Rabe-Hesketh & Skrondal (2008)

2 Variance-components models

2.1 Introduction

Units of observation often fall into groups or clusters. For example, individuals could be nested in families, hospitals, schools, neighborhoods, or firms. Longitudinal data also consist of clusters of observations made at different occasions for the same subject. For two examples of clustered data, the nesting structure is depicted in figure 2.1.

In clustered data, it is usually important to allow for dependence or correlations among the responses observed for units belonging to the same cluster. For example, the adult heights of siblings are likely to be correlated because siblings are genetically related to each other and have usually been raised within the same family. Variance-components models are designed to model and estimate such within-cluster correlations.

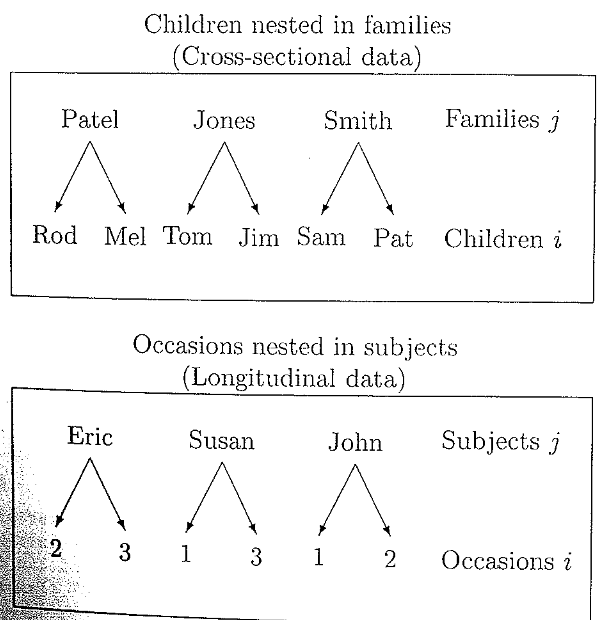

Figure 2.1: Examples of clustered data

In this chapter, we consider the simple situation of clustered data without explanatory variables. This situation is important in its own right and is also useful for introducing and motivating the notions of random effects and variance components. We also describe basic principles of estimation and prediction in this simple setting. However, this means that some parts of the chapter may be a bit demanding, and you may like to skip sections 2.7 and 2.9 on first reading.

We also introduce the Stata commands `xtreg`, `xtmixed`, and `gllamm`, which will be used in subsequent chapters.

2.2 How reliable are peak-expiratory-flow measurements?

The data come from a reliability study conducted by Martin Bland using 17 of his family and colleagues as subjects. The purpose was to illustrate a way of assessing the quality of two instruments for measuring people's peak-expiratory-flow rate (PEFR). The PEFR, which is roughly speaking how strongly subjects can breathe out, is a central clinical measure in respiratory medicine.

The subjects had their PEFR measured twice (in liters per minute) using the standard Wright peak flow meter and twice using the new Mini Wright peak flow meter. The methods were used in random order to avoid confounding practice (prior experience) effects with method effects. If the new method agrees sufficiently well with the old, the old may be replaced with the more convenient Mini meter. Somewhat remarkably, the paper reporting this study (Bland and Altman 1986) is the most cited paper in the *Lancet*, one of the most prestigious medical journals.

In this chapter, we analyze the two sets of measurements using the Mini Wright peak flow meter. Analyses comparing the standard Wright and Mini Wright peak flow meters are discussed in chapter 10.

The data are presented in table 2.1 and are in `pefr.dta` in the same form as in the table, with the following variable names:

- `id`: subject identifier
- `wp1`: Wright peak flow meter, occasion 1
- `wp2`: Wright peak flow meter, occasion 2
- `wm1`: Mini Wright flow meter, occasion 1
- `wm2`: Mini Wright flow meter, occasion 2

Table 2.1: Peak-expiratory-flow rate measured on two occasions using both the Wright and the Mini Wright peak flow meters

	Wright peak flow meter		Mini Wright peak flow meter	
Subject	First	Second	First	Second
1	494	490	512	525
2	395	397	430	415
3	516	512	520	508
4	434	401	428	444
5	476	470	500	500
6	557	611	600	625
7	413	415	364	460
8	442	431	380	390
9	650	638	658	642
10	433	429	445	432
11	417	420	432	420
12	656	633	626	605
13	267	275	260	227
14	478	492	477	467
15	178	165	259	268
16	423	372	350	370
17	427	421	451	443

First, we load the data into Stata using the command

```
. use http://www.stata-press.com/data/mlmus2/pefr
```

The first and second recordings on the Mini Wright peak flow meter can be plotted against the subject identifier with a horizontal line representing the overall mean by using

```
. generate mean_wm = (wm1+wm2)/2
. summarize mean_wm
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     mean_wm |        17    453.9118    111.2912      243.5        650
. twoway (scatter wm1 id, msymbol(circle))
>        (scatter wm2 id, msymbol(circle_hollow)),
> xtitle(Subject id)  xlabel(1/17) ytitle(Mini Wright measurements)
> legend( order(1 "Occasion 1" 2 "Occasion 2")) yline(453.9118)
```

The resulting graph is shown in figure 2.2.

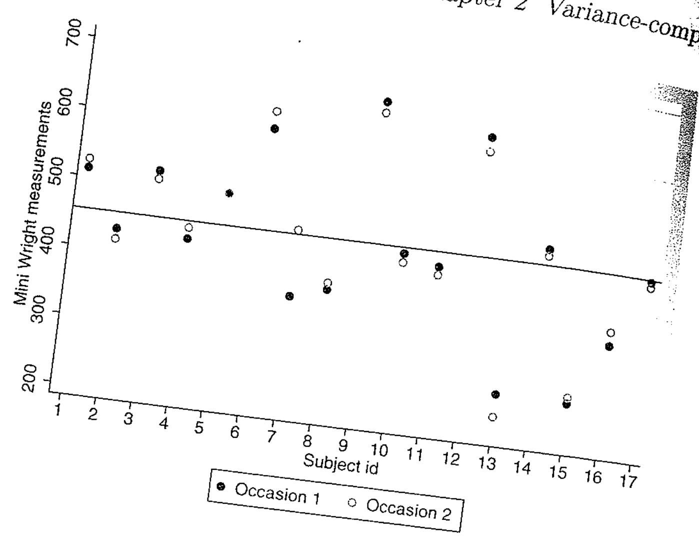

Figure 2.2: First and second recordings of peak expiratory flow using Mini Wright meter versus subject number (the horizontal line represents the overall mean)

It is clear from the figure that repeated measurements on the same subject tend to be closer to each other than to the measurements on a different subject. Indeed, if this were not the case, the Mini Wright peak flow meter would be useless as a tool for discriminating between the individuals in this particular sample. Because there are large differences between subjects (for example, compare subjects 9 and 15) and only small differences within subjects, the responses for occasions 1 and 2 on the same subject tend to lie on the same side of the overall mean, shown as a horizontal line in the figure, and are therefore correlated (see also section 2.3.3). We can also see that there is within-subject dependence by considering prediction of a subject's response at occasion 2 if all we know are the subjects' responses at occasion 1. If the response for a given subject at occasion 2 were independent of his or her response at occasion 1, a good prediction would be the mean response at occasion 1 across all subjects. However, it is clear that a much better prediction here is the subject's own response at occasion 1 because the responses are dependent within subjects. We see that the within-subject dependence is due to between-subject heterogeneity. If all subjects were more or less alike (for example, pick subjects 2, 4, 10, 11, 14, and 17), there would be much less within-subject dependence.

2.3 The variance-components model

2.3.1 Model specification and path diagram

It may be tempting to model the responses y_{ij} of subject j on occasion i using a standard regression model without covariates

$$y_{ij} = \beta + \xi_{ij} \tag{2.1}$$

where ξ_{ij} are residuals or error terms that are independent over both subjects and occasions (the Greek letter ξ is pronounced xi). However, this specification is unreasonable since, as we have seen in figure 2.2, measurements are expected to be more similar within than between subjects or in other words be dependent within subjects.

We can model this dependence by splitting the residual ξ_{ij} into two components: a component ζ_j (ζ is pronounced zeta), which is specific to each subject j and constant across occasions i, and a component ϵ_{ij}, which is specific to each subject j at each occasion i

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij} \tag{2.2}$$

as shown for a single subject in figure 2.3. Here ζ_j is the random deviation of subject j's mean measurement (over a hypothetical population of measurement occasions) from the overall mean β. The component ζ_j, often called a random effect of subject or a *random intercept*, has zero population mean and variance ψ (pronounced psi) over subjects and is assumed to be independent over subjects. The component ϵ_{ij}, often called the residual or within-subject residual, is the random deviation of y_{ij} from subject j's mean. This residual has zero population mean and variance θ (pronounced theta) over occasions and subjects, and is assumed to be independent over both subjects and occasions. We can interpret ψ is the between-subject variance and θ as the within-subject variance.

In classical psychometric test theory, (2.2) represents a measurement model where $\beta + \zeta_j$ is the *true score* for subject j, defined as the long-term mean measurement (e.g., Streiner and Norman 2003).

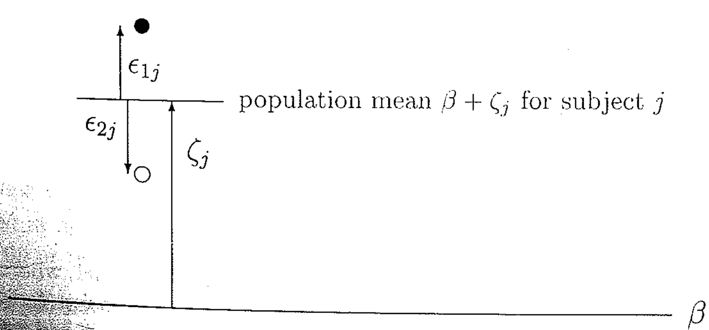

Figure 2.3: Illustration of random-intercept model for a subject j

We assume that the random intercepts ζ_j are normally distributed,

$$\zeta_j \sim N(0, \psi)$$

and that the ϵ_{ij} are normally distributed

$$\epsilon_{ij} \sim N(0, \theta)$$

These distributional assumptions are illustrated for a subject j in figure 2.4. The random intercept ζ_j has a normal distribution with mean zero and variance ψ (see the top distribution in the figure). Drawing a realization from this distribution for subject j determines the mean $\beta+\zeta_j$ of the distributions from which responses y_{ij} for this subject are subsequently drawn. At a given measurement occasion i, a response y_{ij} is therefore sampled from a normal distribution with mean $\beta+\zeta_j$ and variance θ (see the bottom distribution in the figure). Equivalently, a residual (or measurement error) ϵ_{ij} is drawn from a normal distribution with mean zero and variance θ.

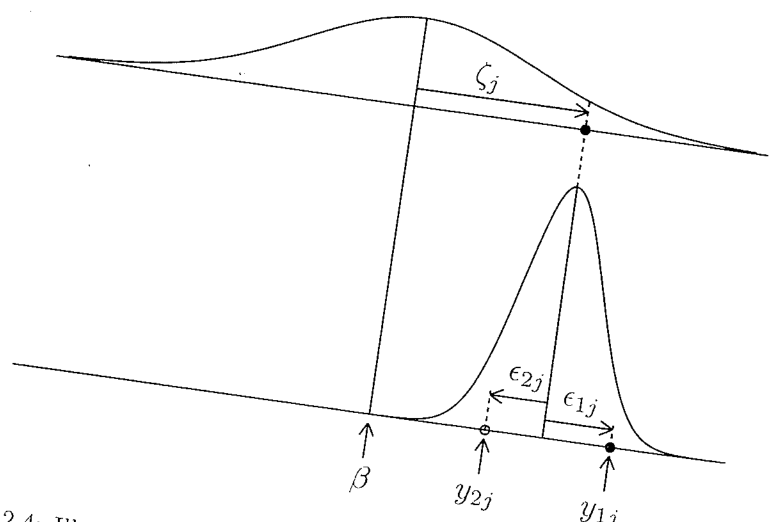

Figure 2.4: Illustration of distributions of error components for a subject j

This hierarchical model is a simple example of a two-level model, where occasions are level-1 units and subjects are level-2 clusters. The random intercept ζ_j is then referred to as the level-2 residual with level-2 (between-subject) variance ψ and ϵ_{ij} as the level-1 residual with level-1 (between-occasion, within-subject) variance θ.

The model is often motivated in terms of two-stage survey sampling where the randomness (between hypothetical repeated samples) is due to two-stage random sampling of clusters with fixed ζ_j from the population of clusters and then units with fixed ϵ_{ij} from the population of units within the clusters. This way of thinking is not useful when the term "population" is interpreted too literally as a *finite* population, which would leave no randomness at level 2 when the clusters are for instance all U.S. states and no randomness at level 1 when the clusters are people's heads and the units are both eyes on each head. In this book we assume that readers want to make inferences regarding data-generating mechanisms or wider populations.

We can display the random part of the model (every term except β) using a path diagram as shown in figure 2.5. Here the rectangles represent the observed responses y_{1j} and y_{2j} for each subject j, where the j subscript is implied by the label "subject j" inside the frame surrounding the diagram. The long arrows from ζ_j to the responses represent regressions with slopes equal to 1, and the short arrows pointing at the responses from below represent the additive level-1 residuals ϵ_{1j} and ϵ_{2j}.

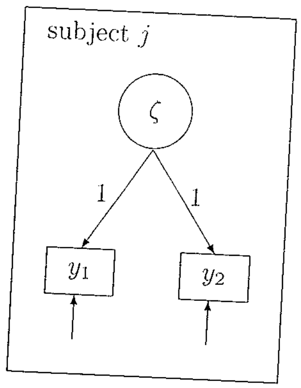

Figure 2.5: Path diagram of random part of random-intercept model

The path diagram makes it clear that the two responses on each subject are *conditionally independent* given ζ_j since there is no arrow directly connecting them. It follows that the responses are conditionally uncorrelated when controlling for ζ_j (i.e., holding it constant across subjects)

$$\text{Cor}(y_{ij}, y_{i'j}|\zeta_j) = 0$$

This can also be seen by imagining that the data in figure 2.2 were generated by the model depicted in figure 2.3, where the dependence is solely due to the measurements being shifted up or down by the shared random intercept ζ_j for each cluster j.

The (marginal) within-subject correlation is induced by ζ_j because this is shared by all responses for the same subject. As we will see in later chapters, path diagrams are useful for conveying the structure of complex models involving several random effects.

2.3.2 Error components, variance components, and reliability

Each response differs from the overall mean β by a total residual or error ξ_{ij}, the sum of two error terms or *error components* ζ_j and ϵ_{ij}

$$\xi_{ij} \;\equiv\; \zeta_j + \epsilon_{ij}$$

The random intercept ζ_j is shared between measurement occasions on the same subject j, whereas ϵ_{ij} is unique for each occasion i (and subject).

Since the error components are independent, the total variance is the sum of the *variance components*

$$\mathrm{Var}(y_{ij}) \;=\; \mathrm{Var}(\beta + \zeta_j + \epsilon_{ij}) = \underbrace{\mathrm{Var}(\beta)}_{0} + \mathrm{Var}(\underbrace{\zeta_j + \epsilon_{ij}}_{\xi_{ij}}) = \psi + \theta$$

the between-subject and within-subject variances. The proportion of the total variance that is between subjects, or due to subjects, is

$$\rho \;=\; \frac{\mathrm{Var}(\zeta_j)}{\mathrm{Var}(y_{ij})} = \frac{\psi}{\psi + \theta} \tag{2.3}$$

In the measurement context, ψ is the variance between subjects' true scores $\beta + \zeta_j$, θ is the *measurement error variance* (the squared *standard error of measurement*), and ρ is a *reliability*, here a test–retest reliability (see, for example, Streiner and Norman 2003). The reliability can be thought of as the proportion of the total variance that is "explained" by subjects, analogously to the coefficient of determination R^2 in linear regression discussed in section 1.5.

2.3.3 Intraclass correlation

Consider first the marginal (not conditional on ζ_j) covariance between the measurements on two occasions i and i' for the same subject, defined as

$$\mathrm{Cov}(y_{ij}, y_{i'j}) \;=\; E[\{y_{ij} - E(y_{ij})\}\{y_{i'j} - E(y_{i'j})\}]$$

The corresponding marginal correlation is the above covariance divided by the product of the standard deviations

$$\mathrm{Cor}(y_{ij}, y_{i'j}) = \frac{\mathrm{Cov}(y_{ij}, y_{i'j})}{\sqrt{\mathrm{Var}(y_{ij})}\sqrt{\mathrm{Var}(y_{i'j})}} \tag{2.4}$$

It follows from the variance-components model that the population means are constrained to be equal to β and the standard deviations to be equal to $\sqrt{\psi + \theta}$ for responses y_{ij} and $y_{i'j}$ at two occasions i and i'. Under the variance-components model, the marginal (not conditional on ζ_j) covariance between the measurements therefore becomes

$$\begin{aligned}\mathrm{Cov}(y_{ij}, y_{i'j}) &= E\{(y_{ij} - \underbrace{\beta}_{E(y_{ij})})(y_{i'j} - \underbrace{\beta}_{E(y_{i'j})})\} = E\{(\zeta_j + \epsilon_{ij})(\zeta_j + \epsilon_{i'j})\} \\ &= E(\zeta_j^2) + \underbrace{E(\zeta_j\epsilon_{i'j})}_{0} + \underbrace{E(\epsilon_{ij}\zeta_j)}_{0} + \underbrace{E(\epsilon_{ij}\epsilon_{i'j})}_{0} = E(\zeta_j^2) = \psi\end{aligned}$$

The corresponding correlation, called the *intraclass correlation*, becomes

$$\mathrm{Cor}(y_{ij}, y_{i'j}) = \frac{\mathrm{Cov}(y_{ij}, y_{i'j})}{\sqrt{\mathrm{Var}(y_{ij})}\sqrt{\mathrm{Var}(y_{i'j})}} = \frac{\psi}{\sqrt{\psi+\theta}\sqrt{\psi+\theta}} = \frac{\psi}{\psi+\theta} = \rho$$

Thus ρ previously given in (2.3) also represents the within-cluster correlation, which cannot be negative because $\psi \geq 0$. We see that between-cluster heterogeneity and within-cluster correlations are different ways of describing the same phenomenon: both are zero when there is no between-cluster variance $\psi = 0$ and both increase when the between-cluster variance increases relative to the within-cluster variance.

The intraclass correlation is estimated by simply plugging in estimates for the unknown parameters

$$\widehat{\rho} = \frac{\widehat{\psi}}{\widehat{\psi} + \widehat{\theta}}$$

Figure 2.6 shows data with an estimated intraclass correlation $\widehat{\rho} = 0.58$ and data with an estimated intraclass correlation $\widehat{\rho} = 0.87$.

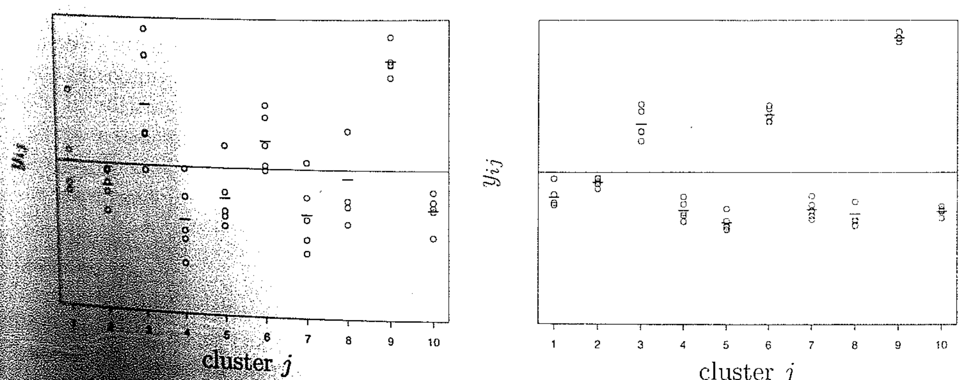

Figure 2.6: Illustration of lower intraclass correlation (left) and higher intraclass correlation (right)

In contrast to the estimated intraclass correlation, the Pearson correlation r is obtained by replacing the expectations in (2.4) by sample means and plugging in *separate* sample means $\overline{y}_{i\cdot}$ and $\overline{y}_{i'\cdot}$ and sample standard deviations s_{y_i} and $s_{y_{i'}}$ for the two occasions,

$$r = \frac{\frac{1}{J-1}\sum_{j=1}^{J}(y_{ij}-\overline{y}_{i\cdot})(y_{i'j}-\overline{y}_{i'\cdot})}{s_{y_i}s_{y_{i'}}}$$

where J is the number of clusters.

To give more insight into the interpretation of the estimated intraclass correlation and Pearson correlation, consider what happens if we alter the second Mini Wright peak flow measurements by adding 100 to them, as shown in figure 2.7. (Such a systematic increase could, for instance, be due to a practice effect.) For the variance-components model it is obvious that the within-cluster variance has increased giving a much smaller intraclass correlation than for the original data (estimated as 0.63 instead of 0.97). In contrast, the Pearson correlation r is 0.97 in both cases (figures 2.2 and 2.7) since it is based on deviations of the first and second measurements from their *respective means*. In contrast, the intraclass correlation is based on deviations from the *overall* or pooled mean.

The Pearson correlation can be thought of as a measure of *relative agreement*, which refers to how well rankings of subjects based on each measure agree and is therefore not affected by linear transformations of the measurements. In contrast, the intraclass correlation is a measure of *absolute agreement*.

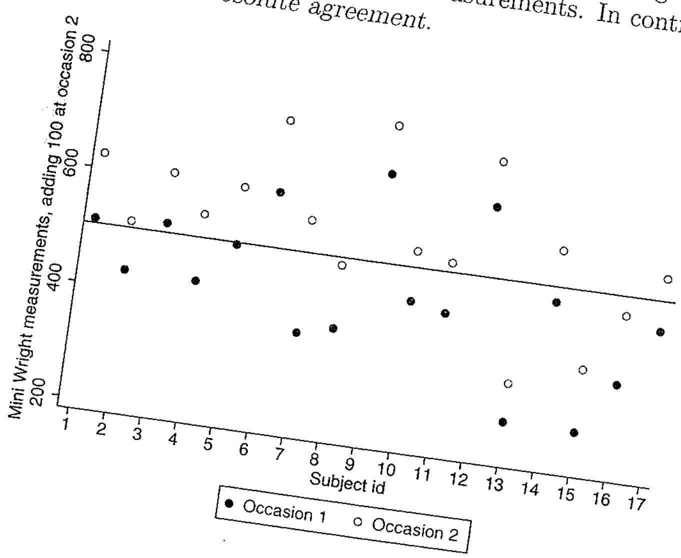

Figure 2.7: First recording of Mini Wright meter and second recording plus 100 versus subject number (the horizontal line represents the overall mean)

The intraclass correlation is useful when the units i are *exchangeable* with identical means and standard deviations. For instance, for twin data, there may not even be such a thing as the first and second twin (presuming that birth order is either irrelevant or unknown), and whereas the Pearson correlation can only be obtained by making an arbitrary assignment to y_{1j} and y_{2j} for each twin-pair, the intraclass correlation does not require this. Twins are an example of exchangeable dyads, where the intraclass correlation is more appropriate, whereas married couples are an example of nonexchangeable or distinguishable dyads where the Pearson correlation between husbands y_{1j} and wives y_{2j} is more appropriate because it is usually difficult to justify that husbands and wives have the same population mean β. Another difference between the intraclass correlation and the Pearson correlation is that the latter is only defined for pairs of variables whereas the former summarizes dependence for clusters of size larger than 2 and clusters of variable sizes; see for example exercise 2.4.

2.4 Fixed versus random effects

Each subject has a different effect ζ_j on the measured peak-expiratory-flow rates. In analysis of variance (ANOVA) terminology (see sec. 1.4 and 1.9), the subjects can therefore be thought of as the levels of a factor or categorical explanatory variable. Since the effects of subjects are random, the variance-components model is therefore sometimes referred to as a one-way random-effects ANOVA model.

The one-way *random*-effects ANOVA model can be written as

$$y_{ij} \;=\; \beta + \zeta_j + \epsilon_{ij}, \qquad \epsilon_{ij}|\zeta_j \sim N(0,\theta) \quad \zeta_j \sim N(0,\psi) \tag{2.5}$$

where ζ_j is a random intercept. In contrast, the one-way *fixed*-effects ANOVA model becomes

$$y_{ij} \;=\; \beta + \alpha_j + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0,\theta) \quad \sum_{j=1}^{J} \alpha_j = 0 \tag{2.6}$$

where α_j are unknown cluster-specific parameters. In the random-effects model, the random intercepts are independent across clusters and independent of the level-1 residuals, and in both models the level-1 residuals are independent across units. Both random-effects models and fixed-effects models include cluster-specific intercepts, ζ_j and α_j respectively, to account for unobserved heterogeneity. Thus a natural question is whether to use a random- or fixed-effects approach.

One way of answering this question is by being explicit about the target of inference, namely, whether interest concerns the *population* of clusters or the particular clusters in the dataset. If we are interested in the variance ψ for the population of clusters or inference regarding the population mean β when both clusters and units are viewed as sampled from respective populations, a random-effects approach must be used. In contrast, if we are interested in the "effects" α_j of the clusters in a particular dataset and inferences regarding β when only units (and not clusters) are taken as sampled, a fixed-effects approach must be used.

Whether the clusters are viewed as sampled from a population mostly affects the estimated standard error of $\widehat{\beta}$ as we will see in section 2.7.2 but can also affect $\widehat{\beta}$ itself.

We can also make inferences regarding the particular clusters in the dataset when using a random-effects approach, as will be discussed in section 2.9.

It is often said that the random-effects approach should only be used if there is a sufficient number of clusters in the sample, typically more than 10 or 20. This is true if the variance components are of interest since ψ will be poorly estimated with a small number of clusters. However, if a random-effects approach is used merely to make appropriate inferences regarding β, a smaller number of clusters may suffice.

Regarding cluster sizes, these should be large in the fixed-effects approach if the α_j are of interest. However, in random-effects models, it is only required that there are a good number of clusters of size 2 or more. It does not matter if there are also "clusters" of size 1. Such singleton clusters do not provide information on the within-cluster correlation or on how the total variance is partitioned into ψ and θ, but they do contribute to the estimation of β and $\psi + \theta$.

In the peak-expiratory-flow application, the one-way fixed-effects ANOVA model has 19 parameters (β, $\alpha_1, \ldots, \alpha_{17}$,$\theta$) and one constraint ($\sum_j \alpha_j = 0$). The one-way random-effects ANOVA model is thus more parsimonious, having only 3 parameters (β, ψ, θ).

2.5 Estimation using Stata

In Stata, maximum likelihood estimates for variance-components models can be obtained using `xtreg` with the `mle` option, `xtmixed` with the `mle` option, or `gllamm`. Restricted maximum likelihood (REML) estimates can be obtained using `xtmixed, reml` (the default method), and generalized least-squares (GLS) estimates can be obtained using `xtreg, re` (the default method). See section 2.7.1 for information on these estimation methods.

`xtreg` is undoubtedly the most computationally efficient procedure for variance-components models. The postestimation command `predict` for `xtmixed` and the postestimation command `gllapred` for `gllamm` are useful for predicting random intercepts. However, at the time of writing this book only `gllapred` produces different kinds of standard errors for the predictions.

We do not generally recommend using `gllamm` for linear variance-components models because `xtreg` and `xtmixed` are more computationally efficient and sometimes more accurate than `gllamm` for such models.

2.5.1 Data preparation

We now set up the data for estimation in Stata. Currently, the responses for occasions 1 and 2 are in *wide format* as two separate variables, `wp1` and `wp2` for the Wright peak flow meter and `wm1` and `wm2` for the Mini Wright peak flow meter:

```
. list if id < 6, clean noobs

 id   wp1   wp2   wm1   wm2   mean_wm
  1   494   490   512   525     518.5
  2   395   397   430   415     422.5
  3   516   512   520   508       514
  4   434   401   428   444       436
  5   476   470   500   500       500
```

However, we need to stack the occasion 1 and 2 measurements using a given meter into one variable. We can use the `reshape` command to obtain such a *long format* with one variable `wp` for both Wright peak flow meter measurements, one variable `wm` for both Mini Wright peak flow meter measurements, and a variable `occasion` (equal to 1 and 2) for the measurement occasion:

```
. reshape long wp wm, i(id) j(occasion)
(note: j = 1 2)

Data                                 wide   ->   long
-------------------------------------------------------------
Number of obs.                         17   ->     34
Number of variables                     6   ->      5
j variable (2 values)                       ->   occasion
xij variables:
                                  wp1 wp2   ->   wp
                                  wm1 wm2   ->   wm
-------------------------------------------------------------
```

The data for the first five subjects now look like this:

```
. list if id < 6, clean noobs

 id   occasion    wp    wm   mean_wm
  1          1   494   512     518.5
  1          2   490   525     518.5
  2          1   395   430     422.5
  2          2   397   415     422.5
  3          1   516   520       514
  3          2   512   508       514
  4          1   434   428       436
  4          2   401   444       436
  5          1   476   500       500
  5          2   470   500       500
```

In the above `reshape` command, `i()` is used to specify clusters, denoted j in this book, and `j()` is used to specify units within clusters, denoted i in this book.

2.5.2 Using xtreg

We can now estimate the parameters of the variance-components model (2.2) using the `xtreg` command with the `mle` option, which stands for maximum likelihood estimation (see sec. 2.7.1).

As in the `regress` command, the response variable `wm` and explanatory variables are listed after the command name. In the variance-components models, the fixed part is just the intercept β, which is included by default, so we do not specify any explanatory variables. The random part includes a random intercept ζ_j whose cluster identifier j can be specified in the `i()` option, here `i(id)`. The level-1 residual ϵ_{ij} need not be specified because it is always included. Therefore the command is

```
. xtreg wm, i(id) mle

Random-effects ML regression                    Number of obs      =        34
Group variable: id                              Number of groups   =        17

Random effects u_i ~ Gaussian                   Obs per group: min =         2
                                                               avg =       2.0
                                                               max =         2

                                                Wald chi2(0)       =         .
Log likelihood  = -184.57839                    Prob > chi2        =      0.00

------------------------------------------------------------------------------
          wm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   453.9118   26.18616    17.33   0.000     402.5878    505.2357
-------------+----------------------------------------------------------------
    /sigma_u |   107.0464   18.67858                      76.0406    150.6949
    /sigma_e |   19.91083   3.414659                      14.2269     27.8656
         rho |   .9665602   .0159494                      .9210943    .9878545
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=    46.27 Prob>=chibar2 = 0.000
```

Instead of specifying the cluster identifier in the `i()` option, we could have declared the data to be "panel data" and `id` to be the "group variable" using the `xtset` command before the `xtreg` command.

We see in the output that there are 34 observations belonging to 17 groups (the clusters, here subjects) and that the log likelihood is -184.58. In the Stata output and in the Stata documentation for `xtreg`, the i subscript is used for clusters (instead of j used in this book), u_i for the random intercept (instead of ζ_j) and the t subscript for occasions (instead of i used in this book). Clusters are called "panels" in the documentation.

The estimate of the overall population mean β, given next to `_cons` in the output, is 453.91. The estimate of the between-subject standard deviation $\sqrt{\psi}$ of the random intercepts of subjects, referred to as `/sigma_u`, is 107.05, and the estimate of the within-subject standard deviation, $\sqrt{\theta}$, referred to as `/sigma_e`, is 19.91. It follows that the intraclass correlation is estimated as

$$\widehat{\rho} = \frac{\widehat{\psi}}{\widehat{\psi} + \widehat{\theta}} = \frac{107.05^2}{19.91^2 + 107.05^2} = 0.97$$

which is referred to as `rho` in the output. This estimate is close to one, indicating that the Mini Wright peak flow meter is very reliable. However, the reliability is not just a characteristic of the instrument; it also depends on the between-subject variance, ψ, and this can differ between populations. The parameter estimates are also given in table 2.2 under "`xtreg, xtmixed`".

2.5.3 Using xtmixed

The variance-components model considered here is a simple special case of a linear mixed-effects models that can be fitted using the `xtmixed` command (available as of Stata 9).

The fixed part of the model, here β, is specified as in any estimation command in Stata (response variable followed by list of explanatory variables). The random part, except the residual ϵ_{ij}, is specified after two vertical bars `||`. To include a random intercept ζ_j, which varies between subjects whose identifier is in the variable `id`, the syntax is simply `id:` because a random intercept ζ_j is included by default (it can be excluded using the `noconstant` option). Finally, we can request maximum likelihood estimation using the `mle` option:

```
. xtmixed wm || id:, mle
Mixed-effects ML regression                     Number of obs      =        34
Group variable: id                              Number of groups   =        17

                                                Obs per group: min =         2
                                                               avg =       2.0
                                                               max =         2

                                                Wald chi2(0)       =         .
Log likelihood = -184.57839                     Prob > chi2        =         .

          wm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   453.9118   26.18616    17.33   0.000     402.5878    505.2357

  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                   sd(_cons) |   107.0464   18.67857      76.0406    150.6949
-----------------------------+------------------------------------------------
                sd(Residual) |   19.91083   3.414679     14.22688    27.86565

LR test vs. linear regression: chibar2(01) =    46.27 Prob >= chibar2 = 0.0000
```

The table of estimates for the fixed part has the same form as that for `xtreg` and all Stata estimation commands. The random part is given under `Random-effects Parameters`. Here `sd(_cons)` is the estimate of the random-intercept standard deviation $\sqrt{\psi}$, and `sd(Residual)` is the estimate of the standard deviation $\sqrt{\theta}$ of the level-1 residuals. All these estimates are identical to the estimates using `xtreg` and are given under "`xtreg`, `xtmixed`" in table 2.2. We could also obtain estimated variances (instead of standard deviations) with their standard errors using the `variance` option.

There are some differences in the terminology and notation used in this book and the Stata documentation for `xtmixed`. Using the usual multilevel or hierarchical modeling conventions, we use the indices i for occasions and j for subjects and call the corresponding levels 1 and 2. In contrast, the `xtmixed` documentation uses i for subjects and j for occasions and calls subjects level 1. Contrary to common terminology where multilevel models for such data are called two-level models, the `xtmixed` documenta-