and two estimated coefficients for each predictor
   ii. The null hypothesis is that each of the 2 beta coefficients (for a given predictor) is equal to zero
   iii. The test compares the log likelihood of the full model with the predictor to that of the reduced model without the predictor. The test is distributed approximately chi-square, with 2 df for each predictor tested

E. The Wald test is used to test the significance of the predictor at a single outcome level. The procedure is analogous to standard logistic regression.

V. Extending the polytomous model to *G* outcomes and *k* predictors (pages 444–449)

A. The model easily extends to include k independent variables.

B. The general form of the model for G outcome levels is

$$\ln\left[\frac{\mathrm{P}(D=g\,|\,\mathrm{X})}{\mathrm{P}(D=0\,|\,\mathrm{X})}\right]=\alpha_g+\sum_{i=1}^{k}\beta_{gi}X_i,$$

where $g=1,2,\ldots,G-1$.

C. The calculation of the odds ratio, confidence intervals, and hypothesis testing using the likelihood ratio and Wald tests remains the same.

D. Interaction terms can be added and tested in a manner analogous to standard logistic regression.

VI. Likelihood function for polytomous model (pages 450–452)

A. For an outcome variable with G categories, the likelihood function is

$$\prod_{j=1}^{n}\prod_{g=0}^{G-1}\mathrm{P}(D=g\,|\,\mathbf{X})^{y_{ig}},\quad\text{where}$$

$$y_{jg}=\begin{cases}1 & \text{if the } j\text{th subject has } D=g\\ 0 & \text{if otherwise}\end{cases}$$

where n is the total number of subjects and $g=0,1,\ldots,G-1$.

VII. Polytomous vs. multiple standard logistic regressions (page 453)

A. The likelihood for polytomous regression takes into account all of the outcome categories; the likelihood for the standard logistic model considers only two outcome categories at a time.

B. Parameter and standard error estimates may differ.

VIII. Summary (page 453)

Practice Exercises

Suppose we are interested in assessing the association between tuberculosis and degree of viral suppression in HIV-infected individuals on antiretroviral therapy, who have been followed for 3 years in a hypothetical cohort study. The outcome, tuberculosis, is coded as none ($D = 0$), latent ($D = 1$), or active ($D = 2$). The degree of viral suppression (VIRUS) is coded as undetectable (VIRUS = 0) or detectable (VIRUS = 1). Previous literature has shown that it is important to consider whether the individual has progressed to AIDS (no = 0, yes = 1), and is compliant with therapy (COMPLIANCE: no = 1, yes = 0). In addition, AGE (continuous) and GENDER (female = 0, male = 1) are potential confounders. Also, there may be interaction between progression to AIDS and compliance with therapy (AIDSCOMP = AIDS × COMPLIANCE).

We decide to run a polytomous logistic regression to analyze these data. Output from the regression is shown below. (The results are hypothetical.) The reference category for the polytomous logistic regression is no tuberculosis ($D = 0$). This means that a descending option was used to obtain the polytomous regression output for the model, so Intercept 1 (and the coefficient estimates that follow) pertains to the comparison of $D = 2$ to $D = 0$, and Intercept 2 pertains to the comparison of $D = 1$ to $D = 0$.

Variable	Coefficient	S.E.
Intercept 1	−2.82	0.23
VIRUS	1.35	0.11
AIDS	0.94	0.13
COMPLIANCE	0.49	0.21
AGE	0.05	0.04
GENDER	0.41	0.22
AIDSCOMP	0.33	0.14
Intercept 2	−2.03	0.21
VIRUS	0.95	0.14
AIDS	0.76	0.15
COMPLIANCE	0.34	0.17
AGE	0.03	0.03
GENDER	0.25	0.18
AIDSCOMP	0.31	0.17

1. State the form of the polytomous model in terms of variables and unknown parameters.
2. For the above model, state the fitted model in terms of variables and estimated coefficients.
3. Is there an assumption with this model that the outcome categories are ordered? Is such an assumption reasonable?

4. Compute the estimated odds ratio for a 25-year-old noncompliant male, with a detectable viral load, who has progressed to AIDS, compared with a similar female. Consider the outcome comparison latent tuberculosis vs. none ($D = 1$ vs. $D = 0$).
5. Compute the estimated odds ratio for a 25-year-old noncompliant male, with a detectable viral load, who has progressed to AIDS, compared with a similar female. Consider the outcome comparison active tuberculosis vs. none ($D = 2$ vs. $D = 0$).
6. Use the results from the previous two questions to obtain an estimated odds ratio for a 25-year-old noncompliant male, with a detectable viral load, who has progressed to AIDS, compared with a similar female, with the outcome comparison active tuberculosis vs. latent tuberculosis ($D = 2$ vs. $D = 1$).

   Note. If the same polytomous model was run with latent tuberculosis designated as the reference category ($D = 1$), the output could be used to directly estimate the odds ratio comparing a male to a female with the outcome comparison active tuberculosis vs. latent tuberculosis ($D = 2$ vs. $D = 1$). This odds ratio can also indirectly be estimated with $D = 0$ as the reference category. This is justified since the OR ($D = 2$ vs. $D = 0$) divided by the OR ($D = 1$ vs. $D = 0$) equals the OR ($D = 2$ vs. $D = 1$). However, if each of these three odds ratios were estimated with three separate logistic regressions, then the three estimated odds ratios are not generally so constrained since the three outcomes are not modeled simultaneously.

7. Use Wald statistics to assess the statistical significance of the interaction of AIDS and COMPLIANCE in the model at the 0.05 significance level.
8. Estimate the odds ratio(s) comparing a subject who has progressed to AIDS to one who has not, with the outcome comparison active tuberculosis vs. none ($D = 2$ vs. $D = 0$), controlling for viral suppression, age, and gender.
9. Estimate the odds ratio with a 95% confidence interval for the viral load suppression variable (detectable vs. undetectable), comparing active tuberculosis to none, controlling for the effect of the other covariates in the model.
10. Estimate the odds of having latent tuberculosis vs. none ($D = 1$ vs. $D = 0$) for a 20-year-old compliant female, with an undetectable viral load, who has not progressed to AIDS.

Test

True or False (Circle T or F)

T F 1. An outcome variable with categories North, South, East, and West is an ordinal variable.

T F 2. If an outcome has three levels (coded 0, 1, 2), then the ratio of $P(D = 1)/P(D = 0)$ can be considered an odds if the outcome is conditioned on only the two outcome categories being considered (i.e., $D = 1$ and $D = 0$).

T F 3. In a polytomous logistic regression in which the outcome variable has five levels, there will be four intercepts.

T F 4. In a polytomous logistic regression in which the outcome variable has five levels, each independent variable will have one estimated coefficient.

T F 5. In a polytomous model, the decision of which outcome category is designated as the reference has no bearing on the parameter estimates since the choice of reference category is arbitrary.

6. Suppose the following polytomous model is specified for assessing the effects of AGE (coded continuously), GENDER (male = 1, female = 0), SMOKE (smoker = 1, nonsmoker = 0), and hypertension status (HPT) (yes = 1, no = 0) on a disease variable with four outcomes (coded $D = 0$ for none, $D = 1$ for mild, $D = 2$ for severe, and $D = 3$ for critical).

$$\ln\left[\frac{P(D = g \mid \mathbf{X})}{P(D = 0 \mid \mathbf{X})}\right] = \alpha_g + \beta_{g1}\,\text{AGE} + \beta_{g2}\,\text{GENDER} + \beta_{g3}\,\text{SMOKE} + \beta_{g4}\,\text{HPT},$$

where $g = 1, 2, 3$.

Use the model to give an expression for the odds (severe vs. none) for a 40-year-old nonsmoking male. (*Note.* Assume that the expression $[P(D = g \mid \mathbf{X} / P(D = 0 \mid \mathbf{X})]$ gives the odds for comparing group g with group 0, even though this ratio is not, strictly speaking, an odds.)

7. Use the model in Question 6 to obtain the odds ratio for male vs. female, comparing mild disease to none, while controlling for AGE, SMOKE, and HPT.

8. Use the model in Question 6 to obtain the odds ratio for a 50-year-old vs. a 20-year-old subject, comparing severe disease to none, while controlling for GENDER, SMOKE, and HPT.

9. For the model in Question 6, describe how you would perform a likelihood ratio test to simultaneously test the significance of the SMOKE and HPT coefficients.

State the null hypothesis, the test statistic, and the distribution of the test statistic under the null hypothesis.

10. Extend the model from Question 6 to allow for interaction between AGE and GENDER and between SMOKE and GENDER. How many additional parameters would be added to the model?

Answers to Practice Exercises

1. Polytomous model:

$$\ln\left[\frac{\mathrm{P}(D = g \mid \mathbf{X})}{\mathrm{P}(D = 0 \mid \mathbf{X})}\right] = \alpha_g + \beta_{g1}\mathrm{VIRUS} + \beta_{g2}\mathrm{AIDS} + \beta_{g3}\mathrm{COMPLIANCE} + \beta_{g4}\,\mathrm{AGE} + \beta_{g5}\mathrm{GENDER} + \beta_{g6}\mathrm{AIDSCOMP},$$

where $g = 1, 2$.

2. Polytomous fitted model:

$$\ln\left[\frac{\hat{\mathrm{P}}(D = 2 \mid \mathbf{X})}{\hat{\mathrm{P}}(D = 0 \mid \mathbf{X})}\right] = -2.82 + 1.35\mathrm{VIRUS} + 0.94\mathrm{AIDS} + 0.49\mathrm{COMPLIANCE} + 0.05\mathrm{AGE} + 0.41\mathrm{GENDER} + 0.33\mathrm{AIDSCOMP},$$

$$\ln\left[\frac{\hat{\mathrm{P}}(D = 1 \mid \mathbf{X})}{\hat{\mathrm{P}}(D = 0 \mid \mathbf{X})}\right] = -2.03 + 0.95\mathrm{VIRUS} + 0.76\mathrm{AIDS} + 0.34\mathrm{COMPLIANCE} + 0.03\mathrm{AGE} + 0.25\mathrm{GENDER} + 0.31\mathrm{AIDSCOMP}.$$

3. No, the polytomous model does not assume an ordered outcome. The categories given do have a natural order however, so that an ordinal model may also be appropriate (see Chap. 10).
4. $\widehat{\mathrm{OR}}_{1\mathrm{vs}0} = \exp(0.25) = 1.28$.
5. $\widehat{\mathrm{OR}}_{2\mathrm{vs}0} = \exp(0.41) = 1.51$.
6. $\widehat{\mathrm{OR}}_{2\mathrm{vs}1} = \exp(0.41)/\exp(0.25) = \exp(0.16) = 1.17$.
7. Two Wald statistics:

$$H_0\colon \beta_{16} = 0; \quad z_1 = \frac{0.31}{0.17} = 1.82; \text{ two-tailed } P\text{-value}: 0.07,$$

$$H_0\colon \beta_{26} = 0; \quad z_2 = \frac{0.33}{0.14} = 2.36; \text{ two-tailed } P\text{-value}: 0.02.$$

The P-value is statistically significant at the 0.05 level for the hypothesis $\beta_{26} = 0$ but not for the hypothesis $\beta_{16} = 0$. Since we must either keep or drop both interaction parameters from the model, we elect to keep both parameters because there is a suggestion of interaction between AIDS and COMPLIANCE. Alternatively, a likelihood ratio test could be performed. The likelihood ratio test has the advantage that only one test statistic needs to be calculated.

8. Estimated odds ratios (AIDS progression: yes vs. no):
   for COMPLIANCE $= 0$: $\exp(0.94) = 2.56$,
   for COMPLIANCE $= 1$: $\exp(0.94 + 0.33) = 3.56$.

9. $\widehat{\text{OR}} = \exp(1.35) = 3.86; 95\%$ CI: $\exp[1.35 \pm 1.96(0.11)]$
   $= (3.11, 4.79)$.
10. Estimated odds $= \exp[-2.03 + (0.03)(20)]$
    $= \exp(-1.43) = 0.24$.