MATH/STAT 301 outcomes:

1. Develop statistical computation skills
    - Use formulas to carry-out statistical analyses
    - Use statistical software to conduct analyses and explore data

2. Develop statistical reasoning
    - Explain statistical processes
    - Fully interpret results of analyses
    - Choose appropriate analysis methods
    - Identify and explain sources of variation
    - Explain and evaluate assumptions for analysis methods
    - Apply knowledge in novel situations

3. Develop statistical communication skills
    - Communicate statistical analyses and results using appropriate terminology and notation

4. Develop statistical thinking
    - Develop models to simulate and explore random phenomena
    - Recognize how, when, and why inferential tools can be used • Make appropriate conclusions from statistical analyses
    - Quantify variability
    - Explore data numerically and graphically
    - Describe common pitfalls & sources of error in statistical analyses
    - Evaluate fit and validity of statistical models

5. Develop statistical literacy
    - Use correct statistical notation
    - Make valid conclusions from visual displays of data
    - Accurately interpret output from statistical applications

MATH/STAT 301 objectives:

Statistical Computing

1. Use R and RStudio to load, manipulate, summarize, visualize, and analyze data
    a. Visualizations: histograms, density plots, stripplots, boxplots, plotDist, dotPlot, error bars, interaction plot, mosaic plot
    b. Packages: mosaic, ggplot2, dplyr, broom, knitr,
2. Explain the importance of reproducibility in research
3. Conduct simulations in R
4. Report results of analyses using Rmarkdown
5. Locate sources to help implement unknown functions in R

Lesson 1: Comparing means of two groups

6. Write out competing statistical models to compare two group means
7. Conduct a randomization-based test to compare 2 group means, generate a randomization distribution, & estimate a p-value
8. Explain how the randomization process works
9. Construct a bootstrap confidence interval for the difference between two group means
10. Explain how the bootstrap method works
11. Evaluate the assumptions/conditions necessary for an independent samples t-test
12. Derive the sampling distribution for the difference between two group means using Satterthwaite's approximation & SE_{pooled}.
13. Evaluate normality via Q-Q plots and Shapiro-Wilks tests
14. Conduct an independent samples t-test (with and without an equal variances assumption) to estimate a p-value
15. Construct a theory-based confidence interval for the difference between two group means
16. Calculate and interpret an effect size (Cohen's d) for the difference between two group means
17. Compare means between two dependent groups (matched pairs) using randomization- and theory-based methods
18. Apply randomization-based methods to compare two group medians

Lesson 2: Comparing variances of two groups

19. Derive a test statistic to compare the variances of two groups
20. Use randomization to compare 2 group variances, generate a randomization distribution, & estimate a p-value
21. Construct a bootstrap confidence interval for the difference between two group variances
22. Derive the sampling distribution for the difference between two group variances (F-distribution)
23. Conduct an F-test to test the difference between two group variances
24. Evaluate the assumptions/conditions necessary to conduct the F-test to compare two group variances

Lesson 3: Comparing means of 2+ groups

25. State appropriate null and alternative hypotheses to compare 3 or more group means
26. Explain why we should not conduct multiple t-tests to compare multiple pairs of group differences
27. Calculate the experiment-wise Type I error rate when comparing all possible pairs of group means in a dataset
28. Derive the SAD and MAD test statistics to measure the differences among 2+ group means
29. Conduct a randomization-based test (using MAD or SAD) to compare 2+ group means
30. Identify limitations with the randomization-based approach
31. Derive the mean square ratio as the test statistic to compare 2+ group means
32. Generate a statistical model to compare 2+ group means
33. Derive formulas for SSA, SSE, and SStotal
34. Determine the appropriate degrees of freedom for each SS value
35. Explain what a mean square represents
36. Using the definition of a mean square, derive simpler formulas for SSE and SStotal
37. Demonstrate how we can partition the total variation in a dataset into between- and within-groups factors
38. Explain the relationship between MSE and SE_{pooled} (from an independent samples t-test)
39. Sketch and label SStotal, SSA, and SSE for distributions under null and alternative hypotheses
40. Construct an ANOVA summary table and interpret all values
41. Calculate and interpret effect sizes for ANOVA (eta-squared, omega-squared)
42. Evaluate the assumptions/conditions necessary to conduct an ANOVA
43. Test for homogeneity of variances using Fmax, Levene's test, and Bartlett's test
44. Conduct an ANOVA in R with and without the homogeneity of variance assumption
45. Conduct a randomization-based test using mean square ratios to compare 2+ group means
46. Explain the relationship between the F-statistic in an ANOVA and a t-statistic in an independent samples t-test

Lesson 4: Post-hoc tests

47. Explain the concept behind the Bonferroni adjustment
48. Explain the concept behind Holm's method
49. Conduct post-hoc tests using the Bonferroni adjustment, Holm's method, Tukey's HSD
50. Evaluate the power and Type I error rates of Bonferroni and Holm's corrections
51. Write out an appropriate contrast to conduct a specified post-hoc test
52. Derive the test statistic (and associated standard error) to test a given contrast
53. Conduct Scheffe's test to compare a linear combination of means

Lesson 5: AxB ANOVA

54. Sketch means plots (interaction plots) to make conclusions about main effects and interaction effects
55. Calculate main effects and interaction effects in a given dataset
56. Explain what a significant *interaction effect* represents in a given scenario
57. Write out the statistical model for an AxB ANOVA
58. Derive formulas for SStotal, SSA, SSB, SSAxB; their associated degrees-of-freedom, and MS values
59. Construct an ANOVA summary table to summarize results from an AxB ANOVA
60. Conduct an AxB ANOVA, testing for significant interaction
61. Conduct a test of main effects, when appropriate
62. Conduct a test of simple effects when appropriate, splitting the data appropriately
63. Evaluate the assumptions/conditions necessary to conduct an AxB ANOVA
64. Use randomization-based methods to conduct an AxB ANOVA

Lesson 6: Repeated Measures & Experimental Design

65. Evaluate the advantages/disadvantages between a completely randomized design and a repeated measures design
66. Explain how total variation is partitioned in one-way, AxB, and repeated measures ANOVA designs
67. Explain how a repeated measures design can increase the statistical power of an analysis
68. Construct a statistical model for a repeated measures design

69. Construct a summary table to report results from an AxS ANOVA
70. Explain the process by which we test a groups-within-treatments design

Lesson 8: Analyzing Categorical Data

71. Explain the development of, and concept behind, Benford's Law
72. Generate expectations for cell frequencies of a discrete distribution
73. Derive the chi-square goodness-of-fit test statistic
74. Conduct a chi-square goodness-of-fit test, using randomization- and theory-based methods
75. Explain what it means when two outcomes are independent
76. Derive a formula to calculate expected frequencies in a contingency table, assuming independence
77. Conduct Fisher's Exact test to test the relationship between two categorical variables
78. Conduct a chi-square test for independence
79. Evaluate the assumptions/conditions necessary to conduct a chi-square test
80. Calculate an effect size from a contingency table, using relative risk or odds ratios
81. Explain the reasoning behind Yates's correction and explain why the use of Yates's correction is not recommended
82. Explain the concept behind likelihood ratio tests
83. Conduct a G-test to test for independence in a 2x2 contingency table

Lesson 9: Correlation

84. Explain what a covariance represents
85. Derive the formula for Pearson's product-moment correlation
86. Calculate Pearson's r for a given pair of variables
87. Demonstrate or explain the following characteristics of Pearson's r
    a. Scale invariance
    b. Only measures linear relationships
    c. Impact of outliers (predict the effect)
    d. Impact of range restriction
88. Provide or explain an example where correlation ≠ causation
89. Interpret correlations for a given pair of variables
90. Conduct a randomization-based test for a correlation
91. Construct a bootstrap confidence interval for a correlation
92. Conduct a t-test for Pearson's r
93. Evaluate the conditions necessary to conduct a t-test for Pearson's r
94. Explain how Spearman's rho is calculated
95. Explain how Kendall's tau is calculated
96. Independently assimilate and apply information about correlations for categorical data:
    a. Phi coefficient
    b. Cramer's V
    c. Biserial correlation
    d. Point-biserial correlation
    e. Polychoric correlation
97. Explain the concept behind distance correlations
98. Calculate Spearman's rho, Kendall's tau, and distance correlations for given variables
99. Evaluate the appropriateness of various correlation coefficients for a given set of variables

Lesson 10: Simple linear regression

100. Given a dataset, a scatterplot, and the least-squares regression line, interpret the slope and y-intercept.
101. Use a regression line to predict values of the dependent variable for given values of an independent variable
102. Calculate prediction error
103. Explain the least squares criterion
104. Explain the variance/bias trade-off (or model overfitting)
105. Use the least-squares criterion to derive formulas for the slope and y-intercept of a regression line
106. Calculate the slope and y-intercept of a regression line by hand and with a computer
107. Calculate and interpret SSE for a given dataset with least-squares regression line
108. Given a scatterplot and prediction line, sketch SStotal, SSregression, SSE, the standard error of estimate (RMSE)
109. Evaluate the fit of a regression line using SSE, Sy|x, and the coefficient of determination
110. Derive the theoretical maximums and minimums of SSE, Sy|x, and the coefficient of determination
111. Explain the advantages and disadvantages of using SSE, Sy|x, and the coefficient of determination to evaluate model fit
112. Evaluate the assumptions or conditions necessary to conduct a linear regression analysis (using tests and visualizations)

113. Interpret output from a regression analysis
114. Use randomization-based methods to test the slope of a regression line and estimate a p-value
115. Use bootstrap methods to construct a confidence interval for the slope of a regression line
116. Construct full and reduced (nested) models for a given research question and dataset
117. Construct an ANOVA summary table and derive all formulas to compare full and reduced regression models
118. Use theory-based methods and formulas to test the slope of a regression line
119. Explain why the test for the slope of a regression line is the same as the test for the correlation between X and Y
120. Demonstrate how the t-test for the slope of a regression line is directly related to F-test from the ANOVA
121. Use the omnibus F-statistic to compare full vs reduced regression models
122. Explain how the likelihood ratio can be used to compare nested regression models
123. Explain the criterion and penalty parameter of AIC and how it can be used to compare regression models
124. Construct, interpret, and contrast confidence and prediction intervals for a regression model

Lesson 11: Multiple regression

125. Interpret the parameters of a multiple regression model
126. Explain what a multiple correlation represents and how it is calculated
127. Use forwards and backwards selection to evaluate regression models
128. Explain how AIC and adjusted R-squared might be preferred to R-squared when comparing regression models
129. Explain why we should not compare the magnitudes of regression coefficients within a model
130. Interpret standardized beta coefficients
131. Explain the cause and effect of multicollinearity in a regression analysis
132. Use the VIF statistic to evaluate multicollinearity
133. Use the Durbin-Watson statistic to evaluate independence of errors
134. Interpret the coefficient of an interaction term in a regression model
135. Strategically select, test, and evaluate competing regression models
136. Explain the concept behind best subsets regression and how it can be used to select a regression model
137. Generate a regression model on a training dataset and evaluate it's fit to a test dataset
138. Explain how k-fold cross validation can be used to select a regression model
139. Explain the problem ridge regression and lasso are designed to address and how they address that problem
140. Demonstrate how to conduct a t-test under a regression framework to compare means after controlling for other variables

Lesson 12: Polynomial, robust, quantile regression; lowess; ANOVA as regression

141. Explain how a lowess curve is generated
142. Fit and evaluate polynomial regression models to a given dataset
143. Explain the differences between regression with robust standard errors and robust estimation of parameters and SEs
144. Use bootstrap methods (bootstrap residuals and bootstrap cases) to get robust estimates of model parameters
145. Explain what a quantile regression analysis attempts to investigate
146. Interpret the parameters of a quantile regression line for a given quantile
147. Interpret multiple quantile regression lines for a given dataset
148. Explain how ANOVA can be conducted under a regression framework using dummy (factor) variable coding

Lesson 13: Generalized Linear Model (logistic, Poisson regression)

149. Explain when a logistic regression model should be used
150. Explain what the Generalized Linear Model represents
151. Interpret output from a logistic regression analysis (converting from coefficients to log-odds, to odds, to probabilities)
152. Evaluate the fit of a logistic regression model to a given dataset
153. Compare competing logistic regression models using a Wald statistic
154. Interpret output from a Poisson regression analysis

Lesson 15: Other topics (if time permits) — machine learning algorithms, R packages, multidimensional scaling, PCA

155. Objectives will vary