14

Categorical Data Analysis with a Psychometric Twist

Carolyn J. Anderson

INTRODUCTION

Variables measured in psychological and social science research are often discrete, such as the choice between two objects in a paired comparison experiment, the response option selected on a survey item, the correct answer on a test question, career choice, gender, the highest degree earned, and many others. The focus of this chapter is on models for discrete data such as those often collected in psychological and educational research. The statistical models presented in this chapter fall within the class of generalized linear models (GLM). The GLM framework provides a unification of numerous models proposed in statistics, medicine, economics, sociology, and psychology.

Many statistical models for categorical data were developed outside the realm of psychology; however, many psychometric models can be fit to data as a GLM. Described in this chapter are special cases of GLMs that correspond to particular psychometric models. These include logistic and probit regression models for dichotomous response variables, conditional multinomial logistic regression models for polytomous responses, and log-linear (Poisson regression) models for counts. In psychometric models of behavoir, observations are often assumed to be due to individuals' values on a latent trait. Even though the categories of a variable may have no inherent or natural ordering, they may be ordered with respect to some underlying or latent trait, such as ability, preference, utility, knowledge, attitude, or prestige.

An advantage of using the GLM framework to fit various psychometric models is that many software packages are available for fitting GLMs to data, including SAS (SAS Institute, 2003), S-Plus (Insightful Corporation, 2007), R (R Core Team, 2006), and others. SAS, R, and ℓEM (Vermunt, 1997) input files for examples reported in this chapter are available at http://faculty.ed.uiuc.edu/cja/homepage/software_index.html.

Before delving into GLM theory, measures of association will be discussed, an understanding of which is key to interpreting model parameters that represent dependency between variables. After an introduction to GLMs, sections on logistic and Poisson regression models provide illustrations of basic modeling procedures using data common in psychological and

AQ: "SAS Institute Inc. (2003)" (As per Reference list.)

AQ: "SAS Institute Inc. (2003)" (As per Reference list.)

educational research. After this foundational material, a general model for values on a latent continuum is presented and is used to describe latent variable models for paired comparisons (i.e., Thurstone's model of comparative judgment and the Bradely–Terry–Luce model), a multinomial discrete choice model (i.e., McFadden's model), and item response models (i.e., the Rasch and two-parameter logistic models). All of these psychometric models can be formulated as special cases of GLMs.

MEASURES OF ASSOCIATION

Many measures of association for categorical data exist, including ones for ordinal and nominal variables, dichotomous and multicategory variables, local and global dependency, and symmetric and asymmetric measures (Altham, 1970; Agresti, 1980; Dale, 1986; Edwardes and Baltzan, 2000; Goodman and Kruskal, 1979). Only the two most common measures of association are discussed here: odds ratios and the correlation coefficient.

For illustration, we use the data in Table 14.1, which consists of a cross-classification of two vocabulary items, A and C, from the 2004 General Social Survey (GSS) (Davis, Smith and Marsdsen, 2004). There are ten vocabulary items on the GSS, which are a sub-set of a test originally developed by Thorndike and Lorge in the 1940's for use on Gallup surveys. The actual items are not available, "to minimize the admittedly small possibility that some form of publicity would affect the public's knowledge of the words…" (Appendix D, page 2028). On each item, a target word is given and the respondent selects the word from a set of five other words that is closest in meaning to the target word. The two items are statistically related (i.e., the Pearson chi-square test of independence yields $X^2 = 9.36$, $df = 1$, $p < .01$).

Table 14.1 Cross-classification of vocabulary items *A* and *C* from the 2004 General Social Survey (Davis, Smith, and Marsdsen, 2004)

	Item C		
Item A	*Correct*	*Incorrect*	*Total*
Correct	259	702	961
Incorrect	32	162	194
Total	291	864	1155

Odds ratio

The odds ratio is a symmetric measure for 2×2 tables. The odds ratio equals the ratio of two odds and an odds is the ratio of two probabilities. For example, the estimate of the probability that an individual answers item A correctly given a correct answer to C equals the proportion $259/291 = .89$, and the estimate of the probability of answering A incorrectly given a correct answer to C equals $32/291 = .11$. The ratio of these two estimated conditional probabilities equals the *odds* that an individual correctly responds to item A given a correct answer to item C, which equals $(259/291)/(32/291) = 259/32 = 8.09$. Likewise, the odds that an individual correctly answers item A given an incorrect answer to item C equals $702/162 = 4.33$. Item A was a relatively easy item (i.e., 83.2% of the $n = 1155$ respondents answered it correctly), so it is not surprising that the two odds of a correct response to item A are both greater than 1.

The odds ratio in our example equals $8.09/4.33 = 1.87$, which means that the odds of a correct response to item A given a correct answer to item C is 1.87 times the odds of a correct answer to item A given an incorrect answer to item C. Since odds ratios are symmetric, our odds ratio can also be interpreted as the odds of a correct response to item C given a correct answer to item A is 1.87 times the odds of a correct answer to C given an incorrect response to A. A correct answer to item A is more likely when an individual is also correct on item C; however, the *probability* is not 1.87 times larger. Odds ratios deal with how many times one *odds* is relative to another.

For generality, let n_{ij} equal the frequency in the (i, j) cell of a cross-classification of two variables. The estimate of a population odds ratio $\hat{\gamma}_{ii',jj'}$ is a 'cross-product ratio' of a 2×2

table; that is,

$$\hat{\gamma}_{ii',jj'} = \frac{n_{ij}/n_{i'j}}{n_{ij'}/n_{i'j'}} = \frac{n_{ij}n_{i'j'}}{n_{i'j}n_{ij'}}, \qquad (1)$$

where i and i' are distinct categories of one variable, and j and j' are distinct categories of the other variable. Note that the odds ratio does not depend on the margins of the table (i.e, the univariate marginal distributions of the variables), but only depends on the values within cells of the table (i.e., the joint distribution of the two variables).

A single odds ratio characterizes the association between two dichotomous variables; however, for larger cross-classifications, a set of odds ratios is needed to completely characterize the relationship between two categorical variables. For example, consider the cross-classification of career choice by gender of $N = 600$ high-school seniors from the High School and Beyond data set (Tatsuoko and Lohnes, 1988) given in Table 14.2. If I and J equal the number of rows and columns respectively, the minimum number of odds ratios needed equals $(I - 1)(J - 1)$ (i.e., the degrees of freedom for testing the independence in a two–way table). Given an appropriate set of odds ratios, all other odds ratios can be found from these. For example, using the data from Table 14.2, the odds ratio of girls (versus boys) choosing clerical versus craftsman equals $48(36)/(2(3)) = 288.00$, the odds ratio for craftsman versus farmer equals $3(9)/(36(2)) = .375$, and the odds ratio for clerical versus farmer equals the product of these two, $288.00(.375) = 108$.

If there is a natural reference or control category for both variables, a basic set of odds ratios that completely captures the association in an $I \times J$ table is set where all odds ratios are formed using the reference cell. When the variables are ordinal, a logical basic set of odds ratios are those formed using adjacent categories.

Table 14.2 Cross-classification of intended career by gender of 600 seniors in the High School and Beyond data set (Tatsuoko and Lohnes, 1988). The row scores are from correspondence analysis.

		Gender		
Career choice	*Row scores*	*Boy 1*	*Girl 0*	*Total*
Clerical	(−0.83)	2	48	50
Craftsman	(0.94)	36	3	39
Farmer	(0.73)	9	2	11
Homemaker	(−0.85)	1	32	33
Laborer	(0.59)	9	3	12
Manager	(0.31)	14	9	23
Military	(0.67)	15	4	19
Operative	(0.51)	17	7	24
Professional 1	(−0.13)	63	98	161
Professional 2	(0.07)	46	48	94
Proprietor	(0.18)	12	10	22
Protective	(1.09)	9	0	9
Sales	(−0.24)	4	8	12
School	(−0.56)	3	14	17
Service	(−0.78)	2	27	29
Technical	(0.59)	27	9	36
Not working	(−0.02)	4	5	9
Total		273	327	600

Correlation

The correlation is the Pearson product moment correlation between the two variables where scores have been assigned to the categories. Suppose that we assigned scores or numerical values to categories of the rows and columns, denoted respectively as u_i and v_j. The correlation for an $I \times J$ table equals

$$r = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(u_i - \bar{u})(v_j - \bar{v})}{\sqrt{\sum_{i=1} n_{i+}(u_i - \bar{u})^2}\sqrt{\sum_{j=1}^{J} n_{+j}(v_j - \bar{v})^2}}, \qquad (2)$$

where $\sum_i n_{ij} = n_{+j}$, $\sum_j n_{ij} = n_{i+}$, $\bar{u} = \sum_i n_{i+}u_i$, and $\bar{v} = \sum_j n_{+j}v_j$. For 2×2 tables, where $u_1 = v_1 = 1$ and $u_2 = v_2 = 0$, the correlation simplifies to the "phi coefficient"

$$r = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1+}n_{+1}n_{2+}n_{+2}}}. \qquad (3)$$

Returning to our vocabulary example

$$\begin{aligned} r &= \big(259(162) - (702)(32)big) \\ &\quad \times \big((259 + 702)(259 + 32) \\ &\quad \times (162 + 702)(162 + 32)\big)^{-1/2} \\ &= 0.9 \end{aligned}$$

which subjectively seems like a small value. The correlation depends both on values in a cross-classification (i.e., joint distribution) as well as margins of a table (i.e., the univariate distributions). As a consequence, $|r|$ can only equal 1 if the margins are equal.

In 2×2 tables, the choice of scores is arbitrary (i.e., the correlation is invariant with respect to linear transformations); however, the choice of category scores matters for larger tables. For Table 14.2, there is no natural ordering of the careers so choosing scores for the rows would appear to make computing r problematic. An upper bound for the maximum possible correlation is $\max(|r|) \leq \sqrt{X^2/n_{++}}$, where X^2 equals the Pearson chi-square test statistic for independence. Equality holds when the number of rows or columns equals 2, which is the case for both of our examples. For Table 14.1 where $X^2 = 9.36$, $r = \sqrt{9.3638/1155} = .09$, and for Table 14.2 where $X^2 = 170.9121$, $r = \sqrt{170.9121/600} = .53$. Correctly interpreting the latter correlation requires knowing the scores (implicitly) used to compute the correlation. The scores used for the career choice by gender correlation, which are given in Table 14.2, were found by performing a simple correspondence analysis. The largest possible correlation is always given by the category scores of the first component from correspondence analysis. The larger scores in our example are associated with traditionally male careers (e.g., protective, laborer), lower scores with traditionally female careers (e.g., homemaker, clerical), and those near 0 are more gender neutral (e.g., professional).

Correlation or odds ratios?

Using correlations for categorical variables presumes an underlying continuum where the observed variables have been measured discretely. This was the position taken by Karl Pearson, and he further assumed that the underlying distribution of variables was bivariate normal. Yule, on the other hand, took the position that some variables are clearly nominal (e.g., death due to smallpox) and that association between categorical variables should be measured by odds ratios or functions of them. The different views of Pearson and Yule lead to two distinct lines of model development for discrete data, as well as to very contentious and nasty exchanges (see Agresti, 2002; 2007).

For most of the statistical models presented in this chapter, odds ratios are functions of model parameters and do not require assuming underlying continua. The psychometric models discussed in this chapter assume an underlying continuum, which suggests that correlations may be more useful as a measure of association; however, this is not the case. To add to the irony, the model for data implied by underlying bivariate normality where variables are measured discretely is a model with association parameters that are most naturally interpreted in terms of odds ratios; however, correlations are functionally related to the association parameters (Goodman, 1981).[1]

GENERALIZED LINEAR MODELS

Generalized linear models (GLM) were introduced by Nelder and Wedderburn (1972) and provide a unification of a wide class of regression models. In the GLM framework, the distribution of the response variable need not be normal but any member of the exponential dispersion family of distributions. Furthermore, the relationship between the mean of the response variable and a linear function of explanatory or predictor variables can be nonlinear. A GLM consists of three components: a random component, a systematic component, and a link function. For more detailed descriptions of GLMs see Dobson (1990), Fahrmeir and Tutz (2001), and Lindsey (1997), McCullagh and Nelder (1983), and specifically for categorical data see Agresti (2002; 2007).

The random component

The random component of a model is specified by identifying the response variable and assuming a distribution for it. The distribution must be in the exponential dispersion family of distributions, which is very general and includes the normal, gamma, beta and others. The two distributions used in this chapter are the Poisson and binomial.

Dichotomous variables are very common response variables. For example, an individual may correctly answer item A on the GSS vocabulary test or a person may choose option A over option B when presented a pair of options. The number of times that an event occurs, y, out of n possible independent cases or trials is a bounded count (i.e., y can be at most n). For dichotomous variables we will use the binomial distribution

$$P(Y_i = y) = \binom{n}{y} \pi_i^y (1 - \pi_i)^{n-y}, \qquad (4)$$

where $P(Y_i = y)$ is the probability that $Y_i = y$; $y = 0, 1, \ldots, n$; n is the number of independent trials or cases for which the event could have occurred; and π_i is the probability that the event occurred on a specific trial. The index i could represent an individual. For example, when $n = 1$, Y_i would equal 1 if person i correctly answers item A and 0 otherwise. Alternatively, when $n > 1$, Y_i could equal the number of individuals who answer item A correctly out of n individuals who gave an answer to item A. The index i could also represent a particular situation. For example, if object pair A and B was presented to n individuals, Y_i could equal the number of times object A is chosen.

The mean of the binomial distribution is $\mu_i = n\pi_i$, and the variance is $n\pi_i(1 - \pi_i)$, which depends on the mean. In GLM terminology, the logarithm of the odds $\log(\pi_i/(1 - \pi_i))$ is the "natural parameter." Interest is typically focused on π_i and models for π_i are specified.

The number of trials for binomial random variables can be as small as 1, in which case the distribution of Y is Bernoulli, but in others, n may be so large that counts are virtually unbounded. When counts are not bounded, the Poisson distribution is often a good model for the distribution of response variables. The Poisson distribution is

$$P(Y_i = y) = \frac{\mu^y e^{-\mu_i}}{y!}, \qquad (5)$$

where values of the response are nonnegative integers (i.e., $y = 0, 1, 2, \ldots$), and μ_i is the mean of the distribution. The variance of a Poisson distribution equals the mean. The natural parameter for the Poisson distribution is $\log(\mu_i)$.

The systematic component

The systematic component of a GLM consists of a linear function of explanatory variables

$$\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}, \qquad (6)$$

where x_{ki} is the value on predictor or explanatory variable k for individual or case i, and the β_ks are the unknown regression coefficients, which are considered fixed in the population. Although the systematic component must be linear in the parameters, it can model nonlinear patterns. For example $x_{3i} = x_{1i}^2$ or $x_{3i} = x_{1i}x_{2i}$. There are no restrictions on the nature of the explanatory variables. For example, they could be dummy or effect codes for qualitative variables, as well as numerical values for metric variables. In all except one class of models discussed in this chapter, linear predictors suffice. The exception is for the two parameter logistic model for item response data.

When analyzing counts in cross-classifications of categorical variables, tables sometimes have *structural zeros*. These are cells where the probability of an observation is zero. For example, in a paired comparison experiment, individuals are never asked to compare an object with itself. In the cross-classification of the number of times objects are chosen (rows) versus objects not chosen (columns), there are no values along the diagonal (e.g., see Table 14.5). Such cells can be handled by putting any number in the empty cells and defining an indicator variable for each of these cells (i.e., $x = 1$ for the empty cell and $x = 0$ for all other cells).

When the indicator variable is included in the linear predictor, one parameter is estimated for the cell, which causes the fitted value to equal the value input for the empty cell. This has the effect of essentially removing the cell from the analysis. This approach is used most often to fit models that exclude diagonal elements from square tables or to deal with anomalous data points (Agresti, 2002; 2007; Fienberg, 1985). This methodology is illustrated in the section 'Poisson regression models for counts' and is used implicitly in the section 'Discrete choice/random utility models'.

The link function

The third component of a GLM is the link function $g(\cdot)$, which connects the random component $E(Y_i) = \mu_i$ with the systematic component η_i; that is

$$g(\mu_i) = \eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}. \quad (7)$$

An important consideration when choosing the link function is to choose one that ensures that fitted (predicted) values stay within the range of possible values for the response variable. For example, probabilities must be in the range of 0 to 1 and counts must be non-negative; however, the linear predictor could take on values outside the permissible range for the outcome variable. For count variables, the natural logarithm, $\log(y)$, keeps counts non-negative and is also the "canonical" link function. Canonical link functions are statistically advantageous, as seen later in this chapter.

For dichotomous variables, common choices of link functions for probabilities are the inverses of cumulative distribution functions of continuous variables, which can take on values from 0 to 1. A cumulative distribution function equals $P(Y \leq y) = F(y)$, and using the inverse as a link function gives us

$$F^{-1}(\pi_i) = \eta_i. \quad (8)$$

Three common distributions used as link functions are plotted in Figure 14.1: The standard normal, logistic, and extreme value or Gumbel distributions. For the normal and logistic distributions, probabilities are symmetric around $\eta_i = 0$. Specifically, for $\eta_i \geq 0$, the rate at which probabilities increase toward 1 as a function of η_i equals the rate at which probabilities decrease toward 0 when

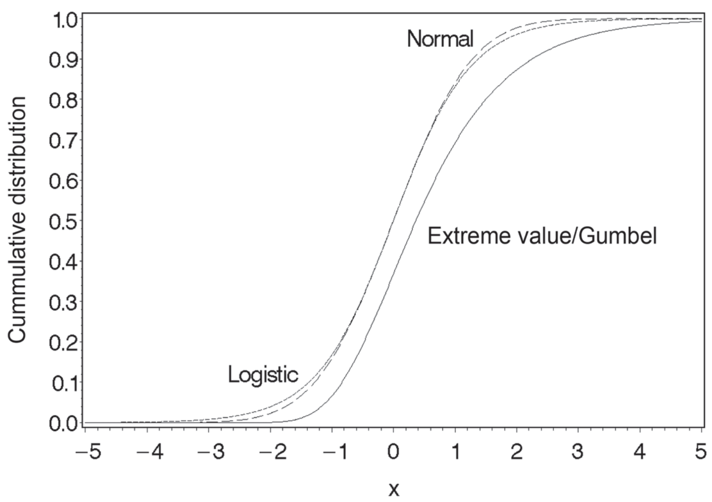

Figure 14.1 Cumulative distribution functions that are often used as link functions: the standard normal distribution, the logistic with dispersion parameter equal to .625, and the extreme value distribution.

$\eta_i \leq 0$. For the extreme value distribution, the rate of increase is not the same as the rate of decrease.

The logit link results from using the logistic distribution, which is the log of the odds

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right). \quad (9)$$

The probit link results from using the standard normal distribution; that is, $\Phi^{-1}(\pi_i)$ where Φ^{-1} is the inverse of the cumulative standard normal. As can be seen from Figure 14.1, the normal and the logistic are very similar[2]. When using the extreme value distribution, the link is $\log(-\log(\pi_i))$ and is known as the "complementary log-log link."

LOGISTIC REGRESSION FOR DICHOTOMOUS VARIABLES

Basic modeling of dichotomous data is illustrated in this section including discussions of model evaluation, statistical inference, and interpretation. Much this also pertains to Poisson regression models, which are illustrated later.

The model components

In this and the next section, the responses to item A from the 2004 vocabulary section of the General Social Survey (Davis, Smith, and Marsdsen, 2004)) are used for illustration. Since the answers to item A are dichotomous, the response variable was coded as

$$Y_i = \begin{cases} 1 \text{ a correct answer} \\ 0 \text{ an incorrect answer.} \end{cases}$$

We assume that Y follows a binomial distribution and use the canonical link function, the logit.

Assuming that there is a latent variable "vocabulary knowledge," which underlies the responses to item A, this latent variable should also influence the responses to the other vocabulary items. In test development, item response functions are often studied by fitting logistic regression models to a target item using a "rest-score" as a predictor variable. A rest-score is the sum score of all the items except for the one being treated as the response variable (Junker and Sijtsma, 2000). Other variables can be included to ascertain whether responses differ due to gender, race, or other variables (Swaminathan and Rogers, 1990). For the systematic component in our example, we consider the respondents' rest-score on four other vocabulary items (i.e., items C, D, E, and F) and the highest degree earned by the respondent. We expect that earning a higher degree is indicative of a higher level of vocabulary knowledge. After some exploratory analyses, highest degree was coded into three categories: no degree (less than a 6th-grade education), elementary-school degree (completed at least 6th grade but not high school), and high-school degree (completed at least 12th grade).

Putting the three components together yields

$$\log\left(\frac{P(Y_i = 1|\text{restscore}_i, \text{degree}_i)}{P(Y_i = 0|\text{restscore}_i, \text{degree}_i)}\right) = \beta_0 + \beta_1(\text{restscore}_i) + \beta_2(\text{hs}_i) + \beta_3(\text{primary}_i). \quad (10)$$

where $P(Y_i = 1|\text{restscore}_i, \text{degree}_i)$ equals the probability that respondent i correctly answered item A given that their rest-score equals "restscore$_i$," and highest degree earned, "degree$_i$" is coded as

$$\text{hs}_i = \begin{cases} 1 \text{ High school} \\ 0 \text{ otherwise} \end{cases}$$

$$\text{primary}_i = \begin{cases} 1 \text{ Primary} \\ 0 \text{ otherwise} \end{cases}.$$

Implicitly, the regression coefficient for those with no degree equals 0.

Model (10) was fit to the data using maximum likelihood estimation, which is standard for GLMs. The observed number of correct and incorrect responses are reported in Table 14.3 along with the observed proportions. The fitted or predicted values in this table are from a slightly revised version of Model (10) described later.

Table 14.3 Observed and fitted values for logistic regression model of correct and incorrect answers to item *A* from the vocabulary section of the 2004 GSS, where the fitted values are from equation (17).

Rest-score	*Highest degree*	*Number correct* y_i	*Predicted correct* $\hat{y}_i$	*Number incorrect* y_i	*Predicted incorrect* $\hat{y}_i$	*Observed proportion* p_i	*Fitted probability* $\hat{\pi}_i$
0	0	2	3.16	6	4.84	0.25	0.40
1	0	11	11.83	11	10.17	0.50	0.54
2	0	25	22.24	8	10.76	0.76	0.67
3	0	29	29.87	9	8.13	0.76	0.77
4	0	4	3.47	0	0.53	1.00	0.87
0	1	5	6.32	8	6.68	0.38	0.49
1	1	19	17.56	9	10.44	0.68	0.63
2	1	79	75.69	22	25.31	0.78	0.75
3	1	276	281.13	58	52.87	0.83	0.84
4	1	94	93.15	9	9.86	0.91	0.90
0	2	5	4.05	2	2.95	0.71	0.58
1	2	3	4.96	4	2.04	0.43	0.71
2	2	45	41.43	6	9.57	0.88	0.81
3	2	227	230.09	33	29.91	0.87	0.88
4	2	137	136.05	9	9.95	0.94	0.93

Table 14.4 Estimated parameters, standard errors, and test statistics for logistic regression models where word *A* is the response variable. The values on the left are those when highest degree earned was treated nominally and those on the right when highest degree earned was treated as a metric variable.

			Degree as nominal		*Degree as metric*	
Effect	*Parameter*	*df*	*Estimate*	*s.e.*	*Estimated*	*s.e.*
Intercept	β_0	1	−0.4071	0.2739	−0.4242	0.2474
Rest-score	β_1	1	0.5767	0.0882	0.5753	0.0876
Highest degree						
High school or more	β_2	1	0.7266	0.2748	0.3694	0.1306
Primary	β_3	1	0.3382	0.2505		
None		0	0.0000	0.0000		

Interpretation

The parameter estimates and corresponding standard errors for (10) are given on the left side of Table 14.4. The parameters are most naturally interpreted in terms of odds ratios. For our model, the odds of answering item A correctly for a fixed value of highest degree completed and two levels of rest-score, say $x+1$ and x, equal

$$\frac{P(Y_i = 1|x+1, \text{degree}_i)}{P(Y_i = 0|x+1, \text{degree}_i)} = \exp\left[\beta_0 + \beta_1(x+1) + \beta_2(\text{hs}_i) + \beta_3(\text{primary}_i)\right] \quad (11)$$

$$\frac{P(Y_i = 1|x, \text{degree}_i)}{P(Y_i = 0|x, \text{degree}_i)} = \exp\left[\beta_0 + \beta_1(x) + \beta_2(\text{hs}_i) + \beta_3(\text{primary}_i)\right], \quad (12)$$

respectively. The ratio of the odds (11) and (12) is an odds ratio, and equals $\exp(\beta_1)$. In our example, $\exp(\hat{\beta}_1) = \exp(0.5753) = 1.78$; that is, for a given degree, the odds of answering item A correctly is 1.78 times larger than the odds when the rest-score is one unit smaller.

With respect to the nominal variable, the odds ratios for highest degree earned are found by taking the ratios of pairs of the following odds:

$$\frac{P(Y_i = 1|\text{restscore}_i, \text{high school})}{P(Y_i = 0|\text{restscore}_i, \text{high school})} = \exp\left[\beta_0 + \beta_1(\text{restscore}_i) + \beta_2\right] \quad (13)$$

$$\frac{P(Y_i = 1|\text{restscore}_i, \text{primary})}{P(Y_i = 0|\text{restscore}_i, \text{primary})} = \exp\left[\beta_0 + \beta_1(\text{restscore}_i) + \beta_3\right] \quad (14)$$

$$\frac{P(Y_i = 1|\text{restscore}_i, \text{none})}{P(Y_i = 0|\text{restscore}_i, \text{none})} = \exp\left[\beta_0 + \beta_1(\text{restscore}_i)\right]. \quad (15)$$

Using the parameter estimates on the left-side of Table 14.4, the odds ratio for answering *A* correctly when completing primary school versus no degree equals the ratio of (14) and (15), which is $\exp(\hat{\beta}_3) = \exp(.3382) = 1.40$. The odds ratio for completing high school relative to elementary school equals the ratio of (13) and (14), which is $\exp(\hat{\beta}_2 - \hat{\beta}_3) = \exp(0.3384) = 1.46$.

The near equality between of the estimated odds ratio for elementary versus none and that for high school versus elementary (i.e., 1.40 and 1.46) suggest that the odds ratios may be equal for adjacent levels of highest degree earned. Restrictions can be placed on the parameters to make the two odds ratios equal; that is, $\exp(\beta_2 - \beta_3) = \exp(\beta_3)$ or equivalently $\exp(\beta_2) = \exp(2\beta_3)$. In other words, rather that treating highest degree earned as a nominal variable, we could treat it as a metric variable and assign equally spaced scores to the categories as follows

$$\text{degree}_i = \begin{cases} 0 & \text{no degree} \\ 1 & \text{elementary school degree.} \\ 2 & \text{high school degree} \end{cases} \quad (16)$$

Using degree as a metric variable yields

$$\log\left(\frac{P(Y_i = 1|\text{restscore}_i, \text{degree}_i)}{P(Y_i = 0|\text{restscore}_i, \text{degree}_i)}\right) = \beta_0 + \beta_1(\text{restscore}_i) + \beta_2(\text{degree}_i). \quad (17)$$

The parameter estimates for (17) are reported on the right side of Table 14.4, and the fitted counts and probabilities are in Table 14.3. The fitted odds ratio of correctly answering item *A* for adjacent categories of degree equals $\exp(0.3694) = 1.45$. Later we will formally test whether imposing the restriction on the parameters for degree in (10) significantly reduced the goodness-of-fit of the model.

Probabilities are often easier to understand than odds ratios. Logit models can be re-written as models for probabilities, because there is a one-to-one relationship between odds and probabilities

$$P(Y_i = 1|\eta_i)/(1 - P(Y_i = 1|\eta_i)) = \exp(\eta_i)$$

$$P(Y_i = 1|\eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}. \quad (18)$$

For our example, the model for the probabilities equals

$$P(Y_i = 1|\text{restscore}_i, \text{degree}_i) = \frac{\exp\left[\beta_0 + \beta_1(\text{restscore}_i) + \beta_2(\text{degree}_i)\right]}{1 + \exp\left[\beta_0 + \beta_1(\text{restscore}_i) + \beta_2(\text{degree}_i)\right]}. \quad (19)$$

Using the estimated parameters, the fitted probabilities are plotted in Figure 14.2 as a function of the rest-scores with a separate curve for each degree. Since the parameter for rest-score is positive, $\hat{\beta}_1 = 0.58$, the curves monotonically increase. The effect of degree can be seen by the horizonal shift of the curves. Since there is no interaction between rest-score and highest degree earned, the curves are parallel (i.e., in Figure 14.2, line segments $a = c$ and $b = d$). Since equally spaced scores were used for degree, the difference between the curves for high-school and elementary is the same as the difference between those for elementary and none (i.e., in Figure 14.2, line segments $a = b$ and $c = d$).

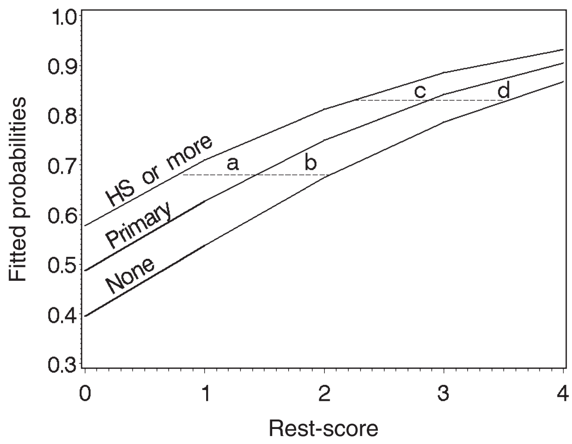

Figure 14.2 Fitted probabilities from (17) of responding correctly to item *A* of the 2004 GSS vocabulary test as a function of rest-score with a separate curve for each degree. The horizontal dashed line segments show characteristics of a model with no interaction and equally spaced scores for a discrete explanatory variable.

Model evaluation

Model goodness-of-fit can be assessed by computing either Pearson's χ^2 or a likelihood ratio statistics G^2, which equal

$$X^2 = \sum_i \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \qquad (20)$$

and

$$G^2 = 2\sum_i y_i \ln(y_i/\hat{y}_i), \qquad (21)$$

where y_i and $\hat{y}_i$ are the observed and fitted counts, and i indexes all possible cells in a cross-classification of the explanatory variables and the response variable. For the two models fit in the previous section, $X^2 = 10.02$ and $G^2 = 10.56$ for model (10), and $X^2 = 10.01$ and $G^2 = 10.58$ for model (17).

The sampling distributions of X^2 and G^2 are approximately chi-square provided that the counts in the cells are large (i.e., most $y_i \geq 5$) and that the table size is fixed (i.e., adding more observations does not increase the number cells in the cross-classification of the response and explanatory variables). These conditions are reasonable for this example.

The degrees of freedom for the model equal

$$\begin{aligned} df &= \text{(number of data points)} - \text{(number of non-redundant parameters)} \\ &= \text{(number of logits)} - \text{(number of unique parameters)}. \end{aligned} \qquad (22)$$

In model (10), there are 15 logits (i.e., the number of rows in Table 14.3) and four parameters (i.e., β_0, β_1, β_2 and β_3), which gives us $df = 15 - 4 = 11$. When comparing $X^2 = 10.02$ and $G^2 = 10.56$ to a chi-square distribution with 11 degrees of freedom, we obtain *p*-values of .44 and .48, respectively. For Model (17), which has one less parameter and $df = 12$, comparing $X^2 = 10.01$ and $G^2 = 10.58$ with a chi-square distribution with 12 degrees of freedom gives us *p*-values of .62 and .57, respectively. Both models fit the data well.

Models may be poor representations of data because the random component, the link function, or the linear predictor are not good choices for the data. To further assess model goodness-of-fit, additional and alternative statistics can be computed, including the Hosmer–Lemshow statistic, information criteria for comparing models (e.g., AIC, BIC), conditional likelihood-ratio statistics, grouping data, and receiver operating characteristic curves. In addition to global fit statistics, standardized residuals and various influence diagnostics should be examined. For these and other methods see Agresti (2002; 2007) or Hosmer–Lemeshow (2000). Possible misclassification on the response variable can be investigated by using range-of-influence statistics (Fay, 2002).

Statistical inference for parameters

Before interpreting the parameters, their significance should be assessed. Both Wald and likelihood ratio statistics can used for hypothesis tests of the parameters, such as $H_o : \beta = 0$. The fact that maximum likelihood estimates of parameters are approximately normal (i.e., $\hat{\beta} \approx \mathcal{N}(\beta, \sigma_{\hat{\beta}}^2)$) is used in Wald tests and confidence intervals. The Wald statistic for $H_o : \beta = 0$ versus $H_a : \beta \neq 0$ is

$$\text{Wald} = \left(\frac{\hat{\beta}}{(se)}\right)^2, \qquad (23)$$

where (se) is the asymptotic standard error of $\hat{\beta}$. For example, the Wald statistics for testing rest-score is $(0.5767/0.0882)^2 = 42.75$ with $df = 1$ and $p < .01$. A 95% confidence interval for the parameters equals:

$$\hat{\beta} \pm z_{(1-\alpha)/2}(se) = 0.5767 \pm 1.96(0.0882) \longrightarrow (0.40, 0.75)$$

and the 95% confidence interval for the odds ratio is

$$(\exp(\hat{\beta} - z_{(1-\alpha)/2}(se)), \exp(\hat{\beta} + z_{(1-\alpha)/2}(se))) \longrightarrow (1.49, 2.12).$$

Wald tests can be constructed to test a set of coefficients (e.g., for highest degree[3], $H_o : \beta_2 = \beta_3 = 0$); however, we will use the slightly more powerful, conditional likelihood ratio test. The conditional likelihood ratio LR test statistic compares the maximum of the likelihood for two models, one of which is nested within the other. For testing $H_o : \boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta}$ is a vector of parameters, the nested model does not include $\boldsymbol{\beta}$. The *LR* test statistic equals

$$LR = -2(\ln(L_{\mathcal{M}_0}) - \ln(L_{\mathcal{M}_1})) = G^2(\mathcal{M}_0) - G^2(\mathcal{M}_1), \qquad (24)$$

where $\ln(L_{\mathcal{M}_0})$ and $\ln(L_{\mathcal{M}_1})$ are the logarithms of the maximum of the likelihood in the nested model (i.e., $\boldsymbol{\beta} = \mathbf{0}$) and full models, respectively. The LR statistic can also be computed using $G^2(\mathcal{M}_0)$ and $G^2(\mathcal{M}_1)$, which are the likelihood ratio global goodness-of-fit statistics[4]. The sampling distribution of the LR is approximately chi-square with the degrees of freedom equal to the difference between the degrees of freedom of the two models. Even when a likelihood ratio goodness-of-fit statistic does not have an approximate chi-square distribution, the sampling distribution of the conditional LR statistic is often reasonably well approximated by a chi-square distribution. In our example, for rest-score, LR $= 42.60$, $df = 1$ and $p < .01$, and for degree earned, LR $= 8.04$, $df = 2$ and $p = .02$.

We can also use the conditional likelihood ratio test to assess whether the restrictions that we placed on the parameters for degree in (10) (i.e., $H_o : (\beta_2 - 2\beta_3) = 0$). The restricted model (17) is a special case of (10), and not surprisingly, LR $= 10.5824 - 10.5611 = .02$, $df = 1$, $p = .99$. Model (17) is preferable because it is simpler and gives a good representation of the data.

POISSON REGRESSION MODELS FOR COUNTS

In this section, we discuss and illustrate the Poisson regression model for counts, show the relationship between Poisson regression and logistic regression, and describe some special models for square tables. When explanatory variables are all categorical, Poisson regression models are often referred to as "log-linear" models.

The model

The GSS vocabulary data in Table 14.3 are used here as an example. The data in Table 14.3 consist of a cross-classification of rest-scores (five levels) by highest degree (three levels) by answer to item A (two levels). We will start by treating all variables as nominal variables and then consider rest-score and degree earned as metric variables.

The response variable y is a count, which in our example is the frequency y_{jkl} in the (j, k, l) cell of Table 14.3 where j, k and l index the levels of rest-score, degree and answer to item A, respectively. We assume that the counts y_{jkl} follow a Poisson distribution and use the canonical link function, $\log(\mu_{jkl})$.

A notational convention often used when analyzing counts in a cross-classification of categorical variables is to represent the categorical variables using an ANOVA-like form. For example, rest-score has five categories, rather than defining four dummy or effect codes, x_1, x_2, x_3, x_4 and writing the linear predictor as $\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$, it may be written more compactly as λ_j^R where $j = 1, 2, 3, 4$. The superscript symbolizes the effect (i.e., R for rest-score), and the subscript j indicates the level of rest-score.

A good starting point is the log-linear model of complete independence

$$\log(\mu_{jkl}) = \lambda + \lambda_j^R + \lambda_k^D + \lambda_l^A, \qquad (25)$$

where λ ensures that the sum of the fitting counts equals the total number of observed counts, λ_j^R, λ_k^D and λ_l^A are the marginal effects for rest-score, highest degree earned, and answer to item A, respectively. Since λ_j^R, λ_k^D and λ_l^A are in the model, the fitted margins for rest-score, degree earned, and answer to item A equal their respective observed margins. If a margin is fixed by design, the corresponding marginal effect term should be included in the model.

More interesting models are those that include interactions, such as the homogeneous association log-linear model, which includes all two-way interactions

$$\log(\mu_{jkl}) = \lambda + \lambda_j^R + \lambda_k^D + \lambda_l^A + \lambda_{jk}^{RD} + \lambda_{jl}^{RA} + \lambda_{kl}^{DA}, \qquad (26)$$

where the terms λ_{jk}^{RD}, λ_{jl}^{RA}, and λ_{kl}^{DA} represent interactions or dependencies between pairs of variables. As the interaction terms are in the model, the fitted and observed two-way margins are equal (e.g., $n_{+kl} = \hat{\mu}_{+kl}$ because λ_{kl}^{DA} is in the model).

Interactions in log-linear models are interpreted in terms of conditional or partial odds ratios. Partial odds ratios, which are the odds ratios in two-way tables of two variables conditional on all other variables, are functions of the interaction parameters. For example, using equation (26), the odds ratio for rest-scores j and j' and answers l and l' to item A given degree k equals

$$\gamma_{jj',ll'(k)} = \frac{\mu_{jkl}\mu_{j'kl'}}{\mu_{j'kl}\mu_{jkl'}} = \exp[\lambda_{jl}^{\mathrm{RA}} + \lambda_{j'l'}^{\mathrm{RA}} - \lambda_{j'l}^{\mathrm{RA}} - \lambda_{jl'}^{\mathrm{RA}}]. \qquad (27)$$

The odds ratios between rest-score and A are the same over the K levels of degree earned. An implication of the homogeneous association model is that the other partial odds ratios are homogeneous. For example,

$$\gamma_{jj',kk'(l)} = \exp[\lambda_{jk}^{RD} + \lambda_{j'k'}^{RD} - \lambda_{j'k}^{RD} - \lambda_{jk'}^{RD}] \qquad (28)$$

$$\gamma_{kk',ll'(j)} = \exp[\lambda_{kl}^{DA} + \lambda_{k'l'}^{DA} - \lambda_{k'l}^{DA} - \lambda_{kl'}^{DA}]. \qquad (29)$$

In Poisson regression models, model parameters for discrete predictor variables require location constraints for identification. Typical ones are to set one level equal to zero (e.g., $\lambda_1^R = 0$, $\lambda_{1l}^{\mathrm{RA}} = \lambda_{j1}^{\mathrm{RA}} = 0$), which corresponds to dummy coding the variable, or zero-sum constraints (e.g., $\sum_j \lambda_j^R = 0$, $\sum_j \lambda_{jl}^{\mathrm{RA}} = \sum_l \lambda_{jl}^{\mathrm{RA}} = 0$), which correspond to using effect coding.

Models (25) and (26) were fit to the data. The model of independence fails (i.e., $G^2 = 209.72$, $df = 22$, $p < .01$); however, the homogeneous association model (26) gives a good representation of the data in Table 14.3 (i.e., $G^2 = 5.86$, $df = 8$, $p = .67$). The conditional LR statistics for the variables are all significant (i.e., for $H_0 : \lambda^{\text{RD}}_{jk} = 0$, LR $= 111.61$, $df = 8$, and $p < .01$; for $H_0 : \lambda^{RA}_{jl} = 0$, LR $= 47.30$, $df = 4$, $p < .01$; and for $H_0 : \lambda^{DA}_{kl} = 0$, LR $= 8.38$, $df = 2$, $p = .02$).

The logit/log-linear model connection

For every logit model there is an equivalent log-linear (Poisson regression) model; however, not every log-linear model corresponds to a logit model. This result is one of the benefits of using the canonical link functions for the Poisson and Binomial distributions. We show the equivalence between logit and log-linear models by example.

In the final logit model for the GSS data (i.e., Equation (17)) item A was the response variable, and rest-score and degree were metric explanatory variables. Also recall that in the logit modeling, the binomial distribution was assumed for the responses to item A and the number of cases for each combination of rest-score and degree earned was treated (implicitly) as fixed (i.e., "n" in the binomial distribution). Therefore, to ensure that the rest-score by degree margin is fit perfectly by the model the term λ^{RD}_{jk} must be in the log-linear model. Numerical variables can be included in Poisson regression models. Since the homogeneous association model fits the data, a special case of it will be fit to the data using rest-scores and degree as metric variables in the interaction terms between each of these and item A; that is, we will replace λ^{RA}_{jl} by the product $\lambda^{\text{RA}}_{l}(\text{restscore}_j)$, and replace λ^{DA}_{kl} by $\lambda^{\text{DA}}_{l}(\text{degree}_k)$, which yields

$$\log(\mu_{jkl}) = \lambda + \lambda^{R}_{j} + \lambda^{D}_{k} + \lambda^{A}_{l} + \lambda^{\text{RD}}_{jk} + \lambda^{\text{RA}}_{l}(\text{restscore}_j) + \lambda^{\text{DA}}_{l}(\text{degree}_k). \tag{30}$$

The goodness-of-fit statistics for this log-linear model are $G^2 = 10.58$ and $X^2 = 10.01$ with $df = 12$, which are exactly the same as our final logit model in the previous section. Logit model (17) and log-linear model (30) are equivalent models.

The correspondence between (17) and (30) can be seen by forming logits using (30):

$$\begin{aligned}\log\left(\frac{\mu_{jk2}}{\mu_{jk1}}\right) &= \log(\mu_{jk2}) - \log(\mu_{jk1}) \\ &= (\lambda^{A}_{2} - \lambda^{A}_{1}) + (\lambda^{RA}_{2} - \lambda^{RA}_{1})(\text{restscore}_i) \\ &\quad + (\lambda^{DA}_{2} - \lambda^{DA}_{1})(\text{degree}_i) \\ &= \beta_0 + \beta_1(\text{restscore}_i) + \beta_2(\text{degree}_i),\end{aligned} \tag{31}$$

where $\beta_0 = (\lambda^{A}_{2} - \lambda^{A}_{1})$, $\beta_1 = (\lambda^{RA}_{2} - \lambda^{RA}_{1})$, and $\beta_2 = (\lambda^{DA}_{2} - \lambda^{DA}_{1})$. Note that the index i in the logit model corresponds to unique combinations of indices j and k in the log-linear model. Only terms that have an A superscript in the log-linear model appear in the logit model.

When there is a single dichotomous response variable, using the logistic regression formulation is preferable because there are fewer parameters and numerous regression diagnostic procedures specifically for logistic regression exist. The diagnostic procedures are especially useful when global goodness-of-fit statistics do not have approximate chi-square distributions. When the response variable has more than two levels or there are multiple response variables, such models can be fit to data as Poisson regression or log-linear models. We will use logit models when modeling choices are made between two objects but use log-linear models when modeling choices among three or more options. When we consider item response models, where each item on a test or survey is a response variable (i.e., multiple response variables), we will use log-linear and related models.

Log-linear models for square tables

Square tables are those where the row and column classifications are the same. Examples of square tables include cross-classifications

Table 14.5 The number of women college students who would prefer to spend an hour with the row celebrity over the column celebrity (Kroeger, 1992).

	Bush	*Reagan*	*Clinton*	*Blair*	*Kersee*	*Caprati*
Barbara Bush	—	64	34	62	45	52
Nancy Reagan	32	—	31	46	40	41
Hillary Clinton	62	65	—	64	54	48
Bonnie Blair	34	50	32	—	36	40
Jackee Joyner-Kersee	51	56	42	60	—	56
Jennifer Caprati	44	55	48	56	40	—

of winning (rows) and losing (columns) baseball teams where cells equal the number of times a row team wins a game against a column team (Fienberg, 1985), the movie ratings given by Roger Ebert (rows) and Gene Siskal (columns) who rate the same movies (Agresti and Winner, 1997), and choices between pairs of cars (Maydeu-Olivares and Böckenholt, 2005). The data used here are choices made by 96 women college students (Kroeger, 1992). Each subject was given pairs of female celebrities and chose which one they would prefer to talk with for an hour. The entries in Table 14.5 equal the number of students who chose the row celebrity over the column celebrity. There are no observations in the diagonal of Table 14.5.

Log-linear models designed to deal with square tables can be used to fit psychometric models to data, and they are interesting in their own right. The models described here are quasi-independence, symmetry, and quasi-symmetry. With square tables, there are often either no entries along the diagonal or the diagonal entries are very large (e.g., if Sikel and Ebert for the most part agree, then we would expect the values on the diagonal to be large). To fit these models as log-linear models requires the use of methodology for dealing with empty cells or structural zeros.

To test whether there is any association between the row and column classifications when there are empty cells, any number is input for the empty cell (e.g., 0) and a parameter is estimated for it. This parameter ensures that the fitted counts and the values input are equal. The resulting quasi-independence model is:

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \delta_i I(i = j) \quad (32)$$

where λ_i^R and λ_j^C are the marginal effects for the row and columns, $I(i = j) = 1$ if $i = j$ and 0 otherwise, and δ_i is the parameter for the ith diagonal cell.

A common hypothesis for square tables is that the table is symmetric (i.e., $\mu_{ij} = \mu_{ji}$). If a matrix is symmetric, then sum of entries in the ith row must equal the sum in the ith column; that is, the margins must be equal. The log-linear model of symmetry for tables with empty cells for the diagonal is

$$\log(\mu_{ij}) = \lambda + \lambda_i + \lambda_j + \lambda_{ij}^{RC} + \delta_i I(i = j), \quad (33)$$

where $\lambda_{ij}^{RC} = \lambda_{ji}^{RC}$, and the marginal effects are the same (i.e., no super-scripts on λ_i and λ_j).

The symmetry model is very restrictive because is specifies both the marginal and joint distributions. Marginal homogeneity can be relaxed by allowing the marginal effect terms to differ while retaining symmetry in the association. This model is known as quasi-symmetry[5]

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC} + \delta_i I(i = j), \quad (34)$$

where the interaction parameters are $\lambda_{ij}^{RC} = \lambda_{ji}^{RC}$, and λ_i^R does not necessarily equal λ_i^C.

The models of quasi-independence and symmetry fail to fit the celebrity data;

however, the quasi-symmetry model gives good representation of the data in Table 14.5 ($G^2 = 9.41$, $df = 10$, $p = .49$). As a preview to psychometric models for choice data discussed in following sections, note that odds for $i \neq j$ based on the model of quasi-symmetry (34) equal

$$\frac{\mu_{ij}}{\mu_{ji}} = \frac{\exp[\lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC}]}{\exp[\lambda + \lambda_j^R + \lambda_i^C + \lambda_{ji}^{RC}]}$$
$$= \exp[(\lambda_i^R - \lambda_i^C) - (\lambda_j^R - \lambda_j^C)]$$
$$= \exp[\psi_i - \psi_j], \qquad (35)$$

which is the Bradley–Terry–Luce model discussed later. At the end of the chapter, we discuss how the Rasch model can be fit as a model of quasi-symmetry.

PSYCHOMETRIC FRAMEWORK

A latent continuum is often assumed to govern the choices made by individuals among objects. The continuum may be utility, preference, skill, achievement, ability, excellence, or some other construct upon which comparisons or decisions are made. Models for choices generally make similar assumptions about the values of objects on a latent continuum. Specifically, let S_{ik} be the “subjective” impression or value on the underlying construct for individual i and object k. This subjective value S_{ik} is typically assumed to be random and thus accounts for variation in responses. Variation in responses may be due to differences made by a single individual on repeated occasions, differences between individuals, or both. The model for S_{ik} used for all our psychometric models is

$$S_{ik} = \psi_{ik} + \epsilon_{ik} \qquad (36)$$

where ψ_{ik} is assumed to be a fixed value on the continuum for object or item k and individual i, and ϵ_{ik} is typically assumed to be random. We further assume that $E(\epsilon_{ik}) = 0$, the variance var(ϵ_{ik}) equals a constant for all i and k, and ϵ_{ik} is independent over objects and individuals.

The psychometric models discussed in the remainder of this chapter can all be fit as a GLM, and Equation (36) provides an unifying framework for our psychometric models. In different psychometric models, assumptions regarding the components of S_{ik} differ as well as how individuals' values on S_{ik} effect observed choices and decisions. These assumptions determine the random, link, and systematic components for each model.

DISCRETE CHOICE/RANDOM UTILITY MODELS

Models for choices between pairs of objects are discussed in the next section, and models for choices among a set of objects are discussed on p. 327. Models for paired comparisons are special cases of those for choices among a set of objects where the set only contains two choice options.

Models for paired comparisons

The models described in this section apply to cases where there is a set of objects (e.g., cars, colleges or universities, celebrities) and an individual is given the choice between two of them on a single trial. Two common psychometric models for such data are Thurstone's law of comparative judgment (Thurstone, 1927; see also Torgenson, 1958) and the Bradley–Terry–Luce (BTL) choice model (Bradley and Terry, 1952; Luce, 1959). Fienberg and Larntz (1976; see also Fienberg, 1985) showed how the BTL model can be fit as quasi-symmetric and quasi-independence log-linear models; however, in this section, we show how the BTL and Thurstone Case V models can be fit as logit and probit models, respectively (e.g., Agresti, 2002; Powers and Xie, 2000).

In both the BTL and Thurstone's model, objects are assumed to be compared with respect to their values of S_{ik}; however, on each occasion the value of the object on the latent continuum may randomly vary around its mean value (i.e., $E(S_{ik}) = \psi_{ik}$). In Thurstone's law of comparative judgement, which builds on Weber's law and Fechner's work in psychophysics, Thurstone proposed "some kind of process in us by which we react differently to several specimens…" (Thurstone, 1927: 274.).

The assumed comparison process involves comparing two objects' values S_{ik} and $S_{ik'}$, and the object with the larger value is chosen. Our measurement model is:

$$Y_{i,kk'} = \begin{cases} 1 \text{ if } (S_{ik} - S_{ik'}) > 0 \\ \qquad \text{(i.e., } k \text{ chosen over } k') \\ 0 \text{ if } (S_{ik} - S_{ik'}) \leq 0 \\ \qquad \text{(i.e., } k' \text{ chosen over } k). \end{cases} \tag{37}$$

The response variable is assumed to be a binomial random variable. Data can be analyzed on un-aggregated or individual level (i.e., $Y_{i,kk'}$), in which case the response is a 0/1 variable with $n = 1$. When the data are analyzed on collapsed data (i.e., $Y_{kk'} = \sum_i Y_{i,kk'}$), the response variable is how many times object k was selected over k' with n equal to the number of individuals who compared the two objects.

The distribution of ϵ_{ik} in Thurstone's model is assumed to be normal, and the distribution in the BTL model is assumed to be Gumbel. These distributional assumptions determine the link function. When the distributions of ϵ_{ik} and $\epsilon_{ik'}$ are symmetric, (e.g., normal or logistic), the probability that k is selected over k' equals

$$\begin{aligned} P(Y_{i,kk'} = 1) &= P(S_{ik} - S_{ik'} > 0) \\ &= P(\epsilon_{ik} - \epsilon_{ik'} > -(\psi_{ik} - \psi_{ik'})) \\ &= P(\epsilon_{ik} - \epsilon_{ik'}) \leq (\psi_{ik} - \psi_{ik'})). \end{aligned} \tag{38}$$

If ϵ_{ik} and $\epsilon_{ik'}$ are both $\mathcal{N}(0, \sigma^2)$ and independent, then $(\epsilon_{ik} - \epsilon_{ik'}) \sim \mathcal{N}(0, 2\sigma^2)$. The right side of (38) is found from a cumulative normal distribution and the link function is the probit. If ϵ_{ik} and $\epsilon_{ik'}$ are independent random variables from an extreme value (Gumbel) distribution with mean 0 and equal scale parameters, then the distribution of their difference $\epsilon_{ik} - \epsilon_{ik'}$ follows a logistic distribution (Nadarajah, 2007; Train, 2003; Yellott, 1977) and the link function is the logit.

The last component of the GLM that needs to be specified is the systematic component. The ψ_{ik}s are the parameters of model, so the systematic component are indicator variables for the objects in the pair compared. When objects k and k' are compared, the values of the explanatory variables are 1 for object k, -1 for object k', and 0 for all of the others. Our model for paired comparisons is a GLM with the Binomial distribution for the random component, the link is either the logit or probit, and the linear predictor equals object indicators. Putting these three components together yields the following GLM:

$$g(P(Y_{i,kk'} = 1)) = \psi_{ik} - \psi_{ik'}, \tag{39}$$

where $g(.)$ is either the probit or logit link. Note that there is no intercept in this model. When the link is the logit (i.e, the BTL model), Model (39) is the same Equation (35), which was derived from the log-linear model of quasi-symmetry.

Both the BTL and Thurstone's models were fit to the celebrity choice data in Table 14.5 as logit and probit models, respectively. Both models give a good representation of the data (i.e., $G^2 = 9.41$ for the BTL and $G^2 = 9.42$ for Thurstone, each with $df = 10$). The estimated parameters and standard errors for the BTL and Thurstone Case V models are reported in Table 14.6. The order of the celebrities in Table 14.6 corresponds to the most preferred to the least preferred celebrity. Hillary Clinton was most likely to be selected and Nancy Reagan and Bonnie Blair were the least. The estimated parameters from the two models are nearly perfectly correlated, but those from the BTL model are slightly larger than those from Thurstone's model, because the standard normal distribution has a smaller variance than the standard logistic distribution (Powers and Xie, 2000).

The BTL model is often presented as

$$P(Y_{kk'} = 1) = \frac{\pi_k}{\pi_k + \pi_{k'}}, \tag{40}$$

Table 14.6 Estimated parameters and standard errors from the Thurstone Case V model and the Bradley–Terry–Luce model fit to the celebrity data.

	Thurstone Case V		*Bradley–Terry–Luce*		
Celebrity	$\hat{\psi}$	*s.e.*	$\hat{\psi}$	*s.e.*	$\hat{\pi}$
Hillary Clinton	0.2224	0.0746	0.3577	0.1203	.24
Jackee Joyner-Kersee	0.0969	0.0743	0.1560	0.1192	.19
Barbara Bush	0.0620	0.0743	0.0991	0.1190	.18
Jennifer Capirati	0.0000	—	0.0000	—	.16
Bonnie Blair	−0.2268	0.0746	−0.3637	0.1201	.11
Nancy Reagan	−0.2357	0.0746	−0.3783	0.1202	.11

where π_ks are probabilities; therefore, we also transformed the $\hat{\psi}_k$ from the BTL into probabilities (i.e., $\hat{\pi}_k = \exp(\hat{\psi}_k)/\sum_h \exp(\psi_h)$). The estimated probabilities are reported in Table 14.6.

If the BTL or Thurstone's model fail to give a good representation of data, possible reasons include un-modeled heterogeneity between individuals, a lack of independence of a person's judgments over pairs, or the homogeneity of variance assumption. Models for the ψ_{ik}s can be proposed that include the measures on individuals or characteristics of the choice objects. Models for ψ_{ik} that include both measures on individuals and choice objects are illustrated in the next section. See Maydeu-Olivares and Böckenholt (2005) for further extensions for Thurstone's model.

Multinomial discrete choice models

Some studies involve a choice among a set of possible options. For example, in the High School and Beyond data, students were asked what career they would like. In this section, McFadden's conditional multinomial choice model is presented (McFadden, 1974; see also Agresti, 2002; Fahrmeir and Tutz, 2001; Long, 1997; Powers and Xie, 2000).

The measurement model for multicategory choices is

$$Y_{ik} = \begin{cases} 1 \text{ if object } k \text{ has the largest value of} \\ \qquad S_{ik} \text{ in the choice set} \\ 0 \text{ otherwise.} \end{cases}$$

Each of the objects in the choice set has different attributes, which could be used to help predict the value ψ_{ik}. Furthermore, decision makers may be heterogeneous and observed measures of individuals' characteristics may also help to predict ψ_{ik}. Suppose that we have a set of q explanatory variable $x_{1ik}, \ldots, x_{qik}$, which may be attributes of objects, characteristics of individuals, and even interactions between them. Using these variables, our model for subjective values is

$$S_{ik} = \sum_p \beta_p x_{pik} + \epsilon_{ik}. \qquad (41)$$

If ϵ_{ik} follows an extreme value or Gumbel distribution, then the probability that individual i selects the option with the largest S_{ik} equals

$$P(Y_{ik} = 1|x_{1ik}, \ldots, x_{qik}) = \frac{\exp[\sum_p \beta_p x_{pik}]}{\sum_k \exp[\sum_p \beta_p x_{pik}]}. \qquad (42)$$

(McFadden, 1974).

Model (42) is equivalent to the following Poisson regression model

$$\log(\mu_{ik}) = \lambda + \lambda_i^{Ind} + \lambda_k^C + \sum_p \beta_p x_{pik}, \qquad (43)$$

where λ ensures that the fitted values sum to the sample size (number of individuals), λ_k^C is the marginal effect for choice option k that ensures that the fitted number of choices for k equals the observed number, and λ_i^{Ind} ensures that the fitted values for individual i sum to one. Since each person makes a single choice, the individual margin must sum to one, which necessitates the λ_i^{Ind} parameters. The λ_i^{Ind}s are nuisance parameters.

If an individual is only given a sub-set of the possible choice options, then the options not offered are structural zeros, which can be handled using indicator variables as was done earlier with the log-linear models for square tables. For simplicity, we assume that all individuals are given the same choice options.

From (43), we get the same model for probabilities as in (42) by recognizing that the fitted values from (43) range from 0 to 1 and sum to 1 (i.e., the $\mu_{ik}/n = \mu_{ik}$ are probabilities)

$$\begin{aligned}\mu_{ik} &= P(Y_i = k|x_{1ik}, \ldots, x_{qik}) \\ &= \exp[\lambda + \lambda_i^{Ind}]\exp\left[\sum_p \beta_p x_{pik}\right] \\ &= \frac{\exp[\sum_p \beta_p x_{pik}]}{\sum_h \exp[\sum_p \beta_p x_{pih})]}. \end{aligned} \tag{44}$$

The term $\exp[\lambda + \lambda_i^{Ind}]$ only depends on individual i and it is in the model to ensure that probabilities sum to one; therefore, it must equal $\sum_h \exp[\sum_p \beta_p x_{pih}]$.

The odds of choosing one object over another based on Model (42) depends on differences between values on the explanatory variables for the options and not on the differences between parameters as was the case for the logistic regression models previously discussed. For example, consider options k and k',

$$\begin{aligned}&\frac{P(Y_i = k|x_{1ik}, \ldots, x_{qik})}{P(Y_i = k'|x_{lik'}, \ldots, x_{qik'})} \\ &\quad = \exp\left[\sum_p \beta_p(x_{pik} - x_{pik'})\right]. \end{aligned} \tag{45}$$

If we add a characteristic of an individual, then $(x_{pik} - x_{pik'}) = 0$. To incorporated individual level variables in the model, we need to create dummy (or effect) codes for the choice objects and take the interaction between these dummy variables and the individual level variables as illustrated below.

To fit (42) to data as log-linear model (44), the data need to be in an expanded format; that is, instead one line of data for each individual, there needs to be K of them, one for each choice option. To incorporate individual level explanatory variables, we will use the HSB data as an example and include student's gender and achievement test scores as explanatory variables. We define dummy codes for the careers as:

$$d_k = \begin{cases} 1 \text{ for career } k \\ 0 \text{ otherwise.} \end{cases} \tag{46}$$

If gender g_i is coded as 0 for male and 1 female, the interactions $g_i d_k$ for all k should be in the design matrix. For achievement a_i, we include the interactions $a_i d_k$ for $k = 1, \ldots, K$. Table 14.7 contains the data lines (design matrix) for individual i for the model predicting career choice with mean achievement scores for each career (a career attribute), and individuals' achievement score and gender (student characteristics). When using the interactions between the dummy codes and individual level variables, different parameters are estimated for each career for students's achievement and gender.

The education required for a particular career may be a good predictor of career choice. Although not ideal, we used the average achievement scores of those who selected a particular career (i.e., $\bar{a}_k$) as a proxy for the education required for that career.

A number of models were fit to the career choices made by seniors in the High School Beyond data set[6]. Individual attributes considered were students' gender, race, and achievement test scores. Race was not a significant predictor in any model fit to the data. The final model includes mean achievement per career $\bar{a}_k$, students' achievement score a_i, and students' gender g_i as predictor variables; that is:

$$\begin{aligned}\log(\mu_{ijk}) = \lambda + \lambda_i^{Ind} + \lambda_k^c + \beta_1\bar{a}_k + \\ + \sum_{k'=1}^{16} \beta_{(1+k')} d_{k'} a_i \\ + \sum_{k'=1}^{16} \beta_{(17+k')} d_{k'} g_i. \end{aligned} \tag{47}$$

Table 14.7 Section of the data file (design matrix) needed to fit the conditional multinomial discrete choice model as a Poisson regression model where $a_i d_1$–$a_i d_{16}$ and $g_i d_1$–$g_i d_{16}$ are variables for student *i*'s achievement test score and gender, respectively.

	Career	Y_{ik} choice	x_{1ik} $\bar{a}_k$	x_{2ik} $a_i d_1$	x_{3ik} $a_i d_2$	... ...	x_{17ik} $a_i d_{16}$	x_{18ik} $g_i d_1$	x_{19ik} $g_i d_2$	... ...	x_{33ik} $g_i d_{16}$
1	Clerical	0	256.78	a_i	0.0	...	0.0	g_i	0	...	0
2	Craftsman	0	238.00	0.0	a_i	...	0.0	0	g_i	...	0
3	Farmer	0	270.23	0.0	0.0	...	0.0	0	0	...	0
4	Homemaker	0	238.60	0.0	0.0	...	0.0	0	0	...	0
5	Laborer	0	230.64	0.0	0.0	...	0.0	0	0	...	0
6	Manager	0	257.65	0.0	0.0	...	0.0	0	0	...	0
7	Military	0	249.00	0.0	0.0	...	0.0	0	0	...	0
8	Operative	0	233.10	0.0	0.0	...	0.0	0	0	...	0
9	Professional 1	0	268.31	0.0	0.0	...	0.0	0	0	...	0
10	Professional 2	1	280.97	0.0	0.0	...	0.0	0	0	...	0
11	Proprietor	0	247.50	0.0	0.0	...	0.0	0	0	...	0
12	Sales	0	267.28	0.0	0.0	...	0.0	0	0	...	0
13	School	0	262.25	0.0	0.0	...	0.0	0	0	...	0
14	Service	0	241.29	0.0	0.0	...	0.0	0	0	...	0
15	Technical	0	272.22	0.0	0.0	...	0.0	0	0	...	0
16	Not Working	0	250.00	0.0	0.0	...	a_i	0	0	...	g_i

Since $d_k = 1$ and $d_{k'} = 0$ for all $k' \neq k$, Model (47) can be written more compactly as

$$\log(\mu_{ik}) = \lambda + \lambda_i^{Ind} + \lambda_k^C + \beta_1 \bar{a}_k, + \beta_{(k+1)} a_i + \beta_{(17+k)} g_i. \quad (48)$$

Although we only have three effects in the model, it is very complex having 33 parameters (16 for achievement, 16 for gender, and 1 for education). This model was simplified by setting specific β parameters equal to 0 if they were not significant, which was done by dropping the corresponding variable from the model (e.g., $d_k a_i$ or $d_k g_i$). Other parameters that were similar in value were set equal by recoding the dummy variable for a particular interactions. For example, the parameters for $d_1 g_i$ (clerical), $d_4 g_i$ (homemaker) and $d_{14} g_i$ (service) were approximately equal. A new dummy variable was defined where $d^G_{1,4,14} = 1$ for clerical, homemaker, and service careers, and 0 otherwise (i.e., $d^G_{i,4,14} = d^G_1 + d^G_4 + d^G_{14}$). The variable $d^G_{1,4,14} g_i$ was added to the model and $d_1 g_i$, $d_4 g_i$ and $d_{14} g_i$ were deleted. The restrictions were tested using likelihood ratio tests. The final model has approximately half the number of parameters (16 versus 33).

The estimated parameter for education required for a career (i.e., $\bar{a}_k$) equals −0.1251 (s.e. = 0.0121, Wald = 107.27, $p < .01$). The values for $\bar{a}_k$ are reported in Table 14.7. The odds of choosing a particular career that requires only a one point higher score is $\exp(-0.1251) = .88$. For example, given a student's gender and their achievement test score, the odds of choosing professional 2 versus clerical equals $\exp[-.1251(280.97-256.78)] = 0.049$. The probability of choosing a career is lower the more education that a career requires.

The parameter estimates and standard errors for student achievement and gender are given in Table 14.8. The careers in the table have been ordered according to the values of parameter estimates to aid interpretation. On the left side of Table 14.8 are the estimated parameters for the students' achievement test scores. Homemaker, laborer and service have the lowest value (i.e., $\hat{\beta} = -0.0105$), which means that students with low achievement test scores tend to be more likely to choose one of these careers. The career with the highest values is professional 2 (i.e., doctors, lawyers, etcetera) with $\hat{\beta} = 0.0213$, so students with higher achievement test scores are more likely to choose this career than any of the other careers. On the right side of Table 14.8 are the estimated parameters for gender. The career craftsman is least likely to be chosen by a female ($\hat{\beta} = -2.57$); whereas

Table 14.8 Parameter estimates from final conditional multinomial logistic model fit to the career choices of high-school seniors

	Achievement			
Career	*Estimate*	*s.e.*	*Wald*	*p*
Homemaker				
Laborer	−0.0105	0.0010	106.80	< .01
Service				
Operative	−0.0056	0.0012	22.90	< .01
Clerical				
Craftsman				
Military	0.0000			
Proprietor				
School				
Not working				
Manager	0.0051	0.0011	23.08	< .01
Sales	0.0074	0.0015	23.12	< .01
Farmer	0.0107	0.0017	37.95	< .01
Professional 1	0.0154	0.0014	116.77	< .01
Technical	0.0174	0.0011	232.63	< .01
Professional 2	0.0213	0.0017	166.27	< .01

	Gender			
Career	*Estimate*	*s.e.*	*Wald*	*p*
Craftsman	−2.5708	0.6018	18.25	< .01
Farmer	−1.7046	0.7870	4.69	.03
Laborer				
Military	−1.0569	0.2589	16.67	< .01
Operative				
Technical				
Manager				
Professional 1				
Professional 2	0.0000			
Proprietor				
Sales				
Not working				
School	2.0045	0.3556	31.78	< .01
Clerical				
Homemaker	2.5366	0.2135	141.20	< .01
Service				

clerical, homemaker and service careers were most likely to be chosen by a female ($\hat{\beta} = 2.54$).

ITEM RESPONSE MODELS

Two main goals of item-response theory (IRT) are to study items on a test or questionnaire for future use on a measurement instrument and to measure individuals' values on some latent trait. Two basic IRT models are discussed in this section using GLMs and the psychometric framework presented earlier.

The specific models discussed here are the Rasch model and the two parameter logistic (2*PL*) model, which are designed for dichotomous responses and assume one underlying trait. A number of researchers have studied the connection between Rasch and log-linear models (Anderson et al. 2007; Cressie and Holland, 1983; de Leeuw and Verhelst, 1986; Kelderman, 1984, 1996; Tjur, 1982), but fewer have studied the connection between the 2*PL* model and log-multiplicative association models (LMA) (Anderson and Yu, 2007; Holland, 1993b). The LMA models are special cases of homogeneous association log-linear models (i.e., log-linear model with two–way interactions between all pairs of variables).

Although the focus is on models for dichotomous responses for one latent trait, the approach presented here, which is based on Anderson and Yu (2007; Anderson, et al. 2007) readily extends to polytomous variables and multiple correlated latent variable (e.g., Anderson, et al. 2007) with or without coviariates (Tettegah and Anderson, 2007).

Item response function

Using the model for S_{ik} given in Equation (36) (i.e., $S_{ik} = \psi_{ik} + \epsilon_{ik}$), the measurement model is

$$Y_{ik} = \begin{cases} 1 \text{ if} S_{ik} > \beta_{ik} \\ \quad \text{(e.g., yes or a correct response)} \\ 0 \text{ if } S_{ik} \leq \beta_{ik} \\ \quad \text{(e.g., no or an incorrect response).} \end{cases} \qquad (49)$$

where β_{ik} is a criterion for individual i on item k. To be identifiable, models can have either random values on the latent variable (i.e., S_{ik}) or random criteria (i.e., β_{ik}), but not both (Powers and Xie, 2000). The standard assumption is that ϵ_{ik} and hence S_{ik} is random, and that β_{ik} is fixed. In this chapter, all individuals are assumed to have the same criteria for a given item[7] (i.e., $\beta_{ik} = \beta_k$). The model for the probability of a correct response to item k is

$$P(Y_{ik}=1|S_{ik})=P(S_{ik}>\beta_k) \\ =P(\epsilon_{ik}>-\psi_{ik}+\beta_k). \quad (50)$$

Guttmans's scalogram (Torgerson, 1958) is obtained by assuming that ϵ_{ik} is fixed (e.g., $\epsilon_{ik} = 0$) and $\psi_{ik} = \psi_i$. With these assumptions, Model (50) is deterministic. For a set of K items, Guttman's scaling model has strong implications in that only a relatively small number of possible response patterns for K items are permissible. As Guttman's model rarely fits in practice, Goodman (1975; see also Fienberg, 1985) proposed a model for data not perfectly consistent with a Guttman scale. In this model, the cross-classification of items is modeled by a quasi-independence log-linear model where those response patterns not consistent with a Guttman scale are treated as structural zeros. Scale values (i.e., β_ks) for the items are computed for the response patterns consistent with a perfect Guttman scale by taking the difference between the observed number of respondents with a consistent pattern and the predicted for that response pattern based on the quasi-independence model.

More commonly, a stochastic or probabilistic item response model is formed by letting ϵ_{ik} be random[8]. If ϵ_{ik} follows a normal or logistic distribution, then $P(Y_{ik} = 1|S_{ik}) = P(\epsilon_{ik} \le \psi_{ik} - \beta_{ik})$. In the case of a logistic distribution, the implied link function is the logit link, which yields

$$P(Y_{ik} = 1|\psi_{ik}) = \frac{\exp[\psi_{ik} - \beta_k]}{1 + \exp[\psi_{ik} - \beta_k]}. \quad (51)$$

In the case of a normal distribution for ϵ_{ik}, the implied link function is probit. Many item response functions are simply logistic or probit regression models where the explanatory variable is latent. In this section, we use the canonical link function for the binomial distribution and take advantage of the connection between logit and Poisson (log-linear) regression models.

Two parameter logistic model

In psychology and education, a person's value on a latent variable is typically assumed to exist prior to its measurement. Consistent with this, ψ_{ik} in (51) is set equal to $\alpha_k\psi_i$, which allows an interaction between a person and an item. This yields the 2PL model

$$P(Y_{ik} = 1|\psi_i) = \frac{\exp[\alpha_k\psi_i - \beta_k]}{1 + \exp[\alpha_k\psi_i - \beta_k]}, \quad (52)$$

where ψ_i is individual i's value on the latent variable, α_k and β_k in the IRT literature are referred to as 'discrimination' and 'difficulty' parameters, respectively.

To obtain a model for responses to K items, the conditional probabilities in (52) for $k = 1, \ldots K$ are assumed to independent given ψ_i (i.e., local independence). The model for the observed data (i.e., response patterns) is

$$P(\mathbf{Y}_i = \mathbf{y}) = \int \prod_{k=1}^{K} P(Y_{ik} = y_k|\psi_i) f(\psi_i) d\psi_i, \quad (53)$$

where $\mathbf{y} = (y_1, \ldots, y_K)$ is a possible response pattern. In practice, numerical integration is typically used to fit the model to data. An alternative approach based on Anderson and Yu (2007; Anderson, et al. 2007) does not require numerical integration and yields either a log-linear or log-multiplicative model for the data.

Given estimates of the α_ks in Equation (52), the weighted sum score $\sum_h \alpha_h y_{ih}$ is a sufficient statistic for ψ_i (Andersen, 1995; Heinen, 1996), where y_{ij} is the score (0 or 1) for individual i on item h. In the model for item k, replacing ψ_i by the weighted sum $\sum_{h\neq k} \alpha_h y_{ih}$, which is based on all items except k, yields

$$P(Y_{ik} = 1|y_{ih}, h \neq k) \\ = \frac{\exp[\alpha_k(\sum_{h\neq k} \alpha_h y_{ih}) - \beta_k]}{1 + \exp[\alpha_k(\sum_{h\neq k} \alpha_h y_{ih}) - \beta_k]}. \quad (54)$$

Justification and precedent for using rest-scores for studying item response functions can be found in Junker and Sijtsma (2000), and was done earlier on page 317.

For a set of K items, there are K item-response functions (54), one for each item. A set of models defined as in (54) for K items over-determines the joint distribution of the responses to the items (i.e., the distribution of response patterns). Restrictions are required to ensure that the conditional models are consistent or compatible with some joint distribution. The compatibility conditions are that the coefficient when predicting responses to item k based on k' is the same as the coefficient when predicting responses to item k' based on k (Joe and Liu, 1996). These conditions are met in (54). When predicting k from k', the coefficient for item k' is $\alpha_k\alpha_{k'}$, which equals $\alpha_{k'}\alpha_k$, the coefficient for k when predicting k' from k.

The joint distribution implied by the set of conditional models defined in (54) is a log-multiplicative association model

$$\log(P(\mathbf{Y}_i = \mathbf{y})) = \lambda + \sum_k \beta_k y_{ik} + \sum_{k<h} \alpha_k \alpha_h y_{ik} y_{ih}. \quad (55)$$

Model (55) is a special case of a homogeneous association log-linear model where the unstructured two-way interaction parameters between items k and h have been replaced by the products $\alpha_k\alpha_h$. This model is a generalization of Goodman's *RC* association model for two variables (Goodman, 1979, 1985). The parameters of the LMA association Model (55) equal those from the 2*PL* model up to a linear transformation.

The LMA in (55) is not truly a GLM because the systematic component contains multiplicative terms; however, maximum likelihood estimation of the parameters of LMAs can be done without numerical integration. As an example, the 2*PL* model was fit to the five vocabulary items from the 2004 GSS (Davis, Smith, and Marsden, 2007) as an LMA model and by marginal maximum likelihood estimation, which is a standard method in IRT applications. The data are given in Table 14.9. Both models were fit to the data using the *LEM* program (Vermunt, 1997). The fit statistics are reported in Table 14.10. In addition to likelihood ratio statistics, the dissimilarity index and the Bayesian information criterion (BIC) are also reported. These can be used to compare non-nested models. The dissimilarity index is

$$D = \frac{\sum_i |y_i - \hat{y}_i|}{2N}, \quad (56)$$

where N is the total sample size. The dissimilarity index is interpretable as the proportion of observations that would need to have to change response patterns (cells of the cross-classification of the items) for the model to fit perfectly. The BIC statistic is

$$\text{BIC} = G^2 - df \ln(N). \quad (57)$$

Smaller values of D and *BIC* indicate better models.

The goodness-of-fit statistics for the 2*PL* model fit to data by the two estimation methods are very similar and the models fit by both of the methods give adequate representations of the data (i.e., for LMA, $G^2 = 27.30$, $df = 21$, $p = .16$ and for MMLE, $G^2 = 26.26$, $df = 21$, $p = .20$). The correlations between parameter estimates for the βs equals $r = .97$ and for the αs equals $r = .99$. In terms of goodness-of-fit and parameter estimation, the closeness of results between the MMLE of the 2*PL* and MLE of the LMA is consistent with the findings of

Table 14.9 Cross-classification of five of the vocabulary items from the 2004 General Social Survey

			E			
			0		1	
			F		F	
A	C	D	0	1	0	1
0	0	0	16	2	4	0
0	0	1	17	18	14	91
0	1	0	1	0	1	0
0	1	1	3	4	5	18
1	0	0	12	6	3	11
1	0	1	23	60	74	513
1	1	0	1	1	0	1
1	1	1	3	8	10	235

Anderson and Yu (2007), who simulated data using (53) where ψ_i was from either a normal or exponential distribution.

The Rasch model

The Rasch model can be viewed as special case of the 2PL where the discrimination parameters are constant over items (i.e., $\alpha_k = \alpha$). Typically α is set equal to 1 for a correct response and 0 otherwise. Following the same procedures used to derive a LMA from the 2PL, we use $\phi \sum_{h \neq k} y_{ih}$ as an estimate of ψ_i when predicting item k from the rest. We have included a weight parameter ϕ because the scale of the latent variable is unknown (Anderson, et al. 2007). Using the rest-score for ψ_i we obtain

$$\log(P(\mathbf{Y}_i = \mathbf{y})) = \lambda + \sum_k \beta_k y_{ik} + \phi \sum_{k<h} y_{ik} y_{ih}. \quad (58)$$

This model is a GLM, specifically, a log-linear by linear association model, which is a Poisson regression model.

Another approach to fitting the Rasch model as a GLM is to fit it as a quasi-symmetric log-linear model, which yields conditional maximum likelihood estimates (Tjur, 1982; see also Agresti, 2002; de Leeuw and Verhelst, 1986; Kelderman, 1996). The Rasch model as a quasi-symmetric log-linear model is

$$P(\mathbf{Y}_i = \mathbf{y}) = \exp\left[\lambda + \sum_k \beta_k + \lambda^{\text{total}}_{y_1+\ldots+y_K}\right], \quad (59)$$

where $\lambda^{\text{total}}_{y_1+\ldots+y_K}$ is the parameter for the test total $y_1 + \ldots + y_K$. In this model, the test total is treated as a nominal variable. This model is quasi-symmetric, because the interaction parameter $\lambda^{\text{total}}_{y_1+\ldots+y_K}$ for response patterns with the same number of correct answers is the same. For example, consider our vocabulary example. The response patterns $(y^A, y^C, y^D, y^E, y^F)$ and $(y^C, y^A, y^D, y^E, y^F)$, where answers to A and C have been permuted, has the same test total and value for λ^{total} in the model.

In the log-linear by linear model, the association between two items, say k and h is symmetric (i.e., $\phi y_j y_h = \phi y_h y_j$). The log-linear by linear model association model can be viewed as a limited information quasi-symmetric Rasch model. The association is symmetric, but it does not permit higher-order associations. Model (59) does allow for higher-order associations.

The Rasch model was fit to the five vocabulary items from the 2004 GSS (Davis et al. 2007) in Table 14.9 as a quasi-symmetric log-linear model, a log-linear by linear association model, and by MMLE assuming a normal distribution for ψ_i. The fit statistics are reported in Table 14.10. Although none of the Rasch models estimated yields a good representation of the data, notable is the difference in goodness-of-fit of the Rasch models under different estimation methods. The model fit by MMLE fits the data the worst, followed by the log-linear by linear model, and the best is the quasi-symmetry model. These results are expected, because, the quasi-symmetry estimation of the Rasch Model (59), referred to as the "extended random Rasch" model (de Leeuw and Verhlest, 1986) needs complicated restrictions placed on the parameters (Lindsay, Clogg, and Grego, 1991;

Table 14.10 Summary of item response models fit to the GSS 2004 vocabulary words

Model	*df*	*G^2*	*p*	*D*	*BIC*
Complete independence	26	334.20	<.01	.1354	150.86
Homegeneous association	16	25.43	.06	.0396	−87.40
Rasch as a quasi-symmetric log-linear model	22	56.32	<.01	.0426	−98.82
Rasch as log-linear by linear	25	66.72	<.01	.0517	−109.58
Rasch by MMLE	25	74.10	<.01	.0589	−102.20
2PL as a log-multiplicative association model	21	27.30	.16	.0402	−120.79
2PL by MMLE	21	26.26	.20	.0380	−121.82

Table 14.11 Summary of generalized linear models and related models that are equivalent to psychometric models, which were described in this chapter

	Generalized linear Model		
Psychometric Model	*Random*	*Systematic*	*Link*
Discrete choice/random utility models			
Thurstone's Law of Comparative judgment, Case V	Binomial	Contrasts	probit
Bradley–Terry–Luce	Binomial	Contrasts	logit
Bradley–Terry–Luce	Poisson	Quasi-symmetry	log
Item-response models			
Rasch	Poisson	Quasi-symmetry	log
Rasch	Poisson	Linear by linear	log
Two-parameter logistic	Poisson	Multiplicative	log

Hout, Duncan, and Sobel, 1987), but these are difficult to impose and were not imposed here. This example illustrates that under MMLE results "may become quite different from CML [conditional maximum likelihood estimation]" (de Leeuw and Verhlest, 1986: 192).

Taking into account the different number of parameters estimated, the sample size and goodness-of-fit, the log-linear by linear model is the best for the Rasch model (i.e., smallest BIC). Even though the goodness-of-fit statistics for the Rasch models estimated as a quasi-symmetric log-linear model and log-linear by linear models differ, the estimated βs are highly correlated, $r > .99$. In this case, a potential advantage of using the log-linear by linear estimates over the quasi-symmetric model is that all of the standard errors from the log-linear by linear model were smaller except for one.

CONCLUSION

Table 14.11 gives a summary of the psychometric models and the corresponding GLMs discussed in this chapter. Following the methods used here, additional psychometric models can be formulated and estimated as standard models for categorical data. For example, a version of Thurstone's Law of Categorical Judgment can be fit as a proportional odds model, which is a logistic regression model for ordinal data. Additionally, multidimensional compensatory IRT models with covariates are an extension of the models presented earlier, and models in a Rasch family for polytomous items and multiple latent variables is described in Anderson et al. (2007). The MLE of log-linear by linear and LMA models is limited to relatively small numbers of terms; however, Anderson et al. (2007) provide an estimation method for the Rasch family that only requires fitting a single logistic regression model to data and modification of this method can handle more general models. Although the relationship between GLMs and psychometric models was emphasized in this chapter, categorical data analysis is very useful in wide range of psychological studies, whether it involves latent variable models or not.

NOTES

1 For underlying multivariate normality, the model implied has association parameters that are functionally related to partial correlations.

2 The parameters of the logistic distribution (mean 0 and dispersion parameter .625) were chosen so that the standard normal and logistic distributions approximate each other.

3 For the curious, the Wald statistic equals 7.97, $df = 2$, and $p = .02$.

4 The global goodness-of-fit statistics G^2 are *LR* statistics that compare a specific model to the statured model (i.e., the data).

5 Marginal homogeneity can be tested by doing a conditional likelihood ratio test because the symmetry log-linear model is nested within the quasi-symmetric model.

6 The career "protective" was omitted from this analysis because no females choose this career.

7 Models for β_k could be specified that include attributes of the items.

8 ϵ_{ik} may be random due to random sampling or random behavior (Holland, 1990a). The source of stochastic responses empirically cannot be determined simply by fitting an IRT model.

REFERENCES

Agresti, A. (1980) 'Generalized odds ratios for ordinal data', *Biometrika,* 36, 59–67.

Agresti, A (2002) *Categorical Data Analysis* (2nd edn.). Hoboken, NJ: Wiley.

Agresti, A (2007) *An introduction to Categorical Data Analysis* (2nd edn.). Hoboken, NJ: Wiley.

Agresti, A. and Winner, L. (1997) 'Evaluating agreement and disagreement among movie raters', *Chance,* 10: 10–14.

Altham, P.M.E. (1970) 'The measurement of association of rows and columns of an r x s contingency table', *Journal of the Royal Statistical Society,* Series B, 32: 63–73.

Anderson, C.J. and Yu, H.T. (2007) 'Log-multiplicative association models as item response bmodels', *Psychometrika,* 72: 5–23. DOI: 10.10071s11336–005–1419–2.

Anderson, C.J., Li, Z., and Vermunt, J.K. (2007) 'Estimation of a family of Rasch models for polytomous items and multiple latent variables', *Journal of Statistical Software,* 20. DOI. Online. Available: http://www.jstatsoft.org/v20/i06/v20i06.pdf

Anderson, E.B. (1995) 'The derivation of polytomous Rasch models', in Fischer, G.H. and Molenaar, I.W. (eds), *Rasch Models: Foundations, Recent Developments, and Applications.* New York: Springer.

Bradley, R.A. and Terry, M.E. (1952) 'The rank analysis of incomplete block designs I: The method of paired comparisons', *Biometrika,* 39: 324–345.

Cressie, N. and Holland, P.W. (1983) 'Characterizing the manifest probabilities of latent trait distributions', *Psychometrika,* 48: 129–141.

Dale, J.R. (1986) 'Global cross-ratio models for bivariate, discrete, ordered responses', *Biometrics*, 42: 909–917.

Davis, J.A., Smith, T.W., and Marsden, P.V. (2004) *General Social Surveys, 1972–2004 Codebook.* Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

de Leeuw, J. and Verhelst, N. (1986) 'Maximum likelihood estimation in generalized Rasch models', *Journal of Educational Statistics,* 11: 183–196.

Dobson, A.J. (1990) *An introduction to Generalized Linear Models.* London: Chapman and Hall.

Edwardes, M.D. and Baltzan, M. (2000) 'The generalization of the odds ratio, risk ratio and risk differences to k $\times$ k tables', *Statistics in Medicine,* 19: 1901–1914.

Fahrmeir, L. and Tutz, G. (2001) *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd edn.). New York: Springer.

Fay, M.P. (2002) 'Measuring a binary response's range of influence in logistic regression', *The American Statistician,* 56: 5–9.

Fienberg, S.E. (1985) *The Analysis of Cross-Classified Data.* Cambridge, MA: MIT Press.

Fienberg, S.E. and Larntz (1976) 'Log linear representations for paired and multiple comparison models', *Biometrika,* 63: 245–254.

Goodman, L.A. (1979) 'Simple models for the analysis of association in cross-classifications having ordered categories', *Journal of the American Statistical Association,* 74: 537–552.

Goodman, L.A. (1981) 'Association models and the bivariate normal distribution for contingency tables with ordered categories', *Biometrika,* 68: 347–355.

Goodman, L.A. (1985) 'The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries', *The Annals of Statistics,* 13: 10–69.

Goodman, L.A. and Kruskal, W.H. (1979) *Measures of Association for Cross-classifications.* New York: Springer.

Heinen, T. (1996) *Latent Class and Discrete Latent Trait Models: Similarities and Differences.* Thousand Oaks, CA: Sage Publications.

Holland, P.W. (1990a) 'On the sampling theory foundations of item response theory models', *Psychometrika,* 55: 577–601.

Holland, P.W. (1990b) 'The Dutch identity: a new tool for the study of item response models', *Psychometrika,* 55: 5–18.

Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression* (2nd edn.). New York: Wiley.

Hout, Dunncan and Sobel, M (1987) 'Association and heterogeneity: Structural models of similarities and differences', *Sociological Methodology,* 17: 145–184.

Insightful Corporation (2007) *S Plus.* Seattle, WA: Insightful Corporation. Online. Available: http://www.insightful.com/products/splus/default.asp

Joe, H. and Liu, Y. (1996) 'A model for multivariate binary response with covariates based on conditionally specified logistic regressions', *Statistics and Probability Letters,* 31: 113–120.

Junker, B.W. and Sijtsma, K. (2000) 'Latent and manifest monotonicity in item response models', *Applied Psychological Measurement,* 24: 65–81.

Kelderman, H. (1984) 'Loglinear rasch model tests', *Psychometrika,* 49: 223–245.

Kelderman, H. (1996) 'Loglinear multidimensional item response models for polytomously scored items', in van der Linden, W.J. and Hambleton, R.K. (eds.) *Handbook for Modern Item Response Theory.* New York: Springer. pp. 287–364.

Kroger, K. (1992) 'Unpublished dataset', *University of Illinois Urbana-Champaign.*

Lindsey, J.K. (1997) *Applying Generalized Linear Models.* New York: Springer.

Lindsey, J.K., Clogg, C.C., and Grego, J. (1991) 'Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis', *Journal of the American Statistical Association,* 86: 96–107.

Long, S.J. (1997) *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage.

Luce, R.D. (1959) *Individual Choice Behavior.* New York: Wiley.

Maydeu-Olivares, A. and Böckenholt, U. (2005) 'Structural equation modeling of paired-comparison and ranking data', *Psychological Methods,* 10: 285–304. DOI: 10.1037/1082– 989X.10.3.285.

McCullagh, P and Nelder, J.A. (1983) *Generalized Linear Models* (2nd edn.). London: Chapman and Hall.

McFadden, D. (1974) 'Conditional logit analysis of qualitative choice behavior', in Zarembka, P. (ed.) *Frontiers of Econometrics.* New York: Academic Press. pp. 105–142.

Nadarajah, S. (2007) 'Linear combinations of Gumbel random variables', *Stochastic Environmental Risk Assessment,* 21: 283–286.

Nelder, J. and Wedderburn, R.W.M. (1972) 'Generalized linear models', *Journal of the Royal Statistical Society,* A (136): 370–384.

Powers, D.A. and Xie, Y. (2000) *Statistical Methods for Categorical Data Analysis.* New York: Academic Press.

R Development Core Team (2006) *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Online. Available: http://www.R-project.org

SAS institute, inc (2003) *Statistical Analysis System, Version 9.1.* Cary, NC: SAS Institute, Inc. Online. Available: http//www.sas.com

Swaminathan, H. and Rogers, H.J. (1990) 'Detecting differential item functioning using logistic regression procedures', *Journal of Educational Measurement,* 27: 361–370.

Tatsuoka, M.M., and Lohnes, P. (1988) *Multivariate Analysis: Techniques for Educational and Psychological Research* (2nd edn.). Basingstoke, UK: Macmillan.

Tettagh, S. and Anderson, C.J. (2007) 'Pre-service teachers' empathy and cognitions: statistical analysis of text data by graphical models'. *Contemporary Educational Psychology,* 32: 48–82.

Thurstone, L.L. (1927) 'A law of comparative judgment', *Psychological Review*, 34:273–286.

Tjur, T. (1982) 'A connection between Rasch's item analysis model and a multiplicative Poisson model', *Scandinavian Journal of Statistics,* 9: 23–30.

Torgerson, W.S. (1958) *Theory and Methods of Scaling.* New York: Wiley.

Train, K.E. (2003) *Discrete Choice Methods with Simulation.* New York: Cambridge University Press.

Vermunt, J.K. (1997) *EM: A General Program for the Analysis of Categorical Data.* Tilburg, the Netherlands. Online. Available: http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html

Yellott, J.I. Jr. (1977) 'The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution', *Journal of Mathematical Psychology,* 15: 109–144.