Midterm Review Sheet

Resources

You must bring your **I-Card** to the exam.

You will need to use **SAS installed on a computer in the classroom**. As stated in class, if you are accustomed to using SAS Studio, make sure you are familiar with the regular SAS interface and are able to generate your report using ods rtf or pdf statements. You will also need to use Word or a PDF editor (if available) on the classroom computer to complete your report.

For reference material, you can only use the SAS help, the chapter code and notes accessed directly from compass during the exam, and notes on one 8 ½'' by 11'' sheet of paper. You can have notes on front and back (or you can staple two sheets together if you only have notes on one side of each sheet). You can use the calculator app on the classroom computer for basic math if needed.

You will need to download the initial data from the course website and submit your exam through compass. Aside from that and the chapter code and notes, you will not be allowed to use any materials from compass or anywhere else on the web. Other reference materials or websites are not allowed, and the instructor is the only person you can ask questions.

You may not use other materials or technology. You may not use USB drives or other computing devices; your phone must be put away; and you must separate yourself from others you are used to talking to.

Material Covered

For all topics, results must be interpreted statistically and in terms of the real world measures and questions of interest.

Chapter 2

Descriptive Statistics and Visualization

- Locations (mean, median), spread (standard deviation, interquartile range, variance), skewness
  - Understand, interpret, and compare across groups
- Histograms, probability plots, box plots
  - Understand and interpret; draw qualitative conclusions about the distribution sampled from
- Correlation
  - Pearson correlation for roughly linear relationships and data without extreme values
  - Spearman correlation for data with extreme values or nonlinear relationships
  - Interpret and draw conclusions based on magnitudes of and hypothesis tests about correlation values

Tests for Normality

- Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises
    - Interpret the results and draw conclusions regarding normality of a population based on a sample

Location Tests

One-sided and two-sided alternatives

- One sample tests
    - T test when sample is not far from normal
    - Signed rank test for non-normal but reasonably symmetric distributions
    - Sign test otherwise
    - Choose correct test based on sample distribution and draw correct conclusion for mean or median of the population
- Two sample tests
    - T test when both samples are not far from normal
        - Test equal variance assumption and choose pooled or Satterthwaite t results based on variance conclusion
        - Correctly interpret results in terms of population means
    - Rank Sum test when one or both samples is far from normal
        - Correctly interpret results in terms of population values

Chapter 3

Contingency Tables and Tests of Association

- Contingency tables (2x2 and higher dimensions)
- Assess indications of association based on observed and expected frequencies
- Test for significant association
    - Pearson and Likelihood Ratio Chi-Square tests for large enough samples
    - Mantel-Haenszel Chi-Square for linear trend between ordinal categorical variables for large enough samples
    - Fisher's Exact Test for small samples (one and two sided alternatives for 2x2 table)

Risk and Risk Differences (risks are column risks in SAS)

- Risk is conditional probability
    - Estimated from row proportions (for column risks)
- Risk differences for 2x2 tables
    - For comparing risks of being in a column group for two different row groups
    - Confidence interval for the difference containing 0 indicates no significant difference of risks

Chapters 4 and 5

Balanced and Unbalanced ANOVA

Balanced Case

- Equal number of observations in each cell
- Use proc anova (also for one-way analyses)

- Test for unequal variance and adjust using Welch's adjustment (if necessary) in the one-way case
- Choose a best model (model selection) and explain selection process
- Interpret ANOVA tables and significance of model and terms in model
- Interpret R^2
- Determine and interpret significant differences between groups

Unbalanced Case

- Unequal number of observations in cells
- Use proc glm (when more than one predictor in the model)
- Choose a best model (model selection) and explain selection process
- Interpret ANOVA tables, type I and III sums of squares and significance of model and terms in model
- Interpret R^2
- Determine and interpret significant differences between groups

Chapters 6 and 7

Linear Regression

- Simple (single predictor) and multiple (multiple predictors) linear regression
- Use proc reg
- Choose a best model (model selection) and explain selection process
- Assess and remedy diagnostic issues (e.g. VIFs, diagnostic panel, residual plots, Cook's distances)
- Interpret ANOVA table and R^2
- Determine significance of parameters and interpret the model