ABSTRACT In this chapter, a number of the tests of model fit for the Rasch model for dichotomous items presented in Chapter 5 are generalized to a class of IRT models for polytomous items. Again, the problem of evaluating model fit is solved in the framework of the general multinomial model, and it is shown that the four types of tests considered in Chapter 5 – generalized Pearson tests, likelihood ratio tests, Wald tests, and Lagrange multiplier tests – can all be adapted to the framework of polytomous items. Apart from providing global measures of overall model fit, test statistics must also provide information with respect to specific model violations. The last section of this chapter gives an example of a testing procedure focussing on the model violation of item bias or differential item functioning.

18.1 Introduction

In Chapter 5, an overview of the various approaches to testing the fit of the Rasch model for dichotomous items was presented. A taxonomy of tests was introduced based on three aspects: the specific assumptions and properties of the model tested, the type of statistic on which the test is based, and the mathematical sophistication of the procedure, particularly, the extent to which the (asymptotic) distribution of the statistic is known. This taxonomy can also be applied to the case of polytomous items. However, for several reasons the actual setup of this chapter significantly differs from that of Chapter 5. Firstly, the Rasch model for polytomous items has quite a few generalizations. Therefore, first a general expression encompassing most of these different models will be given. Another problem is the relative lack of testing procedures. For this reason, this chapter will contain more new than old material. Finally, the third aspect of the taxonomy, that is, the extent to which the (asymptotic) distribution of the statistic is known, has little relevance in the present case, because most existing procedures have firm statistical roots. This may be due to the circumstance that the greater complexity of the models for polytomous items thwarts simple approximations.

It will be shown that most of the mathematically well-founded statistics of Chapter 5 can be generalized to the case of polytomous items. The focus will mainly be on generalized Pearson statistics. The definitions of likelihood ratio statistics, Wald statistics, and Lagrange multiplier statistics given in Chapter 5...
will hardly need adjustment. In the last section of this chapter, an example of a testing procedure focusing on item bias or DIF will be given. The statistics will apply to the framework of the OPLM for polytomous items, but the ideas on which the testing procedure is based apply to all models considered in this chapter.

18.2 A General Formulation of a Rasch Model for Polytomous Items

To code the response of a person to an item, let item \( I_i \) have \( m_i + 1 \) response categories \( C_h \), indexed \( h = 0, 1, ..., m_i \). The response to the item will be represented by an \((m_i + 1)\)-dimensional vector \( \mathbf{x}_i = (x_{i0}, ..., x_{ih}, ..., x_{im_i}) \), where \( x_{ih} \) is defined by

\[
x_{ih} = \begin{cases} 1 & \text{if the response is in category } C_h, \ h = 0, ..., m_i, \\ 0 & \text{otherwise}. \end{cases}
\]

Consider a model where the probability of a response in category \( C_h, \ h = 0, ..., m_i \), as a function of a vector of ability parameters \( \boldsymbol{\theta}' = (\theta_1, ..., \theta_q, ..., \theta_Q) \) and a vector of the parameters of item \( I_i, \boldsymbol{\beta}'_i = (\beta_{i1}, ..., \beta_{iU}, ..., \beta_{iU}) \), is given by

\[
P(X_{ih} = 1|\boldsymbol{\theta}, \beta_i) = \frac{\exp(\sum_{q=1}^{Q} r_{ihq} \theta_q - \sum_{u=1}^{U} s_{ihu} \beta_{iu})}{\sum_{h=0}^{m_i} \exp(\sum_{q=1}^{Q} r_{ihq} \theta_q - \sum_{u=1}^{U} s_{ihu} \beta_{iu})}
\]

where \( r_{ih} = (r_{ih1}, ..., r_{ihq}, ..., r_{ihQ}) \) and \( s_{ih}' = (s_{ih1}, ..., s_{ihu}, ..., s_{ihU}) \) are fixed 'score functions', which are part of the sufficient statistics. Notice that \( r_{ih} \) specifies the relation between the \( Q \) ability dimensions and response category \( C_h \). In the same manner, \( s_{ih} \) defines the relation between the \( U \) parameters of item \( I_i \) and category \( C_h \).

Introducing the matrices of score functions \( R_i = [r_{i0}, ..., r_{ih}, ..., r_{im_i}] \) and \( S_i = [s_{i0}, ..., s_{ih}, ..., s_{im_i}] \) and using (18.2), the probability of response \( \mathbf{x}_i \) can be written as

\[
P(\mathbf{x}_i|\boldsymbol{\theta}, \beta_i) \propto \exp[\mathbf{x}_i'(R_i'\mathbf{\theta} - S_i'\boldsymbol{\beta})].
\]

Let a test consist of \( k \) items. Using (18.3) and local stochastic independence, the probability of response pattern \( \mathbf{x}, \mathbf{x}' = (\mathbf{x}_1', ..., \mathbf{x}_i', ..., \mathbf{x}_k') \), given the ability and item parameters, can be written as

\[
P(\mathbf{x}|\boldsymbol{\theta}, \beta) \propto \exp[\mathbf{x}'(R'\mathbf{\theta} - S'\beta)],
\]

where \( R \) is a matrix of score functions defined by \( R = [R_1, ..., R_i, ..., R_k] \), and \( S \) a matrix of score functions \( S = [S_1, ..., S_i, ..., S_k] \). A model of the universality of