Learning Objectives

Exam 03: Inference with several variables

- LO 1. Use R to compare the difference in mean across two groups using either a parametric or non-parametric test.
 - two sample t-test: a parametric test that relies on the central limit theorem. Provides a confidence interval for the differences in the means.
 - Mann-Whitney U: a non-parametric test for differences in the mean. Does not include an effect size or confidence interval.
- LO 2. Use R to compare the difference in mean across more than two groups.
 - one-way ANOVA: a parametric test that relies on the central limit theory.
 - Kruskal-Wallis test: a non-parametric test to detect whether samples from multiple groups come from the same distribution.
- LO 3. Understand the problem behind multiple hypothesis testing.
- LO 4. Define the **family-wise error rate** (FWER) and the **false discovery rate** (FDR). Be able to control these with the Holm-Bonferroni method or the Benjamini-Hochberg-Yekutieli (BHY) method.
- LO 5. Use R to find pairwise differences in means or proportions between groups, correcting for multiple hypothesis testing.
 - Pairwise T-test
 - Pairwise Mann-Whitney U
 - Pairwise Proportions Test
- LO 6. Understand the problem of confounding variables and the resulting bias caused within the model.
- LO 7. Apply linear regression in R to describe a relationship with a numeric dependent variable, potentially with one or more confounding variables.
- LO 8. Understand the basic assumptions behind linear regression.
 - The 'true' mean of the dependent variable is a linear function of the independent variables.
 - The observations are drawn independently from one another.
 - The variance of the dependent variable is unrelated to the observed values of the independent variables.
 - The central limit theorem is used to make assumptions about the testing statistics.
- LO 9. Apply logistic regression in R to describe a relationship with a two-category dependent variable, potentially with one or more confounding variables.
- LO 10. Produce a data dictionary for a dataset, describing the variable names, data types, and collection methods.
- LO 11. Map aesthetics within the grammar of graphics to influence the size, color, shape, and opacity of the data points.