Chapter 2

Descriptive and explanatory item response models

Mark Wilson
Paul De Boeck

2.1 Introduction

In this chapter we present four item response models. These four models are comparatively simple within the full range of models in this volume, but some of them are more complex than the common item response models. On the one hand, all four models provide a measurement of individual differences, but on the other hand we use the models to demonstrate how the effect of person characteristics and of item design factors can be investigated. The models range from descriptive measurement for the case where no such effects are investigated, to explanatory measurement for the case where person properties and/or item properties are used to explain the effects of persons and/or items.

In the following sections of this chapter we will concentrate on logistic models, but all that is said also applies to normal-ogive models if the logit link is replaced with the probit link. The models we will discuss are all GLMMs with random intercepts and fixed slopes.

2.1.1 The intercept or person parameter

Typically, the intercept in an item response model is one that varies at random over persons. It is therefore called the *person parameter*. In the notation for item response models, it is commonly denoted by θ_p. It is assumed in this chapter that θ_p is normally distributed with mean zero: $\theta_p \sim N(0, \sigma^2_\theta)$.

The random intercept or person parameter fulfills the function that is often the main reason why people are given a test. Person parameters provide a measurement of latent variables such as abilities, achievement levels, skills, cognitive processes, cognitive strategies, developmental stages, motivations, attitudes, personality traits, states, emotional states or inclinations. A general term that we will use for what is measured in a test is *propensity*. Alternatively, another conception of the person parameter is that it can also be (a) a fixed parameter, and/or (b) more than one person parameter (i.e., in a multidimensional model). We will elaborate on these possibilities only later. For now, it suffices to know that the random intercept is a person parameter and that the estimate for an individual person is considered a measurement of the propensity expressed in the test.

As a measurement tool, item response models of the type we are discussing provide more than ordinal quantification. However, an important alternative approach is to restrict quantification to ordinal numbers. Ordinal item response models are often also called nonparametric item response models (Junker & Sijtsma, 2001; Sijtsma & Molenaar, 2002). The important asset of nonparametric models is that they make no assumptions regarding the item response functions, except for monotonicity assumptions. Thus, they are more flexible than parametric item response models. However, the family of nonparametric models has been developed mainly for measurement purposes. It is not yet fully elaborated for explanatory purposes to investigate the effect of person properties and item properties (such as factors in an experimental design). Thus, in this volume, we will concentrate on parametric models.

2.1.2 The weights or item parameters

As in Chapter 1 we will denote the item predictors by an X, with subscript k ($k = 1, \ldots, K$) for the predictors, so that X_{ik} is the value of item i on predictor k. The most typical predictors in an item response model are not real item properties as in Chapter 1, but item *indicators*. This means that as many predictors are used as there are items, one per item, so that $X_{ik} = 1$ if $k = i$, and $X_{ik} = 0$ if $k \neq i$. For example, for a set of six items, the predictor values would be as follows:

item 1:	1	0	0	0	0	0	item 4:	0	0	0	1	0	0
item 2:	0	1	0	0	0	0	item 5:	0	0	0	0	1	0
item 3:	0	0	1	0	0	0	item 6:	0	0	0	0	0	1.

In typical item response modeling applications, the weights of these predictors are fixed parameters since they do not vary over persons. These weights are the slopes of the binary indicators (see Figure 2.1). The values of these indicator weights are called the *item parameters*, commonly denoted by β_i. Since each item has its own predictor, the subscript i is used instead of k.

2.1.3 Resulting models

The resulting equation for the linear component η_{pi} is the following:

$$\eta_{pi} = \beta_i + \theta_p, \tag{2.1}$$

with $\beta_i = \sum_{k=1}^{K} \beta_k X_{ik}$. As noted in Chapter 1, η_{pi} is $\eta_{pi}|\theta_p$, but here and in the following we will omit the conditional notation for η_{pi} (and π_{pi}). Since all X_{ik} with $i \neq k$ equal 0, only one term of this sum has a non-zero value. It is a common practice to reverse the sign of the item parameter, so that the contribution of the item is negative and may be interpreted as the item difficulty in the context of an achievement test. The resulting equation is:

$$\eta_{pi} = \theta_p - \beta_i. \tag{2.2}$$

In order to convey some intuitions about the intercept and coefficients used above, we give, in Figure 2.1 a graphical representation of Equation 2.2 for person p and the kth predictor. The value of X_{ik} is represented on the x-axis. X_{ik} can have two values: 0 and 1. For $k = i$, the value is 1 for item i, and 0 for all other items. This simply means that item i makes no contribution for other items. Note that the intercept of the regression line is the value of θ at $X_{ik} = 0$. Also note that the difference between $X_{ik} = 1$ and $X_{ik} = 0$ is 1, and the difference between the η_{pi} for $X_{ik} = 1$ and $X_{ik} = 0$ is $-\beta_i$, hence the slope of the line (i.e., the regression weight) is also $-\beta_i$. Other persons will have a parallel line, but the intercepts of the line will vary (and we have assumed they follow a normal distribution). Figure 2.1 does not give the full picture since it represents the effect of only one predictor, the item indicator $k = i$. Figure 2.1 is also somewhat imaginary in the sense that our item indicators can have only two values, while the line connecting the two points suggests that intermediate values can also exist.

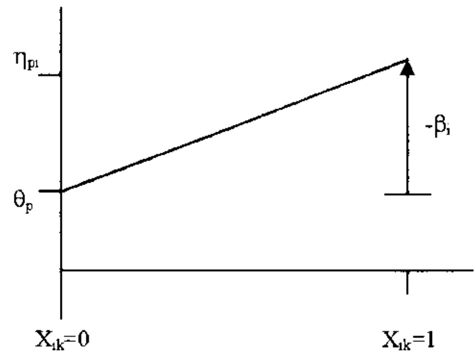

FIGURE 2.1. Linear function for one item predictor in the Rasch model. (Note that in this case $\beta_i < 0$.)

The resulting model of Equation 2.2 (or, equivalently, 2.1) is the Rasch model (Rasch, 1960). The Rasch model is a model that is descriptive for both the person side and the item side of the data matrix. It describes variation in the persons through a person parameter θ_p, which is a random variable as presented here. And it describes the variation in the items through fixed individual item parameters.

2.2 Four item response models

The primary aim of this chapter is to illustrate the distinction between a *descriptive* approach and an *explanatory* approach in the context of item response modeling. In the course of illustrating the distinction, we will present four item response models one of which is the Rasch model from Equations 2.1 and 2.2. The four models differ in whether they are descriptive or explanatory at the person side and the item side.

The four models we have selected to present below are logistic random-intercepts models and therefore belong to the Rasch tradition, but this does not mean we are in this volume restricting our possible models to that approach. In the Rasch tradition, which might also be called *prescriptive measurement* (Rasch, 1960; Fischer & Molenaar, 1995), models include no interactions between persons and items, but just main effects of persons and items – specifically, the random intercept, θ_p, is not weighted depending on the item. If there were such interactions, then the effect of a person parameter would depend on the items, and therefore, by implication, in the inferential step, the measurement outcome would necessarily also depend on the items that are included. This prescriptive measurement approach is only one of two measurement approaches that are commonly followed with item response models (Thissen & Orlando, 2001; Wilson, 2003). The alternative approach might be termed *empirical* in that one seeks to modify the model to fit the data more closely – specifically, the model is expanded by weighting the random intercept by an item parameter α_i (Birnbaum, 1968). Such a model is called the *two-parameter logistic model* (2PL model) Thus, in the empirical tradition, relatively more items will fit the model than in the prescriptive tradition, although there will be items that do not fit well under either tradition.

The basis for selecting these particular models for this second introductory chapter is that they are building blocks which can serve as the basis for the very extensive expansion of the models in the remainder of this volume, and which will include, as one aspect, adding the second item parameter α_i, typical of the empirical tradition. After the model formulation and discussion for each of the four models below, an application will be discussed, making use of the dichotomized example data from Chapter 1.

Table 2.2 shows four types of models, depending on the types of predictors that are included. There are two kinds of item predictors: item indicators, and item properties. And there are also two kinds of person predictors: person indicators, and person properties. Look first at the top left-hand corner of the 2×2 layout of Table 2.1. When each person has his/her own unique effect, unexplained by person properties, and when each item has its own unique effects, unexplained by item properties, we will refer to the model as *doubly descriptive*. Such a model describes the individual effects of the persons and of the items (hence, doubly descriptive), without explaining either of these effects. The Rasch model is an example.

TABLE 2.1. Models as a function of the predictors.

	Person predictors	
Item predictors	Absence of properties	Inclusion of properties (person properties)
Absence of properties	doubly descriptive	person explanatory
Inclusion of properties (item properties)	item explanatory	doubly explanatory

Doubly descriptive models are mostly sufficient for measurement purposes, and are those most commonly seen in practice.

However, if the person parameter is considered to be a random effect, then there may be unwanted consequences if the effect of certain person properties is not taken into account. If a normal distribution is assumed, the result is that the normal distribution no longer applies for the entire subset of persons, but only for subsets of persons who share the same person property values. For example, if gender has an effect, then not one normal distribution applies but two, differentiated by the gender of the person. Thus, when person properties are included in the model to explain the person effects, then the models will be called *person explanatory* (top right-hand corner of Table 2.1).

In a similar way, when item properties are included to explain the item effects, the models will be called *item explanatory* (bottom left-hand corner of Table 2.1). Finally, when properties of both kinds are included, the models will be called *doubly explanatory* (bottom right-hand corner of Table 2.1). See Zwinderman (1997) and Adams, Wilson and Wu (1997) for similar taxonomies and short descriptions of the models. In the verbal aggression example data set from Chapter 1, we have information on person properties as well as on item properties, so that the two types of explanatory models (person and item) can be illustrated.

2.2.1 Summary and notation

A summary of the four models to be explained is given in Table 2.2. The following notation is used in the table and will be followed also in the remainder of this chapter. θ_p is used for the random person parameter, with mean zero and variance σ_θ^2. When person properties are included in the model, the symbol ε_p is used for the unexplained part of the person contribution, with mean zero and variance σ_ε^2. The person properties are denoted with capital Z. The subscript j is used for these predictors, $j = 1, \ldots, J$.

TABLE 2.2. Summary of the four models.

	$\eta_{pi} =$			
Model	Person part	Item part	Random effect	Model type
Rasch model	θ_p	$-\beta_i$	$\theta_p \sim N(0, \sigma_\theta^2)$	Doubly descriptive
Latent reg Rasch model	$\sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p$	$-\beta_i$	$\varepsilon_p \sim N(0, \sigma_\varepsilon^2)$	Person explanatory
LLTM	θ_p	$-\sum_{k=0}^{K} \beta_k X_{ik}$	$\theta_p \sim N(0, \sigma_\theta^2)$	Item explanatory
Latent reg LLTM	$\sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p$	$-\sum_{k=0}^{K} \beta_k X_{ik}$	$\varepsilon_p \sim N(0, \sigma_\varepsilon^2)$	Doubly explanatory

This is a deviation from the GLMM notation where Z is used for predictors with a random effect. The GLMM notation is the notation that is followed in Chapter 4 on the statistical background of this volume and in Chapter 3 on multicategorical data also because that chapter relies more directly on the general GLMM framework. Rather than distinguishing between the predictors on the basis of whether they have a fixed or random effect, we use here a different notation for person predictors and item predictors, because they lead to quite different item response models and because in these models persons and items are not treated in an equivalent way, as will be explained in Sections 2.4.1, 2.5.1, and 2.6.1. This leaves the X for the item predictors, with subscript k, $k = 1, \ldots, K$. Where the effects of person predictors are considered fixed, they are denoted by ϑ_j, and the fixed effects of item predictors by β_k. The random intercepts may be considered the effect of a constant predictor (Z_{p0}, or alternatively X_{i0}).

2.3 A doubly descriptive model: the Rasch model

2.3.1 Formulation of the model

The *Rasch model* was defined earlier in Equations 2.1 and 2.2. We will use Equation 2.2 to obtain an expression for the odds, or $\pi_{pi}/(1-\pi_{pi})$. If on both sides of Equation 2.2 the exponential form is used, then $\exp(\eta_{pi}) = \exp(\theta_p - \beta_i)$. Since $\eta_{pi} = \log(\pi_{pi}/(1 - \pi_{pi}))$, and $\exp(\theta_p - \beta_i) = \exp(\theta_p)/\exp(\beta_i)$, it follows that

$$\pi_{pi}/(1 - \pi_{pi}) = \exp(\theta_p)/\exp(\beta_i). \tag{2.3}$$

Equation 2.3 is the *exponential form* of the Rasch model. As a way to understand Equation 2.3, interpret $\exp(\theta_p)$ as an exponential measure of the ability of person p when taking an achievement test, and interpret

$\exp(\beta_i)$ as an exponential measure of the difficulty of the item i from that test. Then the formula expresses the ratio of the success probability π_{pi} to the failure probability $(1 - \pi_{pi})$ as the ratio of a person's ability to the difficulty of the item.

The intuition reflected in the formula, in an achievement context, is that ability allows one to succeed, while difficulty makes one fail, and that the ratio of both determines the odds of success. Figure 2.2a gives a schematic presentation of this intuitive idea. The figure shows two rectangles on a balance beam – if one weighs more than the other, then the balance will tip that way. Physical balance beams tip one way as soon as the weight on that side is larger than the weight on the other side. Imagine now that tipping one way or the other way in an achievement context is probabilistic as follows. The white rectangle represents the ability and the gray rectangle the difficulty. The ratio of ability to difficulty is 2/1, so that the ratio of the success probability to the failure probability is also 2/1.

From the odds equation, one can derive the equation for the probability. If the numerator on each side of Equation 2.3 is divided by the sum of the numerator and the denominator, it follows that $\pi_{pi}/(\pi_{pi}+(1-\pi_{pi})) = \exp(\theta_p)/(\exp(\theta_p)+\exp(\beta_i))$, and thus that $\pi_{pi} = \exp(\theta_p)/(\exp(\theta_p)+\exp(\beta_i))$. When the numerator and denominator of the latter are each divided by $\exp(\beta_i)$, then the familiar equation for the probability of a 1-response is obtained:

$$\pi_{pi} = \exp(\theta_p - \beta_i)/(1 + \exp(\theta_p - \beta_i)). \tag{2.4}$$

The intuition behind this alternate formula for the Rasch model is that there are two competing responses each of which has a certain attractiveness. Let us denote the attractiveness of $Y_{pi} = 0$ as A and the attractiveness of $Y_{pi} = 1$ as B. The probability of a response may then be considered the ratio of its attractiveness to the sum of the two attractiveness values, or $\pi_{pi} = B/(A+B)$. This is an example of the well-known Bradley-Terry-Luce choice rule: the probability of an alternative depends on the ratio of the attractiveness of that alternative to the sum of the attractiveness values of all alternatives. In Equation 2.4, $A = 1$, and $B = \exp(\theta_p - \beta_i)$. The value of 1 for A is an arbitrarily chosen convention (i.e., the value of π_{pi} is invariant under multiplicative transformations of the attractiveness values, so that one may as well set A equal to 1).

The intuition behind Equation 2.4 is presented in Figure 2.2b. The two attractiveness values are each represented by a section of a rectangle: the gray section for the 0-response, and the white section for the 1-response. The probability of each response is the proportion of the corresponding section in the rectangle. The white section is twice as large as the gray section, so that the resulting probability of a 1-response is $2/(2+1) = .67$.

The link between Figure 2.2a and Figure 2.2b is that the two rectangles of the upper part are first shrunken in equal proportions, and then put next to one another to form one long rectangle. This is a legitimate operation

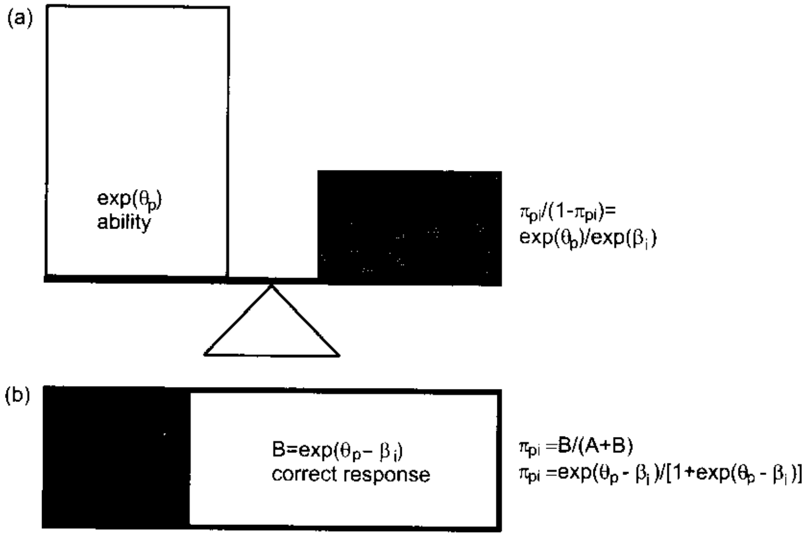

FIGURE 2.2. Illustration of two ideas behind two different formulations of the Rasch model: (a) odds formula, and (b) probability formula.

since π_{pi} is invariant under multiplicative transformations of the rectangles. The transformation illustrates that both $\exp(\theta_p)$ and $\exp(\theta_p - \beta_i)$ may be understood as the attractiveness of a 1-response, and both $\exp(\beta_i)$ and 1 as the attractiveness of a 0-response, depending on whether or not one divides by $\exp(\beta_i)$.

A third metaphor is one of a hurdler (the person) and a series of hurdles (the items). The hurdler is seen as having the ability to leap over hurdles of a certain height (the ability is indicated by θ_p), and the series of hurdles have heights indicated by the series of item difficulties $(\beta_1, \ldots, \beta_I)$. When the hurdler's ability is equal to the height of the hurdle, the leap is successful, with a probability of .50. When the hurdler's ability is different than the height of the hurdle, the leap is successful, with a probability dependent on the difference between them (when the difference is positive, the probability will be greater than .50, and when it is negative, it will be less than .50). This metaphor is possibly better-suited to achievement and ability contexts than other such as attitude variables, but similar interpretations in such contexts are also possible.

In a fourth metaphor, one can represent the heights of the hurdles (the item difficulties) as points along a line, and the ability of the person as a point along the same line. The amount determining the probability of success is then the difference between the two locations, or $(\theta_p - \beta_i)$. This representation is sometimes called an 'item map' or 'construct map.' A generic example is shown in Figure 2.3, where the students are shown on the left-hand side, and the items on the right-hand side. This representation has been used as a way to enhance the interpretability of the results from item response model analyses. Segments of the line can be labeled as exhibiting particular features, for both the persons and the items, and the progress of say, students, through this set of segments, can be interpreted as development in achievement. The placement of the person and item locations in a directly linear relationship has been the genesis of an extensive methodology for interpreting the measure (Masters, Adams, & Wilson, 1990; Wilson, 2003; Wilson, 2005; Wright & Stone, 1979).

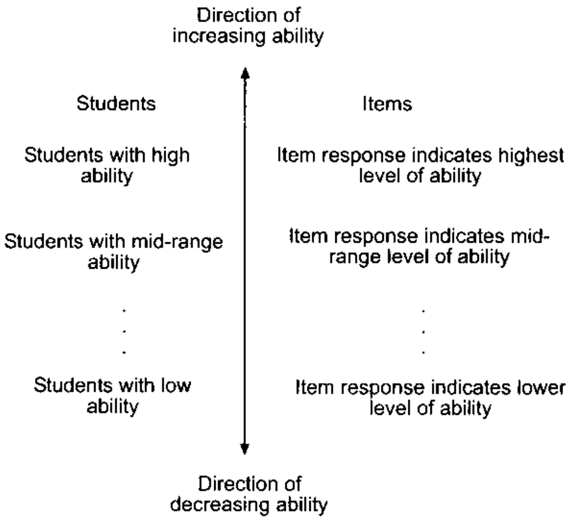

FIGURE 2.3. A generic construct map for an ability.

Item response function

Item response functions or item characteristic curves are item specific functions that map the value of θ_p into the corresponding probability π_{pi}, given the value of β_i. Figure 2.4 shows the item response functions of three items. The shape of Rasch item response functions is the same for all three items, but the location is different. All curves are equally steep, because θ_p is not weighted depending on the item. For all items $\pi_{pi} = .50$ when $\beta_i = \theta_p$, which indicates that β_i locates the curve on the θ-scale.

Graphical representation

The Rasch model is graphically represented in Figure 2.5, following the conventions introduced in the previous chapter. The figure shows the item parameter β_i as the effect of the corresponding item indicator X_{ik} (for $k = i$, the other item indicators are not shown since they don't have an effect), and it shows the person parameter θ_p as the random effect of the constant predictor Z_{p0}. Note that in GLMM notation Z is used for predictors with a random effect, while our notation Z is used for person predictors.

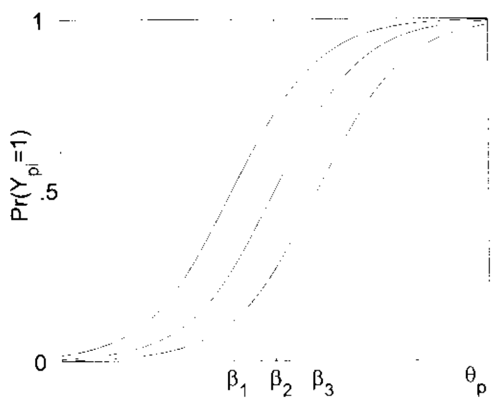

FIGURE 2.4. Item response functions for three items.

Incidentally, the Z_{p0} in Figure 2.5 corresponds with both conventions. It is a constant predictor with a random effect, and it may be considered a person predictor as well, one with a value of 1 for all persons.

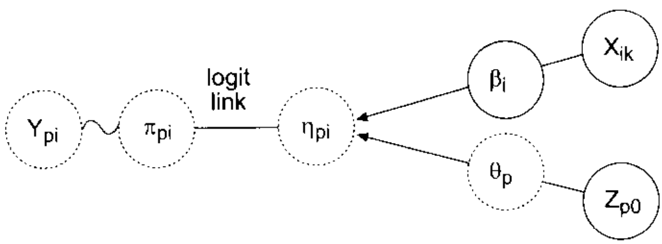

FIGURE 2.5. Graphical representation of the Rasch model. (Note that $k = i$.)

Local independence

An important feature of the model is the so-called *local (or conditional) independence* assumption, meaning that for any response vector $\boldsymbol{y}_p = (y_{p1}, \ldots, y_{pI})'$ (with y_{pi} being the realization of Y_{pi}, (y_{pi} =1 or 0)), the conditional probability of the whole vector is the product of the conditional probabilities of each response. This implies that, for all pairs of items i and i' ($i \neq i'$): $\Pr(Y_{pi} = y_{pi}\ \&\ Y_{pi'} = y_{pi'}|\theta_p) = \Pr(Y_{pi} = y_{pi}|\theta_p) \times \Pr(Y_{pi'} = y_{pi'}|\theta_p)$. Under this assumption, θ_p is the only source of dependence (or correlation) between items – hence, for a given value of θ_p the observations are independent, which means that one dimension or latent trait, θ_p, explains all inter-item correlations. The assumption of local independence underlies all four models in this chapter, and also all models in this volume, except for models with a residual dependence part (see Chapters 7 and 10 for an explanation of that).

Parametrization

Note that the parameters in the above equations appear in two forms: the *exponential form*, using $\exp(\theta_p)$ and $\exp(\beta_i)$, as in Equation 2.3, and the *logarithmic form*, using θ_p and β_i, as in Equation 2.4. We will use the logarithmic form, which is also the most common form. Four different but equivalent parametrizations are possible based on the signs of the person and item expressions:
(1) $\theta_p - \beta_i$;
(2) $\theta_p + \beta_i^*$, with $\beta_i^* = -\beta_i$;
(3) $-\theta_p^* - \beta_i$, with $\theta_p^* = -\theta_p$; and
(4) $\beta_i^* - \theta_p^*$.
The difference between the four is that in some contexts, one of them might work better in terms of interpretation. For example, taking the difference between the item parameter and the person parameter (fourth parametrization) could be useful for the verbal aggression example if the person parameter is seen as a personal aggression threshold (θ_p^*) and the item parameter as the inductive power of the situation-behavior pair (β_i^*). The probability of a verbally aggressive response then grows with the difference between the inductive power of the situation-behavior pair and the personal threshold. In general, the two subtraction parametrizations (1 and 4) lend themselves to metaphors of comparison and competition (e.g., ability and difficulty), and are compatible with the intuitions mentioned above whereas the two addition formulations (2 and 3) are suitable for an intensification metaphor.

Identification

The model as formulated in the previous equations would have an *identification problem* if the mean of the person parameters was not restricted to be zero. The exponential parameters and logarithmic parameters are identified only up to a multiplicative or additive constant, respectively. If one multiplies all exponential parameters with a constant c, then the odds in Equation 2.3 do not change, and if one adds a constant c to all logarithmic parameters, then the probability in Equation 2.4 does not change. Different conventions exist to solve this problem. For instance, one can set the mean θ_p equal to 0, which is the solution we have chosen for this volume, or one can set either a particular β_i or the mean of the β_i equal to 0, which are the most common tactics if θ_p is not considered a random effect.

Variants

The Rasch model exists in three variants named after the formulation of the likelihood to be maximized (Molenaar, 1995). There are three likelihood formulations for the model: the joint maximum likelihood formulation (JML), the conditional maximum likelihood formulation (CML), and the marginal maximum likelihood formulation (MML). The labels of the three formulations refer to a maximization of the likelihood function for estimation purposes. The likelihood function is the probability of the data as a function of the parameters, and, in the case of CML, also of the sufficient statistics for the person parameters. It has been common to consider the three different formulations as no more than three estimation tools, but they can also be considered as being based on different models, as explained in Chapter 12 of this volume. We will follow here the MML formulation, meaning that we assume that the person parameters are sampled from a distribution, so that only the parameters of that distribution (and not the individual person parameters) enter the likelihood that is maximized. If the distribution is the normal distribution, these parameters of the distribution are the mean and the variance. In all applications up to Chapter 10, the normal distribution will be used for person parameters. Other distributions can also be used – for example a histogram distribution can be particularly flexible (Adams, Wilson, & Wu, 1997; de Leeuw & Verhelst, 1986; Follmann, 1988).

MML formulation and estimation of person parameters

For the MML formulation, a more complete way of presenting the model is

$$\begin{aligned}&\pi_{pi} = \exp(\theta_p - \beta_i)/(1 + \exp(\theta_p - \beta_i)),\\ &\theta_p \sim N(0, \sigma^2_\theta),\end{aligned} \tag{2.5}$$

with σ^2_θ being the variance of the θ_p, and assuming local independence. The corresponding marginal likelihood for a full response pattern ($\boldsymbol{y}_p$ as the realization of $\boldsymbol{Y}_p$) is

$$\begin{aligned}&\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p)\\ &= \int_{-\infty}^{+\infty} \prod_{i=1}^{I} (\exp(y_{pi}(\theta_p - \beta_i))/(1 + \exp(\theta_p - \beta_i)))g(\theta_p|\boldsymbol{\psi})d\theta_p,\end{aligned} \tag{2.6}$$

with $g(\theta_p|\boldsymbol{\psi})$ as the normal density of θ_p with parameters $\boldsymbol{\psi}$ (μ_θ and σ^2_θ). For all persons together, the marginal likelihood is the product of the corresponding integrals. The marginal likelihood will not be repeated for the next three models, since one can simply adapt Equation 2.6 based on the equation for π_{pi}. To estimate the model, we need to estimate only the structural parameters $\beta_1, \ldots, \beta_I$, and σ^2_θ (the mean of the distribution is fixed at 0). Therefore, the estimation of θ_p requires a further step beyond the model estimation. A common method for this second step is to calculate empirical Bayes estimates; see Bock and Aitkin (1981), Adams, Wilson and Wang (1997), or Wainer et al. (2001) for a discussion of the concept within the context of item response modeling. These estimates are maximum likelihood estimates given the item responses of the person and the assumed normal distribution with estimated (or fixed) mean and variance. For a discussion and some interesting results on the estimation of person parameters for the Rasch model, see Hoijtink and Boomsma (1995) and Warm

(1989). The issue of estimating person parameters is the same for all four models in this chapter, and in general for all models with a random person parameter.

Comments and literature

The Rasch model is a doubly descriptive model, since it yields only estimates of the individual item and individual person effects. Its great asset is that if it is valid, the person effect does not depend on the item, which is an attractive measurement quality and corresponds to certain notions of what it means to measure (Rasch, 1961). When the ultimate goal is to assign a number to each person in order to measure the person's latent trait, the Rasch model is an excellent model. However, there may be complications in the data that it does not incorporate, and when it comes to *understanding* the responses in terms of person and item properties, the model itself does not help.

The Rasch model is also called the one-parameter logistic (1PL) model because it has only one parameter per item. We will not use this terminology for the Rasch model, since a model with unequal but fixed item weights (discriminations) is also a one-parameter logistic model (OPLM, Verhelst & Glas, 1995). The Rasch model was first described by the Danish mathematician and statistician Rasch (1960, 1961, 1967), and it became known in the psychometric literature thanks to work by Fischer (1968, 1974, 1981) in Europe and Wright (1968, 1977) in the United States. For a history of the Rasch model, see Wright (1997). For a description and discussion of recent developments in the Rasch model and related models, see Fischer and Molenaar (1995) and Rost (2001). A recent introduction has been written by Bond and Fox (2001). A good description of the life and work of Rasch is given by Andersen and Olsen (2001).

2.3.2 Application of the Rasch model

After a dichotomization (i.e., 2 and 1 are mapped to 1), the example data set is analyzed with the NLMIXED procedure of SAS (SAS Institute, 1999), in order to estimate the Rasch model in its MML formulation. The options we chose for all four models discussed in this chapter are: Gaussian quadrature for numerical integration, with 20 quadrature points without adaptive centering (with centering on 0), and Newton Raphson as the optimization method. When adaptive centering was used, essentially the same results were obtained for all four models as with the nonadaptive method – however, it took much longer to run the analysis. For a discussion of estimation methods, see Chapters 4 and 12, and for a discussion of software, see Chapter 12. The use of the NLMIXED procedure of SAS is described in Section 2.8.1.

We will not test this model and the other models with respect to their absolute goodness of fit. Instead we will do two other things. First, we will report the value of three indices: the deviance, the Akaike information criterion (AIC) (Akaike, 1974), and the Bayesian information criterion (BIC) (Schwarz, 1978), with the aim to compare the four models from this chapter on these fit indices. The deviance is $-2\log(L)$, with L being the maximum of the likelihood function given the estimated model. The AIC and BIC are information criteria derived from the deviance, but with a penalty included for the number of parameters: $AIC = -2\log(L) + 2N_{par}$, and $BIC = -2\log(L) + \log(P)N_{par}$, with N_{par} being the number of parameters (for the persons, only the variance is counted as a parameter), and P being the number of persons (see also Bozdogan, 1987; Read & Cressie, 1988). Lower values of the deviance, the AIC, and the BIC indicate a better fit. As a comparison makes sense only when at least one other model is involved, we will start using these indices only in the discussion of the results from the second model; see Section 2.4.2.

Second, we will use significance tests of the likelihood-ratio type and Wald tests. For nested models, we can use *likelihood-ratio tests* (LR tests). The LR test is based on the ratio of two likelihoods. The first likelihood (L_1) belongs to a model that is nested in a second, more general model. The second likelihood (L_2) belongs to this more general model. When the models are estimated with a maximum likelihood method, then minus two times the logarithm of the likelihood ratio, $-2\log(L_1/L_2)$, or the difference between the deviances, is asymptotically distributed as a χ^2 with a number of degrees of freedom (df) equal to the difference between the number of parameters of the two models. Further, we will also use *Wald tests* (Wald, 1941) to determine whether the difference of an estimate with zero is statistically significant. The asymptotic normality of the parameter estimates is the basis for dividing the parameter estimate by its standard error, in order to obtain a statistic that is approximately distributed as a standard normal. For a discussion of adaptations one may consider for this test, depending on the estimation method that is followed, see Verbeke and Molenberghs (2000). The LR test does not apply when one wishes to compare a model with one or more parameters fixed at a boundary value to a model in which these parameters are not fixed but free. For example, the regular LR test does not apply when comparing a model with the person variance fixed to zero and another model where the variance is estimated. For a model with one variance parameter fixed to zero (model 1, likelihood is L1) and a model where that variance is estimated (model 2, likelihood is L2), the LR statistic $-2\log(L1/L2)$ follows a mixture of a $\chi^2(0)$ and a $\chi^2(1)$ distribution (Verbeke & Molenberghs, 2000). Therefore, the regular LR test (which would use $\chi^2(1)$ as the difference in number of parameters is one) is conservative and in fact the p-values must be halved. Given the asymptotic equivalence of the Wald test for a given parameter value and the likelihood-ratio test to test whether the parameter is needed, the Wald test may also be considered conservative. Thus, if the p-value of the Wald test (as shown by NLMIXED) is smaller than the critical value, then the correct p-value certainly is smaller also.

Results

Person variance

The estimated *person variance* is 1.98 on the logit scale. The standard error (SE) of the variance estimate is .21, meaning that the individual differences are statistically significant, with $p < .001$. In general, to interpret an effect a on the logit scale, one should multiply the odds by $\exp(a)$. In order to translate this effect into an effect on the probability, the probability of .50 can be used as a reference value. The size of the person effects can be examined by considering the effect of one standard deviation of θ. Based on Equation 2.3, the odds increase by a factor 4.08 when θ increases by one standard deviation (i.e., 4.08 is $\exp(\sqrt{1.98})$). To illustrate this, suppose a person has a probability of .50 of responding with a 1 ("yes" or "perhaps") on the first item, then someone with a θ-value that is one standard deviation higher has a probability of .80.

Item parameters

The estimated *item parameters* vary from −1.75 to +2.97 on the logit scale, with an average value of .16. The estimates of the item parameters are given in Table 12.3 (Chapter 12). Note that, because of the subtraction in the model equations, lower values of the item parameters imply higher probabilities (i.e., are 'easier' to endorse). The average item value is only slightly higher than the mean of the persons (fixed at zero to identify the model). This means that the average person has a probability of about .50, or more exactly .46, to endorse the average item (responding "yes" or "perhaps"). Note that the effect on the average person is not the average effect, as will be explained in Chapter 4.

Discussion

The rationale of the Rasch model is in the first place to measure persons in this case, to measure the tendency of individual persons to react with verbal aggression. When used for that purpose, the 24 items relating to only four situations are a rather narrow basis for a reliable measurement (but note that Cronbach's $\alpha = .89$). One way to estimate the reliability of the estimates is to derive the standard error (SE) of each of the person parameters. However, since we want to concentrate on the model and not so much on its application for measurement, we will not follow up the reliability of the person measurement at this point (but see Hoijtink & Boomsma, 1995). Instead we will switch to models that can explain person variance and/or item parameters.

2.4 A person explanatory model: the latent regression Rasch model

2.4.1 Formulation of the model

The second model that we consider is the *latent regression Rasch model.* It includes person properties to explain the differences between persons with respect to verbal aggression. Including person properties as predictors is a possibility in GLMMs that we mentioned in Chapter 1, but we did not elaborate on this point there. Recall that person predictors are denoted by Z, and the predictor subscript with j, while the fixed effect is denoted by ϑ. The model differs from the Rasch model in that θ_p is now replaced with a linear regression equation (see also Table 2.2):

$$\theta_p = \sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p, \tag{2.7}$$

so that

$$\eta_{pi} = \sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p - \beta_i. \tag{2.8}$$

in which Z_{pj} is the value of person p on person property j ($j = 1, \ldots, J$),
ϑ_j is the (fixed) regression weight of person property j,
ε_p is the remaining person effect after the effect of the person properties is accounted for, $\varepsilon_p \sim N(0, \sigma_\varepsilon^2)$, which may be considered as the random effect of Z_{p0}, the random intercept.
Note that the ϑ_j that is used in Equation 2.7 as a symbol for the regression weight of a person property is a symbol that differs from θ_p, which is used as the person parameter.

This model is called here the 'latent regression Rasch model', because one can think of the latent person variable θ_p as being regressed on external person variables (Adams, Wilson, & Wu, 1997) such as, for the verbal aggression example, Gender and Trait Anger.

The external person variables are considered as variables with fixed values. When observed person properties are used, the fact that they may include error is ignored in this model (i.e., any errors in the Zs are not modeled). An alternative solution would be a regression on the latent variable that underlies the observed properties (Fox & Glas, 2003; Rabe-Hesketh, Pickles, & Skrondal, 2001). For example, the latent variable underlying the Trait Anger score can function as a latent predictor for the verbal aggression propensity. However, this solution is not part of the latent regression Rasch model formulation in this chapter. In principle, it can be incorporated in the present framework through a multidimensional model with a criterion θ being a function of predictor θs. Depending on the model this may require restrictions on the covariance structure of the θs. For example, when θ_1 has an effect on both θ_2 and θ_3, then this has consequences for the correlation between θ_2 and θ_3.

Graphical representation

Figure 2.6 gives a graphical representation of the latent regression Rasch model. The difference with Figure 2.5 is that the person parameter θ_p is explained in terms of person properties (the Zs) and their effects (the ϑs), and that the unexplained part or error term is the random effect of the constant predictor. One can also connect the two right-most arrows directly to η_{pi}, omitting θ_p, in correspondence with Equation 2.8.

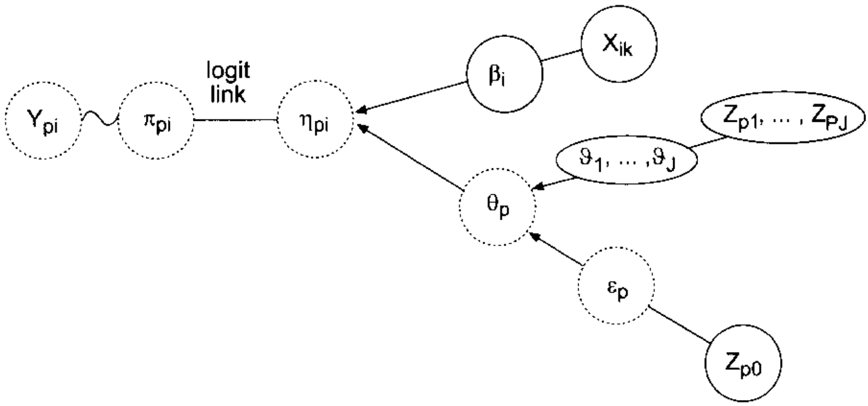

FIGURE 2.6. Graphical representation of the latent regression Rasch model. (Note that $k = i$.)

Literature

The latent regression Rasch model was first described by Verhelst and Eggen (1989) and Zwinderman (1991). This latter author used the term 'generalized Rasch model for manifest predictors' for the global model, and 'structural model' for the latent regression part of the model. Similar models have been presented by Mislevy (1987) for the 2PL or Birnbaum model. For a rather brief but thorough discussion of this model in the broader context of the models of this chapter, see Zwinderman (1997).

2.4.2 Application of the latent regression Rasch model

Two person properties will be used in the application ($J = 2$): the Trait Anger score ($j = 1$) and Gender ($j = 2$). A *dummy coding* is used for Gender, with a 1 for males, and a 0 for females. Of the 316 respondents 243 are males, and 73 are females. For Trait Anger, the raw score is used as a person property; as reference points, the mean score is 20.00 and the standard deviation is 4.85. The use of the NLMIXED procedure for this application is described in Section 2.8.2.

Table 2.3 shows the goodness of fit of the latent regression Rasch model, and also of the Rasch model. The lower the value of these indices, the better the fit of the model. One should of course take into account the number of parameters to make an evaluation, which is why the AIC and the BIC are important criteria. As explained earlier, the penalty for number of parameters is larger in the BIC. It can be noted from Table 2.3 that

TABLE 2.3. Goodness-of-fit indices for the four models.

Model	deviance	AIC	BIC
Rasch	8072	8122	8216
latent regression Rasch	8060	8114	8215
LLTM	8232	8244	8266
latent regression LLTM	8220	8236	8266

the latent regression Rasch model has a better fit than the Rasch model, although the difference is rather small, especially for the BIC. Based on a LR test, the difference is significant ($\chi^2(2) = 12.6$, $p < .01$) meaning that the goodness of fit of the Rasch model is lower.

Person property effects and residual person variance

There are a number of ways to express the results indicated by the estimated parameters. We mention several of them in the following paragraphs.

The estimated *effect of Trait Anger* is .057 on the logit scale, with a SE of .016, so that the effect is highly statistically significant ($p < .001$). The value of .057 is the change one would expect, given a change of one unit on the Trait Anger score – it corresponds to a multiplication of the odds ratio by 1.06. An alternative framework is provided by the standard deviation. An increase of one standard deviation (SD) in Trait Anger (instead of one unit) represents a multiplication of the odds by 1.32, and the difference between $-2SD$ and $+2SD$ represents a multiplication of the odds by 3.02. The effect of $+1SD$ on a .50 probability is to raise this probability to .57.

The estimated *effect of Gender* is .29 on the logit scale, with a SE of .20, so that the effect is not statistically significant. Males are not significantly more inclined to verbal aggression than females, but the odds for male students are nevertheless 1.34 times larger than the odds for female students. The effect of being male on a probability of .50 is to raise this probability to .57.

Since Trait Anger and Gender explain part of the original person variance, the *residual person variance* may be expected to be lower than the one estimated with the Rasch model. The estimated value of the person variance is 1.84, with a SE of .19, so that we must conclude that the individual differences that are not explained by Trait Anger and Gender are still highly statistically significant ($p < .001$). We note that the person variance is smaller than it was for the Rasch model.

In comparison with the residual person variance, the variance that is explained by Trait Anger is rather small: the variance of Trait Anger multiplied by the squared effect of Trait Anger is $(4.85^2 \times .057^2 =).08$, which is 4% when added to the residual person variance. This percentage represents a correlation of .20 between Trait Anger and the verbal aggression propensity as measured in a small set of specific situations. This low correlation is not surprising since typically situational behavior is not correlated higher than approximately .20 to .30 with trait measures (Mischel, 1968). The variance explained by Gender is even much smaller: the variance of Gender multiplied with the squared effect of Gender is $(.42^2 \times .29^2 =).02$, which is not significant. Thus, in terms of effect size, the effect of Trait Anger is small to moderate and the effect of Gender is small to vanishing.

Item parameters

The estimated *item parameters* vary from $-.57$ to $+4.16$. To interpret these values one needs to know the actual mean of the person effects. This mean is the result of adding three terms: (1) the mean of the normal distribution of ε (which is zero), (2) the average Trait Anger score (20.00) times the Trait Anger effect (.057), and (3) the average of Gender (the proportion of males: .23) times the effect of Gender (.29). The sum of these three terms is 1.20. When this reference value of 1.20 is subtracted from the original range ($-.57$ to $+4.16$), the result is -1.77 to $+2.96$, which is very close to the range obtained with the estimates from the Rasch model. This short discussion demonstrates how the parameter values are identified only up to an additive constant.

2.5 An item explanatory model: the LLTM

2.5.1 Formulation of the model

In the third model, the *linear logistic test model* (LLTM), item properties are used to explain the differences between items in terms of the effect they have on η_{pi}, and therefore on π_{pi}. The model differs from the Rasch model in that the contribution of item i is reduced to the contribution of the item properties and the values they have for item i (see also Table 2.2):

$$\eta_{pi} = \theta_p - \sum_{k=0}^{K} \beta_k X_{ik}, \tag{2.9}$$

in which X_{ik} is the value of item i on item property k ($k = 0, \ldots, K$), and β_k is the regression weight of item property k. Comparing Equation 2.9 with the corresponding equation for the Rasch model (see Equation 2.2), one can see that the item parameter β_i is replaced with a linear function:

$$\beta'_i = \sum_{k=0}^{K} \beta_k X_{ik}. \tag{2.10}$$

Note that in general β'_i will not equal β_i as the prediction will not be perfect.

Because the mean of the person distribution is fixed to zero, a property with a value of 1 for all items is needed (a constant predictor) to act as the intercept in Equation 2.10. Hence, we need an item predictor for $k = 0$, with $X_{i0} = 1$ for all values of i, so that β_0 is the item intercept. An alternative is to estimate the mean of the θ_p, and to omit the contribution of the constant predictor, so that in Equations 2.9 and 2.10 k would run from 1 to K. These remarks apply also to the fourth model: see Section 2.6.1.

The model in Equation 2.9 is called the 'linear logistic test model' (LLTM; Fischer, 1973) because the model is based on a logit link and on a linear combination of item properties in the linear component, and because it was first used for test data. Instead of estimating individual item effects, the effects of item properties are estimated. The term 'logistic' in the label of the model does not mean that the principle of a linear combination of item properties cannot be used for normal-ogive models. Substituting a probit link instead of a logit link is all that is needed to obtain the normal-ogive equivalent of the LLTM.

The LLTM also allows for interactions between the item properties. If one is interested in the interaction between two item properties, their product can be added as an additional item property.

Graphical representation

A graphical representation of the LLTM is given in Figure 2.7.

The difference between Figure 2.5 for the Rasch model and Figure 2.7 for the LLTM is that the contribution of each item is explained through the item properties (the Xs) and their fixed effects (the βs from 1 to K, and a constant β_0, the effect of the constant item predictor). The constant predictor is represented twice, as X_{i0} and Z_{p0}, because it is also used twice: for the fixed LLTM intercept (β_0) and for the random intercept (θ_p).

Comments and literature

Note that there is no error term in Equations 2.9 and 2.10 and hence, the prediction is assumed to be perfect. The model implies that the item effects can be perfectly explained from the item properties, that β_i from the Rasch model equals β_i' from Equation 2.10. This is a strong assumption, and it

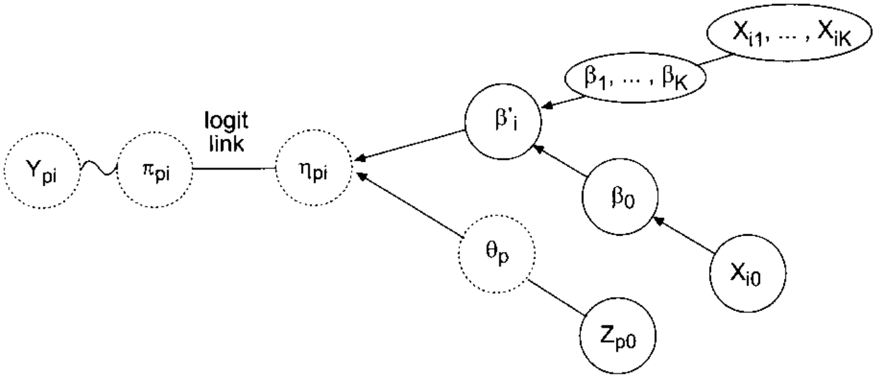

FIGURE 2.7. Graphical representation of the LLTM.

makes the model highly restrictive. But this constraint may be relaxed in more complex models. In Chapter 6, models are presented with an error component added to Equations 2.9 and 2.10.

The LLTM was developed by Fischer (1973, 1983). For an early application of regressing the item parameters on item properties, although the latter were not incorporated in the model, see Scheiblechner (1972). Fischer (1977) has presented a LLTM for multidimensional items, and later he described a general framework for designs with multidimensional items and different points in time, possibly with different subsets of items for different occasions (Fischer, 1989). For an overview of LLTM developments, see Fischer (1995).

2.5.2 Application of the LLTM

Three item properties are used in the LLTM for the verbal aggression data: Behavior Mode, Situation Type, and Behavior Type. The three properties are coded into four X-variables ($k = 1$ to 4), complemented with the constant item predictor ($k = 0$). We chose the coding given in Figure 2.8.

Behavior Mode		
predictor 1	Do = 1	Want = 0
Situation Type		
predictor 2	Other-to-blame = 1	Self-to-blame = 0
Behavior Type		
predictor 3	Curse, Scold = 1/2	Shout = -1
predictor 4	Curse, Shout = 1/2	Scold = -1

FIGURE 2.8. Coding scheme for the LLTM

Note that the coding scheme as presented in Figure 2.8 differs from the one used for the simple linear regressions in Chapter 1, since, except for the Behavior Type, dummy coding is used. This illustrates how alternative coding schemes are possible. For the Behavior Type, contrast coding with centering on the overall mean is used as in Chapter 1, because we are still interested in the effect of the behavioral features (Blaming, Expressing) in comparison with the mean. However, we will also report the estimates using dummy coded factors for Behavior Type: one for Curse versus the other two behaviors, and one for Scold versus the other two behaviors (Shout is the reference level). The Behavior Mode is coded as a dummy variable: Do is coded as 1, and Want as 0. Also the Situation Type is coded as a dummy variable, with Other-to-blame coded as 1, and Self-to-blame as 0. In order to include an intercept, an item predictor is added with a value of one for all items ($k = 0$).

The goodness-of-fit values of the LLTM are given in Table 2.3. The values are clearly inferior to those of the previous models. The LR test comparing the LLTM to the Rasch model is significant $\chi^2(19) = 159.6$ ($p < .001$), meaning that the goodness of fit of the LLTM is lower. The reason is that the 24 parameters for item effects are now reduced to only five, corresponding to the five item predictors (including the constant predictor). But see our discussion below regarding the estimates, where we conclude that the item properties have a very high explanatory value. This illustrates how choosing to use an explanatory model can be at the cost of a statistically significant lower goodness of fit even when the explanation is rather successful. See Chapter 6 for a solution to this by defining the item parameters as a random variable. As for the other models, we first discuss the results regarding the person variance.

Person variance

The estimated *person variance* is 1.86, with a SE of .20 and thus significant ($p < .001$). Note that the variance is smaller than for the Rasch model (where it was 1.98). This illustrates how the estimates for the person mode are slightly affected by a different approach for the item mode (explanatory instead of descriptive). This phenomenon can be explained as a scaling effect (Snijders & Bosker, 1999, pp. 227–228), which was also discussed in Chapter 1. The effect is due to the less than perfect explanation of the item parameters on the basis of the item properties (see next paragraph).

Item property effects

We no longer have estimates of the individual item parameters but instead we have estimates of the effects of the item properties. To find out the effect per item, the sum of the effects of the corresponding item property variables must be made, as will be illustrated below.

The estimated *effect of the Behavior Mode* is .67, with a SE of .06, so that this effect is also highly statistically significant ($p < .001$) – when going from wanting to doing, the odds are reduced to about half of their value for wanting. The odds decrease with (are divided by) a factor of about two, more precisely 1.96. If the probability of wanting were .50, then the reduction would yield a probability of .34.

The estimated *effect of the Situation Type* is -1.03, with a SE of .06, so that the effect is highly statistically significant ($p < .001$). The effect implies that when others are to blame, verbal aggression is more common than when oneself is to blame. When others are to blame, the odds increase by a factor 2.80. The effect on a probability of .50 would be to raise it to .74.

Recall that for the *effect of the Behavior Type* two predictors were used. The effect of the first (Curse and Scold vs Shout) is -1.36, with a SE of .05; and the effect of the second (Curse and Shout vs Scold) is $-.70$, with a SE of .05. Both effects are highly statistically significant ($p < .001$). From these effects it may be concluded that for the situations under investigation the blaming aspect of a behavior has a larger effect on its occurrence than the expression aspect. When both effects are combined, the values for the three behaviors are: $-1.36/2 - .70/2 = -1.03$ for Curse, $-1.36/2 + .70 = .02$ for Scold, and $1.36 - .70/2 = 1.01$ for Shout. Using odds to describe the effect size, the odds of cursing are 2.86 times higher than those of scolding, and the odds of scolding are in turn 2.69 times higher than those of shouting. The odds roughly increase with a factor of almost three when going from shouting to scolding, and when going from scolding to cursing. If the probability of scolding were .50 in a given situation, then the corresponding probabilities of cursing and shouting would be .74, and .27, respectively. Equivalent results are obtained with the dummy coding. The effects are -2.04 (SE is .07) for Curse, and $-.99$ (SE is .07) for Scold. Finally, the estimated effect of the *constant predictor* is .31, the estimation of the fixed intercept using the coding scheme of Figure 2.8. Given the mixed coding (contrast coding and dummy coding) this effect has no easy interpretation.

In order to reconstruct the individual item parameters from the LLTM, one has to add up the effects that correspond to the four item property variables and the constant. For example, the reconstructed parameter for "A bus fails to stop for me. I would want to scold" is .02 (Scold) + .00 (Want is the reference level) -1.03 (Other-to-blame) $+.31$ (constant) = $-.70$. The parameter as estimated on the basis of the Rasch model is $-.57$. The correlation between the item parameters as estimated with the Rasch model and the parameters as reconstructed from the LLTM is .94. Thus, although the LLTM fits significantly worse in a statistical sense, it does very well in explaining the item parameters, so that we may say it has a large effect size in this respect.

2.6 A doubly explanatory model: the latent regression LLTM

2.6.1 Formulation of the model

Finally, one can carry out both of the previous extensions by combining Equations 2.7 and 2.10 into the equation for the Rasch model (Equation 2.2), assuming that $\beta_i{}'$ is used in place of β_i. This yields the *latent regression LLTM*, a model that is explanatory for both the person mode and the item mode (see also Table 2.2):

$$\eta_{pi} = \sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p - \sum_{k=0}^{K} \beta_k X_{ik}. \tag{2.11}$$

As for the previous models, the model of Equation 2.11 has two parts: a person contribution and an item contribution. The person contribution is explained in terms of person properties and has an error term, while the item contribution is explained in terms of item properties and does not include an error term. This asymmetric construction is not a necessity, as will be seen in Chapter 6.

The model in Equation 2.11 is a GLMM with both person predictors and item predictors, each having a fixed effect, and a random intercept, which is the error term of the person contribution. The previous three models in this chapter can be obtained from Equation 2.11. Two kinds of modifications are needed to obtain the other three models: (a) to obtain the LLTM, the Zs are omitted, so that ε_p can be expressed as θ_p; and (b) to obtain the latent regression Rasch model, the Xs are just the item indicators ($X_{ik} = 1$ if $i = k$, $X_{ik} = 0$ otherwise, and $K = I$), so that for $k = i$ it holds that $\beta_k \mathrm{X}_{ik} = \beta_i$, and for $k \neq i$ it holds that $\beta_k X_{ik} = 0$. For the Rasch model both modifications are needed. Alternatively, these three models can be seen as being built up by adding complications to the basic building block of the Rasch model.

Graphical representation

Figure 2.9 gives a graphical representation of the latent regression LLTM. The difference with Figure 2.5 (the Rasch model) is that in Figure 2.9 for the latent regression LLTM both the contribution of each item and of each person is explained through properties, item properties with a fixed effect β_k, and person properties with a fixed effect ϑ_p, respectively. For the items, the effect of the constant predictor is β_0, while for the persons the effect of the constant predictor is a random effect, which appears as an error term ε_p. This is why both X_{i0} and Z_{p0} are included in the representation. Note that the circles containing $\beta_i{}'$ and θ_p are not needed. A direct connection of the arrows from the Xs and the Zs to η_{pi} is a more parsimonious but perhaps less interpretable representation.

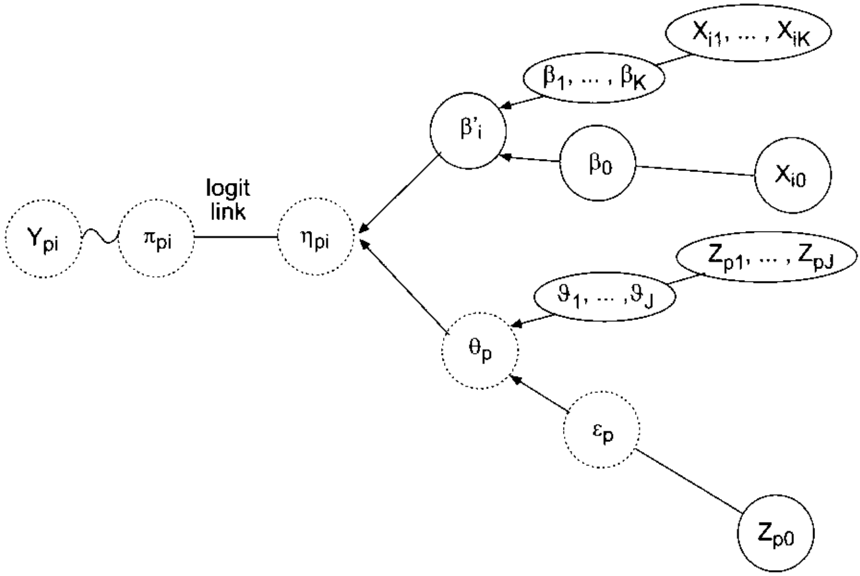

FIGURE 2.9. Graphical representation of the latent regression LLTM.

Literature

The latent regression LLTM is simply a combination of the latent regression idea with the LLTM, and this is why we call this combined model here the 'latent regression LLTM'. It is described theoretically in Zwinderman (1997), and Adams, Wilson and Wu (1997).

2.6.2 Application of the latent regression LLTM

The fit indices for the latent regression LLTM are given in Table 2.3. The goodness of fit is slightly better than for the LLTM, for the same reasons that the latent regression Rasch model had a slightly better goodness of fit than the Rasch model. The LR test comparing the latent regression LLTM to the LLTM is significant ($\chi^2(2) = 12.6$, $p < .001$). We will not note the specific effect estimates here, as the estimated *person property effects* are about the same as those obtained with the latent regression Rasch model, and also the estimated *item property effects* are about the same as those obtained with the LLTM.

It is noteworthy that the *residual person variance*, after the estimated effect of Trait Anger and Gender is accounted for, amounts to 1.73 in the latent regression LLTM, while it was 1.84 in the corresponding latent regression Rasch model. Again, the more flexible the model is for the estimation of the item effects, the larger the variance is of the (residual) person effects, as could be expected from the scaling effects discussed earlier.

2.7 Enlarging the perspective

The four models we have presented are chosen to illustrate the contrast between descriptive and explanatory models. They are only an introductory selection. In order to cover the broad variety of item response models, we need an enlargement of the perspectives. In principle the extensions can relate to the three parts of a GLMM: the random component, the link function, and the linear component.

Regarding the first two parts, the extension of the models to multicategorical data has consequences for the link function and the random component. We will not go as far as extending the models also to count data, however, which would require a logarithmic link and a Poisson distribution for the random component. Regarding the linear component, the extensions concern not only the type of predictors and the type of effects, but also the linear nature of the component, since some of the item response models are not generalized linear mixed models but nonlinear mixed models. Examples of nonlinear mixed models are the two- and the three-parameter logistic models (2PL and 3PL models), and the multidimensional two-parameter models. Finally, the assumption of local independence will be relaxed.

For all these models, the parameters can either be descriptive parameters or explanatory parameters. Explanatory parameters are effects of properties, or in other words, of external variables. Descriptive parameters are either random effects or fixed effects of predictors that are not properties but indicators. This distinction, which is at the basis of the presentation of four models in this chapter, will be extrapolated in the following chapters.

Chapter 3 discusses extensions to multicategorical data. Other extensions are presented from Part II on. Chapter 4 describes more thoroughly than the previous chapters the statistical background for this volume.

2.8 Software

2.8.1 Rasch model (verbal aggression data)

The basic options that were used are described in Section 2.3.2. In later chapters, the basic options are reported in the sections on software.

Code

```
PROC NLMIXED data=aggression_dich method=gauss
technique=newrap noad qpoints=20;
PARMS b1-b24=1 sd0=1;
beta= b1*x1+b2*x2+b3*x3+b4*x4+b5*x5+b6*x6+b7*x7
+b8*x8+b9*x9+b10*x10+b11*x11+b12*x12+b13*x13+b14*x14
+b15*x15+b16*x16+b17*x17+b18*x18+b19*x19+b20*x20
```

```
+b21*x21+b22*x22+b23*x23+b24*x24;
ex=exp(theta-beta);
p=ex/(1+ex);
MODEL y ~ binary(p);
RANDOM theta ~ normal(0,sd0**2) subject=person;
ESTIMATE 'sd0**2' sd0**2;
RUN;
```

Comments

1. The data set is called **aggression_dich** (see website mentioned in the Preface). The data matrix contains the data in one long string and the values of the design factors corresponding with each observation (see Chapter 12).

2. In the PARMS statement, the parameters are introduced together with their initial values.

3. Next, the formula for the probability is built up from two ingredients: beta and theta. The beta part is based on the 24 item indicators (x1 to x24) and their weights (b1 to b24). The theta part is just a single term (θ_p, but see the software for the next application). With the basic ingredients of theta and beta, the formula for the probability is constructed. Instead of building up the formula in steps, one can as well give the formula in one step.

4. In the MODEL statement, it is specified that the observations follow a Bernoulli distribution (binary) with parameter p (π_{pi}).

5. In the RANDOM statement the distribution of theta is specified, over persons (subject=person), with mean zero and a variance that is the squared value of sd0 (σ_θ). The value that is estimated is therefore the SD and not the variance.

6. This is why an ESTIMATE statement is added, so that also the variance is estimated, with label 'sd0**2' (the label may differ from the symbol in the software; e.g., vartheta would be another label).

7. The code for the LLTM will not be shown, but is analogous: x1 to x24 is replaced with x1 to x5 (the coded design factors) with their weights.

2.8.2 Latent regression Rasch model (verbal aggression data)

The options are the same as for the Rasch model.

Code

```
PROC NLMIXED data=aggression_dich method=gauss
technique=newrap noad qpoints=20;
PARMS b1-b24=1 sd0=1 g1-g2=0;
theta=eps + g1*anger + g2*male;
beta= b1*x1+b2*x2+b3*x3+b4*x4+b5*x5+b6*x6+b7*x7
+b8*x8+b9*x9+b10*x10+b11*x11+b12*x12+b13*x13+b14*x14
+b15*x15+b16*x16+b17*x17+b18*x18+b19*x19+b20*x20
+b21*x21+b22*x22+b23*x23+b24*x24;
ex=exp(theta-beta);
p=ex/(1+ex);
MODEL y ~ binary(p);
RANDOM eps ~ normal(0,sd0**2) subject=person;
ESTIMATE 'sd0**2' sd0**2;
RUN;
```

Comments

The two differences with the estimation of the Rasch model are:
1. `theta` is now defined as a sum of the Gender effect, the Trait Anger effect, and a random term `eps`, in correspondence with how `theta` is defined in the latent regression Rasch model. The person properties are `anger` and `male` (the Zs), and their weights `g1` and `g2` (the ϑs).
2. It is now the distribution of `eps` that is defined, instead of the distribution of `theta`.

2.9 Exercises

1. Why is no intercept (β_0) used in the Rasch model?

2. Redraw Figure 2.5 for a model with fixed person effects and random item effects.

3. How should one interpret the intercept in the LLTM? Suppose the intercept would be fixed to zero, while the mean of the θ-distribution is free. What would be the consequence of this? How do β_0 and the mean of θ relate to one another?

4. Suppose that for Do vs Want not a dummy coding would have been used but contrast coding (Do $= 1$, Want $= -1$). What would then have been the weight of this predictor?

5. θ_p can be removed from Figure 2.6. How would the new figure look

like then? Would ε_p be the random intercept? If yes, how can an error term be the measure of a latent trait, and how would the trait be defined?

2.10 References

Adams, R.J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1–23.

Adams, R.J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22*, 47–76.

Akaike, M. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Andersen, E.B., & Olsen, L.W. (2001). The life of Georg Rasch as a mathematician and as a statistician. In A. Boomsma, M.A.J. van Duijn & T.A.B. Snijders (Eds), *Essays on Item Response Theory* (pp. 3–24). New York: Springer.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds), *Statistical Theories of Mental Test Scores* (pp. 394–479). Reading, MA: Addison-Wesley.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika, 46*, 443–459.

Bond, T., & Fox, C. (2001). *Applying the Rasch Model: Fundamental Measurement in Human Sciences.* Mahwah, NJ: Lawrence Erlbaum.

Bozdogan, H. (1987). Model selection for Akaike's information criterion (AIC). *Psychometrika, 53*, 345–370.

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

Fischer, G.H. (1968). Neue Entwicklungen in der psychologischen Testtheorie. In G.H. Fischer (Ed.), *Psychologische Testtheorie* (pp. 15–158). Bern: Huber.

Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 3*, 359–374.

Fischer, G.H. (1974). *Einführung in die Theorie Psychologischer Tests.* Bern: Huber.

Fischer, G.H. (1977). Linear logistic trait models: Theory and application. In H. Spada & W.F. Kampf (Eds), *Structural Models of Thinking and Learning* (pp. 203–225). Bern: Huber.

Fischer, G.H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika, 46*, 59–77.

Fischer, G.H. (1983). Logistic latent trait models with linear constraints. *Psychometrika, 48*, 3–26.

Fischer, G.H. (1989). An IRT-based model for dichotomous longitudinal data. *Psychometrika, 54*, 599–624.

Fischer, G.H. (1995). Linear logistic models for change. In G.H. Fischer & I. Molenaar (Eds), *Rasch Models. Foundations, Recent Developments and Applications* (pp. 157–201). New York: Springer.

Fischer, G.H., & Molenaar, I. (1995) (Eds), *Rasch Models. Foundations, Recent Developments, and Applications.* New York: Springer.

Follmann, D.A. (1988). Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika, 53*, 553–562.

Fox, J.P., & Glas, C.A.W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika, 68*, 169–191.

Hoijtink, H., & Boomsma, A. (1995). On person parameter estimation in the dichotomous Rasch Model. In G.H. Fischer & I. Molenaar (Eds), *Rasch Models. Foundations, Recent Developments and Applications* (pp. 53–68). New York: Springer.

Junker, B., & Sijtsma, K. (2001). Nonparametric item response theory. Special issue. *Applied Psychological Measurement, 25.*

Masters, G.N., Adams, R.A., & Wilson, M. (1990). Charting of student progress. In T. Husen & T.N. Postlewaite (Eds), *International Encyclopedia of Education: Research and Studies. Supplementary Volume 2.* (pp. 628–634) Oxford: Pergamon Press.

Mischel, W. (1968). *Personality Assessment.* New York: Wiley, 1968.

Mislevy, R.J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement, 11*, 81–91.

Molenaar, I. (1995). Estimation of item parameters. In G.H. Fischer & I. Molenaar (Eds), *Rasch Models. Foundations, Recent Developments and Applications* (pp. 39–57). New York: Springer.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM Manual.* Technical Report 2001/01. Department of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen, Denmark: Danish Institute for Educational Research.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Theory of Probability* (Vol. IV, pp. 321–333). Berkeley, CA:

University of California Press.

Rasch, G. (1967). An informal report on a theory of objectivity in comparisons. In L.J.Th. van der Kamp & C.A.J. Vlek (Eds), *Psychological Measurement Theory* (pp. 1–19). Proceedings of The NUFFIC international summer session. Leiden: University of Leiden.

Read, T.R.C., & Cressie, N.A.C. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data.* New York: Springer.

Rost, J. (2001). The growing family of Rasch models. In A. Boomsma, M.A.J. van Duijn & T.A.B. Snijders (Eds), *Essays on Item Response theory* (pp. 25–42). New York: Springer.

SAS Institute (1999). *SAS OnlineDoc (Version 8)* (software manual on CD-ROM). Cary, NC: SAS Institute.

Scheiblechner, H. (1972). Das Lernen and Lösen komplexer Denkaufgaben. *Zeitschrift für Experimentelle und Angewandte Psychologie, 3*, 456–506.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

Sijtsma, K., & Molenaar, I. (2002). *Introduction to Nonparametric Item Response Theory.* Thousand Oaks, CA: Sage.

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis.* London: Sage.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds), *Test Scoring* (pp. 73–140). Mahwah, NJ: Lawrence Erlbaum.

Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer.

Verhelst, N.D., & Eggen, T.J.H.M. (1989). Psychometrische en statistische aspecten van peilingsonderzoek (PPON rapport 4). Arnhem: Cito.

Verhelst, N.D., & Glas, C.A.W. (1995). The one parameter logistic model. In G.H. Fisher & I. Molenaar (Eds), *Rasch Models. Foundations, Recent Developments, and Applications* (pp. 215–237). New York: Springer.

Wainer, H., Vevea, J.L., Camacho, F., Reeve, B.B., Rosa, K., Nelson, L., Swygert, K., & Thissen, D. (2001). Augmented scores – "Borrowing strength" to compute scores based on small numbers of items. In D. Thissen & H. Wainer (Eds), *Test Scoring* (pp. 343–387). Mahwah, NJ: Lawrence Erlbaum.

Wald, A. (1941). Asymptotically most powerful tests of statistical hypotheses. *Annals of Mathematical Statistics, 12*, 1–19.

Warm, T.A. (1989). Weighted likelihood estimation of ability in item response models. *Psychometrika, 54*, 427–450

Wilson, M. (2003). On choosing a model for measuring. *Methods of Psychological Research - Online, 8*(3), 1-22.

Wilson, M. (2005). *Constructing Measures: An Item Modeling Approach.* Mahwah, NJ: Lawrence Erlbaum.

Wright, B. (1968). Sample-free test calibration and person measurement. *Proceedings 1967 Invitational Conference on Testing* (pp. 85–101). Princeton: ETS.

Wright, B. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*, 97–116.

Wright, B. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice, 16*, 33–52.

Wright, B., & Stone, M. (1979). *Best Test Design.* Chicago, IL: MESA.

Wu. M.L., Adams, R.J., & Wilson, M. (1998). *ACERConquest.* Hawthorn, Australia: ACER Press.

Zwinderman, A.H. (1991). A generalized Rasch model for manifest predictors. *Psychometrika, 56*, 589–600.

Zwinderman, A.H. (1997). Response models with manifest predictors. In W.J. van der Linden & R.K. Hambleton (Eds), *Handbook of Modern Item Response Theory* (pp. 245–256). New York: Springer.