Essential topics (nearly 100% chance to appear on the exam)

- Representing categorical predictors in regression using one-hot encoding to create dummy variables. How the model coefficients associated with these dummy variables are interpreted (ie: difference relative to the reference category)
- Basic workflow of a one-way ANOVA analysis: visualize data for preliminary assumption check -> global test (either ANOVA or Kruskal-Wallis) -> local test (only if global test is significant, either Tukey HSD or Dunn).
- Marginal vs. adjusted effects – how each are defined, when/why they might be similar or different, scenarios in which each would be preferred

High-priority topics (very likely to appear)

- Describing relationships in a scatter plot (form, direction, and strength)
- Pearson's vs. Spearman's correlation, including when to use each and how they relate to trends in a scatterplot.
- Theoretical (population-level) vs. fitted (estimated from sample data) models
- Hypothesis tests for single components of a regression model (ie: T-Test for H0: beta_j = 0)
- Models involved in one-way ANOVA (null model = single overall mean, alt model = group-specific means), how their fit is expressed via sums of squared residuals, and how the F-test compares fit.
- Residual diagnostic plots, including residual vs. fitted plots and QQ-plots
- Knowing how to sketch various models on a scatterplot, such as simple linear regression, quadratic regression, one-way ANOVA, multivariable regression with a one categorical and one quantitative predictor.

Medium priority topics (moderate chance to appear)

- Impact of outliers on regression models
- F-test for nested model comparisons (outside of the one-way ANOVA setting)
- The 2x standard deviation rule of thumb for the equal variability across groups assumption of ANOVA
- R^2, including how it is defined, what it means, how it relates to adjusted R^2
- Relationship between Pearson's correlation, R^2, and the T-test of H0: beta_1 = 0 for simple linear regression
- Components in an ANOVA table and how they relate to each other and the F-statistic
- Visually assessing within group vs. between group variability from side-by-side boxplots in the context of a one-way ANOVA analysis.

Low-priority topics (one or two topics from this list might appear)

- Log-transforming outcomes in a regression model, including why you'd do it and how to interpret the resulting model
- Quadratic regression, specifically how things like R^2 and residual diagnostics might improve under this model when a predictor has a non-linear relationship with the outcome.
- Ecological fallacy as an issue that sometimes arises during correlation analysis
- Extrapolation as an issue that sometimes arises during regression analysis
- The F-distribution used in the ANOVA F-test – knowing it is strictly positive and right-skewed, and that the p-value is calculated as the area to the right of the observed F-statistic
- The concept of multicollinearity and the problems it causes in multivariable regression