Quantitative Understanding in Biology I
Q1-2 2018
Course Syllabus

Thursday, September 6th, 2018 – Thursday, December 13th, 2018
Tuesdays and Thursdays, 5:30 PM – 7:00 PM
Location: Weill Auditorium (or as indicated; n.b.: first day meets in A-950)
1300 York Avenue; 2nd Floor, Room C-200 (unless otherwise noted)
Midterm Exam: Thursday, November 1st, 2018 (5:30 pm – 7:00 pm)
Course Directors:
• Jason Banfelder (jbanfelder@rockefeller.edu)
• Luce Skrabanek (las2017@med.cornell.edu)
This course will be fully graded: (Honors, High Pass, Low Pass, Fail)
Course web site: [http://physiology.med.cornell.edu/people/banfelder/qbio](http://physiology.med.cornell.edu/people/banfelder/qbio)
Teaching assistants:
• Shana Bergman (shb2037@med.cornell.edu)
• Nebojsa Jukić (nej2011@med.cornell.edu)
• Scott Klum (sdk2004@med.cornell.edu)

Summary

This course will prepare students to apply quantitative techniques to the analysis of experimental data. To emphasize both practical and theoretical skills, the course will involve several hands-on workshops, and the completion of several projects will be required. Students will be well positioned to meet the emerging requirements of funding agencies for formally planned experiments and fully reproducible and documented data analysis methods.
Specific topics include: practical aspects of data formatting and management; graphical, mathematical and verbal communication of quantitative concepts; a review of statistics, with emphasis on the selection of appropriate statistical tests, the use of modern software packages, the interpretation of results, and the design of experiments; the formulation, evaluation, and analysis of mathematical models of biological function, with an emphasis on linear and non-linear regression, determination of model parameters, and the critical comparison of alternative models with regard to over-parameterization.

Class Sessions

1. Quantifying a Sample Distribution – part I
   Thursday, September 6th, 2018
   (Note: class meets in A-950)
   summary statistics, quantiles, SD vs. SEM

2. Introduction to Swirl; Quantifying a Sample Distribution – part II
   Tuesday, September 11th, 2018
   summary statistics, measurement variation vs. biological variation

3. Probability Density Functions and the Normal Distribution – part I
   Thursday, September 13th, 2018
   binomial, Poisson, and normal distributions, testing for normality (part I): an introduction to formal statistical tests

4. Probability Density Functions and the Normal Distribution – part II
   Tuesday, September 18th, 2018
   testing for normality (part II): qqplots vs. formal tests

5. Lab #1: Practical R
   Thursday, September 20th, 2018
   introduction to R, common data structures (vectors, factors and matrices); strategizing Jeop-R-dy

6. Jeop-R-dy heats
   Monday-Friday, Sept 24-28, 2018
   (Note: times and locations TBA)

7. Confidence Intervals and Contingency Tables
   Tuesday, October 2nd, 2018
t-tests, working with proportional data, why CIs are more informative than p-values, study types (retrospective, prospective, and cross sectional), working with rare events

8. Final Jeop-R-dy  
Thursday, October 4th, 2018

9. p-Values and Formal Statistical Testing  
Tuesday, October 9th, 2018  
duality between p-values and CIs, statistical vs. biological significance, choosing an appropriate α, type I and type II errors

10. Lab #2: Practical R  
Thursday, October 11th, 2018  
common data structures continued (lists and dataframes); understanding Type I error rates

11. Statistical Power and Experimental Design  
Tuesday, October 16th, 2018  
why you cannot just add a few N to your dataset when \( p > 0.05 \), appreciating the economics (in time and money) of experimental design, statistical vs. biological significance revisited, designing experiments around hard-to-obtain samples

12. Multiple Hypothesis Testing and Non-parametric Tests  
Thursday, October 18th, 2018  
from Bonferroni to False Discovery Rate, dealing with non-normal data, t-tests revisited

13. Midterm Exam Review (led by TAs)  
Tuesday, October 23rd, 2018

14. Bayesian Methods  
Thursday, October 25th, 2018  
how to incorporate prior knowledge into statistical models

15. Midterm Exam  
Thursday, November 1st, 2018

16. Lab #3: Practical R  
Tuesday, November 6th, 2018  
the role of simulation in statistics, non-parametric tests, exploring the optimal stoping problem
17. Correlation vs. Linear Regression
   Thursday, November 8th, 2018
   introduction to modeling in R; why $r^2$ is not the whole story

18. Code Quality
   Tuesday, November 13th, 2018

19. Fitting Model Parameters to Data
   Thursday, November 15th, 2018
   non-linear regression, a statistical view of curve fitting, confidence intervals revisited

20. Thanksgiving Break: NO CLASS
   Week of November 20th, 2018
   Tuesday, November 27th, 2018

21. Quantitative Comparison of Models and ANOVA – part I
   Thursday, November 29th, 2018
   how to avoid over-fitting, F-test and AICs, F-tests as a means of parameter estimation

22. Quantitative Comparison of Models and ANOVA – part II
   Tuesday, December 4th, 2018

23. Principal Component Analysis
   Thursday, December 6th, 2018
   (Note: class meets in A-250)
   data transformation and data reduction methods

24. Introduction to ggplot2
   Tuesday, December 11th, 2018
   (Note: class meets in A-250)
   publication-quality graphs and plots with ggplot

25. Spare day in schedule
   Thursday, December 13th, 2018

Books and Materials

Students will need a laptop computer on which they can install software (R\(^{[1]}\) and R Studio\(^{[2]}\)), and bring to class. Both packages are free, and run on recent versions of Linux, Mac OS

\(^{[1]}\)https://www.r-project.org/
\(^{[2]}\)https://www.rstudio.com/
While the course does not require the use of a specific textbook, the following resources are recommended.

- **Intuitive Biostatistics**, Harvey Motulsky
  One of the most accessible introductions to statistics.

- **The Art of R Programming**, Norman Matloff
  One of the more comprehensive introductions to R.

- **R for Everyone**, Jared Lander
  Less in-depth than the above, but covers both basic use of R and basic statistics in a single, accessible text.

- **Practical Computing for Biologists**, Haddock and Dunn
  Covers many computing topics not covered in this class. Recommended for students considering a computational lab for a rotation or thesis.

- **R for Data Science**, Grolemund and Wickham
  Introduction to the tidyverse, focusing on importing, wrangling, exploring, and modeling your data and communicating the results. Some experience with R advised. Also available online at [http://r4ds.had.co.nz/](http://r4ds.had.co.nz/).

**Assessment**

This class will be graded according to the usual WCGS scale (Honors, High Pass, Low Pass, Fail).

Grades will be determined based on several take-home problem sets, and a midterm exam.

All students will also be asked to complete a survey at the end of the semester, soliciting feedback on the course to inform its content and format in future years.