Mixed-Model ANOVA: One Repeated and One Non-Repeated Factor

Thomas W. Pierce
Department of Psychology
Radford University

We've gone through an example of a case where the independent variable is manipulated within-subjects. In that case the error we were dealing with was the fact that the independent variable couldn't explain why the subjects had different patterns of change over time. In other words, we had a situation where every subject was treated exactly alike, but knowing these subjects did not all respond in the same way to the treatment. Some subject's scores went up a lot, while some subjects had scores that hardly changed at all. We discussed the fact that the amount of error present in the study can be measured by calculating the sum of squares for the interaction of the independent variable we manipulated with subjects – the AXS interaction.

<u>The mixed-design (one between-subjects independent variable and one within-subjects independent variable)</u>

Now let's say that I knew that just following my experimental group over time probably wasn't going to give me a very definitive answer about whether my treatment worked. For example, let's say that I demonstrated that the scores on the dependent variable (perceived ability to cope with stress) improved significantly over time. Why couldn't a journal editor come back and say that I don't have any evidence that the improvement that I saw couldn't have been caused just being a subject in an experiment. Maybe just having an experimenter pay attention to them was enough to elicit more positive feelings about their ability to cope.

In my original pilot study I didn't have a group of caregivers to serve as a control group. The variable I needed to control for was "being paid attention to by a experimenter". So, in a second experiment I had two groups of subjects. One group of caregivers got the same manipulation that I described last class (i.e., I measured their scores before they got the computers, 2 months after they got the computers, and 4 months after they got the computers). A second group of caregivers filled out the same questionnaires at the same points in time (i.e., at an initial testing session, two months later, and then four months later). The second group of caregivers didn't have the opportunity to participate in the electronic-mail support group.

If its just "being paid attention to" that's resulting in the scores improving, then the two group should display the same pattern of change across the three times of testing. In other words, the subjects getting the computers may show significant improvements, but I will only be able to say that its because of my support group if the amount of improvement for the caregivers with computers is a lot larger than the amount of improvement (if any) shown by caregivers without the computers.

So there we have it. There are two independent variables in this new study. Let's say that factor A refers to the two groups of caregivers (Computer Group). Factor A is a between-subjects variable because each caregiver provides data in only one of these conditions. Let's say that factor B refers to the Time of Testing (with three levels). Factor B is a within-subjects variable because each subject provides data in all three levels of this independent variable. The dependent variable is a score between one and fifteen that reflects a subject's perceived ability to cope with stress. Here are the data…

			Time of Testing (B)		
			Before b1	2 months b2	4 months b2
Caregiver Group (A)	Computers (A1)	S1	3	6	9
		S2	4	5	10
		S3	6	5	9
	No Computers (A2)	S4	7	5	3
		S5	6	4	4
		S6	8	6	2

I made the following predictions before I collected the data.

1. I thought that the experimental group would display changes over time, but that the control group would not. So I predicted that the Computer Group by Time of testing interaction would be significant.

2. I predicted that there would a significant simple effect of Time of Testing for subjects in the experimental group. Anticipating this significant simple effect I predicted that, for subjects in the experimental group, scores would increase significantly between Time 1 and Time 2. I also predicted that the scores would continue to increase between Time 2 and Time 3.

3. I predicted that no simple effect of Time of Testing would be observed for subjects in the control group.

NOW, I need to be able to test each of the three overall effects in the study (AXB, main effect for A, and main effect for B). The question is, WHAT IS THE APPROPRIATE ERROR TERM FOR TESTING AXB? FOR TESTING THE MAIN EFFECT OF THE REPEATED FACTOR? FOR TESTING THE MAIN EFFECT FOR THE NON-REPEATED FACTOR?

Jeez, I don't know. Let me think about it. Error, for a statistician is what the effect is supposed to account for, but can't. Let's start with the non-repeated factor (A).

<u>Error term for testing the main effect of the non-repeated factor</u>

Think about it. If you were testing… No, <u>really</u>, I want you to think about this… If you had never had the repeated factor in the first place. If you had only compared caregivers with and without computers on their perceived ability to cope with stress. What kind of design would you have? think, think, think, think… Well, of course. You'd simply have a design with one independent variable that is a between-subjects (non-repeated) factor. That independent variable would have three different subjects in each of two levels.

When testing the main effect of an independent variable you are pretending for the moment that you only had that one independent variable. What would the error term be in this case? think, think, think. Well, of course! The error term would be the variability of subjects scores within the groups, or S/A. This independent variable cannot explain why subjects that are all treated alike do not all have the same scores.

That's not so bad. You are probably already more familiar than is healthy with S/A as an error term.

<u>Error term for testing the main effect for the repeated factor</u>

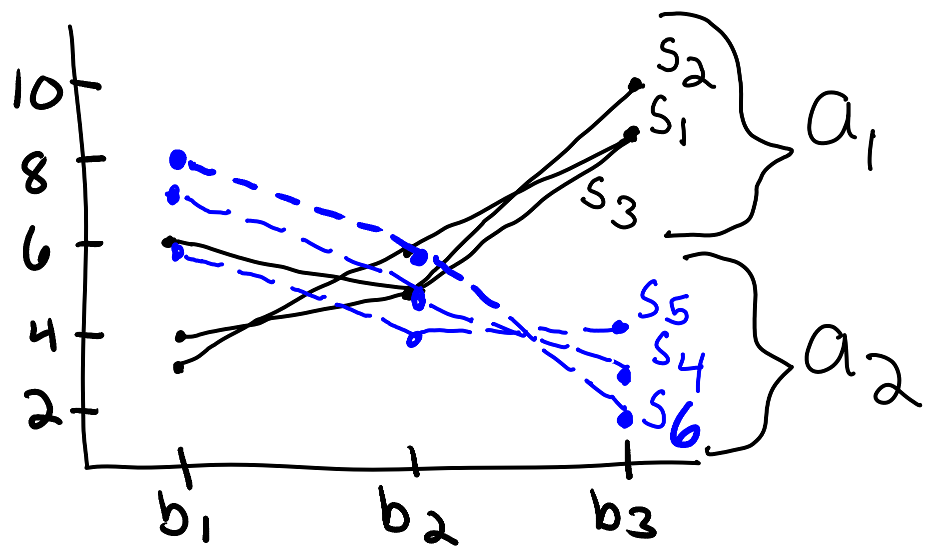
Think about it. If you had never had the non-repeated factor in the first place. You only had one independent variable that was repeated. What kind of design would this be? Obviously, it's a design with one independent variable that is repeated (within-subjects). What's the error term to test the effect of a within-subjects factor? What variability is the independent variable trying to account for but cannot?

Remember, when you only have one within-subjects factor you know that every subject was treated exactly alike when you got their scores. The independent variable is trying to explain why subject's <u>score change over time</u>. If every subject's scores changed by exactly the same amount as you look from one time of testing to the next, you would have to conclude that their was no error as far as the within-subjects factor was concerned. It would have accounted for everything that it needed to. There would have been absolutely no inconsistency in the effects of the independent variable (every subject would have displayed exactly the same pattern of change over time).

But it would be very naïve to think that this is what would actually happen in real life. It's much more likely that <u>the pattern of change in the scores across the levels of the repeated factor would not be the same for the different subjects</u>. The degree to which the subjects do not display the same pattern of change across the levels of the independent variable is measured by the interaction between the within-subjects factor and subjects. The symbol for that in this study would be BXS. So if we had never known whether the subjects had a computer or not – we just had six subjects measured at three different points in time – the error term would be BXS.

But that's not the end of the story. Take a look at the graph below. It's a graph of data from the present study. Three subjects from level A1 and three subjects from level A2 provided data. Each subject was measured three times (at B1, B2, and B3).

[Figure 1]

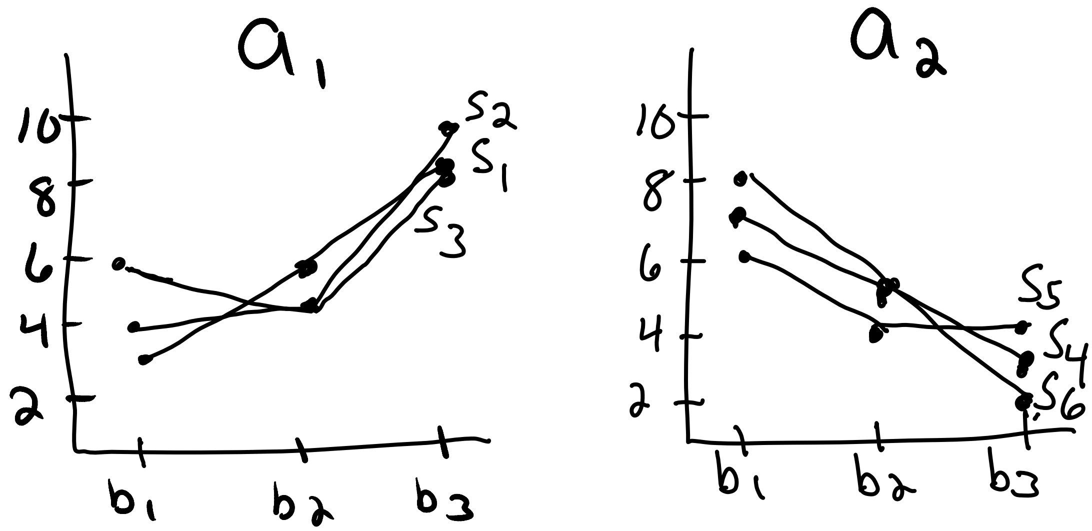


Does it look like there's an interaction between Factor B and Subjects. Of course it does, the six lines in the graph are clearly not parallel. So it looks like the error term to test the effect of the repeated factor, B, should be a large number. That means that we're not going to be in very good shape when it comes to trying to reject the null hypothesis.

But wait. We can explain some of the reason for why the six lines are not parallel. We know that the three lines from the group that got the computers (A1) are all going up at about the same rate. And we know that the three lines from the group that did not get the computers (A2) are all going down at pretty much the same rate. I can explain a big chunk of the BXS interaction. In other words, I know that if you have a computer the scores go up and if you don't have a computer the scores go down. I know what makes the three lines for A1 different from the three lines for A2. <u>Any amount of the original BXS interaction that I can explain does not have to be included as error.</u> This is because error is anything that the independent variables cannot explain.

The easiest way to think about what the error term should be is to acknowledge that what we have here are two entirely separate BXS interactions – one for each of the two levels of A. We have BXS interactions <u>within</u> each level of A. The error is not the degree to which the six lines are not parallel to each other. Because I know exactly what makes the three subjects in A1 different from the three subjects in A2, the error used to test the main effect of B is the <u>average</u> of the BXS interaction within the first level of factor A and the BXS interaction within the second level of A.

[Figure 2]



Think about it. Rate the BXS interaction at A1 on a scale of one to ten. It will be a very small number. Now rate the BXS interaction at A2. It will also be a very small number. When you average these two BXSs together you get a small number. The BXS interaction taking all six subjects (all six lines) into account would be a very large number.

The error here is reflected by the degree to which we have a BXS interaction within each level of A. The symbol for this amount of variability is **BXS/A** (BXS within each level of A).

Error term for testing the interaction between the within-subjects factor and the between-subjects factor.

The error term for testing the interaction between A and B in this study is the same error term used to test the main effect of the within-subjects independent variable. Basically, any of the overall effects that involve the within-subjects factor (B) are tested using BXS/A as the error term.

One way think of thinking about why this might make sense is to recognize that BXS/A includes the type of error found in between-subjects designs (S/A) and also the type of error found in within-subjects designs (BXS). BXS/A is a combination of the two types of error.

Simple effects involving the effect of the within-subjects factor

Let's say that the AXB interaction reaches significance at the .05 level. The appropriate thing to do at that point would be a set of simple effects. Let's say that the investigator elects to look at the effect of Time of Measurement at each level of Computer Group separately. In other words, we want to look at the effects of B at each level of A.

Factor B is a within subjects variable. There are two simple effects in this set (B at A1 and B at A2). What error term or error terms should be used to test these simple effects. Remember this general rule: Every effect involving the within-subjects factor needs its own specific error term.

When we test the effect of B at A1, the only data that are relevant to that effect are the data for the three subjects in level A1. So the appropriate error term is the amount of inconsistency in the effect of factor B within that one level of A. This would be the amount of variability in BXS, but only for the five subjects in level A1. This error term would be referred to as **BXS at A1**.

Applying this same reasoning the error term to test the effect of B at A2 would be **BXS at A2**.

Separate error terms are needed to test each of the two separate simple effects in this set. But why not use BXS/A to test each of the two simple effect of B at levels of A. Remember, BXS/A is the average BXS interaction at the different levels of A. There is no guarantee that the size of the BXS interaction at A1 will be the same as the size of the BXS interaction at A2. Because these two BXSs could be radically different from each other, the most appropriate thing to do is to use the BXS interaction for that specific effect, <u>not</u> the average of both of them.

<u>Simple comparisons involving the within-subjects factor.</u>

When a simple effect if significant the investigator should do a set of simple comparisons. Each of the simple effects of the within-subjects factor at levels of the between-subjects factor (B at each level of A) is nothing more than a one-way within-subjects ANOVA. As we discussed in this last section of the course, every comparison involving the within-subjects factor requires its own error term.

Let's say that B at a1 is significant so the investigator decides to do two simple comparisons: b1 versus b2 at A1 and b2 versus b3 at A1. The error terms for each of these two simple comparisons would consist of a BXS interaction computed only using the data that are relevant for that particular comparison. In other words the error term for the first simple comparison in this set would be calculated on a data set from the three subjects at level A1 and their scores in levels b1 and b2.

Once we get to level of simple comparisons and simple effects, everything is exactly the same as in our previous example of a one-way within-subjects ANOVA, with the exception that now the symbol for the independent variable being tested is B rather than A.

An ANOVA table for data from this study.

Source	SS	df	MS	F(obs)	F(critical)
AXB	63.0	2	31.5	23.7	4.46
B	3.0	2	1.5	1.12	4.46
BXS/A	10.67	8	1.33		
A	8.0	1	8.0	24.6	7.71
S/A	1.3	4	0.325		
B at A1	42.0	2	21.0	15.78	6.94
(BXS) at A1	5.34	4	1.33		
(b1 vs b2) at A1	1.5	1	1.5	0.75	18.5
BXS(comparison)	4.0	2	2		
(b2 vs b3) at A1	24.0	1	24	48.0	18.5
BXS(comparison)	1.0	2	0.5		
B at A2	24.0	2	12	8.88	6.96
(BXS) at A2	5.4	4	1.35		
Total	87.97	17			

Degrees of freedom for the sources of variability in the ANOVA table

Here's how one would know whether the degrees of freedom reported by an investigator are correct for this design. Start with what you know.

1. Main effect for A: levels of A – 1 = 2 –1 = 1
2. Main effect for B: levels of B – 1 = 3 – 1 = 2
3. AXB = (df for A)(df for B) = (1)(2) = 2
4. S/A = (a)(n-1) = 2(3 – 1) = (2)(2) = 4
5. BXS/A = (df for B)(df for S/A) = (2)(4) = 8. The symbol for the error term in this case – BXS/A – tells you what to do. Multiply the degrees of freedom for the main effect for B by the number of degrees of freedom for S/A.
6. B at a1: levels of B – 1 = 3 - 1 = 2
7. BXS at A1: (levels of B –1)(number of subjects at that level of A –1) = (2)(2) = 4
8. Simple comparison of (b1 vs b2) at A1: two levels for the comparison – 1 = 2 – 1 = 1
9. BXS for the simple comparison: (2 levels of B – 1)(number of subjects at that level of A – 1) = 2