Multiple Linear Regression II

Image source: https://commons.wikimedia.org/wiki/File:Autobunnskr%C3%A4iz-RO-A201.jpg

Lecture 8

Survey Research & Design in Psychology

James Neill, 2018

Creative Commons Attribution 4.0

Overview

1. Summary of MLR I
2. Semi-partial correlations
3. Residual analysis
4. Interactions
5. Analysis of change
6. Writing up an MLR
7. Summary of MLR II
8. MLR II quiz – Practice questions

2

Readings

1. As per Lecture 7 - MLR I, plus
2. **Howitt & Cramer (2014)**
   - Ch 39: Moderator variables and relationships between two variables

3

Summary of MLR I

Purpose of MLR:

To examine linear relations between:
- two or more predictors (IVs; X) and
- a single outcome variable (DV; Y)

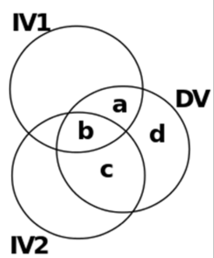
Model:

Develop a theoretical model:
- Path diagram and/or Venn diagram
- Express as one hypothesis per IV

4

Semi-partial correlation

Venn Diagrams depict variance and shared variance:
- a + b + c + d = variance in DV
- a + b + c = R^2 (variance in the DV explained by IV1 and IV2)
- a + c = uniquely explained variance (sr^2s)
- b = non-uniquely explained variance



a and c are **semi-partial correlations squared (sr^2)**:
a = sr^2 between *IV1* and DV after controlling for (partialling out) the influence of IV2
c = sr^2 between *IV2* and DV after controlling for (partialling out) the influence of IV1

Image source: http://www.gseis.ucla.edu/courses/ed230bc1/notes1/con1.html

Summary of MLR I

Interpret output:
- **Overall**:
  - R, R^2, Adjusted R^2
  - Changes in R^2 (if hierarchical)
  - Significance (F, p)
- **For each IV**:
  - Standardised coefficient
    - size, direction and significance
  - Unstandardised coefficient
    - report equation (if useful)
  - Semi-partial correlations (sr^2)

6

Summary of MLR I

Choose type:

- Standard
- Hierarchical
- Stepwise, Forward, Backward

7

Summary of MLR I

Check assumptions:

- Level of measurement
- Sample size
- Normality
- Linearity
- Homoscedasticity
- Multicollinearity
- Multivariate outliers
- Residuals should be normally distributed

8

Semi-partial correlations in MLR

When interpreting MLR coefficients:

- Draw a Path diagram or Venn diagram
- Compare zero-order (*r*) and semi-partial correlations (s*r*) for each IV to help understand relations amongst the IVs and the DV:
  - A semi-partial correlation (*sr*) will be less than or equal to the correlation (*r*)
  - If a *sr* equals the *r*, then the IV independently predicts the DV
  - To the extent that a *sr* is less than the *r*, the IV's explanation of the DV is shared with other IVs
  - An IV may have a significant *r* with the DV, but a non-significant *sr*. This indicates that the unique variance explained by the IV in the target population could be 0, so the IV is not significant.
- Compare the relative importance of the predictors using betas and/or *sr*s

9

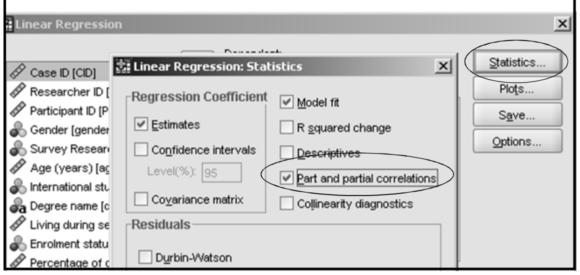
Semi-partial correlations in MLR

- SPSS provides semi-partial correlations (optional) (Note: The *sr*s are somewhat misleadingly labelled "part")
- Square the *sr* values to get sr^2.
- sr^2 indicates the % of variance in the DV which is uniquely explained by an IV.
- Compare each sr^2 with the r^2 (or *sr* with the *r*) – do they differ – why?

10

Semi-partial correlations for MLR in SPSS - Example

Analyze - Linear Regression - Statistics - Part and partial correlations



Semi-partial correlations for MLR in SPSS - Output

Coefficients[a]

http://web.psych.unimelb.edu.au/jkanglim/correlationandregression.pdf

Model		Correlations		
		Zero-order	Partial	Part
1	Zscore: GA: General Ability Total	.285	.107	.090
	Zscore: PSA: Perceptual Speed Ability Total	.419	.319	.280
	Zscore: PMA: Psychomotor Ability Total	.275	.034	.028
	QBK: Total Knowledge of Text Editing Keys (% correct)	.386	.368	.329

a. Dependent Variable: Typing Test: Speed (Words Per Minute)

Semi-partial correlations (*sr*) indicate the relative importance of each of the predictors; sr^2 indicates the % of variance uniquely explained by each predictor.

Semi-partial correlations for MLR in SPSS - Output

Coefficients[a]

http://web.psych.unimelb.edu.au/jkanglim/correlationandreggression.pdf

Model		Correlations Zero-order	Partial	Part
1	Zscore: GA: General Ability Total	.285	.107	.090
	Zscore: PSA: Perceptual Speed Ability Total	.419	.319	.280
	Zscore: PMA: Psychomotor Ability Total	.275	.034	.028
	QBK: Total Knowledge of Text Editing Keys (% correct)	.386	.368	.329

a. Dependent Variable: Typing Test: Speed (Words Per Minute)

General Ability has a small positive *r* with the DV, but a near 0 *sr*, so it doesn't uniquely explain much variance in the context of the other IVs.

Semi-partial correlations for MLR in SPSS - Output

Coefficients[a]

http://web.psych.unimelb.edu.au/jkanglim/correlationandreggression.pdf

Model		Correlations Zero-order	Partial	Part
1	Zscore: GA: General Ability Total	.285	.107	.090
	Zscore: PSA: Perceptual Speed Ability Total	.419	.319	.280
	Zscore: PMA: Psychomotor Ability Total	.275	.034	.028
	QBK: Total Knowledge of Text Editing Keys (% correct)	.386	.368	.329

a. Dependent Variable: Typing Test: Speed (Words Per Minute)

Perceptual Speed Ability has a small to moderate positive *r* with the DV, and a small *sr*, so it uniquely explains some variance in the context of the other IVs.

Semi-partial correlations for MLR in SPSS - Output

Coefficients[a]

http://web.psych.unimelb.edu.au/jkanglim/correlationandreggression.pdf

Model		Correlations Zero-order	Partial	Part
1	Zscore: GA: General Ability Total	.285	.107	.090
	Zscore: PSA: Perceptual Speed Ability Total	.419	.319	.280
	Zscore: PMA: Psychomotor Ability Total	.275	.034	.028
	QBK: Total Knowledge of Text Editing Keys (% correct)	.386	.368	.329

a. Dependent Variable: Typing Test: Speed (Words Per Minute)

Psychomotor Ability has a small positive *r* with the DV, but a near 0 *sr*, so it doesn't uniquely explain much variance in the context of the other IVs.

Semi-partial correlations for MLR in SPSS - Output

Coefficients[a]

http://web.psych.unimelb.edu.au/jkanglim/correlationandreggression.pdf

Model		Correlations Zero-order	Partial	Part
1	Zscore: GA: General Ability Total	.285	.107	.090
	Zscore: PSA: Perceptual Speed Ability Total	.419	.319	.280
	Zscore: PMA: Psychomotor Ability Total	.275	.034	.028
	QBK: Total Knowledge of Text Editing Keys (% correct)	.386	.368	.329

a. Dependent Variable: Typing Test: Speed (Words Per Minute)

Knowledge of Text Editing Keys has a small to moderate positive *r* with the DV, as does the *sr*, so it uniquely explains the most variance.

Semi-partial correlations in MLR

In a different example, where R^2 = .34:

- 16% is uniquely explained by the IVs (sum of sr^2s)
- 18% is explained by the combination of the IVs (34% - 16%) (R^2 - sum of sr^2s)
- 64% is unexplained variance (100% - R^2)

Independent variables	sr^2
Not Coping (NC)	.07
Somatic Coping (SC)	.06
Wishful Thinking (WT)	.02
Worry (W)	.01

NC & SC are the strongest predictors.

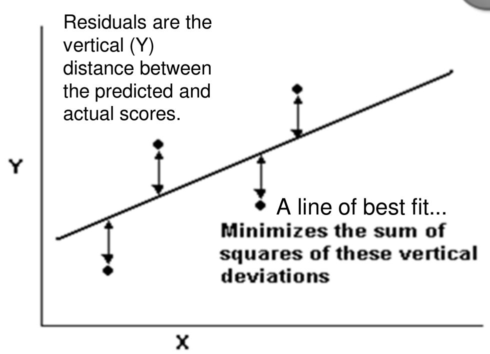
17

Summary: Semi-partial correlations in MLR

1. In MLR, *sr* is labelled "part" in the regression coefficients table SPSS output
2. Square these values to obtain sr^2, the unique % of DV variance explained by each IV
3. Discuss the extent to which the explained variance in the DV is due to unique or shared contributions of the IVs

18

Residual analysis



Residual analysis

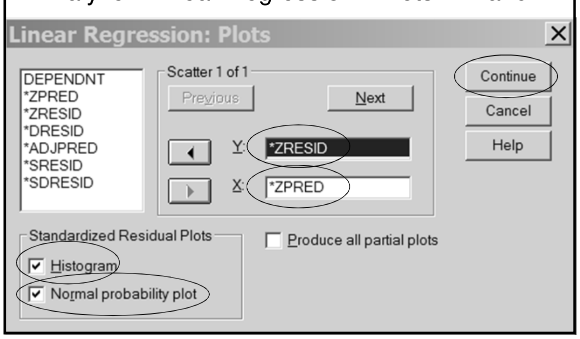
Assumptions about residuals:

- Sometimes +ve, sometimes -ve but, on average, 0
- Error is random
- Normally distributed about 0

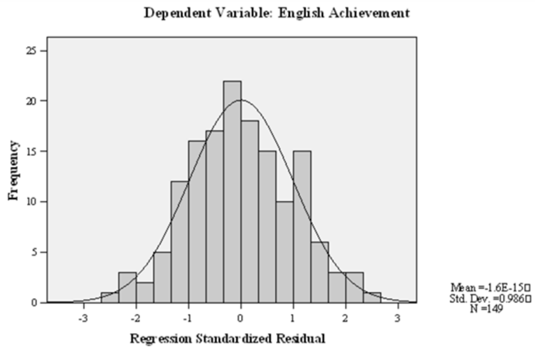
20

Residual analysis

Analyze - Linear Regression - Plots – Y and X



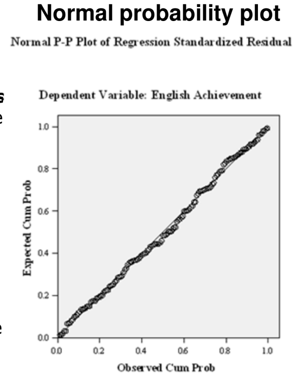
Residual analysis: Histogram



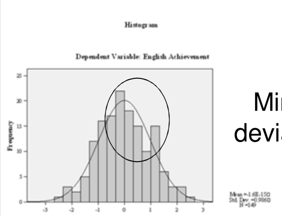
Residual analysis:

Normal probability plot

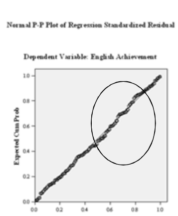
"The **Normal P-P** (Probability) **Plot of Regression Standardized Residuals** can be used to assess the assumption of normally distributed residuals. **If the points cluster reasonably tightly along the diagonal line (as they do here), the residuals are normally distributed.** Substantial deviations from the diagonal may be cause for concern."
Allen & Bennett, 2008, p. 183



Residual analysis: Histogram and NPP

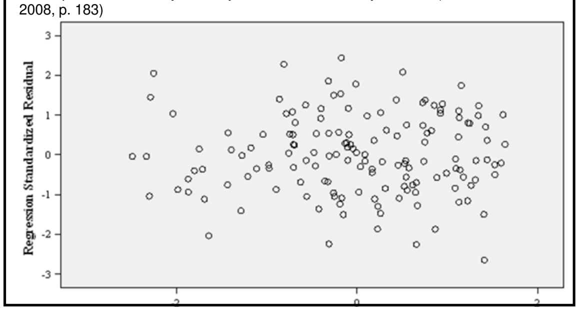


Minor deviation



Residual analysis: Scatterplot of predicted vs. actual values

The absence of any clear patterns in the spread of points indicates that the MLR assumptions of normality, linearity and homoscedasticity are met. (Allen & Bennett, 2008, p. 183)



Residuals - Why the big fuss?

↑ assumption violation

→

↑ Type I error rate

(i.e., more false positives)

26

Residuals - Why the big fuss?

- Standard error formulae (used for confidence intervals and sig. tests) work when residuals are well-behaved.
- If the residuals don't meet assumptions these formulae tend to underestimate coefficient standard errors giving lower *p*-values and more Type I errors.

27

Summary: Residual analysis

1. Residuals are the difference between predicted and observed Y values
2. MLR assumption is that residuals are normally distributed.
3. Examining residuals also helps to assess:
   1. Linearity
   2. Homoscedasticity

28

Interactions

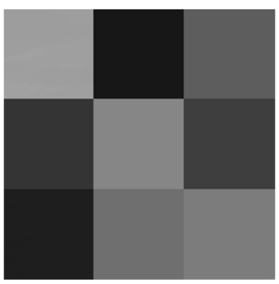


Image source: http://commons.wikimedia.org/wiki/File:Color_icon_orange.png

Interactions

- Additivity - when IVs act independently on a DV they do not interact.
- Alternatively, there may be interaction effects - i.e., the magnitude of the effect of one IV on a DV varies as a function of a second IV.
- Also known as a moderation effect.

30

Interactions: Example 1

Some drugs interact with each other to reduce or enhance other's effects e.g.,

Pseudoephedrine → ↑ Arousal
Caffeine → ↑ Arousal
Pseudoeph. X Caffeine → ↑↑↑ Arousal

X

31

Interactions: Example 2

Physical exercise (IV1)
and
Natural environments (IV2)
may provide
multiplicative benefits
in reducing stress e.g.,

Natural environment → ↓ Stress
Physical exercise → ↓ Stress
Natural env. x Phys. ex. → ↓↓↓ Stress

32

Interactions: Example 3

University student satisfaction (IV1)
and
Level of coping (IV2)
may predict level of stress e.g.,

Satisfaction → ↓ Stress
Coping → ↓ Stress
Satisfaction x Coping → ↓↓↓ Stress
(Dissatisfaction x Not coping → ↑↑↑ Stress)

33

Interactions

Test interactions in MLR by computing a cross-product term e.g.,:

- Pseudoephedrine (IV1)
- Caffeine (IV2)
- Pseudoephedrine x Caffeine (IV3)

cross-product

34

Interactions

cross-product term

$$Y = b_1x_1 + b_2x_2 + b_{12}x_{12} + a + e$$

- b_{12} is the product of $b_1 \times b_2$
- b_{12} can be interpreted as the amount of change in the slope of the regression of Y on b_1 when b_2 changes by one unit.

35

Interactions

- Conduct Hierarchical MLR
  - Step 1:
    - Pseudoephedrine
    - Caffeine
  - Step 2:
    - Pseudo x Caffeine (cross-product)
- Examine ΔR^2, to see whether the interaction term explains additional variance above and beyond the additive effects of Pseudo and Caffeine.

36

Interactions

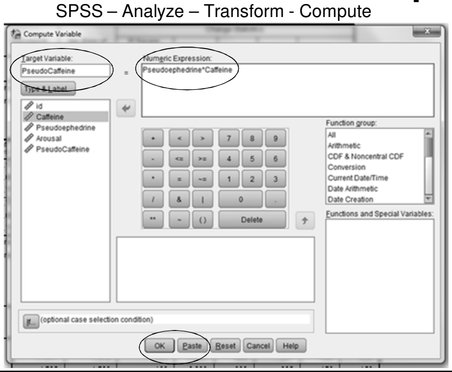
Possible effects of Pseudo and Caffeine on Arousal:

- None
- Pseudo only (↑/↓)
- Caffeine only (↑/↓)
- Pseudo + Caffeine (↑↑/↓↓)
- Pseudo x Caffeine (↑↑↑ (synergistic))
- Pseudo x Caffeine (↓↓↓ (antagonistic))

37

Interactions - SPSS example

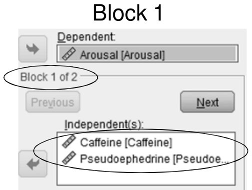
SPSS – Analyze – Transform - Compute



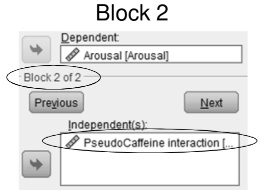
Interactions - SPSS example

SPSS – Analyze – Regression - Linear

Block 1



Block 2



Interactions - SPSS example

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change
1	.543[a]	.295	.288	17.34399	.295
2	.547[b]	.299	.289	17.33241	.004

a. Predictors: (Constant), Caffeine, Pseudoephedrine

b. Predictors: (Constant), Caffeine, Pseudoephedrine, PseudoCaffeine interaction

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	57.705	5.109		11.296	.000
	Pseudoephedrine	11.511	1.510	.464	7.625	.000
	Caffeine	4.735	1.780	.162	2.660	.008
2	(Constant)	42.730	14.137		3.023	.003
	Pseudoephedrine	18.137	6.025	.732	3.010	.003
	Caffeine	10.126	5.069	.346	1.998	.047
	PseudoCaffeine interaction	-2.313	2.036	-.380	-1.136	.257

Pseudoephedrine and Caffeine are each significantly, positively associated with Arousal, however **there is no interaction**.

Note: This data was fabricated for demonstration purposes.

Interactions

- Cross-product interaction terms may be highly correlated (collinear) with the corresponding simple IVs, creating problems with assessing the relative importance of main effects and interaction effects.
- An alternative approach is to conduct separate regressions for each level of one of the IVs e.g.,
  - What is the effect of caffeine on Arousal? (without taking any pseudoephedrine)?
  - What is the effect of caffeine on Arousal? (when also taking pseudoephedrine)?

41

Summary: Interactions

1. In MLR, IVs may interact to:
   1. Have no effect
   2. Increase the IVs' effect on the DV
   3. Decrease the IVs' effect on the DV
2. Model interactions using hierarchical MLR:
   1. Step 1: Enter IVs
   2. Step 2: Enter cross-product of IVs
   3. Examine change in R^2

42

Analysis of Change

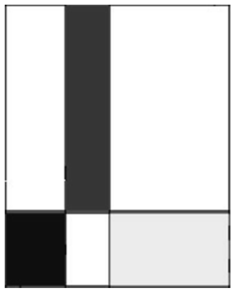


Image source: http://commons.wikimedia.org/wiki/File:PMinspirert.jpg, CC-by-SA 3.0

Analysis of change

- MLR can be used to analyse the variance of changes in an outcome measures over time (e.g., an intervention study using pre and post tests).
- Two main approaches:
  1. Standard regression:
     1. Compute post-pre difference (or change) scores in the outcome measure and use these change scores as the DV in a standard regression.
  2. Hierarchical MLR:
     1. DV is the post-intervention measure
     2. Step1: "Partial out" the baseline by entering the pre-intervention score as an IV.
     3. Step 2: Enter the IVs and examine the change in R^2.

44

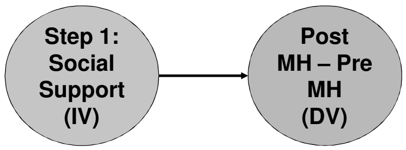
Analysis of change: Example

Does the quality of social support explain changes in mental health between the beginning and end of an intervention ?

45

Analysis of change: Option 1

Calculate difference scores between pre and post Mental Health.

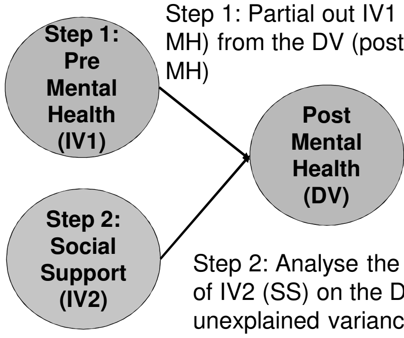


Use regression to explain the extent to which the IV (SS) can explain variance in the DV (change in MH).

46

Analysis of change: Option 2

Step 1: Partial out IV1 (pre-MH) from the DV (post-MH)



Step 2: Analyse the effect of IV2 (SS) on the DV (i.e., unexplained variance in MH)

47

Analysis of change: Option 2

Hierarchical MLR

- **DV** = Post-intervention mental health
- **Step 1**
  - **IV1** = Pre-intervention mental health
- **Step 2**
  - **IV2** = Social support from group members

48

Analysis of change: Option 2

Results of interest

- Change in R^2 – amount of additional variance in the DV (Post-MH) explained by IV2 (Social support) in Step 2 (after IV1 (Pre-MH) variance has been accounted for in Step 1)
- Regression coefficients - IV2 (Social support) in Step 2 indicates variance explained in the DV (Post-MH) after controlling for IV1 (Pre-MH) in Step 1

49

Summary: Analysis of change

Analysis of changes over time can be assessed by:

1. Standard regression
   1. Calculate difference scores (Post-score minus Pre-score) and use as a DV
2. Hierarchical MLR
   1. Step 1: "Partial out" baseline scores
   2. Step 2: Enter other IVs to help predict variance in changes over time.

50

Writing up an MLR

Image source: http://commons.wikimedia.org/wiki/File:Writing.svg, Public domain

Writing up an MLR: Intro

- What are the constructs of interest and why are they important?
- Explain how / why the IVs may be related to the DV, drawing on theory and research
- State hypotheses (one per IV)
  - Null hypothesis or
  - Alternative hypothesis (directional or non-directional)

52

Writing up an MLR: Results

- Establish the purpose of the analysis
- Univariate descriptive statistics (*M*, *SD*, Skewness, Kurtosis) for the IVs and DV
- Type of MLR
- Key assumptions
- Examine correlations
  - Between IVs
  - Between IVs and DV
- Overall model: R^2, Adjusted R^2, F, p
- Regression coefficients
- Table of regression coefficients (unstandardised and standardised), t, p, and sr^2 values

53

Writing up an MLR: Discussion

- Summarise the results in relation to each of the hypotheses and the overall research question, paying attention to:
  - Direction
  - Size
  - Statistical significance
- Interpret the results in the context of relevant psychological theory and research
- Entertain alternative explanations
- Consider strengths and limitations
- Discuss implications and recommendations

54

Table of correlations and regression coefficients

Table 2

Standard Multiple Linear Regression of Non-productive Coping Strategies on Psychological Distress Reported by Adolescents During Outdoor Education Programs

Variables	PD[a] (DV)	NC	WT	W	SC	B (unique)	β	sr^{2b}
Not Coping (NC)	.43					.41*	.29	.07
Wishful Thinking (WT)	.38	.36				.15*	.15	.02
Worry (W)	.36	.25	.40			.13*	.13	.01
Somatic Coping (SC)	.42	.22	.30	.38		.26*	.26	.06
Intercept						.60*		

Note. * $p < .05$; [a] PD = Psychological Distress; [b] sr^2 = the squared semipartial correlations indicate the unique variance predicted by the independent variable.

Summary: Writing up an MLR

1. Introduction
   1. Establish purpose
   2. Describe model and hypotheses
2. Results
   1. Univariate descriptive statistics
   2. Correlations
   3. Type of MLR and assumptions
   4. Regression coefficients
3. Discussion
   1. Summarise and interpret, with limitations
   2. Implications and recommendations

56

Questions

?

57

MLR II Quiz - Practice question 1

In an MLR, if the *r* between the two IVs is 1, *R* will equal the *r* between one of the IVs and the DV.
[Hint: Draw a Venn Diagram]
a) True
b) False

58

MLR II Quiz - Practice question 2

In MLR, if two IVs are somewhat correlated with the DV and with one another, the *sr*s between the IVs and the DV will be _______ in magnitude than the *r*s:
[Hint: Draw a Venn Diagram]
a) Equal
b) Smaller
c) Larger
d) Impossible to tell

59

MLR II Quiz - Practice question 3

In MLR, the unique variance in the DV explained by an IV is estimated by its:
a) Zero-order correlation square (r^2)
b) Multiple correlation coefficient squared (R^2)
c) Semi-partial correlation squared (sr^2)

60

MLR II Quiz - Practice question 4

Interaction effects can be tested in MLR by using IVs that represent:
a) Cross-products between the IVs and DV in a hierarchical regression
b) Cross-products of IVs
c) Semi-partial correlations squared (sr^2)

61

MLR II Quiz - Practice question 5

To assess the extent to which social support from group members during an intervention program explain changes in mental health (MH) between the beginning and end of the intervention, what MLR design could be used?
a) Hierarchical with pre-MH in Step 1
b) Hierarchical with cross-products of IVs in Step 2

62

References

- Allen, P. & Bennett, K. (2008). *SPSS for the health and behavioural sciences*. South Melbourne, Victoria, Australia: Thomson.
- Francis, G. (2007). *Introduction to SPSS for Windows: v. 15.0 and 14.0 with Notes for Studentware* (5th ed.). Sydney: Pearson Education.
- Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, CA: Wadsworth.
- Howitt, D. & Cramer, D. (2011). *Introduction to statistics in psychology* (5th ed.). Harlow, UK: Pearson.

63

Next lecture

Power and effect sizes
- Statistical power
- Significance testing
- Inferential decision making
- Effect sizes
- Confidence intervals
- Publication bias
- Academic integrity

64