Research Design - - Topic 14
MRC and Split Plot Factorial Analysis of Variance with Categorical and continuous Factors

- General Description, Purpose and Example
  (We will consider only Model I using Effect coding)
- Analysis Using MRC (Multiple Regression)
  Performing the Post hoc tests
- Analysis Using GLM (Repeated)
  Performing the Post hoc tests

1

General Description

This analysis is similar to Topics 12 and 13 but one of the factors is based on repeated measures. Thus, there will be separate error terms for the between subject and within subject variability. MRC can be used if the Between Subject Factor is categorical or continuous. Our example will focus on the case where it is continuous, though the procedure would be the same if it were categorical. As before, we will see how to do the analysis using either SPSS multiple regression or GLM repeated.

The following table presents data for a continuous centred Between Subjects factor (C) and a three level Repeated Measures factor (B).

2

A Centred Continuous Between Subjects Factor (C) and a Categorical Repeated Measures Factor (B)

C	B1	B2	B3
2	7	9	5
0	2	6	10
9	4	4	7
1	3	4	5
0	2	5	12
8	5	2	5
-8	8	5	12
-3	3	4	11
-1	4	5	10
-7	3	7	13
-1	2	6	11

3

Effect coding for the first subject

B	c	b1	b2	X	S	b1c	b2c
1	2	1	0	7	21	2	0
2	2	0	1	9	21	0	2
3	2	-1	-1	5	21	-2	-2

Note. As with the repeated measures analysis in Topic 12, a subject factor is formed as the sum of an S's scores across B as opposed to using (n-1) Subject vectors.

4

Purpose

- The analysis is concerned with assessing the:
  - **Main Effect of the C factor.** Does the mean slope differ significantly from 0?
  - **Main Effects of the B factor.** Do the intercepts for the B conditions vary more than can be reasonably attributed to chance?
  - **Interaction Effects.** Do the slopes for the B conditions differ more than can be reasonably attributed to chance?

5

To perform this analysis it is necessary to compute 7 squared multiple correlations in order to calculate the squared semi-partial multiple correlations. They are:

$$R^2_{C,B,BC} = .67476 \qquad R^2_{B,BC} = .57279$$

$$R^2_{C,B} = .59657 \qquad R^2_{C,BC} = .18016$$

$$R^2_{C,S} = .16451 \qquad R^2_{C} = .10197$$

$$R^2_{B,BC,S} = .73730$$

6

Computing squared semi-partial multiple correlations:

Effects of Interest

$$\hat{R}^2_{C} = R^2_{C,B,BC} - R^2_{B,BC} = .67476 - .57279 = .10197$$

$$\hat{R}^2_{B} = R^2_{C,B,BC} - R^2_{C,BC} = .67476 - .18016 = .49460$$

$$\hat{R}^2_{BC} = R^2_{C,B,BC} - R^2_{C,B} = .67476 - .59657 = .07819$$

Error Terms

$$\hat{R}^2_{S/C} = R^2_{C,S} - R^2_{C} = .16451 - .10197 = .06254$$

$$\hat{R}^2_{BS/C} = 1 - R^2_{B,BC,S} = 1.000 - .73730 = .26270$$

7

Computing the F-ratios

Between Subjects Factor

$$F_C = \frac{\hat{R}^2_C/1}{\hat{R}^2_{S/C}/v_1} = \frac{.10197/1}{.06254/9} = 14.674 \quad p < .01$$

Within Subjects Factor

$$F_B = \frac{\hat{R}^2_B/2}{\hat{R}^2_{BS/C}/v_2} = \frac{.49460/2}{.26270/18} = 16.945 \quad p < .01$$

$$F_{BC} = \frac{\hat{R}^2_{BC}/2}{\hat{R}^2_{BS/C}/v_2} = \frac{.07819/2}{.26270/18} = 2.680 \quad ns$$

8

Interpreting the F-ratios

- The significant F-ratio for C indicates that the mean slope differs significantly from 0.
- The significant F-ratio for B indicates that the intercepts for the three levels of B differ more than can be reasonably attributed to chance. Post hoc tests would involve comparing intercepts across the levels of B.
- If the F-ratio for the interaction was significant that would indicate that the three slopes vary more than can be reasonably attributed to chance. If it were significant, post hoc tests would involve comparing slopes across the levels of B.

9

Computing Sums of Squares

The total Sum of Squares is:

$$SS_{Total} = \sum (X_{abi} - \overline{G})^2 = 336.72736$$

The sums of squares for the effects are:

$$SS_C = \hat{R}_C^2 \, SS_{Total} = (.10197)(336.72736) = 34.336$$

$$SS_B = \hat{R}_B^2 \, SS_{Total} = (.49460)(336.72736) = 166.545$$

$$SS_{BC} = \hat{R}_{BC}^2 \, SS_{Total} = (.07819)(336.72736) = 26.329$$

$$SS_{S/C} = \hat{R}_{S/C}^2 \, SS_{Total} = (.06254)(336.72736) = 21.059$$

$$SS_{BS/C} = \hat{R}_{BS/C}^2 \, SS_{Total} = (.26270)(336.72736) = 88.458$$

10

Regression Coefficients for the Effects of Interest

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	6.09091	0.35059		17.37317	0.00000
	C	-.20438	0.07025	-0.31933	-2.90947	0.00716
	b1	-2.18182	0.49581	-0.55769	-4.40048	0.00015
	b2	-.90909	0.49581	-0.23237	-1.83354	0.07777
	b1c	.17883	0.09934	0.22814	1.80014	0.08302
	b2c	.06569	0.09934	0.08381	0.66128	0.51404

a. Dependent Variable: X

Note. The regression equation for the main and interaction effects does not include the Subject vectors. It can be used to compute the slopes and intercepts for each level of the repeated measures factor.

11

Table of Intercepts and Slopes

$$INTERCEPT = 6.09091 + (-2.18182)B1 + (-.90909)B2$$

$$SLOPE = (-.20438) + (.17883)B1 + (.06569)B2$$

	b1	b2	b3	means
Intercept	3.90909	5.18182	9.18182	6.09091
Slope	-.02555	-.13869	-.44890	-.20438

12

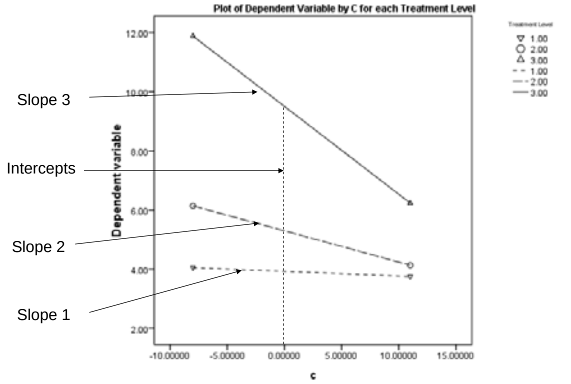
The Precise Meaning of the Regression Coefficients from Effect Coding

Constant:	B = mean intercept = 6.09091
c:	B = mean slope = -.20438
b1:	B = Intercept 1 - mean intercept = -2.18182
b2:	B = Intercept 2 - mean Intercept = -.90909
b1c:	B = Slope 1 - mean slope = .17883
b2c:	B = Slope 2 - mean slope = .06569

Note that none of these values represent any measure of variation among the slopes or the intercepts. Furthermore, they do not take into account that B is based on repeated measures. In short, the tests of significance of the regression coefficients are not appropriate post hoc tests.

Moral: Never interpret a regression coefficient if you don't know its precise meaning.

13



14

Running the Analysis Using SPSS GLM Repeated

The analysis could also be run in SPSS GLM Repeated by defining B as a repeated measures factor and C as a covariate. In this simple case, no changes are necessary to the Syntax because SPSS GLM Repeated includes the interaction between the covariate and the repeated measures factor as a source of variation. For more complex designs, it would be necessary to add missing sources to the custom model (as discussed in earlier topics).

The next two slides present the summary table from the SPSS GLM Repeated run. Note that it contains the same values as on Slide 8 (F-ratios) and Slide 10 (Sums of Squares).

15

Summary Tables from SPSS GLM Repeated

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
b	Sphericity Assumed	166.545	2	83.273	16.945	.000
	Greenhouse-Geisser	166.545	1.707	97.580	16.945	.000
	Huynh-Feldt	166.545	2.000	83.273	16.945	.000
	Lower-bound	166.545	1.000	166.545	16.945	.003
b * c	Sphericity Assumed	26.328	2	13.164	2.679	.096
	Greenhouse-Geisser	26.328	1.707	15.426	2.679	.107
	Huynh-Feldt	26.328	2.000	13.164	2.679	.096
	Lower-bound	26.328	1.000	26.328	2.679	.136
Error(b)	Sphericity Assumed	88.459	18	4.914		
	Greenhouse-Geisser	88.459	15.361	5.759		
	Huynh-Feldt	88.459	18.000	4.914		
	Lower-bound	88.459	9.000	9.829		

16

Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1224.273	1	1224.273	523.239	.000
c	34.336	1	34.336	14.675	.004
Error	21.058	9	2.340		

The following table presents the Parameter Estimates yielded by SPSS GLM if requested. Note that the values in the table are in fact the intercepts and slopes presented in slide 12. They are not the regression coefficients you would obtain if you were to use dummy coding and multiple regression. I show those values (obtained with multiple regression) on the next slide.

Parameter Estimates

Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
b1	Intercept	3.90909	0.64142	6.094	.000	2.45810	5.36008
	c	-0.02555	0.12852	-.199	.847	-0.31627	0.26518
b2	Intercept	5.18182	0.53528	9.681	.000	3.97092	6.39271
	c	-0.13869	0.10725	-1.293	.228	-0.38131	0.10393
b3	Intercept	9.18182	0.63898	14.370	.000	7.73635	10.62729
	c	-0.44891	0.12803	-3.506	.007	-0.73853	-0.15928

Moral: Beware the regression coefficient you don't know explicitly.

These are the regression coefficients that are obtained if you use dummy coding and run the full model in multiple regression. As before these can be used to compute the intercepts and slopes using standard dummy coding with group 3 given all 0's. If so, you will obtain the values on Slide 12.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	9.18182	0.60724		15.12046	0.00000
	C	-.44891	0.12167	-0.70138	-3.68952	0.00100
	d1	-5.27273	0.85877	-0.77812	-6.13983	0.00000
	d2	-4.00000	0.85877	-0.59030	-4.65780	0.00008
	cd1	.42336	0.17207	0.38189	2.46041	0.02056
	cd2	.31022	0.17207	0.27984	1.80289	0.08256

a. Dependent Variable: X

The Precise Meaning of the Regression Coefficients from Dummy Coding

Constant:	B = Intercept 3 = 9.18182
c:	B = Slope 3 = -.44891
d1:	B = Intercept 1 - Intercept 3 = -5.27273
d2:	B = Intercept 2 - Intercept 3 = -4.0000
d1c:	B = Slope 1 - Slope 3 = .42336
d2c:	B = Slope 2 - Slope 3 = .31022

Note that these values refer to Condition 3 and deviations from Condition 3. The tests of significance of these coefficients in Slide 19 test hypotheses as defined here, however, they do not take into account that B is based on repeated measures. Thus, they are meaningless.

Moral: Beware the regression coefficient you don't know explicitly.

Performing Post hoc Tests

Post hoc tests with repeated measure factors are more complex than for completely randomized designs. The regression coefficients for both Effect coding and Dummy coding are not appropriate because they do not take into account that the contrasts are based on repeated measures, in addition to their other limitations. Also, the Pattern Estimates from SPSS GLM do not describe contrasts, though they do provide tests of significance of the individual intercepts or slopes from 0.

21

Example of the contrasts between B1 and B3.

The only true way of testing pairwise contrasts with repeated measures is to run the analysis with two levels at a time. This can be done with MRC and effect coding providing the data are correctly recoded, including new Subject vectors. The F-ratios for the main effect of B and the interaction are tests of the difference between the two intercepts and two slopes respectively.

An easier method is to run the data through SPSS GLM Repeated for two levels of the repeated measures factor at a time. The next slide shows the results for contrasting B1 with B3.

22

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
b1b3	Sphericity Assumed	152.909	1	152.909	22.364	.001
	Greenhouse-Geisser	152.909	1.000	152.909	22.364	.001
	Huynh-Feldt	152.909	1.000	152.909	22.364	.001
	Lower-bound	152.909	1.000	152.909	22.364	.001
b1b3 * c	Sphericity Assumed	24.555	1	24.555	3.591	.091
	Greenhouse-Geisser	24.555	1.000	24.555	3.591	.091
	Huynh-Feldt	24.555	1.000	24.555	3.591	.091
	Lower-bound	24.555	1.000	24.555	3.591	.091
Error(b1b3)	Sphericity Assumed	61.536	9	6.837		
	Greenhouse-Geisser	61.536	9.000	6.837		
	Huynh-Feldt	61.536	9.000	6.837		
	Lower-bound	61.536	9.000	6.837		

23

The significant F-ratio for b1b3 indicates that the difference between intercept 1 and intercept 3 (3.90909-9.18182) is significant. The associated t-test is the square root of 22.364 = 4.729.
The F-ratio for b1b3*c = 3.591 is not significant indicating that the difference between slope 1 and slope 3 (-.02555--.44890) is not significant. The associated t-test is 1.895. In each case, the df for the t-test is 9.

Alternatively, the F-ratios can be calculated using the error term from the full analysis. That is:

$$F_{b1b3} = \frac{152.909}{4.914} = 31.12 \quad @\ df = 1,18 \ p < .01$$

$$F_{b1b3^*c} = \frac{24.555}{4.914} = 4.997 \quad @\ df = 1,18,\ p < .05$$

24

References

Cohen, J. & Cohen, P. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (Second Edition), Chapter 11 (pp. 428-451). Hillsdale NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (Third Edition), (pp. 573-578). Hillsdale NJ: Lawrence Erlbaum.