Statistics
Chapter Nine

· Pearson's R
→ A descriptive index that summarizes magnitude and nature (direction) of a relationship b/wn two high-level variables in a sample
→ Can also be used to make infrences about relationships in the population

· Basic Hypotheses
→ Correlational hypotheses are about p (rho), the population correlation coefficient
→ Basic null hypothesis: rho is zero
→ The alternative hypothesis is the opposite
  → H_1: $p \neq .00$ (positive or negative)

· Sampling Distribution (Theoretical distribution)
→ The mean of a sampling distribution of the correlation coefficient is rho, the population coefficient
→ When the null hypothesis is true (rho = .0):
  → The theoretical sampling distribution is centred on .00
  → The sampling distribution is approximately normal

· Assumptions and Requirements
→ Pearson's R is suitable for interval & ratio (high level) variables and detecting linear relationships
→ Pearson's R can be used inferentially:
  → If the variables have an underlying distribution that is bivariate normal. Pearsons R is robust to this if sample size > 15
  → If values of both variables are homoscedastic (for each value of X, variability of Y scores about the same) tested using Levene's test

Chapter Nine

- Testing Significance

→ Value of computed r must be compared to critical values in a table for which df is known & significance is established

→ Pearson's r df : N - 2

- r to z Transformation

→ Testing differences btwn two correlations requires that the two correlation coefficiences be transformed

→ Then use normal distribution to see if values are significantly different

- Magnitude of Effect

→ r^2 = magnitude of effect

→ Tells you the proportion of the variability in one variable that is explained by the other variable

→ r^2 is sometimes called the coefficient of determination

→ r^2 is analogous to eta^2 : it represents the ratio of explained varience to total varience

→ $r^2 = SS_{Explained} \div SS_{Total}$

- Precision and Pearson's r

→ Confidence intervals can be built around the value of r to indicate the precision of the population estimate

→ Eg $\bar{w}$ a sample of 50, the 95% CI around r = .26 is -.02 to .50

→ This includes the possibility that the population correlation is zero - the null hypothesis

Chapter Nine

· Power and Pearson's r

→ In power analysis, r = effect size index

→ Table can be used to estimate sample size needed (to minimize the risk of a Type II error)

→ As a last resort, small, medium, and large effects corresponds to rs of .10, .30, and .50 respectively

→ This corresponds to needed Ns of 785, 85, and 29

· The Factors Affecting r

→ The magnitude of r can be affected by:

→ Having variables w̄ restricted ranges of values

→ Using groups at both extremes of a distribution of values

→ Having outliers in the data

→ Measuring the variables w̄ instruments having low reliability

· Nonparametric Correlations

→ Spearman's rho (r_s): Based on ranks of the original data values

→ Kendall's tau (τ): A complex formula, sometimes a preferred index bc of its statistical properties

→ Both range from -1.00 (through .00) to 1.00

· Pearson's r Formula

→ $$r_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[\Sigma(x-\bar{x})^2][\Sigma(y-y)^2]}}$$

Chapter Nine

- Regression: Technique used to analyze relationships btwn variables and to make predictions about values of variables

- Simple linear regression involves regressing one variable (Y) on another variable (X)
  - → Y: the dependent (outcome) variable · trying to predict
  - → X: the independent variable · predictor variable
  - → The goal is to predict new values of Y based on values of X

- linear regression builds on the equation for a straight line because the relationship btwn the two is assumed to be linear
  - → A straight line should yield the best "fit" of the data points in a scatterplot

- Any straight line can be described by:
  - → $Y = a + bX$
    - → Y: values on one variable
    - → X: Values on the other variable
    - → a: the intercept concept (where line crosses Y axis)
    - → b: the slope of the line

- Regression Equation
  - → $Y' = a + bX$
    - → Y' = predicted value of Y
    - → X = actual value of X
    - → a = intercept constant
    - → b = slope of the line (regression coefficient)

- $b = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$ · $a = \bar{Y} - b\bar{X}$

Solve 1st

Chapter Nine

- Prediction and Regression
  → Regression equations yield predictions of new values of Y based on known values of X
    → Compare predicted Y to actual Y

- Errors of Prediction: differences btwn actual and predicted values of Y, also called residuals, symbolized as e

- The regression equation uses a least-squares criterion in solving for a & b
  → The squares of the errors of prediction (e^2) are minimized