

MULTIPLE REGRESSION (HIERARCHICAL) ANALYSIS

Does the model fit the data well?

1. F ratio is large (at least > 1) and statistically significant
2. R^2 is large and significantly improves between models (using an F-test)
3. Akaike information criterion (AIC) is small (compared between models using the same data)
4. No substantial outliers
 - a. 95% of z scores are in between -1.96 and 1.96
 - b. 99% of z scores are in between -2.58 and 2.58
 - c. 99.9% of z scores are in between -3.29 and 3.29
5. No substantial influential cases
 - a. Problem if Cook's distance score > 1
 - b. Problem if leverage value $> 3 [(\text{predictor \#} + 1)/n]$
 - c. Problem if Covariance Ratio $> 1 + 3[(\text{predictor \#} + 1)/n]$ (i.e. upper limit) or if Covariance Ratio $< 1 - 3[(\text{predictor \#} + 1)/n]$ (i.e. lower limit)
 - d. DFBeta (plot and inspect)
 - e. DFFit (plot and inspect)

Can the model be generalised to the population of interest?

Model assumptions need to be satisfied

1. **Variable type:** continuous or dichotomous nominal
2. **Linearity** of relationship between the outcome and predictor variables: Plot of standardised residuals and standardised predicted values should not resemble a curve
3. **Low multicollinearity among predictor variables:** Any single VIF should not be greater than 10, and the average VIF should not be substantially greater than 1
4. **Independence of residuals** (residuals should not be substantially correlated with one another): Durbin-Watson statistic should be close to 2
5. **Homoscedasticity** (equality in the variance of the residuals): Plot of the standardised residuals and standardised predicted values should be randomly dispersed, rather than a funnel shape
6. **Normal distribution of residuals:** Histogram of the standardised residuals should resemble a bell curve, and the normal probability (Q-Q) plot of the standardised residuals should resemble a diagonal line
7. **Predictor variables should not be substantially correlated with variables outside of the analysis**
8. **Predictor variables should have some variation**
9. **Each participant should not provide more than one dataset**

Note: This checklist is not exhaustive and should not be the only source of information used when conducting multiple regression analysis. This is merely a guide to help you get started. It is recommended that you learn what each of these checklist points mean, opposed to blindly following the rules specified above. Remember - there are no blanket rules/answers in statistics that apply to every situation.