

## **Chi-squared Testing**

High priority:

- When it is used (nominal categorical data)
- Differences between Goodness of Fit and Test of Independence, including when to use each test
- Null hypotheses and p-value interpretation

Medium priority:

- Test statistic (sum of observed – expected squared divided by expected)
- Standardized residuals
- Sample size assumption (expected counts of at least 5) and Fisher's exact test

Low priority:

- Shape of the Chi-squared distribution and degrees of freedom
- Relationship between Chi-squared distribution and Normal curve
- Cramer's V as a measure of strength of association

## **One-way ANOVA**

High priority:

- When it is used (categorical explanatory variable and quantitative response)
- Null and alternative models (single overall mean vs. group-specific means) and how they relate to hypotheses
- p-value interpretation
- Workflow of an ANOVA analysis – compare models using F-test, check assumptions (Normal residuals and approx. equal standard deviations across groups), post-hoc testing (pairwise tests to identify which pairs of groups are different)

Medium priority:

- Measuring model fit via sum of squared residuals
- How the F-test statistic relates to sum of squares

Low priority:

- Kruskal Wallis test and Dunn test as non-parametric alternatives when assumptions are violated and log-transformation doesn't work

## **Linear Regression**

High priority:

- When it is used (one or more explanatory variables and a quantitative response variable)
- Workflow of a regression analysis - comparing nested models using F-tests, checking assumptions (residuals are independent, have constant variance, and follow a Normal distribution), inference on coefficients (t-tests and confidence intervals)
- The null hypothesis and p-value interpretation for an F-test of nested models
- The null hypothesis and p-value interpretation for a t-test on a model coefficient

Medium priority:

- Interpreting coefficients when the outcome variable has been log-transformed
- The visual relationships between certain models and their fit as measured by sum of squared residuals.

Low priority:

- Knowing that the F-test can be used to assess polynomial effects.

## **Logistic Regression**

High priority:

- When it is used (one or more explanatory variables and a binary categorical response variable)
- Visual differences from linear regression and a basic visual understanding of the model (S-shaped logistic curve determined by slope coefficients)
- Interpretation of coefficients (intercept, coefficients on continuous predictors, coefficients on categorical predictors encoded via dummy variables) and the need to apply inverse transformations

Medium priority:

- Likelihood ratio testing to compare nested models – know the null hypothesis and p-value interpretation, as well as when this test is used
- Confidence intervals and hypothesis tests for model coefficients

Low priority:

- Understanding the likelihood function and how it reflects the fit of a model (do not need to know the form of the function)