Statistics
Chapter Seven

- Analysis of Varience (ANOVA): used to compare group means when t-tests are not appropriate eq. when three or more groups are being compared
  - → Dependent variable must be interval or ratio level (high-level)
    - → Eq. heart rate, scores on self esteem scale
  - → Independent variable is nominal (eq. three racial/ethnic groups) or could be ordinal w̄ a small number of categories (eq. normal weight, overweight, obese)

- ANOVA null hypothesis: group means are equal
  - → Eq. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
  - → ANOVA alternate hypothesis: at least one of the group means are not equal
    - → Eq. H_1 : not H_0

- Assumptions for ANOVA
  - → Random sampling from the populations
  - → Dependent variable is normally distributed in the populations
  - → Variences in the populations are equal
    - → Can be tested w̄ Levene's Statistic
  - → Robust to violation of normality assumptions if n per group is greater than 20
    - → Eq. if variables are not normally distributed it doesn't matter as long as there's at least 20 people
  - → Robust to homogeneity assumption if ns are similar

Chapter Seven

· Partitioning Varience
→ Varience is associated w̄ different things
→ Involves isolating "reasons" why people's scores might differ from one another
→ Some of the reasons people differ is because of their individual variability
→ But another reason might be because of the independent variable

· Sources of Variation
→ Between-group varience: differences between the groups being compared
→ Within-group varience: Individual differences among people in the groups

· In ANOVA the computed statistic is the F-statistic
→ F = Between-group varience / within-group varience

or

→ F = Effect of IV + Sampling error / Sampling error
→ Sampling distributions of F are asymmetric around the value of 1.0

· One-way ANOVA
→ Used for comparing means of three or more independent groups
→ Computations involve deviations of scores from the group means and the overall grand means

· Step 1 in calculating ANOVA: Sum of squares
→ Sum of squares within: total of the squared deviations of each persons score from their group mean, $\Sigma (x - \bar{x})^2$
→ take each value, subtract the mean, square it, add the values up
→ Sum of squares between: the total of the squared deviations of each group mean from the grand mean

Chapter Seven

· Step 2 in Calculating ANOVA: Mean Square
→ Mean Squares within (MS_W): the sum of squares within divided by df_{within}
  → df_{within} = N - number of groups
→ Mean Squares between (MS_B): the sum of squares between divided by $df_{between}$
  → $df_{between}$ = number of groups - 1

· Step 3 in Calculating ANOVA: Testing F
→ $F = MS_B / MS_W$
→ This computed F is then compared to a table of critical values for the appropriate degrees of freedom and significance criterion
→ If calculated F ≥ than tabled F the results are statistically significant
  → The null hypothesis can be rejected

* Sum of square between extra step
→ Take the mean from each group and subtract the grand mean, then square that value, multiply by number of people in that group

· If anova calculated in spss, & significance level is .05 or smaller it is significant and the null hypothesis can be rejected

· Multifactor ANOVA: ANOVA can be extended to situations in which there are multiple independent variables
→ Most typical is a two way ANOVA which involves 2 distinct IVs
→ In two-way ANOVA each IV is sometimes called a factor

Chapter Seven

· Hypothesis in Two-Way ANOVA
→ Null hypothesis for main effects
→ Null hypothesis for Factor A: $H_0: \mu_{A1} = \mu_{A2}$
→ Null hypothesis for Factor B: $H_0: \mu_{B1} = \mu_{B2}$

· The interaction hypothesis is about the joint effects of the factors - it concerns whether the effect of one IV is consistent at all levels of a second IV
→ The null hypothesis is that means are equal for all cells
→ $H_0: \mu_{A1B1} = \mu_{A2B1} = \mu_{A1B2} = \mu_{A2B2}$

· Repeated Measures ANOVA: the same thing as a two-sample t-test for dependent groups (RM-Anova)
→ Three sources of variation
→ Total variability (SS_{Total}) is comprised of:
→ A "treatment" effect ($SS_{Treatment}$) same as between factor
→ Error variation (SS_{Error}) same as within
→ A "subjects" factor ($SS_{subjects}$)

· Sums of Squares Subjects:
→ Variability for subjects ($SS_{subjects}$) reflects individual differences the tendency of people to be different from each other, but consistent across different conditions or at different times
→ Individual differences (subject variability) can be captured because the same people are measured multiple times
→ $SS_{subjects}$ can be removed from the "error" term for calculating the F statistic making it a more powerful test

· F-Test for RM-ANOVA
→ $F = MS_{Treatment} / MS_{Error}$

Chapter Seven

- RM-ANOVA has an additional assumption of sphericity.
  - → It is assumed that the varience of the difference scores in the populations are equal
  - → Difficult assumption to meet bc RM-ANOVA are not robust to violations
  - → Test for sphericity: The Mauchly Test
  - → When sphericity cannot be assumed then df must be adjusted using:
    - → Huynh-Feldt Epsilon
    - → Greenhouse-Geisser Epsilon

- ANOVA only tells us whether the null is true it does not pinpoint the means that are significantly different from one another
  - → Multiple Comparison Procedures must be used to compare individual pairs of group means

- Multiple Comparison Procedures
  - → Tukey's HSD Test is most common amongst nurse researchers

- Magnitude of Effects in ANOVA
  - → The overall relationship btwn the IV and DV can be estimated through the effect-size index called eta-squared
  - → eta-squared (η^2) expresses the percentage of variability in the DV "explained" or accouted for by the IV
  - → eta-squared is often used to estimate sample size needs for ANOVA situations