

Research Methods II

Multiple Regression

Stacey Coleman



● Recap so far

- Correlation to linear regression
- Correlation:
 - Parametric test
 - Non-experimental design
- Linear regression:
 - Used to discover the effect of one variable (x) on another variable

● Recap so far

- Linear regression is similar to correlation
 - However:
 - Correlation allows us to conclude how strongly 2 variables relate to each other (both in terms of magnitude and direction),
 - Linear Regression also tells us by HOW MUCH one variable will change if the other variable changes.
 - It aims to predict scores on one variable based on scores from another variable.

Terms to remember with regression techniques

CRITERION VARIABLE

- Same as the DV

PREDICTOR VARIABLE

- The same as the IV; the variable that is predicting a change in the DV

PREDICT

- Always looking to predict something with regression, not looking for differences

- **b** = The amount by which y changes, when x changes by 1 unit – this is represented by the slope or tilt of our straight line in a scatterplot.

- **a** = the intercept (also called the Constant) – the value of Y' when X is 0.

● Multiple Regression

- Used to explore the relationship between one criterion and **more than one** predictor variables
 - Standard multiple regression
 - Hierarchical (entered in Blocks)
 - Stepwise (to be avoided.. Adds/deleted predictors based on strength)
- MR techniques are ideal for investigating complex real-life research questions
 - Less used in lab-based experimental designs
 - Ideal for longitudinal prospective designs

● Multiple Regression

- MR often used to test how 'good' a theory is
 - To what extent do the variables identified by a theory explain a given behaviour?
 - Can add additional variables / a new element to the theory and test this

● Multiple Regression

○ Research questions:

- Which of a **range** of variables are most strongly related to a specific variable of interest?
- Do these variables do better than chance or simple mean scores at **predicting** scores on a variable of interest?
- How might these variables **combine** to predict a variable?

● Multiple Regression

- *Why not just carry out many simple correlations?*
 - We could – this would provide us with several Bivariate associations (correlation between two variables)
 - Still only a non-causal correlation produced
 - Will not indicate the overall success with all the variables used

● Multiple Regression

- Important extension of 'regression' over 'correlation':
 - **Combined** association effect of the predictor variables = multiple regression!
 - Treating the predictors (separate variables) as a combined variable in relation to the criterion
 - How 'associated' they are as a whole
 - How 'predictive' they are of the criterion as a whole
 - (nb. Also looks at each in isolation...)



Assumptions for data when conducting multiple regression

● Multiple Regression: Assumptions

- MR is one of the fussier statistical techniques and makes a number of assumptions about data
 - And not all that forgiving if they are violated

1. Continuous data is best (interval or ratio)

- Criterion must be interval or ratio
- Predictors should ideally be interval or ratio; Nominal data can be used, but some caution around this

2. Normal distribution of data (normality)

- Ensure check for outliers

Multiple Regression: Assumptions

3. Outliers

- MR particularly sensitive to outliers
- Always check for extreme scores during cleaning and screening (for predictors and criterion)
- Tabachnick and Fidell (2013): outliers are those with standardised residuals above 3.3 (or less than -3.3)

4. Linear relationship between predictors and criterion (linearity)

- Non-linear relationships will require transformation of data

5. Homogeneity of variance (Homoscedasticity)

- Refers to the variance of errors being the same across all IV's
- Check using residuals scatterplot

Multiple Regression: Assumptions

6. Multicollinearity

- Exists when predictors are highly correlated with one another ($r = .90+$)
- Ideally should be avoided as created problems when trying to infer relative contribution of individual predictors
- We want predictors to be correlated with criterion, but not necessarily with other predictors

7. Singularity

- When one predictor is actually a combination of other predictors (e.g., when subscale AND total scores are used)

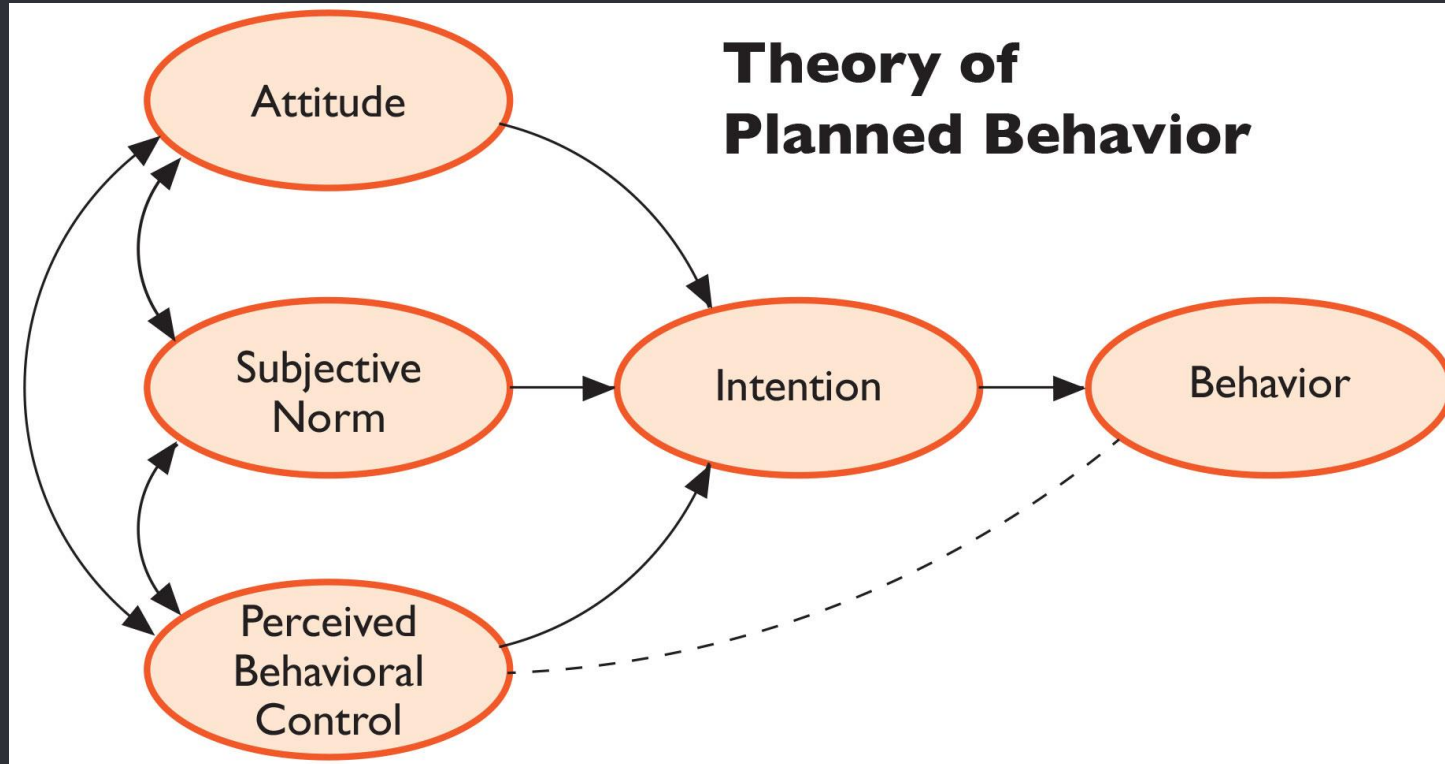
8. Sample size / number of observations

- Estimate of *adjusted R square* in regression techniques depends on the number of participants and the number of predictor variables
- Stevens (1996): approx. 15 participants per predictor
- Tabachnick and Fidell (2013) $N > 50 + 8m$

● How do you choose which variables to use as 'predictors' in a regression design?

- Regression techniques can be used to analyse a more 'causal' relationship between variables
- Regression analyses the amount of change in the criterion based on change in the predictor variables
- Can be used to make **predictions** for the criterion based on scores from a range of predictors

Multiple Regression Example



Testing the theory of planned behaviour components
to **predict** intentions to engage in sun-safety
behaviour

● Multiple Regression: Assumptions

- Study example: Do the TPB components predict intentions to engage in sun-safety behaviour?
 1. To what extent does **TPB model as a whole** predict intentions to engage in sun-safety behaviour?
 2. To what extent do each of the **individual predictor variables** predict intentions?
 3. Which predictor variable is the **BEST** predictor?

● Conduct study, run analysis, CHECK ASSUMPTIONS!

- **Multicollinearity** (correlations table)
 - Your predictors should show at least some relationship with your criterion (preferably **above 0.3**)
 - You don't really want your predictors to be too highly correlated with one another – anything **below 0.7** in this case is considered ok
- Collinearity diagnostics can also be requested and performed by SPSS – presented in the coefficients table labelled TOLERANCE and VIF
 - Tolerance – value above .10
 - VIF – value below 10

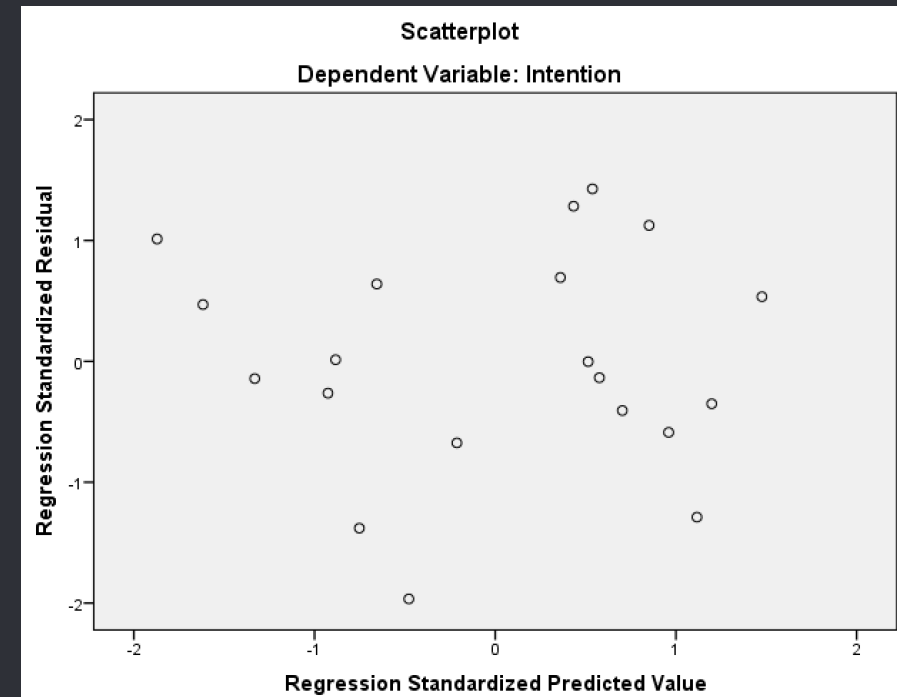
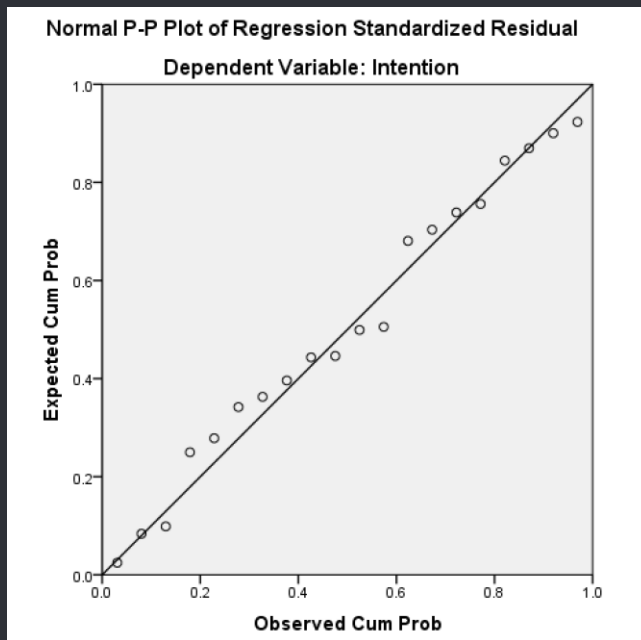
Coefficients ^a						
Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
			Lower Bound	Upper Bound	Tolerance	VIF
	.302	.767	-4.139	5.513		
.741	6.578	.000	.879	1.716	.766	1.305
.147	-1.037	.315	-.972	.333	.488	2.051
.243	1.579	.134	-.117	.798	.410	2.439

Tolerance:
Values should be above .10 – not violated assumption of collinearity

VIF:
Values should be below 10 – not violated assumption of collinearity

Next: deal with outliers and normality

- Multiple regression is very sensitive to outliers!
- Inspect and interpret this assumption mainly using the plots created by SPSS



● Main outputs for Multiple Regression

○ 1. MODEL SUMMARY TABLE

What percentage of the variance is explained by the model as a whole?

2. ANOVA TABLE

Is the model significantly better at predicting intentions to engage in sun-safety behaviour than simply making predictions using the mean score for the DV?

3. COEFFICIENTS TABLE

Which of the individual predictor variables predicts intentions to engage in sun-safety behaviour? How much change in each predictor variable equates to change in the criterion (DV)

● Main outputs for Multiple Regression

1. MODEL SUMMARY TABLE

What percentage of the variance is explained by the model as a whole?

- MR indicates how good a 'fit' all the predictors are in explaining scores in the criterion
 - Can assess the degree to which the criterion can be predicted from the set of predictors
- R – multiple regression coefficient
 - Between 0 and 1
 - The greater the R , the less residuals; stronger measure of the correlation between the observed value and value of the criterion

● Main outputs for Multiple Regression

○ 2. ANOVA TABLE

Is the model significantly better at predicting intentions to engage in sun-safety behaviours than simply making predictions using the mean score for the DV?

- Report whether the model is significant in its prediction of criterion scores
- Reported as an ANOVA

● Main outputs for Multiple Regression

○ 3. COEFFICIENTS TABLE

Which of the individual predictor variables predicts intentions to engage in sun-safety behaviour? How much change in each predictor variable equates to change in the criterion (DV)

- Beta values represent the amount of contribution for each of the predictor variables
 - How good a fit
- Can determine which of the predictors (if any) are **significant unique predictors**, and also which predictor is the **strongest predictor**

● Main outputs for Multiple Regression

○ 3. COEFFICIENTS TABLE

Which of the individual predictor variables predicts intentions to engage in sun-safety behaviour? How much change in each predictor variable equates to change in the criterion (DV)

- *Unstandardised Beta values or Standardised Beta values??*
 - *Unstandardised values* useful for interpretation of the effect of the predictors on the DV, but cannot provide us with a way of comparing the effect of one predictor with another predictor in the model
 - Standardised values are useful when there are different units of measurements, and measured in units of standard deviation.
Beta value indicates the number of SD that scores in the criterion changes if there was a 1SD change in the predictor.
Multiple predictors – use standardised values – allows comparison of strength

- **Main outputs for Multiple Regression**

- Tutorial video and worksheets available on Moodle
- Practice!!