

## **Material Assessment Review – Summer 2024**

Identify type of variable (Quantitative, Ordinal or Categorical)

Dummy coding for Categorical (Reference Coding)

Difference between treating Ordinal Variable as Quantitative versus Categorical

Different measures of location (impact of skewness or outliers on these measures)

Different measures of spread (impact of skewness or outliers on these measures)

Describe shape of distributions (modality, skewness, center, spread, kurtosis)

Characteristics of a Normal distribution

Empirical Rule

When graphs should be used for distributions/relationships (for example, histograms illustrate the distribution of a quantitative variable; barplots illustrate distribution of a categorical variable; boxplots can illustrate relationships between a categorical and quantitative variable)

QQ plots to assess normality (identify if there is skewness or kurtosis in a QQ plot).

What values are used to create a boxplot

How to identify potential outliers

Statistics versus Parameters and Sample versus Population

Central Limit Theorem

When do sample means follow a normal distribution?

Standard error versus Standard deviation

Sampling distribution of sample mean

Confidence interval and terminology (confidence, MOE, standard error)

Interpret Confidence Intervals

Hypothesis testing

Terminology and definitions (Null Hypothesis, Alternative Hypothesis, Type I errors, Type II errors, significance level, test statistic, p-value, how to state conclusions)

Identify null and alternative hypotheses for different situations

When to reject the null hypothesis

Perform one sample test (state hypotheses, know assumptions, make conclusions, read output from R, test statistic, calculate p-value)

Relationship between a two-tailed hypothesis test and confidence intervals

Perform two sample test (state hypotheses, know assumptions, make conclusions, read output from R, test statistic, calculate p-value, when to use unequal variance t-test versus pooled t-test....how to test for this)

Perform Wilcoxon Rank sum (what it is and what is it testing, read output and make conclusion)

Explanatory versus Predictive Modeling

ANOVA (what it is, what it is testing, assumptions, hypotheses, make conclusions, number of predictions, setup...what type of response variable and what type of explanatory variable)

Welch's ANOVA (when to use it)

Kruskal-Wallis (when to use it)

Tukey-Kramer and Dunnett's post hoc test and Games-Howell or Dunn's test (why do we do them, what are they used for...also understand experiment-wise error rate and why important in multiple hypothesis testing)

What is Honest Assessment?

What the training data, validation data and test data is used for (and when it should be used)

When to create them

EDA (univariate and bivariate) and graphs to do this

Pearson's correlation (properties and hypothesis test...how does this test relate to test of slope in SLR)

Correlation versus Causation versus slope in SLR

SLR (write out equation, make predictions, calculate residuals, interpret slope and intercept, interpret confidence intervals, do hypothesis test and interpret, assumptions)

Two-Way ANOVA (and n-way ANOVA)

Interactions and Sliced ANOVA

Differences in MLR and SLR

What does LINEAR mean in linear regression

Global F-test

F-distribution

Assumptions for MLR (adding no perfect collinearity)

Adjusted  $R^2$

Categorical Predictor Variables

Interpretation

Best Linear Unbiased Estimator

What is an unbiased estimator

What is meant by best

Regularized regression

Ridge, LASSO, Elastic Net

Compare and contrast

Cross-validation

AIC, BIC (what they are, penalties ....and how affected by number of variables)

Forward selection, backward elimination and stepwise selection (what they are and how they work, issues with these procedures)

p-values and sample size

How to assess assumptions in MLR (Misspecified model, lack of constant variance, lack of normality, correlated error terms)

How to identify potential outliers and influential points (what measures are used for each)

Multicollinearity (what it is and how to identify it)

Nominal vs. Ordinal categorical variables

Association of categorical variables

Pearson's Chi-square

Likelihood Ratio Chi-square

Mantel-Haenzsel Chi-square

Chi-square distribution

Assumptions to tests of association

Measures of association

Odds ratios, odds vs probability

Cramer's V

Spearman's correlation

Logistic regression for any type of categorical target variable

Why is least squares bad?

Logit link function

Interpretation on coefficients

Concordant, discordant, tied pairs

Variable selection