Advanced research notes

Chapter 18: Bivariate correlation

Bivariate correlation

- Relationship between two variables
- Most common concern linear relationships

Types of relationships

- Positive
- Negative
- No relationship
- Perfect

Three possible correlations

- X causes y
- Y causes x
- Another variable

Correlation coefficients

- Statistical measures of correlation
- Pearson's r
  - is the most common measure of a linear relationship
  - used when both variables have at least interval levels of measurement
  - is a parametric linear correlation coefficient
  - has the assumption of normality

Dichotomous variables

- Only two possible values ie gender

Nonparametric correlation coefficient

- Spearman's rho
- Used when one or both variables have only an ordinal measurement level and normality assumption is violated

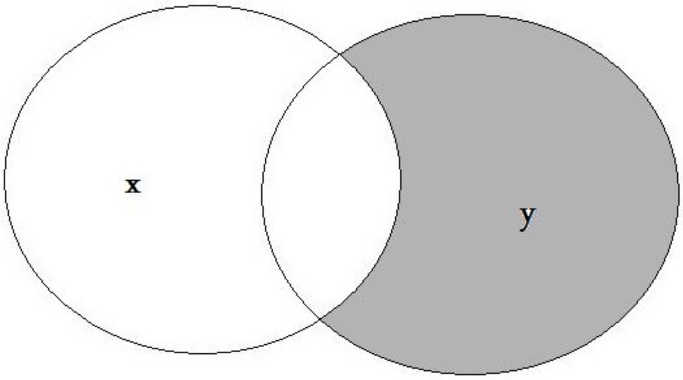
Strength and direction of correlation coefficients

- -1 to 1 (number = magnitude of strength of relationship, sign = direction)
  - .00-.09 – weak/negligable
  - .10 - .29 – small
  - .3 - .49 – moderate
  - .5 - .69 - strong
  - .7 – 1 – very strong

The squared correlation coefficient/coefficient of determination

- r^2 = indicates the proportion of variance in one variable predicted by the other and vice versa
- if r = .5, r^2 = .25 (25% explained by IV, 75% other variables)

- r = sample population, p = population correlation
- represented in a venn diagram – overlap = shared variance



Calculating a correlation coefficient

- Deviation score formula: provides best conceptual understanding of what the coefficient actually does
    - Step 1: combined variance (covariance) calculated
    - Step 2: covariance is standardised
    - Degrees of freedom = N-1
- Raw score formula: most suitable for manual calc

Assumptions

- relationship is linear
- distribution is equal, that is relationship between variables should be homoscedastic not heteroscedastic
- no restricted range'
- no outliers
- not using extreme groups
- participants randomly sampled and scores independent of each other
- N > 30

Reporting results

- Correlation reported as r (df of N – 2) = correlation coefficient without leading 0, p = sig value (with alpha previously specified)
- R(12) = .87, p = < .001
- Means and SD reported to indicate range
- Same follows for bivariate

Chapter 19: Bivariate Regression

- When variables share a relationship, it is possible to make predictions
- Prediction is based on calculating line of best fit/regression line
- Regression line: line on scatterplot that is closest on average to all the observation points
- Outcome/criterion: variable you wish to predict (y axes)
- Predictor: the variable used to predict (x axes)

Predicting scores from a regression line

- The smaller the correlation the more inaccurate the prediction
- The least regression line: the line calculated so that it minimises the sum of the squared risiduals

Line of best fit equation

Y = a + bX

- a = constant/intercept (point where the line intercepts the y axes
- b = b weight/regression coefficient (the slope of the line)

The standard error of the estimate

- Final figure given in the SPSS model summary
- Similar to SD in univariate distributions
- Corresponds to the average amount of error in predicted Y scores
- The higher the correlation the smaller the standard error prediction

Predicting Y from X and X from Y

- Normally predict Y from X
- But to reverse we use a separate regression analysis and calculate a new scatterplot

Assumptions and uses of linear regression

- Linear regression is an extension of correlation analyses and has similar assumptions
- relationship is linear
- distribution is equal, that is relationship between variables should be homoscedastic not heteroscedastic
- no restricted range'
- no outliers
- not using extreme groups
- participants randomly sampled and scores independent of each other
- $N > 30$

Percentage of variance

- R2 = indicates the proportion of variance in one variable explained by the other and vice versa