MSIS 5503 – Statistics for Data Science

Fall 2021 – Exam 2 Time, Coverage and Format
25 points

On-campus students: will be in Room 070 on **Monday December 6**, **2021** from **8:00 to 9:50 p.m. (Monday Section)** and **Tuesday December 7**, **2021** from **8:00 to 9:50 p.m. (Tuesday Section)**

Online (Distance) students: You will have to take your exam between December 6, 8:00 a.m. and December 7, 8:00 p.m.

Bring your calculators. I will provide a handout.

The coverage will be Lecture 3D, Lecture 4 (A, B, C, D and E), and Lecture 5.

The exam will be in *two* sections:

Section 1 will have 20 – 30 **multiple choice questions** on theory. (Includes, but not limited to):

- Types of tests for differences in means and proportions for two populations
- Definition of correlation, partial correlation and auto-correlation
- Meaning of Least squares estimate.
- Calculation of slope and intercept for simple regression.
- Relationship between standardized model and the unstandardized model in regression.
- Understanding of tests of slopes, test of model fit, R-squared and adjusted R-squared.
- Assumptions in regression and their tests – linearity, heteroscedasticity, normality of error terms and independence of residuals.
- Consequence of violation of normality of error term.
- Consequence of violation of independence of errors.
- Multicollinearity definition, assessment, consequence and how to fix.
- Outlier analysis – terms and definitions
- Types of missing data handling
- Relationship between regression with a categorical predictor and two-sample t-tests.
- Regression with categorical and continuous data
- Understanding and interpreting interaction
- Autocorrelation and the Durbin-Watson Statistic
- Effects of Autocorrelation in a regression model
- Chi-square tests for independence, homogeneity and goodness of fit
- Understanding odds ratio and log-odds in logistic regression
- Types of variables vs types of analysis

Section 2 will have 5 to 10 **problems**. The problems will include:

- Two-sample hypothesis tests (Lecture 3D)
- Calculating missing entries in Regression outputs (Lectures 4A, 4B and 4D).
- Interpreting slopes in Regression.
- Calculating predicted Values in Regression outputs (Lectures 4A, 4B and 4D)
- Performing Significance Tests for Regression model and slopes (Lectures 4A, 4B and 4D)
- Calculating Contingency Table Entries and Performing Hypothesis Tests using Chi-square Tables (Lecture 5)
- Using Contingency Table to obtain joint, marginal and conditional probabilities.
- Calculating missing entries in Logistic Regression outputs (Lectures 5).
- Calculating Predicted Probabilities in Logistic Regression (Lecture 5)