5

INTRODUCTION TO RANDOM EFFECTS AND MIXED MODELS

> This class includes all problems of estimating, and testing to determine whether to infer the existence of, components of variance ascribable to random deviation of the characteristics of individuals of a particular generic type from the mean values of these characteristics in the "population" of all individuals of that generic type, etc. In a sense, this is the true analysis of variance, and the estimation of the respective components of the over-all [sic] variance of a single observation requires further steps beyond the evaluations of the entries of the analysis-of-variance table itself.
>
> (Eisenhart, 1947, p. 4)

The researcher of the previous two chapters, having discovered an effect of melatonin dosage on sleep onset, now ponders the following question:

> **Is sleep onset a function not only specific doses, but of melatonin dosage in general? That is, if we randomly sampled 3 dosages from a population of potential doses, would these differing doses account for variation in sleep onset?**

In this situation, the researcher is not interested specifically in any particular set of doses. Rather, the researcher would like to draw the conclusion that **differing dose level is associated with differing sleep onset**. The effect for dose in this case would be considered a **random effect**, since levels of dose are randomly drawn from a wider population of possible doses. The subset of dosages randomly sampled for the given experiment is used to make a generalization to the population of dosage levels. This type of design calls for the **random effects analysis of variance model**.

Upon further thought, not only is the researcher interested in randomly sampling three dosage levels for use in his experiment, but just as he did for the two-way model of the previous chapter, he also wants to include ambient noise as a factor in his design. For this factor, he is only interested in

Applied Univariate, Bivariate, and Multivariate Statistics: Understanding Statistics for Social and Natural Scientists, With Applications in SPSS and R, Second Edition. Daniel J. Denis.

comparing levels **no noise** to **some noise** and hence keeps the factor **fixed**. He is not interested in generalizing to the population of noise levels. Hence, the researcher will now have one random factor (dose) and one fixed factor (noise) in his experiment. This type of design calls for a **mixed effects factorial analysis of variance model**.

In this chapter, we survey the random effects and mixed effects analysis of variance models. As we did for one-way and factorial fixed effects, we develop the conceptual basis and then move on to a consideration and development of suitable *F*-ratios to test effects. As we saw in previous chapters, in a fixed effects model, whether one-way, two-way, or higher-order, expected mean squares revealed that MS error was the correct error term for testing main effects and interactions. As we will see in the random effects and mixed models, **MS error is not always the most suitable error term for testing effects**. We will survey some of the theories as to why other error terms are more suitable in these situations. We also provide software examples of random effects and mixed effects models in R. For fitting mixed models in R, readers should consult Gelman and Hill (2007). Pinheiro and Bates (2000) provide an excellent treatment of the wider mixed effects model in S-Plus. Demidenko (2004) provides a very technical treatment along with some applications.

5.1 WHAT IS RANDOM EFFECTS ANALYSIS OF VARIANCE?

Recall that in the fixed effects models studied in previous chapters, what made the effects in these models "fixed" was the fact that over theoretical repetitions of the experiment, levels of the independent variable were to remain **constant**. For example, in the melatonin experiment, the fixed factor of dosage was so named because the researcher had a specific interest in the dosages tested. The idea of a random effects model is that over theoretical repetitions of the experiment, treatment effects are no longer assumed to remain fixed. Rather, treatment effects are considered to be random, and hence over numerous theoretical replications of the experiment (i.e., if we were to perform them), it is reasonable to assume that we will obtain different treatment levels when sampling each time. In a random effects model then, **the levels of a random factor are randomly sampled from a population of possible levels that could have been included in the given experiment**. When a factor is a random factor, it implies that there is a probability distribution of levels associated with that factor, and what you are using in your experiment is but a **sample** of levels from a wider set of potential levels that could have been used. In the language of sets, the levels randomly sampled are but a **proper subset** of the wider set of population levels. As Casella (2008) noted:

> … by the very nature of a random factor, we are not really interested in estimating the levels of the factor that are in the experiment. Why? Because if the factor is truly random, the levels in the experiment are nuisance parameters, and only the variance of the factor is meaningful for inference. (p. 101)

Historically, nobody better described the concept of a random effects model than Eisenhart (1947):

> … when an experimenter selects two or more treatments, or two or more varieties, for testing, he rarely, if ever, draws them at random from a population of possible treatments or varieties; he selects those that he believes are most promising. Accordingly Model I [fixed effects] is generally appropriate where treatment, or variety comparisons are involved. On the other hand, when an experimenter selects a sample of animals from a herd or a species, for a study of the effects of various treatments, he can insure that they are a random sample from the herd, by introducing randomization into the sampling procedure, for example, by using a table of random numbers. But he may consider such a sample to be a random sample from the species, only by making the assumption that the herd itself is a random sample from the species. In such a case, if several herds (from the same species) are involved, Model II [random effects] would clearly be appropriate with respect to the variation among the animals from each of the respective herds, and might be appropriate with respect to the variation of the herds from one another. (p. 19)

The random effects model has sometimes historically been called a **components of variance** model (Searle, Casella, and McCulloch, 1992) because unlike the fixed effects model in which the primary interest is in testing null hypotheses about specific differences between population means, the primary interest in a random effects model is in estimating **variance** in the dependent variable that can be attributed to main effects or interactions. This estimate of variance accounted for will apply not only to the levels actually sampled but to the larger set of possible levels (i.e., population) from which our sample was drawn. Hence, in random effects models, our primary goal is to estimate **components of variance** rather than test null hypotheses about equalities among population means as was the case in the fixed effects model.

5.2 THEORY OF RANDOM EFFECTS MODELS

Insight into the random effects model can be gleamed from a brief discussion of its assumptions, and then by comparing these assumptions to those made in the previously studied fixed effects model. Recall the one-way analysis of variance model of Chapter 3:

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

where μ is the grand mean, α_j is a population effect estimated by the sample effect $\bar{y}_j - \bar{y}.$, and ε_{ij} is the error associated with observation i in group j. We first list the assumptions for the one-way random effects model that parallel those of the fixed effects model:

- For any treatment j, the errors ε_{ij} are normally distributed, with a mean of 0 (i.e., $E(e_{ijk}) = 0$) and variance σ_e^2, which is identical for each possible treatment j. That is, $N\left(0, \sigma_e^2\right)$. Notice that this assumption parallels the assumption of normality in the fixed effects model.
- The values of the random variable e_{ijk} are all independent (as was also assumed in the fixed effects model). In cases of naturally-occurring, or imposed hierarchical nesting structures, errors within groups may be related (see Section 5.17 and Chapter 6 for details on blocking and nesting).
- $\sigma_{e_{ijk}}^2 < \infty$, that is, the variance of the errors is some finite number (which, as was true in the one-way and two-way models, implies that it is less than infinity).
- $\sigma_{jk=1}^2 = \sigma_{jk=2}^2 = \sigma_{jk=JK}^2$, that is, the variances across cell populations are equal (recall this is called the **homoscedasticity** assumption and is essentially the same as in the fixed effects models studied previously).

Where the random effects model differs from the fixed effects model is in the following assumptions:

- a_j is a **random variable** having a distribution with mean 0 and variance σ_A^2. That is, unlike the fixed effects model, the sample treatment effects a_j are no longer considered to be **constant** across replications. Analogous to how we can reach into a bag and take a sample of 10 objects and calculate a sample mean on them, the sample mean can be considered to be a random variable that can vary from experiment to experiment. We now need to treat a_j as possibly fluctuating from sample to sample or from experiment to experiment. They are no longer **fixed** as they were in the fixed effects model.
- The values of the random variable a_j occurring in the experiment are all independent of each other (Hays, 1994).

- Each pair of random variables a_j and e_{ij} are independent. That is, the sample effects are independent of error (or if you wish, the error **effects**).

Note that the assumptions for a random effects model are for two different distributions, one for the distribution of the random variable a_j, and the other for e_{ij}. In the fixed effects model, we only made a probability assumption about e_{ij}, since we assumed a_j to be fixed across theoretical replications. Since the sample effects were assumed to be fixed, it made no sense to associate them with a probability distribution.

5.3 ESTIMATION IN RANDOM EFFECTS MODELS

There have been, historically, several different methods of estimating parameters in random effects and mixed models. The classic method in which one computes **expected values of mean squares** is historically known as **ANOVA estimation** (Wu, Yu, and Liu, 2009). This methodology has some flaws and drawbacks, and in part because of the advances in computing power, other methods of estimation have come into vogue, which include **maximum-likelihood (ML)**, **restricted maximum-likelihood (REML)**, and **minimum norm quadratic unbiased estimation**. Of these, ML and REML are dominant today in the estimation of variance components in both random effects and mixed models. These methods of estimation, however, are quite complex and require **iteration** for their solution.

As we did in prior chapters, we focus on the method of taking expectations (ANOVA estimation), largely because under certain conditions, results of ANOVA estimation match those of the iterative methods. Also, a brief study of expectations in ANOVA models, I believe, goes a long way to demystifying the theory behind estimation in general, and opens the door for the reader to understand more complex methods for estimating parameters.

In what follows then, we begin with the principles developed in previously studied fixed effects models and derive expected mean squares for random effects models. Our discussion and derivation is based largely on the work of Hays (1994), Kempthorne (1975), Searle, Casella, and McCulloch (1992), and Scheffé (1999), who all present thorough accounts of random effects ANOVA.

5.3.1 Transitioning from Fixed Effects to Random Effects

Recall the quantities of **MS Between** and **MS Within** as first derived in the fixed effects model of Chapter 3:

$$\text{MS between} = \frac{\text{SS between}}{J-1} = \frac{\sum_j n_j\left(\bar{y}_j - \bar{y}.\right)^2}{J-1}$$

$$\text{MS within} = \frac{\text{SS within}}{N-J} = \frac{\sum_j\sum_i\left(y_{ij} - \bar{y}_j\right)^2}{N-J}$$

Should we expect derived EMS on these values to be the same in a random effects model? Not necessarily. The reason is that now we are randomly selecting the J different factor levels. They are no longer fixed. Because of this, as we will see, our expected mean squares will change. They will change because we are no longer interested in population mean differences. We are interested, rather, in estimating **variances**.

Because we are randomly sampling the levels of our factor in a random effects model, we can write the mean of the sample random effects as

$$\bar{a} = \frac{\sum_i a_j}{J}$$

where a_j is, as before, the sample effect $(\bar{y}_j - \bar{y}.)$ for a given group $J = j$. This is the mean of the sample effects for the given experiment we are conducting. Theoretically, if we were to conduct the experiment again, and obtain new levels, we would obtain another $\bar{a}$ for that particular experiment, and so on for additional repetitions of the experiment. What is key to understand here is that **this mean will surely vary from sample to sample due to sampling error** (i.e., the error generated simply by the process of sampling) inherent in the random effect. This is why a_j has now become a random variable. We can have some certainty, however, that the mean effect over **all infinite samples that could be drawn from the population** will equal to zero. More formally, we say that the expected value of $\bar{a}$ will be 0, $E(\bar{a}) = 0$. However, the value of $\bar{a}$ in any given sample need not be equal to the long-run expectation. That is, as noted by Hays (1994, p. 530), "… although the mean of the effects **over all the possible treatments** [emphasis added] must be 0, the mean $\bar{a}$ of the sample effects present in a given set of data need not be 0."

Theoretically then, in any particular experiment, the value of $\bar{a}$ is not constrained to equal 0 as it was in the fixed effects model. The major point is that in any given model with a random effects term (other than the obvious e_{ij} effect, which is indeed also a random effect), we must somehow deal with the fact that these **treatment effects a_j are now random**. **Being random, their values will undoubtedly change from experiment to experiment**. This change in assumption figures prominently in the derivation of the expected mean squares. We will see that because of this random quality of the sample effects, the expected mean squares in the random effects model are quite different than in the fixed. Likewise, **null hypotheses** will be defined differently as well.

5.3.2 Expected Mean Squares for MS Between and MS Within

Recall once more the reason for taking expectations of mean squares. It is to learn what parameter our given mean squares is estimating. By calculating EMS, we can then use these to generate suitable F-ratios to test various effects of interest, whether they be main effects or interactions.

As Hays (1994) does, we begin our derivation by conceptualizing the mean of the errors for any group j in a one-way random effects ANOVA as

$$\bar{e}_j = \frac{\sum_i e_{ij}}{n}$$

where $\bar{e}_j$ is the mean error for a given group, $\sum_i e_{ij}$ is the sum of errors across all groups j, and n is the sample size per group (as before, we are assuming a balanced design). If we take this for the entire sample across J groups, we will have

$$\bar{e} = \frac{\sum_j \sum_i e_{ij}}{N} = \frac{\sum \bar{e}_j}{J}$$

which means that the average overall error is equal to the mean error, $\bar{e}_j$, per group. Given this, and just as we did in previous chapters where we wrote out model equations, we can write the deviation of any group mean $\bar{y}_j$ from the grand sample mean $\bar{y}.$ as

$$(\bar{y}_j - \bar{y}.) = (a_j - \bar{a}) + (\bar{e}_j - \bar{e}) \tag{5.1}$$

Why does it make sense to write the deviation of a group mean from the grand mean as in (5.1)? This makes sense, because we just mentioned that we can calculate a "mean of errors" term over all groups. If this is the case, then it stands to reason that for a given group *j*, the mean error for that particular group minus the overall mean error for the entire data will give us the "effect" of error for that particular group, just as $(\bar{y}_j - \bar{y}.)$ gives us the sample "effect" of being in a particular group *j*. Notice that the sum of effects for $(a_j - \bar{a})$ will sum to 0, and the sum of effects for $(\bar{e}_j - \bar{e})$ will also sum to 0; so, as usual, we take the **squared deviations**, otherwise the entire right-hand side of (5.1) will always sum to zero (this idea of the sum of unsquared effects always equaling zero should be becoming familiar territory by now). Squaring (5.1), summing, and taking expectations, we get (Hays, 1994, p. 531):

$$E\left[\sum_j (\bar{y}_j - \bar{y}.)^2\right] = E\left[\sum_j (a_j - \bar{a})^2\right] + E\left[\sum_j (\bar{e}_j - \bar{e})^2\right] \tag{5.2}$$

From (5.2), we have the expected mean squares for SS between:

$$\begin{aligned} &E\left[\sum_j (\bar{y}_j - \bar{y}.)^2\right] = E\left[\sum_j (a_j - \bar{a})^2\right] + E\left[\sum_j (\bar{e}_j - \bar{e})^2\right] \\ &E(\text{MS between}) = n\sigma_{\text{A}}^2 + \sigma_e^2 \end{aligned} \tag{5.3}$$

where *n* is the number of subjects (or objects) per group, σ_A^2 is the variance attributable to varying levels of factor A, and σ_e^2 is the variance of the error. That is, the sum of squares for between is equal to a source of variability for factor A, $n\sigma_A^2$, and a source of variability represented by the error term, σ_e^2.

The **expectation for error**, as was true for the fixed effects model, is the average error per group:

$$E(\text{MS within}) = \sum \frac{\sigma_e^2}{J} = \sigma_e^2$$

That is, **MS within**, just as was the case for the fixed effects ANOVA, is an unbiased estimate of error variance, and **only** error variance.

5.4 DEFINING NULL HYPOTHESES IN RANDOM EFFECTS MODELS

In the random effects model, null hypotheses are stated differently than in a fixed effects model. A null hypothesis in a random effects model is not really about means. It is more about **variances**. Or to be even more precise, **variance components**. The null hypothesis for the one-way random effects model is given by

$$H_0 : \sigma_A^2 = 0$$

where σ_A^2 is the variance attributable to differing levels of factor A. If changing levels of the factor is not associated with any change in the dependent variable in our sample, then it stands that the variance explained, sampling error aside, should equal to 0. And since the purpose of conducting the investigation is usually to show that varying levels of the factor is associated with variance explained in the dependent variable, our alternative hypothesis is given by:

$$H_1 : \sigma_A^2 > 0$$

Notice that the alternative hypothesis is specified in terms of a **positive** value. The **greater than** sign denotes that σ_A^2 cannot be zero or negative given a rejection of the null hypothesis. This is reasonable, since we know variance, by definition, is a positive quantity. If there are treatment effects, either for those treatments sampled or across all treatment levels in the population, we would expect the variance attributable to our factor to be greater than 0. For the one-way random effects model then, there are two "**components of variance**" that need to be obtained. One is σ_A^2, the other is σ_e^2. Both of these components add up to the total variance σ_y^2 in the dependent variable. That is, $\sigma_y^2 = \sigma_A^2 + \sigma_e^2$. We will discuss shortly why this is the case.

5.4.1 *F*-Ratio for Testing H_0

How do we come up with a suitable ratio for testing $H_0 : \sigma_A^2 = 0$? We do so by considering the derived expected mean squares. As was the case for the fixed effects model, we want to isolate that part of the expected mean squares that represents the "effect" we are interested in. In $n\sigma_A^2 + \sigma_e^2$, that part is $n\sigma_A^2$. That is, if our experimental treatment "worked," (in some sense) we would expect $n\sigma_A^2$ to be large relative to σ_e^2. Notice that once we have isolated the part we are interested in, as was true for the fixed effects models of the previous chapters, the correct error term quite naturally reveals itself. Since we do not want our effects to be "polluted" by σ_e^2, we will divide $n\sigma_A^2 + \sigma_e^2$ by σ_e^2. But what is σ_e^2? This is the expectation of MS within. Hence, the *F*-ratio we want to produce is one which takes $n\sigma_A^2 + \sigma_e^2$ in the numerator and divides it by σ_e^2. That is, our *F*-ratio for the one-way random effects model is:

$$F = \frac{n\sigma_A^2 + \sigma_e^2}{\sigma_e^2}$$

At first glance, it may appear that we can simply cross out σ_e^2 in the numerator and σ_e^2 in the denominator. However, recall from the rules of algebra that we cannot do this since the numerator is a **sum** and not a **product**. Had the numerator been $\left(n\sigma_A^2\right)\left(\sigma_e^2\right)$, where the parentheses denote multiplication, then crossing out σ_e^2 would have worked. But since we are dealing with addition, we cannot eliminate σ_e^2 in this way.

Returning to our *F*-ratio, we can appreciate why it makes good sense to construct it as we did. If there are no treatment effects for our factor, then $n\sigma_A^2$ will be 0, since σ_A^2 would equal 0, and any n (i.e., sample size per group in a balanced design) multiplied by 0 will equal 0. Under this condition, we are simply left with σ_e^2 in the numerator, and our *F*-ratio would be equal to approximately

$$F = \frac{n\sigma_A^2 + \sigma_e^2}{\sigma_e^2} = \frac{n(0) + \sigma_e^2}{\sigma_e^2} = \frac{\sigma_e^2}{\sigma_e^2} \approx 1$$

That is, we can state more formally that under H_0,

$$E\left(\frac{n\sigma_A^2 + \sigma_e^2}{\sigma_e^2}\right) \approx 1$$

If, on the other hand, H_0 is false, then this implies the alternative hypothesis, $\sigma_A^2 > 0$, and so $n\sigma_A^2$ will be some quantity **larger** than 0. Our expectation then for our ensuing *F*-ratio would be

$$E\left(\frac{n\sigma_A^2 + \sigma_e^2}{\sigma_e^2}\right) > 1$$

under a false null hypothesis. As was the case for the fixed effects model, we evaluate F on $J - 1$ and $N - J$ degrees of freedom. A statistically significant F-statistic suggests that the variance attributable (or "accounted for") by our factor (i.e., either the levels represented in the sample or by the population of levels) is not equal to 0 in the population from which these data were drawn. That is, a rejection of the null hypothesis implies that the variance in our dependent variable that is accounted for by our factor is greater than 0.

5.5 COMPARING NULL HYPOTHESES IN FIXED VERSUS RANDOM EFFECTS MODELS: THE IMPORTANCE OF ASSUMPTIONS

It would do well at this point to emphasize and reiterate the fact that a rejection of the null hypothesis in a random effects analysis of variance tells us something different than a rejection of a null hypothesis in the fixed effects models of the previous chapters. In the fixed effects model, we tested hypotheses about **means**. In the random effects model, we are testing hypotheses about **variances**. A rejection of the null hypothesis in a fixed effects model hints to us that somewhere among the population means, it looks like there is a mean difference. A rejection of the null hypothesis in the random effects model tells us that changing levels of the independent variable has the effect of explaining or accounting for variance in the dependent variable. These two null hypotheses are **not the same**.

What we have noticed, however, is that the error terms used for testing both hypotheses in the one-way fixed effects and one-way random effects model are the same. In both cases, **MS error is the correct error term**. Why are they the same? They are the same (so far) because in both cases, the one-way fixed and one-way random effects, MS error "gets the job done" in terms of isolating the term in the numerator that we are interested in. Recall that in the one-way fixed effects model, the **expectation for MS between** was equal to

$$E(\text{MS between}) = \sigma_e^2 + \frac{\sum_j n_j \alpha_j^2}{J - 1}$$

The **expectation for MS within** was equal to σ_e^2, and so because we were interested in isolating

$$\frac{\sum_j n_j \alpha_j^2}{J - 1}$$

since it contained any treatment effects present, it made sense to use σ_e^2 as the error term. I want to emphasize that this is why we used MS within as the error term, because it made sense to do so in terms of what we wished to isolate in the numerator. This is the general logic of choosing error terms in ANOVA, whether in simple designs or more complex. Deciding on a correct error term is not a "mysterious" process once you have the expected mean squares at your disposal (on the other hand, **deriving** EMS can be somewhat difficult).

The expectation for MS within is again equal to σ_e^2 in our current random effects model, and so because we are interested in isolating $n\sigma_A^2$, it again makes sense to use MS error as the error term. Also, be sure to note that the phrase **error term** and **MS error** are not synonymous with one another. Under our current discussion, MS error is the appropriate error term. As we will see for the two-factor random effects model, the correct error term will be other than MS error. **It is extremely important to not get into the habit of automatically associating "error term" with "MS error." MS error is, under many circumstances and models, the appropriate error term, but under other models, it no longer is. In those cases, we will seek an error term other than MS error**.

5.6 ESTIMATING VARIANCE COMPONENTS IN RANDOM EFFECTS MODELS: ANOVA, ML, REML ESTIMATORS

Once we have computed the analysis of variance, whether in the one-way or two-way (to be discussed shortly) or higher-order analyses, our next job is to estimate **variance components** for such models. Note that in our computations of analysis of variance so far, we have not yet addressed just how quantities such as σ_A^2 and σ_e^2 are estimated. All we have considered thus far is how to use these quantities to help us derive suitable *F*-ratios. We first consider **ANOVA estimators**, and then move on to a brief consideration of **maximum likelihood** and **restricted maximum likelihood**.

5.6.1 ANOVA Estimators of Variance Components

ANOVA estimators are easily computed, and in some cases can be used as starting values to other forms of estimation. They are also the most historically relevant in the evolution of variance component estimation. Recall once more the expectation for MS between found in (5.3). We can solve for σ_A^2 and get an unbiased estimate of σ_A^2:

$$\begin{aligned} E(\text{MS between}) &= n\sigma_A^2 + \sigma_e^2 \\ n\sigma_A^2 + \sigma_e^2 &= E(\text{MS between}) \\ n\sigma_A^2 &= E(\text{MS between}) - \sigma_e^2 \end{aligned}$$

We can then obtain our estimate of the variance component σ_A^2 quite simply:

$$\hat{\sigma}_A^2 = \frac{\text{MS between-MS within}}{n}$$

where MS between and MS within are obtained from the ANOVA, and n is the sample size per group in a balanced design. The next question is how to use this component. By itself, it simply represents a quantity of variance. What we would like to obtain is a **proportion of variance** attributable to our factor relative to the total variance in our dependent variable. To obtain this estimate, we need to know that the variance of our dependent variable y can be written as a function of two components in the one-way random effects model. The first component is σ_A^2, while the second component is σ_e^2. That is,

$$\sigma_y^2 = \sigma_A^2 + \sigma_e^2$$

This tells us that the total variance in a population for a one-factor experiment is composed of variability due to our factor, σ_A^2, and variability not due to our factor, which is relegated to the error component, σ_e^2.

The question now becomes how to estimate the total variance σ_y^2 in the random effects model. We do so by (Hays, (1994, p. 534)):

$$\begin{aligned} \hat{\sigma}_y^2 &= \hat{\sigma}_A^2 + \hat{\sigma}_e^2 \\ &= \frac{\text{MS between} + (n-1)\text{MS within}}{n} \end{aligned}$$

where $\hat{\sigma}_y^2$, $\hat{\sigma}_A^2$, and $\hat{\sigma}_e^2$ are respective estimates of variances σ_y^2, σ_A^2, and σ_e^2. Having estimated the respective components of variance, we can now assess the proportion of variance due to, or accounted for, by our factor. We take the following ratio, called the **intraclass correlation coefficient**:

$$\hat{\rho} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_e^2} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_y^2} \tag{5.4}$$

The intraclass correlation coefficient measures the proportion of variance due to the grouping factor, and like all proportions, ranges from 0 to 1. As noted by Kirk (1995), it is generally considered to be the most popular measure of **effect size** for random effects.

A second, related interpretation of the intraclass correlation, is that it is the **bivariate correlation coefficient between any two randomly selected observations within a given level of the independent variable** (Fox, 2016). That is, we can define ρ as

$$\rho = \mathrm{cor}\left(y_{ij}, y_{ij'}\right)$$

where y_{ij} and $y_{ij'}$ are two distinct observations in a given group j. Intraclass correlations are useful in measuring proportions of variance explained in applications of random effects and mixed models of the current chapter as well as blocking and repeated measures models of the following chapter.

5.6.2 Maximum Likelihood and Restricted Maximum Likelihood

As discussed by Searle, Casella, and McCulloch (1992), ANOVA estimation has some weaknesses, including the fact that **negative variance estimates** are possible. According to Casella (2008, p. 143), negative variance estimates are often the fault of the estimation procedure rather than the model. Further, Casella notes that a negative variance component should not in itself imply a conclusion that $\sigma_A^2 = 0$, and that when negative estimates occur, one should try a better estimation procedure, such as **restricted maximum likelihood** (REML), which is a variation of **maximum likelihood** (ML).

Maximum likelihood estimation has its recent history beginning with a paper by Hartley and Rao (1967) in which ML equations were derived, but required iterative calculations to estimate variance components. At first, these computations were quite laborious, but with the advent of high-speed computing, iterations are now performed with relative ease and speed. Closed-form solutions for ML estimation are usually heavily dependent on normality assumptions.

Restricted maximum likelihood estimation focuses on maximizing the likelihood which is invariant (i.e., does not change) to the fixed effects of the model (called the **location parameters** of the model). REML estimates variance components as a function of **residuals** that are left over after estimating the fixed effects by least-squares (Searle, Casella, and McCulloch, 1992). For balanced data, REML solutions are identical to ANOVA estimators. For unbalanced data, ML and REML are generally preferable over ANOVA estimators (Searle, Casella, and McCulloch, 1992). Choosing between ML and REML is not straightforward, and our best advice is to follow the recommendation of Searle, Casella, and McCulloch (1992):

> As to the question "ML or REML?" there is probably no hard and fast answer. Both have the same merits of being based on the maximum likelihood principle – and they have the same demerit of computability requirements. ML provides estimators of fixed effects, whereas REML, of itself, does not. But with balanced data REML solutions are identical to ANOVA estimators which have optimal minimum variance properties – and to many users this is a sufficiently comforting feature of REML that they prefer it over ML. (p. 255)

5.7 IS ACHIEVEMENT A FUNCTION OF TEACHER? ONE-WAY RANDOM EFFECTS MODEL IN R

Recall once more the `achiev` data of Chapters 3 and 4 (reproduced in Table 5.1). In those chapters, we designated teacher as a **fixed effect**. In the current analysis, we will consider it to be a **random effect**.

Imagine the following scenario—You are the parent of Taylor, an 11-year old child in sixth grade elementary education. Taylor is not performing as well as you would like in school, and based on a few verbal reports from your daughter and parents of other children, you suspect it may have something to do with Taylor's teacher. The principal of the school, however, comes to the teacher's defense and makes the following claim to you: "Student achievement is not associated with teacher. Whether a student has one teacher or another makes no difference in how the child performs."

In advancing your argument, you would like to accumulate some evidence to help substantiate that teacher does play a "role" in academic achievement. You **randomly sample** four teachers from your city and obtain mathematics achievement scores from the children in those classes, scored from 0 to 100, where higher scores are indicative of greater achievement. Ideally, children would also be randomly assigned to teacher, but for now, our focus is simply on understanding how teacher can be considered a random effect. Even if not by experimental design, it is most likely that children were randomly assigned to teacher from the outset (i.e., unless of course a school designates particular students for particular teachers, in which case, random assignment is not taking place). For our purposes here, again, we focus simply on teachers being randomly selected from a wider set of teachers.

Notice that your hypothesis calls for a **one-way random effects model**, since levels of teacher were randomly sampled. Surely, you are not interested in showing differences (i.e., mean differences) between **these particular teachers you have sampled**. Rather, you would like to draw the conclusion that variance in achievement is a function of different teachers, of which these four in your design constitute a random sample of teachers for the given study. We thus have the perfect setup for a one-way random effects model. Should your study be "successful" in that you obtain evidence that variance in achievement accounted for by teacher is greater than 0, you would be in a position to respond to the principal of the school arguing that **varying teachers is associated with variance explained in achievement**, which would stand contrary to the principal's initial claim that regardless of teacher, students achieve to the same degree.

We run the model using the function `lmer` (linear mixed effects models) in the package `lme4` (Bates et al., 2014) specifying **teacher** as a **random effect**. To request maximum likelihood estimation, we include the statement `REML = FALSE` (i.e., by default, `lmer` will run REML):

TABLE 5.1 Achievement as a Function of Teacher

Teacher			
1	2	3	4
70	69	85	95
67	68	86	94
65	70	85	89
75	76	76	94
76	77	75	93
73	75	73	91
$M = 71.00$	$M = 72.5$	$M = 80.0$	$M = 92.67$

```
> library(lme4)
> fit.random <- lmer(ac ~ 1 + (1|f.teach), achiev, REML = FALSE)
```

About the above model specification:

- `~ 1` fits an intercept to the model.
- `(1|f.teach)` specifies `f.teach` as a random factor.
- `achiev` is the name of the dataframe in which the data are contained (i.e., the .txt file we loaded into R).
- `REML = FALSE` tells R to bypass the default estimation method (REML) and to fit the model by maximum likelihood

```
> fit.random
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: ac ~ 1 + (1 | f.teach)
   Data: achiev
     AIC      BIC   logLik deviance df.resid
157.1869 160.7211 -75.5935 151.1869       21
Random effects:
 Groups   Name        Std.Dev.
 f.teach  (Intercept) 8.388
 Residual             4.341
Number of obs: 24, groups:  f.teach, 4
Fixed Effects:
(Intercept)
      79.04
```

Features of the output include the following:

- **AIC** is equal to 157.19, and recall is useful for comparing models. Lower values of AIC indicate a better-fitting model than do larger values. Recall that AIC jointly considers both the goodness-of-fit as well as the number of parameters required to obtain the given fit, essentially "penalizing" for increasing the number of parameters unless they contribute to model fit. If we were to build on the current model by potentially adding terms, then we could observe the extent to which AIC changes and use this in our global assessment of model fit.
- **BIC** yields a value of 160.72, which is also useful for comparing models. Lower values of BIC are also generally indicative of a better-fitting model than are larger values. As was true for AIC, if we were to fit additional parameters to the model, we would want to see a drop in BIC values to justify, on a statistical basis, the addition of the new parameters.
- **Deviance** of 151.19, defined as $-2[\log_e L_{Model} - \log_e L_{Saturated}]$, where L_{Model} is the likelihood of the current model and $L_{Saturated}$ is the likelihood of the saturated model. Here we assume $\log_e L_{Saturated}$ is equal to 0, hence we can also write the deviance as $-2[\log_e L_{Model}]$. Smaller values than not are indicative of better fit.
- The variance component for `f.teach` is equal to the square of the standard deviation. That is, $(8.388)^2 = 70.36$.

- The variance component for residual is equal to the square of the standard deviation. That is, $(4.341)^2 = 18.84$.
- The only fixed effect for this model is the intercept term, and is equal to 79.04. This is the grand mean of achievement for all observations and is not of immediate interest to us.

We could also request a summary of the fitted model (`summary(fit.random)`), which will provide us with similar output as above, with the exception that variance components are included (so we do not have to square the standard deviations ourselves).

5.7.1 Proportion of Variance Accounted for by Teacher

Having fit the model, we can now compute the proportion of variance accounted for by `f.teach`. Recall that the variance component for `f.teach` was equal to 70.36, while the variance component for residual was equal to 18.84. It is important to emphasize that these are **variance components**, they are not **proportions of variance** (that they are not proportions should be evident in itself since proportions range from 0 to 1).

Since $\sigma_y^2 = \sigma_A^2 + \sigma_e^2$, we can compute the estimated proportion of variance accounted for by our independent variable, the **intraclass correlation**, as:

$$\frac{\hat{\sigma}_A^2}{\hat{\sigma}_y^2} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_e^2} = \frac{70.36}{70.36 + 18.84} = \frac{70.36}{89.20} = 0.79$$

That is, approximately 79% (we rounded up) of the variance in achievement is accounted for by teacher.

Of course, this is an extremely large measure of association for data of this kind. If you actually did find such an effect for teacher, what would it suggest? Consistent with our interpretation of the random effects model, it would imply that 79% of students' achievement variance in school is associated with varying teachers, either those teachers selected in the sample or those in the population from which the sampled levels were drawn. Does this mean that one's teacher is somehow **responsible** for one's achievement? Surely not, at least not so based on our statistical analysis.

Still, the finding of 79%, if it were actually true, could serve as a strong counter-argument against that of the principal's who claimed that teacher had no "impact" on students' achievement. Again, we must be cautious with our interpretation, because we certainly have no evidence for anything remotely close to **causal**. The word "impact" is used purposely in quotes here. Concluding that teachers "impact" student achievement implies a directional causal-like claim, and hence must be used with great care, if used at all.

However, such data are still rather strong evidence that changing teachers might be a good idea for Taylor given that she is struggling in school. And the benefit of conducting a random effects model instead of a fixed effects one is that our inferences are not restricted to generalizing to only the levels sampled for the given analysis. We can generalize to the **population** of levels of which the ones featured in the given analysis were merely a random sample. Because you conducted a random effects model, the principal cannot rebuke your evidence by accusing you of "handpicking" certain teachers over others. Your finding of 79% is generalizable to the population of teachers of which the ones you tested were but a random sample. This is what gives random effects their power to draw rather far-reaching generalizations, not unlike when we randomly sample subjects, of which the particular subjects you obtained in your experiment are but a sample of a larger population. Because of the way subject "levels" were sampled, we feel more confident about generalizing to a wider population of subjects.

5.8 R ANALYSIS USING REML

We now fit the one-way random effects model using REML estimation, and briefly compare the output to the previous analysis using ML. To fit by REML, simply exclude the statement `REML = FALSE` from our previous model statement (`fit.random <- lmer(ac ~ 1 + (1|f.teach), achiev, REML = FALSE)`):

```
> fit.random.reml <- lmer(ac ~ 1 + (1|f.teach), achiev)
> summary(fit.random.reml)

REML criterion at convergence: 146.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6056 -0.8696  0.2894  0.7841  1.3893

Random effects:
 Groups   Name        Variance Std.Dev.
 f.teach  (Intercept) 94.87    9.740
 Residual             18.84    4.341
Number of obs: 24, groups:  f.teach, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)    79.04       4.95   15.97
```

We see that the output using REML is very similar to that using ML. The variance components for teacher and residual are 94.87 and 18.84 respectively, for a proportion of variance due to teacher equal to 0.83 (i.e., 94.87/(94.87 + 18.84) = 94.87/113.71 = 0.83), a figure slightly higher than that using maximum likelihood. We could have also obtained the standard deviations by `VarCorr (fit.random.reml)`.

5.9 ANALYSIS IN SPSS: OBTAINING VARIANCE COMPONENTS

We now conduct the identical analysis using SPSS's `VARCOMP` function. We will demonstrate using both maximum likelihood (ML) and restricted maximum likelihood (REML), and briefly compare our results to those obtained using R.

To run the one-way random effects model using ML, we request in SPSS:

```
VARCOMP ac BY teach
 /RANDOM=teach
 /METHOD=ML
```

The remainder of the syntax should include a limit on the number of times you wish the algorithm to iterate (for our example, we have chosen 50), the criteria for convergence (choosing a relatively small number is recommended, or just use the default in SPSS as we have done), and the history of the iteration:

```
/CRITERIA = ITERATE(50)
/CRITERIA = CONVERGE(1.0E-8)
/PRINT = HISTORY(1)
```

Select output from the VARCOMP procedure follows. As we can see, much of it is essentially analogous to that obtained using R (ML).

Iteration History			
Iteration	Log-Likelihood	Var(teach)	Var(Error)
0	−83.415	98.007	89.207
1	−76.353	31.190	18.842
2	−75.593	70.365	18.842
3	−75.593[a]	70.365	18.842

Dependent variable: ac.

Method: maximum-likelihood estimation.

[a]Convergence achieved.

First, we see the **iteration history**, showing the number of times the algorithm took to converge on a log-likelihood statistic having requested convergence criteria (recall our criteria was 1.0E-8). Though the numbers are rounded, we can see that from iteration 2 to iteration 3 the difference between log-likelihood statistics is extremely small (too small to be noticeable in SPSS's report due to rounding, both values are equal to −75.593 in the output). We can also see that SPSS settled on variance components of 70.365 for teach and 18.842 for error. These are the same as those estimated in R.

Next, SPSS reports the variance component estimates that appeared at the last stage of the iteration (i.e., under iteration 3 above):

Variance Estimates	
Component	Estimate
Var(teach)	70.365
Var(Error)	18.842

Dependent variable: ac.

Method: maximum-likelihood estimation.

As we did in the analysis via R, we can compute the **proportion of variance** explained by teacher by 70.365/(70.365 + 18.842) = 70.365/89.207 = 0.79, which is the same figure we obtained in our analysis using R.

We next briefly demonstrate the syntax and output for the same model fit in SPSS, this time fit by REML. To conserve space, only the final variance component estimates are given:

```
VARCOMP ac BY teach
 /RANDOM=teach
 /METHOD=REML [note the change from ML to REML]
/CRITERIA = ITERATE(50)
/CRITERIA = CONVERGE(1.0E-8)
/PRINT = HISTORY(1)
```

Variance Estimates	
Component	Estimate
Var(teach)	94.867
Var(Error)	18.842

Dependent variable: ac.

Method: restricted maximum-likelihood estimation.

Using REML as our method of estimation, we see that teacher accounts for approximately 83% of the variance in achievement (i.e., 94.867/(94.867 + 18.842) = 94.867/113.709). These results parallel those found in R using REML.

5.10 FACTORIAL RANDOM EFFECTS: A TWO-WAY MODEL

Having discussed the one-way random effects model and having come to the conclusion through expected mean squares that the correct error term was indeed MS error, we now turn to consideration of the **two-way random effects model**. In this case, both factors are random, which again implies that the levels for a given experiment are sampled levels from a wider population of levels. As was true for the one-way model, we are not interested specifically in mean differences. Rather, we are interested in variance in the dependent variable attributable to each factor, and potentially also to their interaction.

For example, suppose that instead of merely hypothesizing an association between teacher and achievement, we hypothesize that hours of homework is also related to achievement. However, as was the case for teacher, we are not interested in only **particular** hours of homework (levels), but rather would like to randomly sample a few hours (levels) in an effort to generalize our findings to a **population** of homework hours. Such would designate hours of homework to be a random effect. In this model then, both teacher and homework would be random effects, giving us the two-way random effects model:

$$y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

where μ is the population grand mean, α_j (i.e., a_j, its estimate) is the random variable for row sample effects, β_k (i.e., b_k) is the random variable for column sample effects, $(\alpha\beta)_{jk}$ (i.e., $(ab)_{jk}$) is the random interaction effect for a given cell *jk*, and ε_{ijk} (i.e., e_{ijk}), as before, is the error component, this time for a given individual *i* in a given cell *jk*. Notice that the only part of the model that is not random in the two-way random effects model is the grand mean (Hays, 1994). The rest of the model consists of random variables, including the error component ε_{ijk}.

The assumptions for the two-way random effects model parallel those of the one-way random model, though we now have to generally assume interaction effects, $(\alpha\beta)_{jk}$, to be normally distributed with mean 0 and variance σ^2_{AB}, as well as assuming α_j, β_k, $(\alpha\beta)_{jk}$, and ε_{ijk} are all pairwise independent (Hays, 1994).

In terms of partitioning variability, the arithmetical computations for the two-way analysis of variance under the random effects model are exactly the same as those for the two-way analysis of variance under the fixed effects model. However, as was true for the one-way model, the mean squares will be different. Consequently, this will imply that we construct our *F*-ratios differently than in the fixed effects model. As we will see, and for very good theoretical reasons, the **error term for each factor in the two-way random effects model will be MS interaction**, and no longer MS error. This may seem counterintuitive at first consideration, but our derivation of the EMS will prove our intuition wrong.

We begin by considering the expected mean squares. As was true for the one-way random effects model mean squares, our starting point for considering these for the two-way model begins with recalling features of the fixed effects model. Recall that in the two-way fixed effects model, the row and column effects each summed to 0, that is, $\sum_j a_j = 0$ and $\sum_k b_k = 0$. The interaction effects, $(ab)_{jk}$, also summed to zero across rows, columns, and cells. What this meant is that in the fixed effects models, when considering relevant row and column effects, we did not need to concern ourselves with interaction effects being "picked up" along the way in our computation of row or column effects, since they summed to 0 in each case. The only thing that was being accumulated in our summation was the usual error term, $\bar{e}_j$. For instance, a given row effect a_j could be written as follows:

$$(\bar{y}_j - \bar{y}.)^2 = (a_j + \bar{e}_j - \bar{e})^2 \tag{5.5}$$

The major point of (5.5) is to emphasize that when taking squared deviations from the grand mean in the fixed effects model, the deviation reflects only the fixed effect a_j and mean error (i.e., $\bar{e}_j - \bar{e}$). **Notice that the interaction effect does not contribute to the sums of squares for rows, because the sum of the interaction effects equals 0 in the fixed effects model as we sum across columns**. Or, again, if you prefer, one could say that the interaction effect **is** included in the sum of squares for the fixed effect a_j, but that since it equals 0, it drops out of the fixed effect term. A similar situation applies to columns. There is simply no interaction effect (i.e., the interaction effect will equal to 0) included in the column effect. This is an extremely important point to grasp in order to understand the random effects model under discussion, and the mixed model to be surveyed later. When generating F-ratios for fixed effects, we were not "picking up" interaction variance, and hence had no need to consider interaction in generating suitable F-ratios to test main effects. That is, they did not figure in the expected mean squares.

5.11 FIXED EFFECTS VERSUS RANDOM EFFECTS: A WAY OF CONCEPTUALIZING THEIR DIFFERENCES

As an aside and prior to our development of the two-way model, there is a way to understand the difference between a fixed effect and a random one, and that is in drawing on our knowledge of an "effect" we are already very much familiar, that of e_{ijk}.

Recall that the effects a_j in any given sample will not necessarily equal their long-run expectation in a random effects model. Yes, while it is true that $E(\bar{a}_j) = 0$, when we simply take a random sample from the set of all possible levels, there is no guarantee, theoretically, that a given sample will match that long-run expectation. A similar situation applies for the b_k column effects. Likewise, the sample values for interaction effects $(ab)_{jk}$, because they are now too **random**, do not have to match their expected values in the sample of levels selected for the given experiment.

If you compare this with the behavior of the error term, e_{ijk}, you will notice that the error term behaves in a similar fashion. Yes, the long-run expectation of the error is equal to 0, that is, the mean of the error over an infinite number of repeated samples is expected to be 0. However, in any given experiment, in any given **sampling** of e_{ijk}, there is no reason to suspect that e_{ijk} will equal that long-run expectation. This is why e_{ijk} is quite naturally regarded as a **random effect** (even before we knew what random effects were!). Its "levels" (i.e., the values of e_{ijk} occurring in a given experiment) are randomly sampled from a larger population of potential "levels" (i.e., from a larger population of potential errors).

As we will see, it is this element of randomness of both a_j and b_k that will have influential consequences on ensuing expected mean squares and generation of suitable F-ratios to test effects of interest.

5.12 CONCEPTUALIZING THE TWO-WAY RANDOM EFFECTS MODEL: THE MAKE-UP OF A RANDOMLY CHOSEN OBSERVATION

To explain how things work in a two-way random effects model, we begin with the idea that we have been tracing since our first look at the one-way fixed effects ANOVA in Chapter 3, that of the "make-up" of a given observation for the model under consideration. We again borrow quite heavily from the work of Hays (1994), Kempthorne (1975), Kirk (1995), Searle, Casella, and McCulloch (1992), and Scheffé (1999) in what follows.

For the two-way model, we begin by conceiving that the grand sample mean $\bar{y}_{..}$ will consist of average row effects, column effects, interaction effects, and mean error:

$$\bar{y}_{..} = \bar{a}_{..} + \bar{b}_{..} + \left(\overline{ab}\right) + \bar{e}_{..}$$

The mean $\bar{y}_{j.}$ of any row will consist of the effect of that row, the mean of the column effects (because we are summing across columns), the mean of the interaction effects within that row, and the mean error in that row:

$$\bar{y}_{j.} = a_j + \bar{b}_{..} + \left(\overline{ab}\right)_{j.} + \bar{e}_{j.}$$

That is, notice that to calculate the mean of any row, $\bar{y}_{j.}$, aside from a row effect, a_j (which is what we actually **want** to obtain), we are also "picking up" mean column effects, mean interaction effects, and mean error. As Hays (1994, p. 542) notes, the difference between the row mean and the grand mean (which we want to calculate as usual to get a row effect, $\bar{y}_{j.} - \bar{y}_{..}$), will not include any column effects (we will see that it drops out of the equation), but it does include average interaction effects as well as row effects and error:

$$\begin{aligned} \bar{y}_{j.} - \bar{y}_{..} &= a_j + \bar{b}_{..} + \left(\overline{ab}\right)_{j.} + \bar{e}_{j.} - \bar{a}_{..} - \bar{b}_{..} - \left(\overline{ab}\right)_{..} - \bar{e}_{..} \\ &= \left(a_j - \bar{a}_{..}\right) + \left[\left(\overline{ab}\right)_{j.} - \left(\overline{ab}\right)_{..}\right] + \left(\bar{e}_{j.} - \bar{e}_{..}\right). \end{aligned} \tag{5.6}$$

Notice that when we take deviations from the grand mean, of the form $\bar{y}_{j.} - \bar{y}_{..}$, which by the above is the quantity "$a_j + \bar{b}_{..} + \left(\overline{ab}\right)_{j.} + \bar{e}_{j.}$" minus "$\bar{a}_{..} + \bar{b}_{..} + \left(\overline{ab}\right) + \bar{e}_{..}$," this difference does not include any column effects, because in (5.6), $\bar{b}_{..}$ dropped out of the final solution. It canceled out, since $\bar{b}_{..} - \bar{b}_{..} = 0$. The final solution does, however, contain row effects and interaction effects. That is, to get a row effect $\bar{y}_{j.} - \bar{y}_{..}$, we also get the "unwanted" interaction effects. We will need a way of dealing with these unwanted effects when we build our F-ratio. In the fixed effects models, we did not have to worry about picking up "nuisance effects" (other than error) when computing row or column effects. Why not? Because these nuisance factors did not exist in fixed effects models (or equivalently, they did exist, but were equal to 0).

Similarly, for the deviation of any column mean from the grand mean, we can define a column effect as containing an effect for that particular column, $b_{.k}$, the mean of the row effects, $\bar{a}_{..}$(because we are summing this time across rows), a mean interaction effect, $\left(\overline{ab}\right)_{.k}$, and the mean error in that column, $\bar{e}_{.k}$:

$$\bar{y}_{.k} = b_{.k} + \bar{a}_{..} + \left(\overline{ab}\right)_{.k} + \bar{e}_{.k}$$

Therefore, when we take $\bar{y}_{.k}$ deviations about the grand mean, $\bar{y}_{.k} - \bar{y}_{..}$, we end up with

$$\begin{aligned}\bar{y}_{.k} - \bar{y}_{..} &= b_{.k} + \bar{a}_{..} + \left(\overline{ab}\right)_{.k} + \bar{e}_{.k} - \bar{a}_{..} - \bar{b}_{..} - \left(\overline{ab}\right)_{..} - \bar{e}_{..} \\ &= \left(b_{.k} - \bar{b}_{..}\right) + \left[\left(\overline{ab}\right)_{.k} - \left(\overline{ab}\right)_{..}\right] + \left(\bar{e}_{.k} - \bar{e}_{..}\right) \end{aligned} \tag{5.7}$$

That is, a column deviation from the grand mean contains a column effect, average interaction effects, and average error, but no row effect, because similar to (5.6) when the column effect dropped out of the equation for the row effect, here, the row effect $\bar{a}_{..}$ drops out of the equation. Notice that $\bar{a}_{..} - \bar{a}_{..} = 0$ in (5.7).

In summary then, we need to find a way to produce our F-ratios such that the interaction in the row and column effects is accounted for. As we will see, **for the two-way random effects model, this will call for a test of main effects MS against the interaction term instead of the MS error term as in the fixed effects ANOVA**. To understand why this is so, however, we need to once more consider the expected mean squares.

5.13 SUMS OF SQUARES AND EXPECTED MEAN SQUARES FOR RANDOM EFFECTS: THE CONTAMINATING INFLUENCE OF INTERACTION EFFECTS

Let us see how the interaction involvement of (5.6) and (5.7) will influence the sums of squares for rows in the two-way random effects factorial model. Recall we derived, for the two-way fixed effects model, the effect for row to be

$$\text{SS}\,A = \text{SS between rows} = \sum_j Kn\left(\bar{y}_{j.} - \bar{y}_{..}\right)^2$$

Now, when we substitute $\left(\bar{y}_{j.} - \bar{y}_{..}\right)$ with

$$\left(a_j - \bar{a}_{..}\right) + \left[\left(\overline{ab}\right)_{j.} - \left(\overline{ab}\right)_{..}\right] + \left(\bar{e}_{j.} - \bar{e}_{..}\right)$$

of (5.6), we obtain

$$\text{SS}\,A = \text{SS between rows} = \sum_j Kn\left\{\left(a_j - \bar{a}_{..}\right) + \left[\left(\overline{ab}\right)_{j.} - \left(\overline{ab}\right)_{..}\right] + \left(\bar{e}_{j.} - \bar{e}_{..}\right)\right\}^2 \tag{5.8}$$

which we can now reduce to, in terms of expected mean squares:

$$E\left(\text{MS}\,A\right) = E\left(\text{MS between rows}\right) = Kn\sigma_A^2 + n\sigma_{AB}^2 + \sigma_e^2 \tag{5.9}$$

We notice (5.9) contains the interaction term, $n\sigma_{AB}^2$. What this means is that when we consider the construction of a suitable F-ratio to isolate σ_A^2, we are going to need a denominator that includes $n\sigma_{AB}^2$ so that we can account for it being a part of the numerator of our F-test. Likewise, for factor B (columns), in terms of EMS, we have:

$$E\left(\text{MS between columns}\right) = Jn\sigma_B^2 + n\sigma_{AB}^2 + \sigma_e^2 \tag{5.10}$$

Again, the term $n\sigma_{AB}^2$ appears in (5.10), whereas in the fixed effects model, this term did not appear (or, again, if you like, it did appear, but was equal to 0). Analogous to our test of the row effect, this will call for a different F-ratio for testing the column effect than what we had in the fixed effects model. In the fixed effects model of the previous chapter, we simply did not have to deal with the "contamination" of $n\sigma_{AB}^2$.

Finally, the expectation mean squares for the interaction term ends up being $n\sigma_{AB}^2 + \sigma_e^2$, and as usual, the expectation for MS error is σ_e^2. See Searle, Casella, and McCulloch (1992) for how this expectation is obtained.

5.13.1 Testing Null Hypotheses

As was true for the one-way random effects model, the null for factor A is given by $H_0 : \sigma_A^2 = 0$. This null hypothesis, if "true," would imply that $Kn\sigma_A^2 = 0,$ and so all that is left from the expected mean squares is

$$Kn\sigma_A^2 + n\sigma_{AB}^2 + \sigma_e^2$$
$$0 + n\sigma_{AB}^2 + \sigma_e^2$$

What if we naively decided to use good 'ol MS error as our error term for testing this effect? Under the null hypothesis that $\sigma_A^2 = 0$, we would have:

$$\frac{n\sigma_{AB}^2 + \sigma_e^2}{\sigma_e^2}$$

Notice that had we used MS error, we would still have an interaction term unaccounted for in the numerator, which would mean that even if there are no effects for factor A, we might still obtain an F appreciably greater than 1. This would be because interaction variance, $n\sigma_{AB}^2$ is making its way into the numerator and we are not effectively isolating $Kn\sigma_A^2$. Therefore, this calls for us to use a new error term to test the main effect for such a random effect. Which error term shall we choose to "get rid of" $n\sigma_{AB}^2 + \sigma_e^2$? We notice that this term is actually the **mean square for interaction**, since recall that this is what we found the expectation for interaction to be.

Now, everything should be beginning to fall into place. The test for factor A must be against MS interaction as it allows us to isolate the effect of interest in the numerator:

$$F = \frac{\text{MS A}}{\text{MS A} \times \text{B interaction}} = \frac{Kn\sigma_A^2 + n\sigma_{AB}^2 + \sigma_e^2}{n\sigma_{AB}^2 + \sigma_e^2}$$

We lose a degree of freedom for row and one for column, so the degrees of freedom on which the above F will be tested are equal to $(J-1)$ and $(J-1)(K-1)$.

Likewise, for factor B, to evaluate the null hypothesis $H_0 : \sigma_B^2 = 0$, since there is interaction variance again "contaminating" the effect, $E(\text{MS between columns}) = Jn\sigma_B^2 + n\sigma_{AB}^2 + \sigma_e^2$, the appropriate denominator for testing this effect (on $(K-1)$ and $(J-1)(K-1)$ degrees of freedom) is once more $n\sigma_{AB}^2 + \sigma_e^2$:

$$F = \frac{\text{MS B}}{\text{MS A} \times \text{B interaction}} = \frac{Jn\sigma_B^2 + n\sigma_{AB}^2 + \sigma_e^2}{n\sigma_{AB}^2 + \sigma_e^2}$$

If $H_1 : \sigma_B^2 > 0$, then the term $Jn\sigma_B^2$ will reflect this effect, and the F-statistic will be appreciably greater than 1.0. Otherwise, we will be left with simply

$$\begin{aligned} F &= \frac{Jn\sigma_B^2 + n\sigma_{AB}^2 + \sigma_e^2}{n\sigma_{AB}^2 + \sigma_e^2} \\ &= \frac{0 + n\sigma_{AB}^2 + \sigma_e^2}{n\sigma_{AB}^2 + \sigma_e^2} \\ &= \frac{n\sigma_{AB}^2 + \sigma_e^2}{n\sigma_{AB}^2 + \sigma_e^2} \end{aligned}$$

and our expectation for F would be approximately 1.0 under the null hypothesis $\sigma_B^2 = 0$.

What is the appropriate denominator for testing $H_0 : \sigma_{AB}^2 = 0$? This one is easy. Since we found the expected mean squares to be $n\sigma_{AB}^2 + \sigma_e^2$, it is quite evident that the correct denominator in this case actually is **MS error**, evaluated on $(J-1)(K-1)$ and $JK(n-1)$ degrees of freedom. That is,

$$F = \frac{\text{MS interaction}}{\text{MS error}} = \frac{n\sigma_{AB}^2 + \sigma_e^2}{\sigma_e^2}$$

In summary then, we have found that in the two-way random effects model, both random effects are to be tested against MS interaction, while the interaction term is to be tested against MS error.

5.14 YOU GET WHAT YOU GO IN WITH: THE IMPORTANCE OF MODEL ASSUMPTIONS AND MODEL SELECTION

Even if you should never venture into models with random effects (other than, of course, the **error term** in a fixed effects model, which is virtually always present), a survey of random effects is pedagogically instructive because it serves to illustrate that the conclusions one draws from an analysis of data are very much contingent on the **assumptions** and **sampling** one enters with into the model-building process. The actual arithmetic of the ANOVA may very well be the same in many cases, but **the construction of F-ratios will differ based on the assumptions you make at the very beginning of your experiment**. We summarize this idea with the following:

> **If you use a fixed effects model, when really, you are interested in interpreting a random effects model, you will be restricted to making inferences only about the levels of the independent variable that are present in your experiment. Your substantive conclusions are intimately tied to the model you have tested.**

There are many research papers across the sciences where researchers, after conducting a fixed effects analysis of variance, regularly, and perhaps inadvertently, generalize their findings to levels of the independent variable(s) not tested in the model. As emphasized by Searle, Casella, and McCulloch (1992, p. 22), "**Users of computer packages that have *F*-values among their output must be totally certain that they know precisely what the hypothesis is that can be tested by each such *F*-value**."

Let us shed a bit more perspective on Searle et al.'s warning. Consider the following scenario: As a researcher in sensation and perception, suppose you are interested in the variability explained in pupil size (i.e., dependent variable) when looking at various playing cards. If you select two playing cards, say a king of spades and a jack of hearts, measure pupil size, and find there is a statistically significant difference between pupil size for king of spades versus pupil size for jack of hearts, under the fixed effects ANOVA, you will only be able to conclude mean differences for **these two card-types only**, since you are assuming that in replications of the experiment, only these two cards would be used again and again. Now, had you used a random effects model, **and randomly sampled these two cards from the deck**, you could have concluded that differences in cards, either those selected randomly for the given experiment or those in the population of potential cards that could have been selected, accounts for a given amount of variance in pupil size. That is, you would be able to make a more **general** statement in the random effects model. You would be able to say something about playing cards **in general**, rather than just the two kinds you selected.

As a general guideline, when you interpret an ANOVA, always ask yourself whether the investigator is assuming a fixed or random effects model, and then critically evaluate whether the data were analyzed and interpreted in correspondence with these assumptions. Be sure to verify whether conclusions outlined in results and discussion sections agree with the model actually analyzed. If they line up, then great. If they do not, then at least you will have a sense of the limitations imposed by the analysis in relation to the potentially much more broad conclusions drawn in the discussion of the paper. **Researchers often like to overstate conclusions in discussion sections despite the fact that their statistical analyses do not support such conclusions**.

5.15 MIXED MODEL ANALYSIS OF VARIANCE: INCORPORATING FIXED AND RANDOM EFFECTS

Suppose that instead of merely wanting to demonstrate that teacher is associated with variance in achievement, you also wanted to show that the lesson plan used by the teacher is also associated with achievement. Suppose you were interested in specifically comparing five different lesson plans. Hence, teacher remains **random**, but lesson is now **fixed**. When we have a mix of fixed and random factors, we have the **mixed model analysis of variance**. Pinheiro and Bates (2000) do a nice job of summarizing the applied rationale of a mixed model:

> Mixed-effects models are primarily used to describe relationships between a response variable and some covariates in data that are grouped according to one or more classification factors. Examples of such **grouped data** include **longitudinal data**, **repeated measures data**, **multilevel data**, and **block designs**. By associating common random effects to observations sharing the same level of a classification factor, mixed-effects models flexibly represent the covariance structure induced by the grouping of the data. (p. 3)

Purely random effects models are relatively rare. Fixed effects models are much more common across the social, economic, and medical sciences. However, a study of random effects such as we have undergone is quite useful, not only because it provides an understanding of the random effects model itself, but also because it serves as a “bridge” to the mixed model, which is quite popular.

As we did for both the fixed effects and random effects models, we consider the expected mean squares for the mixed model. When we obtain effects for the fixed factor, we will need to sum across a **random factor**. Just as we summed across random factors in the two-way random effects model, we will once again conclude that this factor (i.e., the **fixed** one, not the random one) be tested against **MS interaction** and not MS error.

To help better understand the denominators we will use for testing fixed and random effects, consider the layout in Table 5.2. In this layout, the fixed factor, represented by rows, has six levels, and the random factor, represented by columns, has three levels.

TABLE 5.2 Cell Layout for 6 × 3 Mixed Model Analysis of Variance

		Random Factor (B)			
		I	II	III	Row Means
Fixed Factor (A)	I	y_{ijk}	y_{ijk}	y_{ijk}	$\bar{y}_{j.}$
	II	y_{ijk}	y_{ijk}	y_{ijk}	$\bar{y}_{j.}$
	III	y_{ijk}	y_{ijk}	y_{ijk}	$\bar{y}_{j.}$
	IV	y_{ijk}	y_{ijk}	y_{ijk}	$\bar{y}_{j.}$
	V	y_{ijk}	y_{ijk}	y_{ijk}	$\bar{y}_{j.}$
	VI	y_{ijk}	y_{ijk}	y_{ijk}	$\bar{y}_{j.}$
	Column means	$\bar{y}_{.k}$	$\bar{y}_{.k}$	$\bar{y}_{.k}$	$\bar{y}_{..}$

In the layout of Table 5.2, we will have the following effects for the fixed factor and random factor:

- **Row effects**, denoted by $\bar{y}_{j.} - \bar{y}_{..}$, represent the effect of being in one row versus being in other rows on levels of the fixed factor.
- **Column effects**, denoted by $\bar{y}_{.k} - \bar{y}_{..}$, represent the effect of being in one column versus being in other columns on levels of the random factor.

The questions we need to ask ourselves about Table 5.2 are the following:

- What kind of information went into producing the row effects, $\bar{y}_{j.} - \bar{y}_{..}$? Notice that to get these row effects, we need to sum across a **random** factor. How will this summing across a random factor impact the makeup of the given row effect?
- What kind of information went into producing the column effects, $\bar{y}_{.k} - \bar{y}_{..}$? Notice that to get these column effects, we need to sum across a **fixed** factor. How will this summing across a fixed factor impact the makeup of the given column effect?

To get a given row effect, $\bar{y}_{j.} - \bar{y}_{..}$, because we are needing to sum across a random effect, we have every reason to believe that **the sum of interaction effects,** $(ab)_{jk}$**, will not equal to 0** (Hays, 1994). Hence, we will need to account for this source of variation when constructing our F-ratio. That is, within any row of the fixed effect, we can expect there to be an average interaction effect, unequal to zero (and possibly different from row to row), that we are "picking up" as we sum across the given row. These row totals then, and their corresponding effects, will not only reflect row effects, but rather will also be reflective of average **interaction effects**. To the contrary, to get a given column effect, $\bar{y}_{.k} - \bar{y}_{..}$, because we are summing across a **fixed** effect, we have good reason to believe that the sum of interaction effects, $(ab)_{jk}$, **will** equal to 0. Hence, we do not need to account for this source of variation when constructing our F-ratio (or equivalently, we can account for it, but it will be equal to zero each time).

How are the expected mean squares impacted by all this? For the fixed effect, factor A, EMS is equal to

$$E(\text{MS A}) = \sigma_e^2 + n\sigma_{AB}^2 + \frac{Kn\left(\sum_j \alpha_{j.}^2\right)}{J-1}$$

Notice that included in this EMS is interaction variance, $n\sigma^2_{AB}$, which is **unwanted**. For the random effect, EMS is equal to

$$E(\text{MS B}) = \sigma^2_e + Jn\sigma^2_B$$

Notice that the only **unwanted** variation in this EMS is that of σ^2_e. The EMS for the interaction term ends up being, quite simply

$$E(\text{MS AB}) = \sigma^2_e + n\sigma^2_{AB}$$

We now have all the information necessary to build our F-ratios. For the fixed effect, under the null hypothesis of no effect, we get

$$\begin{aligned} E(\text{MS A}) &= \sigma^2_e + n\sigma^2_{AB} + \frac{Kn\left(\sum_j \alpha^2_{j.}\right)}{J-1} \\ &= \sigma^2_e + n\sigma^2_{AB} + \frac{Kn\left(\sum_j (0)^2_{j.}\right)}{J-1} \\ &= \sigma^2_e + n\sigma^2_{AB} \end{aligned}$$

which suggests that the correct denominator for testing the fixed effect must be MS interaction:

$$F = \frac{\text{MS A}}{\text{MS AB}} = \frac{\sigma^2_e + n\sigma^2_{AB} + \frac{Kn\left(\sum_j (0)^2_{j.}\right)}{J-1}}{\sigma^2_e + n\sigma^2_{AB}}$$

Notice that it is the fixed factor (not the random factor) that is tested against the interaction term in the mixed model.

Under the hypothesis of no column effect (random factor), $\sigma^2_B = 0$, since $E(\text{MS B}) = Jn\sigma^2_B + \sigma^2_e$ we end up with simply σ^2_e. Thus, the F-ratio for the random factor is given by

$$F = \frac{Jn\sigma^2_{\text{B}} + \sigma^2_e}{\sigma^2_e}$$

Notice that it is the random factor (not the fixed factor) that is tested against MS error in the mixed model.

As a recap of what we have done, we have seen that in a two-way mixed model, to produce the F-test for the random effect, we divide by MS error. The reason for this is that to produce the column means, we have to sum across the fixed factor. Those respective sums are not expected to contain anything but variability due to levels of the random factor along with error.

For the fixed effect, however, what went into the sums for rows? That is, when we produce the sum (or the mean) for each row (fixed effect, in our layout), what kind of variability went into each of these row sums? There is surely (hopefully) variability due to the effect of being in that particular row and not other rows, and there is variability due to error, as usual. But, there is another source of variability, and that is interaction variance. Why? Because when we tally up the cell totals for a level of the fixed factor, we are summing across only a **sample** of possible levels of the random factor. Hence, if we were to do the experiment over, and presumably sampled different levels of the random factor, the effect we would obtain for the given level of the fixed effect might change by the very nature of summing across the random factor in question. Hence, we have "unwanted" interaction variance in the rows and have to account for this when generating the corresponding *F*-ratio. If we produced our *F*-ratio by dividing by MS error, we would still have an interaction effect left over in the numerator, and thus we would have failed to isolate the effect of interest (i.e., row effect). We would have failed to test our null hypothesis of interest.

5.15.1 Mixed Model in R

Having laid out some of the theory for mixed models, we now estimate a mixed model on the achievement data, this time specifying textbook as a fixed factor and teacher as a random effect (Table 4.1). Of course, there is much more to the fitting of a mixed model than shown here (e.g., plots, diagnostics to verify assumptions, etc.). Our purpose here is only to briefly demonstrate how such a model can be fit in R.

We use the package `nlme` (Pinheiro et al., 2014), and fit our model using REML (partial output shown below):

```
> library(nlme)
> mixed <- lme(ac ~ f.text, data = achiev, random = ~1 | f.teach)
> summary(mixed)

Random effects:
 Formula: ~1 | f.teach
        (Intercept) Residual
StdDev:    9.733736 4.423571

Fixed effects: ac ~ f.text
               Value Std.Error DF   t-value p-value
(Intercept) 78.58333  5.031607 19 15.617940  0.0000
f.text2      0.91667  1.805915 19  0.507591  0.6176
```

In the code, `random = ~1 | f.teach)` designates the random effect. The coefficient for `f.text2` is a mean contrast between the first and second textbooks (i.e., 79.50 – 78.58 = 0.92). The effect for textbook is not statistically significant ($p = 0.6176$). The variance component for `f.teach` is equal to the square of 9.73, which is 94.67. Since the square of the residual is equal to 19.57, the proportion of variance accounted for by `f.teach` is 94.67/(94.67 + 19.57) = 94.67/114.24 = 0.83 (rounded up from 0.829). Confidence intervals for effects can also be obtained via `intervals(mixed)`.

5.16 MIXED MODELS IN MATRICES

Having briefly introduced the mixed model for the simplest case, we now briefly consider the mixed model in its most general matrix form:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{ZU} + \mathbf{E} \tag{5.11}$$

where, **Y** is a response matrix, **X** is a model matrix associated with the fixed effects in **B**, **B** is a vector of parameters corresponding to the fixed effects, **Z** is the model matrix associated with the random effects in **U**, and **E** is a vector of errors, what is left over from the model after prediction of **Y**. We assume that $\mathbf{U} \sim N(\mathbf{0}, \mathbf{\Sigma}_z)$ and $\mathbf{E} \sim N(\mathbf{0}, \mathbf{\Sigma}_\varepsilon)$, where $\mathbf{\Sigma}_z$ is the covariance matrix for the random effects and $\mathbf{\Sigma}_\varepsilon$ is the covariance matrix for the errors contained in **E**. This formulation of the model often goes by the name of the **Laird-Ware form**, after the seminal paper "**Random-Effects Models for Longitudinal Data"** (Laird and Ware, 1982) that provided the very general form of the mixed model. Because of **Y**, the model in (5.11) can also accommodate more than a single response variable, giving us the **multivariate mixed model** (Timm, 2002), of which all other univariate mixed models can be considered special cases.

5.17 MULTILEVEL MODELING AS A SPECIAL CASE OF THE MIXED MODEL: INCORPORATING NESTING AND CLUSTERING

Our study of the mixed model lends itself well to introducing a class of modeling methodologies that is increasing in popularity in the social and natural sciences, that of **multilevel** or **hierarchical modeling**. As we discuss in the chapter to follow, mixed models are also useful for addressing problems of **repeated measurements**, which usually can also be conceptualized as having a "multilevel" or "hierarchical" structure.

The topic of multilevel modeling is beyond the scope of this book. Our goal here is to simply conclude this chapter with a **foot-in-the-door** commentary as to how these models can be conceptualized as a special case of the more general mixed model. Indeed, as Pinheiro and Bates (2000) note:

> This model with two sources of variation, b_i and ε_{ij}, is sometimes called a **hierarchical** model … or a multilevel model. The b_i are called **random** effects because they are associated with the particular experimental units […] that are selected at random from the population of interest. They are **effects** because they represent a deviation from an overall mean. …Because observations made on the same [level of the independent variable] share the same random effect b_i, they are correlated. The covariance between observations on the same [level] is σ_b^2 corresponding to the correlation of $\sigma_b^2/\left(\sigma_b^2+\sigma^2\right)$. (p. 8)

To properly discuss the multilevel model, it helps first to recall where we have been. Recall the one-way fixed effects analysis of variance model of Chapter 3:

$$y_{ij} = \overline{y} + a_j + e_{ij}$$

In this model, we assumed the treatment effects a_j to be fixed and e_{ij} to be random and normally distributed. In specifying a_j as fixed, it implied that we were only interested in mean differences as represented by the factor levels actually included in the given experiment. If we were interested in the population of levels of which the ones showing up in our experiment constituted a random sample, then we specified a_j as random, and had the one-way random effects model, which is the same as the fixed effects model, only that now, sample effects are considered randomly sampled from a larger population.

This type of model in which we allow a_j to be random instead of fixed can, in many cases, actually be conceived as a very simple version of what is known as the **multilevel** or **hierarchical** model. What are the levels of the "hierarchy?" The observations y_{ij} constitute **level 1**, and the "grouping" random treatment effect a_j constitutes **level 2**. We say that observations y_{ij} are **nested** within level 2.

For instance, suppose that in our achievement example, instead of randomly assigning students to teacher, we simply sampled students **as they were**, and **as already associated with a given teacher**. In such a case, school children y_{ij} would be considered **nested** within teacher. If we then randomly sampled a number of teachers (say, four, as in our previous example), but wished to generalize to a wider population of teachers, then teacher becomes a random effect. But how is this also a multilevel or hierarchical model? Such models emphasize the fact that observations often occur in a **natural hierarchy** or as a result of one being imposed through a sampling plan (such as **blocking**). For our student observations, there is expected to be a **likeness** about students who share the same teacher. **Observations "within teacher" are more likely to be similar than observations between teachers, not necessarily because of any external treatment condition imposed, but simply because these students share the same teacher**. And having the same teacher means they share the same teaching style, etc., and all of the other infinite innumerable (and potentially even **immeasurable**) elements that may be related to sharing the same teacher. And though there is nothing technically inherent in the definition of "multilevel modeling" that prevents us from designating all effects as **fixed effects** (e.g., studying and generalizing to mean differences between teachers), when we speak of multilevel or hierarchical models, we are usually implicitly invoking the idea that we have one or more **random effects**. For our example, we are usually interested in generalizing to more teachers than we have sampled for our study, making it, as we have seen, a random effect.

Our point is that **multilevel structures** are often analyzed via mixed models. There is nothing inherent in such a hierarchical structure that "demands" such data be analyzed as such, but for reasons of both wanting to account for **likeness** of observations within levels of the hierarchy as well as generalizing to levels of the treatment effect, these typically necessitate the use of such models. For a classic introduction to multilevel and hierarchical data, see Raudenbush and Bryk (2002). Snijders and Bosker (1999) also provide a very readable treatment.

5.18 CHAPTER SUMMARY AND HIGHLIGHTS

- In the traditional **fixed effects model**, the specific levels of the independent variable(s) chosen by the experimenter are of interest, and population inferences are made about those, and only those, levels used in the experiment. Null hypotheses are tested of the sort $H_0 : \mu_1 = \mu_2 = \mu_3 \cdots = \mu_J$.
- In the **random effects model**, the experimenter is not interested specifically in the levels chosen for the particular experiment. Instead, the levels chosen are merely regarded as a **random sample of potential levels** that could have been chosen. The experimenter is interested in testing a null hypothesis that the variance in the dependent variable accounted for by the given factor is equal to 0, that is, $H_0 : \sigma_A^2 = 0$.
- The conceiving of **sample effects** as **random** rather than **fixed** has important implications for the construction of F-ratios.
- In the one-way random effects model, **MS error** is a suitable **error term** for constructing the F-ratio for a test of the random effect. **Variance components** may be estimated using ANOVA estimation, ML, or REML. REML is often the estimator of choice in random effects and mixed models.
- In the two-way random effects model, because each effect is computed by summing across a random effect, the expected mean squares dictate **MS interaction** to be the correct error term for each effect in the generation of F-ratios.
- When a model has a mixture of fixed and random effects (in addition, naturally, to the error term), the model is a **mixed model**. EMS for a two-way mixed model reveals that it is the **fixed effect** that is tested against **MS interaction**. The random effect is tested against MS error.

- Understanding that ε_{ijk} is always a **random effect**, whether in fixed effects, random effects, or mixed models, helps one to better appreciate the nature of random effects in general, realizing that their behavior will be governed by similar random processes as is true of ε_{ijk}.
- An understanding of basic mixed model theory coupled with the idea of **nesting structures** lends itself to conceiving the **multilevel** or **hierarchical linear model**.
- **Random effects** and **mixed models** can be fit in R using `lme4` or `nlme`. SPSS's `VARCOMP` can also be used to estimate variance components.

REVIEW EXERCISES

5.1. Discuss why a researcher may wish to conduct a **random effects analysis of variance** instead of a **fixed effects** ANOVA.

5.2. Elaborate on the statement "**Random effects ANOVA is not about means, it is about variances**."

5.3. Distinguish between a **random effects** model and a **mixed effects** model.

5.4. Give an example of three research scenarios that would necessitate the fitting of a **random effects** model.

5.5. Give an example of three research scenarios that would necessitate the fitting of a **mixed effects** model.

5.6. Distinguish the **assumptions** for a one-way **fixed effects** model from those of a one-way **random effects** model. How are they similar? Different?

5.7. How are a_j and e_{ij} similar in a **random effects** model but different in a **fixed effects** model?

5.8. How can it be said that, technically, **virtually all ANOVA models are either random effects or mixed models**, and that purely fixed effects models rarely exist?

5.9. What are three common ways of estimating parameters in a **random effects model**?

5.10. What is the **expected mean squares** for the **random factor** in a one-way random effects model? What implication does this EMS have on the construction of the corresponding F-ratio?

5.11. How does the **null hypothesis** for a one-way **random effects** model differ from that of a one-way fixed effects model?

5.12. Given the F-ratio for a one-way **random effects model**, what is the expectation for F under the null hypothesis, and why?

5.13. Define the **intraclass correlation** coefficient, its meaning, and its purpose.

5.14. In the chapter example of **achievement as a function of teacher**, explain how the interpretation of findings would have changed had teacher been regarded as a **fixed effect** rather than a random one. Would this have helped or hindered the cause of the parent in responding to the principal's claim? How so?

5.15. Consider the following hypothetical data in Table 5.3 on **factor A** (three levels) and **factor B** (six levels). Factor A is a **fixed factor** while factor B is a **random factor**. Within each cell is a single observation.

TABLE 5.3 Cell Layout of Data on Factors A and B

	Factor A (F)			
Factor B (R)	1	2	3	Means
1	11	27	57	31.67
2	12	29	45	28.67
3	14	31	65	36.67
4	16	26	95	45.67
5	51	36	54	47.00
6	24	35	46	35.00
Means	21.33	30.67	60.33	37.44

Estimate a two-way mixed model in R using REML. How much variance is accounted for by the random effect?