Research Design - - Topic 12

MRC Analysis and Two Factor Designs: Completely Randomized and Repeated Measures

1. General overview

2. Completely Randomized Two Factor Designs
   - Model I
     - Effect Coding
     - Regression Equation and Means
   - Model II
     - Dummy Coding
     - Regression Equation and Means
   - Model III

3. Single Factor Repeated Measures Designs

The General Linear Model Using MRC Analysis

- **The Model** (with 2 levels of A and 3 levels of B)

$$X_i' = b_0 + b_1A_1 + b_2B_1 + b_3B_2 + b_4A_1B_1 + b_5A_1B_2$$

The general linear model is a least squares approach to the analysis of variance. For a factorial design with equal sample sizes the results obtained are identical to those obtained with the Experimental Design model. When sample sizes are not equal, different ways of expressing the general linear model will produce different models and different answers. Overall and Spiegel (1969) identified these as Model I (Unique Sums of Squares), Model II (General Experimental) and Model III (Hierarchical). They can be run on SPSS GLM Univariate by selecting SPSSTYPE3, SPSSTYPE2, and SPSSTYPE 1, respectively.

Two Factor Designs

- General Description. Two factor analysis of variance permits you to study the simultaneous effects of two factors. Consider the data for a 2X3 design, in which there are an unequal number of observations in each cell, and each level of the A factor appears in combination with each level of the B factor.

	B1	B2	B3
A1	31 33 37 29 37 31	32 36 30 33 37	38 29 36 32
A2	36 37 39 33 34	36 27 28 30	36 37 36 46 45 42

- Table of means

	B1	B2	B3	Unweighted A-means	Weighted A-means
A1	33.0	33.6	33.75	33.45	33.40
A2	35.8	30.25	40.33	35.46	36.13
Unweighted B-means	34.40	31.93	37.04	34.46	
Weighted B-means	34.27	32.11	37.70		34.77

- Questions to ask of the Data

Main Effects of A

Do the A-means vary more than you would expect on the basis of chance?

Main Effects of B

Do the B-means vary more than you would expect on the basis of chance?

Interaction Effects of A and B

Do the AB means vary from what you would expect given the values of the A-means and the B-means?

SPSS GLM Univariate Output

Tests of Between-Subjects Effects

Dependent Variable: x

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	302.533[b]	5	60.507	4.612	.004	.490	23.062	.932
Intercept	34652.977	1	34652.977	2641.624	.000	.991	2641.624	1.000
a	29.514	1	29.514	2.250	.147	.086	2.250	.302
b	121.059	2	60.529	4.614	.020	.278	9.228	.725
a * b	115.540	2	57.770	4.404	.023	.268	8.808	.703
Error	314.833	24	13.118					
Total	36879.000	30						
Corrected Total	617.367	29						

a. Computed using alpha = .05

b. R Squared = .490 (Adjusted R Squared = .384)

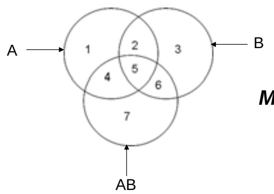
5

Effect Coding of a 2X3 (A*B) Factorial Design (Showing first observation for each cell)

A level	B level	A	B1	B2	AB1	AB2	X
1	1	1	1	0	1	0	31
1	2	1	0	1	0	1	32
1	3	1	-1	-1	-1	-1	38
2	1	-1	1	0	-1	0	36
2	2	-1	0	1	0	-1	36
2	3	-1	-1	-1	1	1	36

6

Venn Diagrams and Models in Analysis of Variance



Model I   Unique SS

SSA = 1
SSB = 3
SSAB = 7

Model II Classical Experimental

SSA = 1 + 4
SSB = 3 + 6
SSAB = 7

Model III Hierarchical

SSA = 1 + 2 + 4 + 5
SSB = 3 + 6
SSAB = 7

7

Squared multiple correlations based on the Effect Coded variables needed to compute the relevant squared multiple semipartial correlations for the three models

$R^2_{A,B,AB}$ = .49004
$R^2_{A,B}$ = .30289
$R^2_{A,AB}$ = .29395
$R^2_{B,AB}$ = .44223
R^2_{A} = .09076
R^2_{B} = .24652

Note. These two are not needed for Model 1

Note. Effect coding can be used for all three models, whereas Dummy coding can be used only for Models II and III.

8

Semipartial R² estimates for the three models

- **Model I** [1]

$$\hat{R}^2_A = R^2_{A,B,AB} - R^2_{B,AB}$$
$$\hat{R}^2_B = R^2_{A,B,AB} - R^2_{A,AB}$$
$$\hat{R}^2_{AB} = R^2_{A,B,AB} - R^2_{A,B}$$

- **Model II**

$$\hat{R}^2_A = R^2_{A,B} - R^2_B$$
$$\hat{R}^2_B = R^2_{A,B} - R^2_A$$
$$\hat{R}^2_{AB} = R^2_{A,B,AB} - R^2_{A,B}$$

- **Model III**

$$\hat{R}^2_A = R^2_A$$
$$\hat{R}^2_B = R^2_{A,B} - R^2_A$$
$$\hat{R}^2_{AB} = R^2_{A,B,AB} - R^2_{A,B}$$

1 Only Effect Coding can be used with Model I. Dummy Coding will produce wrong values for R^2_A and R^2_B.

Computing squared multiple semipartial correlations for Model I

$$\hat{R}^2_A = R^2_{A,B,AB} - R^2_{B,AB} = .49004 - .44223 = .04781$$

$$\hat{R}^2_B = R^2_{A,B,AB} - R^2_{A,AB} = .49004 - .29395 = .19609$$

$$\hat{R}^2_{AB} = R^2_{A,B,AB} - R^2_{A,B} = .49004 - .30289 = .18715$$

F-ratios for Model I

$$F_A = \frac{\hat{R}^2_A/(a-1)}{(1-\hat{R}^2_{A,B,AB})/(N-p-1)} = \frac{.04781/1}{.50996/24} = 2.250$$

$$F_B = \frac{\hat{R}^2_B/(b-1)}{(1-\hat{R}^2_{A,B,AB})/(N-p-1)} = \frac{.19609/2}{.50996/24} = 4.614$$

$$F_{AB} = \frac{\hat{R}^2_{AB}/(a-1)(b-1)}{(1-\hat{R}^2_{A,B,AB})/(N-p-1)} = \frac{.18715/2}{.50996/24} = 4.404$$

Note. The general form of the F-ratio is:

$$F = \frac{\hat{R}^2_{effect}/v_1}{(1-R^2_{total})/(N-p-1)}$$

Where: v_1 = number of vectors for the effect, N-p-1 = degrees of freedom for error, and p is the number of vectors necessary to calculate R^2_{total}.

Calculating the Analysis of Variance Summary Table

If it were desired to construct the Summary Table for the Analysis of Variance, the Sums of Squares could be calculated by multiplying the R² values by the Total Sums of Squares as follows:

$$SS_A = \hat{R}^2_A\, SS_{Total} = .04781 * 617.367 = 29.516$$

$$SS_B = \hat{R}^2_B\, SS_{Total} = .19609 * 617.367 = 121.059$$

$$SS_{AB} = \hat{R}^2_{AB}\, SS_{Total} = .18715 * 617.367 = 115.540$$

$$SS_{S/AB} = (1 - R^2_{total})\, SS_{Total} = .50996 * 617.367 = 314.833$$

Dividing by the appropriate degrees of freedom yields the Mean Squares.

Regression Coefficients and Regression Equation

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	34.456	.670		51.397	.000
	a	-1.006	.670	-.222	-1.500	.147
	b1	-.056	.922	-.010	-.060	.952
	b2	-2.531	.970	-.444	-2.608	.015
	ab1	-.394	.922	-.072	-.428	.673
	ab2	2.681	.970	.467	2.763	.011

a. Dependent Variable: x

Regression Equation

$$X'_{abi} = b_0 + b_1 A + b_2 B_1 + b_3 B_2 + b_4 AB_1 + b_5 AB_2$$

$$X'_{abi} = 34.456 + (-1.006)A + (-.056)B_1 + (-2.531)B_2 + (-.394)AB_1 + 2.681AB_2$$

This equation will produce the cell means and marginal unweighted means presented in Slide 4.

Dummy Coding of a 2X3 (A*B) Factorial Design (Showing first observation for each cell)

A level	B level	A	B1	B2	AB1	AB2	X
1	1	1	1	0	1	0	31
1	2	1	0	1	0	1	32
1	3	1	0	0	0	0	38
2	1	0	1	0	0	0	36
2	2	0	0	1	0	0	36
2	3	0	0	0	0	0	36

Squared multiple correlations based on the Dummy coded variables needed to compute the squared multiple semipartial correlations for Models II and III

$R^2_{A,B,AB}$ = .49004

$R^2_{A,B}$ = .30289

$R^2_{A,AB}$ = .09343

$R^2_{B,AB}$ = .32155

R^2_{A} = .09076

R^2_{B} = .24652

Note that $R^2_{A,AB}$ and $R^2_{B,AB}$ differ from the values obtained with Effect coding

Squared semipartial multiple correlations and Regression coefficients for Model II at Step 1

Step 1. Compute:

$$\hat{R}^2_A = R^2_{A,B} - R^2_B = .30289 - .24652 = .05637$$

$$\hat{R}^2_B = R^2_{A,B} - R^2_A = .30289 - .09076 = .21213$$

Regression Coefficients and Regression Equation

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	38.570	1.420		27.167	.000
	a	-2.176	1.501	-.240	-1.450	.159
	b1	-3.111	1.791	-.330	-1.737	.094
	b2	-5.250	1.884	-.530	-2.787	.010

a. Dependent Variable: x

$$X'_{abi} = 38.570 + (-2.176)A + (-3.111)B_1 + 5.250B_2$$

The regression equation produces the following cell means. Dummy coding does not permit calculation of marginal means, but they can be estimated as means of the cell means.

Cell Means for Step I of Model II

	B1	B2	B3	A-means
A1	33.283	31.144	36.394	33.607
A2	35.459	33.320	38.570	35.783
B-means	34.371	32.232	37.482	34.695

Note that these means do not correspond to any of the means in slide 4. They are estimated assuming no interaction. Thus, the "main" effects tested in slide 18 refer to variation in the marginal means assuming no interaction.

17

F-ratios for Main Effects for Model II

Cohen, Cohen, Aiken & West (2003, p. 171) refer to two different error terms that can be used. They define Model 1 error as the residual at step 1 (i.e., 1-$R^2_{A,B}$) and Model 2 error as the residual for the full model. The general form is:

$$F = \frac{\hat{R}^2_{effect} / v_1}{(residual\ error)/(N-p-1)}$$

where: v1 = number of vectors for the effect, N-p-1 = degrees of freedom for error, and p is the number of vectors necessary to calculate residual error

	Model 1 error	Model 2 error
F_A	$\frac{.05637/1}{(1-.30289)/26} = 2.103$	$\frac{.05637/1}{(1-.49004)/24} = 2.653$
F_B	$\frac{.21213/2}{(1-.30289)/26} = 3.956$	$\frac{.21213/2}{(1-.49004)/24} = 4.992$

18

Step 2. $\hat{R}^2_{AB}$ would be as computed in slide 9 and have the same value as in slide 10. Moreover, the F-ratio would be the same as in slide 11.

Regression coefficients for the full model using Dummy Coding

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	40.333	1.479		27.278	.000
	a	-6.583	2.338	-.726	-2.816	.010
	b1	-4.533	2.193	-.482	-2.067	.050
	b2	-10.083	2.338	-1.019	-4.313	.000
	ab1	3.783	3.206	.334	1.180	.249
	ab2	9.933	3.372	.816	2.946	.007

a. Dependent Variable: x

The regression equation is:

$$X'_{abi} = 40.333 + (-6.583)A + (-4.533)B_1 + (-10.083)B_2 + 3.783AB_1 + 9.933AB_2$$

This equation produces the cell means from slide 4. As before, marginal means cannot be computed with Dummy coding but they can be calculated as the means of the cell means (i.e., unweighted means).

19

- Model III.

  Step 1. Compute the squared multiple correlation for one of the factors (e.g., A) $\hat{R}^2_A = R^2_A = .09076$

  The two F-ratios are:

	Model 1 error	Model 2 error
F_A	$\frac{.09076/1}{(1-.09076)/28} = 2.795$	$\frac{.09076/1}{(1-.49004)/24} = 4.271$

If you were to obtain the regression coefficients and solve for the A means at this point, you would obtain the weighted A means from Slide 4.

Step 2 yields results for B identical to those from Step 1 for Model II.

Step3 yields results for AB identical to those from Step 1 for Model I and step 2 for Model II.

20

Definition of Regression Coefficients for Effect Coding and Dummy Coding

Vector	Coefficient	Effect coding	Dummy coding
Constant	b_0	$\overline{G}$	$\overline{X}_{a2b3}$
A	b_1	$\overline{X}_{a1} - \overline{G}$	$\overline{X}_{a1b3} - \overline{X}_{a2b3}$
B1	b_2	$\overline{X}_{b1} - \overline{G}$	$\overline{X}_{a2b1} - \overline{X}_{a2b3}$
B2	b_3	$\overline{X}_{b2} - \overline{G}$	$\overline{X}_{a2b2} - \overline{X}_{a2b3}$
AB1	b_4	$\overline{X}_{a1b1} - \overline{X}_{a1} - \overline{X}_{b1} + \overline{G}$	$\overline{X}_{a1b1} - \overline{X}_{a1b3} - \overline{X}_{a2b1} + \overline{X}_{a2b3}$
AB2	b_5	$\overline{X}_{a1b2} - \overline{X}_{a2} - \overline{X}_{b2} + \overline{G}$	$\overline{X}_{a1b2} - \overline{X}_{a1b3} - \overline{X}_{a2b2} + \overline{X}_{a2b3}$

21

Summary Points

The three models differ in terms of how contrasts are defined.

- **Model I** contrasts each set from all others in the study.
- **Model II** contrasts each set from others at the same and lower levels.
- **Model III** contrasts each set from others at the lower levels, and in a specified order in each set.

The type of coding does have an influence. This is discussed by Cohen, Cohen, Aiken & West (2003, p.362) who refer to them as Type III, Type II, and Type I respectively, and by Gardner (2008), who uses the above labelling. In short:

- **Effect** coding can be used for all models.
- **Dummy** coding can be used for models II and III.

22

Single Factor Repeated Measures Designs

- The Model (with 8 subjects and 4 treatments)

$$X_i' = b_0 + b_1A_1 + \ldots + b_3A_3 + b_4S_1 + \ldots b_{10}S_7 + b_{11}A_1S_1 + \ldots + b_{31}A_3S_7$$

Using the logic we used for the two factor design, this would require 3 vectors for the 4 treatment conditions, 7 for the 8 subjects, and 21 for the product terms. When using the multiple regression approach, it is not necessary to form the 21 product vectors because if this were done, we would have accounted for all the variation. In this type of analysis, the product terms are treated as residual variation.

23

An Example Using the Data from Kirk, p. 270

	A_1	A_2	A_3	A_4	$\overline{P}_i$
	3	4	4	3	3.50
	2	4	4	5	3.75
	2	3	3	6	3.50
	3	3	3	5	3.50
	1	2	4	7	3.50
	3	3	6	6	4.50
	4	4	5	10	5.75
	6	5	5	8	6.00
Means	3.00	3.50	4.25	6.25	$\overline{G} = 4.25$
Variances	2.29	0.86	1.07	4.50	$\overline{S^2} = 2.18$

Major Question to ask of the data:
Do the A-means vary more than can be reasonably attributed to chance?

24

Analysis of Variance for these data from Topic 7

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
A	Sphericity Assumed	49.000	3	16.333	11.627	.000
	Greenhouse-Geisser	49.000	1.859	26.365	11.627	.001
	Huynh-Feldt	49.000	2.503	19.578	11.627	.000
	Lower-bound	49.000	1.000	49.000	11.627	.011
Error(A)	Sphericity Assumed	29.500	21	1.405		
	Greenhouse-Geisser	29.500	13.010	2.268		
	Huynh-Feldt	29.500	17.520	1.684		
	Lower-bound	29.500	7.000	4.214		

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	578.000	1	578.000	128.444	.000
Error	31.500	7	4.500		

25

- The following slide shows the Effect coding for the first two subjects and the last subject for the sample data from Slide 24.
- **Note** there are no vectors representing the 8 subjects. Rather there is 1 Subject vector (P) which contains the sum of each subject's score on the dependent variable. Pedhazur (1977) showed that the correlation of this vector with the dependent variable was identical to the multiple correlation of the Subject vectors with the dependent variable. It is viewed as a multiple correlation based on (n-1) vectors.
- **Note too**, that there are no product vectors. They are not needed; they constitute the residual term.

26

A1	A2	A3	P	X
1	0	0	14	3
0	1	0	14	4
0	0	1	14	4
-1	-1	-1	14	3
1	0	0	15	2
0	1	0	15	4
0	0	1	15	4
-1	-1	-1	15	5
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1	0	0	24	6
0	1	0	24	5
0	0	1	24	5
-1	-1	-1	24	8

27

Relevant R^2 and F values

R^2_A = .44545 R^2_S = .28636 $R^2_{A,S}$ = .73182

$$F_A = \frac{R^2_A/(a-1)}{(1-R^2_{A,S})/(N-(a-1)-(n-1)-1} = \frac{.44545/3}{(1-.73182)/(32-3-7-1)} = 11.627$$

Note that this value is the same as that obtained in Slide 25.

Dummy coding would yield the same R^2 values because there is no product term. Of course, the regression coefficients for the Constant and A vectors would be different and in both types of coding the S vector yields the within conditions regression coefficient.

28

Calculating the Analysis of Variance Summary Table

If it were desired to construct the Summary Table for the Analysis of Variance, the Sums of Squares could be calculated for Between Subjects, A, and Residual by multiplying the R^2 values by the total Sums of Squares as follows:

$$SS_S = R_S^2\ SS_{Total} = .28636*110.0 = 31.50$$

$$SS_A = R_A^2\ SS_{Total} = .44545*110.0 = 49.00$$

$$SS_{AS} = (1 - R_{A,S}^2)\ SS_{Total} = .26818*110.0 = 29.50$$

The degrees of freedom would be (n-1) = 7, (a-1) = 3, and (a-1)(n-1) = 21, respectively.

29

References

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

Gardner, R. C. (2008). 2X2 analysis of variance and multiple regression: Coding does make a difference. Unpublished manuscript, University of Western Ontario. Available at http://publish.uwo.ca/~Gardner/DataAnalysisDotCalm/.

Overall, J.E. & Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin, 72,* 311-322.

Pedhazur, E.J. (1977). Coding subjects in repeated measures designs. *Psychological Bulletin, 84,* 298-305.

30