**Terms:**

In regression, independent variables are also referred to as predictor or explanatory variables and dependent variables as outcome, response or criterion variables.

**Types of regression we will experiment with:**

1. Multiple Linear Regression:

Used when we have two or more independent variables, and we want to determine their contribution to the dependent variable (how each predictor variable predicts increases/decreases in the outcome variable).

1. Hierarchical Multiple Linear Regression (HMR):

This type of regression is similar to multiple linear regression but with some added components. In non-alien words, hierarchical multiple regression explores how **each** predictor variable (IV) associates with the outcome variable (DV) whist accounting for the variation of other predictor variables.

For example, a study investigated if dreams impact our mood the next day. They collected data on dream recall frequency (how much of your dreams you remember), negative and positive dreams, and mood.

The correlations showed that participants who reported low dream recall frequency also reported a positive mood regardless of the nature of the dream, and those who reported a high dream recall frequency also reported a negative mood (again regardless of dream). Does this mean that dreams negatively affect mood? Here is where HMR comes in. HMR can control for the predictor variable ‘dream recall frequency’ so that we can isolate the effect of positive and negative dreams on mood. After controlling for dream recall frequency, they found that positive dreams were associated with positive mood and negative dreams with negative mood.

**Multiple Regression**

**Assumptions**

* Fairly large sample size: need 20 records for each predictor variables. This rule only applies if the dependent variable is normally distributed. If it is not, you would need more than 20.
* The dependent variable is normally distributed
* Absence of outliers in all the variables (Cook’s distance is less than 1)
* A linear relationship between the independent and the dependent variables
* Absence of multicollinearity between the independent variables

All these assumptions can be checked as part of the linear regression procedure except for the dependent variable been normally distributed. However, we already do **some** of these checks when we screen the data.

Multiple Regression in SPSS:

1. Analyse- Regression – Linear
2. Move dependent variable to dependent box
3. Predictors to independent(s) box
4. Under method: Enter
5. Under statistics: Check the boxes for Estimates, Confidence intervals (95%), Model fit, R squared change, Descriptives, Part and partial correlations, Casewise diagnostics
6. Click continue
7. Plots: ZRESID to Y-axis and ZPRED to X-axis
8. Check box for histogram and normal probability plot
9. Under save: Click on Cook’s
10. Click continue and ok

**Results in SPSS output**

**Checking for assumptions**

Most of the assumptions are covered through the data screening process so I am just going to focus on some.

The ‘Correlations’ table helps us to check for the assumptions of the linear multiple regression.

* Outliers: No outliers if the maximum Cook’s distance is less than 1
* Multicollinearity **between the predictors**: if a correlation (‘Pearson Correlation’) is greater than 0.7 then there is multicollinearity
* Relationship between independent and dependent variables: A linear relationship is determined by a value of 0.3 (or higher) of the Pearson correlation

Cook’s distance example table:

Table

Description automatically generated

In this example Cook’s distance is .168 which is less than 1 so we don’t have outliers. If we did have outliers, we would go back and remove them.

Looking for multicollinearity and linear relationships example:

Table

Description automatically generated

In the example there is no multicollinearity between the predictor variables. Also, there is a linear relationship between the predictor variables and the outcome variable.

**Interpretation of Main Results**

Model Summary

The ‘Model Summary’ table shows r square adjusted and adjusted r square.

* R2adj: tells you how much of the variance in the dependent variable your model explains. A value of 0.3 is a good fit.

Example:

Table

Description automatically generated

**Reporting:** Taken as a set, the predictors (list them) depression, anxiety and substance use account for 33% of the variance in functioning (outcome variable).

ANOVA

The ‘ANOVA’ table tells you if the r value (reported above) is significant. If it is, (p <0.05) we can reject the null hypothesis.

Example:

Table

Description automatically generated

**Reporting:** The overall regression model was significant, *F*(Regression df for F, and Residual df for F) = F value, *p*-value, *R*2adj = (R square adjusted value from model summary table)

e.g., The overall regression model was significant, *F*(3,96) = 16.01, *p*<.001, *R*2adj = .31

Coefficients

The ‘Coefficients’ table tells you how important each predictor is to the outcome. Basically, it tells you if a predictor on its own is significant.

From this table we take the p-value for each independent variable.

Example:

Table

Description automatically generated

**Reporting:** Is the amount of unique variance a predictor accounts for statistically significant? If p < .05 then yes, it is.

e.g.,

How to do multiple regression with dichotomous independent variables (e.g., gender):

* If you have two categories in that variable (e.g., males and females) then follow the same procedure as above.
* If you have more than two (e.g., ethnicity) then you need to re-express that variable prior to entering it in the regression
* This can be done by splitting that variable into multiple variables based on its number of categories
* E.g., if you have ethnicity with 4 categories, you would need to create a variable for each category

**Important note:** To do this, the dependent variable needs to be quantitative (scale). If it was categorical, then logistic or multinomial regression would need to be run.

What happens when r-square is significant but none of the predictors are significant?

* This means that we had multicollinearity in the predictors (they were correlated with each other)

**How to report:**

1. Describe the test you carried out and why you did it. Mention all of your variables.

* e.g., *A multiple regression was conducted to see if intelligence level and extroversion level predicted the total value of sales made by salespeople per week*

Were the variables entered simultaneously? (i.e., Method: Enter). – this should be included in the statistical analysis sections in the methods

**Hierarchical Multiple Regression**

* Predictor variables are entered into the regression model in an order determined by past research and expectations. The variables are entered in steps (known as ‘blocks’) with each variable being assessed on what it adds to the prediction of the outcome variable.
* The variables entered in first, are the ones being controlled for.

**Assumptions:**

* The same as multiple linear regression

HMR in SPSS:

1. Analyse – Regression – Linear
2. Move the outcome variable to the Dependent box
3. Move the predictor variable(s) you want to control to the Independent(s) box
4. Click next (this creates block 1)
5. Now move your predictor variables you want to isolate to the Independent(s) box
6. Click next (this creates block 2)
7. Under Statistics check the boxes for: Estimates, Model fit, R squared change, Descriptives, Part and partial correlations, Collinearity diagnostics, Casewise diagnostics and click continue
8. Under Plots move the ZRESID to the Y-axis and the ZPRED to the X-axis and check the box for Normal probability plot.
9. Click next and move DEPENDENT to Y and ADJPRED to X - continue
10. Under Save, Check the box for Cook’s distance (and leave the rest as they are)
11. Click ok

The output is similar to that of normal multiple linear regression so you can check for your assumptions.

**Main findings**

1. Model summary table

This time you will get more than 1 models, depending on your number of blocks

One of the things we want to look for (other than those shown in the multiple regression above) is the R square change between the models. This will tell us how much of the variance the controlled predictors accounted for and how much of the variance was accounted by all the predictors together.

Example:

Table

Description automatically generated

Model 1: Shows the R-values for when controlling a predictor variable (i.e., Family support)

Model 2: Shows the R-values for all the predictor variables (i.e., Family support, Depression, Anxiety)