1

Person regression models

Solutions to the exercises

1. For each person, several responses are observed. On each observation occasion, the person answers one specific item. The occasion therefore is characterized by the item that is answered by the person. Note that persons typically answer each item only once, but this is not necessarily the case. It may for instance also be true that the same item is answered twice. In this case, the values for the item indicators are the same for the two observation occasions.
   In the hierarchical generalized linear models discussed in this chapter, the first level equation may describe the variation of the covert continuous variable as a function of occasion characteristics and a residual term (as in Equation 5.3). The residual term is distributed logistically. Alternatively, the first level model may describe the variation of the logit of the probability of a correct answer as a function of occasion characteristics. In the last case, the first level model does not explicitly include a residual term. The random fluctuation at the first level nevertheless is implicitly taken into account by modeling the probability. It is only when the probability of a correct answer is zero or one, that there is no random fluctuation. More specifically, Y_{pi} ~Bernoulli(π_{pi}) and the variance of Y equals π_{pi}(1-π_{pi}).

2. To use PQL1, click on 'Nonlinear' at the bottom of the Equations-window. Select 1st order linearisation and PQL. Parameter estimates are given in the table below. The variance estimate, the standard errors and the absolute value of the fixed parameter estimates are systematically smaller compared to the corresponding values when using PQL2 (see Chapter 12 for a discussion of the differences between the estimation methods).

3. If the school-level is ignored, the between-pupil variance estimate (0.86) is almost equal to the sum of the between-pupil and the between-school variance estimates of the complete model (0.63 + 0.29). It can be shown for balanced group sizes that, when ignoring the highest level, the variance estimate on this level is added to the variance estimate on the level below. The SE associated with the between-pupil variance increases. The fixed item parameters are only slightly affected, but the SEs of these parameters decreased substantially (from around .11 to around .06).

TABLE 1.1. Parameter estimates and standard errors for the 2-level latent regression model (Verbal aggression data)

	Estimate (SE)
Fixed effects of	
Item1 (Bus - Want - Curse)	0.01 (0.36)
Item2 (Bus - Want - Scold)	-0.61 (0.36)
Item3 (Bus - Want - Shout)	-1.08 (0.36)
Item4 (Train -Want - Curse)	0.52 (0.37)
Item5 (Train - Want - Scold)	-0.48 (0.36)
Item6 (Train -Want - Shout)	-1.14 (0.36)
Item7 (Grocery - Want - Curse)	-0.65 (0.36)
Item8 (Grocery - Want - Scold)	-1.81 (0.36)
Item9 (Grocery - Want - Shout)	-2.61 (0.37)
Item10 (Call - Want - Curse)	-0.12 (0.36)
Item11 (Call - Want - Scold)	-1.49 (0.36)
Item12 (Call - Want - Shout)	-2.15 (0.36)
Item13 (Bus - Do - Curse)	0.01 (0.36)
Item14 (Bus - Do - Scold)	-0.78 (0.36)
Item15 (Bus - Do - Shout)	-1.99 (0.36)
Item16 (Train - Do - Curse)	-0.32 (0.36)
Item17 (Train - Do - Scold)	-1.21 (0.36)
Item18 (Train - Do - Shout)	-2.57 (0.37)
Item19 (Grocery - Do - Curse)	-1.36 (0.36)
Item20 (Grocery - Do - Scold)	-2.59 (0.37)
Item21 (Grocery - Do - Shout)	-4.00 (0.40)
Item22 (Call - Do - Curse)	-0.48 (0.36)
Item23 (Call - Do - Scold)	-1.52 (0.36)
Item24 (Call - Do - Shout)	-3.07 (0.37)
Gender	0.30 (0.18)
Trait Anger	0.05 (0.02)
Variance components	
Level 2 (persons)	1.63 (0.15)

4. To allow a different between-pupil variance in the three types of schools, the effect of the type of school is defined to be random over persons. To do this, you need three new binary variables indicating the type of school, for instance with the following names: PUBLIC, PRIVATE and CATHOLIC. Replace the line 'random intercept/sub=student(school);' in the program given in 5.9.2 by 'random public private catholic /sub=student(school);'. To estimate the parameters, you will also have to include the three binary variables in the fixed part of the model (this is in the model statement). Unfortunately, SAS may experience problems when estimating the parameters of this complex model for the large data set. In MLwiN, define three indicator variables for the type of school. Define the coefficients of these variables as randomly varying at the student level, instead of the randomly varying intercept. Although the variables do not need to be included in the fixed part when using MLwiN, it may be a good idea not to assume that there are no overall differences between the types of schools. To ease the comparison with the results of the descriptive model with a homogeneous student level variance, we estimate the model with RIGLS and PQL1, and with including the indicator variables in the random part only. This results in a between-student variance of 0.56 (Catholic schools), 0.49 (private schools), and 0.64 (public schools), providing evidence that differences between students are largest in public schools, smallest in private schools. Corresponding standard error estimates are 0.095, 0.093 and 0.029. The standard error for public schools is substantially smaller, since the majority of schools included in the sample are public schools. Covariance estimates are equal to zero, since persons belong to one single school. Other parameter estimates are comparable with those of the model with a homogeneous level-2 variance (Table 5.3).

5. First define a new variable, indicating the science items. This variable, together with the indicator for the mathematics items is included in the model as a covariate with a coefficient that varies randomly at the student level. The random coefficient for the constant at the student level is removed. To ease the comparison with the results of the descriptive 3-level model given in Table 5.3, we use RIGLS and PQL1 to estimate the model parameters. Fixed parameters and corresponding standard errors hardly change. Also the between-school variance estimate and corresponding standard error estimate remain approximately the same. The between-student variance estimate is 0.72 with a standard error of 0.03 for the math items, 0.70 with a standard error of 0.03 for the science items. Mathematics therefore cannot be regarded as an additional source of variance. The estimate of the covariance at the student level between the mathematics and science items is 0.55, with an estimated standard error of 0.028. Di-

viding the covariance by the squared variance estimates, results in a correlation estimate of .77. This means that in general students who perform relatively good for mathematics, also perform relatively good for science. Note that one could also model variance heterogeneity at the third level, by estimating the between-school variance for mathematics and science items separately. The covariance estimate gives an idea about the correlation over schools between the mathematics performance and the science performance.