17

MODELLING ORDINAL AND MULTINOMIAL DATA

OBJECTIVES

After reading this chapter, you should be able to:

1. Select an appropriate model from the following based upon the objectives of your study and the nature of your data:
   - multinomial logistic model
   - proportional-odds model
   - adjacent-category model
   - continuation-ratio model.
2. Fit all of the models listed above.
3. Evaluate the assumptions on which the models are based, and use one or more tests to compare different models.
4. Interpret *OR* estimates from each of the models.
5. Compute predicted probabilities from each of the models.

17.1 INTRODUCTION

In some studies, the outcome of interest might be categorical but have more than 2 categories (*ie* multinomial). These data could be recorded on either a nominal or ordinal scale. Nominal data arise when the outcome categories have no specific ordering (*eg* reason for death might be natural causes, accident, suicide, or homicide). Ordinal data arise when the outcome categories have a distinct order to them (*eg* severity of disease might be classified as absent, mild, moderate, or severe). Clinical outcome data may be better analysed by treating the results as ordinal data rather than dichotomising the result (Norris *et al*, 2006; Valenta *et al*, 2006).

Nominal data can be analysed using log-linear models or multinomial logistic regression models. Log-linear models can simultaneously evaluate the effects of multiple predictors on multiple outcomes, but are limited to the evaluation of categorical variables (predictors and outcomes). Log-linear models are used less frequently than regression-type models in epidemiology, so they will not be discussed further.

An overview of a variety of regression models applicable to nominal and ordinal data is presented in Section 17.2. Each of the 4 models introduced in that section is described in more detail in Sections 17.3 to 17.7. All of the examples used in this chapter are based on the Apgar scores in the birth weight dataset). The Apgar scores were initially recoded into 4 categories (1–6, 7, 8, and 9–10) for the purpose of Figs. 17.1–17.4, but were then recoded into 3 categories (1–6, 7–8 and 9–10) for all subsequent analyses. The main focus of each analysis is to determine the effect of the number of prenatal visits on Apgar scores while controlling for 3 potential confounders (-white-, -gest-, and -male-). The original dataset (bw5k) is described more fully in Chapter 31, but the main variables used in this chapter are shown in Table 17.1.

Table 17.1 Main variables used in the evaluation of factors affecting Apgar scores

obs	observation number
apgar_c4	Apgar score in 4 categories (0=1–6, 1=7, 2=8, 3=9–10)
apgar_c3	Apgar score in 3 categories (0=1–6, 1=7–8, 2=9–10)
previs_c3	number of prenatal visits in 3 categories (0=1–5, 1=6–11, 2=≥12) (also dichotomised at <6, ≥6)
white	mother's race (0=other, 1=white)
gest	gestation length (weeks)
male	gender of baby (0=female, 1=male)

17.2 OVERVIEW OF MODELS

An overview of the 4 models to be discussed in this chapter is presented here. In each case, we will assume that the outcome has J categories, with j being used to designate the categories from 1 to J (*ie* j=1,...,J). For the sake of simplicity, we will assume that there is a single dichotomous predictor in the model, but these models can easily be extended to multiple predictors. A simple example, based on the data in Table 17.2, will be used to demonstrate most of the models. All models discussed in this chapter are presented as logistic models; they can be

fit as other binomial models (*eg* probit, complementary log log) but these are beyond the scope of this text. More details about these models can be found in Hilbe (2009), Long (1997), and Long and Freese (2006).

Table 17.2 Cross-tabulation of Apgar score categories vs number of prenatal visits (dichotomised at <6 and ≥6)

Category	Apgar scores	<6 prenatal visits	≥6 prenatal visits	Totals
0	1–6	14	58	72
1	7	7	92	99
2	8	45	493	538
3	9–10	235	4056	4291
	Total	301	4699	5000

17.2.1 Multinomial logistic model

Nominal data can be analysed using a **multinomial logistic model** which relates the probability of being in category *j* to the probability of being in a baseline category (which we will refer to as category 1). The model can be written as follows.

$$\ln \frac{\mathrm{p}(Y=j)}{\mathrm{p}(Y=1)} = \beta_0^{(j)} + \beta_1^{(j)} X \qquad \textit{Eq 17.1}$$

A complete set of coefficients (β_0 and β_1) is estimated for each of the *J*-1 levels being compared with the baseline (these are designated as $\beta^{(j)}$). Graphically, the effect of the predictor can be seen in Fig. 17.1.

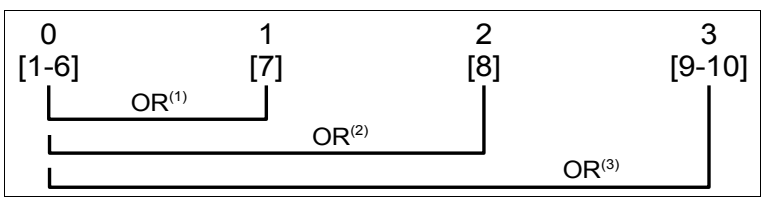


Fig. 17.1 Multinomial logistic regression (Apgar scores in [ ])

Based on the data in Table 17.2, the odds ratio (*OR*) for a high prenatal visit baby (*ie* baby from a pregnancy in which there were ≥6 prenatal visits) being in category 1 (Apgar=7) (compared with category 0) is:

$$OR^{(1)} = \frac{92 * 14}{58 * 7} = 3.17$$

Similarly, the *OR* for category 2 (Apgar=8) compared with category 0 (Apgar=1–6) is:

$$OR^{(2)} = \frac{493 * 14}{45 * 58} = 2.64$$

and similarly for the *OR* for category 3.

17.2.2 Proportional-odds model

The multinomial model does not make any assumptions about the ordering of the categories. An approach for analysing ordinal data is to use a proportional-odds model which relates the probability of being at or above a category to the probability of being in any lower category. The model assumes that this relationship is the same at each of the categories. The model can be written as follows.

$$\ln \frac{\mathrm{p}(Y \geq j)}{\mathrm{p}(Y < j)} = \beta_0^{(j)} + \beta_1 X \qquad \textit{Eq 17.2}$$

Fitting this model requires that J-1 intercepts (β_0) be estimated, but only a single β_1. Graphically, the effects of the predictor can be seen in Fig. 17.2.

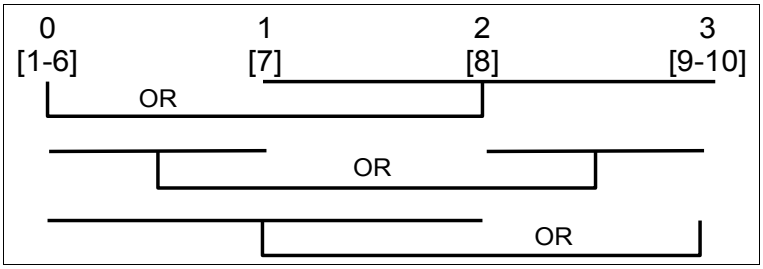


Fig. 17.2 Proportional-odds model (Apgar scores in [ ])

Based on the data in Table 17.2, the *OR* associated with being a high prenatal visit baby for categories 1, 2, or 3 (compared with category 0) is:

$$OR^{(1)} = \frac{(92 + 493 + 4056) * 14}{(7 + 45 + 235) * 58} = 3.90$$

while the *OR* associated with being a high prenatal visit baby for categories 2 or 3 (compared with 0 or 1) (*ie* 8–10 vs 1–7) is:

$$OR^{(2)} = \frac{(14 + 7) * (493 + 4056)}{(45 + 235) * (58 + 92)} = 2.27$$

Because the 2 *OR*s are not very close, the assumption of proportional odds may not be valid. This will be further investigated later in this chapter.

17.2.3 Adjacent-category model

If the categories are ordered, and in some sense 'equidistant', then a constrained multinomial model, or **adjacent-category model,** can be fit to the data. This model is based on the assumption that the predictor increases (or decreases) the log odds of a category occurring by a fixed amount as you go up through the categories. Consequently, the model can be written as follows.

$$\ln \frac{\mathrm{p}(Y = j)}{\mathrm{p}(Y = j-1)} = \beta_0^{(j)} + (J-1)\beta_1 X \qquad \textit{Eq 17.3}$$

Fitting this model requires that J-1 intercepts (β_0) be estimated, but only a single β_1. Graphically, the effects of the predictor can be seen in Fig. 17.3.

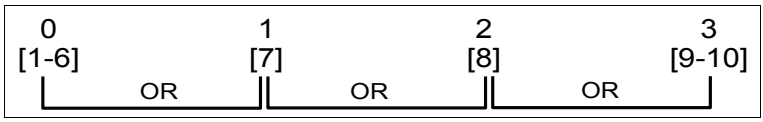


Fig. 17.3 Adjacent-category model (Apgar scores in [ ])

The estimate of β_1 cannot be derived easily from the data in Table 17.2. Based on an adjacent-category model, β_1=0.409 (*OR*=1.51). However, simple estimates of *OR*s comparing Apgar=7 vs 1–6, Apgar=8 vs 7 and Apgar=9–10 vs 8 are 3.17, 0.83, and 1.58, respectively, suggesting that this model may not be appropriate for these data.

17.2.4 Continuation-ratio model

An alternative for analysing ordinal data is to use a continuation-ratio model which relates the probability of being in a category to the probability of being in any lower category. The model can be written as follows.

$$\ln \frac{\mathrm{p}(Y=j)}{\mathrm{p}(Y<j)} = \beta_0^{(j)} + \beta_1^{(j)} X \qquad \textit{Eq 17.4}$$

A complete set of coefficients (β_0 and β_1) is estimated for each of the J-1 categories above the baseline. Graphically, the effect of the predictor can be seen in Fig. 17.4.

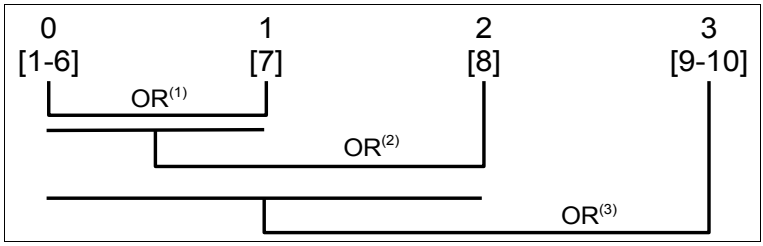


Fig. 17.4 Continuation-ratio model (Apgar scores in [ ])

Based on the data in Table 17.2, the *OR* associated with being a high prenatal visit baby for category 1 (compared with category 0) is:

$$OR^{(1)} = \frac{92*14}{58*7} = 3.17$$

while the *OR* associated with being a high prenatal visit baby for category 2 (compared with being <2) (*ie* 8 vs 1–7) is:

$$OR^{(2)} = \frac{493*(14+7)}{45*(58+92)} = 1.53$$

17.3 Multinomial logistic regression

In multinomial logistic regression of an outcome that has J categories, the probability of membership in each of the outcome categories is computed by simultaneously fitting J-1 separate logistic models (with 1 category serving as the baseline or reference category). Consequently, for a dependent variable with 4 levels (leaving the first level as the baseline category), we estimate 3 sets of coefficients ($\beta^{(1)}$, $\beta^{(2)}$, $\beta^{(3)}$) corresponding to the remaining outcome categories. Because $\beta^{(0)}=0$, the predicted probability that an observation is in category 0 is:

$$p(Y=0)=\frac{1}{1+\exp(X\beta^{(1)})+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})} \qquad \textit{Eq 17.5}$$

while the probability of being in category 1 is:

$$p(Y=1)=\frac{\exp(X\beta^{(1)})}{1+\exp(X\beta^{(1)})+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})} \qquad \textit{Eq 17.6}$$

and similarly for categories 2 and 3.

17.3.1 Odds ratios

For any given predictor (*eg* -white-), there is a separate estimate of the effect of that predictor on each outcome (relative to the base level). Exponentiation of the coefficients from a multinomial regression model produces odds ratios as a measure of effect. **Note** Strictly speaking, these effect measures are not odds ratios. They are actually the ratio of 2 relative risks (or risk ratios), with each relative risk describing the probability of the outcome in the category of interest relative to the baseline category. Consequently, it would be more appropriate to refer to them as relative risk ratios, and some computer programs do so. However, the term odds ratio is commonly applied and will be used in this chapter.

Example 17.1 shows a very simple model for Apgar scores (3 categories—all analyses from now on will be based on the 3-level Apgar score variables) with -previs_c2- as the single predictor and the baseline level set to Apgar scores of 9–10. The odds ratios indicate that a high prenatal visit baby was 0.24 and 0.65 times as likely to have a score between 1 and 6 or between 7 and 8 (compared with 9–10) as a low prenatal visit baby.

Both of the *OR*s in Example 17.1 suggest that frequent prenatal visits reduced the risk of a lower Apgar score and this effect was clearly statistically significant (see Section 17.3.3 for assessment of significance).

As with ordinary logistic regression, multinomial logistic regression can be extended to model the effects of multiple predictors that might be categorical or continuous in nature. Example 17.2 shows a model for Apgar scores including additional potential confounders with results presented as coefficients.

Example 17.1 Simple multinomial logistic regression
data = bw5k

A simple multinomial logistic regression of Apgar scores (3 levels) was carried out with -previs_c2- as the sole predictor. The baseline (referent) level was babies with scores of 9–10.

The first table presents the results in terms of coefficients of the logistic models.

Number of obs = 5000
LR chi2 (2) = 21.94
Prob > chi2 = <0.001
Log likelihood = -2263.01

Apgar score		Coef	SE	Z	P	95% CI	
1–6							
	previs_c2	-1.427	0.305	-4.67	0.000	-2.025	-0.829
	constant	-2.821	0.275	-10.25	0.000	-3.360	-2.281
7–8							
	previs_c2	-0.428	0.160	-2.68	0.007	-0.741	-0.115
	constant	-1.508	0.153	-9.84	0.000	-1.809	-1.208

Having a high number of prenatal visits reduced the logit of the probability of having Apgar scores ≤6 and 7–8 by 1.42 and 0.43 units, respectively.

The second table presents the results in terms of odds ratios.

Apgar score		OR	SE	95% CI	
1–6					
	previs_c2	0.240	0.073	0.132	0.437
7–8					
	previs_c2	0.652	0.104	0.477	0.891

Babies from high prenatal visit pregnancies were 0.24 and 0.65 times as likely to have Apgar scores of 1–6 or 7–8, respectively, compared with low prenatal visit babies.

17.3.2 Interpretation of coefficients

Estimates (coefficients or *OR*s) from multinomial logistic regression models are interpreted in a manner similar to those from ordinary logistic regression. The *OR* for the predictor -previs_c2- in Example 17.1 suggests that, for mothers with ≥6 prenatal visits, the odds of having a baby with an Apgar score of 1–6 goes down by a factor of $e^{-1.427}$=0.24 (76% reduction), while the odds of having a score of 7–8 goes down by a factor of $e^{-0.428}$=0.65 (35% reduction). In Example 17.2, all of the predictors have more pronounced effects on the 1–6 vs 9–10 comparison compared with the 7–8 vs 9–10 comparison. This was expected given the ordinal nature of the data, but nothing in the model guarantees this. This pattern would not be expected if unordered nominal data were being analysed.

Example 17.2 Multiple multinomial logistic regression
data = bw5k

Prediction of Apgar score category (n=3) based on the number of prenatal visits (3 categories: 0–5, 6–11, 12+), mother's race (white vs other), gestation length (weeks) and baby gender. The Apgar score category 9–10 served as the baseline.

Number of obs = 5000
LR chi2 (10) = 108.2
Prob > chi2 < 0.001
Log likelihood = -2219.97

Apgar score	Coef	SE	Z	P	95% CI	
1–6						
previs_c3=1	-1.025	0.358	-2.87	0.004	-1.726	-0.324
previs_c3=2	-0.944	0.361	-2.61	0.009	-1.653	-0.236
white	0.429	0.253	1.70	0.090	-0.067	0.925
gest	-0.216	0.032	-6.79	0.000	-0.278	-0.153
male	0.715	0.260	2.74	0.006	0.204	1.225
constant	4.271	1.142	3.74	0.000	2.032	6.510
7–8						
previs_c3=1	-0.295	0.170	-1.74	0.083	-0.627	0.038
previs_c3=2	-0.336	0.170	-1.98	0.048	-0.669	-0.003
white	0.279	0.088	3.16	0.002	0.106	0.453
gest	-0.103	0.017	-6.11	0.000	-0.136	-0.070
male	0.015	0.086	0.17	0.862	-0.153	0.183
constant	2.166	0.642	3.38	0.001	0.909	3.424

When compared with the highest Apgar scores, higher numbers of prenatal visits reduced the risk of lower scores with the effect on the lowest category of score (1–6) being most pronounced. Longer gestation lengths reduced the risk of both lower categories of Apgar score while being a male baby and having a white mother significantly increased the risks of having scores in the ranges 1–6 and 7–8, respectively.

17.3.3 Testing significance of predictors

The significance of predictors can be assessed using either a Wald test or a likelihood ratio test (*LRT*). Overall tests of significance can be carried out (for all logistic models fit) as well as tests for coefficients within individual models. Note, however, that tests of significance for a predictor within a given logistic model (*eg* for Apgar score=1–6) will change if the baseline category is changed. Consequently, overall tests of significance provide a better estimate of the significance of the predictor. All of the factors in the model in Example 17.2 had significant Wald tests ($P<0.05$).

17.3.4 Obtaining predicted probabilities

Predicted probabilities of the occurrence of each outcome category can be computed from the multinomial logistic regression (see Eqs 17.5 and 17.6). These will, of course, vary with the values of the predictors for the individual. Table 17.3 shows those values for a selected number of babies based on the model in Example 17.2.

Table 17.3 Predicted probabilities from a multinomial logistic regression model

						probability of score category		
obs	Apgar score category	prenatal visits	mother's race	gestation length	baby gender	0 (1–6)	1 (7–8)	2 (9–10)
655332	2	6–11	white	33	male	0.048	0.216	0.736
1358363	2	≥12	white	29	female	0.055	0.279	0.666
2926875	1	<6	white	36	male	0.069	0.209	0.722
3464037	0	<6	non-white	25	male	0.284	0.289	0.427
3586653	2	≥12	white	40	female	0.007	0.118	0.875

The sum of the probabilities for each individual equals 1. The baby with the very short gestation (25 weeks) had a relatively high probability of a very low Apgar score (category 0), and did indeed have the lowest score category.

17.3.5 Assumption of independence of irrelevant alternatives (IIA)

The multinomial regression model is based on an assumption that the odds of 1 level of the outcome being observed is independent of what other alternatives are available. For the Apgar score data discussed, this would mean that if for any individual the odds of a score of 9–10 were twice those of a score of 7–8, they should always be twice, regardless of how many other alternatives were possible.

Two of the most commonly used tests of this assumption are the Hausman and McFadden (1984), and Small-Hsiao (1985) tests of IIA. Both are based on the principle of fitting a full model and comparing the coefficients from that model to a model with 1 or more of the alternatives deleted (partial model). The null hypothesis is that the coefficients from the full model are the same as from the partial model. If the P-value of the test is >0.05, there is insufficient evidence to reject the null hypothesis (*ie* the assumption has been met). For the Hausman test, the statistic may be negative, which is also assumed to support the null hypothesis.

Unfortunately, the 2 tests often give conflicting results, and recent simulation studies (cited in Long and Freese (2006)) suggest that they may be of limited use in determining whether the assumption has been met. In the face of conflicting results, the best advice may be from the early statement of McFadden cited in Long and Freese that multinomial models should only be used when the alternatives 'can plausibly be assumed to be distinct and weighted independently in the eyes of the decision-maker'. Given the semi-quantitative and partially objective nature of Apgar scores, it seems likely that the assigning of an Apgar score would be independent of the

range of choices available. Example 17.3 contains the results of these 2 tests and they support the notion that the assumption is satisfied.

It is also possible to statistically evaluate (using a Wald or likelihood ratio test) whether any of the outcome levels are not significantly different from other levels. If some are not, one might want to consider combining those levels (see Example 17.3).

17.3.6 Regression diagnostics

Specialised diagnostics for multinomial logistic regression are not as readily available as they are for ordinary logistic regression. One approach is to fit ordinary logistic models for pairs of comparisons (*eg* grade=1–6 vs 9–10 and 7–8 vs 9–10) and evaluate the regression diagnostics for those models. An overall goodness-of-fit test has recently been developed (Fagerland *et al*, 2008) but, at the time of writing, was not readily available in standard software packages.

17.3.7 Models for outcomes with alternative specific data

In the Apgar score example, none of the predictors vary across outcome alternatives (*ie* the gestation length of the baby was constant regardless of the Apgar score category). This is not always the case. Consider the situation in which a patient has to choose among 3 options for dealing with a recently diagnosed case of cancer. The options might include: treatment at their local hospital, treatment at a regional referral hospital or treatment at a specialised cancer clinic. Factors which might influence their decision might include age of patient, income level, and distance to various clinics. While the first 2 factors (age and income) are independent of the alternatives, the last (distance) varies with the alternatives being considered (*eg* the local clinic is closer than the other alternatives). Various approaches for dealing with this situation exist (one of which is conditional logistic regression—Section 16.15) and the reader is referred to Hilbe (2009), and Long and Freese (2006) for an explanation of how to structure the data and fit an appropriate model for this situation.

17.4 Modelling ordinal data

Ordinal data can arise in a variety of ways. For example, an observed continuous variable might be divided into categories. Alternatively, levels of an ordinal variable could represent categories of an unobserved (but hypothesised) continuous variable (*eg* opinions ranging from strongly agree to strongly disagree, or disease severity ranging from absent to severe). Finally,

Example 17.3 Evaluating assumption of independence of irrelevant alternatives (IIA)
data = bw5k

The P-values for the Hausman test of IIA were 0.969 and 1.000 if levels 1–6 or 7–8, respectively, were left out. Both values strongly support the notion that the assumption of IIA had been satisfied. The Small-Hsiao test of IIA produces different estimates each time it is run (due to a random element in the calculations) and the results are somewhat unstable but generally support the null hypothesis that the assumption was met.

A likelihood ratio test of whether or not any of the levels could be combined produced P-values <0.001 for all pairwise combinations of levels, suggesting that no pairs of outcome levels could be combined.

categories might represent total values of a composite variable made up of a series of scored variables (*eg* a hygiene score that represents the sum of scores from several questions about hygiene in an operating room).

While the multinomial models described above can also be used to analyse ordinal data, they ignore the fact that the categories fall in a logical, ordered sequence. There are a number of ways to fit ordinal models. We will consider 3 of them: proportional-odds models (Section 17.5), adjacent-category models (Section 17.6), and continuation-ratio models (Section 17.7).

17.5 Proportional-odds model (constrained cumulative logit model)

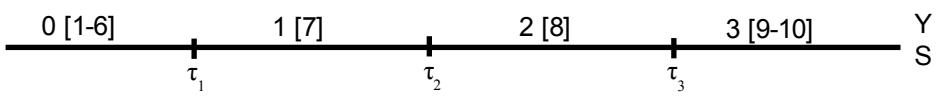
This is the most commonly encountered type of ordinal logistic model. In a proportional-odds model, the coefficients measure the effect of a predictor on the log odds of being at or above a specified level compared with the log odds of being below the specified level. It is based on the assumption that the coefficients do not depend upon the outcome level, so only a single coefficient for each predictor is estimated. A graphic representation of this model is presented in Fig. 17.2.

A proportional-odds model assumes that the ordinal outcome variable represents categories of an underlying continuous latent (unobserved) variable. Assume that the value of the underlying latent variable (or 'score') (S_i) is a linear combination of predictor variables.

$$S_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad \textit{Eq 17.7}$$

where ε_i is a random error term from a continuous distribution.

The latent variable (S) is divided by cutpoints (τ_j) so that the i^{th} individual is classified as category 0 (1–6) if $S_i \leq \tau_1$ and is classified as category 1 (7) if $\tau_1 < S_i \leq \tau_2$, and so on.



The probability of observing outcome j in the i^{th} individual is:

$$p(\text{outcome}_i = j) = p(\tau_{j-1} < S_i < \tau_j) \qquad \textit{Eq 17.8}$$

If the random term (ε_i) is assumed to have a logistic distribution (with a mean of 0 and a variance of $\pi^2/3$), then

$$\text{p}(S_i < \tau_j) = \frac{1}{1 + \text{e}^{S_i - \tau_j}} \qquad \textit{Eq 17.9}$$

Note Assuming the latent variable has a normal distribution gives rise to an ordinal probit model, but those are not discussed in this chapter.

The model fit by assuming a logistically distributed latent variable can also be written as (presented with a single predictor X for simplicity):

$$\text{logit}(\text{p}(Y \leq j)) = \beta_{0j} + \beta X$$

where the β_{0j} are intercepts and β is the effect (slope) of the predictor. Thus, the model is one in which the log odds of the outcome can be viewed as being represented by a series of parallel lines with different intercepts.

Example 17.4 presents a proportional-odds model for the Apgar score data.

17.5.1 Predicted probabilities

The probabilities of each Apgar score category for the 5 selected babies (and the values of their predictor variables) are shown in Table 17.4. The first observation listed in Table 17.4 is a male baby from a white mother who had 6–11 prenatal visits during her 33-week gestation. For this baby, the latent variable (S_1) is:

$$S_i = 4.292$$

Consequently, the probability of this baby being in category 0 (Apgar score=1–6) (from Eq 17.9) is:

$$\mathrm{p}(Y=0) = \frac{1}{1+\mathrm{e}^{4.292-(0.868)}} = 0.032$$

Similarly, the probabilities of this baby being in categories 7–8 and 9–10 are 0.245 and 0.724, respectively.

Example 17.4 Proportional-odds model
data = bw5k

A proportional-odds model was fit to the Apgar score data with the same predictors as used in Example 17.2 and 17.3.

Number of obs = 5000
LR chi2 (5) = 90.90
Prob > chi2 < 0.001
Log likelihood = -2228.53

	Coef	SE	Z	P	95% CI	
previs_c3=1	0.441	0.157	2.81	0.005	0.134	0.748
previs_c3=2	0.464	0.157	2.96	0.003	0.157	0.771
white	-0.297	0.085	-3.51	0.000	-0.463	-0.131
gest	0.128	0.016	8.12	0.000	0.097	0.159
male	-0.091	0.082	-1.11	0.269	-0.252	0.070
cutpoint 1	0.868	0.605			-0.318	2.055
cutpoint 2	3.328	0.601			2.150	4.506

According to this model, the odds ratio associated with having ≥12 prenatal visits, compared with <6 visits is:

$$\mathrm{e}^{0.464}=1.59$$

This suggests that having ≥12 visits (compared with <6) increases the odds of being at or above any given Apgar category compared with being below that category by 1.59 times. In other words, compared with babies with 0–5 prenatal visits, babies with ≥12 visits were approximately 1.5 times as likely to have an Apgar score ≥7 as <7 and also 1.5 times as likely to have a score ≥9 as <9. It measures the overall increased chance of a higher score that is associated with having many prenatal visits.

Table 17.4 Values of predictor variables, latent variables (S_i) and predicted probabilities of each of the Apgar score categories from the proportional-odds model

							probability of score		
obs	Apgar score category	prenatal visits	mother's race	gestation length	baby gender	S	0 (1–6)	1 (7–8)	2 (9–10)
655332	2	6–11	white	33	male	4.292	0.032	0.245	0.724
1358363	2	≥12	white	29	female	3.892	0.046	0.316	0.637
2926875	1	<6	white	36	male	4.237	0.033	0.254	0.713
3464037	0	<6	non-white	25	male	3.121	0.095	0.457	0.448
3586653	2	≥12	white	40	female	5.305	0.012	0.110	0.878

The effect of a single predictor (-gest-) on the predicted probability can best be viewed by generating smoothed curves of the probability of each Apgar score category against -gest-. Fig. 17.5 shows a graph of lowess smoothed probabilities (smoothed with a bandwidth of 50%) of each category against the gestation length. **Note** As the probability of each outcome depends on the value of all predictors in the model, the smoothed curves shown in Fig. 17.5 represent average probabilities of the category as -gest- changes. As can be seen, the probability of a baby having a low Apgar score (either 1–6 or 7–8) goes down as the gestation length goes up. On the other hand, the probability of a high score (9–10) goes up substantially.

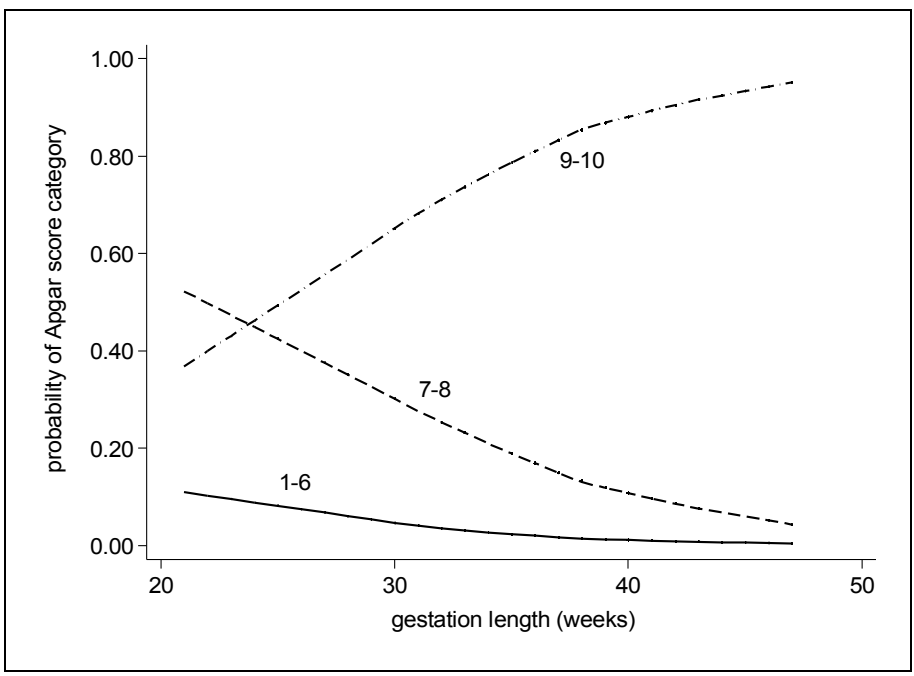


Fig. 17.5 Smoothed mean probabilities of Apgar score categories (based on proportional odds model)

17.5.2 Evaluating the proportional-odds assumption

A rough assessment of the assumption of proportional odds can be obtained by comparing the log likelihood of the ordered logit model (L_1) with one obtained from the multinomial logit model (L_0) using a likelihood ratio test. If there are k predictors (not counting the intercept) and J categories of outcome, the multinomial model will fit $(k+1)(J-1)$ parameters, while the proportional-odds model will fit $k+(J-1)$, so the difference in degrees of freedom is $k(J-2)$. Consequently, $-2(L_1-L_0)$ should have an approximate χ^2 distribution with $k(J-2)$ degrees of freedom. **Note** This is only an approximate test because the proportional-odds model is not nested within the multinomial model. However, it gives a rough assessment of the proportional-odds assumption.

In our example, the log likelihoods of the multinomial and proportional odds models were -2220.0 and -2228.5, respectively so the *LRT* is:

$$LRT = -2(-2220.0-[-2228.5]) = 17.0$$

The χ^2 statistic has $k(J-2)=5$ df which yields a P-value of 0.004. Consequently, there is strong evidence that the proportional-odds assumption does not hold. As an alternative to comparing the ordinal logistic model with a multinomial model, the comparison can be made with a generalised ordinal logistic model (described below—Section 17.5.3). This comparison yields a χ^2 of 16.7 (P=0.005).

An alternative approximate *LRT* based on fitting J-1 separate binary models has been developed (Wolfe and Gould, 1998). The models are fit first assuming the βs are constant across all models (proportional-odds assumption) and the sum of these log likelihoods is compared with the sum of those obtained by fitting the models without the assumption of constant βs. For the Apgar data model, this test produces a χ^2 value of 16.8 (P=0.005).

The likelihood ratio tests described above are omnibus tests which evaluate the assumption of proportional odds over all predictors. A Wald test which will provide an overall assessment as well as an evaluation of the assumption for each predictor separately is available (Brant, 1990). The results of this test for the model fit in Example 17.4 are presented in Table 17.5.

Table 17.5 Brant (Wald) test of proportional-odds assumption

Variable	χ^2	P	df
all	19.49	0.002	5
previs_c3=1	2.81	0.094	1
previs_c3=2	1.75	0.186	1
white	0.13	0.717	1
gest	6.15	0.013	1
male	6.37	0.012	1

The P-value of the overall Wald test is comparable to the approximate likelihood ratio tests described above. The proportional-odds assumption is most clearly violated for -gest- and -male-. Other tests of the proportional-odds assumption are available but there are no clear guidelines for choosing one test over another. In general, if any of the tests discussed above yields a significant result, the assumption should be investigated further.

17.5.3 Dealing with non-proportional odds

In the event that 1 or more predictors appears to violate the assumption of proportional odds, there are a number of potential approaches to dealing with the problem. A **generalised ordinal logistic regression model** is 1 in which a complete set of coefficients is estimated for each cutpoint in the ordinal model (*eg* 1–6 vs 7–10 and 1–8 vs 9–10). Consequently, it is no more parsimonious than the multinomial model, but it does take into account the ordering of the categories. The log-likelihood of this model can be compared with that of a model assuming proportional odds to see if the assumption is valid (see Section 17.5.2).

If the proportional odds assumption appears to hold for some predictors, but not all, it is possible to fit a **partial proportional-odds model** in which the assumption of proportional odds is removed for selected predictors. For our example, -gest- and -male- were the predictors which most clearly violated the proportional odds assumption (Table 17.5). If the coefficients for these 2 predictors are allowed to vary across cut-points, but the remainder are constrained to be constant (proportional odds), the log-likelihood for the model is -2221.861 which, when compared with the proportional odds model, yields a likelihood ratio test χ^2 of 13.3 (P=0.001), providing clear evidence that the proportional odds model is superior. There is no evidence that the generalised ordinal model is superior to the partial proportional odds model (χ^2 of 3.32 (P=0.345).

Two other approaches for dealing with non-proportional odds are the **stereotype logistic model** and the **heterogeneous choice logistic model**. These are beyond the scope of this text and the reader is referred to Hilbe (2009) and Long and Freese (2006) for details.

17.5.4 Regression diagnostics

As with multinomial models, regression diagnostics for ordinal models are not well developed. Hosmer and Lemeshow (2000) suggest fitting ordinary logistic models to data based on the cutpoints in the ordinal data (*eg* 1 model which compares 1–6 with 7–10, and one which compares 1–8 with 9–10). Residuals from these models can be evaluated using techniques described in Chapter 16.

17.6 Adjacent-category model

In an adjacent-category logistic regression model, each coefficient measures the effect of a factor on the logit of the probability of being in a specified level compared with the probability of being in the level below. For any given predictor, this results in the estimation of a single effect that expresses how the predictor influences the log odds of the outcome moving up to the next (adjacent) category. This model is also known as a constrained multinomial model, because it is estimated as a multinomial model with the constraint that the coefficient for categories *n* levels apart be *n* times the coefficient for adjacent categories. (Alternatively, the *OR* for categories *n* levels apart will be the *OR* for adjacent levels raised to the power *n*.) This model is based on the assumption that, as you go from one level to the next, the *OR* is constant. A graphic representation is shown in Fig. 17.3.

Example 17.5 presents an adjacent-category model based on the multinomial model fit in Example 17.2. A likelihood ratio test can be used to compare this ‘constrained multinomial

Example 17.5 Adjacent-category model
data = bw5k

An adjacent-category model was fit using the same predictors presented in Example 17.2. The constraint that coefficients for categories 2 levels apart be twice those of the adjacent categories reduces the number of parameters which need to be estimated.

Number of obs = 5000
LR chi2 (5) = 110.09
Prob > chi2 < 0.001
Log likelihood = -2223.35

		Coef	SE	Z	P	95% CI	
1–6	**baseline**						
7–8							
	previs_c3=1	0.398	0.130	3.07	0.002	0.144	0.652
	previs_c3=2	0.413	0.130	3.17	0.002	0.158	0.668
	white	-0.259	0.075	-3.48	0.001	-0.405	-0.113
	gest	0.107	0.013	8.52	0.000	0.082	0.131
	male	-0.120	0.072	-1.67	0.095	-0.261	0.021
	constant	-1.935	0.462	-4.19	0.000	-2.841	-1.030
9–10							
	previs_c3=1	0.796	0.259	3.07	0.002	0.288	1.304
	previs_c3=2	0.826	0.260	3.17	0.002	0.316	1.337
	white	-0.518	0.149	-3.48	0.001	-0.811	-0.226
	gest	0.213	0.025	8.52	0.000	0.164	0.262
	male	-0.240	0.144	-1.67	0.095	-0.523	0.042
	constant	-4.284	0.924	-4.63	0.000	-6.096	-2.472

Note It is easier to think about an adjacent category model as you move up the outcome scale, so for this example the Apgar score category=1–6 has been set as the baseline (referent) level.

Note The coefficient for each predictor for score=9–10 is twice that for score=7–8 because it is 2 categories away from 1–6. For example, for each additional week of gestation, the log odds of being 7–8 goes up by 0.107 units and the log odds of being 9–10 goes up by 0.213 units.

model' with the usual multinomial model. If the test is significant, it suggests that the multinomial model is superior. The *LRT* for the model in Example 17.5 had a χ^2 of 6.76 with 5 df (because 5 fewer coefficients were estimated) with a P-value of 0.239, suggesting that adjacent-category model would be a valid (and simpler) alternative to the multinomial model.

17.7 CONTINUATION-RATIO MODEL

In continuation-ratio models, the log *OR* measures the effect of a factor on the odds of being in a specified level compared with the odds of being in any of the lower levels. This type of model

is useful in situations where the dependent variable represents the number of attempts required to achieve an outcome (*eg* number of applications to a medical school required for admission). The individual must pass through all lower levels to reach the current level (you can't have your 3rd application until you have had your 1st and 2nd); hence the term 'continuation-ratio.' A graphic representation is shown in Fig. 17.4.

This model can be fit as a series of simple logistic models in which the dependent variable (Y) is recoded to be 1 for the level of interest, 0 for all lower levels and missing for all higher levels. For example, a continuation-ratio model evaluating the effects of predictors on the probability of medical school admission for up to 4 attempts would require 3 separate logistic regressions. The data would be recoded as shown in Table 17.6.

Table 17.6 Coding of data for a continuation-ratio model of effect of predictors on number of attempts before gaining admission to medical school

	number of attempts			
	1	**2**	**3**	**4**
Y1	0	1	missing	missing
Y2	0	0	1	missing
Y3	0	0	0	1

In this example, the coefficient for a predictor represents the effect of the factor on the log odds of being admitted on the j^{th} attempt, conditional on not being admitted on any previous attempts.

The model contains the same number of parameters as the multinomial model presented in Section 17.3. Consequently, the model is no more 'parsimonious', but it results in estimates of the *OR* which have different interpretations than those from a multinomial logistic regression model. A constrained continuation-ratio model can be fit with the *OR* for each predictor constrained to be equal for each increment in the outcome. A likelihood ratio test, comparing the constrained and unconstrained models, can be used to evaluate the assumption of equal *OR*s.

The *OR* from the separate logistic models for the Apgar score data are not presented because it does not make biological sense to fit these data with a continuation-ratio model (*ie* movements between categories are not sequential events).

References

Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics. 1990;46(4):1171–8.

Fagerland MW, Hosmer DW, Bofin AM. Multinomial goodness-of-fit tests for logistic regression models. Stat Med. 2008;27(21):4238–53.

Hausman J, McFadden D. Specification tests for the multinomial logit model. Econometrica. 1984;52:1219–40.

Hilbe JM. Logistic Regression Models. Boca Raton FL: CRC Press; 2009.

Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd Ed. New York: John Wiley and Sons; 2000.

Long JS. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks CA: Sage Publications; 1997.

Long JS, Freese J. Regression Models for Categorical Dependent Variables Using Stata. College Station, TX: Stata Press; 2006.

Norris CM, Ghali WA, Saunders LD, Brant R, Galbraith D, Faris P, *et al*. Ordinal regression model and the linear regression model were superior to the logistic regression models. J Clin Epidemiol. 2006;59(5):448–56.

Small KA, Hsiao C. Multinomial logit specification tests. International Economic Review. 1985;26:619–27.

Valenta Z, Pitha J, Poledne R. Proportional odds logistic regression—effective means of dealing with limited uncertainty in dichotomizing clinical outcomes. Stat Med. 2006;25(24):4227–34.

Wolfe R, Gould W. An approximate likelihood-ratio test for ordinal response models. Stata Technical Bulletin. 1998;42:24–7.