Quantitative Understanding in Biology I
Q1-2 2016
Course Syllabus

Tuesday, September 13th, 2016 – Thursday, December 15th, 2016
Tuesdays and Thursdays, 5:30 PM – 7:00 PM
Location: Weill Auditorium
1300 York Avenue; 2nd Floor, Room C-200 (unless otherwise noted)
Midterm Exam: Thursday, November 3rd (5:30 pm – 7:00 pm) in Room A-250
Final Exam: Thursday, December 15th (5:30 pm – 7:00 pm) in Room C-200
Course Directors:

- Jason Banfelder (✉ jrb2004@med.cornell.edu)
- Luce Skrabanek (✉ las2017@med.cornell.edu)
- Michael LeVine (✉ mil2037@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`

Teaching assistants:

- Nimra Asi (✉ nia2015@med.cornell.edu)
- Gabriele Campanella (✉ gac2010@med.cornell.edu)
- Kevin Hadi (✉ keh2019@med.cornell.edu)
- Mehtap Isik (✉ mei2007@med.cornell.edu)

additional TAs TBD; five to six total depending on enrollment

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
   Tuesday, September 13th, 2016

   summary statistics, quantiles, SD vs. SEM, measurement variation vs. biological variation

2. Probability Density Functions and the Normal Distribution
   Tuesday, September 20th, 2016

   binomial, Poisson, and normal distributions, testing for normality (part I): an introduction to formal statistical tests, testing for normality (part II): qqplots vs. formal tests

3. Lab #1:Practical R (part I)
   Thursday, September 22nd, 2016

   introduction to R, common data structures (vectors, factors, lists, dataframes), importing data

4. Confidence Intervals and Contingency Tables
   Tuesday, September 27th, 2016

   t-tests, working with proportional data, why CIs are more informative that p-values, study types (retrospective, prospective, and cross sectional), working with rare events

5. p-Values and Formal Statistical Testing
   Tuesday, October 4th, 2016

   duality between p-values and CIs, statistical vs. biological significance, choosing an appropriate α, type I and type II errors

6. Lab #2: Practical R (part II)
   Thursday, October 6th, 2016

   libraries and `ggplot`, producing publication quality plots, data driven graphics

7. Statistical Power and Experimental Design
   Tuesday, October 11th, 2016

   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

8. Multiple Hypothesis Testing and ANOVA
   Tuesday, October 18th, 2016

   from Bonferroni to False Discovery Rate, dealing with non-normal data, t-tests revisited, fitting with categorical data, dealing with hierarchical and time-series data

9. Lab #3: Practical R (part III)
   Thursday, October 20th, 2016

   the role of simulation in statistics, non-parametric tests, exploring the optimal stoping problem

10. Bayesian Methods
    Tuesday, October 25th, 2016

    how to incorporate prior knowledge into statistical models

11. Midterm Exam
    Thursday, November 3rd, 2016
    The midterm exam will be held in Room A-250

12. Correlation vs. Linear Regression
    Tuesday, November 8th, 2016

    introduction to modeling in R; why r^2 is not the whole story

13. Lab #4: Correlation and Regression Lab
    Thursday, November 10th, 2016

14. Principal Component Analysis
    Tuesday, November 15th, 2016

data transformation and data reduction methods

15. Thanksgiving Break: No class
    Week of November 21st, 2016

16. Fitting Model Parameters to Data
    Tuesday, November 29th, 2016

    non-linear regression, a statistical view of curve fitting, confidence intervals revisited

17. Lab #5: Practical R (part IV)
    Thursday, December 1st, 2016

    data wrangling with `melt`, `dcast`, and `plyr`

18. Quantitative Comparison of Models
    Tuesday, December 6th, 2016

    how to avoid over-fitting, F-test and AICs, F-tests as a means of parameter estimation

19. Lab #6: Model Comparison Lab
    Thursday, December 8th, 2016
    Lab #6 will be held in Room A-250

20. Final Exam
    Thursday, December 15th, 2016

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 X, and Microsoft Windows.

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.

[1] `https://www.r-project.org/`
[2] `https://www.rstudio.com/`

- *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.

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 and final 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.