The 'Practical Statistics for Experimentalists' Workshop

Hugo Bowne-Anderson, Yale University, Molecular Biophysics & Biochemistry Department

Instructor: Hugo Bowne-Anderson (hugo dot bowne-anderson at yale dot edu). *Teaching Assistant*: Neal Ravindra, Olivier Trottier (alphabetical order). *Host*: Jonathon Howard. *Location*: Science Hill (specifics TBA).

Date & Time: January 30th & 31st, 2016, 0900-1700. Breakfast, coffee & lunch provided.

1 Outline

The Workshop will be a mixture of statistics and data analysis and the focus will be on solving practical problems, although to do this we will need to know a bit of theory! Each day will consist of 2 modules (see below) and will be split roughly into half a day of learning statistical & data analytic techniques & half of hands-on practical statistical coding in the programming language R. Coding work will be interspersed through the Workshop & the final 2 hours of each day will be dedicated to Workshop participants getting hands-on, assisted practice in performing data analysis on real-world datasets in R. No background in R is required. Each module, however, will build on those preceding it.

The general outline is as follows:

Day 1:

1. Plotting data, summary statistics and basic statistical rules of thumb; questions considered will include "which summary statistics should I report in which contexts (e.g. when do I report SD and when SEM)?" & "how many data points are required to be confident in my estimate of the mean and/or the standard deviation?";
2. Statistical hypothesis testing (e.g. normality tests, t-tests, F-tests, ANOVA); questions considered will include "how many data points will I need to to find a statistically significant result?" & "what type of experimental set-up is required in order to retrieve the signal-to-noise ratio necessary to validate anticipated results?"

Day 2:

3. Model fitting and model selection: linear regression, maximum likelihood estimation and nonlinear least squares curve fitting; questions considered will include "when is a particular model, such as linear regression, appropriate & when is it not?" & "How do I compare the performance of two models that I have fit to the same data?"
4. Time series analysis: exploratory time series data analysis, filtering techniques & smoothing data, error reduction & noise analysis, regression in time series, correlation of time series (including autocorrelation), Fourier analysis, Autoregressive Moving Average (ARMA) models; questions considered will include "how can I de-noisify my signal to more readily see trends?" and "how do I analyze and report the concurrent phenomena of noise, trends & periodic behavior in time series data?"