BMES 441/641 Survival Guide v0.1

Christian D'Andrea

September 2024

1 Introduction

Reports show me that you know how to think logically about the experiment performed. That you can envision an experiment and think of relevant, specific, and testable hypotheses. That you can provide background about why this work is relevant or useful, what is to be learned, and what the objective is to complete this learning. That you can describe your methods such that the process could be reproduced. That your results transparently describe the outcomes without interpretation (yet). That the discussion provides an evidence-backed interpretation of your results with respect to your hypotheses. If you can prove to me that you succeed at this within the required formatting guidelines, you will likely achieve a grade with which you'll be content!

2 Presentation and Report Submission Tips

1. Submit presentations and reports as a PDF directly to BBLearn. Do NOT submit the PDF in a zip file.
2. Code will be submitted either in a zip file or to GitHub (TBD)

3 Statistics

Here is a choose-your-own-adventure guide to performing statistics on your data. This should cover anything you run into during these courses. I'm not a statistician, and am learning this art along with you, so if you disagree with any recommendation below, please email me and we can figure it out.

1. Pre-tests
   (a) Normality Test
      i. For small samples ($n < 50$): Shapiro-Wilk test
      ii. For large samples ($n \geq 50$): Kolmogorov-Smirnov test
      iii. Visual inspection (Q-Q plots, histograms)
   (b) Homogeneity of Variance (for parametric tests)
      i. Levene's test
      ii. Bartlett's test (if data is normal)
   (c) Outlier Detection
      i. Visual methods (box plots, scatter plots)
      ii. Statistical methods (Z-scores, Interquartile Range)
2. Test Selection Guide
   (a) Two Groups
      i. Small Sample Size ($n < 30$ per group)
         A. Normal Distribution & Equal Variances
            - Independent: Independent samples t-test
            - Paired: Paired samples t-test

B. Non-Normal Distribution OR Unequal Variances
- Independent: Mann-Whitney U test
- Paired: Wilcoxon signed-rank test

ii. Large Sample Size ($n \geq 30$ per group)

A. Normal Distribution
- Independent: Independent samples t-test
- Paired: Paired samples t-test

B. Non-Normal Distribution
- Independent: Mann-Whitney U test (or t-test if $n > 50$)
- Paired: Wilcoxon signed-rank test (or paired t-test if $n > 50$)

(b) More than Two Groups

i. Small Sample Size ($n < 30$ per group)

A. Normal Distribution & Equal Variances
- Independent: One-way ANOVA
- Paired/Repeated Measures: Repeated measures ANOVA

B. Non-Normal Distribution OR Unequal Variances
- Independent: Kruskal-Wallis H test
- Paired/Repeated Measures: Friedman test

ii. Large Sample Size ($n \geq 30$ per group)

A. Normal AND Non-Normal Distributions
- Independent: One-way ANOVA
- Paired/Repeated Measures: Repeated measures ANOVA

3. Post-hoc Tests

(a) For ANOVA (parametric)
- Tukey's HSD (Honestly Significant Difference): For all pairwise comparisons when sample sizes are equal and variances are homogeneous
  - Default in MATLAB's multcompare() function
  - Balances Type I error mitigation and statistical power
  - Suitable for most situations with equal sample sizes
- Bonferroni: Conservative approach for mitigating Type I error, suitable for a small number of comparisons, or else it may get *too* conservative. Consider using this when the real-life consequences of a Type I error are high.
- Dunnett's test: When comparing multiple groups to a single control group
- This is not an exhaustive list, but likely covers most cases. Feel free to use another if justified.

(b) For Kruskal-Wallis H test (non-parametric)
- Dunn's test: For pairwise comparisons after a significant Kruskal-Wallis result
- Mann-Whitney U tests with Bonferroni correction: Alternative to Dunn's test, more conservative

(c) For Friedman test (non-parametric repeated measures)
- Nemenyi test: For all pairwise comparisons
- Wilcoxon signed-rank tests with Bonferroni correction: Alternative to Nemenyi test, more conservative

4. Other Cases

(a) Categorical Data
- Two categories: Chi-square test of independence
- More than two categories: Chi-square test or Fisher's exact test (for small expected frequencies)

(b) Correlation

- Normal Distribution: Pearson correlation
- Non-Normal Distribution: Spearman rank correlation

(c) Regression

- Linear relationship & Normal residuals: Linear regression
- Non-linear relationship or Non-normal residuals: Non-linear regression or Generalized Linear Models / Multiple-linear regression

4 Useful Tools

This class involves a solid amount of coding work, no doubt. The important part is not that you know how to change a plot's line width in MATLAB, it's that you, with your brain, can architect an flow which allows you to perform a valid analysis of the data and intuitively visualize it, and communicate this effectively to me/your readers, in a way that proves you have designed and understand your approach. This approach still requires knowledge of how to write and generate well-structured code, but it's wasted time to get stuck on small details which don't enable learning biomechanics. So, here are some helpful tools to this end, which I use in my research:

1. Miro, a flowchart/diagramming tool
2. Cursor, an AI-enabled code editor. You can install a MATLAB extension for editing the code in Cursor, and can run this code through the usual MATLAB IDE.
3. GitHub, a Software Versioning and Collaboration tool
4. Perplexity, a new search engine - great for gathering initial evidence, leading more quickly to scholarly sources
5. NotebookLM, an interface to upload resarch and chat with it, as well as generate podcast-like audio summaries of the material. Great for gaining familiarity with a paper without having to read it.