Seminar 7
How to conduct and interpret the hierarchical multiple regression

Quantitative Research Methods
Dr. Naser Alziyadat

Murdoch UNIVERSITY

MULTIPLE REGRESSION

- The prediction of individual income may depend on a combination of education, job experience, gender, age, etc.
- Multiple regression employs the same rationale as simple regression and the formula is a logical extension of that for linear regression:

$$Y = b_0 + b_1X_1 + b_2 X_2 + b_3 X_3 + \ldots$$

etc

Assumptions of Multiple Regression

- Linearity: it is assumed when it comes to our work in this unit.
- Normality of the residuals: To be checked after we run the test as it will be part of the output.
- Homoscedasticity: To be checked after we run the test as it will be part of the output.
- *Multicollinearity:* Very high correlations between IV's should be avoided. To be checked after we run the test as it will be part of the output.

TYPES OF MULTIPLE REGRESSION

Murdoch UNIVERSITY

There are three types of Multiple Regression

1. Standard Multiple regression
2. Hierarchical Multiple Regression
3. Stepwise Multiple Regression

Example: from the last workshop

- We will be exploring the impact of respondents' perceptions of control on their levels of perceived stress. The literature in this area suggests that if people feel that they are in control (IVs) of their lives, they are less likely to experience 'stress' (DV) (tpstress)
- Control = IVs =
  - the Mastery scale, which measures the degree to which people feel they have control over the events in their lives (tmast)
  - the Perceived Control of Internal States Scale (PCOISS), which measures the degree to which people feel they have control over their internal states
- Example of research questions
  - 1. How well do the two measures of control (mastery, PCOISS) predict perceived stress? How much variance in perceived stress scores can be explained by scores on these two scales?
  - 2. Which is the best predictor of perceived stress: control of external events (Mastery scale), or control of internal states (PCOISS)?

Procedure for standard multiple regression

- 1. From the menu at the top of the screen click on: Analyze, then click on Regression, then on Linear.
- 2. Click on your continuous dependent variable (e.g. total perceived stress: tpstress) and move it into the **Dependent** box.
- 3. Click on your independent variables (total mastery: tmast; total PCOISS: tpcoiss) and move them into the **Independent** box.
- 4. For **Method**, make sure **Enter** is selected (this will give you standard multiple regression).
- 5. Click on the Statistics button:

  - Tick the box marked **Estimates, Confidence Intervals, Model fit, Descriptives, Part and partial correlations and Collinearity diagnostics.**
  - In the **Residuals section** tick the **Casewise diagnostics and Outliers outside 3 standard deviations.**
  - Click on **Continue**.

Procedure for standard multiple regression

6. Click on the **Options** button. In the **Missing Values section** click on **Exclude cases pairwise**.

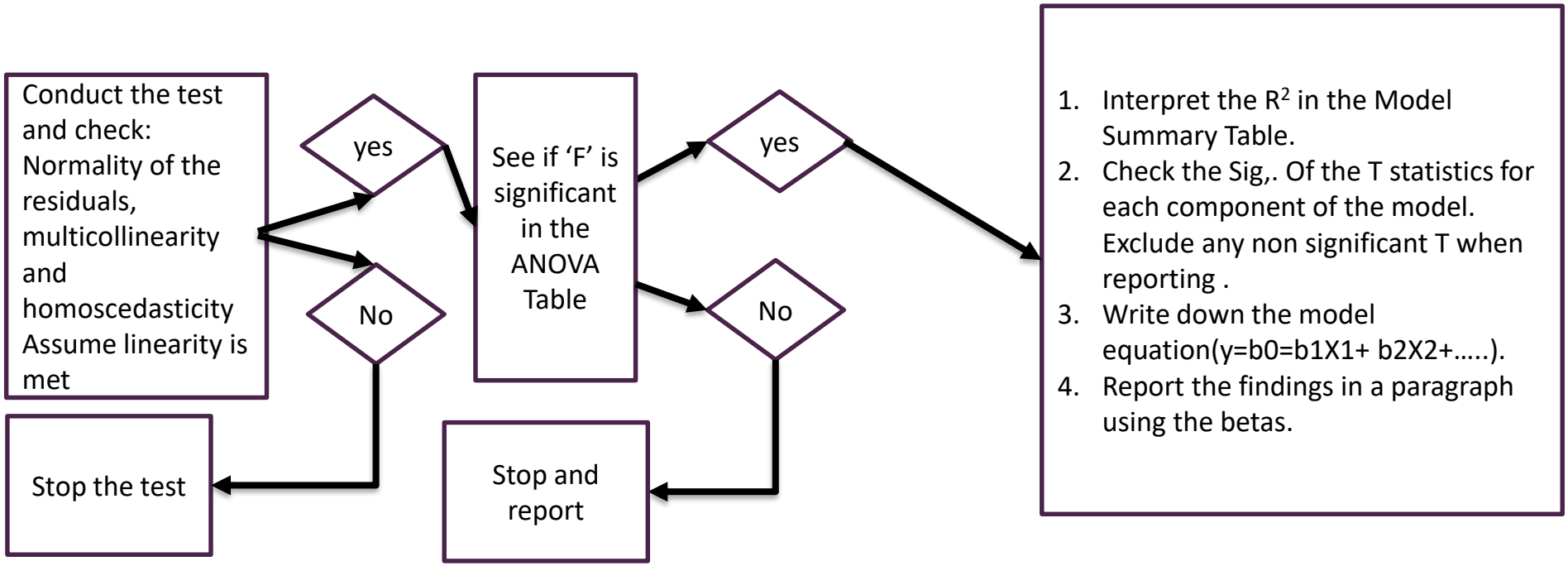
7. Click on the **Plots** button.
   - Click on ***ZRESID** and the arrow button to move this into the Y box.
   - Click on ***ZPRED** and the arrow button to move this into the X box.
   - In the section headed Standardized Residual Plots, tick the **Normal probability plot** option.
   - Click on **Continue**.

8. Click on the Save button.

- • In the section labelled **Distances** tick the **Mahalanobis** box (this will identify multivariate outliers for you) and **Cook's**, then Click on **Continue**.

9. Click **OK**

Multiple Linear Regression



Check for the assumption of **Multicollinearity**

* Go for the Correlation table.
* Check that your independent variables show at least some relationship with your dependent variable (above 0.3 preferably).
* Also check that the correlation between each of your independent variables is not too high (no more than 0.7).
* In cases with more than (0.7), you need to omit one of the variables.
* Go for table “collinearity diagnostics” .
* A condition index greater than 15 indicates a possible problem
* An index greater than 30 suggests a serious problem with collinearity.

Correlations

		Total perceived stress	Total PCOISS	Total Mastery
Pearson Correlation	Total perceived stress	1.000	-.581	-.612
	Total PCOISS	-.581	1.000	.521
	Total Mastery	-.612	.521	1.000
Sig. (1-tailed)	Total perceived stress	.	.000	.000
	Total PCOISS	.000	.	.000
	Total Mastery	.000	.000	.
N	Total perceived stress	433	426	433
	Total PCOISS	426	430	429
	Total Mastery	433	429	436

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Total PCOISS	Total Mastery
1	1	2.965	1.000	.00	.00	.00
	2	.019	12.502	.62	.80	.01
	3	.016	13.780	.38	.20	.99

a. Dependent Variable: Total perceived stress

Check for the assumption of **Multicollinearity** >>> Cont.

Murdoch UNIVERSITY

Coefficients[a]

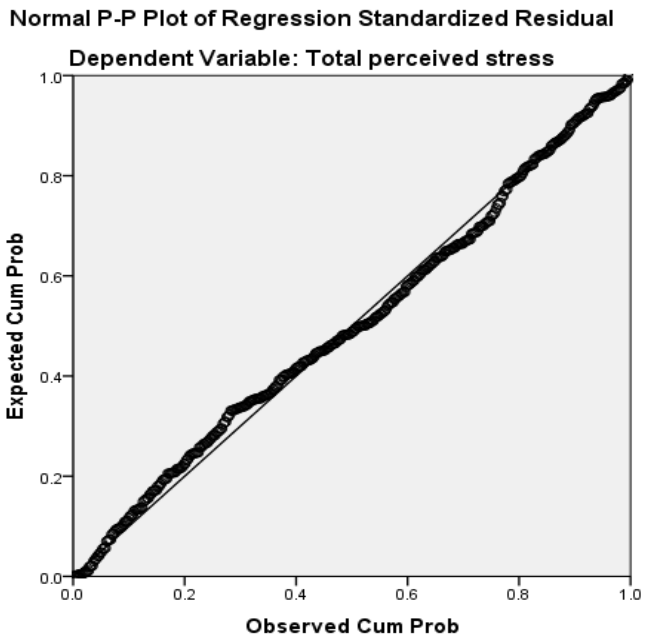
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	50.971	1.273		40.035	.000	48.469	53.474					
	Total PCOISS	-.175	.020	-.360	-8.660	.000	-.215	-.136	-.581	-.388	-.307	.729	1.372
	Total Mastery	-.625	.061	-.424	-10.222	.000	-.745	-.505	-.612	-.445	-.362	.729	1.372

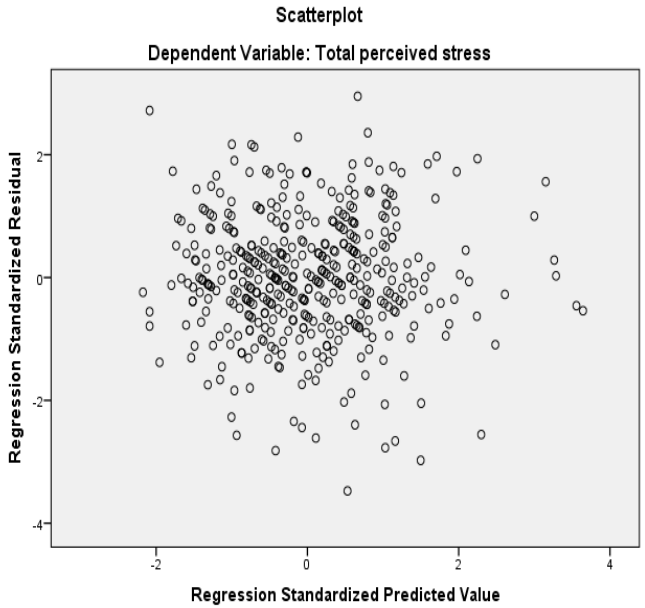
a. Dependent Variable: Total perceived stress

is an indicator of how much of the variability of the specified independent is not explained by the other independent variables in the model and is calculated using the formula 1–R2 for each variable. If this value is very small (less than .10), it indicates that the multiple correlation with other variables is high, suggesting the possibility of multicollinearity.

The other value given is the VIF (Variance inflation factor), which is just the inverse of the Tolerance value (1 divided by Tolerance). VIF values above 10 would be a concern here, indicating multicollinearity.

Check for the assumption of **normality and homoscedasticity**

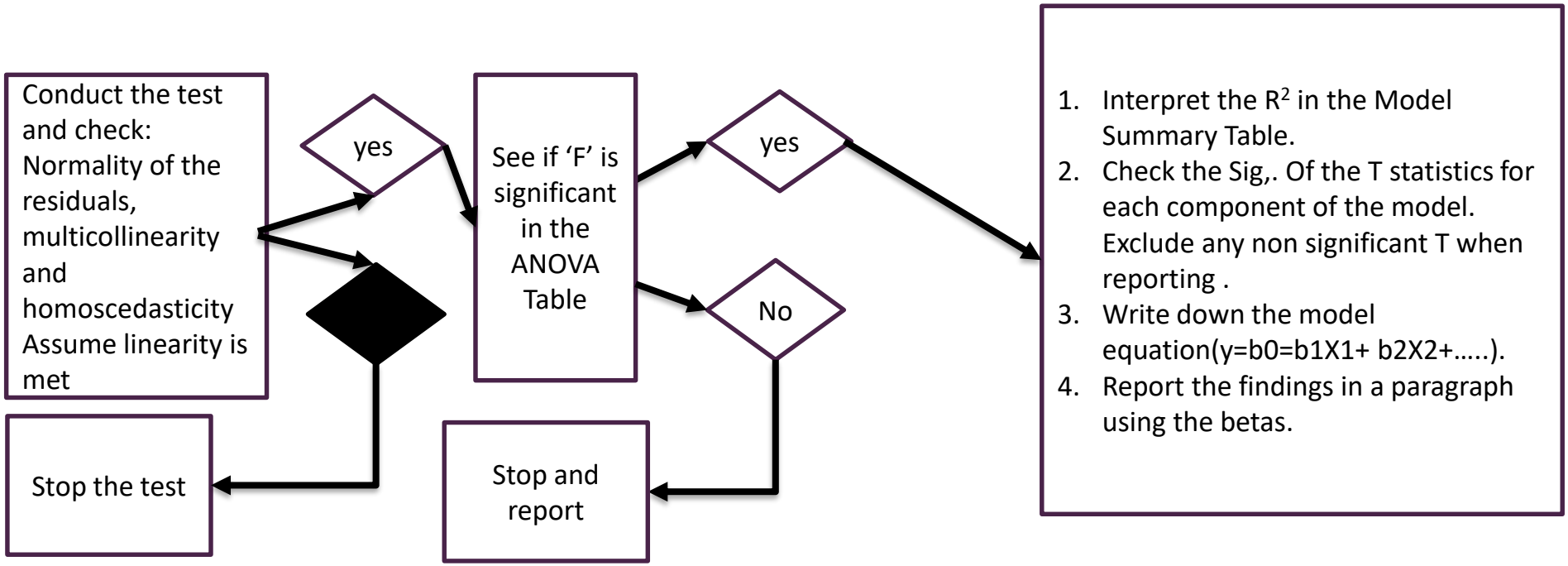




In the Normal Probability Plot you are hoping that your points will lie in a reasonably straight diagonal line from bottom left to top right.

In the Scatterplot of the standardised Residuals you are hoping that the residuals will be roughly rectangularly distributed, with most of the scores concentrated in the centre (along the 0 point).

Multiple Linear Regression



Interpretation of the output

- Evaluating the Model

ANOVA[a]

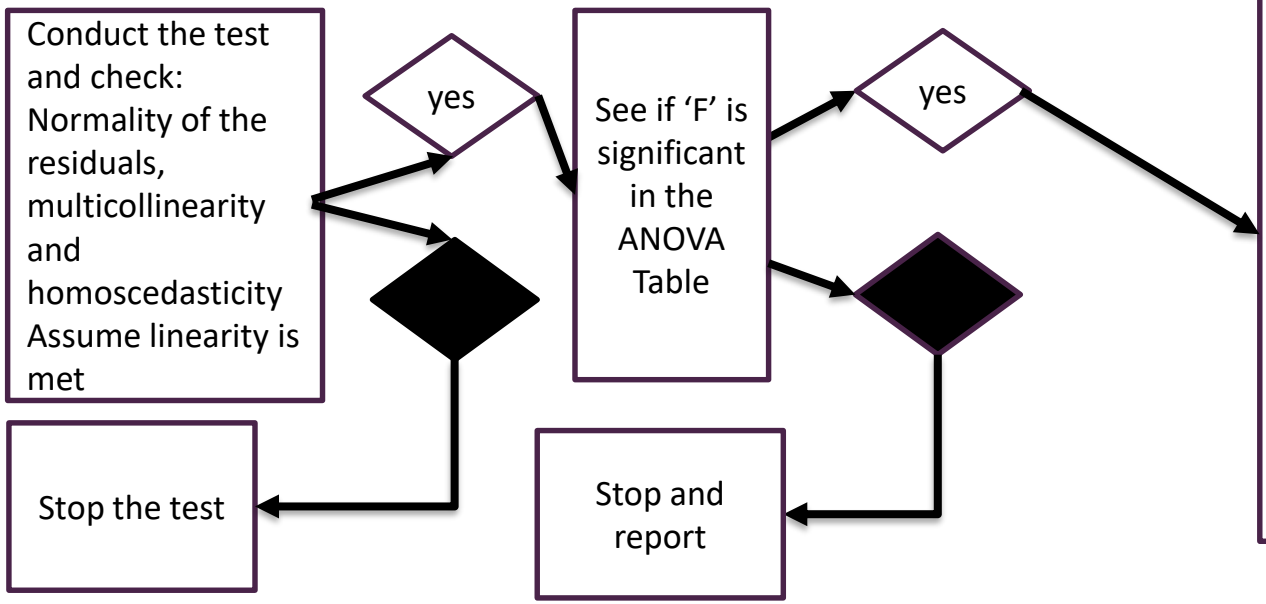
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6806.728	2	3403.364	186.341	.000[b]
	Residual	7725.756	423	18.264		
	Total	14532.484	425			

a. Dependent Variable: Total perceived stress

b. Predictors: (Constant), Total Mastery, Total PCOISS

(Sig = .000, this really means p<.0005).

Multiple Linear Regression



1. Interpret the R^2 in the Model Summary Table.
2. Check the Sig,. Of the T statistics for each component of the model. Exclude any non significant T when reporting .
3. Write down the model equation(y=b0=b1X1+ b2X2+…..).
4. Report the findings in a paragraph using the betas.

Interpretation of the output

- Evaluating the Model

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.684[a]	.468	.466	4.274

a. Predictors: (Constant), Total Mastery, Total PCOISS

b. Dependent Variable: Total perceived stress

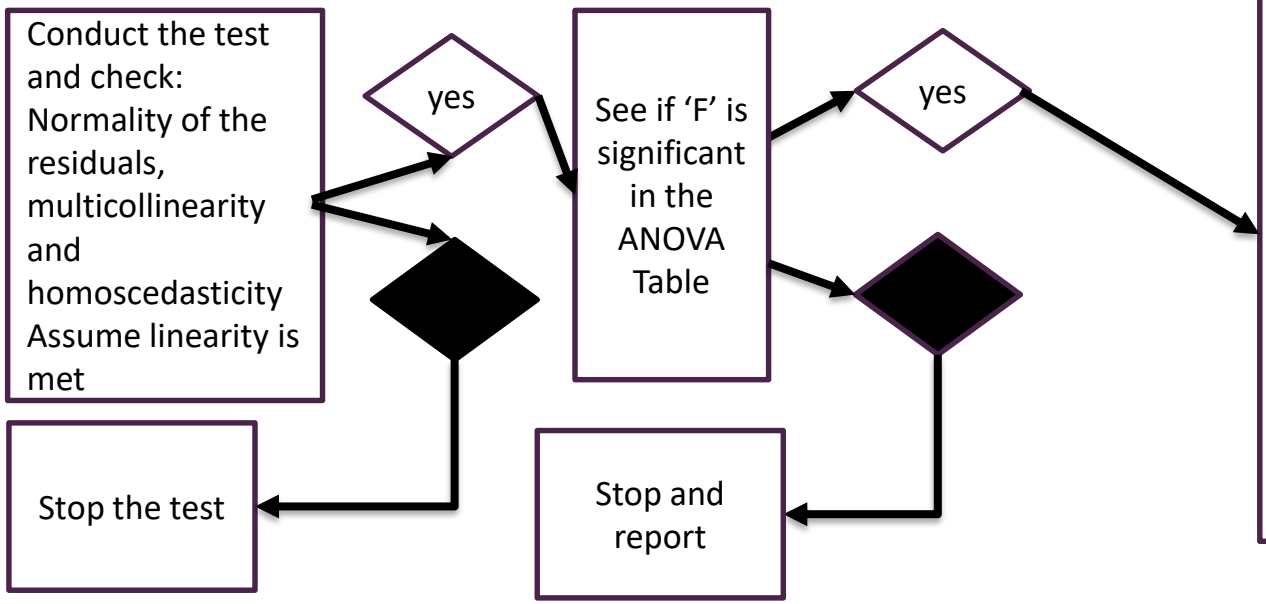
How much of the variance in the DV (stress) is explained by the model.

When a small sample is involved, the R square value in the sample tends to be a rather optimistic overestimation of the true value in the population. The Adjusted R square statistic 'corrects' this value to provide a better estimate of the true population value. If you have a small sample you may wish to consider reporting this value, rather than the normal R Square value.

Research question

1. How well do the two measures of control (mastery, PCOISS) predict perceived stress? How much variance in perceived stress scores can be explained by scores on these two scales?

Multiple Linear Regression



1. Interpret the R^2 in the Model Summary Table.
2. Check the Sig,. Of the T statistics for each component of the model. Exclude any non significant T when reporting .
3. Write down the model equation(y=b0=b1X1+ b2X2+…..).
4. Report the findings in a paragraph using the betas.

Interpretation of the output

- Evaluating each of the independent variables

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	50.971	1.273		40.035	.000	48.469	53.474					
	Total PCOISS	-.175	.020	-.360	-8.660	.000	-.215	-.136	-.581	-.388	-.307	.729	1.372
	Total Mastery	-.625	.061	-.424	-10.222	.000	-.745	-.505	-.612	-.445	-.362	.729	1.372

a. Dependent Variable: Total perceived stress

Research question
2. Which is the best predictor of perceived stress: control of external events (Mastery scale), or control of internal states (PCOISS)?

If you square this value (whatever it is called) you get an indication of the unique contribution of that variable to the total R squared

(Sig = .000, this really means p<.0005) for each IV

Standardized coefficients refer to how many standard deviations a dependent variable will change, per standard deviation increase in the predictor variable. Standardization of the coefficient is usually done to answer the question of which of the independent variables have a greater effect on the dependent variable in a multiple regression analysis, when the variables are measured in different units of measurement (for example, income measured in dollars and family size measured in number of individuals).

Interpretation of the output

- More interpretations

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	50.971	1.273		40.035	.000	48.469	53.474					
	Total PCOISS	-.175	.020	-.360	-8.660	.000	-.215	-.136	-.581	-.388	-.307	.729	1.372
	Total Mastery	-.625	.061	-.424	-10.222	.000	-.745	-.505	-.612	-.445	-.362	.729	1.372

a. Dependent Variable: Total perceived stress

This relationship is in the original units (scores of PCOISS, and scores of Mastery). This is useful for predicting things in the real world, but it is difficult to compare different predictors. Predictors might have large B values just because they are measured on a larger scale (compare minutes to hours in the above example).

Interpreting Estimated Coefficient

tpstress = 50.971 + **(**-.175 × tPCOISS) + (-.625 × t Mastery)

This is the model of prediction of the tpstress by the two variables.

Reporting the results (APA style)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6806.728	2	3403.364	186.341	.000[b]
	Residual	7725.756	423	18.264		
	Total	14532.484	425			

a. Dependent Variable: Total perceived stress

b. Predictors: (Constant), Total Mastery, Total PCOISS

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.684[a]	.468	.466	4.274

a. Predictors: (Constant), Total Mastery, Total PCOISS

b. Dependent Variable: Total perceived stress

Multiple regression analysis was used to test if the two measures of control (mastery, PCOISS) significantly predicted the perceived stress predicted. The results of the regression indicated the two predictors explained 46.8% of the variance (R^2=.468, $F(2,423)=186.341$, $p<.05$). It was found that Mastery significantly predicted total perceived stress (β= -.424, p<.05), as did PCOISS (β= -.36, p<.05).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	50.971	1.273		40.035	.000	48.469	53.474					
	Total PCOISS	-.175	.020	-.360	-8.660	.000	-.215	-.136	-.581	-.388	-.307	.729	1.372
	Total Mastery	-.625	.061	-.424	-10.222	.000	-.745	-.505	-.612	-.445	-.362	.729	1.372

a. Dependent Variable: Total perceived stress

Hierarchical Multiple Regression

- While standard multiple regression considers all the predictors at once, hierarchical multiple regression is used when researchers are interested in looking at the influence of several predictor variables in a **sequential** way.
- That is, they want to know what the prediction will be for the first predictor variable, and then how much is added to the overall prediction by including a second predictor variable, and then perhaps how much more is added by including a third predictor variable, and so on.

Hierarchical (sequential) multiple regression

- Question 3: If we control for the possible effect of age and socially desirable responding, is our set of variables (Mastery, PCOISS) still able to predict a significant amount of the variance in perceived stress?
- Procedure
- - From the menu at the top of the screen click on: Analyze, then click on Regression, then on Linear.
- - Click on your continuous dependent variable (e.g. total perceived stress: tpstress) and move it into the **Dependent** box.
- - Move the variables you wish to control for into the **Independent** box (e.g. age, total social desirability). This will be the first block of variables to be entered in the analysis (Block 1 of 1).
- - Click on the button marked **Next**. This will give you a second independent variables box to enter your second block of variables into (you should see Block 2 of 2).
- - Choose your next block of independent variables (e.g. total mastery, Total PCOISS).

Hierarchical (sequential) multiple regression

- Procedure>>> Cont.
- - In the **Method** box make sure that this is set to the default (**Enter**). This will give you standard multiple regression for each block of variables entered
- - Click on the **Statistics** button. Tick the boxes marked **Estimates**, **Model fit**, **R squared change, Descriptives**, **Part and partial correlations** and **Collinearity diagnostics**. Click on **Continue**.
- - Click on the **Options** button. In the **Missing Values** section click on **Exclude cases pairwise**.
- Click on the **Save** button. Click on **Mahalonobis** and **Cook's**. Click on **Continue** and then **OK**.

Evaluation of the model

The new variables only

All the variables

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	.238[a]	.057	.052	5.69	.057	12.711	2	423	.000
2	.688[b]	.474	.469	4.26	.417	166.873	2	421	.000

a. Predictors: (Constant), AGE, Total social desirability

b. Predictors: (Constant), AGE, Total social desirability, Total Mastery, Total PCOISS

c. Dependent Variable: Total perceived stress

- Check the R^2 for the two models.
- Check the change in the R^2 in the two models.
- Check for the significance of the models in the ANOVA Table

Research question

If we control for the possible effect of age and socially desirable responding, is our set of variables (Mastery, PCOISS) still able to predict a significant amount of the variance in perceived stress?

Evaluation of the model

The new variables only

All the variables

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	.238[a]	.057	.052	5.69	.057	12.711	2	423	.000
2	.688[b]	.474	.469	4.26	.417	166.873	2	421	.000

a. Predictors: (Constant), AGE, Total social desirability

b. Predictors: (Constant), AGE, Total social desirability, Total Mastery, Total PCOISS

c. Dependent Variable: Total perceived stress

Mastery and PCOISS explain an additional 41.7 per cent (.417 × 100) of the variance in perceived stress, even when the effects of age and socially desirable responding are statistically controlled for. This is a statistically significant contribution, as indicated by the Sig. F change value for this line (.000). The ANOVA table indicates that the model as a whole (which includes both blocks of variables) is significant [F (4, 421)=94.78, p<.0005).

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	823.865	2	411.932	12.711	.000[a]
	Residual	13708.620	423	32.408		
	Total	14532.484	425			
2	Regression	6885.760	4	1721.440	94.776	.000[b]
	Residual	7646.724	421	18.163		
	Total	14532.484	425			

Evaluation of the independent variables

Murdoch UNIVERSITY

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	31.076	.983		31.605	.000					
	total social desirability	-.599	.140	-.209	-4.271	.000	-.228	-.203	-.202	.928	1.08
	age	-.031	.022	-.070	-1.438	.151	-.127	-.070	-.068	.928	1.08
2	(Constant)	51.922	1.366		38.008	.000					
	total social desirability	-.149	.108	-.052	-1.373	.171	-.228	-.067	-.049	.871	1.15
	age	-.021	.017	-.047	-1.239	.216	-.127	-.060	-.044	.860	1.16
	total mastery	-.641	.062	-.435	-10.286	.000	-.612	-.448	-.364	.699	1.43
	total PCOISS	-.160	.022	-.327	-7.373	.000	-.581	-.338	-.261	.635	1.57

a. Dependent Variable: total perceived stress

Scanning the Sig. column, there are only two variables that make a statistically significant contribution (less than .05). In order of importance they are: Mastery (beta=–.44, after rounding up) and Total PCOISS (beta=–.33, after rounding). Neither age nor social desirability made a unique contribution.