6

The Logistic Model for Dichotomous Data and Proportions

The models discussed so far assume a continuous dependent variable and a normal error distribution. If the dependent variable is a scale in which the responses to a large number of questions are summated to one score, the data generally approximate normality. However, there are situations in which the assumption of normality is clearly violated. For instance, in cases where the dependent variable is a single dichotomous variable, both the assumption of continuous scores and the normality assumption are obviously untrue. If the dependent variable is a proportion, the problems are less severe, but both the assumptions of continuous scores and normality are still violated. Also, in both cases, the assumption of homoscedastic errors is violated.

The classical approach to the problem of non-normally distributed variables and heteroscedastic errors is to apply a transformation to achieve normality and reduce the heteroscedasticity, followed by a traditional analysis with ANOVA or multiple regression. To distinguish this approach with the generalized linear modeling approach explained later, where the transformation is part of the statistical model, it is often referred to as an empirical transformation. Some general guidelines for choosing a suitable transformation have been suggested for situations in which a specific transformation is often successful (e.g., Kirk, 1968; Mosteller and Tukey, 1977). For proportions some recommended transformations include, for instance: the arcsine transformation that is given by $f(p) = 2 \text{ arcsine}(\sqrt{p})$, the logit or logistic

transformation f(p) = logit(p) = ln(p/(1-p)), and the probit or inverse Normal transformation f(p)=Φ^{-1}(p), where Φ^{-1} is the inverse Normal distribution. Thus, for proportions, we can use the logit transformation, and use standard regression procedures on the transformed variable:

$$\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (6.1)$$

When the dependent variable is a frequency count of events with a small probability, such as the number of errors made in a school essay, the data tend to follow a Poisson distribution, which can often be normalized by taking the square root of the scores: f(x) = $\sqrt{x}$. When the data are highly skewed, which is usually the case if, for instance, reaction time is the dependent variable, a logarithmic transformation is often used: f(x) = ln(x), or the reciprocal transformation: f(x)=1/x. For reaction times the reciprocal transformation has the nice property that it transforms a variable with an obvious interpretation: reaction time, into another variable with an equally obvious interpretation: reaction speed.

Such transformations have the disadvantage that they seem ad hoc, and may encounter problems in specific situations. For instance, if we model dichotomous data, both the logistic and the probit transformations break down, because for values 0 and 1 these functions are not defined. In fact, *no* transformation can ever transform a dichotomous variable, which takes on only two values, into any resemblance of a Normal distribution.

6.1 GENERALIZED LINEAR MODELS

The modern approach to the problem of non-normally distributed variables is to include the necessary transformation and the choice of the appropriate error distribution (not necessarily a Normal distribution) explicitly in the statistical model. This class of statistical models is called *generalized linear models* (McCullagh & Nelder, 1989; Gill, 2000). Generalized linear models are defined by three components:

1) an outcome variable y with a specific error distribution, that has mean μ,

2) a linear additive regression equation, which predicts an unobserved outcome variable η,

3) and a *link function* that links the predicted values for η to the observed values of y by the function $\eta=f(\mu)$.

If the link function is the identity function ($f(x)=x$) and the error distribution is normal, the generalized linear model simplifies to ordinary multiple regression analysis. The ordinary multiple regression model can be specified as a generalized linear model by stating that:

1) the probability distribution is $N(\mu,\sigma^2)$ with mean μ and variance σ^2,
2) the linear predictor is the multiple regression equation for η, e.g., $\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2$,
3) the link function is the identity function given by $\eta=f(\mu)=\mu$,

(cf. Aittkin, Francis, Anderson & Hinde, 1989, chap. 2). This is nearly the same as writing the normal regression equation as $(y + \varepsilon) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ with $\varepsilon \sim N(0,\sigma^2)$. The difference is, that specifying a model with an additive error term $(y + \varepsilon)$ as in the common regression model, does not work well if we assume non-normal distributions or non-linear models. There, ε may not have a simple distribution, or its variance may depend on the mean. If we transform the response variable, we are assuming that on the transformed scale, the error distribution is normal. Using the structure of a generalized linear model, the error distribution is separated from the (non-linear) link function.

The formulation of the regression model as a generalized linear model gives us considerable freedom to specify nonlinear models for non-normal data. For instance, a generalized linear model for proportions is given by:

1) the probability distribution is binomial (μ) with mean μ,
2) the linear predictor is the multiple regression equation for η, e.g., $\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2$,
3) the link function is the logistic function given by $\eta=\text{logit}(\mu)$.

Other link functions and error distributions are discussed by McCullagh & Nelder (1989) and Aitkin et al. (1989). The estimation method uses the inverse of the link function to predict the response variable. The inverse function for the logit is $g(x)=\exp(x)/(1+\exp(x))$. We could call this the *expit*, and write the model for proportions as $p = \text{expit}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$. However, this

does not show that the outcome has a binomial distribution, and in reporting the results of an analysis with a generalized linear model it is more usual to list the three components explicitly. But the inverse equation p = expit ($\beta_0 + \beta_1 X_1 + \beta_2 X_2$) makes clear why modeling dichotomous data now works; generalized linear modeling does not apply a logistic transformation to the observed values 0 and 1, which cannot work, but it applies the inverse transformation to the predicted values.

In principle many different error distributions can be used with any link function. But each distribution has a specific link function for which sufficient statistics exist, which is called the *canonical link* function. Table 6.1 presents some canonical link functions and the corresponding error distribution. The canonical link has some desirable statistical properties, and McCullagh and Nelder (1989, chap 2.) express a mild preference for using it. However, there is no compelling reason to use only canonical links, and other link functions may be better in some circumstances.

Table 6.1 Some canonical link functions and their corresponding error distribution

Response	link	name	distribution
continuous	$\eta=\mu$	identity	Normal
proportion	$\eta=\ln(\mu/(1-\mu))$	logit	binomial
count	$\eta=\ln(\mu)$	log	Poisson
positive	$\eta=\mu^{-1}$	inverse	gamma

Figure 6.1 shows the relation between the value p for a proportion and the transformed values using a logit or probit transformation.

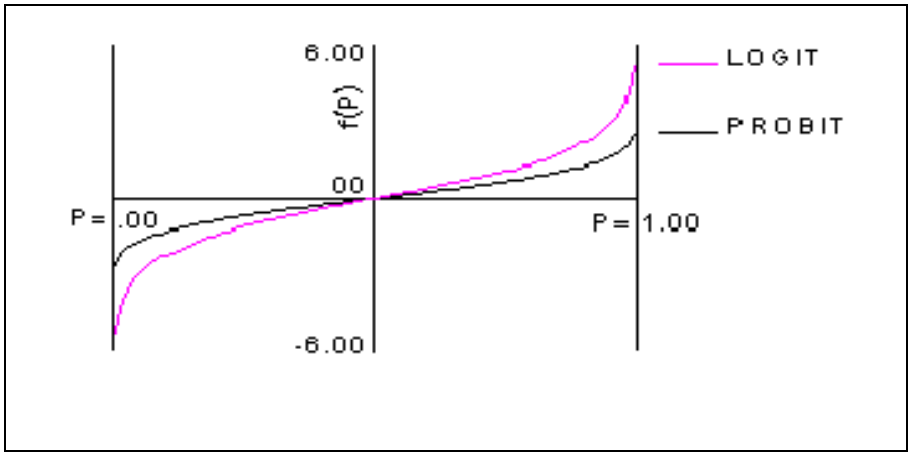

Figure 6.1 Plot of Logit and Probit transformed proportions P

Figure 6.1 makes clear, that the logit transformation spreads the proportions close to 0.00 or 1.00 over a larger range on the transformed scale than the probit transformation. If the proportions are close to zero or one, and the interest is in these extreme proportions, the logit is a better transformation, because it gives these proportions more weight in the estimation. There are other transformations, such as the log-log transformation, given by f(p) = -log(-log(p)), and the complementary log-log transformation, given by f(p) = log(-log(1-p)), that are sometimes used. These functions have the interesting property that they are asymmetric. For instance, for

proportions larger than 0.5, the log-log function behaves much like the logit, while for proportions smaller than 0.5, it behaves more like the probit. The complementary log-log function behaves in the opposite way. Again, McCullagh and Nelder (1989, p 108-110) express a mild preference for the logit link function. When the modeled proportions are all between 0.1 and 0.9, the difference between the link functions is negligible.

6.2 MULTILEVEL GENERALIZED MODELS

Multilevel generalized models have been described by Wong and Mason (1985), Longford (1990), Mislevy and Bock (1989), and Goldstein (1991, 1995). In generalized multilevel models, the multilevel structure appears in the linear regression equation of the generalized linear model. Thus, a two-level model for proportions is written as follows (cf. equation 2.5):

$$Y_{ij} = \pi_{ij} \qquad \pi \sim \text{Binomial}\,(n_{ij}, \mu) \tag{6.2}$$

$$\pi_{ij} = \text{logit}(\gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j}) \tag{6.3},$$

which states that we use a logit link function, and that conditional on the predictor variables, π_{ij} is assumed to have a error binomial distribution. Note that, in general, the lowest level residual variance e_{ij} is not in the model equation, because it is part of the generalized linear model specification. If the error distribution is binomial, the residual error variance is a function of the (unknown) population proportion π_{ij}: $\sigma^2 = (\pi_{ij}/(1-\pi_{ij})$. Some software allows the specification or estimation of a scale factor for the lowest level variance. If the scale factor is set at one, the assumption is made that the observed errors follow the theoretical error distribution exactly. If the scale factor is significantly higher or lower than one, there is *overdispersion* or *underdispersion*. Although the inclusion of a scale factor for the error distribution improves the fit, and in this way takes care of the problem, if the scale factor is very different from one it is good practice to examine the problem. Overdispersion can occur if the data have a strong grouping structure that is not included in the model, or if there are extreme outliers. Underdispersion often indicates a misspecification of the model, such as the omission of large

interaction effects.

6.2.1 Estimating generalized multilevel models

The parameters of generalized linear models are estimated using maximum likelihood methods. Multilevel models are generally also estimated using maximum likelihood methods, and it is clear that combining multilevel and generalized linear models leads to complex models and estimation procedures. The prevailing approach, implemented e.g., in MlwiN, HLM, and Prelis, is to approximate the nonlinear function by an almost linear function, and to embed the multilevel estimation in the generalized linear model. This approach is a quasi-likelihood approach, and it confronts us with two choices that must be made. The nonlinear function is linearized using an approximation that is known as Taylor series expansion. Taylor expansion approximates a nonlinear function by an infinite series of terms. Often only the first term of the series is used, which is referred to as a first order approximation. When the second term is also used, the approximation is better. So the first choice is whether to use a first order or a second order approximation. The second choice also involves the Taylor expansion. Taylor linearization of a function depends on the values of its parameters. Maximum likelihood estimation proceeds iteratively, starting with approximate values, which are then improved in each successive iteration. Thus, the parameter values change during the iterations. In consequence, the Taylor expansion must be repeated at each iteration, using the current estimated values of the model parameters. And this presents us with the second choice. The Taylor expansion can use the current values of the fixed part only, which is referred to as marginal quasi-likelihood (MQL), or it can be improved by using the current values of the fixed part plus the residuals, which is referred to as penalized (or predictive) quasi-likelihood (PQL).

Estimation procedures for generalized multilevel models are discussed by Goldstein (1995, chapters 5 & 7), including procedures to model extra variation at the lowest level. Rodriguez and Goldman (1995) show in simulated data sets with a dichotomous response variable that if the groups at the lowest level are small, both the fixed and the random effects are severely underestimated by the standard first order MQL method. Goldstein and Rasbash (1996) show that using PQL and second order estimates produce estimates that have only a small bias.

Browne (1998) repeats their analysis, using a much larger simulation setup. The extend of the bias can be judged from Table 6.2, which summarizes some of Browne's findings.

Judging from the results in Table 6.2, first order MQL estimation appears almost useless, especially regarding the second level variance estimate. However, Goldstein and Rasbash (1996) argue that the data structure of this simulation is extreme, because there are large variances in combination with very small groups. In less extreme data sets, the bias is much smaller, and even first order MQL produces acceptable estimates. Goldstein (1995) also warns that using second order PQL sometimes creates problems in estimation. This explains or choice problem. If second order estimation and penalized quasi-likelihood are always better, then why not always use these? The problem is that with complex models and/or small data sets there may be convergence problems, and we may be forced to use first order MQL. Goldstein and Rasbash (1996) suggest using bootstrap methods to improve the quasi-likelihood estimates, and Browne (1998) explores Bayesian methods. These approaches will be treated in chapter XX.

Table 6.2 Summary of simulation comparing MQL and PQL (Browne, 1998)

True value	**MQL - 1**	**PQL – 2**
β_0=0.65	0.47	0.61
β_1=1.00	0.74	0.95
β_2=1.00	0.75	0.96
β_3=1.00	0.73	0.94
σ_e^2=1.00	0.55	0.89
σ_u^2=1.00	0.03	0.57

It is important to stress that the approach described above is a quasi-likelihood method. Since the likelihood that is maximized is not the real likelihood, the test procedures based on comparing the deviances of the model (which equal minus 2 times the log likelihood) are not very accurate. The Wald test, or procedures based on bootstrap or Bayesian methods are preferred.

Some software does not use the quasi-likelihood approach. Hedeker's programs (cf. Hedeker & Gibbons, 1996a, 196b) use numerical integration to maximize the correct likelihood and HLM 5.0 uses a different method to accomplish the same. If these programs are used, test procedures based on the deviance are appropriate.

6.3 EXAMPLE: ANALYZING PROPORTIONS

The example concerns data from a meta-analysis of studies that compared face-to-face, telephone, and mail surveys on various indicators of data quality (De Leeuw, 1992; for a more thorough analysis see Hox & De Leeuw, 1994). One of these indicators is the response rate; the number of completed interviews divided by the total number of eligible sample units. Overall, the response rates differ between the three data collection methods. In addition, the response rates also differ across studies, which makes it interesting to analyze what study characteristics account for these differences.

These meta-analysis data have a multilevel structure. The lowest level is the 'condition-level,' and the higher level is the 'study-level.' There are three variables at the condition level: the proportion of completed interviews in that specific condition, the number of eligible respondents in that condition, and a categorical variable indicating the data collection method used. The categorical data collection variable has three categories: `face-to-face', `telephone' and `mail.' To use it in the regression equation, it is recoded into two dummy variables: a `telephone-dummy' and a `mail-dummy.' In the `mail' condition, the mail-dummy equals one, and in the other two conditions it equals zero. In the `telephone' condition, the telephone-dummy equals one, and in the other two conditions it equals zero. The face to face condition is the reference category, indicated by a zero for both the telephone- and the mail-dummy. There are three variables at the study level: the year of publication (0 = 1947, the oldest study), the saliency of the questionnaire topic (1 = not salient, 3 = highly salient), and the way the response rate is calculated. If the response rate is calculated by dividing the response by the total sample size, we have the completion rate, if the response rate is calculated by dividing by the sample size corrected for sampling frame errors, we have the response rate. Most studies compared only two of the three data collection methods; a few compared all three. Omitting missing values, there are 47 studies in which a total of 105 data collection conditions are compared. The data set is described in the appendix.

The dependent variable is the response rate. This variable is a proportion: the number of completed interviews divided by the number of eligible respondents. If we model these

proportions directly by normal regression methods, we encounter two critical problems. The first problem is the fact that proportions do not have a normal distribution, but a binomial distribution, which (especially with extreme proportions and/or small samples) invalidates several assumptions of the normal regression method. The second problem is that a normal regression equation might easily predict values larger than 1 or smaller than 0 for the response rate, which are impossible values for proportions. Using the generalized linear (regression) model for the proportion p of potential respondents that are responding to a survey solves both problems, which makes it a more appropriate model for these data.

As I outlined above, the generalized linear model has three distinct components: 1) a specific error distribution, 2) a linear regression equation, and 3) a link function. The customary link function for binomial data is the *logit* function: logit(x) = ln(x/(1-x). The corresponding canonical error distribution is the binomial distribution.

The hierarchical generalized linear model for our response rate data can be described as follows. In each condition we have a number of individuals who may or may not respond. The population probability of responding is given by π_{ij}, that is, for each individual r in each condition i of study j the probability of responding is the same. Note that we could have a model where each individual's probability of responding varies, with individual level covariates to model this variation. Then, we would model this as a three-level model, with binary outcomes at the lowest (individual) level.[1] Since in this meta-analysis example we do not have individual data, the lowest level is the condition-level, with conditions (data collection methods) nested within studies.

Let p_{ij} be the observed proportion respondents in condition i of study j. At the lowest level, we use a linear regression equation to predict logit (p_{ij}). The simplest model, corresponding to the intercept-only model in ordinary multilevel regression analysis is given by:

$$\text{logit}\,(p_{ij}) = \beta_{0j} \qquad (6.4)$$

Note again that the usual lowest level error term e_{ij} is not included in equation (6.4). In the binomial distribution the variance of the observed proportion depends only on the population

[1]Binary or dichotomous data follow a binomial distribution with n=1 for the sample size. This special case of the binomial distribution is sometimes referred to as the *Bernoulli* distribution.

proportion π_{ij}. As a consequence, in the model described by equation (6.4) the lowest level variance is determined completely by the predicted value for p_{ij}, therefore it does not enter the model as a separate term.[1] The variance of p is modeled by

$$\text{VAR}(p_{ij}) = \sigma^2 (\pi_{ij} * (1-\pi_{ij})) / n_{ij} \tag{6.5}$$

In equation (6.5), σ^2 is not a variance, but a scale factor. Choosing the binomial distribution fixes σ^2 to a default value of 1.00. As explained above, this means that the binomial model is assumed to hold precisely, and the value 1.00 reported for σ^2 is not interpreted. It is possible to estimate σ^2, to model under- or overdispersion.

The model in equation (6.5) can be extended to include an explanatory variable X_{ij} (e.g., a variable describing the condition as a mail or face-to-face condition) at the condition level:

$$\text{logit}\,(p_{ij}) = \beta_{0j} + \beta_{1j}\, X_{ij} \tag{6.6}$$

The regression coefficients beta are assumed to vary across studies, and this variation is modeled by the study level variable z_j in the usual second level regression equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\, z_j + u_{0j} \tag{6.7}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\, z_j + u_{1j} \tag{6.8}$$

By substituting (6.7) and (6.8) into (6,6) we get the multilevel model:

$$\text{logit}\,(p_{ij}) = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} X_{ij}\, Z_j + u_{0j} + u_{1j} X_{ij} \tag{6.9}$$

The interpretation of the regression parameters in (6.9) is *not* in terms of the response proportions we want to analyze, but in terms of the underlying variate defined by the logit transformation $\text{logit}(p) = \ln(p/(1-p))$. The logit link function transforms the proportions, which are between 0.00

[1] This is similar to the meta-analysis model in chapter 5. In both cases the lowest level variance is known. However, in the meta-analysis model this variance must be supplied, while in the model for proportions it is automatically supplied because it is a function of the estimate for p_{ij}.

and 1.00 by definition, into values on a logistic scale that range from -∞ to +∞. The logit link is nonlinear, and in effect assumes that near the extremes of 0.00 and 1.00 it becomes more difficult to produce a change in the dependent variable (the proportion). For a quick examination of the analysis results we can simply inspect the regression parameters as calculated by the program. To understand the implications of the regression coefficients for the proportions we are modeling, we must transform the predicted logit values back to the proportion scale.

For our analysis, we analyze response rates where available, and if these are not available in the report, the completion rate is used. The null model for our example data is not the 'intercept-only' model, but a model that contains a dummy variable that codes whether the response proportion is a response rate or a completion rate. The lowest level regression model is:

$$\text{logit}\,(p_{ij}) = \beta_{0j} + \beta_{1j}\,resptype \tag{6.10}$$

where the random intercept coefficient β_{0j} is modeled by

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{6.11}$$

and the slope for the variable *resptype* by

$$\beta_{1j} = \gamma_{10} \tag{6.12}$$

which leads by substitution to:

$$\text{logit}\,(p_{ij}) = \gamma_{00} + \gamma_{10}\,resptype + u_{0j} \tag{6.13}$$

Since in meta-analysis the accurate estimation of the variance terms is an important goal, the estimation method uses the restricted maximum likelihood methods, with second order PQL approximation. For comparison, table 6.3 presents the parameter estimates for the model given by equation (6.13), for both first order MQL and second order PQL approximation.

Table 6.3 Null model for response rates

Fixed part	**MQL - 1**	**PQL – 2**
Predictor	coeff. (s.e.)	coeff. (s.e.)
intercept	0.45 (.12)	0.59 (.15)
resptype	0.68 (.06)	0.71 (.06)
Random part		
intercept_1	1.00	1.00
intercept_2	0.67 (.14)	0.93 (.20)

The PQL-2 method estimates the expected response rate as 1.30, and the MQL-1 methods as 1.13. As noted before, this refers to the underlying distribution established by the logistic link function, and *not* to the proportions themselves. To determine the expected proportion, we must use the inverse transformation for the logistic link function, given by $g(x)=\exp(x)/(1+\exp(x))$. Using this inverse function we find an expected response rate of 0.79 for PQL-2 estimation, and 0.76 for MQL-1 estimation. This is not precisely equal to the value of 0.78 that we get as the mean of the response rates, weighted by sample size. However, this is as it should be, since we are using a nonlinear link function, and the value for the intercept refers to the intercept of the underlying variate. Transforming that value back to a proportion is *not* the same as computing the intercept for the proportions themselves. Nevertheless, when the proportions are not very close to 1 or 0, the difference is usually rather small.

The value of precisely 1.00 for the variance at the lowest level looks peculiar. As I explained above, in the binomial distribution (and also in the Poisson and gamma distributions), the lowest level variance is completely determined when the mean (which in the binomial case is the proportion) is known. Therefore, in these models σ^2 has no useful interpretation; it defines the scale for the underlying normal variate. By default σ^2 is fixed at 1.00, which is equivalent to the assumption that the binomial (Poisson, gamma) distribution holds exactly. In some applications the variance of the error distribution turns out to be much larger than expected; there is *overdispersion* (cf. McCullagh & Nelder, 1989; Aitkin et al., 1989). In fact, if we allow σ^2 to be estimated, we get a value of 21.9 with standard error of 2.12. This is simply enormous. However, we already know that important explanatory variables, such as the data collection mode, are not in the model. Also, given the small samples at the condition level (on average 2.2 conditions per study), this unusual estimate for the extrabinomial variation is probably highly inaccurate. For

the moment we ignore the extrabinomial variation.

It is tempting to use the value of 1.00 as a variance estimate to calculate the intraclass correlation for the null model in table 6.3. However, 1.00 is a scale factor. The variance of a logistic distribution with scale factor 1 is $\pi^2/3 = 3.29$ (Evans, Hastings & Peacock, 1993). So the intraclass correlation for the null-model is $\rho_I = 0.93/(0.93+3.29) = 0.22$.

The next model adds the condition level dummy variables for the telephone and the mail condition, assuming fixed regression slopes. The equation at the lowest (condition) level is:

$$\text{logit}\,(p_{ij}) = \beta_{0j} + \beta_{1j}\, resptype_{ij} + \beta_{2j}\, tel_{ij} + \beta_{3j}\, mail_{ij} \qquad (6.14)$$

and at the study level:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (6.15)$$

$$\beta_{1j} = \gamma_{10} \qquad (6.16)$$

$$\beta_{2j} = \gamma_{20} \qquad (6.17)$$

$$\beta_{3j} = \gamma_{30} \qquad (6.18)$$

By substituting (6.15) to (6.18) into (6.14) we obtain:

$$\text{logit}\,(p_{ij}) = \gamma_{00} + \gamma_{10}\, resptype_{ij} + \gamma_{20}\, tel_{ij} + \gamma_{30}\, mail_{ij} + u_{0j} \qquad (6.19)$$

Until now, we have treated the two dummy variables as fixed. One could argue that it doesn't make sense to model them as random, since the dummy variables are simple dichotomies that code for our three experimental conditions. The experimental conditions are under control of the investigator, and there is no reason to expect their effect to vary from one experiment to another. But some more thought leads to the conclusion that the situation is more complicated. If we conduct a series of experiments, we would expect identical results only if the research subjects were all sampled from exactly the same population, and if the operations that define the experimental conditions were all carried out in exactly the same way. In the present case, both assumptions are questionable. In fact, some studies have sampled the general population, while others sample special populations such as college students. Similarly, although most articles give

only a very short description of the procedures that were actually used to implement the data collection methods, it is highly likely that they were not all identical. As a consequence, even if we don't know all the details about the populations sampled and the procedures used, we may expect much variation between the conditions in the way they actually were implemented. This should result in random regression coefficients in our model. Thus, we analyze a model in which the slope coefficients for the dummy variables for the telephone-and the mail condition are assumed to be random across studies.

The model with random slopes for the telephone and mail condition is given by

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + u_{3j}$$

which gives

$$\text{logit}\,(p_{ij}) = \gamma_{00} + \gamma_{10}\, resptype_{ij} + \gamma_{20}\, tel_{ij} + \gamma_{30}\, mail_{ij} + u_{0j} + u_{2j}\, tel_{ij} + u_{3j}\, mail_{ij} \qquad (6.20)$$

The results for the models specified by (6.19) and (6.20) are given in table 6.4.

Table 6.4 Models for response rates in different conditions

Fixed part	**conditions fixed**	**conditions random**
Predictor	coeff. (s.e.)	coeff. (s.e.)
intercept	0.90 (.14)	1.17 (.21)
resptype	0.53 (.06)	0.20 (.23)
telephone	-0.16 (.02)	-0.20 (.10)
mail	-0.49 (.03)	-0.58 (.16)
Random part		
$intercept_1$	1.00	1.00
$intercept_2$	0.86 (.18)	0.87 (.20)
telephone		0.26 (.08)
mail		0.59 (.20)

The intercept represents the condition in which all explanatory variables are zero. When the

telephone-dummy and the mail-dummy are both equal to zero, we have the face-to-face condition. Thus, the values for the intercept in table 6.4 estimate the expected completion rate in the face-to-face condition, 0.90 in the fixed model. The intercept plus the slope for *resptype* equals 1.43. These values on the logit scale translate to an expected completion rate of 0.71 and response rate of 0.81 for the average face-to-face survey. The negative values for the slope coefficients for the telephone and mail dummy-variables indicate that in these conditions the expected response is lower. To find out how much lower, we must use the regression equation to predict the response in the three conditions, and transform these values (which refer to the underlying variate) back to proportions. For the telephone conditions we expect an outcome of 1.26, and for the mail condition 0.94. These values on the logit scale translate to an expected response rate of 0.78 for the telephone and 0.72 for the mail condition.

The study level variances of the intercept and the conditions are obviously significant, and we may attempt to explain these using the known differences between the studies. In our example data we have two study level explanatory variables: year of publication, and the salience of the questionnaire topic. Since not all studies compare all three data collection methods, it is quite possible that study level variables also explain between condition variance. For instance, if older studies tend to have a higher response rate, and the telephone method is only included in the more recent studies (telephone interviewing is, after all, a relatively new method), the telephone condition may seem to be characterized by low response rates. In that case, however, after correcting for the year of publication, the telephone response rates should look better. We cannot inspect the condition level variance to see if the higher level variables explain condition level variability, because the condition level variance is always fixed at 1.00.

Both study level variables make a significant contribution to the regression equation, but only the year of publication interacts with the two conditions. Thus, the final model for these data is given by

$$\text{logit}\,(p_{ij}) = \beta_{0j} + \beta_{1j}\, resptype_{ij} + \beta_{2j}\, tel_{ij} + \beta_{3j}\, mail_{ij}$$

at the lowest (condition) level, and at the study level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\, year_j + \gamma_{02}\, saliency_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}\, year_j + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31}\, year_j + u_{3j}$$

which produces the combined equation

$$\text{logit}\,(p_{ij}) = \gamma_{00} + \gamma_{10}\, resptype_{ij} + \gamma_{20}\, tel_{ij} + \gamma_{30}\, mail_{ij} + \gamma_{01}\, year_j + \gamma_{02}\, saliency_j + \gamma_{21}\, tel_{ij}\, year_j + \gamma_{31}\, mail_{ij}\, year_j + u_{0j} + u_{2j}\, tel_{ij} + u_{3j}\, mail_{ij} \qquad (6.21)$$

The results for the model specified by equation (6.21) are given in table 6.5.

Table 6.5 Models for response rates in different conditions, model with random slopes and cross-level interactions

Fixed part	**no interactions**	**with interactions**
Predictor	coeff. (s.e.)	coeff. (s.e.)
intercept	0.93 (.43)	1.26 (.46)
resptype	0.33 (.21)	0.28 (.21)
telephone	-0.17 (.10)	-0.91 (.34)
mail	-0.58 (.15)	-1.45 (.46)
year	-0.02 (.01)	-0.03 (.01)
saliency	0.63 (.17)	0.64 (.17)
tel * year		0.02 (.01)
mail * year		0.03 (.01)
Random part		
intercept_1	1.00	1.00
intercept_2	0.63 (.14)	0.63 (.14)
telephone	0.26 (.07)	0.24 (.07)
mail	0.57 (.19)	0.43 (.15)

Compared to the earlier results, in the model without the interactions the regression coefficients are about the same, but in the model with the cross-level interactions some values are different. This is not informative, because the slopes for the conditions are involved in an interaction with the year of publication, and should therefore be considered together.

As I noted above, the regression coefficients have to be interpreted in terms of the underlying variate. Also, the logit transformation implies that raising the response becomes more

difficult as we approach the limit of 1.00. To show what this means, I present the predicted response for the three methods as logits (in parentheses) and proportions in the next table, both for a very salient (saliency=2) and a non-salient (saliency=0) questionnaire topic. To compute these numbers we must fill in the regression equation implied by the last column of table 6.5 and use the inverse logit transformation given earlier to transform the predicted logits back to proportions. The year is coded as 1947=0, but for the calculations in table 6.6 the year was set at 1990 (year=43). As the expected response rates in table 6.6 show, the differences between the three modes are small, while the effect of the saliency of the topic is much larger.

Table 6.6 Response rates (logits) in 1990 based in the cross-level interaction model

Topic	**face to face**	**telephone**	**mail**
not salient	0.56 (.25)	0.55 (.20)	0.52 (.09)
salient	0.71 (.89)	0.70 (.84)	0.68 (.73)

To gain a better picture understanding of the development of response rates over the years, it is useful to predict the response rates from the model and plot these predictions over the years.

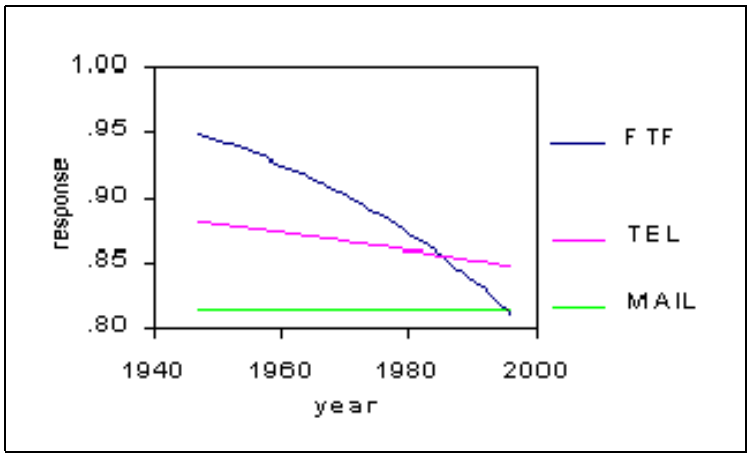

Figure 6.2 Predicted response rates over the years

Figure 6.2 presents such predictions, based on the cross-level interaction model, for the response

rates, with the saliency set to intermediate. The oldest study was published in 1947, the latest in 1992, so the values after that are an extrapolation. Figure 6.2 show that the cross level model implies that at the beginning, in 1947, the differences between the three data collection modes were large. After that, the response rates for face-to-face and telephone surveys have declined, while those for mail surveys stayed stable. As a result, the response rates for the three data collection modes have become more similar. Hox and De Leeuw (1994) analyze the same data; using a larger model with more explanatory variables and 1st order MQL estimation, they report a similar pattern.

As a final analysis, we can estimate the cross-level interaction model again, using a different link function. For the cross-level interaction model, the deviance is estimated as -136.8. For the probit link, the deviance is 138.1, and for the log-log link -138.9. Since deviance using quasi-likelihood estimation is not estimated very accurately, these values should not be used for formal comparisons. However, they are close, so the choice of a specific link function is not very important for these data. The estimates with different link functions are also close.

6.4 ANALYZING COUNTS

Nonlinear relations also result in the analysis of data that are counts. If we count the occurrences of relatively rare events, such as fires in precincts or accidents on road stretches, the data follow a Poisson distribution. In a generalized linear model, such data are usually analyzed using a Poisson error distribution, and a log link. The issues in analyzing count models are much the same as in analyzing proportions. The usual approach is to apply quasi-likelihood estimation with Taylor linearization, with some software allowing a choice between first order or second order Taylor linearization, MQL or PQL estimation, and possible modeling overdispersion. An alternative to the Poisson distribution is the negative binomial distribution, which can incorporate overdispersion by estimating an extra parameter in the variance function (cf. McCullagh & Nelder, 1989, p373). Note that the Poisson distribution assumes relatively rare events. Counts of relatively frequent events, such as the number of cigarettes smoked by a smoker, have different properties. The most important property is that counts are always positive integeres. If thre data do not have a normal distribution, the usual link function is the inverse function, coupled with a

gamma error distribution. Current multilevel software does not support this, but using macros it could be modeled in MlwiN. Alternatively, an empirical transformation can be used. An example of modeling counts is a study by Pickery and Loosveldt, who model the number of questions missed in a questionnaire. Since missing a question is a relatively rare event, the count of missed questions follows a Poisson distribution nicely. Pickery and Loosveldt use a generalized multilevel model to analyze the effect of respondent and interviewer characteristics.