Notes on Maxwell & Delaney
(updated on 26-November-2019)

PSY710

10 Random or Nested Factors

10.1 Random Effects

So far in this course we have been using statistical methods that are appropriate for cases in which the experimenter is interested in knowing the effects of *the particular levels* of the independent variables included in the experiment. Such factors are called **fixed** because the same levels would be used in replications of the experiment, and the models used to analyzed such data are called fixed-effects models. However, in other experimental situations it may not be possible or appropriate to include all levels of a variable in a single experiment, and therefore we may be forced to select a subset, or sample, of the population of levels for inclusion in our experiment. The particular levels of the factor therefore are likely to vary across replications of the experiment. Factors in which levels have been selected randomly are called **random** factors, and the statistical models used to analyze experiments that use such factors are called *random-effects models*. If an experiment contains both fixed and random factors, then the data are analyzed with a so-called *mixed model*.

Why do we need a different kind of model to analyze experiments that include random factors? In fixed-effects models, the only source of *random* variation among scores is error variance. However, experiments that include random factors have multiple sources of variability, or variance components: in addition to error variance, the effects associated with random factors are conceptualized as random variables drawn from a population of effects. As we shall see, these extra sources of variation alter the way we evaluate main effects and interactions. Furthermore, the *goal* of evaluating random effects differs slightly from the goal of analyzing fixed effects. In the case of fixed-effects models, conclusions reached about significance (or lack thereof) of main effects and interactions are restricted to the particular levels of the factors that were included in the experiment. Random-effects models, on the other hand, enable us to generalize our conclusions to the entire *population* of levels of the random factor, even ones not included in our experiment.

10.2 one-factor case

Data from an experiment that uses a single random factor are analyzed by comparing the following two nested models:

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij} \quad (1)$$

$$Y_{ij} = \mu + \epsilon_{ij} \quad (2)$$

As before, the errors are assumed to be independent and normally distributed with a zero mean and a variance of σ_ϵ^2. Note that these models look identical to the ones used to evaluate one-factor, fixed effects experiments. However, in a random effects model the α's are assumed to be random variables selected from a normal distribution with a mean of zero and a variance of σ_α^2. Furthermore, α_j and ϵ_{ij} are assumed to be independent. With these assumptions, the expected value of the scores is μ, and the variance of the Y_{ij}'s is the sum of two **variance components**:

$$\mathrm{Var}(Y_{ij}) = \sigma_\alpha^2 + \sigma_\epsilon^2 \quad (3)$$

The random effects models in Equations 1 and 2 can be used to evaluate the following null and alternative hypotheses:

$$\text{H0}: \sigma_\alpha^2 = 0 \tag{4}$$
$$\text{H1}: \sigma_\alpha^2 > 0 \tag{5}$$

Note how these hypotheses differ from those evaluated by fixed-effects models. Also note that the conclusions drawn from the analysis apply to all possible levels of the random factor, not just the ones used in the current experiment. Despite these differences, the actual statistical test of a random effect in a one-factor design is equivalent to the test used to evaluate a single fixed factor.

As for the fixed-effects case, the expected value of the mean-square residuals, or mean-square within group, for the full model is

$$\xi\left[\frac{E_F}{df_F}\right] = \xi\left[\text{MS}_{WG}\right] = \sigma_\epsilon^2, \tag{6}$$

where E_F is the sum of squared residuals from the full model, and df_F – the degrees of freedom – equals the total number of scores minus the number of groups ($df = N - a$). Equation 6 sometimes is referred to as the ANOVA estimate of the variance component σ_ϵ^2. Equation 6 is true regardless of whether the null hypothesis is true or false, but the between-group variation does depend on the status of the null hypothesis. When it is true, $\sigma_\alpha^2 = 0$, error variance is the only source of variation among the scores, and therefore the expected value of the mean-square *between* groups equals the expected value of MS_{WG}:

$$\xi\left[\text{MS}_{BG}\right] = \xi\left[\text{MS}_{WG}\right] = \sigma_\epsilon^2 \tag{7}$$

When the null hypothesis is false, the variation among scores reflects σ_α^2 and σ_ϵ^2, and the expected value of MS_{BG} is

$$\xi\left[\text{MS}_{BG}\right] = \xi\left[\frac{E_R - E_F}{df_R - df_F}\right] = \sigma_\epsilon^2 + n'\sigma_\alpha^2 \tag{8}$$

where E_R is the sum of squared residuals from the reduced model, df_R equals the total number of scores minus one ($df = N - 1$), and

$$n' = [1/(a-1)]\left[\sum n_j - (\sum n_j^2 / \sum n_j)\right] \tag{9}$$

When there are an equal number of subjects per group, $n' = n$. When there are unequal numbers of subjects per group, $0 < n' < \bar{n}$.

When the null hypothesis is true, $\sigma_\alpha^2 = 0$, $\text{MS}_{BG} \approx \text{MS}_{WG}$, and the expected value of $\text{MS}_{BG}/\text{MS}_{WG}$ is one. In this case, the ratio $\text{MS}_{BG}/\text{MS}_{WG}$ is a random variable that is distributed as F with $a - 1$ and $N - a$ degrees of freedom. However, when the null hypothesis is false, $\sigma_\alpha^2 > 0$ and the expected value of $\text{MS}_{BG}/\text{MS}_{WG}$ increases. This fact forms the basis of the null hypothesis test: we assume that H0 is true, and therefore that the ratio of mean-squares is distributed as F with $a - 1$ and $N - a$ degrees of freedom. If the observed ratio is unusually large – i.e., if the probability of obtaining an F that is at least as large as the observed value is $\leq .05$ or .01 – then we reject the null hypothesis in favor of the alternative.

10.2.1 strength of association & variance components

When dealing with a random factor, the proper measure of association strength is the **intraclass correlation**, which is denoted as ρ_I. The intraclass correlation represents the proportion of population variance accounted for by a random effect:

$$\rho_I = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2} \tag{10}$$

When there are equal n per group, the expected value of MS_{BG} is $\sigma_\epsilon^2 + n\sigma_\alpha^2$. Taking advantage of the fact that $\xi(\text{MS}_{WG}) = \sigma_\epsilon^2$, it is possible to show that

$$\hat{\sigma}_\alpha^2 = (1/n) \times (\mathrm{MS}_{BG} - \mathrm{MS}_{WG}) \tag{11}$$

Equation 11 sometimes is referred to as the ANOVA estimate of the *variance component* σ_α^2. Using this last equation, we can show that an unbiased estimate of the intraclass correlation is given by

$$\hat{\rho}_I = \frac{\mathrm{MS}_{BG} - \mathrm{MS}_{WG}}{\mathrm{MS}_{BG} + (n-1)\mathrm{MS}_{WG}} \tag{12}$$

or, equivalently,

$$\hat{\rho}_I = \frac{F_{BG} - 1}{(n-1) + F_{BG}} \tag{13}$$

Notice that $\hat{\rho}_I < 0$ when $F_{BG} < 1$. On such occasions, it is standard practice to set $\hat{\rho}_I$ to zero[1].

10.2.2 power

To estimate power, we start by defining

$$f_{rand} = \frac{\sigma_\alpha}{\sigma_\epsilon} \tag{14}$$

f_{rand} is analogous to Cohen's f, which is a measure of effect size for fixed effects. It is possible to show that the ratio of the expected values of the between- and within-group means squares is

$$\frac{\xi(\mathrm{MS}_{BG})}{\xi(\mathrm{MS}_{WG})} = (1 + nf_{rand}^2)$$

where n is the number of subjects per group. (N.B. I will consider only the case where there are equal n per group.)

When the null hypothesis is true, the ratio of between- and within-group mean squares will be distributed as F with $a-1$ and $a(n-1)$ degrees of freedom, where a is the number of groups. Let us define $F_{critical}$ as the value of F that would lead to a rejection of the null hypothesis for a given level of α. In R, $F_{critical}$ can be calculated using the command

```
> F.crit <- qf(1-alpha, df1=a-1, df2=a*(n-1))
```

Power refers to the probability that the observed value of F, or F_{obs}, exceeds $F_{critical}$ when the null hypothesis is false. When the null hypothesis is false, F_{obs} is a random variable that is distributed as a multiple of a regular F statistic that has $a-1$ and $a(n-1)$ degrees of freedom. Specifically,

$$P[F_{obs} \geq X] = P\left[\frac{\xi(\mathrm{MS}_{BG})}{\xi(\mathrm{MS}_{WG})} F_{a-1,\ a(n-1)} \geq X\right] = P[(1 + nf_{rand}^2)F_{a-1,\ a(n-1)} \geq X]$$

If we let $X = F_{critical}$, power can be calculated as

$$\text{power} = P\left[(1 + nf_{rand}^2)F_{a-1,\ a(n-1)} \geq F_{critical}\right]$$

and therefore

$$\text{power} = P\left[F_{a-1,\ a(n-1)} \geq F_{critical}/(1 + nf_{rand}^2)\right]$$

In R, this probability can be calculated with the `pf()` command:

```
> k <- ( 1+n*(f.rand^2) )
> the.power <- 1-pf(F.crit/k, df1=a-1, df2=a*(n-1))
```

[1] In some circumstances it may be better to keep the negative value. For example, if you are averaging values across experiments then setting negative values to zero would introduce a bias and therefore it might be best to include the negative values in the average.

10.2.3 linear contrasts

Given the fact that levels of a random factor are selected randomly, and given the nature of the null and alternative hypotheses that are evaluated with a random-effects model, it is not clear if it is meaningful to compare individual levels of a random factor. Certainly it is hard to imagine a case where planned comparisons are appropriate: the levels were selected *randomly*, so why would we plan to do a specific comparison? Also, post-hoc tests strike me as odd because the statistical tests are being used to evaluate the variance of the population effects, rather than estimating the α's for the particular levels of the factor used in the experiment. Nevertheless, there may be a situation where you feel compelled to do comparisons among the group means. In that case, the methods are the same as those used to evaluate contrasts of fixed effects.

Table 1: Hypothetical population of A and B scores.

	b_1	b_2	b_3	b_4	b_5	b_6	Mean
a_1	4	6	3	1	2	2	3
a_2	2	0	3	5	4	4	3
mean	3	3	3	3	3	3	3

10.2.4 R example

This example uses the `Dyesuff` data frame, which is part of the `lme4` package:

```
> library(lme4) # load lme4 into memory
> data(Dyestuff) # load data frame into memory
> head(Dyestuff)

  Batch Yield
1     A  1545
2     A  1440
3     A  1440
4     A  1520
5     A  1580
6     B  1540

> with(Dyestuff,tapply(Yield,Batch,length))

A B C D E F
5 5 5 5 5 5
```

The data frame contains the yield of dyestuff (Naphthalene Black 12B) from five preparations from each of six randomly selected batches of an intermediate chemical product (H-acid). We know that the yield varies across preparation. Presumably, the variation among yields is due to variation among batches as well as variation among preparations. We will conduct a one-way, random effects ANOVA to estimate the two variance components:

```
> options(contrasts=c("contr.sum","contr.poly"))
> dye.aov <- aov(Yield~Batch,data=Dyestuff)
> summary(dye.aov)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Batch        5  56358   11272   4.598 0.0044 **
Residuals   24  58830    2451
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> var.comp.error <- 2451 # variance component preparations (i.e., error)
> n <- 5
> ( var.comp.batch <- (11272-2451) / 5 ) # variance component batch
```

```
[1] 1764.2
```

```
> var.comp.batch / var.comp.error
```

```
[1] 0.7197878
```

```
> var.comp.batch / (var.comp.batch + var.comp.error)
```

```
[1] 0.4185329
```

```
> sqrt(var.comp.batch + var.comp.error) # estimated population stand dev of yield
```

```
[1] 64.92457
```

```
> sd(Dyestuff$Yield) # sample stand dev of yield
```

```
[1] 63.02367
```

The effect of `Batch` was significant ($F(5, 24) = 4.59$, $p = 0.004$), so we reject the null hypothesis that the variance among batches is zero. Also, the estimated variance component for batch (1764.2) is 72% the size of the variance among preparations (2451): in other words, variation among batches accounts for $\approx 42\%$ of the total variance of yields. Finally, we note that the estimated standard deviation of yields, $\hat{\sigma}_{yield} = 64.9 = \sqrt{\hat{\sigma}^2_{batch} + \hat{\sigma}^2_{error}}$ is similar to the sample standard deviation (63.02).

10.3 two-way factorial designs

10.3.1 mixed model

In this section we consider the analysis of an experiment that uses two factors that are crossed in a balanced factorial design. Previously, we have considered the case where both factors are fixed; here we consider the case where at least one of the factors is random. The full linear model for analyzing data collected in such experiments is

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk} \tag{15}$$

When one factor (A) is fixed and the other factor (B) is random, Equation 15 represents a mixed model. For the mixed model (A fixed), the intercept, μ, is estimated by the grand mean of the scores. The effect of treatment a_j (i.e., α_j) is defined as in the fixed-effects case and is subject to the constraint that the sum of α's is zero. The effect of treatment b_k is a random variable that is distributed normally with a mean of zero and a population variance of σ^2_β. The interaction effect, $(\alpha\beta)_{jk}$, is a random variable distributed normally with a mean of zero and a variance of $\sigma^2_{(\alpha\beta)}$, with the constraint that the sum of interaction effects *across the levels of the fixed factor* is zero:

$$\sum_{j=1}^{a} (\alpha\beta)_{jk} = 0$$

The random variables β_k and $(\alpha\beta)_{jk}$ are independent of each other, and also are independent of ϵ_{ijk} which is distributed normally with a mean of zero and a variance of σ_e^2.

The sum of squares for the main effects and interaction are calculated by comparing the sum of squares residuals for the full model (Equation 15) with reduced models that have set the α's, β's, and $(\alpha\beta)$'s to zero. For the case where there are equal n per cell

$$SS_A = bn\frac{\sum_j \hat{\alpha}_j^2}{a-1} \quad (16)$$

$$SS_B = an\frac{\sum_k \hat{\beta}_j^2}{b-1} \quad (17)$$

$$SS_{A\times B} = n\frac{\sum_j \sum_k \left(\hat{\alpha\beta}\right)_{jk}^2}{(a-1)(b-1)} \quad (18)$$

The mean squares are obtained by dividing each SS value with the appropriate degrees of freedom ($df_A = a-1$; $df_B = b-1$; $df_{A\times B} = (a-1)(b-1)$).

The methods for calculating SS and MS values for mixed models are the same as those used to evaluate fixed-effects models. However, the expected values of the means squares, and therefore the method of evaluating the statistical significance of the effects, differ from the fixed-effects case. To illustrate this difference, consider the situation in which two factors, A and B, are crossed and there are no main effects: the marginal means of A and the marginal means of B are equal. Such a situation is illustrated in Table 1. Now consider what would happen if we randomly sampled two levels of B from the population of six levels, creating a set of four scores. One example is shown in Table 2. Notice that the marginal means of A are no longer equal. In other words, sampling the levels of B resulted in an apparent main effect of A. How did this occur? Inspection of Table 1 indicates that the effect of A *depended* on the level of B: There was an $A \times B$ interaction. These interaction effects summed to zero across all levels of B, and therefore there was no main effect of A. However, the interaction effects are not guaranteed to sum to zero across a random subset of the levels of B, and therefore the $A \times B$ interaction can masquerade as a main effect of A. This effect of subsampling the levels of B alters the expected value of MS_A.

Table 2: Scores resulting after sampling levels of factor B.

	b_1	b_2	Mean
a_1	4	6	5
a_2	2	0	1
mean	3	3	3

The expected values of the mean squares for a design that includes one fixed factor (A) and one random factor (B) are shown in Table 3. Note that $\xi(\mathrm{MS}_B)$ and $\xi(\mathrm{MS}_{A\times B})$ are as expected: the expected values are the sum of σ_e^2 and the variance of the appropriate effect (i.e., σ_β^2 or $\sigma_{(\alpha\beta)}^2$). Also note that both B and $A \times B$ are considered random, not fixed, effects. Because these expected mean squares are of the expected form, the B main effect and the $A \times B$ interaction can be evaluated by in the usual way (i.e., estimating F by dividing MS_B and $\mathrm{MS}_{A\times B}$ by the mean square residual term, or MS_R). However, such an F test is *not* appropriate for evaluating the effect of A. If $\sigma_{(\alpha\beta)}^2 > 0$, then $\xi(\mathrm{MS}_A) > \sigma_e^2$ even if all of the fixed effects (i.e., the $\alpha's$) are zero. In other words, comparing MS_A to MS_R would produce a biased test of the null hypothesis that $\alpha_j = 0$ for every group j, and our Type I error would be greater than the nominal value. On the other hand, comparing MS_A to $\mathrm{MS}_{A\times B}$ would produce an unbiased test of the null hypothesis, because the expected values of MS_A and $\mathrm{MS}_{A\times B}$ are equal when the null hypothesis of all α_j's being zero is true. Therefore, when evaluating the

Table 3: Expected mean squares for mixed and random factorial designs.

Source	$\xi(MS)$ A fixed; B random	$\xi(MS)$ A & B random
A	$\sigma_e^2 + n\sigma_{(\alpha\beta)}^2 + bn\sum_{j=1}^{a}\alpha_j^2/(a-1)$	$\sigma_e^2 + n\sigma_{(\alpha\beta)}^2 + bn\sigma_\alpha^2$
B	$\sigma_e^2 + an\sigma_\beta^2$	$\sigma_e^2 + n\sigma_{(\alpha\beta)}^2 + an\sigma_\beta^2$
$A \times B$	$\sigma_e^2 + n\sigma_{(\alpha\beta)}^2$	$\sigma_e^2 + n\sigma_{(\alpha\beta)}^2$
Residuals	σ_e^2	σ_e^2

main effect of a fixed factor in a mixed design, the mean square of the fixed factor (MS_A) is compared to the mean square of the interaction ($\text{MS}_{A\times B}$), not mean square residuals (MS_R).

The null and alternative hypotheses for the main effect of the fixed factor, A, are

$$\begin{aligned} \text{H0}: \ \alpha_j &= 0 \text{ for all levels } j \\ \text{H1}: \ \alpha_j &\neq 0 \text{ for at least one level } j \end{aligned}$$

The null and alternative hypotheses for the random factor, B, are

$$\begin{aligned} \text{H0}: \ \sigma_\beta^2 &= 0 \\ \text{H1}: \ \sigma_\beta^2 &> 0 \end{aligned}$$

Finally, the F test for the $A \times B$ interaction evaluates the following hypotheses:

$$\begin{aligned} \text{H0}: \ \sigma_{(\alpha\beta)}^2 &= 0 \\ \text{H1}: \ \sigma_{(\alpha\beta)}^2 &> 0 \end{aligned}$$

10.3.2 random-effects model

We now consider the situation where the experiment contains two crossed, random factors. The full random-effects model is Equation 15, the same as the mixed model. However, unlike the mixed model, the effect of α_j is a random variable distributed normally with a zero mean and a variance σ_α^2. Table 3 also shows the expected mean squares for the random-effects model. In this case, both MS_A and MS_B contain a variance component related to the $A \times B$ interaction (i.e., $n\sigma_{(\alpha\beta)}^2$), and therefore neither main effect can be evaluated by comparing the mean square to MS_R. Instead, both random main effects must be evaluated by comparing the mean squares to $\text{MS}_{A\times B}$. The interaction term in a two-way, factorial random design can still be evaluated by comparing $\text{MS}_{A\times B}$ to MS_R. The F test for the main effect of A evaluates the following null and alternative hypotheses:

$$\begin{aligned} \text{H0}: \ \sigma_\alpha^2 &= 0 \\ \text{H1}: \ \sigma_\alpha^2 &> 0 \end{aligned}$$

The F tests for B and $A \times B$ evaluate the same null hypotheses that were tested in the mixed model.

10.3.3 strength of association

For two-factor designs, the recommended indices of association strength for fixed and random factors are, respectively, partial omega-squared and partial intraclass correlation. If A is fixed and B is random, then partial omega-squared for the fixed factor A is

$$\omega_{A,partial}^2 = \frac{\sum_{j=1}^{a}(\alpha_j^2/a)}{\sigma_e^2 + \sum_{j=1}^{a}(\alpha_j^2/a)}$$

and is estimated by

$$\hat{\omega}^2_{A,partial} = \frac{(a-1)(F_A-1)}{(a-1)(F_A-1)+nab}$$

The partial intraclass correlations for the random components B and $A \times B$ are

$$\rho_{I:B,partial} = \frac{\sigma^2_\beta}{\sigma^2_\beta + \sigma^2_e}$$

$$\rho_{I:AB,partial} = \frac{\sigma^2_{(\alpha\beta)}}{\sigma^2_{(\alpha\beta)} + \sigma^2_e}$$

Estimates of the partial intraclass correlations can be obtained from estimates of the variance components, which in turn can be derived from values in the ANOVA table. When A is fixed and B is random, the variance components for B, $A \times B$, and the error term are

$$\hat{\sigma}^2_\beta = \frac{\text{MS}_B - \text{MS}_R}{na}$$

$$\hat{\sigma}^2_{(\alpha\beta)} = \frac{\text{MS}_{A\times B} - \text{MS}_R}{n}$$

$$\hat{\sigma}^2_e = \text{MS}_R$$

When A and B are both random, the expected values of the mean squares change – now the values for *both* main effects are influenced by the $A \times B$ interaction (see Table 3) – and therefore the variance components are calculated with the following formulae:

$$\hat{\sigma}^2_\alpha = \frac{\text{MS}_A - \text{MS}_{A\times B}}{nb}$$

$$\hat{\sigma}^2_\beta = \frac{\text{MS}_B - \text{MS}_{A\times B}}{na}$$

$$\hat{\sigma}^2_{(\alpha\beta)} = \frac{\text{MS}_{A\times B} - \text{MS}_R}{n}$$

$$\hat{\sigma}^2_e = \text{MS}_R$$

Note that it is possible for these so-called ANOVA estimates of variance components to be negative, and therefore for intraclass correlations to be less than zero. Such values cannot correspond to the true population values because variances must be equal to or greater than zero, and the true intraclass correlation must be between zero and 1 (inclusive). Therefore, when the estimated association strength is less than zero, it is standard practice to set it to zero.

10.3.4 R example

This example uses data shown in Table 5 of your textbook (page 482). The data come from a fictitious study that examined the effectiveness of two programs to prepare high-school students for US-college entrance exams. One study program uses a traditional package of written materials, whereas the other is based on an interactive computer program. The experiment compared the two study programs in four randomly selected schools. The type of program is represented by the fixed factor `study`, but `schools` is a random factor. Hence, we will use a mixed model to analyze these data.

```
> options(contrasts=c("contr.sum","contr.poly"))
> mw.act <- read.table("chapter_10_table_5.dat")
> names(mw.act)<-c("study","school","score")
> mw.act$study<-factor(mw.act$study,labels=c("computer","standard"))
> mw.act$school<-factor(mw.act$school,labels="s")
> with(mw.act,tapply(score,list(study,school),length))

         s1 s2 s3 s4
computer  5  5  5  5
standard  5  5  5  5

> summary(aov(score~study+school+study:school,data=mw.act))

             Df Sum Sq Mean Sq F value   Pr(>F)
study         1    360   360.0  19.931 9.35e-05 ***
school        3    100    33.3   1.845    0.159
study:school  3     80    26.7   1.476    0.240
Residuals    32    578    18.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The anova table shows the F and p values for a fixed-effects model: all F values are calculated using MS_R for the denominator in the means square ratio. However, `schools` is a random factor, and therefore the mean square for `study` (i.e., the fixed factor) should be compared to the mean square value for the `study` × `school` interaction (see Table 3). The correct values of F and p for the main effect of `study` are

```
> (F.study <- 360/26.67)

[1] 13.49831

> (p.study <- 1-pf(F.study,1,3))

[1] 0.03490257
```

Therefore, we do not reject the null hypotheses that $\sigma^2_{school} = 0$ and $\sigma^2_{study:school} = 0$, but we do reject the null hypothesis that the marginal means of the two levels of `study` are equal.

The variance components are

$$\begin{aligned}
\hat{\sigma}^2_{school} &= \frac{33.33 - 18.06}{2 \times 5} = 1.53 \\
\hat{\sigma}^2_{study:school} &= \frac{26.67 - 18.06}{5} = 1.72 \\
\hat{\sigma^2_e} &= 18.06
\end{aligned}$$

Obviously, the variance components associated with `school` and `study:school` are much smaller than the error variance[2]. Partial intraclass correlations for the two random effects are

$$\begin{aligned}
\rho_{I:school,partial} &= \frac{1.53}{1.53 + 18.06} = 0.078 \\
\rho_{I:study:school,partial} &= \frac{1.72}{1.72 + 18.06} = 0.087
\end{aligned}$$

[2]Note that this finding – i.e., that variation on our dependent variable that is related to schools was *small* compared to error variance – might be interesting and important. In fact it might be more important than the result showing that the effectiveness of the two programs of study differ.

The association strength for the fixed effect is

$$\hat{\omega}^2_{study,partial} = \frac{(2-1)(13.49-1)}{(2-1)(13.49-1)+5\times 2\times 4} = 0.237$$

and the effect size is

$$\hat{f}_{study} = \sqrt{\frac{0.237}{1-.237}} = 0.56$$

This is a large effect.

10.4 Nested Factors

A factor is said to be *nested within another factor* if each level of the first factor occurs in conjunction *with only one* level of the second factor. So B is nested within A if each level of B occurs within only one level of A. An example of a design in which B is nested within A that uses $n = 6$ subjects per cell is shown in Table 4. Notice that b_{1-3} occur only with a_1, whereas b_{4-6} occur only with a_2. A nested design can be thought of as a factorial design that is missing some cells. Obviously, the fact that some cells are missing means that some parameters of the full model for factorial designs (Equation 1) cannot be estimated. For instance, there is no way to estimate the $A \times B$ interaction because some of the combinations of A and B are missing.

The full model for a nested design (i.e., B nested within A) is

$$Y_{ijk} = \mu + \alpha_j + \beta_{k/j} + \epsilon_{ijk} \tag{19}$$

α_j is defined as it has been in all previous models: it represents the difference between the marginal mean of level a_j and the grand mean (i.e., $\alpha_j = \mu_{j.} - \mu$). The term for the nested effect, $\beta_{k/j}$, is defined as

$$\beta_{k/j} = \mu_{jk} - \mu_{j.}$$

which is the difference between a cell mean and the marginal mean for the non-nested factor in which the cell appears. Another way of describing $\beta_{k/j}$ is to say that it is the simple effect of the k-th level of B (the nested factor) in the j-th level of A (the non-nested factor). The error term, ϵ_{ijk}, is a normally distributed random variable with a mean of zero and a population variance of σ^2_e. Notice that Equation 19 does not include an interaction term.

Table 4: Example of a nested design.

	b_1	b_2	b_3	b_4	b_5	b_6
a_1	6	6	6	x	x	x
a_2	x	x	x	6	6	6

The sum of squares for the effects in Equation 19 are calculated the standard way, namely by constructing reduced models that have the effects of interest set to zero, and then calculating the change in the sum of squared residuals. For the nested term, the sum of squares is denoted by $SS_{B/A}$ and equals

$$SS_{B/A} \sum_j \sum_k n\hat{\beta}^2_{k/j}$$

which can be shown to equal to be equivalent to

Table 5: Expected mean squares for nested designs (B nested within A).

Source	$\xi(MS)$ A fixed; B random	$\xi(MS)$ A & B random	df
A	$\sigma_e^2 + n\sigma_\beta^2 + bn\sum_{j=1}^{a}\alpha_j/(a-1)$	$\sigma_e^2 + n\sigma_\beta^2 + bn\sigma_\alpha^2$	$a-1$
B/A	$\sigma_e^2 + n\sigma_\beta^2$	$\sigma_e^2 + n\sigma_\beta^2$	$\sum_{j=1}^{a}(b-1) = a(b-1)$
Residuals	σ_e^2	σ_e^2	$ab(n-1)$

$$SS_{B/A} = \sum_{j=1}^{a} SS_{B\ w\ A_j}$$

Hence, the sum of squares for the nested factor is obtained by pooling the sums of squares of the simple effect of B at each level of A.

The expected mean squares and degrees of freedom for mixed and random-effects nested designs are shown in Table 5. Note that the degrees of freedom for the nested factor is $a(b-1)$, where b equals the number of levels of B *in each level of* A. The values of $\xi(MS)$ make it clear that the main effect of A should be evaluated by comparing MS_A to $\text{MS}_{B/A}$, and the effect of B/A is evaluated by comparing $\text{MS}_{B/A}$ to MS_R. When B is random, the null hypothesis is that $\sigma_\beta^2 = 0$.

10.4.1 variance components

When A is fixed and B is random, the variance of our dependent variable will have two components, σ_ϵ^2 and σ_β^2. The ANOVA estimate of σ_e^2 is given by the Mean Square Within-cell (or Mean Square Residuals). The ANOVA estimate of σ_β^2 is

$$\sigma_\beta^2 = \frac{MS_{B/A} - MS_R}{n} \tag{20}$$

where MS_R is the Mean-Square Residuals and n is the number of observations per cell (which is assumed to be constant). Note that this definition differs from the one given in Table 10.10 in your book. Equation 20 is correct; the formula in Table 10.10 has a typographical error.

When A and B are both random, the variance of our dependent variable will have three components, σ_e^2, σ_β^2, and σ_α^2. The ANOVA estimates of σ_ϵ and σ_β are given by the Mean Square Within-cell and Equation 20, respectively. The ANOVA estimate of σ_α^2 is

$$\sigma_\alpha^2 = \frac{MS_A - MS_{B/A}}{bn} \tag{21}$$

where $MS_{B/A}$ is the Mean-Square for B nested within A, b is the number of levels of B within each level of A, and n is the number of observations per cell.

10.4.2 homogeneity of variance assumption

The assumptions underlying our analysis of a nested design are similar to those used in previous models. The errors are assumed to be distributed normally with constant variance. When A or B is random, then the effects are assumed to be distributed normally with variances σ_α^2 and σ_β^2, and the random factors are assumed to be independent of each other and ϵ. There is, however, one additional assumption that we make when we analyze a nested design: Specifically, we must assume that the $\beta_{k/j}$ terms **have the same variability at each level of factor** A. In other words, σ_β^2 does not depend on, or interact with, A. Note that we have to make this assumption because there is no obvious way to evaluate the $A \times B$ interaction in this design.

10.4.3 R examples

The first example uses data shown in Table 9 of your textbook (page 500). The data come from a fictitious that examined the effect of gender on the severity ratings that clinical psychology graduate students assigned to patients. Three male and three female clinical trainees were selected randomly to participate, and four patients were assigned randomly to each trainee. In this experiment, `gender` is a fixed factor and `trainee` is a random factor that is nested within `gender`. Here is the analysis:

```
> mw10.9<-read.csv("chapter_10_table_9.csv")
> mw10.9

   gender trainee score
1       1       1    49
2       1       1    40
3       1       1    31
4       1       1    40
5       1       2    42
6       1       2    48
7       1       2    52
8       1       2    58
9       1       3    42
10      1       3    46
11      1       3    50
12      1       3    54
13      2       4    53
14      2       4    59
15      2       4    63
16      2       4    69
17      2       5    44
18      2       5    54
19      2       5    54
20      2       5    64
21      2       6    58
22      2       6    63
23      2       6    67
24      2       6    72

> mw10.9$gender<-factor(mw10.9$gender,labels=c("male","female"))
> mw10.9$trainee<-factor(mw10.9$trainee,labels="t")
> with(mw10.9,tapply(score,list(gender,trainee),length))

       t1 t2 t3 t4 t5 t6
male    4  4  4 NA NA NA
female NA NA NA  4  4  4

> mw10.9.aov <- aov(score~gender+trainee,data=mw10.9)
> summary(mw10.9.aov)

            Df Sum Sq Mean Sq F value  Pr(>F)
gender       1   1176  1176.0   25.82 7.8e-05 ***
trainee      4    472   118.0    2.59  0.0716 .
Residuals   18    820    45.6
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> (F.gender <- 1176/118)

[1] 9.966102

> (p.gender <- 1-pf(F.gender,df1=1,df2=4) )

[1] 0.03428551
```

There are several things to note. First, an examination of the data frame shows that each trainee is given a unique number, so the three trainees in the second level of `gender` are designated as trainees 4, 5, and 6 rather than 1, 2, and 3. This coding is important because it tells R that the two sets of trainees are different. Second, the linear model lacks an interaction term, which is appropriate for this design. Third, the ANOVA table lists the second effect as `trainee` rather than `trainee/gender`. Fourth, the F and p values for `gender` were obtained by comparing the means square for gender to MS_R, and therefore these values are incorrect. The correct values are obtained by comparing the mean square for `gender` to the mean square for `trainee`. The following code calculates the variance component for `trainee`:

```
> MS.TraineeInGender <- 118
> MS.resid <- 45.56
> n <- 4;
> (Var.Comp.Trainee <- (MS.TraineeInGender - MS.resid)/n)

[1] 18.11
```

The estimated variance component for `trainee` (18.11) is approximately 40% of the estimated error variance (45.56). Finally, we can estimate the strength of association between gender and the dependent variable by calculating the variance components for trainee and error and then using those values to calculate omega squared and partial omega squared for gender using equations in Table 10.10 and on page 509 in your textbook. Note that those equations have typographical errors; the corrections are shown at
`www.psychology.mcmaster.ca/bennett/psy710/notes/errata-table10_10-page_509.pdf`.

```
> a <- 2 # 2 genders
> b <- 3 # levels on trainee factor within each level of gender
> n <- 4 # measures per cell
> MS.gender <- 1176
> MS.trainee.in.gender <- 118
> MS.error <- 45.6
> # variance components
> var.comp.error <- MS.error
> (var.comp.trainee <- (MS.trainee.in.gender - MS.error)/n )

[1] 18.1

> # using equations on page 509 in textbook:
> ( sigma.alpha.squared <- (a-1)/a * (MS.gender - MS.trainee.in.gender)/(b*n) )

[1] 44.08333

> (omega.squared <- sigma.alpha.squared / (sigma.alpha.squared + var.comp.trainee + var.comp.error)
```

```
[1] 0.4089995

> (p.omega.squared <- sigma.alpha.squared / (sigma.alpha.squared +  var.comp.error) )

[1] 0.4915443
```

The second example uses data from Table 10.16.1 (p 286) in (Snedecor and Cochran, 1967). Four turnip plants were selected at random, then three leaves were selected randomly from each plant. Finally, two 100 mg samples were taken from each leaf, and the calcium concentration was measured for each sample. The data are shown in Table 6. Notice that plant and leaf are random factors, and the leaf is nested within plant. We want to use ANOVA to estimate the variance components for plant, leaf, and samples (i.e., error).

```
> head(turnips)

  plant leaf spot calcium
1    p1   L1   s1    3.28
2    p1   L1   s2    3.09
3    p1   L2   s3    3.52
4    p1   L2   s4    3.48
5    p1   L3   s5    2.88
6    p1   L3   s6    2.80

> with(turnips,tapply(calcium,list(plant,leaf),length)) # balanced!

   L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12
p1  2  2  2 NA NA NA NA NA NA  NA  NA  NA
p2 NA NA NA  2  2  2 NA NA NA  NA  NA  NA
p3 NA NA NA NA NA NA  2  2  2  NA  NA  NA
p4 NA NA NA NA NA NA NA NA NA   2   2   2

> n <- 2
> a <- 4
> b <- 3
> summary(aov(calcium~plant+leaf,data=turnips) )

            Df Sum Sq Mean Sq F value   Pr(>F)
plant        3   7.56  2.5201  378.73 3.80e-12 ***
leaf         8   2.63  0.3288   49.41 5.09e-08 ***
Residuals   12   0.08  0.0067
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R prints the ANOVA table under the assumption that the two factors are fixed, and therefore the F's are calculated by dividing the mean square values for each factor by the mean square residuals. The resulting F and p values for leaf are correct, but the we need to recalculate F and p for plant using the mean square for leaf:

```
> (F.plant <- 2.5201/.3288) # denominator from leaf

[1] 7.664538

> (p.plant <- 1-pf(F.plant,3,8) )
```

```
[1] 0.009727413
```

The F's for plant and leaf are both significant, so we reject the null hypotheses that the variance components are zero. Finally, we calculate the variance components from values listed in the ANOVA table, and then use those values to compute the partial intraclass correlations for plant and leaf:

```
> var.comp.error <- .0067
> ( var.comp.leaf <- (.3288-.0067) / n )

[1] 0.16105

> ( var.comp.plant <- (2.5201 - .3288) / (b*n) )

[1] 0.3652167

> # estimate of pop stand dev for calcium:
> sqrt(var.comp.plant + var.comp.leaf + var.comp.error)

[1] 0.7300457

> # compare estimate to sample st dev
> sd(turnips$calcium)

[1] 0.6682357

> (part.icc.leaf <- var.comp.leaf / (var.comp.leaf + var.comp.error) )

[1] 0.9600596

> (part.icc.plant <- var.comp.plant / (var.comp.plant + var.comp.error) )

[1] 0.9819852
```

Our analysis suggests that the variation among samples within a leaf is very small compared to variation among leaves and plants.

Table 6: Data from Table 10.16.1 in (Snedecor and Cochran, 1967).

Plant	Leaf	Samples
1	1	3.28, 3.09
1	2	3.52, 3.48
1	3	2.88, 2.80
2	4	2.46, 2.44
2	5	1.87, 1.92
2	6	2.19, 2.19
3	7	2.77, 2.66
3	8	3.74, 3.44
3	9	2.55, 2.55
4	10	3.78, 3.87
4	11	4.07, 4.12
4	12	3.31, 3.31

References

Snedecor, G. W. and Cochran, W. G. (1967). *Statistical methods*. Iowa State University Press, Ames, 6th ed edition.