SUMMARY OF WHAT EACH ANALYSIS TYPE DOES:

Correlations: Used to measure and describe the *linear relationship* between two continuously scored variables (can have 1 dichotomous – categorical variable with 2 levels)

Bivariate Regressions: Used to determine the ability for one variable's scores to *predict* scores on another variable

Multiple Regressions: Used to test 1+ IVs (predictors) against 1 DV (outcome variable) (DV can only be continuous, IVs can be continuous or dichotomous)

One-Way ANOVAs: Used to measure differences between >2 groups (using group means) DV = continuous, IV = categorical

Multiple Comparisons: Used to determine which specific groups differ from each other

Factorial ANOVAs: Used to compare means across 2 or more IVs against a DV

Repeated Measures ANOVA: Used to remove individual differences between participants which gives you an overall smaller error term and increased statistical power

WEEK THREE: MULTIPLE REGRESSION

Why is Multiple Regression Preferred over Bivariate Regression?

- More realistic modelling of relationships between variables (can show overlaps between IVs)
- Produces ONE equation that best predicts the DV from a set of IVs
- Allows us to determine:
    1. The combined influence of IVs on prediction of DV
    2. The relative contributions of IVs for prediction of DV
    3. Level of improvement in prediction if we add another IV (or several additional IVs)

General Multiple Regression Equation

- $Y_i = (b_0 + b_1X_{1i} + b_2X_{2i} + \ldots + b_nX_{ni}) + \varepsilon_i$
    - b_0 = y-intercept (a)
    - b = unstandardized regression coefficients (independent (or x) variables
        - Size of b is determined by:
            - The strength of the *unique* relationship between IV and DV
            - The scale if measurement of IV
        - Not useful for comparing relative importance of IVs or for determining unique variance explained by an IV
        - B ranges from $-\infty$ to $+\infty$
        - To see is B statistic is significant, need to convert it into a t value
        - *Beta (β) weights are the standardised version (z-scores)*
            - Useful for comparing the relative contributions of IVs for predicting the DV
            - Doesn't tell us the unique variance of IVs
            - Eg. $\beta = 0.53$: A one standard deviation increase in the IV results in a 0.53 standard deviation increase in the predicted DV score
- Multiple regression produces an equation that best predicts the Y or dependent variable (criterion variable) from several independent variables (predictor variables)
- If we used the regression equation with beta weights (standardized coefficients), the predicted values for the DV would be in standardized format (Z scores, with Mean = 0 and SD = 1), and we could get DV values that are plausible as standardized values, but don't make sense in light of the original metric of the DV
- The unstandardized regression equation is used for predicting scores on the DV in the original (and meaningful) scale

Multiple Correlation Coefficient (R^2)

- "Squared multiple correlation" – determines how well the linear regression equation fits the data (ie. the amount of variance predicted in the DV)
    - *Multiple R* – correlation between the IVs and the DV
    - To see if multiple R is significant, need to convert it to an F statistic
- Unsquared, the multiple correlation (R) is the correlation between the observed and predicted Y values of the dependent variable

- The multiple correlation can be expressed as $R=r_{yy'}$, where $r_{yy'}$ is simply the correlation between actual values on y (i.e., the values participants gave) and predicted values (y')
- Mathematically, total variance can be calculated by: $SS_T = \sum(Y - \overline{Y})^2$
  - SS_T = total sum of squares difference
  - Y = individual participants' scores on the DV
  - $\overline{Y}$ = represents the mean score for the DV
  - *If we didn't square the mean difference, the sum of differences would equal 0*
- $SS_M = \sum(\hat{Y} - \overline{Y})^2$
- $SS_R = \sum(Y - \hat{Y})^2$
- $R^2 = \frac{SS_M}{SS_T}$

Adjusted R^2

- Modified measure of R^2
- R^2 is known to be artificially inflated by: – adjusted R^2 aims to correct this
  - Large number of IVs in the model
  - Small sample size
- Gives a projected R^2 value that you might expect from the population that the sample drew upon
- *adj* $R^{*2} = 1 - \frac{(1-R^2)(N-1)}{N-k-1}$
  - N = number of participants
  - K = number of predictor variables (IVs)
- If we are interested in estimating the effect size for a different sample from the same population, Field recommends that this formula is used instead:
  - adj $R^{*2} = 1 - \left[\left(\frac{n-1}{n-k-1}\right)\left(\frac{n-2}{n-k-2}\right)\left(\frac{n+1}{n}\right)\right](1-R^2)$

Research Uses of Multiple Regression

- Regression analysis is used to evaluate how a dependent variable is related to several other independent variables
- Can identify the contribution of each individual independent variable to the overall relationship
- Can control for the effects of certain independent variables to determine the subsequent contribution to the equation from other variables of interest
- Designed primarily for continuous IVs but regression can handle categorical data if converted into dichotomous variables
- *Dummy variable coding* – to turn the IV into dichotomous variable