Research Design - - Topic 11

MRC Analysis and Single Factor Designs

1. General Overview of MRC and GLM
2. Example of a single factor design from Kirk (and Topic 3)
   Effect Coding
   Dummy Coding
3. Running SPSS Regression (Linear) using Effect coding
4. Running SPSS Regression (Linear) using Dummy coding
5. Assumptions

1

Applications and Implications

It is often said that multiple correlation can be used to identify good predictors. This is not the case. Multiple correlation does not identify predictors of a criterion. It identifies variables that add to prediction. There is a difference. Note that:

The Pearson product moment correlation between a variable and the criterion can be considered a measure of prediction. The correlation coefficient is the regression coefficient in standard score form.

The regression coefficient in multiple regression is a measure of the extent to which a variable adds to the prediction of a criterion, given the other variables in the equation. It is not a correlation coefficient.

2

General Linear Model Approach Using MRC

The Model: Analysis of variance can be seen as an instance of the general linear model.

Thus, Cohen & Cohen (1983, p. 4) state "Technically, AV/ACV and conventional multiple regression analysis are special cases of the "general linear model" in mathematical statistics. It thus follows that any data analyzable by AV/ACV may be analyzed by MRC, while the reverse is not the case".

In a more recent edition of the book, Cohen, Cohen, West & Aiken (2003, p. 4) state" The description of MRC in this book includes extensions of conventional MRC analysis to the point where it is essentially equivalent to the general linear model. It thus follows that any data analyzable by ANOVA/ANCOVA may be analyzed by MRC, whereas the reverse is not the case".

3

MRC Analysis

Cohen (1968) noted that, if group membership is defined in terms of a series of arbitrary variables (A), analysis of variance can be viewed as a special case of multiple regression. Thus, one can write a regression equation as:

$$X_i = b_o + b_1 A_1 + b_2 A_2 + ... + \varepsilon_i$$

where the number of arbitrary variables is one less than the number of treatment levels. The predicted value for each individual is the mean of the treatment condition for that individual and the A variables are codes defining the treatment levels.

4

Types of Coding: There are many types of coding. Each yield the same multiple correlation but the regression coefficients differ. We will consider two types, **Dummy Coding** and **Effect Coding**. Following are two examples involving 4 treatment levels of a factor.

Dummy Coding

Treatment	A1	A2	A3
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

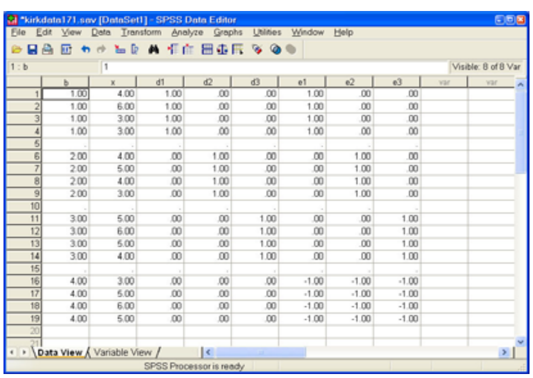
Effect Coding

Treatment	A1	A2	A3
1	1	0	0
2	0	1	0
3	0	0	1
4	-1	-1	-1

Example from Kirk (1995, p.230) used in Topic 3

	A_1	A_2	A_3	A_4	
	4	4	5	3	
	6	5	6	5	
	3	4	5	6	
	3	3	4	5	
	1	2	3	6	
	3	3	4	7	
	2	4	3	8	
	2	3	4	10	
Means	3.00	3.50	4.25	6.25	4.25
Variances	2.286	.857	1.071	4.500	2.179

Data Editor with first 4 Subjects for each treatment showing both Dummy and Effect coding



	b	x	d1	d2	d3	e1	e2	e3
1	1.00	4.00	1.00	.00	.00	1.00	.00	.00
2	1.00	6.00	1.00	.00	.00	1.00	.00	.00
3	1.00	3.00	1.00	.00	.00	1.00	.00	.00
4	1.00	3.00	1.00	.00	.00	1.00	.00	.00
5	.	.	.	.	.	.	.	.
6	2.00	4.00	.00	1.00	.00	.00	1.00	.00
7	2.00	5.00	.00	1.00	.00	.00	1.00	.00
8	2.00	4.00	.00	1.00	.00	.00	1.00	.00
9	2.00	3.00	.00	1.00	.00	.00	1.00	.00
10	.	.	.	.	.	.	.	.
11	3.00	5.00	.00	.00	1.00	.00	.00	1.00
12	3.00	6.00	.00	.00	1.00	.00	.00	1.00
13	3.00	5.00	.00	.00	1.00	.00	.00	1.00
14	3.00	4.00	.00	.00	1.00	.00	.00	1.00
15	.	.	.	.	.	.	.	.
16	4.00	3.00	.00	.00	.00	-1.00	-1.00	-1.00
17	4.00	5.00	.00	.00	.00	-1.00	-1.00	-1.00
18	4.00	6.00	.00	.00	.00	-1.00	-1.00	-1.00
19	4.00	5.00	.00	.00	.00	-1.00	-1.00	-1.00

Analysis Using Effect Coding

```
GET
  FILE='F:\PSYCH540\kirkdata171.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT x
  /METHOD=ENTER e1 e2 e3  .
```

Descriptive Statistics

	Mean	Std. Deviation	N
x	4.2500	1.88372	32
e1	.0000	.71842	32
e2	.0000	.71842	32
e3	.0000	.71842	32

Correlations

		x	e1	e2	e3
Pearson Correlation	x	1.000	-.620	-.524	-.381
	e1	-.620	1.000	.500	.500
	e2	-.524	.500	1.000	.500
	e3	-.381	.500	.500	1.000
Sig. (1-tailed)	x	.	.000	.001	.016
	e1	.000	.	.002	.002
	e2	.001	.002	.	.002
	e3	.016	.002	.002	.
N	x	32	32	32	32
	e1	32	32	32	32
	e2	32	32	32	32
	e3	32	32	32	32

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.667[a]	.445	.386	1.47600

a. Predictors: (Constant), e3, e2, e1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	49.000	3	16.333	7.497	.001[a]
	Residual	61.000	28	2.179		
	Total	110.000	31			

a. Predictors: (Constant), e3, e2, e1

b. Dependent Variable: x

Note. The analysis of variance summary table from the multiple regression analysis agrees with that from the analysis of variance from Topic 3 as shown below.

Tests of Between-Subjects Effects

Dependent Variable: x

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	49.000[a]	3	16.333	7.497	.001
Intercept	578.000	1	578.000	265.311	.000
b	49.000	3	16.333	7.497	.001
Error	61.000	28	2.179		
Total	688.000	32			
Corrected Total	110.000	31			

a. R Squared = .445 (Adjusted R Squared = .386)

The meaning of the regression coefficients with Effect Coding

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1.000	(Constant)	4.25000	.26092		16.28838	.00000
	e1	-1.25000	.45193	-.47673	-2.76591	.00994
	e2	-.75000	.45193	-.28604	-1.65955	.10817
	e3	.00000	.45193	.00000	.00000	1.00000

a. Dependent Variable: x

$$b_o = \overline{G} = 4.25$$

$$b_1 = \overline{X}_1 - \overline{G} = 3.00 - 4.25 = -1.25$$

$$b_2 = \overline{X}_2 - \overline{G} = 3.50 - 4.25 = -.75$$

$$b_3 = \overline{X}_3 - \overline{G} = 4.25 - 4.25 = 0.$$

Analysis Using Dummy Coding

If dummy coding were used instead:

1. The descriptive statistics would differ from those in slide 8.
2. Correlations would differ from those in slide 9.
3. The Model Summary and the ANOVA tables would be the same as in slides 9 and 10.
4. The Regression coefficients would differ from those in slide 11.

The meaning of the regression coefficients with Dummy Coding

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	6.250	.522		11.977	.000
	d1	-3.250	.738	-.759	-4.404	.000
	d2	-2.750	.738	-.642	-3.726	.001
	d3	-2.000	.738	-.467	-2.710	.011

a. Dependent Variable: x

$$b_o = \overline{X}_4 = 6.25$$

$$b_1 = \overline{X}_1 - \overline{X}_4 = 3.00 - 6.25 = -3.25$$

$$b_2 = \overline{X}_2 - \overline{X}_4 = 3.50 - 6.25 = -2.75$$

$$b_3 = \overline{X}_3 - \overline{X}_4 = 4.25 - 6.25 = -2.00$$

Relation Between the Experimental Design and GLM Models

The two models are:

$$X_{ai} = \mu + \alpha_a + \varepsilon_{ai} \qquad X_i = b_0 + b_1 A_1 + \ldots + \varepsilon_i$$

Therefore $\mu + \alpha_a = b_0 + b_1 A_1 + \ldots$

	Dummy Coding	Effect Coding
For a = 1	$\mu + \alpha_1 = b_0 + b_1$	$\mu + \alpha_1 = b_0 + b_1$
For a = 2	$\mu + \alpha_2 = b_0 + b_2$	$\mu + \alpha_2 = b_0 + b_2$
For a = 3	$\mu + \alpha_3 = b_0 + b_3$	$\mu + \alpha_3 = b_0 + b_3$
For a = 4	$\mu + \alpha_4 = b_0$	$\mu + \alpha_4 = b_0 - b_1 - b_2 - b_3$

Understanding Regression Coefficients: Dummy Coding

Given $\mu + \alpha_4 = b_0$

Therefore $b_0 = \mu + \alpha_4 = \mu + \mu_4 - \mu = \mu_4$

Given $\mu + \alpha_1 = b_0 + b_1$

Therefore $b_1 = \mu + \alpha_1 - b_0 = \mu_1 - \mu_4$

Given $\mu + \alpha_2 = b_0 + b_2$

Therefore $b_2 = \mu + \alpha_2 - b_0 = \mu_2 - \mu_4$

Given $\mu + \alpha_3 = b_0 + b_3$

Therefore $b_3 = \mu + \alpha_3 - b_0 = \mu_3 - \mu_4$

Understanding Regression Coefficients: Effect Coding

Given the four equations

$$\mu + \alpha_1 = b_0 + b_1 \quad \therefore b_0 + b_1 = \mu_1$$

$$\mu + \alpha_2 = b_0 + b_2 \quad \therefore b_0 + b_2 = \mu_2$$

$$\mu + \alpha_3 = b_0 + b_3 \quad \therefore b_0 + b_3 = \mu_3$$

$$\mu + \alpha_4 = b_0 - b_1 - b_2 - b_3 \quad \therefore b_0 - b_1 - b_2 - b_3 = \mu_4$$

Summing yields $4b_0 = \mu_1 + \mu_2 + \mu_3 + \mu_4 = 4\mu$

Therefore $b_0 = \mu$

And $b_1 = \mu_1 - \mu$

$b_2 = \mu_2 - \mu$

$b_3 = \mu_3 - \mu$

Major Observations

1. Either type of coding yields a multiple correlation of .667, and the test of significance produces an $F(3,28) = 7.497$.
2. The results are identical to those obtained using an analysis of variance program.
3. The meaning of the regression coefficients differs for every type of coding.

17

Assumptions Underlying the General Linear Model

Independence of residuals. The residuals (errors of prediction) are uncorrelated.

Homoscedasticity of residuals. The variances of the residuals are constant in the treatment populations.

Normality of residuals. The residuals are normally distributed in the treatment populations.

Null Hypothesis: The treatment population means are equal.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$$

18

References

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70,*426-443.

Cohen, J. & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (Second Edition).* 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).* Hillsdale, NJ: Lawrence Erlbaum.

19