SUMMARY OF STATISTICAL PROCEDURES

1. **DESCRIPTIVE STATISTICS** - what does the data look like?

   A. Graphical representation
   (frequency polygons, histograms, ogives, stem-and leaf charts)

   B. Numerical representation

      1) Measures of central tendency
      (mean, median, mode)

      2) Measures of variability
      (standard deviation, variance, quartiles, percentiles, outliers, boxplots)

2. **PROBABILITY** - what are the chances that a specific event will occur?

   A. Sample space size (combinations, permutations, factorials, multiplication rule)

   B. Events (simple & complex probability, conditional, complement, “at least one”)

   C. Distributions (normal, binomial, poisson)

3. **CONFIDENCE INTERVALS** - cushion around a sample statistic to estimate a population parameter

   A. Normal (one mean or difference of two means)

   B. Binomial (one proportion or difference of two proportions)

4. **SAMPLE SIZE** - how many should I sample given a level of confidence and an error bound ± value?

   A. Normal (to collect data for a mean - population standard deviation known)

   B. Binomial (to collect data for a proportion - population proportion known or unknown)

5. **HYPOTHESIS TESTING** - does the data collected reflect our hypothesis and the population value? includes H_0 and H_1, alpha level, Type I (α) level, critical value, test value, decision rule, rejecting or not rejecting H_0)

   A. Normal
      1. one mean compared to a standard
      2. two means compared to each other (independent samples with equal or unequal variances, dependent (matched pairs))

   B. Proportion
      1. one proportion compared to a standard
      2. two proportions compared to each other

6. **CORRELATION** - is there a positive or negative significant relationship between two variables?
    1. Pearson's correlation coefficient (r value)
    2. Hypothesis test for significant correlation

7. **LINEAR REGRESSION** - assuming that the correlation between the variables is significant, what is the equation for the line that best fits the data?

    1. $y' = a + bx$, where x is independent variable and y is dependent variable
    2. given an x value near the original data range, a predicted y' value can be determined

8. **CHI-SQUARE** - how well does observed counts of percentages of data compare to expected values?

    1. test for goodness of fit (data fits pattern)
    2. test of independence (row and column variables are independent/dependent)
    3. test of homogeneity of proportions (column proportions are equal or not)

9. **ANOVA (Analysis of Variance)** - comparing three or more means to see if they are equal or at least one is significantly different from the other

OTHER STATISTICAL TECHNIQUES

10. **MULTIPLE COMPARISON TESTS** - assuming ANOVA shows at least one mean not equal to others, which means actually are significantly different from the others?

11. **NON-PARAMETRIC TESTS** - if the data is the results of surveys, counts, or possibly unreliable, how can any significant trends or differences be detected?