Research Methodology Statistics Comprehensive Exam Study Guide

References

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology (3rd ed.)*. Boston: Allyn and Bacon.

Gravetter, F. J. (2006). *Statistics for the behavioral sciences : analysis and application for the public sector (7th ed.)*. Belmont, MO: Wadsworth.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis : a researcher's handbook*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Lomax, R. G. (2001). *Statistical concepts : a second course for education and the behavioral sciences (2nd ed.)*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Shavelson, R. J. (1996). *Statistical reasoning for the behavioral sciences (3rd ed.)*. Boston, Mass.: Allyn and Bacon.

Outline

- Basic definition
  - Measurement Scale
  - Population vs. sample
  - Discrete vs continuous variables
  - Qualitative vs quantitative variables
  - Sampling
  - Randomization
  - Sampling error
- Descriptive statistics
  - General
  - Central Tendency
    - Mean
    - Median
    - Mode
  - Variability
    - Range
    - Interquartile Range / Semi-Interquartile Range
    - Variance
    - Standard Deviation.
  - Distribution
    - Frequency Distribution
    - Percentile Ranks/Percentiles

  - Standard Scores
    - T-scores
    - z-scores
    - Linear Transformation
  - Graphing data
    - Bar Graph
    - Polygon
    - Histogram
    - Boxplot
    - Stem-and-leaf
    - Ogive

- Theoretical distribution
  - General
  - Normal Distribution
  - t
  - F
  - Chi-square
  - Binomial
  - Relationships among theoretical distributions

- Correlation
  - General
  - Covariance
  - Scatterplot
  - Point-biserial correlation
  - Pearson product-moment correlation
  - Spearman Rank correlation
  - Phi correlation

- Simple Linear Regression
  - Linear regression model
  - Ordinary least square
  - R-square
  - Standard error of estimate

- Probability
  - General
  - Independent events

  - Mutual exclusive
  - Sample space
  - Sampling with/without replacement

- Sampling distribution
  - General
  - Central Limit Theorem
  - Logic of sampling distribution
  - Standard error

- Hypothesis testing
  - Concepts
    - Logic of hypothesis testing
    - One-tailed two-tailed test
    - p-value
    - Type I and Type II error
    - Power
    - Sample Size
  - z-test
  - One-sample t-test
  - Paired-sample t-test
  - Independent-sample t-test
  - Test of Correlation Coefficient

- Estimation (confidence intervals)
  - Point estimate
  - Confidence Interval
  - Relationship between hypothesis testing and confidence interval

- Nonparametric (chi-square)
  - One-way Chi-square
  - Two-way Chi-square

- One-way between subjects ANOVA
  - Statistical model
  - Main effect
  - Assumption
  - Comparisons
    - Planned

        - Post-hoc adjustments
        - Orthogonality
    - Effect Size (eta-square & omega-square)
    - Power

- Two-way between subjects ANOVA
    - Statistical model
    - Main effect
    - Interaction effect
    - Simple effects
        - Simple main effect
        - Simple comparisons
        - Marginal comparisons
        - Interaction contrast
        - Planned
        - Post-hoc adjustments
    - Assumption
    - Effect Size
    - Power

- One-way within subjects ANOVA
    - Statistical model
    - Main effect
    - Assumption
        - Sphericity
        - Compound Symmetry
        - Greenhouse-Geiser
        - Huynh-Feldt
    - Comparison
        - Planned
        - Post-hoc adjustments
        - Orthogonality
    - Effect Size (eta-square & omega-square)
    - Power

- Two-way within subjects ANOVA
    - Statistical model
    - Main effect
    - Interaction effect

  - Simple effects
    - Simple main effect
    - Simple comparisons
    - Marginal comparisons
    - Interaction contrast
    - Planned
    - Post-hoc adjustments
  - Assumption
    - Sphericity
    - Compound Symmetry
    - Greenhouse-Geiser
    - Huynh-Feldt
  - Effect Size (eta-square & omega-square)
  - Power

- Two-way mixed ANOVA
  - Statistical model
  - Main effect
  - Interaction effect
  - Simple effects
    - Simple main effect
    - Simple comparisons
    - Marginal comparisons
    - Interaction contrast
    - Planned
    - Post-hoc adjustments
  - Assumption
    - Sphericity
    - Compound Symmetry
    - Greenhouse-Geiser
    - Huynh-Feldt
  - Effect Size
  - Power

- One-Way Between-Subjects ANCOVA
  - Statistical model
  - Main effect
  - Comparisons
    - Planned
    - Post-hoc adjustments
  - Adjusted means
  - Assumption
    - Homogeneity of regression

Terminology and Notation - ANOVA

- A and B will denote IVs
- Y will denote a DV
- SS - sum of squares
- MS - mean squares

Between-subjects ANOVA

subjects appear only in one group/cell

- One-way between-subjects ANOVA: there is only 1 IV (A)
  - model

$$Y_{ij} = \mu_{\cdot} + \alpha_j + \varepsilon_{ij} \tag{1}$$

    where
    - $\mu_{\cdot}$ - grand mean
    - Y_{ij} is the dependent variable for subject i in j^{th} group
    - $\alpha_j = \mu_j - \mu_{\cdot}$ where μ_j is the mean of a DV for group j
    - $\varepsilon_{ij} \sim N(0, \sigma^2)$
  - sum of squares

$$SS_{\text{total}} = SS_A + SS_{S/A} \tag{2}$$

    - SS_A - sum of squares of A (aka sum of squares between groups)
    - $SS_{S/A}$ - sum of square error (aka sum of squares within groups)
  - main effect of A
- two-way between-subjects ANOVA: there are 2 IVs (A and B)
  - model

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk} \tag{3}$$

    - grand mean: $\mu_{\cdot\cdot}$
    - main effect of A: $\alpha_j = \mu_{j\cdot} - \mu_{\cdot\cdot}$
    - main effect of B: $\beta_k = \mu_{\cdot k} - \mu_{\cdot\cdot}$
    - interaction effect $A \times B$: $\alpha\beta_{jk} = \mu_{jk} - \mu_{\cdot\cdot} - \alpha_j - \beta_k = \mu_{jk} - \mu_{j\cdot} - \mu_{\cdot k} + \mu_{\cdot\cdot}$
    - individual difference: $\varepsilon_{ijk} = y_{ijk} - \mu_{jk}$
  - sum of squares

$$SS_{\text{total}} = SS_A + SS_B + SS_{A\times B} + SS_{S/AB} \tag{4}$$

    - sum of squares between-groups breaks up into SS_A, SS_B, and $SS_{A\times B}$
    - $SS_{S/AB}$ - sum of squares error (aka sum of squares within groups)

- main effect of A, main effect of B, and interaction of AB
- simple effects
  - simple main effects
  - simple comparisons
  - marginal comparisons
  - interaction contrasts (tetrad differences)

Within-subjects ANOVA

aka repeated-measures ANOVA. Subjects appear in all the cells.

- one-way within-subjects ANOVA: there is only 1 IV
  - model

$$Y_{ij} = \mu_{..} + \alpha_j + \mathcal{S}_i + \alpha\mathcal{S}_{ij} + \varepsilon_{ij} \tag{5}$$

    - grand mean: $\mu_{..}$ - mean of the DV
    - iv effect: $\alpha_j = \mu_{.j} - \mu_{..}$
    - individual difference: $\mathcal{S}_i = \mu_{i.} - \mu_{..}$
    - individual by iv: $\alpha\mathcal{S}_{ij} = y_{ij} - \mu_{..} - \alpha_j - \mathcal{S}_i$. idiosyncratic response of the subject in a particular condition. differences in skill, ability, or predilection make some subjects perform better in one condition, others in another.
    - variability of the individual observation: ε_{ij}
  - sum of squares

$$SS_{\text{total}} = SS_A + SS_S + SS_{A\times S} \tag{6}$$

    - sum of squares between-subjects, SS_S
    - sum of squares within-subjects breaks up into SS_A and $SS_{A\times S}$
- two-way within-subjects ANOVA: there are 2 IVs
  - model

$$Y_{ijk} = \mu_{...} + \alpha_j + \beta_k + \alpha\beta_{jk} + \mathcal{S}_i + \alpha\mathcal{S}_{ij} + \beta\mathcal{S}_{ik} + \alpha\beta\mathcal{S}_{ijk} + \varepsilon_{ijk} \tag{7}$$

    - grand mean: $\mu_{...}$
    - main effect of A: $\alpha_j = \mu_{.j.} - \mu_{...}$
    - main effect of B: $\beta_k = \mu_{..k} - \mu_{...}$
    - interaction effect of AB: $\alpha\beta_{jk} = \mu_{.jk} - \alpha_j - \beta_k - \mu_{...}$
    - $\mathcal{S}_i = \mu_{i..} - \mu_{...}$
    - $\alpha\mathcal{S}_{ij} = \mu_{ij.} - \alpha_j - \mathcal{S}_i - \mu_{...}$
    - $\beta\mathcal{S}_{ik} = \mu_{i.k} - \beta_k - \mathcal{S}_i - \mu_{...}$
    - $\alpha\beta\mathcal{S}_{ijk} = y_{ijk} - \alpha\beta_{jk} - \alpha\mathcal{S}_{ij} - \beta\mathcal{S}_{ik} - \mu_{...} - \alpha_j - \beta_k - \mathcal{S}_i$
  - sum of squares

$$SS_{\text{total}} = SS_A + SS_B + SS_{A\times B} + SS_S + SS_{A\times S} + SS_{B\times S} + SS_{A\times B\times S} \tag{8}$$

    - sum of squares between-subjects, SS_S
    - sum of squares within-subjects breaks up into SS_A, SS_B, $SS_{A\times B}$, $SS_{A\times S}$, $SS_{B\times S}$, $S_{A\times B\times S}$

Two-way mixed ANOVA

Subjects appear in some of the cells (more than 1, less than all). There are 2 IVs (1 within-subjects, 1 between-subjects)

- model

$$Y_{ijk} = \mu_{...} + \alpha_j + \beta_k + \alpha\beta_{jk} + \mathcal{S}_{ij} + \beta\mathcal{S}_{ijk} + \varepsilon_{ijk} \tag{9}$$

  - grand mean: $\mu_{...}$
  - main effect of A: $\alpha_j = \mu_{.j.} - \mu_{...}$
  - main effect of B: $\beta_k = \mu_{..k} - \mu_{...}$
  - interaction effect of AB: $\alpha\beta_{jk} = \mu_{.jk} - \alpha_j - \beta_k - \mu_{...}$
  - between-subjects error: $\mathcal{S}_{ij} = \mu_{ij.} - \mu_{.j.}$
  - within-subjects error: $\beta\mathcal{S}_{ijk} = y_{ijk} - \alpha\beta_{jk} - \mathcal{S}_{ij} - \mu_{...} - \alpha_i - \beta_k$

- sum of squares

$$SS_{\text{total}} = SS_A + SS_B + SS_{A\times B} + SS_{S/A} + SS_{B\times S/A} \tag{10}$$

  - sum of squares between-groups breaks up into SS_A and $SS_{S/A}$
  - sum of squares within-groups breaks up into SS_B, $SS_{A\times B}$ and $SS_{B\times S/A}$