tion calls these models single-level models because the lowest level, here occasions, is not considered a level.

2.5.4 Using gllamm

We now introduce the `gllamm` command, which will be used extensively for models with categorical or discrete responses in later chapters.

To check if `gllamm` is installed on your computer, type

```
. which gllamm
```

If the following message appears,

```
command gllamm not found as either built-in or ado-file
```

install `gllamm` (assuming that you have a net-aware Stata) by using the `ssc` command:

```
. ssc install gllamm
```

Occasionally, you should update `gllamm` using `ssc` with the `replace` option:

```
. ssc install gllamm, replace
```

The `gllamm` command for fitting the variance-components model (2.2) resembles the `xtreg` command. We add two options to ensure accurate estimates: the `nip(12)` and `adapt` options (see sec. 6.11.1 for more details).

```
. gllamm wm, i(id) nip(12) adapt
number of level 1 units = 34
number of level 2 units = 17

Condition Number = 152.64774

gllamm model

log likelihood = -184.57839
```

wm	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	453.9116	26.18394	17.34	0.000	402.592	505.2312

```
Variance at level 1
------------------------------------------------------------------------------
  396.70879 (136.11609)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id)

   var(1): 11456.828 (3997.7689)
------------------------------------------------------------------------------
```

The output from `gllamm` first shows the number of units at each level, here 34 units at level 1 (the total number of measurements) and 17 units at level 2 (the subjects). If the `Condition Number` is very large, the model may not be well identified, but here it is not alarming.

Next the maximized log likelihood is given as -184.58 followed by a regression table giving the estimated fixed regression coefficient $\widehat{\beta}$ next to `_cons`.

Estimates and standard errors for the random part of the model are given under the headings "`Variance at level 1`" for the variance θ of the level-1 residuals ϵ_{ij} and "`Variances and covariances of random effects`" and "`***level 2 (id)`" for the variance ψ of the random intercept ζ_j. Variance estimates from `gllamm` are presented under "`gllamm`" in table 2.2.

`xtreg` and `xtmixed` display the estimated standard deviations instead of variances. We can convert these standard deviations to variances $\widehat{\theta} = 19.91083^2 = 396.44115$ and $\widehat{\psi} = 107.0464^2 = 11,458.932$, which differ slightly from the estimates using `gllamm`. The reason for the discrepancy is that `gllamm` uses numerical integration, whereas `xtreg` and `xtmixed` exploit the closed form of the likelihood for random-effects models with normally distributed continuous responses. The accuracy of the `gllamm` estimates can be improved by increasing the number of integration points (see section 6.11.1) using the `nip()` option.

Table 2.2: Maximum likelihood estimates for Mini Wright peak flow meter

	Est	(SE)
Fixed part		
β	453.91	(26.18)
Random part		
`xtreg`, `xtmixed`		
$\sqrt{\psi}$	107.05	
$\sqrt{\theta}$	19.91	
`gllamm`		
ψ	11,456.83	
θ	396.71	
Log likelihood	-184.58	

Slight discrepancies between estimates due to numerical integration in `gllamm`

Before doing further analyses, we save the `gllamm` estimates using `estimates store`:

```
. estimates store RI
```

We can `restore` these estimates later without having to refit the model. This will be useful in section 2.9.3, where we use `gllamm`'s prediction command `gllapred`. Storing estimates means that they remain available during a Stata session; if we require the estimates again in a later Stata session, we can save them in a file using `estimates save` *filename* (a command introduced in Stata release 10).

2.6 Hypothesis tests and confidence intervals

2.6.1 Hypothesis test and confidence interval for the population mean

In the regression tables produced by `xtreg`, `xtmixed`, and `gllamm`, z statistics are reported for β instead of the t statistics given by the `regress` command.

As the t statistic in ordinary linear regression, the z statistic is given by

$$z = \frac{\widehat{\beta}}{\widehat{\mathrm{SE}}(\widehat{\beta})}$$

(where the standard error takes a different form than in linear regression as discussed in section 1.5).

The reason this statistic is called z instead of t is that a standard normal sampling distribution is assumed under the null hypothesis that $\beta = 0$ instead of a t distribution. The t distribution is a finite sample distribution whose shape depends on the degrees of freedom. For the variance-components model, the finite sample distribution does not have a simple form, so Stata's commands use the asymptotic (large-sample) sampling distribution. (Some other software packages approximate the finite-sample distribution by a t distribution where the degrees of freedom are some function of the data.) The null hypothesis that the population mean β is zero is not of interest in the peak-expiratory-flow example.

An asymptotic 95% confidence interval is given by

$$\widehat{\beta} \pm z_{.975}\widehat{\mathrm{SE}}(\widehat{\beta})$$

where $z_{.975}$ is the 97.5th percentile of the standard normal distribution, i.e., $z_{.975} = 1.96$. This kind of confidence interval based on assuming a normal sampling distribution is often called a Wald confidence interval. In the Mini Wright application, the 95% Wald confidence interval for the population mean β is from 402.59 to 505.24, as shown for instance in the output from `xtreg` on page 64.

As for linear regression there are two versions of estimated standard errors: a model-based version and a "robust" version based on the so-called sandwich estimator. The latter can be obtained for the GLS estimator in `xtreg` using the `vce(robust)` option. Robust standard errors for maximum likelihood estimates are produced by `gllamm` when the `robust` option is specified.

2.6.2 Hypothesis test and confidence interval for the between-cluster variance

We now consider testing hypotheses regarding the between-cluster variance ψ. In particular, we are often interested in the hypotheses

$$H_0\text{: } \psi = 0 \quad \text{against} \quad H_a\text{: } \psi > 0$$

This null hypothesis is equivalent to the hypothesis that $\zeta_j = 0$ or that there is no random intercept in the model. If the null hypothesis is true, we can use ordinary regression instead of a variance-components model.

Likelihood-ratio tests are typically used with the test statistic

$$L = 2(l_1 - l_0)$$

where l_1 is the maximized log likelihood for the variance-components model (which includes ζ_j) and l_0 is the maximized log likelihood for a model without ζ_j. Importantly, the distribution of L under H_0 is not χ^2 with 1 degree of freedom as usual. This is because the null hypothesis is on the boundary of the parameter space since $\psi \geq 0$, which renders standard statistical test theory invalid.

If the variance-components model is used for replicated datasets generated under the null hypothesis, we would expect positive correlations among the responses about half of the time and negative correlations the other half of the time. Thus ψ would be estimated as positive half of the time and as zero (since negative correlations cannot be produced by nonnegative ψ in the variance-components model) the other half of the time. The correct sampling distribution under the null hypothesis hence takes a simple form, being a 50:50 mixture of a spike at 0 and a χ^2 with 1 df. The correct p-value can be obtained by simply dividing the "naive" p-value, based on the χ^2 with 1 df, by 2.

This p-value is given at the bottom of the `xtreg` and `xtmixed` output, where the correct sampling distribution is referred to as `chibar2(01)` (click on `chibar2(01)`, which is shown in blue in the Stata output window to find an explanation). We can also perform the likelihood-ratio test ourselves by fitting the variance-components model, storing the estimates, then fitting the model without the random intercept, and comparing the models using the `lrtest` command:

```
. quietly xtmixed wm || id:, mle
. estimates store m1
. quietly xtmixed wm, mle
. lrtest m1 .
Likelihood-ratio test                          LR chibar2(01)  =     46.27
(Assumption: . nested in m1)                   Prob > chibar2  =    0.0000
```

Here the `quietly` prefix command is used to suppress output from `xtmixed`. In the `lrtest` command, `m1` refers to the estimates stored under that name and "." refers to the current (or last) estimates. For the peak-expiratory-flow application, we see that the null hypothesis $\psi = 0$ has a very small p-value and the null hypothesis is rejected at standard significance levels.

An alternative tests for ψ is the *score test* (sometimes called the Lagrange multiplier test) which is based on a quadratic approximation of the likelihood at $\psi = 0$, and can also be used to construct confidence intervals. Score tests and confidence intervals for variance components are not provided by official Stata at the time of writing but Bottai and Orsini (2004) have provided the postestimation command `xtvc` for `xtreg` with the `mle` option. We first install the command by typing

```
. ssc install xtvc
```

and then run it by typing

```
. quietly xtreg wm, mle
. xtvc
```

wm	ML Estimate	[95% Conf.	Interval]
/sigma_u	107.0464	84.36725	200.3493

Score test of sigma_u=0: chi2(1)= 30242.81 Prob>=chi2 = 0.000

The null hypothesis is clearly rejected, and the estimated 95% confidence interval for the standard deviation $\sqrt{\psi}$ is from 84.37 to 200.35. This confidence interval is different from that produced by `xtreg`. The latter confidence interval is obtained by exponentiating the limits of the Wald confidence interval for the log standard deviation.

We can also base the test for unexplained between-cluster heterogeneity on the fixed-effects model, testing the null hypothesis that the fixed effects α_j of the clusters are all zero. This test, described by for instance Wooldridge (2002), can be obtained using `xtreg` with the `fe` option:

```
. xtreg wm, i(id) fe
Fixed-effects (within) regression               Number of obs      =        34
Group variable: id                              Number of groups   =        17

R-sq:  within  = 0.0000                         Obs per group: min =         2
       between = 0.0000                                        avg =       2.0
       overall = 0.0000                                        max =         2

corr(u_i, Xb)  = 0.0000                         F(0,17)            =      0.00
                                                Prob > F           =         .

------------------------------------------------------------------------------
          wm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   453.9118   3.414679   132.93   0.000     446.7074    461.1161
-------------+----------------------------------------------------------------
     sigma_u |  111.29118
     sigma_e |  19.910831
         rho |  .96898482   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(16, 17) =    62.48                 Prob > F = 0.0000
```

We see from the bottom of the output that the null hypothesis is clearly rejected using the F test. (Use of the `xtreg` command with the `fe` option, which stands for "fixed effects", is discussed in more detail in section 3.7.2.)

In this subsection, we have used all the terminology usually associated with estima
ing model parameters. However, it is important to remember that ζ_j is not a parame
in the original model. It is only for the purpose of assigning values to ζ_j that we refor
mulate the problem by treating the original parameters as known constants and the ζ_j
as unknown parameters.

2.9.2 Empirical Bayes prediction

Having obtained estimates $\widehat{\beta}$, $\widehat{\psi}$, and $\widehat{\theta}$ of the model parameters and treating them as the
true parameter values, we can predict values of the random intercepts ζ_j for individual
clusters (subjects in the application). Here we continue to treat ζ_j as a random variable,
not as a fixed parameter as in maximum likelihood estimation.

Maximum likelihood estimation of ζ_j uses the responses y_{ij} for subject j as the only
information about ζ_j by maximizing the likelihood of observing these particular values

$$\text{Likelihood}(y_{1j}, y_{2j}|\zeta_j)$$

In contrast, empirical Bayes prediction also uses the *prior distribution* of ζ_j, summariz-
ing our knowledge about ζ_j before seeing the data for subject j

$$\text{Prior}(\zeta_j)$$

This prior distribution is just the normal distribution specified for the random intercept
with zero mean and estimated variance $\widehat{\psi}$. It represents what we know about ζ_j before
we have seen the responses y_{1j} and y_{2j} for subject j. For instance, the most likely value
of ζ_j is zero. (Obviously, we have already used all responses to obtain the estimate $\widehat{\psi}$,
but we now pretend that ψ is known and not estimated.)

Once we have observed the responses, we can combine the prior distribution with
the likelihood to obtain the *posterior distribution* of ζ_j given the observed responses y_{1j}
and y_{2j}

$$\text{Posterior}(\zeta_j|y_{1j}, y_{2j}) \propto \text{Prior}(\zeta_j) \times \text{Likelihood}(y_{1j}, y_{2j}|\zeta_j)$$

where $\propto$ means "proportional to". The posterior of ζ_j represents our updated knowledge
regarding ζ_j after seeing the data y_{1j} and y_{2j} for subject j.

The empirical Bayes prediction is just the mean of the posterior distribution with
parameter estimates ($\widehat{\beta}$, $\widehat{\psi}$, and $\widehat{\theta}$) plugged in. In a linear model with normal error
terms, the posterior is normal and the mean is thus equal to the mode.

Figure 2.9 shows the prior, likelihood, and posterior for a hypothetical example of a
subject with $n_j = 2$ responses. In both panels, the estimated total residuals $\widehat{\xi}_{ij}$ are 3 and
5, and the estimated total variance is $\widehat{\psi}+\widehat{\theta} = 5$. In the top panel, 80% of this variance is
due to within-subject variability, whereas in the bottom panel, 80% is due to between-
subject variability. In both cases, the likelihood (dotted curve) has its maximum at
$\zeta_j = 4$, i.e., the mode is 4 (see vertical dotted lines). The maximum likelihood estimate
therefore is $\widehat{\zeta}_j^{\text{ML}} = 4$. In contrast, the mode (and mean) of the posterior depends on the

Figure 2.9: Pri
$n_j = 2$ respons
(and means) of

relative sizes of the variance components and is 1.33 in the top panel and 3.56 in the bottom panel (see vertical dashed lines). The mean of the posterior lies between the mean of the prior (zero, vertical solid lines) and the mode of the likelihood.

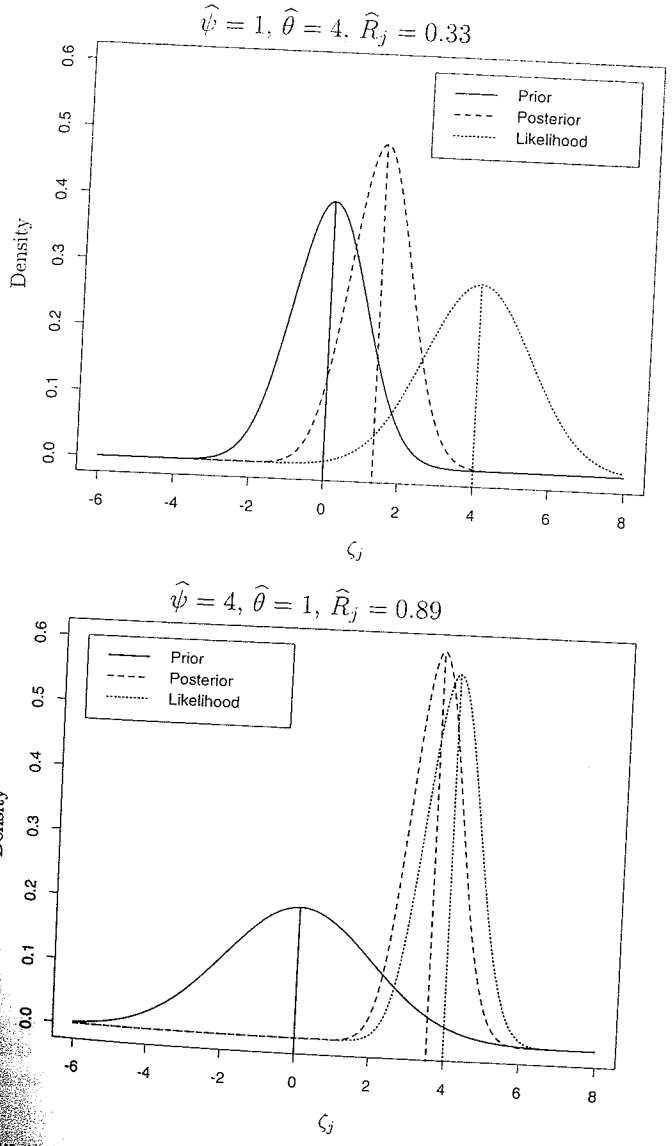

Figure 2.9: Prior, likelihood (normalized), and posterior for a hypothetical subject with $n_j = 2$ responses with total residuals $\widehat{\xi}_{1j} = 3$, $\widehat{\xi}_{2j} = 5$ (the vertical lines represent modes (and means) of the distributions)

In fact, there is a simple formula relating the empirical Bayes prediction $\widetilde{\zeta}_j^{\mathrm{EB}}$ to the maximum likelihood estimator $\widehat{\zeta}_j^{\mathrm{ML}}$ in linear random-intercept models

$$\widetilde{\zeta}_j^{\mathrm{EB}} = \widehat{R}_j \widehat{\zeta}_j^{\mathrm{ML}}, \qquad \text{where} \qquad \widehat{R}_j = \frac{\widehat{\psi}}{\widehat{\psi} + \widehat{\theta}/n_j}$$

Here $\widehat{R}_j$ is similar to the estimated intraclass correlation, except that we divide the estimated level-1 variance $\widehat{\theta}$ by the number of responses n_j. $\widehat{R}_j$ therefore represents the *reliability* of the ML estimator of ζ_j, the mean of n_j residuals. $\widehat{R}_j$ is also known as the *shrinkage factor* because $0 < \widehat{R}_j < 1$ so that the empirical Bayes prediction is shrunken toward 0 (the mean of the prior). There will be more shrinkage (i.e., greater influence of the prior) if we have

- a small random-intercept variance $\widehat{\psi}$ (an informative prior)
- a large level-1 residual variance $\widehat{\theta}$ (uninformative data)
- a small cluster size n_j (uninformative cluster)

A nice feature of empirical Bayes prediction is that the prediction error, defined as the difference $\widetilde{\zeta}_j - \zeta_j$ between the prediction and the truth, has zero mean over repeated samples of ζ_j and ϵ_{ij} (or repeated samples of clusters and units from clusters), as well as having the smallest possible variance (treating the model parameters as fixed and known). In linear mixed models, the empirical Bayes predictor is therefore also known as the best linear unbiased predictor (BLUP). However, empirical Bayes predictions are conditionally biased in the sense that their mean over repeated samples of ϵ_{ij} for a given ζ_j (or repeated samples of units from the same cluster) will be too close to zero (due to shrinkage). In contrast, the maximum likelihood estimator is conditionally unbiased but has a greater prediction-error variance. In most applications, shrinkage is desirable because it only affects clusters that provide little information and effectively downplays their influence, "borrowing strength" from other clusters.

For the Mini Wright meter measurements, the shrinkage factor can be calculated from the estimates of $\sqrt{\widehat{\psi}}$ and $\sqrt{\widehat{\theta}}$ in table 2.2:

```
. display 107.05^2/(107.05^2+(19.91^2)/2)
.98299832
```

We can now obtain empirical Bayes predictions in two ways: either by multiplying the maximum likelihood estimates obtained in section 2.9.1 by the shrinkage factor

```
. generate eb1 = .98299832*ml
```

or by using the `predict` command with the `reffects` option after estimation using `xtmixed`:

```
. quietly xtmixed wm || id:, mle
. predict eb2, reffects
```

```
. sort id
. format eb1 eb2 %8.2f
. list id eb1 eb2 if occasion==1, clean noobs
```

id	eb1	eb2
1	63.49	63.49
2	-30.88	-30.88
3	59.07	59.07
4	-17.61	-17.61
5	45.30	45.30
6	155.89	155.89
7	-41.20	-41.20
8	-67.74	-67.74
9	192.75	192.75
10	-15.15	-15.15
11	-27.44	-27.44
12	158.84	158.84
13	-206.83	-206.83
14	17.78	17.78
15	-187.17	-187.17
16	-92.31	-92.31
17	-6.79	-6.79

Both methods give identical results to two decimal places.

There are no simple formulas for empirical Bayes predictions for the models for noncontinuous responses considered in later chapters. For such models, we would have to use the `gllapred` command with the `u` option (after restoring the `gllamm` estimates):

```
estimates restore RI
gllapred eb, u
```

The predictions would then be placed in the variable `ebm1` and would be close to the values in `eb1` and `eb2`.

2.9.3 ❖ Empirical Bayes variances

There are several different kinds of variances for empirical Bayes predictions that can be used to express uncertainty regarding the predictions. However, it is important to remember that these variances do not take into account the uncertainty in the parameter estimates since they are treated as known in empirical Bayes prediction.

Posterior and prediction-error variances

The posterior variance is the variance of the random intercept ζ_j given the observed responses, the variance of the posterior distribution shown as a dashed curve in figure 2.6. For the model considered here, the posterior variance is

$$\mathrm{Var}(\zeta_j|y_{1j},y_{2j}) = \frac{\widehat{\theta}/n_j}{\widehat{\psi}+\widehat{\theta}/n_j}\,\widehat{\psi} = (1-\widehat{R}_j)\,\widehat{\psi}$$

As expected, the posterior variance is smaller than the estimated prior variance $\widehat{\psi}$ due to the information gained regarding the random intercept by knowing the responses y_{1j} and y_{2j}.

In the linear random-intercept model, the posterior variance equals the variance of the prediction errors

$$\mathrm{Var}(\widetilde{\zeta}_j^{\mathrm{EB}} - \zeta_j)$$

over repeated samples of ζ_j and ϵ_{ij} (or repeated samples of clusters j and units i), but with parameters held constant and equal to the estimates. The square root of this variance is sometimes referred to as the *comparative standard error* because it can be used for inferences regarding differences between subjects' true random intercepts.

At the time this text was printed, there was no `xtmixed` postestimation command for estimating posterior variances. We therefore restore the `gllamm` estimates stored under `RI` and use the `gllapred` command with the `u` option:

```
. estimates restore RI
(results RI are active now)
. gllapred eb, u
(means and standard deviations will be stored in ebm1 ebs1)
```

As indicated by the output, the required posterior standard deviations are stored in the variable `ebs1`. These standard deviations are identical for all clusters because the clusters have the same size $n_j = 2$ so that R_j is the same for all clusters. We therefore display only the value for the first observation:

```
. display ebs1[1]
13.963476
```

Marginal sampling variances

For linear random-intercept models, the sampling distribution of the empirical Bayes predictions (over repeated samples of ζ_j and ϵ_{ij}, or of clusters and units from clusters) is normal with mean 0 and variance

$$\mathrm{Var}(\widetilde{\zeta}_j^{\mathrm{EB}}) = \frac{\widehat{\psi}}{\widehat{\psi} + \widehat{\theta}/n_j}\,\widehat{\psi} = \widehat{R}_j\,\widehat{\psi}$$

This variance is useful for deciding if the empirical Bayes prediction for a given subject is aberrant. For instance, 95% of predictions should be no larger in absolute value than about two sampling standard deviations. Thus the sampling standard deviation is often called the *diagnostic standard error*. We can compute this standard error from the estimated prior variance $\widehat{\psi}$ and posterior variance $\mathrm{Var}(\zeta_j|y_{1j}, y_{2j})$ obtained earlier from `gllapred`, using the relationship

$$\mathrm{Var}(\widetilde{\zeta}_j^{\mathrm{EB}}) = \widehat{R}_j\widehat{\psi} = \widehat{\psi} - (1 - \widehat{R}_j)\,\widehat{\psi} = \widehat{\psi} - \mathrm{Var}(\zeta_j|y_{1j}, y_{2j})$$

The corresponding standard deviation can be obtained using

```
. display sqrt(.98299832*107.05^2)
106.13608
```

Conditional sampling variances

The variance of the empirical Bayes prediction over repeated samples of ϵ_{ij} for a given ζ_j (or repeated samples of units from the same cluster) is given by

$$\text{Var}(\widetilde{\zeta}_j^{\text{EB}}|\zeta_j) \;=\; \widehat{R}_j(1-\widehat{R}_j)\,\widehat{\psi}$$

We can compute the conditional sampling standard deviation using

```
. display sqrt(.98299832*(1-.98299832)*107.05^2)
13.839136
```

Comparing the empirical Bayes variances

If $\widehat{R}_j > 0.5$, as is usually the case in practice, we obtain the following relations among the empirical Bayes variances:

$$\text{Var}(\widetilde{\zeta}_j^{\text{EB}}|\zeta_j) < \text{Var}(\zeta_j|y_{1j},y_{2j}) = \text{Var}(\widetilde{\zeta}_j^{\text{EB}}-\zeta_j) < \text{Var}(\widetilde{\zeta}_j^{\text{EB}})$$

As we would expect, these relations are satisfied for the Mini Wright data since $\widehat{R}_j = 0.98$.

2.10 Summary and further reading

In this chapter, we introduced the idea of decomposing the total variance between responses into variance components, specifically the between-cluster variance ψ and the within-cluster variance θ. This was accomplished by specifying a model that includes corresponding error components, a level-2 random intercept ζ_j for clusters, and a level-1 residual ϵ_{ij} for units within clusters. The random intercept induces correlations among responses for units in the same cluster, known as the *intraclass correlation*.

These concepts underlie all multilevel or hierarchical modeling. By considering the simplest case of a multilevel model, we have provided some insight into estimation of unknown model parameters and prediction of random effects. We have also shown how to conduct hypothesis testing and construct confidence intervals for variance-components models. Although the expressions for estimators and predictors become more complex for the models discussed in later chapters, the basic ideas remain the same.

For further reading about variance-components models, we recommend Snijders and Bosker (1999), as well as many of the books referred to in later chapters. Streiner and Norman (2003), Shavelson and Webb (1991), and Dunn (2004) are excellent books on measurement models.