4

FACTORIAL ANALYSIS OF VARIANCE: MODELING INTERACTIONS

> The assignable sources of variation in a manufacturing process may be divided into two categories. First, there are those factors which introduce variation in a random way. Lack of control at some stage of production very often acts in this manner, and the material itself usually exhibits an inherent random variability. The other type of factor gives rise to systematic variation.
>
> (Daniels, 1939, p. 187, *The Estimation of Components of Variance*)

The researcher of Chapter 3 who studied the effect of melatonin dosage on sleep onset is interested now in learning whether these effects are consistent across ambient noise levels present during sleep. For this experiment, the researcher again randomly assigns 25 individuals to a control group, 25 more to a group receiving 1 mg of melatonin, and 25 more to a group receiving 3 mg of melatonin. In addition, within each of these conditions, half of the participants receive either no ambient noise or a low amount of ambient noise at the moment of melatonin ingestion and lasting throughout the night (for instance, a slight buzzing sound). The researcher would like to test whether sleep onset is a function of dosage, ambient noise, and a potential **combination** of the two factors. That is, the researcher is interested in detecting a potential **interaction** between dose and noise level. He is only interested in generalizing his findings to these particular doses of melatonin and to these particular noise levels. Such a research design calls for a **two-way fixed effects factorial analysis of variance**.

4.1 WHAT IS FACTORIAL ANALYSIS OF VARIANCE?

In the one-way ANOVA of the previous chapter, we tested null hypotheses about equality of population means of the kind:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_J$$

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

In the two-way and higher-order analysis of variance, we have more than a single factor in our design. As we did for the one-way analysis, we will test similar **main effect** hypotheses for each individual factor, but we will also test a new null hypothesis, one that is due to an **interaction** between factors.

In the two-factor design on melatonin and ambient noise level, we are interested in the following effects:

- **Main effect** due to drug dose in the form of mean sleep differences across dosage levels.
- **Main effect** due to ambient noise level in the form of mean sleep differences across noise levels.
- **Interaction** between drug dose and noise level in the form of mean sleep differences on drug not being consistent across noise levels (or vice versa).

It does not take long to realize that science is about the discovery not of main effects, but of interactions. Yes, we are interested in whether melatonin has an effect, but we are even more interested in whether melatonin has an effect **differentially** across noise levels. And beyond this, we may be interested in even higher-order effects, such as **three-way interactions**. Perhaps melatonin has an effect, but mostly at lower noise levels, and mostly for those persons aged 40 and older. This motivates the idea of a three-way interaction, drug dose by noise level by age. One will undoubtedly remark the tone of **conditional probability** themes in the concept of an interaction.

As another example of an interaction, consider Table 4.1 and corresponding Figure 4.1. The plot features the achievement data of the previous chapter, only that now, in addition to students being randomly assigned to one of four teachers (`f.teach`), they were also randomly assigned to the study of one of two mathematics textbooks (`f.text`).

What we wish to know from Figure 4.1 is whether textbook differences (1 versus 2) are **consistent** across levels of teacher. For instance, at teacher = 1, we ask whether the same textbook "story" is being told as at teachers 2, 3, and 4. What this "story" is, are the distances between cell means, as emphasized in part (b) of the plot. Is this distance from textbook 1 to textbook 2 consistent across teachers, or do such differences depend in part on which teacher one has? These are the types of questions we need to ask in order to ascertain the presence or absence of an interaction effect. And though it would appear that mean differences are not equal across teacher, the question we really need to ask is whether these sample differences across teacher are large enough to infer population mean differences. These questions will be addressed by the test for an **interaction effect** in the two-way fixed effects analysis of variance model.

TABLE 4.1 Achievement as a Function of Teacher and Textbook

	Teacher			
Textbook	1	2	3	4
1	70	69	85	95
1	67	68	86	94
1	65	70	85	89
2	75	76	76	94
2	76	77	75	93
2	73	75	73	91

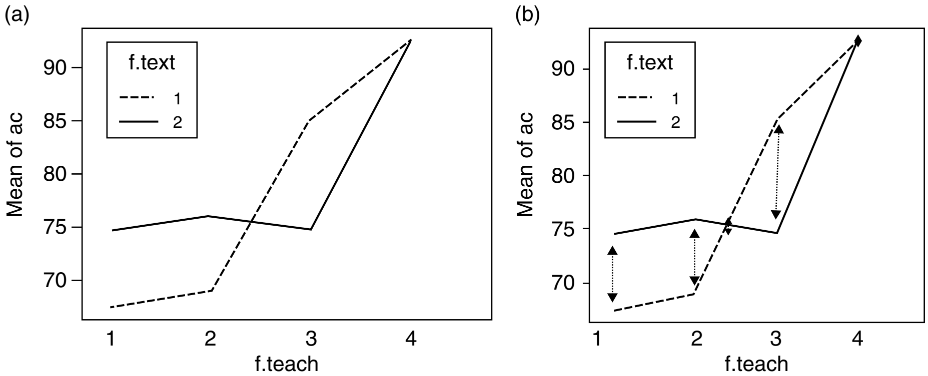


FIGURE 4.1 (a) Cell means for teacher∗textbook on achievement. (b) Distances between cell means as depicted by two-headed arrows. (where `f.text` is the factor name for textbook and `f.teach` is the factor name for teacher).

4.2 THEORY OF FACTORIAL ANOVA: A DEEPER LOOK

As we did for the one-way analysis of variance, we develop the theory of factorial ANOVA from fundamental principles which will then lead us to the derivation of the sums of squares. The main difference between the simple one-way model and the two-way model is the consideration of **cell effects** as opposed to simply **sample effects**. Consider, in Table 4.2, what the two-way layout might look like for our melatonin example in the factorial design.

We are interested in both **row** mean differences, summing across melatonin dose, as well as **column** mean differences, summing across noise level. We ask ourselves the same question we asked in the previous chapter for the one-way model:

Why does any given score in our data deviate from the mean of all the data?

Our answer must now include four possibilities:

- An effect of being in one melatonin-dose group versus others.
- An effect of being in one noise level versus others.
- An effect due to the combination (**interaction**) of dose and noise.

TABLE 4.2 Cell Means of Sleep Onset as a Function of Melatonin Dose and Noise Level (Hypothetical Data)

	Melatonin Dose			
Noise Level	0 mg	1 mg	3 mg	Row Means
High	15	11	8	11.3
Low	12	10	4	8.7
Column means	13.5	10.5	6.0	10.0

Column

Row

	1	2	3	•	•	•	•	•	j	•	•	C	Row means
1	x_{11}	x_{12}	x_{13}	•	•	•	•	•	x_{ij}	•	•	x_{1c}	$x_{1\cdot}$
2	x_{21}	x_{22}	x_{23}	•	•	•	•	•	x_{2j}	•	•	x_{2c}	$x_{2\cdot}$
•	•	•	•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•	•	•	•
i	x_{i1}	x_{i2}	x_{i3}	•	•	•	•	•	x_{ij}	•	•	x_{ic}	$x_{i\cdot}$
•	•	•	•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•	•	•	•
r	x_{i1}	x_{i2}	x_{i3}	•	•	•	•	•	x_{rj}	•	•	x_{rc}	$x_{r\cdot}$
Col means	$x_{\cdot 1}$	$x_{\cdot 2}$	$x_{\cdot 3}$	•	•	•	•	•	$x_{\cdot j}$	•	•	$x_{\cdot c}$	$x_{\cdot\cdot}$

FIGURE 4.2 Generic two-way analysis of variance layout. The two-way factorial analysis of variance has row effects, column effects, and interaction effects. Each value within each cell represents a data point. Row and column means are represented by summing across values of the other factor. Source: Eisenhart (1947). Reproduced with permission from John Wiley & Sons.

- Chance variation that occurs within each **cell** of the design. Notice that this 4^{th} possibility is now the **within-group** variation of the previous one-way model of Chapter 3, only that now, the "within group" is, in actuality, **within cell**. The error variation occurs within the cells of a factorial design.

In the spirit of history, we show an earlier and more generic layout of the two-way model diagramed by Eisenhart (1947) and reproduced in Figure 4.2, where entries in the cells depict data points for each row and column combination. Note the representation of row means and column means. These will aid in the computation of main effects for each factor.

4.2.1 Deriving the Model for Two-Way Factorial ANOVA

We now develop some of the theory behind the two-way factorial model. As always, it is first helpful to recall the essentials of the one-way model, then extend these principles to the higher-order model. Recall the one-way fixed effects model of the previous chapter:

$$y_{ij} = \bar{y}_{.} + a_j + e_{ij}$$

where the sample effect a_j was defined as $a_j = \left(\bar{y}_j - \bar{y}.\right)$. The sample effect a_j denoted the effect of being in one particular sample in the layout. Recall that in the one-way layout, $\sum_j n_j a_j = 0$, which in words meant that the sum of weighted sample effects, where n_j was the sample size per group, summed to zero. For this reason, we squared these treatment effects, which provided us with a measure of the sums of squares between groups:

$$\text{SS between} = \sum_j n_j a_j^{\,2}$$

TABLE 4.3 Cell Means Layout for 2 × 3 Factorial Analysis of Variance

	Factor 2		
Factor 1	Level 1	Level 2	Level 3
Level 1	$\bar{y}_{jk}$	$\bar{y}_{jk}$	$\bar{y}_{jk}$
Level 2	$\bar{y}_{jk}$	$\bar{y}_{jk}$	$\bar{y}_{jk}$

It turned out as well that the sample effect a_j was an **unbiased estimator** of the corresponding population effect, α_j. That is, the expectation of a_j is equal to α_j, or, more concisely, $E(a_j) = \alpha_j$. Recall that the sample effect represents the effect or influence of group membership in our design. For instance, for an independent variable having three levels, we had three groups ($J = 3$) on which to calculate respective sample effects. In the factorial two-way analysis of variance, we will have more than J groups because we are now **crossing** two variables with one another. For example, the layout for a 2×3 (i.e., 2 rows by 3 columns) design is given in Table 4.3.

Notice that now, we essentially have six "groups" in the 2×3 factorial model, where each combination of factor levels generates a mean $\bar{y}_{jk}$, where j designates the row and k designates the column. The "groups" that represent this combination of factor 1 and factor 2 we will refer to as **cells**. This is why we have been putting "groups" in quotation marks, because what these things really are in the factorial design are cells. **The heart of partitioning variability in a factorial design happens between cells**. In addition to defining the sample effects associated with each factor (i.e., a_j and b_k), we will now also need to define what is known as a **cell effect**.

4.2.2 Cell Effects

A **sample cell effect** (Hays, 1994, p. 477) is defined as:

$$[ab]_{jk} = \left(\bar{y}_{jk} - \bar{y}..\right)$$

and represents a measure of variation for being in one cell and not others. Notice that to compute the cell effect, we are taking each cell mean $\bar{y}_{jk}$ and subtracting the grand mean, $\bar{y}..$ (we carry **two periods** as subscripts for the grand mean now to denote the summing across j rows and k columns). But why do this? We are doing this similar to why we took the group mean and subtracted the grand mean in a simple one-way analysis of variance. In that case, in which we computed $a_j = \left(\bar{y}_j - \bar{y}.\right)$, we were interested in the "effect" of being in one group versus other groups (which was represented by subtracting the overall mean).

Likewise, in computing cell effects, we are interested in the effect of being in one cell versus other cells, because now, in the two-way factorial model, in addition to both main effects for row and column, it is the cell effect that will represent our interests in there possibly existing an interaction between the two factors. We will need to compute an interaction effect to do this, but getting the cell effect is the first step toward doing so.

As it was true that the sum of sample effects in the one-way model was equal to 0, $\sum_j n_j a_j = 0$, it will also be true that the sum of cell effects is equal to 0 for any given sample. That is,

$$\sum_j \sum_k [ab]_{jk} = \sum_j \sum_k \left(\bar{y}_{jk} - \bar{y}..\right) = 0$$

TABLE 4.4 Deviations Featured in One-way and Two-way Analysis of Variance

Deviation	In Words	Solution is Squaring Deviations
$\sum_{i=1}^{n}(y_i - \bar{y}.) = 0$	The sum of score deviations around a mean equals 0	$\sum_{i=1}^{n}(y_i - \bar{y}.)^2 > 0$
$\sum_{i=1}^{n}(\bar{y}_j - \bar{y}..) = 0$	The sum of row sample mean deviations around a grand mean equals 0	$\sum_{i=1}^{n}(\bar{y}_j - \bar{y}..)^2 > 0$
$\sum_{i=1}^{n}(\bar{y}_k - \bar{y}..) = 0$	The sum of column sample mean deviations around a grand mean equals 0	$\sum_{i=1}^{n}(\bar{y}_k - \bar{y}..)^2 > 0$
$\sum_{i=1}^{n}(\bar{y}_{jk} - \bar{y}..) = 0$	The sum of cell mean deviations around a grand mean equals 0	$\sum_{i=1}^{n}(\bar{y}_{jk} - \bar{y}..)^2 > 0$

In each case, the sum of deviations equals to 0.

where the double summation represents the summing across k columns first, then across j rows. We can easily demonstrate this by computing the cell effects for Table 4.2 across each row of noise level. For the first cell mean of 15 in row 1, column 1, the cell effect is computed as $15 - 10 = 5$. For row 1, column 2, the cell effect is $11 - 10 = 1$. The remaining cell effects are computed analogously (−2, 2, 0, −6). The sum of these cell effects is easily demonstrated to be zero $((5 + 1 + (-2) + 2 + 0 + (-6) = 0)$. But why would this be true? It is true for the same reason why summing sample effects equals 0. We are taking deviations from the grand mean, and by definition, the grand mean is the "center of gravity" of all means (in a balanced design). So, it is reasonable then that the sum of deviations around that value should be equal to 0. To avoid this, just as we did for the ordinary variance and for the variances derived in the one-way analysis of variance, we square deviations.

To better conceptualize deviations from means across the one-way and two-way factorial designs, it is helpful to compare and contrast the four scenarios featured in Table 4.4.

We can see from Table 4.4 that the solution in each case is to **square** respective deviations. This is precisely why in the case of cell effects, as we did for single deviations and mean deviations, we will likewise square them. We will call this sum of squared cell effects by the name of **SS AB cells**:

$$\text{SS AB cells} = \sum_{j}\sum_{k} n\left([ab]_{jk}\right)^2$$

where n is the number of observations per cell, which we assume to be equal for our purposes.

4.2.3 Interaction Effects

Having defined the meaning of a cell effect, we are now ready to define what is meant by an **interaction effect**. These interaction effects are the reason why we computed the cell effects in the first place. The **sample interaction effect** for each cell jk is given by

$$\begin{aligned}
(ab)_{jk} &= \text{interaction effect of cell } jk \\
&= \text{cell effect for cell } jk - \text{effect for row } j - \text{effect for column } k \\
&= [ab]_{jk} - a_j - b_k \\
&= \bar{y}_{jk} - \bar{y}.. - (\bar{y}_j - \bar{y}..) - (\bar{y}_{.k} - \bar{y}..) \\
&= \bar{y}_{jk} - \bar{y}_{j.} - \bar{y}_{.k} + \bar{y}..
\end{aligned}$$

A few things to remark about sample interaction effects:

- A sample interaction effect $(ab)_{jk}$ exists for each cell in the design.
- The sample interaction effect is defined by the cell effect minus the row and column effects (i.e., $[ab]_{jk} - a_j - b_k$); this makes sense, since it is reasonable that we are interested in the effect of being in a particular cell over and above the corresponding row and column effects.
- The sample interaction effect can also be defined as taking the mean of each cell, $\bar{y}_{jk}$, and subtracting out row means and column means (i.e., $\bar{y}_{j.}$ and $\bar{y}_{.k}$, respectively), then adding on the grand mean, $\bar{y}_{..}$.

As we did for sample effects, we will square the interaction effects so that they do not always sum to zero:

$$\text{SS A} \times \text{B interaction} = \sum_j \sum_k n(ab)_{jk}^{\ 2}$$

4.2.4 Cell Effects Versus Interaction Effects

It is useful at this point to emphasize an important distinction and to clarify something that may at first be somewhat confusing. We have introduced the ideas of cell effects and interaction effects. It is important to recognize that these are not the same things, as evidenced by their different computations. To help clarify, let's compare the two concepts:

$$\text{Cell Effect } [ab]_{jk} = \left(\bar{y}_{jk} - \bar{y}_{..}\right) \text{ versus Interaction Effect } (ab)_{jk} : [ab]_{jk} - a_j - b_k$$

Notice that the interaction effect $(ab)_{jk}$ uses the cell effect in its computation. In our operationalization of the two-way ANOVA, the cell effect is just the starting point to computing the interaction effect. The cell effect simply measures the deviation of a cell mean from the grand mean. It is the **interaction effect** that takes this deviation value and then subtracts further the row and column effects. **Be sure not to confuse cell effects and interaction effects as they are not one and the same**.

4.2.5 A Model for the Two-Way Fixed Effects ANOVA

Having now defined the sample interaction effect, which again, is the distinguishing feature between a one-way fixed effects model and a two-way fixed effects model, we can now state a general linear model for the two-way, one that includes an interaction term:

$$y_{ijk} = \bar{y}_{..} + a_j + b_k + (ab)_{jk} + e_{ijk}$$

where a_j is the sample effect of membership in row j, b_k is the sample effect of membership in column k, $(ab)_{jk}$ is the interaction effect associated with the cell jk, and e_{ijk} is the error associated with observation i in cell jk. In words, what the model says is that any given randomly selected observation from the two-way layout, represented by y_{ijk}, individual i in cell jk, can be theorized to be a function of the grand mean of all observations, $\bar{y}_{..}$, an effect of being in a particular row j, a_j, an effect of being in a particular column k, b_k, the effect of being in a particular cell combination, jk, which is expressed via the interaction effect $(ab)_{jk}$, and an effect unique to individuals within each cell jk, e_{ijk}, for which we either did not account for in our design, or, we concede is due to random variation which we will call by the name of "chance." Either way, e_{ijk} represents our inability to model y_{ijk} perfectly in a truly functional manner. Just as was true for the one-way model, e_{ijk} is the effect that makes our model truly probabilistic.

4.3 COMPARING ONE-WAY ANOVA TO TWO-WAY ANOVA: CELL EFFECTS IN FACTORIAL ANOVA VERSUS SAMPLE EFFECTS IN ONE-WAY ANOVA

It is pedagogical at this point to compare, side by side, the one-way model of the previous chapter to the two-way model of the current chapter. Recall the overall purpose of writing out a model equation. It is an attempt to "explain," in as functional a way as possible, the makeup of a given observation. In the one-way model, we attempted to explain observations by theorizing a single grouping factor along with within-group variability. Our sample model was

$$y_{ij} = \bar{y}_{.} + a_j + e_{ij}$$

Notice that for such a model, it was not appropriate to append the additional subscript k to y_{ij} such as in y_{ijk}, because we did not have "cells" in the one-way ANOVA. Defining the idea of a "cell" did not make a whole lot of sense, since we were simply dealing with a single grouping variable. Subscripting y_{ij} to represent individual i in group j was enough. Indeed, if we were to "pretend" for a moment that we were dealing with cells, we could write the one-way model as

$$\begin{aligned} y_{ij} &= \bar{y}. + \left(\bar{y}_j - \bar{y}.\right) + e_{ij} \\ y_{ij} &= \bar{y}. + [ab]_j + e_{ij} \end{aligned} \tag{4.1}$$

Nothing has changed in (4.1) except for equating "groups" with "cells." Why do this? Simply to note how the factorial model compares with that of the one-way model. Notice that the difference between the one-way model and the two-way model in terms of cell effects is that instead of hypothesizing y_{ijk} to be a function of $a_j = \bar{y}_j - \bar{y}.$, we are now hypothesizing y_{ijk} to be a function of $\bar{y}_{jk} - \bar{y}..$. In both cases, whether $a_j = \bar{y}_j - \bar{y}.$ for the one-way model or $[ab]_{jk} = \bar{y}_{jk} - \bar{y}..$ for the two-way model, **the total systematic variation in the data is represented by either of these, depending on whether there is one factor or two. Sample effects represent the systematic variation in a one-way model, and cell effects represent the systematic variation in a two-way model.** If you understand this concept, then generalizing these ANOVA models to higher-order models (e.g., three-way, four-way, and potentially higher) will not be intimidating, because you will realize at the outset that the systematic variation in the entire model is "housed" in the cell effects, regardless of the complexity of the model. To reiterate, we can say as a general principle of fixed effects analysis of variance models that

> **In the fixed effects analysis of variance model, the systematic variation is housed in the cell effects. In the special case where we have only a single independent variable, the cell effects are equivalent to the sample (group) effects.**

4.4 PARTITIONING THE SUMS OF SQUARES FOR FACTORIAL ANOVA: THE CASE OF TWO FACTORS

Just as we did for the one-way model, we will now work out the partition of the sums of squares for the two-way factorial model. Remember, the reason why we are partitioning "sums of squares" and not simply unsquared effects, is because if we attempted to partition unsquared effects (e.g.,

$a_j = \bar{y}_j - \bar{y}.$ or $[ab]_{jk} = \bar{y}_{jk} - \bar{y}..$), these effects would always sum to 0 (unless of course there is no variation in the data, then whether squared or not, they **will** sum to 0 regardless).

When we partitioned the sums of squares for the one-way model, we started out by hypothesizing what any single observation in our data, y_i, could be a function of. After a process of deliberate reasoning, we concluded that y_i was a function of **between** variation and **within** variation. Upon squaring deviations, we arrived at the identity:

$$\sum_{j=1}^{J}\sum_{i}^{n}(y_{ij} - \bar{y}.)^2 = \sum_{j}^{J} n_j(\bar{y}_j - \bar{y}.)^2 + \sum_{j=1}^{J}\sum_{i}^{n}(y_{ij} - \bar{y}_j)^2$$

which we called the partition of sums of squares for the one-way fixed effects analysis of variance model. We called it an "**identity**" simply because it holds true for virtually any given data set having a continuously measured dependent variable and a categorically-defined independent variable.

Likewise, in the two-way factorial model, we again want to consider how the partition of the sums of squares works out and can be derived. As we did for the one-way model, we follow a very logical process in determining this partition.

4.4.1 SS Total: A Measure of Total Variation

Just as we did in deriving the total sums of squares for the one-way model, instead of simply considering the makeup of y_{ijk}, we will consider the makeup of deviations of the form $y_{ijk} - \bar{y}..$, which when we incorporate into the model, we obtain, quite simply:

$$y_{ijk} = \bar{y}.. + [ab]_{jk} + e_{ijk}$$
$$y_{ijk} - \bar{y}.. = [ab]_{jk} + e_{ijk}$$

Notice that similar to how we did for the one-way model, in which $(y_{ij} - \bar{y}.) = a_j + e_{ij}$ was true, for the two-way model, we likewise claim that the makeup of any given observation is of two "things," systematic variation as represented by $[ab]_{jk}$ (in the one-way model the systematic variation was represented by a_j), and random variation as represented by e_{ijk} (in the one-way model the random variation was represented by e_{ij} — note the subscripts, we did not have **cells** in the one-way, so we did not need to append the subscript k). Instead of squaring $a_j + e_{ij}$ as is done in the one-way model, we will square $[ab]_{jk} + e_{ijk}$. When we take these squares and sum them, as given in Hays (1994, p. 481), we get:

$$\begin{aligned}
\text{SS total} &= \sum_{i}\left([ab]_{jk} + e_{ijk}\right)^2 \\
&= \sum_{j}\sum_{k}\sum_{i}\left([ab]_{jk}^2 + 2[ab]_{jk}e_{ijk} + e_{ijk}^2\right) \\
&= \sum_{j}\sum_{k}\sum_{i}\left([ab]_{jk}^2 + 2\sum_{j}\sum_{k}[ab]_{jk}\sum_{i}e_{ijk} + \sum_{j}\sum_{k}\sum_{i}e_{ijk}^2\right) \\
&= \sum_{j}\sum_{k}n[ab]_{jk}^2 + \sum_{j}\sum_{k}\sum_{i}e_{ijk}^2
\end{aligned}$$

Notice that the term $2\sum_j \sum_k [ab]_{jk} \sum_i e_{ijk}$ dropped out of the above summation (3rd line of the equation). What happened to this term? Since the cell effects $[ab]_{jk}$ sum to zero and the errors within any given cell $\sum_i e_{ijk}$ sum to 0, the term $2\sum_j \sum_k [ab]_{jk} \sum_i e_{ijk}$ drops out of the derivation, since $2\sum_j \sum_k [ab]_{jk} \sum_i e_{ijk} = 0$. Hence, we are left simply with:

$$\text{SS total} = \sum_j \sum_k n[ab]_{jk}^2 + \sum_j \sum_k \sum_i e_{ijk}^2$$

What we have just found is that the total variation in the two-factorial model is a function of the sum of **squared cell effects** and **random variation**. Once we have accounted for the systematic variation in $[ab]_{jk}$, then whatever is leftover must be random error, or otherwise denoted, the variation **within** the cells. Also, because the cell effects, $[ab]_{jk}$, contain all **systematic** variation, it makes sense that within these cell effects will be "hidden" a main effect for A, main effect for B, and interaction effect, A × B. That is, if you take the sums of squares for a cell effect which by itself contains all the systematic variation, it seems reasonable that we could break this down further into the **SS for factor A**, **SS for factor B**, and the **SS for the A x B** interaction, such that:

$$\text{SS AB cells} = \text{SS factor A} + \text{SS factor B} + \text{SS A} \times \text{B interaction}$$

If we put these two partitions together, we end up with the following identities:

$$\text{SS total} = \text{SS AB cells} + \text{SS within cells}$$

$$\text{SS total} = \text{SS factor A} + \text{SS factor B} + \text{SS A} \times \text{B interaction} + \text{SS within}$$

In considering now the main effects for the two-way factorial model, as in the one-way ANOVA, the sample main effect of any level *j* of the row factor A is given by $a_j = \bar{y}_{j.} - \bar{y}_{..}$, where a_j as before represents the effect for a particular row, and $\bar{y}_{j.} - \bar{y}_{..}$ represents the given row mean minus the grand mean of all observations. As in the one-way, the sum of the fixed sample main effects for factor A will be 0, $\sum_j a_j = 0$. Notice again here we are specifying the word "fixed." This is because for a fixed effects model, the sum of effects for a main effect sum to 0. However, in the following chapter, when we consider **random** and **mixed models**, we will see that this is not necessarily the case for certain factors. This will have important implications in how we construct *F*-ratios.

The sums of squares for factor A is thus $\sum_j Kn\left(a_j\right)^2$, where *K* is the number of columns, and *n* is the number of observations per cell. For the column main effect (i.e., factor B), the sample main effect is $b_k = \bar{y}_{.k} - \bar{y}_{..}$, where $\bar{y}_{.k}$ is the sample mean corresponding to a particular column *k*. As with the sample effects for a_j, the sum of the column sample effects, b_k, will also be 0, $\sum_k b_k = 0$. The sums of squares for factor B is thus $\sum_k Jn(b_k)^2$, where *J* is the number of rows.

4.4.2 Model Assumptions: Two-Way Factorial Model

The assumptions for a two-way fixed effects analysis of variance are similar to those of the one-way analysis of variance model, only now, because we have cells in our design, these are the "groups" about which we have to make assumptions when involving the interaction term:

- $E(\varepsilon_{ijk}) = 0$, that is, the expectation of the error term is equal to 0. Note the extra subscript on e_{ijk} to reflect not only the j^{th} population but also the jk^{th} cell.
- ε_{ijk} are $NI\left(0, \sigma_e^2\right)$, that is, the errors are normally distributed and independent of one another. Just as we did for the one-way, we are using ε_{ijk} to denote the corresponding population parameter of the sample quantity e_{ijk}.
- $\sigma_{e_{ijk}}^2 < \infty$, that is, the variance of the errors is some finite number (which, as was true in the one-way model, 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).
- Measurements on the dependent variable are observed values of a random variable that are distributed about true mean values that are **fixed** constants. This is the same assumption made for the one-way model in which we were interested in the fixed effects. This assumption will be relaxed when we contemplate **random effects** models in chapters to come.

We could also add the assumption, as we did for the one-way model, that the model is correctly **specified**, in that there are reasonably no other sources acting on the dependent variable to an appreciable extent. If there were, and we did not include them in our model, we would be guilty of a **specification error** or of more generally misspecifying our model.

4.4.3 Expected Mean Squares for Factorial Design

In deriving F-ratio tests for the various effects in the two-way ANOVA, just as we did for the one-way ANOVA, we need to derive the expectations for the various sums of squares, and then divide these by the appropriate degrees of freedom to produce a mean square for the given factor or interaction. Hence the phrase, "expected mean squares." We adapt the following derivations from Hays (1994), Kempthorne (1975), and Searle, Casella, and McCulloch (1992). We begin with the expected mean squares for within cells (Hays, 1994, p. 485):

$$\begin{aligned} E(\text{SS within cells}) &= E\left[\sum_k \sum_j \sum_i \left(y_{ijk} - \bar{y}_{jk}\right)^2\right] \\ &= \sum_k \sum_j E\left[\sum_i \left(y_{ijk} - \bar{y}_{jk}\right)^2\right] \\ &= \sum_k \sum_j (n-1)\sigma_e^2 \\ &= JK(n-1)\sigma_e^2 \end{aligned} \tag{4.2}$$

Why does $\sum_k \sum_j E\left[\sum_i \left(y_{ijk} - \bar{y}_{jk}\right)^2\right]$ equal $\sum_k \sum_j (n-1)\sigma_e^2$? To understand this, recall in the one-way layout:

$$E\,(\text{SS within}) = E\left[\sum_j \sum_i \left(y_{ij} - \bar{y}_j\right)^2\right]$$

However, for any given sample group j, we know that we have to divide SS by $n - 1$ in order to get an **unbiased** estimate of the error variance. That is, we know that $E\left[\sum_j \sum_i \left(y_{ij} - \bar{y}_j\right)^2\right]$ does not "converge" to σ_e^2, but that $E\left[\sum_i \frac{\left(y_{ij} - \bar{y}_j\right)^2}{n_j - 1}\right]$ does. So, we can rearrange this slightly to get

$$E\sum_i \left(y_{ij} - \bar{y}_j\right)^2 = \left(n_j - 1\right)\sigma_e^2$$

Finally, how did we go from $\sum_k \sum_j (n-1)\sigma_e^2 = JK(n-1)\sigma_e^2$ in the final term of (4.2)? By the rules of summation, $\sum_j^J y = Jy$, and so $\sum_k^K \sum_j^J y = JKy$, in which in our case $(n-1)\sigma_e^2$ acts as "y."

Now that we have the expectation for SS error, that of $JK(n-1)\sigma_e^2$ of (4.2), let us consider what we have to divide this sum of squares by to get MS error. That is, we need to determine the **degrees of freedom for error**. Since there are $J \times K$ cells, we will lose 1 degree of freedom **per cell**, which gives us degrees of freedom = $JK(n - 1)$. So, MS error is equal to:

$$\begin{aligned} \text{MS error } &= \frac{\text{SS error}}{JK(n-1)} \\ &= \frac{JK(n-1)\sigma_e^2}{JK(n-1)} \\ &= \sigma_e^2 \end{aligned}$$

That is, as was the case in the one-way ANOVA, MS error is simply equal to the error variance alone in a two-way fixed effects ANOVA.

What about the mean square for factor A? When determining an appropriate mean square for any term, recall that it is essential to consider what goes into the numerator. For the error term, as we just saw, all that goes into the calculation of error is simply σ_e^2. When considering the effect for factor A, we need to recall that in any given row J, both the column effects b_k and the interaction effects sum to 0. That is, $\sum_k b_k = 0$ and $\sum_k (ab)_{jk} = 0$. Notice that we are summing over k columns to get the row effect. Why is this important? It is important because it tells us what we can leave out of the mean square for factor A. Because we know $\sum_k b_k = 0$ and $\sum_k (ab)_{jk} = 0$, we become aware that these terms will not be part of the mean square for factor A. If you prefer, we might say they **will** still be part of the term, but since they sum to 0, why include them in the mean square for factor A at all? Both ways of thinking about it gets us to the same place in that we do not have to incorporate them when computing our mean squares.

Recall that the sums of squares for factor A are given by

$$Kn\sum_j a_j^2 = Kn\sum_j \left(\bar{y}_{j.} - \bar{y}_{..}\right)^2$$

Given this, and the fact that $\sum_k b_k = 0$ and $\sum_k (ab)_{jk} = 0$, the expectation for MS factor A in which factor A is fixed, is:

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

In words, the expectation is equal to error variance, σ_e^2, plus a term containing variability due to factor A, $\frac{Kn\sum_j \alpha_j^2}{J-1}$. Given the expected mean square, we would like to produce an F-ratio to test the main effect for factor A of $H_0 : \alpha_j = 0$ versus $H_1 : \alpha_j \neq 0$ for at least some population as specified by the levels of factor A. If there is absolutely no effect, we will have:

$$E(\text{MS A}) = \sigma_e^2 + \frac{Kn\sum_j 0}{J-1}$$

and hence

$$E(\text{MS A}) = \sigma_e^2$$

And so it is easy to see that the following F-ratio will be a suitable one for testing the effect due to factor A:

$$F = \frac{\text{MS A}}{\text{MS error}}$$

on $J - 1$ and $JK(n - 1) = N - JK$ degrees of freedom. That is, in the two-way fixed effects analysis of variance, MS error is the correct error term for testing the effect of factor A.

A similar argument applies to the factor B mean square. Since $\sum_j a_j = 0$ and $\sum_j (ab)_{jk} = 0$, we will only expect variability due to that in columns when considering factor B, since the effects for A and interaction effects will both sum to 0 in the fixed effects model we are currently considering (they will not necessarily in random and mixed models of the following chapter). Therefore, the relevant expectation is:

$$E(\text{MS B}) = \sigma_e^2 + \frac{Jn\sum_k \beta_k^2}{K-1}$$

where similar to the case for factor A, the term $Jn\sum_k \beta_k^2$ simply comes from the derivation of the sums of squares for factor B, that of:

$$\text{SS B} = Jn\sum_k b_k^2 = Jn\sum_j (\bar{y}_{.k} - \bar{y}_{..})^2$$

Under the null hypothesis, it will be the case that $\beta_k^2 = 0$, and so we are left with σ_e^2. Hence, the appropriate F ratio is:

$$F = \frac{\text{MS B}}{\text{MS error}}$$

TABLE 4.5 ANOVA Summary Table for Two-Way Factorial Design

Source	Sums of Squares	df	Mean Squares	F
A (rows)	SS A	$J-1$	SS A/$J-1$	MS A/MS error
B (columns)	SS B	$K-1$	SS B/$K-1$	MS B/MS error
A × B	SS AB cells−SS A−SSB	$(J-1(K-1)$	SS A × B/$(J-1)(K-1)$	MS A × B/MS error
Error	SS total−(SSA + SSB + SS A × B)	$N-JK$	SS error/$(N-JK)$	
Total	SS total	$N-1$		

on $K-1$ and $JK(n-1)=N-JK$ degrees of freedom. That is, in the two-way fixed effects analysis of variance, MS error is the correct error term for testing the effect of factor B.

Finally, what of the expected mean squares for interaction? In generating the mean square, we follow a similar argument as when producing the terms for factor A and factor B. That is, we ask ourselves, **what went into the interaction term?** Well, we know that for the sample cell effect, $[ab]_{jk}$, we saw that it was composed of variability due to factor A, factor B, and the A × B interaction. What goes into the interaction term $(ab)_{jk}$ is simply variability due to an interaction between factor A and factor B. Thus, for the interaction, we have:

$$E(\text{MS interaction}) = \sigma_e^2 + \frac{n\sum_j\sum_k(\alpha\beta)_{jk}^2}{(J-1)(K-1)}$$

If the interaction effects end up being 0, that is, if $n\sum_j\sum_k(\alpha\beta)_{jk}^2=0$, then we will wind up with simply σ_e^2. Hence, the appropriate F-ratio is MS interaction/MS error on $(J-1)(K-1)$ and $JK(n-1)=N-JK$ degrees of freedom. The summary table for the two-way factorial design is given in Table 4.5.

4.4.4 Recap of Expected Mean Squares

Recall that the practical purpose behind deriving expected mean squares, whether in the one-way or higher-order ANOVA models, is to be able to generate meaningful F-ratios and test null hypotheses of interest to us. In our discussion of mean squares, we have justified the use of F-ratios for testing the main effect of A, main effect of B, and the interaction of A × B. Notice that in each case, MS error is the appropriate denominator in the **fixed effects model of analysis of variance**. When we consider **random** and **mixed effects models** in chapters to follow, we will see that, and more importantly understand why, MS error is not always the appropriate denominator for testing effects.

4.5 INTERPRETING MAIN EFFECTS IN THE PRESENCE OF INTERACTIONS

Typically, if one has found evidence for an interaction in an ANOVA, one can still interpret main effects, so long as one realizes that the main effects no longer “tell the whole story.” As noted by Kempthorne (1975, p. 483), however, “the testing of main effects in the presence of interaction, without additional input, is an exercise in fatuity.”

As an illustration, suppose the researcher investigating the effect of melatonin did find an effect, but that the drug was only truly effective in conditions of very low noise. If ambient noise is elevated, melatonin no longer reduces sleep onset time. In other words, an interaction is present. In light of this interaction, if we interpreted **by itself the effect of dosage** without also including noise level in our "story," then we would be potentially misleading the reader who may mistakenly conclude taking melatonin could help him get to sleep faster even if in a college dormitory (which is relatively noisy, even at night). **The take-home message is clear—if you have evidence for an interaction in your data, it is the interaction that should be interpreted first**. Interpreting main effects second is fine, so long as you caution your reader that they do not tell the **whole story**. The more complete story is housed in the interaction term.

4.6 EFFECT SIZE MEASURES

Recall that for the one-way fixed effects analysis of variance model, we computed

$$\eta^2 = \frac{\sum_{j}^{J} n_j \left(\bar{y}_j - \bar{y}.\right)^2}{\sum_{j=1}^{J} \sum_{i=1}^{n} \left(y_{ij} - \bar{y}.\right)^2}$$

as a measure of effect size in the sample. It revealed the proportion of variance in the dependent variable that was accounted for by knowledge of the independent variable.

In the factorial design, we can likewise compute η^2, but this time for each factor and interaction. That is, we will have, for respective main effects and interaction,

$$\eta_A^2 = \frac{\text{SS A}}{\text{SS total}} \quad \eta_B^2 = \frac{\text{SS B}}{\text{SS total}} \quad \eta_{A \cdot B}^2 = \frac{\text{SS A} \times \text{B}}{\text{SS total}}$$

Each of these, as was true for the one-way model, will give us an estimate of the variance explained in the dependent variable given the particular source of variation. As was true for the fixed effects model, these measures of η^2 are all **descriptive** measures of what is going on in the particular sample. Measures of η^2 are **biased upward**, and hence the true strength of association in the corresponding population parameters is usually less than what values of η^2 suggest.

In factorial designs, since we are modeling more than a single effect, one can also compute $\eta^2_{Partial}$, defined as:

$$\eta^2_{Partial} = \frac{\text{SS effect}}{\text{SS effect} + \text{SS error}}$$

A look at $\eta^2_{Partial}$ reveals that the denominator contains not the **total variation** as in η^2, but rather SS for the effect we are considering in addition to what is "left over" from the ANOVA in terms of error. For the one-way ANOVA, $\eta^2 = \eta^2_{Partial}$. Some authors (e.g., see Tabachnick and Fidell, 2007) recommend the reporting of $\eta^2_{Partial}$ for the reason that the size of η^2 will depend on the **complexity** of the model. That is, for a given effect, η^2 will typically be smaller in a model containing many effects than in a simpler model as a result of the total variation being larger in the former case. In the case of $\eta^2_{Partial}$, we are not allowing all of these effects to be a part of our denominator, and so $\eta^2_{Partial}$, all else equal, will be greater than η^2.

Analogous to the one-way model, ω^2 can also be computed in factorial models such that it provides a better approximation of the strength of association in the population and yields a more accurate estimate compared to η^2. Estimates of ω^2 can be obtained for both main effects and interactions, though ω^2 is less common in most software than is η^2 and $\eta^2_{Partial}$. For derivation and computation details, see Vaughan and Corballis (1969).

4.7 THREE-WAY, FOUR-WAY, AND HIGHER MODELS

The cases of three or more independent variables are a natural extension of the case for two. The only difference in terms of the partition is that in higher-order models, in addition to subtracting out SS A and SS B, etc., (depending on how many factors we have) from the cells term, we also need to subtract out all **two-way** interaction terms as well, since they are also naturally "part" of the cells term. Hence, for a three-way model, we would have:

$$\mathbf{SS\ A \times B \times C = (SS\ ABC\ cells) - (SS\ A) - (SS\ B) - (SS\ C) - (SS\ A \times B) - (SS\ A \times C) - (SS\ B \times C)}$$

This is nothing new. The principle is the same as for the two-way. Because cell terms contain all systematic effects in an experiment, we need to subtract all effects that may have "gone into" this term. This includes main effects and two-way interactions, which is why we include them in the subtraction.

4.8 SIMPLE MAIN EFFECTS

Given the presence of an interaction, the examination of **simple main effects** allows us to study the effect associated with some level of a given factor when the level of another factor is prespecified. We will usually want to perform simple effects analysis for any statistically significant interaction, and the precise number of simple effects we perform should align at least somewhat with our theoretical predictions as to not unduly inflate type I error rates (or at minimum, we could use a Bonferroni-type correction on α_{FW} to attempt to keep the family-wise error rate at a nominal level).

To understand simple main effects, we begin first by reconsidering factor A with J levels. Recall that the main effect associated with this factor in a two-way factorial model is $a_j = \bar{y}_{j\cdot} - \bar{y}_{\cdot\cdot}$. That is, the effect a_j is defined as the difference between the mean for that particular row, $\bar{y}_{j\cdot}$ and the grand mean of $\bar{y}_{\cdot\cdot}$ (Recall that the periods following the letters are simply used as "placeholders" for columns k when considering $\bar{y}_{j\cdot}$ and for rows j and columns k when considering the grand mean, $\bar{y}_{\cdot\cdot}$). In the presence of a two-way interaction, if we chose only one level k of factor B, and examined only the effects of factor A **within a given level of factor B**, each of these effects would be called **simple main effects**. They are analogously derived for column effects. They are effects (usually main effects, but as we will see, they can also be **interaction effects** in the case of a three-way or higher ANOVA) of a factor at one level of another factor. They allow us to "tease apart" an interaction to learn more about what generated the interaction in the first place.

As a visualization to better understand the concept of a simple main effect, consider once more Figure 4.1 given at the outset of this chapter, only now, with a simple main effect indicated at the level of the first teacher (Figure 4.3). It is the simple main effect of mean achievement differences on textbook at the first teacher.

We can define the simple main effect in Figure 4.3 as:

$$\bar{y}_{jk} - \bar{y}_{.k}$$

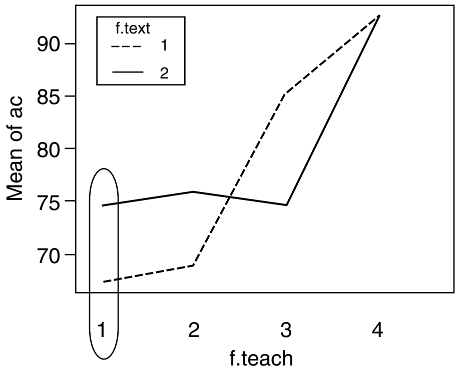


FIGURE 4.3 A simple main effect: Mean difference of textbook at level 1 of teacher.

where $\bar{y}_{jk}$ is the mean for a given textbook cell and $\bar{y}_{.k}$ is the mean for teacher 1, collapsing across textbooks. We can define a number of other simple main effects:

- textbook 1 versus textbook 2 @ teacher 2
- textbook 1 versus textbook 2 @ teacher 3
- textbook 1 versus textbook 2 @ teacher 4

We could also define simple main effects the other way (though not as easily visualized in Figure 4.3):

- teacher 1 versus teacher 2 versus teacher 3 versus teacher 4 @ textbook 1
- teacher 1 versus teacher 2 versus teacher 3 versus teacher 4 @ textbook 2

We carry out analyses of simple main effects in software toward the conclusion of the chapter, where much of this will likely make more sense in the context of a full analysis.

4.9 NESTED DESIGNS

Up to this point in the chapter, our idea of an interaction for the achievement data has implied that all teachers were **crossed** with all textbooks. The layout of 2×4 (i.e., 2 textbooks by 4 teachers) of both Table 4.1 and Figure 4.1 denotes the fact that all combinations of textbook and teacher are represented and analyzed in the ANOVA.

Nesting in experimental design occurs when **particular levels of one or more factors appear only at particular levels of a second factor**. For example, using the example of teachers and textbooks, if only teachers 1 and 2 used the first textbook but teachers 3 and 4 used the second textbook, then we would say the **factor teacher is nested within the factor textbook** (Table 4.6). These types of designs are sometimes referred to as **hierarchical designs** (e.g., see Kirk, 1995, p. 476). Though we do not consider nested designs in any detail in this book, it is important to understand how such designs (should you be confronted with one) differ from the classical factorial design in which all levels are crossed. For further details on nested designs, see Casella (2008), Kirk (1995), Mead (1988), and Montgomery (2005).

TABLE 4.6 Nested Design: Teacher is Nested Within Textbook

Textbook 1		Textbook 2	
Teacher 1	Teacher 2	Teacher 3	Teacher 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
Mean = 71.0	Mean = 72.5	Mean = 80.0	Mean = 92.7

4.9.1 Varieties of Nesting: Nesting of Levels Versus Subjects

It is well worth making another point about nesting. Recall that in our brief discussion of Chapter 3, nesting was defined as a **similarity** of objects or individuals within a given group, whether it be those women receiving mammographies or those exhibiting smoking behavior, or those children within the same classroom, classrooms within the same school, etc. It should be noted at this point that the nesting featured in Table 4.6 in relation to factor levels, other than for a trivial similarity, is not of the same kind of nesting as that of subjects within groups. The word "nesting" is used interchangeably in both circumstances, and much confusion can result from equating both designs.

To illustrate the important distinction, let us conceptualize a design in which the same subject is measured successively over time. These are so-called **repeated-measures designs**, to be discussed at some length in Chapter 6. Consider the data in Table 4.7 in which rats 1 through 6 were each measured a total of three times, once for each trial of a learning task. For this hypothetical data, rats were tested to measure the elapsed time it took to press a lever in an operant conditioning chamber. The response variable is the time (measured in minutes) it took for them to learn the lever-press response. We would expect that if learning is taking place, the time it takes to press the lever should **decrease** across trials.

In such a layout, it is often said that "trials are nested within subject" (in this case, the rats). That is, **measurements from trial 1 to 3 are more likely to be similar within a given rat than between rats**. If a rat performs poorly at trial 1, even if it improves by trials 2 and 3, we could probably still expect a relatively lowered performance overall. On the other hand, if a rat performs very well at trial 1, this information probably will tell us something about its performance at trials 2 and 3. That is, because observations occur **within rat**, we expect trials to be **correlated**.

TABLE 4.7 Learning as a Function of Trial (Hypothetical Data)

	Trial			
Rat	1	2	3	Rat Means
1	10.0	8.2	5.3	7.83
2	12.1	11.2	9.1	10.80
3	9.2	8.1	4.6	7.30
4	11.6	10.5	8.1	10.07
5	8.3	7.6	5.5	7.13
6	10.5	9.5	8.1	9.37
Trial means	$M = 10.28$	$M = 9.18$	$M = 6.78$	

This is one crucial difference when we speak of nesting. On the one hand, we have nested designs in which factor levels of one factor are nested within factor levels of a second factor. This is the nesting featured in Table 4.6. On the other hand, we have **nested measurements**, in which factor levels usually remain the same from subject to subject (or “block to block” as we will see in Chapter 6), but that several measurements are made on each subject. These two types of nesting are not quite the same. The only way the two types of nesting do converge is if we consider **subject to be simply another factor**. In hierarchical and multilevel models, for instance, we say that students are nested within classroom. But what are students? In the sense of nesting, students are but another factor of which we sample many different **levels** (i.e., many different subjects). Likewise, different classrooms have different students, and if there is more similarity among students within the same classroom than between, then we would like this similarity to be taken into account in the statistical analysis. Nesting of this sort is a characteristic of **randomized block designs** and **multilevel sampling**. We discuss this topic further when we survey random effects and mixed models in the next two chapters. For now, it is enough to understand that when the word “nesting” is used, it is important to garner more details about the design to learn exactly how it applies. Half of the battle in understanding statistical concepts is often in appreciating just how the word is being used in the given context.

4.10 ACHIEVEMENT AS A FUNCTION OF TEACHER AND TEXTBOOK: EXAMPLE OF FACTORIAL ANOVA IN R

Having surveyed the landscape of factorial analysis of variance, we now provide an example to help motivate the principles aforementioned. We once more use the hypothetical achievement data for our illustration. As discussed, instead of only randomly assigning students to one of four teachers, we also randomly assign students to one of two textbooks. We are only interested in generalizing our findings to these four teachers and these two textbooks, making the fixed effects model appropriate.

Our data of Table 4.1 appears below in R:

```
> achiev.2 <- read.table("achievement2.txt", header = T)
> achiev.2
> some(achiev.2)

   ac teach text
1  70     1    1
2  67     1    1
3  65     1    1
```

First, as usual, we identify teacher and text as factors:

```
> attach(achiev.2)
> f.teach <- factor(teach)
> f.text <- factor(text)
```

We proceed with the 2×2 factorial ANOVA:

```
> fit.factorial <- aov(ac ~ f.teach + f.text + f.teach:f.text,
data = achiev.2)
```

```
> summary(fit.factorial)

               Df Sum Sq Mean Sq F value   Pr(>F)
f.teach         3 1764.1   588.0 180.936 1.49e-12 ***
f.text          1    5.0     5.0   1.551    0.231
f.teach:f.text  3  319.8   106.6  32.799 4.57e-07 ***
Residuals      16   52.0     3.3
—
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We note that the main effect for teacher is statistically significant, while the main effect for text is not. The interaction between teacher and text is statistically significant ($p = 4.57\text{e-}07$). The identical model can be tested in SPSS (output not shown) using:

```
UNIANOVA ac BY teach text
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /CRITERIA=ALPHA(0.05)
  /DESIGN= teach text teach*text.
```

To look at means more closely, we may use the package `phia` (Rosario-Martinez, 2013), and request cell means for the model:

```
> library(phia)
> (fit.means <- interactionMeans(fit.factorial))

  f.teach f.text adjusted mean
1       1      1      67.33333
2       2      1      69.00000
3       3      1      85.33333
4       4      1      92.66667
5       1      2      74.66667
6       2      2      76.00000
7       3      2      74.66667
8       4      2      92.66667
```

We reproduce the cell means in Table 4.8.

Remember, when trying to discern whether an interaction exists, we ask ourselves the following question—**At each level of one independent variable, is the same "story" being told at each level**

TABLE 4.8 Achievement Cell Means Teacher * Textbook

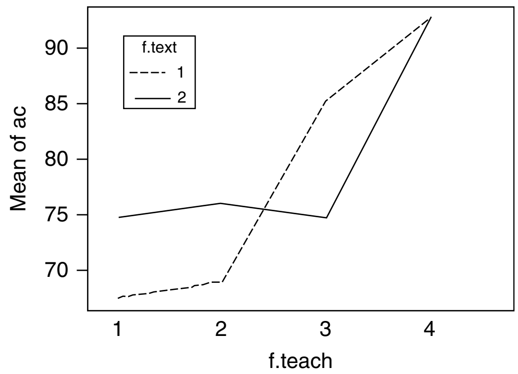
	Teacher				
Textbook	1	2	3	4	Row Means
1	$\bar{y}_{jk} = \bar{y}_{11} = 67.33$	$\bar{y}_{jk} = \bar{y}_{12} = 69.00$	$\bar{y}_{jk} = \bar{y}_{13} = 85.33$	$\bar{y}_{jk} = \bar{y}_{14} = 92.67$	$\bar{y}_{j.} = \bar{y}_{1.} = 78.58$
2	$\bar{y}_{jk} = \bar{y}_{21} = 74.67$	$\bar{y}_{jk} = \bar{y}_{22} = 76.00$	$\bar{y}_{jk} = \bar{y}_{23} = 74.67$	$\bar{y}_{jk} = \bar{y}_{24} = 92.67$	$\bar{y}_{j.} = \bar{y}_{2.} = 79.50$
Column Means	$\bar{y}_{.k} = \bar{y}_{.1} = 71.00$	$\bar{y}_{.k} = \bar{y}_{.2} = 72.5$	$\bar{y}_{.k} = \bar{y}_{.3} = 80.0$	$\bar{y}_{.k} = \bar{y}_{.4} = 92.67$	$\bar{y}_{..} = 79.04$

of the other independent variable? What such a question begs us to do is look at means at the level of one factor **conditioned** on levels of the other factor.

For example, examine the mean teacher differences at textbook 1 in Table 4.8. We note the means to be 67.33, 69.00, 85.33, and 92.67 for the first, second, third, and fourth teachers, respectively. Notice how these means represent a continuous increase from teachers one through four. This is what we mean by the "story" being told at the level of textbook = 1. The actual "story" is not the actual **values** of the means, but rather the **differences** between means. That is, the story is the **magnitude and direction on which these cell means differ**. We can see the story for textbook = 2 is similar, yet not the same as for textbook = 1 (for example, from teacher 2 to 3 denotes a mean **decrease**, not an **increase**).

Trying to discern all this in a table of cell means is quite difficult, and we are better off graphing these cell means, which we can do via an **interaction plot** in R as we did in Figure 4.1 to open this chapter. We reproduce the plot here:

```
> interaction.plot(f.teach, f.text, ac)
```



Be sure you are able to match up the interaction plot with the cell means in Table 4.8. The plot provides a much better picture of what is really going on in the achievement data than a table of numbers only could ever reveal. Is the same mean difference story of textbook differences on achievement being told at each level of teacher? The plot helps to answer such questions. It would appear from the plot that for the first and second teachers, textbook 2 is more effective than textbook 1. But for teacher 3, textbook 1 is more effective than textbook 2. That is, there is a **reversal** of means from teacher 2 to teacher 3. For teacher 4, it appears that achievement is equal regardless of which textbook is used.

Of course, visualizing mean differences in a plot is one thing and provides strong evidence for an interaction in the **sample data**. However, simply because we are seeing that mean differences of teacher across textbook are not equal is not reason in itself to reject the null hypothesis of no interaction and infer the alternative hypothesis that there is one in the **population** from which these data were drawn. We need to conduct the formal test of significance to know if rejecting the null of no interaction is warranted.

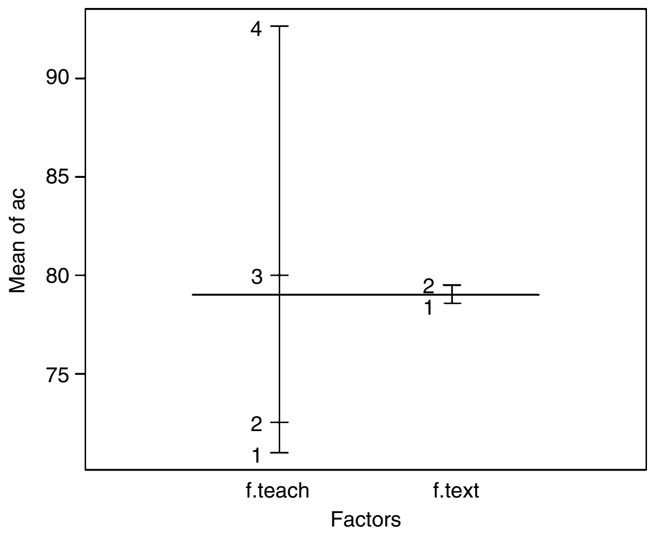
Always remember that differences and effects in sample data may not generalize to actual differences and effects in the populations from which the sample data were drawn. This is the precise point of the inferential significance test and associated p-value, to make a decision as to whether observed differences or effects potentially seen in the sample can be inferred to the population.

Recall also that **as sample size $n \to \infty$, that is, as it grows without bound, even for miniscule sample effects or sample interaction effects, statistical significance is assured**. This may make it sound like it is sample size that is dictating whether we "find something or not." And this is precisely true if we are foolish enough to consider the *p*-value as the "be all and end all" of things. As we pointed out in Chapter 2, when interpreting statistical and scientific evidence, the *p*-value should be used as only **one** indicator of the potential presence of a scientific finding. The other indicator is **effect size**.

To reiterate and emphasize, **distinguishing between statistical significance and effect size is not only a good idea, it is essential if you are to evaluate scientific evidence in an intelligent manner**. If you are of the mind that *p*-values, and *p*-values alone, should be used in the evaluation of scientific evidence, then you should not be interpreting scientific evidence in the first place. Being able to distinguish between what a *p*-value tells you and what an effect size tells you is **that** mandatory. It is not merely a preferred or "fashionable" custom, it is absolutely necessary for quality interpretation of scientific findings.

Another way to visualize the interaction is through R's `plot.design`, where we notice that means across teacher are quite disperse and means across textbook are quite close to one another:

```
> plot.design(ac ~ f.teach + f.text + f.teach:f.text, data = achiev.2)
```



The plot allows us to see the main effects for teacher and textbook. Recall, however, that in the presence of an interaction effect, it is the interaction effect that should be emphasized in interpretation, not the main effects, as these latter effects do not tell us the "whole story."

4.10.1 Comparing Models Through AIC

A model is considered nested within another model if it estimates a subset of the parameters estimated in the larger model. **Akaike's information criteria**, introduced in Chapter 2, is a useful measure when comparing the fit of nested models. It can also be used for comparing the fit of non-nested models as well, however, it is commonly used for comparing nested models. Because the main-effects-only model can be considered a model nested within the higher-order interaction model, computing AIC for each model can also give us a measure of improvement in terms of how much "better" the interaction model is relative to the main-effects-only model. We first compute AIC for the main-effects model:

```
> fit.main <- aov(ac ~ f.teach + f.text, data = achiev.2)
> AIC (fit.main)
[1] 145.8758
```

We next compute AIC for the model containing the interaction term:

```
> fit.int <- aov(ac ~ f.teach + f.text + f.teach:f.text, data = achiev.2)
> AIC (fit.int)
[1] 104.6656
```

Recall that a **decrease** in AIC values denotes an improvement in model fit. The AIC value for the main-effects-only model is 145.88, while AIC for the model containing the interaction term is 104.67, which helps statistically substantiate our obtained evidence for an interaction effect.

Collapsing across cells, the sample means for teacher are computed:

```
> library(phia)
> interactionMeans(fit.factorial, factors = "f.teach")
  f.teach adjusted mean std. error
1       1      71.00000  0.7359801
2       2      72.50000  0.7359801
3       3      80.00000  0.7359801
4       4      92.66667  0.7359801
```

As before, these means for teacher are found by summing across the means for textbook. Are there mean differences for teacher? Our **sample** definitely shows differences, and based on our obtained p-value for teacher, we also have statistical evidence to infer this conclusion to the population from which these data were drawn. Suppose we decided to **not** control for **per comparison error rate** and decided to simply run independent samples t-tests. In R, we can use the `pairwise.t.test` function and for `p.adj`, specify "none" to indicate that we are not interested in adjusting our per comparison error rate:

```
 > pairwise.t.test(ac, f.teach, p.adj = "none")

        Pairwise comparisons using t tests with pooled SD

data:  ac and f.teach

  1       2       3
2 0.5562  -       -
3 0.0018  0.0072  -
4 3.4e-08 1.1e-07 6.1e-05

P value adjustment method: none
```

What is reported in the table are the *p*-values associated with the pairwise differences. We note the *p*-value for comparing teacher 1 to teacher 2 is equal to 0.5562, which is not statistically significant at the 0.05 level. The *p*-value for comparing teacher 1 to teacher 4 is equal to 3.4e-08, and hence, is statistically significant. The *p*-value for comparing teacher 2 to teacher 3 is equal to 0.0072 and is also statistically significant. The remaining *p*-values for comparing teacher 2 to 4 and 3 to 4 are likewise very small and hence the differences are statistically significant.

We now perform the same comparisons, but this time using a **Bonferroni correction** to adjust the per comparison error rate. We do this by requesting `p.adj = "bonf"`:

```
> pairwise.t.test(ac, f.teach, p.adj = "bonf")

        Pairwise comparisons using t tests with pooled SD

data:  ac and f.teach

  1       2       3
2 1.00000 -       -
3 0.01095 0.04316 -
4 2.1e-07 6.4e-07 0.00036

P value adjustment method: bonferroni
```

Though we notice all pairwise differences that were statistically significant (at 0.05) without using a correction are still significant after using a Bonferroni correction, we note the increase in *p*-values for each comparison. Comparison 2 versus 3 now yields a *p*-value of 0.04316, which for instance, would no longer be statistically significant if evaluated at the 0.01 level of significance. This is because the Bonferroni, through its adjustment of the significance level for each comparison, is making it a bit "harder" to reject null hypotheses in an effort to keep the overall type I error rate across comparisons at a nominal level.

We can also obtain means for the textbook factor:

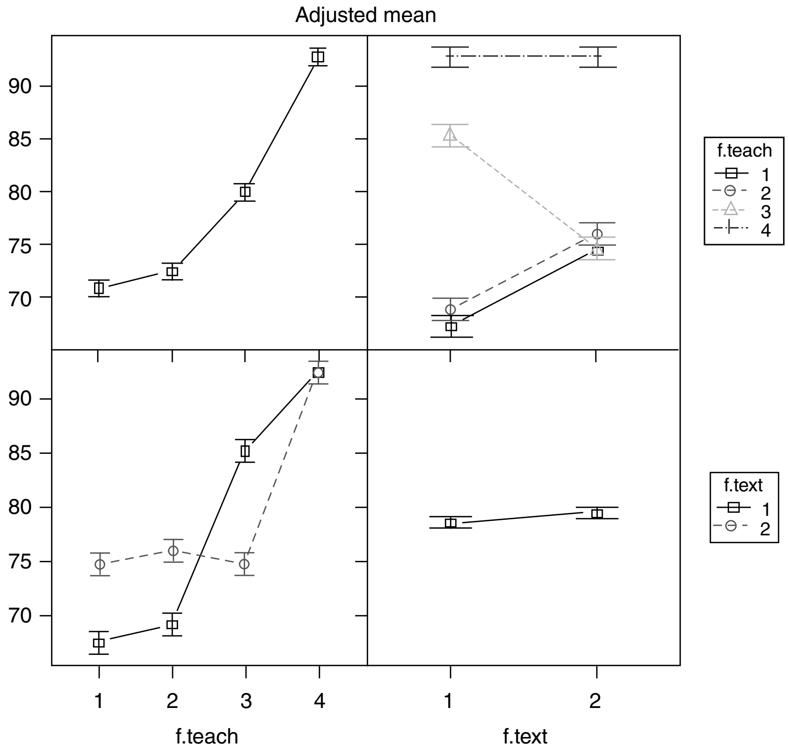
```
> library(phia)
> interactionMeans(fit.factorial, factors = "f.text")
  f.text adjusted mean std. error
1      1      78.58333  0.5204165
2      2      79.50000  0.5204165
```

Since there are only two levels to the textbook factor, conducting a post-hoc test on it makes no sense. There is no type I error to adjust since there is only a single comparison. The problem of "multiple comparisons" does not exist.

4.10.2 Visualizing Main Effects and Interaction Effects Simultaneously

A very nice utility in the `phia` package is its ability to generate a graph for which one can visualize both main effects and potential interaction effects simultaneously. We obtain this with `plot (fit.means)`:

```
> library(phia)
> plot(fit.means)
```



In the quadrants running from top left to lower right are shown the main effects for teacher and textbook, respectively. In the quadrants running from top right to lower left are shown the sample interaction effects for teacher∗textbook. Both of the interaction graphs are yielding the same essential information but in the one case (lower left), teacher is plotted on the *x*-axis while in the other (upper right), textbook is plotted on the *x*-axis. In both graphs, an interaction effect is evident.

4.10.3 Simple Main Effects for Achievement Data: Breaking Down Interaction Effects

Recall that the purpose of conducting simple main effects is to break an interaction effect down into components to better understand it, to learn what is promoting there to be an interaction in the first place. They are essentially reductions of the sample space in order to zero in on analyses that tease apart the interaction effect.

Ideally, a researcher should usually only test the simple main effects of **theoretical** or **substantive** interest. Otherwise, the exercise becomes not one of scientific hypothesis-testing but rather one of data-mining and exploration (and potentially, "fishing"). Data mining and exploration are not "bad" things by any means, only be aware that if you do "exploit" your data, you increase the risk of committing inferences that may turn out to be wrong if replication (or cross-validation) is not performed. If you do decide to test numerous simple main effects, then using a correction on the type I error rate (e.g., Bonferroni) is advised. At minimum, you owe it to your audience to tell them which findings resulted from your predictions, and which were stumbled upon in exploratory searches. From a scientific perspective, especially when working with messy high-variability data, the two are not one and the same.

We evaluate mean differences of textbook across teacher:

```
> library(phia)
> testInteractions(fit.factorial, fixed = "f.teach", across = "f.text")

F Test:
P-value adjustment method: holm
            Value Df Sum of Sq      F    Pr(>F)
1         -7.3333  1    80.667 24.820 0.0004071 ***
2         -7.0000  1    73.500 22.615 0.0004299 ***
3         10.6667  1   170.667 52.513 7.809e-06 ***
4          0.0000  1     0.000  0.000 1.0000000
Residuals         16    52.000
—
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R generates the **Holm test**, which is a multistage test, similar in spirit to the Bonferroni, but in splitting up α per comparisons c, adjusts c depending on the number of null hypotheses remaining to be tested (see Howell, 2002, pp. 386–387 for details). The Holm test is thus generally more powerful than the Bonferroni. The value of the first contrast is the mean difference between textbook 1 versus textbook 2 at teacher 1 (i.e., 67.33 – 74.67 = –7.33), and is statistically significant. The value of the second contrast is the mean difference between textbook 1 versus textbook 2 at teacher 2 (i.e., 69.00 – 76.00 = –7.00), also statistically significant. The third contrast is the mean difference between textbook 1 versus textbook 2 at teacher 3 (i.e., 85.33 – 74.67 = 10.67), and the fourth contrast is the mean difference between textbook 1 versus textbook 2 at teacher 4 (i.e., 92.67 – 92.67 = 0.00). The last of these, of course, is not statistically significant.

Simple main effects of text differences within each teacher can also be tested in SPSS using:

```
UNIANOVA
ac BY teach text
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/EMMEANS = TABLES(teach*text) COMPARE (text) ADJ(BONFERRONI)
/CRITERIA = ALPHA(.05)
/DESIGN = teach text teach*text.
```

One could also test for corresponding teacher differences within each textbook by adjusting the above code appropriately (i.e., `COMPARE (teach)`).

4.11 INTERACTION CONTRASTS

Whereas simple main effects analyze mean differences on one factor at a single level of another factor, **interaction contrasts** constitute a comparison, not of means, but of **mean differences** (i.e., a contrast of contrasts). That is, they compare a **mean difference on one factor to a mean difference on a second factor**. We can obtain values for all interaction contrasts in one large set:

```
> testInteractions(fit.factorial)
```

```
F Test:
P-value adjustment method: holm
             Value Df Sum of Sq       F    Pr(>F)
1-2 : 1-2  -0.3333  1     0.083  0.0256 0.8747843
1-3 : 1-2 -18.0000  1   243.000 74.7692 1.196e-06 ***
1-4 : 1-2  -7.3333  1    40.333 12.4103 0.0084723 **
2-3 : 1-2 -17.6667  1   234.083 72.0256 1.278e-06 ***
2-4 : 1-2  -7.0000  1    36.750 11.3077 0.0084723 **
3-4 : 1-2  10.6667  1    85.333 26.2564 0.0004079 ***
Residuals          16    52.000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The value of the first contrast is the difference between mean differences teacher 1 and teacher 2 for textbook 1 (67.33 – 69.00 = –1.67) and teacher 1 and teacher 2 for textbook 2 (74.67 – 76.00 = –1.33). That is, it is the difference –1.67 – (–1.33) = –0.33. This comparison is not statistically significant (p = 0.87). The value of the second contrast is the difference between mean differences teacher 1 versus teacher 3 for textbook 1 (67.33 – 85.33 = –18.00) and teacher 1 versus teacher 3 for textbook 2 (74.67 – 74.67 = 0). That is, it is the difference –18.00 – 0 = –18.00. This comparison is statistically significant (p = 1.196e-06). Remaining contrasts are interpreted in an analogous fashion.

4.12 CHAPTER SUMMARY AND HIGHLIGHTS

- **Factorial analysis of variance** is a suitable statistical method to test both **main effects** and **interactions** in a model where the dependent variable is continuous and the independent variables are categorical.
- The benefit of using factorial ANOVA over separate one-way ANOVAs is the ability to test for **interactions** between factors.
- Whereas **sample effects** constituted the basis of the one-way ANOVA model, **sample cell effects** constitute the systematic variation in the factorial ANOVA model.
- **Interaction effects** are computed by subtracting row and column effects from the cell effect.
- It is important to understand that **cell effects** are not equal to **interaction effects**. Rather, cell effects are used in the computation of interaction effects.
- Just as was true in the one-way model, the **error term** ε_{ijk} accounts for variability not explained by effects in the model. In the case of a two-way factorial, the error term corresponds to **within-cell** unexplained variation.
- A comparison of the **one-way model** to the **two-way model** is useful so that one can appreciate the conceptual similarities between **sample effects** and **cell effects**.
- In a **two-way model**, the sums of squares for cells partition into row, column, and interaction effects.
- The **assumptions** of the two-way factorial model parallel those of the one-way model, except that now, errors ε_{ijk} are distributed within cells, hence the requirement of the additional subscript k.
- **Expected mean squares** for factors A, B, and A × B reveal that MS error is a suitable denominator for all F-ratios.
- Interpreting **main effects** in the presence of **interaction effects** is permissible so long as one is clear to the fact that an interaction was also detected. Ideally, interaction terms should be interpreted before any main effect findings are discussed.

- A suitable **effect size** measure for terms in a factorial model is η^2, though it suffers from similar problems in the factorial model as it does in the one-way model. For a less biased estimate, ω^2 is usually recommended.
- A **simple main effect** is the effect of one factor at a particular level of another factor. Simple main effects are useful in following up a statistically significant interaction effect.
- **Interaction contrasts** can also be tested in factorial designs. These are comparisons of mean differences on one factor to mean differences on a second factor. They are "contrasts of contrasts."
- Factorial analysis of variance can be very easily performed using R or SPSS. Using the `phia` package in R, one can generate useful interaction graphs to aid in the interpretation of findings.

REVIEW EXERCISES

4.1. Define what is meant by a **factorial analysis of variance**, and discuss the purpose(s) of conducting a factorial ANOVA.

4.2. Explain, in general, what are meant by **main effects** and **interaction effects** in factorial ANOVA.

4.3. Invent a research scenario where a **two-way factorial ANOVA** would be a useful and appropriate model.

4.4. In a **two-way factorial ANOVA**, explain the three reasons why a given randomly sampled data point might differ from the **grand mean** of all the data.

4.5. Define what is meant by a **cell effect**, and why summing cell effects will always result in a **sum of zero**. What do we do to cell effects so that they do not sum to zero for every data set?

4.6. Define an **interaction effect**.

4.7. What is the difference between a **cell effect** and an **interaction effect**?

4.8. To help make the conceptual link between the **one-way model** and the **two-way**, why is it permissible (and perhaps helpful) to think of a_j as cell effects in $y_{ij} = \bar{y}. + a_j + e_{ij}$? Explain.

4.9. What is the **expected mean squares** for **MS within** in the two-factor model? Does this expectation differ from the one-way model? Why or why not?

4.10. What are the **expected mean squares** for **factor A** and **factor B** in the two-way factorial model? How do these compare to the expectations for the one-way model?

4.11. What is the **expected mean squares** for the **interaction term** in the two-way model? Under the null hypothesis of no interaction effect, what do you expect MS interaction to be?

4.12. In constructing ***F*-ratios**, what are the correct error terms for factor A, B, and A × B in the two-way model? What argument says that this is correct?

4.13. Given the presence of an interaction effect in a two-way model, **argue for** and **against** the interpretation of main effects.

4.14. Define what is meant by a **simple main effect**.

4.15. Discuss how an **interaction graph** can display a **sample interaction**, but that evidence might not exist to infer a **population interaction effect**.

4.16. Suppose a researcher wants to test all **simple main effects** in his or her data. Discuss potential problems with such an approach, and how that researcher might go about protecting against such difficulties.

4.17. In our computation of **interaction contrasts**, we interpreted two of them. Interpret the remaining interaction contrasts for the achievement analysis:

```
1-4 : 1-2  -7.3333  1    40.333 12.4103 0.0084723 **
2-3 : 1-2 -17.6667  1   234.083 72.0256 1.278e-06 ***
2-4 : 1-2  -7.0000  1    36.750 11.3077 0.0084723 **
3-4 : 1-2  10.6667  1    85.333 26.2564 0.0004079 ***
```

4.18. In our analysis of the achiev.2 data, we computed the **simple main effects** of textbook across teacher. Compute and interpret the simple main effects of teacher across textbook.

4.19. One way to conceptualize the testing of an interaction effect in ANOVA is to compare **nested models**. Recall a model is considered nested within another if it estimates a subset of parameters of the first model. For the achiev.2 data, though the significance test for interaction indicated the presence of an interaction, compare the **main-effects-only models** to that of the model containing an **interaction term** through the following:

(a) Test the main-effects-only model for teacher. Name the object `main.effects.teacher` in R.

(b) Test the main-effects-only model for teacher and textbook. Name the object `main.effects.textbook` in R.

(c) Test the interaction model. Name the object `interaction.effect` in R.

(d) Compare the models in R using: `anova(main.effects.teacher, main.effects.textbook, interaction.effect)`. Was adding the textbook and interaction effect worth it to the model?