Factorial ANOVA for Mixed Designs

Purpose

As we have seen, ANOVA can be used to test between-subjects differences as well within-subjects (repeated-measures) differences, and the factorial ANOVA framework allows for combining these two types of comparisons. A very common application is for analyzing an experimental (or a non-equivalent control group) design that has a pretest and a posttest. Such a design is called a “mixed factorial ANOVA” because it is a mix of between-subjects and within-subjects design elements. For such a 2 × 2 mixed design, the main effect for the between-subjects factor compares the two groups overall, combining pretest and posttest scores. The main effect for the within-subjects factor compares pretest and posttest scores, combining the two groups. The interaction provides information about whether the change from pretest to posttest differs in the two comparison groups. Note that because a comparison of two matched (correlated) scores in a paired t test, and, thus, a one-way within-subjects ANOVA, is testing whether the average difference score, say $\bar{Y}_{diff} = \bar{Y}_{post} - \bar{Y}_{pre}$, is greater or less than 0, the mixed factorial interaction represents a comparison of whether the difference score mean differs across the two groups. Therefore, a statistically equivalent test is to compare the difference scores in the two between-subjects groups with a independent-samples t test or between subjects one-way ANOVA (i.e., is $\bar{Y}_{diff1}$ equal to $\bar{Y}_{diff2}$).

Notation

In the following hypothetical example, I examine the effects of the educational context on vocabulary in 5th grade students. Vocabulary (number of words correct on a vocabulary test) before and after the lecture (Pre and Post) is compared for three lecture types (physical science, social science, history). Thus, the design is a 3 × 2 factorial design where Lecture Type is a between-subjects factor and Time (pre/post) is a within-subjects factor. In the following table, I represent the between-subjects factor, Lecture Type, as Factor A, and the within-subjects factor, Time, as Factor B to illustrate the design and notation.

Notation: Data Matrix

Physical science ($a1$)

	Pre (b_1)	Post (b_2)	
s1	Y_{ijk}	Y_{ijk}	$\bar{Y}_{11.}$
s2	Y_{ijk}	Y_{ijk}	$\bar{Y}_{21.}$
s3	Y_{ijk}	Y_{ijk}	$\bar{Y}_{31.}$
s4	Y_{ijk}	Y_{ijk}	$\bar{Y}_{41.}$
s5	Y_{ijk}	Y_{ijk}	$\bar{Y}_{51.}$
s6	Y_{ijk}	Y_{ijk}	$\bar{Y}_{61.}$
	$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.1.}$

Social science ($a2$)

	Pre (b_1)	Post (b_2)	
s7	Y_{ijk}	Y_{ijk}	$\bar{Y}_{72.}$
s8	Y_{ijk}	Y_{ijk}	$\bar{Y}_{82.}$
s9	Y_{ijk}	Y_{ijk}	$\bar{Y}_{92.}$
s10	Y_{ijk}	Y_{ijk}	$\bar{Y}_{10\,2.}$
s11	Y_{ijk}	Y_{ijk}	$\bar{Y}_{11\,2.}$
s12	Y_{ijk}	Y_{ijk}	$\bar{Y}_{12\,2.}$
	$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.2.}$

History ($a3$)

	Pre (b_1)	Post (b_2)	
s13	Y_{ijk}	Y_{ijk}	$\bar{Y}_{13\,3.}$

s14	Y_{ijk}	Y_{ijk}	$\bar{Y}_{14\,3.}$
s15	Y_{ijk}	Y_{ijk}	$\bar{Y}_{15\,3.}$
s16	Y_{ijk}	Y_{ijk}	$\bar{Y}_{16\,3.}$
s17	Y_{ijk}	Y_{ijk}	$\bar{Y}_{17\,3.}$
s18	Y_{ijk}	Y_{ijk}	$\bar{Y}_{18\,3.}$
	$\bar{Y}_{.3\,1}$	$\bar{Y}_{.3\,2}$	$\bar{Y}_{.3.}$

Notation: Summary of Means

		B Pre (b_1)	Post (b_2)	
	Physical science (a_1)	$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.1.}$
A	**Social science (a_2)**	$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.2.}$
	History (a_3)	$\bar{Y}_{.31}$	$\bar{Y}_{.32}$	$\bar{Y}_{.3.}$
		$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{...}$

Factorial ANOVA—Test of Main Effects and Interaction

The interpretation and general procedures for testing the main effects and the interaction are the same in the mixed factorial as they are in the between-subjects factorial ANOVA. Different error terms, however, are used for the test of the between-subjects main effect and the within-subjects main effect. *Note that the error term used to test the interaction is the same as the error term used to test the within-subjects main effect*. The table below summarizes the overall analysis. Y_{ijk} is an individual vocabulary score, $\bar{Y}_{.jk}$ is a cell mean, $\bar{Y}_{.j.}$ is a marginal mean for a level of the lecture factor, $\bar{Y}_{..k}$ is a marginal mean for the pretest or posttest, $\bar{Y}_{i..}$ is a mean for the individual student, and $\bar{Y}_{...}$ is the grand mean.

Description	***SS*** (definitional formula)	***df***	***MS***	***F***
A main effect (between-subjects)	$SS_A = (b)(n)\sum\left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2$	$a-1$	$\frac{SS_A}{df_A}$	$\frac{MS_A}{MS_{S/A}}$
Error term for *A* main effect	$SS_{S/A} = b\sum\sum\left(\bar{Y}_{.jk} - \bar{Y}_{.j.}\right)^2$	$a(n-1)$	$\frac{SS_{s/A}}{df_{s/A}}$	
B main effect (within-subjects)	$SS_B = (a)(n)\sum\sum\left(\bar{Y}_{..k} - Y_{...}\right)^2$	$b-1$	$\frac{SS_B}{df_B}$	$\frac{MS_B}{MS_{BxS/A}}$
Interaction	$SS_{AxB} = n\sum\sum\sum\left(\bar{Y}_{.jk} - \bar{Y}_{.j.} - \bar{Y}_{..k} + \bar{Y}_{...}\right)^2$	$(a-1)(b-1)$	$\frac{SS_{AxB}}{df_{AxB}}$	$\frac{MS_{AxB}}{MS_{BxS/A}}$
Error term for *B* main effect and interaction	$SS_{BxS/A} = \sum\sum\sum\left(Y_{ijk} - \bar{Y}_{.jk} - \bar{Y}_{i..} - \bar{Y}_{.j.}\right)^2$	$a(b-1)(n-1)$	$\frac{SS_{BxS/A}}{df_{BxS/A}}$	
Total	$SS_T = \sum\sum\sum\left(Y_{ijk} - \bar{Y}_{...}\right)^2$	$(a)(b)(n)-1$		

If the interaction is not significant, the main effects are examined. If the interaction is significant, the simple effects should be examined using overall simple effects tests or simple contrasts (i.e., single *df*-contrasts). When examining simple effects, however, there are a few complications, because the within-subjects factor requires different error terms for different situations.

Between-Subjects Main Effects

Assuming there is no significant interaction, one can examine the between-subjects main effect (i.e., differences in lecture types). In the above example, we would be examining the variation of the mean for each level of A ($\bar{Y}_{.1.}$, $\bar{Y}_{.2.}$, $\bar{Y}_{.3.}$) around the total mean ($\bar{Y}_{...}$). This is exactly the same as the procedure for the between-subjects one-way ANOVA.

		B Pre (b_1)	Post (b_2)	
	Physical science (a_1)	$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.1.}$
A	Social science (a_2)	$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.2.}$
	History (a_3)	$\bar{Y}_{.31}$	$\bar{Y}_{.32}$	$\bar{Y}_{.3.}$
		$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{...}$

If there are more than two levels of A, as there are in this example, one could follow the main effect analysis with a single-df contrast to compare particular marginal means for lecture types. For example, one might compare the marginal means for social science and history. These analyses examine the overall effect of A, collapsing or averaging across the levels of B. In this example, we average the pretest and posttest measures. Note that averaging the pre and post may not always be very meaningful or useful.

		B Pre (b_1)	Post (b_2)	
	Physical science (a_1)	$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.1.}$
A	Social science (a_2)	$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.2.}$
	History (a_3)	$\bar{Y}_{.31}$	$\bar{Y}_{.32}$	$\bar{Y}_{.3.}$
		$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{...}$

Within-Subjects Main Effects
Assuming there is no significant interaction, one can examine the main effect for B, a within-subjects factor. Here, we average all three of the lecture types conditions to see if there is an overall increase in the number of words correct from the pretest to the posttest.

		B Pre (b_1)	Post (b_2)	
	Physical science (a_1)	$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.1.}$
A	Social science (a_2)	$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.2.}$
	History (a_3)	$\bar{Y}_{.31}$	$\bar{Y}_{.32}$	$\bar{Y}_{.3.}$
		$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{...}$

There are only two levels of Factor B in this example, so the single-df contrast is the same as the main effect. If there were more than two levels of Factor B, however, we could compare just two of the conditions in a single-df contrast. This proceeds just as it does in the within-subjects one-way ANOVA, in which only two of the conditions are compared using the error term for two conditions only (i.e., a paired t-test or within-subjects F-test for two repeated measures).

Follow-up Analyses to a Significant Interaction
Assuming the interaction is significant, one would want to examine the simple effects. There are two approaches, the overall simple effect which compares all means within a level of one of the independent variables, and the simple contrast (or "comparison").

Simple effect for the between-subjects factor. To compute the simple effects within each level of the within-subjects factor, we are comparing groups between-subjects. In this example, we might compare all three of the lecture conditions (physical science, social science, and history) for the pretest. To do this, we just pretend that we have a one-way design with the number of vocabulary words before the lecture (pretest) as the dependent variable. We forget about the posttest for that analysis. Thus, using SPSS, *we would just run a one-way between-subjects ANOVA to test A at $b1$ or A at $b2$*.

Simple contrast for the between-subjects factor. In this analysis, we would be interested in comparing two means (e.g., a_2 vs. a_3) within one level of the B factor. For example, we might compare Social Science ($\bar{Y}_{.22}$) with History ($\bar{Y}_{.23}$) for the Posttest only. This contrast is conducted just as it is with the one-way between-subjects contrasts (using contrast weights (w_i) and Ψ). The error term used is the one used with the overall simple effect for between subjects (e.g., $MS_{s/A}$ at b_2)—*the error term from the one-way ANOVA analyzing Posttest*.

Simple effect for the within-subjects factor. In this analysis, we would be interested in comparing the means of the within-subjects factor for one of the levels of the between subjects factor (e.g. B at a_1). For example, we would compare pretest and posttest scores for Physical Sciences only. The error term used is the $B \times S$ term *only for cases in level a1*. In SPSS, one would just run a within-subjects ANOVA after selecting the students in the physical sciences lecture.

		B		
		Pre (b_1)	Post (b_2)	
	Physical science (a_1)	$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.1.}$
A	Social science (a_2)	$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.2.}$
	History (a_3)	$\bar{Y}_{.31}$	$\bar{Y}_{.32}$	$\bar{Y}_{.3.}$
		$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{...}$

Simple contrast for the within-subjects factor. In this analysis, we would compare two means of the within-subjects factor (here, the B factor) within one level of the between-subjects factor (the A factor in this example). For example, we would compare pretest and posttest for the physical sciences group. In our present example, there are just two levels of the B factor, so the simple contrast is equal to the test of the overall simple effect. (Table illustration is the same as above simple effect for B within a_1 for the two-group case).

If there were more than two levels of the B factor, we could conduct several single-*df* contrasts, comparing two means of the within-subjects factor (e.g., b_2 vs. b_3 at a_2). This analysis follows the same procedure as the contrasts with the one-way within-subjects ANOVA, *but it is done only with subjects in one of the A groups (e.g., only a_2)*. To do this, we just select out the cases for that level of the between-subjects factor (select all cases in a_2), and then conduct a within-subjects ANOVA for our two cells or a paired t-test.

		B			
		Pre (b_1)	Post (b_2)	Follow-up (b_3)	
	Physical science (a_1)	$\bar{Y}_{.11}$	$\bar{Y}_{.12}$	$\bar{Y}_{.13}$	$\bar{Y}_{.1.}$
A	Social science (a_2)	$\bar{Y}_{.21}$	$\bar{Y}_{.22}$	$\bar{Y}_{.23}$	$\bar{Y}_{.2.}$
	History (a_3)	$\bar{Y}_{.31}$	$\bar{Y}_{.32}$	$\bar{Y}_{.33}$	$\bar{Y}_{.3.}$
		$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{..3}$	$\bar{Y}_{...}$

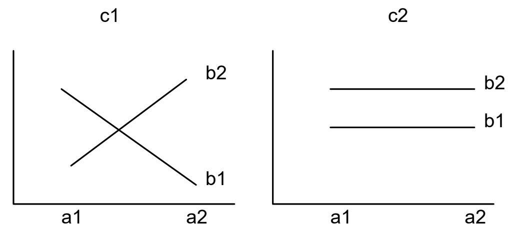
General Comments

Naturally, we could have any combination of within-subjects factors and between-subjects factors in a factorial ANOVA. A design with two or more within-subjects factors and no between-subjects factors is a "pure" within-subjects factorial. As you can imagine, the analyses for these designs become increasingly complex, but "pure" within-subjects factorials are not used too much in applied social science literature. They do occur quite frequently in laboratory studies of memory, communication, or exercise science, because researchers have more complete control over the independent variables.

When describing a mixed factorial design, researchers will often state that they have a 2 × 3 design with the first factor between-subjects and the second factor within-subjects, for example. A three-factor design might be described like this: "our design was a 2 × 2 × 4 design with the first two factors as between-subjects factors and the last factor a within-subjects factor." Such a design has two between-subjects factors with two levels each and a four-level within-subjects factor.

With three-way factorial designs, things become much more complex. We might call the third factor "C", so that a three-way design is an $A \times B \times C$ design. For example, we might examine whether two-way interaction between Lecture Type and Time is different for boys and girls (i.e., Factor C would be gender). A three-way interaction would indicate that the two-way interaction between A and B is not the same at different levels of C. This means that if C has two levels, there might be a significant interaction between A and B at C_1 but not at C_2. Or it might mean that the $A \times B$ interaction takes a different form at C_1 than it does at C_2. Of course, we also have three possible two-way interactions to deal with: $A \times B$, $A \times C$, and $B \times C$. Plus, there are possible main effects for A, B, or C.

Just for kicks, here is a graphical representation of a three-way interaction for a 2 ×2 × 2 design:



In the above example, there is a two-way interaction between *A* and *B* at *c1*, but not at *c2*. At *c2*, there seems to be just a main effect for *B*.

Further Reading

A good source for details on complex ANOVA designs is Keppel, G., & Wickens, T.D. (2004). *Design and analysis: A researcher's handbook* (4th Edition). Upper Saddle River, NJ: Prentice Hall.

A good source for testing complex ANOVA designs and conducting follow-up tests in SPSS is Page, M.C, Braver, S.L., & MacKinnon, D.P., (2003). *Levine's guide to SPSS for analysis of variance* (2nd Edition). Mahwah, NJ: Erlbaum.