

$$y_{ij}^* = \beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij} + \underbrace{\zeta_j + \epsilon_{ij}}_{\xi_{ij}}$$

The total residual variance is

$$\text{Var}(\xi_{ij}) = \psi + \pi^2/3$$

estimated as $\hat{\psi} + \pi^2/3 = 16.08 + 3.29 = 19.37$, which is much greater than the residual variance of about 3.29 for an ordinary logistic regression model. As we have already seen in figure 6.4 for probit models, the slope in the model for y_i^* has to increase when the residual standard deviation increases to produce an equivalent curve for the marginal probability that the observed response is 1. Therefore, the regression coefficients of the random-intercept model (representing subject-specific effects) must be larger than those of the ordinary logistic regression model (representing population-averaged effects) to obtain a good fit of the model-implied marginal probabilities to the corresponding sample proportions. See section 6.13 for the predicted subject-specific and population-averaged probabilities for the toenail data.

Having described subject-specific and population-averaged probabilities or expectations of y_{ij} , for given covariate values, we now consider the corresponding variances. The subject-specific or conditional variance is

$$\text{Var}(y_{ij} | \mathbf{x}_{ij}, \zeta_j) = \text{Pr}(y_{ij} = 1 | \mathbf{x}_{ij}, \zeta_j) \{1 - \text{Pr}(y_{ij} = 1 | \mathbf{x}_{ij}, \zeta_j)\}$$

and the population-averaged or marginal variance is

$$\text{Var}(y_{ij} | \mathbf{x}_{ij}) = \text{Pr}(y_{ij} = 1 | \mathbf{x}_{ij}) \{1 - \text{Pr}(y_{ij} = 1 | \mathbf{x}_{ij})\}$$

The random-intercept variance ψ does not affect the relationship between the marginal variance and the marginal mean. This is in contrast to models for counts described in chapter 9 where a random intercept (with $\psi > 0$) produces so-called overdispersion, with a larger marginal variance for a given marginal mean than the model without a random intercept ($\psi = 0$). It is important to note that, contrary to common belief, overdispersion is impossible for dichotomous responses (Skrondal and Rabe-Hesketh 2007).

6.10 Measures of dependence and heterogeneity

6.10.1 Conditional or residual intraclass correlation of the latent responses

Returning to the latent-response formulation, the dependence among the dichotomous responses for the same subject (or the between-subject heterogeneity) can be quantified by the *conditional intraclass correlation* or *residual intraclass correlation* ρ of the latent responses y_{ij}^* given the covariates

$$\rho \equiv \text{Cor}(y_{ij}^*, y_{i'j}^* | \mathbf{x}_{ij}, \mathbf{x}_{i'j}) = \text{Cor}(\xi_{ij}, \xi_{i'j}) = \frac{\psi}{\psi + \pi^2/3}$$

6.10.2 Median

Substituting the intraclass correlation for the observed

The reason for the intraclass correlation for the observed

For probit $\pi^2/3$ replaced by

6.10.2 Mediar

Larsen et al. (2007) heterogeneity. The values and intercept with is given by expected ratios across re

The median distribution function

$\text{Pr}\{\exp\}$

If the cumulative

Solving this equation

Plugging in

displacement
45.8335

When two subjects are in the same group, the comparison, the same value (treatment 18) is used for ex

6.10.2 Median odds ratio

Substituting the estimated variance $\hat{\psi} = 16.08$, we obtain an estimated conditional intraclass correlation of 0.83, which is large even for longitudinal data. The estimated intraclass correlation is also reported next to rho by xtlogit.

The reason why the degree of dependence is often expressed this way, in terms of the intraclass correlation for the latent responses y_{ij}^* , is that the intraclass correlation for the observed responses y_{ij} varies according to the values of the covariates.

For probit models, the expression for the intraclass correlations is as above with $\pi^2/3$ replaced by 1.

6.10.2 Median odds ratio

Larsen et al. (2000) and Larsen and Merlo (2005) suggest an alternative measure of heterogeneity. They consider repeatedly sampling two subjects with the same covariate values and forming the odds ratio comparing the subject with the larger random intercept with the other subject. For a given pair of subjects j and j' , this odds ratio is given by $\exp(|\zeta_j - \zeta_{j'}|)$ and heterogeneity is expressed as the median of these odds ratios across repeated samples.

The median and other percentiles $a > 1$ can be obtained from the cumulative distribution function

$$\Pr\{\exp(|\zeta_j - \zeta_{j'}|) \leq a\} = \Pr\left\{\frac{|\zeta_j - \zeta_{j'}|}{\sqrt{2\psi}} \leq \frac{\ln(a)}{\sqrt{2\psi}}\right\} = 2\Phi\left\{\frac{\ln(a)}{\sqrt{2\psi}}\right\} - 1$$

If the cumulative probability is set to $1/2$, a is the median odds ratio OR_{median}

$$2\Phi\left\{\frac{\ln(OR_{\text{median}})}{\sqrt{2\psi}}\right\} - 1 = 1/2$$

Solving this equation gives

$$OR_{\text{median}} = \exp\{\sqrt{2\psi}\Phi^{-1}(3/4)\}$$

Plugging in the parameter estimates, we obtain $\widehat{OR}_{\text{median}}$:

```
display exp(sqrt(2*16.08)*invnormal(3/4))
45.833581
```

When two subjects are chosen at random at a given time point from the same treatment group, the odds ratio comparing the subject with the larger odds to the subject with the smaller odds will exceed 45.83 half the time, which is a very large odds ratio. For comparison, the estimated odds ratio comparing two subjects at 20 months, who had the same value of the random intercept but one of whom received itraconazole (treatment=0) and the other of whom received terbinafine (treatment=1), is about $\exp(-1/\exp(-.161 + 20 \times -.137))$.

The bottom panel of figure 6.11 shows an improved approximation, known as *adaptive quadrature*, where the locations are rescaled and translated

$$e_{rj} = a_j + b_j e_r \quad (6.10)$$

to fall under the peak of the integrand, where a_j and b_j are cluster-specific constants. This transformation of the locations is accompanied by a transformation of the weights w_r that also depends on a_j and b_j . The method is called *adaptive* because the quadrature locations and weights are adapted to the data for the individual clusters.

To maximize the likelihood, we start with a set of initial or starting values of the parameters and then keep updating the parameters until the likelihood is maximized. The quantities a_j and b_j needed to evaluate the likelihood are functions of the parameters (as well as the data) and must therefore be updated or “readapted” when the parameters are updated.

There are two different implementations of adaptive quadrature in Stata that differ in the values used for a_j and b_j in (6.10). The method implemented in `gllamm` and the default method in `xtlogit` (as of Stata 10) use the posterior mean of ζ_j for a_j and the posterior standard deviation for b_j . However, obtaining the posterior mean and standard deviation requires numerical integration so adaptive quadrature sometimes does not work when there are too few quadrature points (e.g., fewer than 5). Details of the algorithm are given in Rabe-Hesketh, Skrondal, and Pickles (2002, 2005) and Skrondal and Rabe-Hesketh (2004).

The method implemented in `xtmelogit`, and available in `xtlogit` with the option `method(aghermite)`, uses the posterior mode of ζ_j for a_j and for b_j the standard deviation of the normal density whose logarithm has the same curvature as the log posterior of ζ_j at the mode. An advantage of this approach is that it does not rely on numerical integration and can therefore be implemented even with one quadrature point. With one quadrature point, this version of adaptive quadrature becomes a Laplace approximation.

6.11.2 Some speed considerations

As discussed in section 6.11.1, the likelihood involves integrals that are evaluated by numerical integration. Even with the best approach, adaptive quadrature, the likelihood itself, as well as the maximum likelihood estimates, are therefore only approximate. We can assess whether the approximation is adequate in a given situation by repeating the analysis with a larger number of quadrature points. If we get essentially the same result, the lower number of quadrature points was adequate. Such checking should always be done before estimates are taken at face value.

Also due to numerical integration, estimation can be slow, especially if there are many random effects. The time it takes to fit a model is approximately proportional to the product of the number of quadrature points for all random effects (although this seems to be more true for `gllamm` than for `xtmelogit`). For example, if there are two random effects at level 2 (a random intercept and slope) and 8 quadrature points are used for each random effect, the time will be approximately proportional to 64.

Therefore, using 4 quadrature points for each random effect will take only about one-fourth (16/64) as long as using 8. The time is also approximately proportional to the number of observations, and for programs using numerical differentiation (gllamm and xtmelogit), to the square of the number of parameters. (For xtlogit, computation time increases less dramatically when the number of parameters increases because it uses analytical derivatives.)

For large problems, it may be advisable to estimate how long estimation will take before starting work on a project. In this case, we recommend fitting a similar model with fewer random effects, fewer parameters (e.g., fewer covariates), or fewer observations, and using the above approximate proportionality factors to estimate the time that will be required for the larger problem.

For random-intercept models, by far the fastest command is xtlogit (because it uses analytical derivatives). However, xtlogit cannot fit random-coefficient models, or higher-level models introduced in chapter 10. For such models, xtmelogit or gllamm must be used. The quickest way of obtaining results here is using xtmelogit with one integration point, corresponding to the Laplace approximation. Although this method sometimes works well, it can produce severely biased estimates, especially if the clusters are small and the (true) random-intercept variance is large as for the toenail data. For these data, we obtain the following:

```
. xtmelogit outcome treatment month trt_month || patient:, intpoints(1)
Mixed-effects logistic regression
Group variable: patient
```

| | | |
|--------------------|---|--------|
| Number of obs | = | 1908 |
| Number of groups | = | 294 |
| Obs per group: min | = | 1 |
| avg | = | 6.5 |
| max | = | 7 |
| Wald chi2(3) | = | 131.96 |
| Prob > chi2 | = | 0.0000 |


```
Integration points = 1
Log likelihood = -627.80894
```

| outcome | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|-----------|-----------|-----------|-------|-------|----------------------|-----------|
| treatment | -.3070179 | .6899612 | -0.44 | 0.656 | -1.659317 | 1.045281 |
| month | -.4000919 | .047059 | -8.50 | 0.000 | -.4923258 | -.307858 |
| trt_month | -.1372598 | .0695865 | -1.97 | 0.049 | -.2736469 | -.0008728 |
| _cons | -2.523352 | .788292 | -3.20 | 0.001 | -4.068376 | -.9783276 |

| Random-effects Parameters | Estimate | Std. Err. | [95% Conf. Interval] | |
|---------------------------|----------|-----------|----------------------|---------|
| patient: Identity | | | | |
| sd(_cons) | 4.570918 | .7199338 | 3.356885 | 6.22401 |

LR test vs. logistic regression: chibar2(01) = 560.40 Prob>=chibar2 = 0.0000
Note: log-likelihood calculations are based on the Laplacian approximation.

We see that the estimated intercept and coefficient of *treatment* are different from the estimates in section 6.7.1 using adaptive quadrature with 30 quadrature points. As mentioned in the previous section, gllamm cannot be used with only 1 quadrature point, and adaptive quadrature typically requires at least 5 quadrature points.