Hierarchical Multiple Regression

Dr Daniel Boduszek
University of Huddersfield
d.boduszek@hud.ac.uk

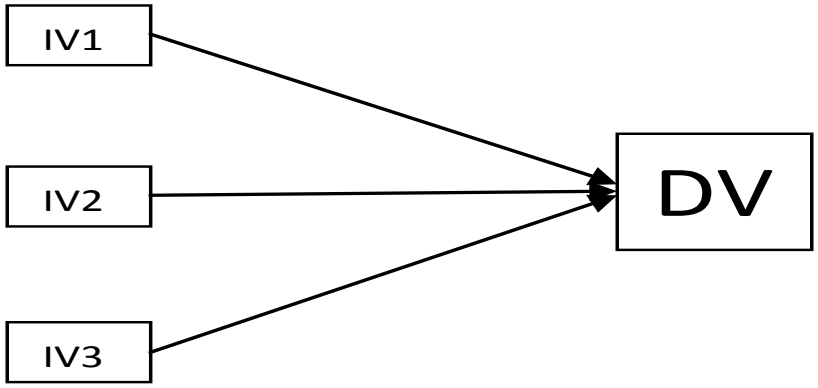
University of HUDDERSFIELD

Presentation Outline

- **Introduction to Multiple Regression**
  - Types of MR
  - Assumptions of MR
- **SPSS procedure of Hierarchical MR**
  - Example based on prison data
- **Interpretation of SPSS output**
- **Presenting results from HMR**
  - Tables
  - Presenting results in text

Intro to Multiple Regression (MR)

- MR is used to explore the relationship between set of IVs or predictors (usually continuous)and one DV



- Don't use MR as a fishing expedition!
- Theoretical or conceptual reason for MR analysis

Types of research questions that MR can be used to address:

- How well a set of IVs is able to predict a particular outcome (DV)
- Which IV is the best predictor of an outcome
- Whether a particular predictor variable is still able to predict an outcome when the effects of another variable is controlled for

Two Major Types of MR

- Standard Multiple Regression
  - All IVs are entered into equation simultaneously
  - Amount of variance in DV explained by set of IVs <u>as a group or block</u>
  - Identifies the strongest predictor variable within the model
- Hierarchical Multiple Regression
  - IVs are entered into equation in the <u>order</u> specified by the researcher based on <u>theoretical grounds</u>
  - IVs are entered in steps (blocks)
  - Each IV is assessed in terms of what it adds to the prediction of DV after the previous IVs have been controlled for.
  - Overall model and relative contribution of each block of variables is assessed

Assumptions of Multiple Regression

- **Sample size** – results with small sample do not generalise (cannot be repeated) with other sample ⟹ little scientific value
  - 15 participants per predictor (Stevens, 1996)
  - Formula for calculating sample size (Tabachnick & Fidell, 2007)

$$N > 50 + 8m$$

N = number of Participants

m = number of IVs

- **Multicollinearity and singularity**
  - Multicollinearity – high correlation between IVs (r = .9 and above)
  - Singularity – One IV is a combination of other variable (e.g. when both subscale scores and the total score of scale are included)

Assumptions of Multiple Regression

- **Outliers** – extreme scores should be deleted from the data set or given a score that is high but not different from the remaining cluster of scores

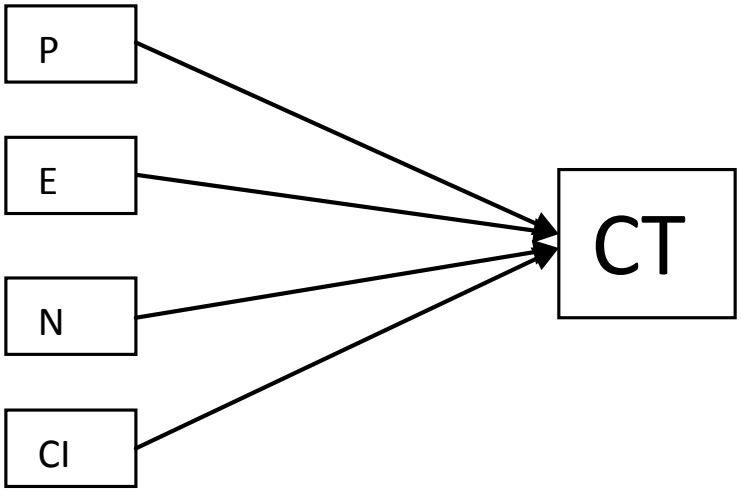
- **Normality, Linearity, Homoscedasticity** – refer to distribution of scores and the nature of the relationship between the variables

- Request standardised residual plot in MR (residuals are the differences between the obtained and the predicted DV scores)

  - **Normality** – the residuals should be normally distributed
  - **Linearity** – the residuals should have a straight-line relationship with predicted DV scores
  - **Homoscedasticity** – the variance of the residuals about predicted DV scores should be the same for all predicted scores.

Example of Hierarchical MR

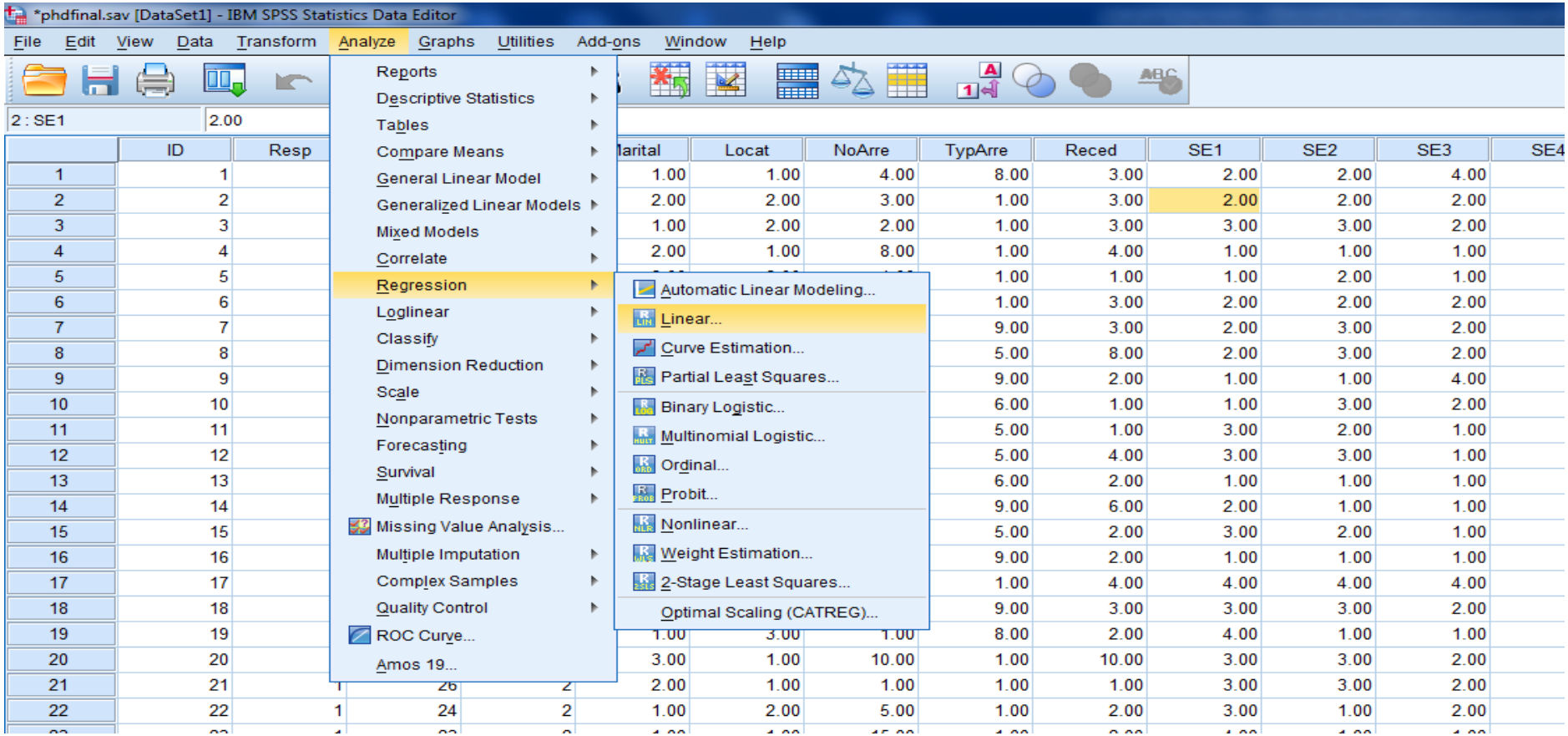
- Research Question: If we control for the possible effect of criminal social identity (CI) is personality (psychoticism (P), extraversion (E), and neuroticism (N)) still able to predict a significant amount of the variance in criminal thinking style (CT).



- Hierarchical Multiple Regression
  - Block 1: psychoticism, extraversion, and neuroticism
  - Block 2: criminal identity

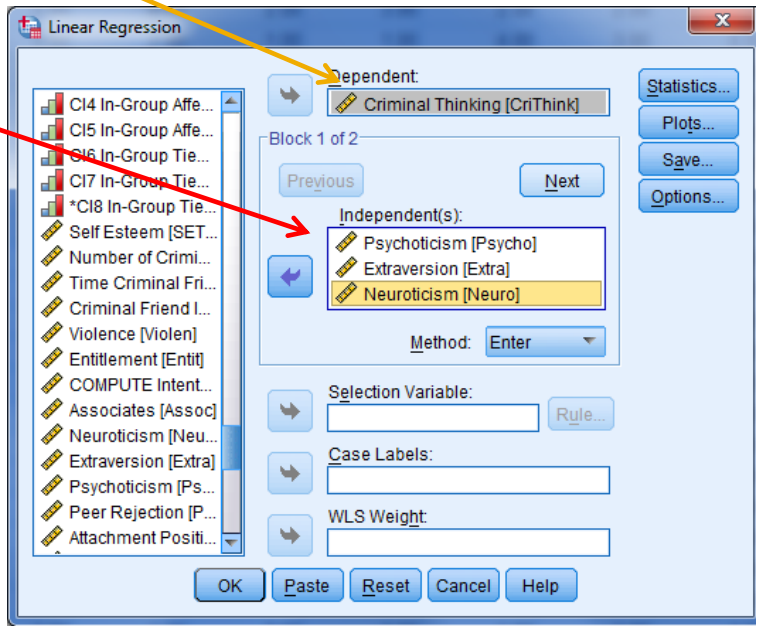
SPSS procedure for HMR

- From the menu at the top of the screen click **Analyze**, then select **Regression**, then **Linear**



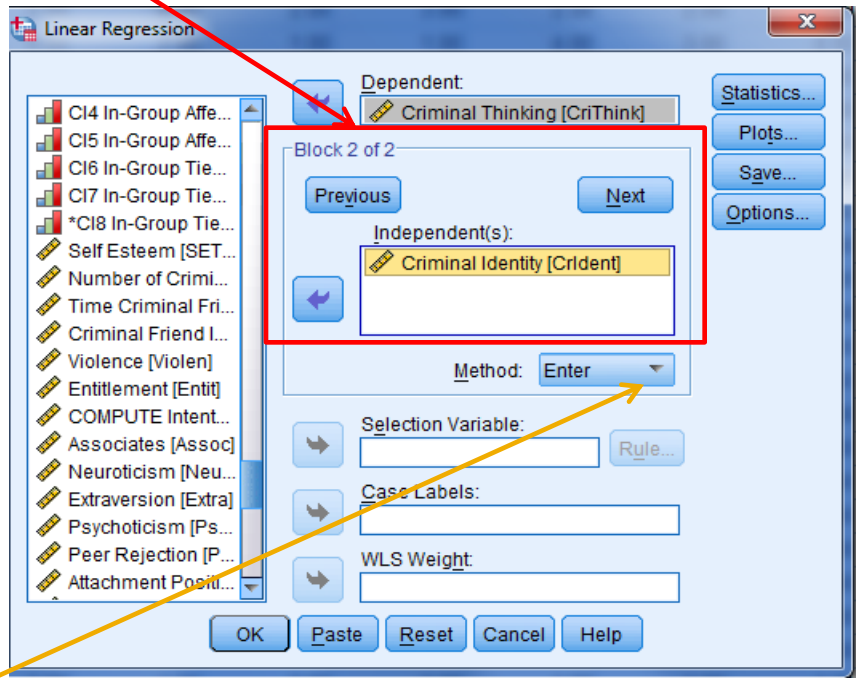
SPSS procedure for HMR

- Choose you continuous DV (Criminal Thinking) and move it into the **Dependent** box
- Move the IVs you wish to control for (P, E, N) into **Independent** box. This is the first block of IVs
- Click **Next**



SPSS procedure for HMR

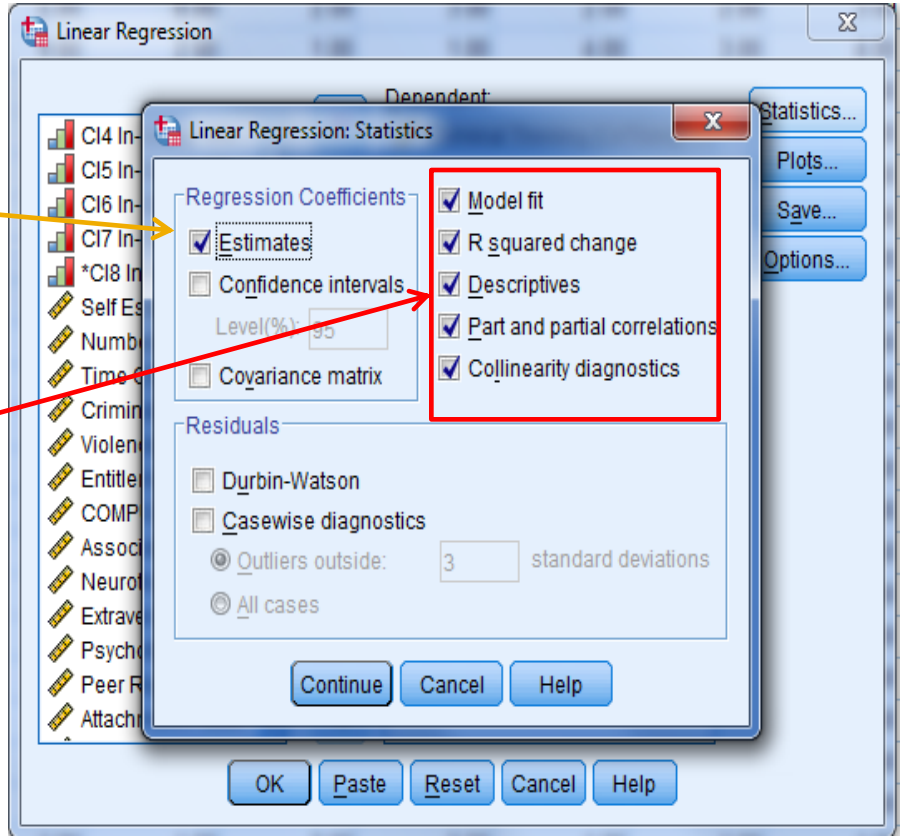
- Block 2 – enter your second set of IVs (this example has only one - Criminal Identity)



- **Method** box - make sure is set to the default (**Enter**)

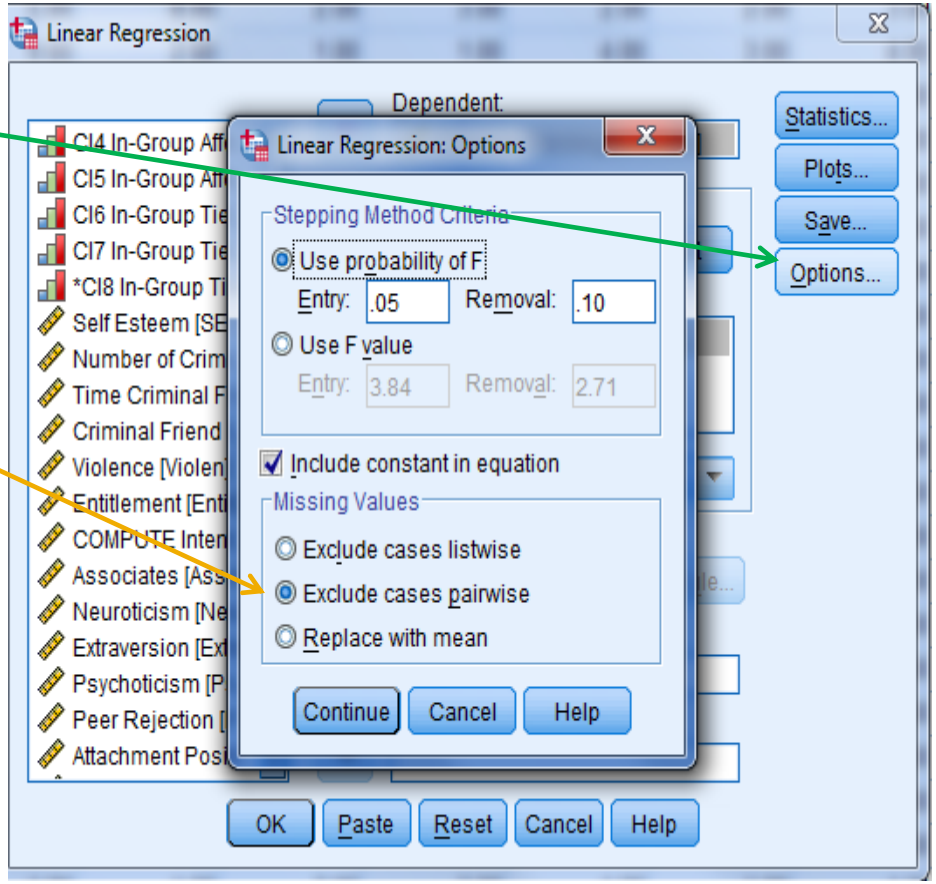
SPSS procedure for HMR

- Click on the **Statistics** button. Select the following:
  - **Estimates**
  - **Model fit**
  - **R squared change**
  - **Descriptives**
  - **Part and partial correlations**
  - **Collinearity diagnostics**
- Click **Continue**



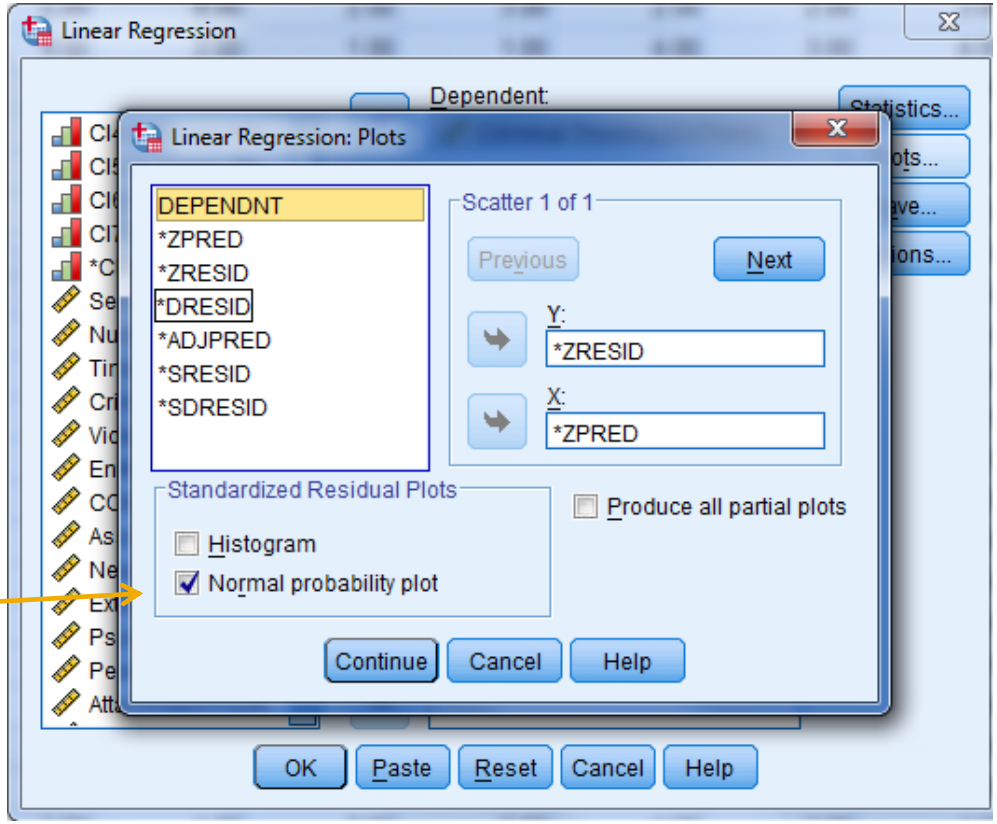
SPSS procedure for HMR

- Click on the **Options** button.
- In the **Missing Values** section, click on **Exclude cases pairwise**.
- Click on **Continue**.



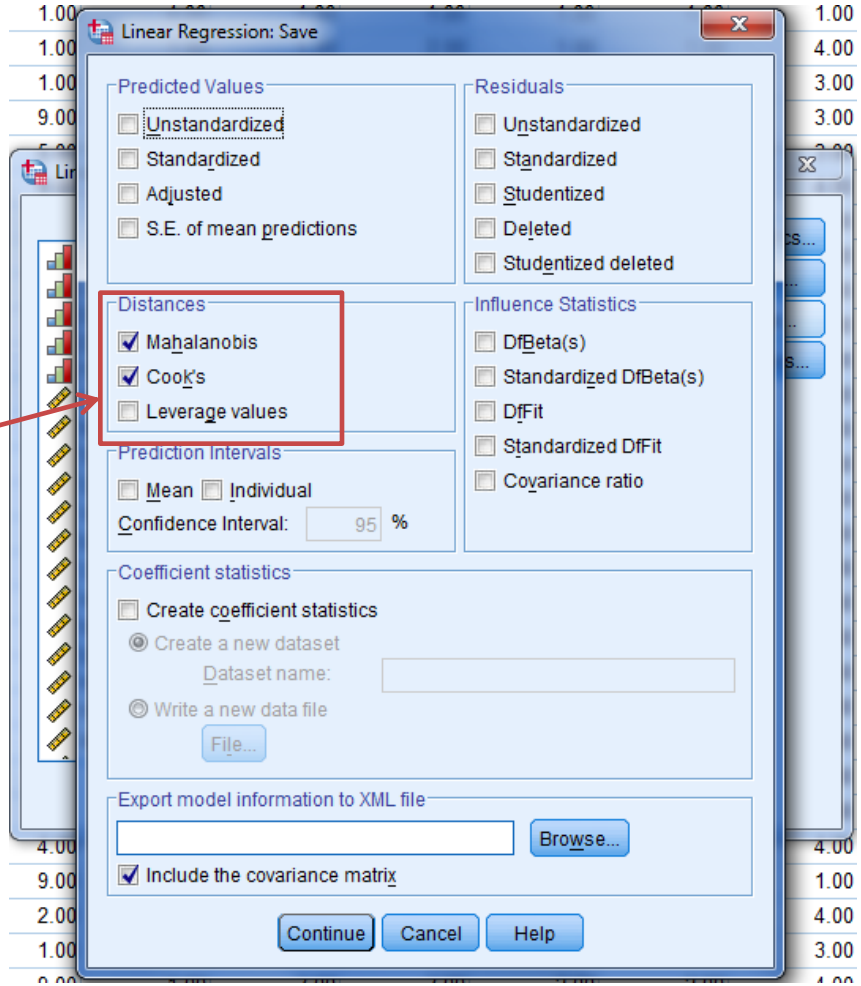
SPSS procedure for HMR

- Click on the **Plots** button.
- Click on the ***ZRESID** and the arrow button to move this into **Y** box.
- Click on the ***ZPERD** and the arrow button to move this into **X** box.
- In the **Standardized Residual Plots**, tick the **Normal probability plot** option.
- Click **Continue**.



SPSS procedure for HMR

- Click on the **Save** button.
- Click on **Mahalonobis** and **Cook's**.
- Click on Continue and then **OK**.



SPSS procedure for HMR

- The syntax from this procedure is:

```
REGRESSION
 /DESCRIPTIVES MEAN STDDEV CORR SIG N
 /MISSING PAIRWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN
TOL CHANGE ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT CriThink
 /METHOD=ENTER Psycho Extra Neuro
 /METHOD=ENTER CrIdent
 /SCATTERPLOT=(*ZRESID ,*ZPRED)
 /RESIDUALS NORMPROB(ZRESID)
 /SAVE MAHAL COOK.
```

Interpretation of the SPSS output

- **Step 1: Checking the assumptions**
  - Multicollinearity

Correlations

		Criminal Thinking	Psychoticism	Extraversion	Neuroticism	Criminal Identity
Pearson Correlation	Criminal Thinking	1.000	.465	.128	.246	.459
	Psychoticism	.465	1.000	.006	.155	.231
	Extraversion	.128	.006	1.000	-.174	-.039
	Neuroticism	.246	.155	-.174	1.000	.352
	Criminal Identity	.459	.231	-.039	.352	1.000
Sig. (1-tailed)	Criminal Thinking	.	.000	.012	.000	.000
	Psychoticism	.000	.	.459	.003	.000
	Extraversion	.012	.459	.	.001	.247
	Neuroticism	.000	.003	.001	.	.000
	Criminal Identity	.000	.000	.247	.000	.
N	Criminal Thinking	312	312	312	312	303
	Psychoticism	312	312	312	312	303
	Extraversion	312	312	312	312	303
	Neuroticism	312	312	312	312	303
	Criminal Identity	303	303	303	303	303

Interpretation of the SPSS output

- **Coefficients** table gives **Tolerance** and **Variance Inflation Factor (VIF)**
  - **Tolerance** value less than.10 – possible multicollinearity
  - **VIF** value above 10 – possible multicollinearity
  - If you exceed these values, you should remove one of the IVs

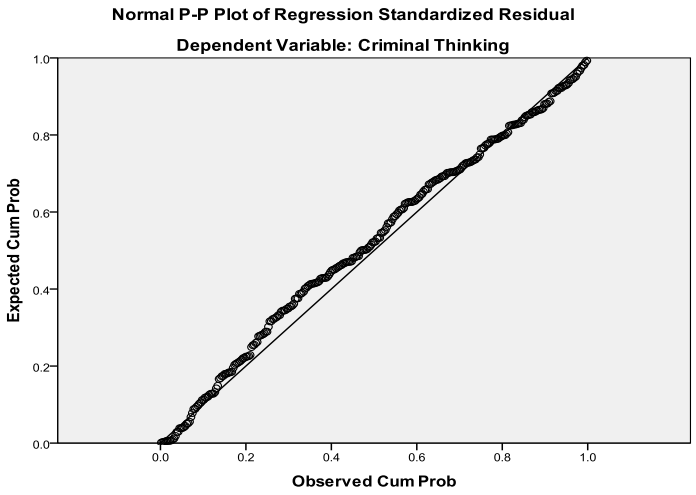
Coefficients[a]

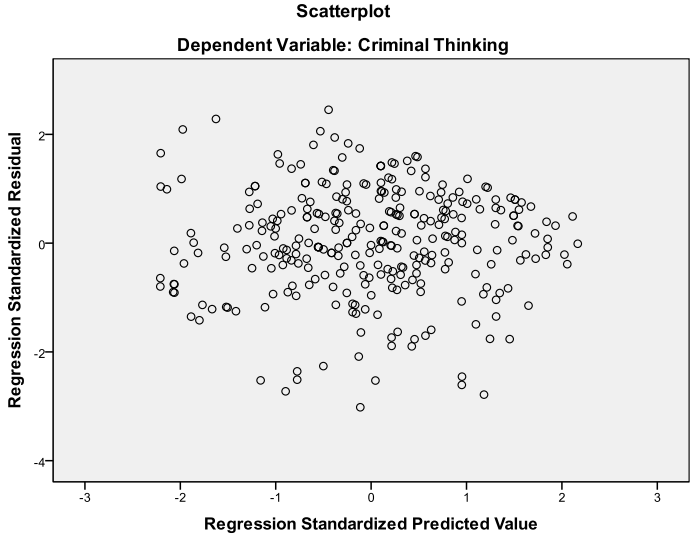
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	18.997	1.431		13.278	.000					
	Psychoticism	2.581	.299	.432	8.636	.000	.465	.447	.426	.975	1.026
	Extraversion	.766	.237	.162	3.231	.001	.128	.184	.159	.969	1.032
	Neuroticism	.804	.197	.207	4.079	.000	.246	.230	.201	.945	1.058
2	(Constant)	12.007	1.668		7.198	.000					
	Psychoticism	2.207	.283	.369	7.801	.000	.465	.412	.358	.940	1.064
	Extraversion	.739	.220	.156	3.353	.001	.128	.191	.154	.968	1.033
	Neuroticism	.365	.194	.094	1.880	.061	.246	.108	.086	.844	1.184
	Criminal Identity	.437	.063	.346	6.941	.000	.459	.373	.318	.844	1.185

a. Dependent Variable: Criminal Thinking

Interpretation of the SPSS output

- **Outliers, normality, linearity, homoscedasticity.**
  - Check **Normal Probability Plot (P-P) of the Regression Standardised Residual** and the **Scatterplot**.
  - In the **Normal P-P plot** – points should lie in reasonably straight diagonal line from bottom left to top right
  - In the **Scatterplot** - the residuals should be rectangularly distributed, with most of the scores concentrated in the centre (along the point 0)
  - Standardized residuals of more than 3.3 or less than -3.3 indicates outlyers (Tabachnick & Fidell, 2007)





Interpretation of the SPSS output

- **Step 2: Evaluating the model**

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.522[a]	.272	.265	7.01495	.272	37.252	3	299	.000
2	.611[b]	.373	.365	6.51941	.101	48.181	1	298	.000

a. Predictors: (Constant), Neuroticism, Psychoticism, Extraversion

b. Predictors: (Constant), Neuroticism, Psychoticism, Extraversion, Criminal Identity

c. Dependent Variable: Criminal Thinking

- Check the **R Square** in the **Model Summary** box. Variables entered in **Block 1** (P, E, N) explained **27% of the variance** (.27 x 100) in DV.
- After **Block 2** variable (criminal identity) has been included , the model as a whole explained **37% of variance** in DV.
- In the column labelled **R Square Change** (on the line marked Model 2) – Criminal Identity explained additional **10 % of the variance** in DV.
- This is significant contribution, as indicated by **Sig. F Change** value for this line **(.000)**

Interpretation of the SPSS output

- The **ANOVA** table indicates that the model as a whole (which includes both blocks of variables) is significant

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5499.479	3	1833.160	37.252	.000[a]
	Residual	14713.631	299	49.209		
	Total	20213.110	302			
2	Regression	7547.286	4	1886.822	44.393	.000[b]
	Residual	12665.824	298	42.503		
	Total	20213.110	302			

a. Predictors: (Constant), Neuroticism, Psychoticism, Extraversion

b. Predictors: (Constant), Neuroticism, Psychoticism, Extraversion, Criminal Identity

c. Dependent Variable: Criminal Thinking

$$F(4, 298) = 44.39, \; p < .0005$$

Interpretation of the SPSS output

- **Step 3: Evaluating each of the IVs**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	18.997	1.431		13.278	.000					
	Psychoticism	2.581	.299	.432	8.636	.000	.465	.447	.426	.975	1.026
	Extraversion	.766	.237	.162	3.231	.001	.128	.184	.159	.969	1.032
	Neuroticism	.804	.197	.207	4.079	.000	.246	.230	.201	.945	1.058
2	(Constant)	12.007	1.668		7.198	.000					
	Psychoticism	2.207	.283	.369	7.801	.000	.465	.412	.358	.940	1.064
	Extraversion	.739	.220	.156	3.353	.001	.128	.191	.154	.968	1.033
	Neuroticism	.365	.194	.094	1.880	.061	.246	.108	.086	.844	1.184
	Criminal Identity	.437	.063	.346	6.941	.000	.459	.373	.318	.844	1.185

a. Dependent Variable: Criminal Thinking

- Check Standardized Coefficient (**Beta** values) and **Sig.** box in Model 2
- Only Neuroticism does not make a unique significant contribution ($p > .05$)
- The best predictor of Criminal Thinking Style is Psychoticism ($\beta = .40$) followed by Criminal Identity ($\beta = .35$) , and Extraversion ($\beta = .16$).

Interpretation of the SPSS output

- The Standardized Beta values indicate also the number of SDs that scores in the DV would change if there was 1 SD unit change in the predictor (IV)

- E.g. If we could increase criminal identity scores by 1 SD (which is **6.49**, from **Descriptive Statistics** table) the criminal thinking scores would be likely to increase by .35 standard deviation units.

Descriptive Statistics

	Mean	Std. Deviation	N
Criminal Thinking	30.2436	8.18113	312
Psychoticism	2.0545	1.36807	312
Extraversion	4.2179	1.73036	312
Neuroticism	3.3750	2.10582	312
Criminal Identity	21.4125	6.48672	303

Presenting the results from HMR

- **Descriptive Statistics and Correlations**

Table 1. Descriptive statistics, reliability, and correlations for all continuous variables (N = 312)

Variables	CT	CI	P	E	N
Criminal Thinking (CT)	1				
Criminal Identity (CI)	.46***	1			
Psychoticism (P)	.47***	.23***	1		
Extraversion (E)	.13*	-.04	.01	1	
Neuroticism (N)	.25***	.35***	.16**	-.17**	1
Means	23.49	21.41	2.05	4.22	3.38
Standard Deviations	6.87	6.47	1.37	1.73	2.11
Range	6-34	6-34	0-6	0-6	0-6
Possible Range	0-36	0-36	0-6	0-6	0-6
Cronbach's Alpha	.86	.87	.62	.73	.71

Note. Statistical significance: *$p < .05$; **$p < .01$; ***$p < .001$

Presenting the results from HMR

R^2 = amount of variance explained by IVs

R^2 Change = additional variance in DV

B = Unstandardized coefficient

β = Standardized coefficient (values for each variable are converted to the same scale so they can be compared)

SE = Standard Error

t = estimated coefficient (B) divided by its own SE. If **t < 2** the IV does not belong to the model

Table 3. Hierarchical Regression Model of Criminal Thinking Style

	R	R^2	R^2 *Change*	*B*	*SE*	*β*	*t*
Step 1	.52	.27***					
Psychoticism				2.58	.30	.43***	8.63
Extraversion				.77	.24	.16**	3.23
Neuroticism				.80	.20	.21***	4.08
Step 2	.61	.37***	.10***				
Psychoticism				2.21	.28	.37***	7.80
Extraversion				.74	.22	.16**	3.35
Neuroticism				.37	.19	.09	1.88
Criminal Identity				.44	.06	.35***	6.94

Note. Statistical significance: *$p < .05$; **$p < .01$; ***$p < .001$

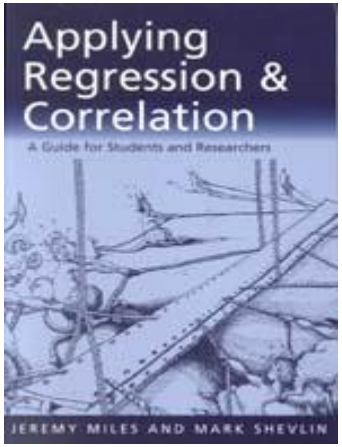
Presenting the results from HMR

Hierarchical multiple regression was performed to investigate the ability of personality factors (psychoticism, extraversion, and neuroticism) to predict levels of criminal thinking style, after controlling for criminal social identity. Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. Additionally, the correlations amongst the predictor variables (criminal identity; psychoticism, neuroticism, and extraversion) included in the study were examined and these are presented in Table 1. All correlations were weak to moderate, ranging between $r = .16$, $p < .01$ and $r = .35$, $p < .001$. This indicates that multicollinearity was unlikely to be a problem (see Tabachnick and Fidell, 2007). All predictor variables were statistically correlated with criminal thinking style which indicates that the data was suitably correlated with the dependent variable for examination through multiple linear regression to be reliably undertaken. The correlations between the predictor variables and the dependent variable (criminal thinking style) were all weak to moderately strong, ranging from $r = .13$, $p < .05$ to $r = .47$, $p < .001$.

In the first step of hierarchical multiple regression, three predictors were entered: psychoticism, extraversion, and neuroticism. This model was statistically significant $F (3, 299) = 37.25$; $p < .001$ and explained 27 % of variance in criminal thinking style. All three personality factors made a significant unique contribution to the model (see Table 2). After entry of criminal social identity at Step 2 the total variance explained by the model as a whole was 37% ($F (4, 298) = 44.39$; $p < .001$). The introduction of criminal social identity explained additional 10% of variance in criminal thinking style, after controlling for psychoticism, extraversion, and neuroticism (R^2 Change = .10; $F (1, 298) = 48.18$; $p < .001$). In the final adjusted model three out of four predictor variables were statistically significant, with psychoticism recording a higher Beta value ($\beta = .37$, $p < .001$) than the criminal social identity ($\beta = .35$, $p < .001$) and extraversion ($\beta = .16$, $p < .01$)

Multiple Regression - Books

- Miles, J., & Shevlin, M. (2001). *Applying Regression and Correlation*. London: Sage Publication.



- Tabachnick, B.G., & Fidell, L.S. (2007). *Using Multivariate Statistics, Fifth Edition*. Boston: Pearson Education, Inc.