Statistics and Data Analysis in Neuroscience
(INP/NBIO 599)

Course Director: Daeyeol Lee
Lecturers: Steve Chang, Damon Clark, Alex Kwan, Ifat Levy, John Murray
Time:

1. **Lectures - 3-5:30 pm Wednesday (3-5:30 pm on Monday and Thursday for Week 1)**
2. **"Hackathon" - (optional) 5-7 pm Monday, JEH H216 (from Week 2)**

Location: SHM B145 (exceptions: 1/26, 2/1, and 3/8, see below).

Description: This course focuses on practical applications of various statistical models and tests commonly used in neuroscience research. It covers basic probability theory, hypothesis testing, and maximum likelihood estimation, as well as model comparison. The specific models and tests covered include ANOVA, regression, time series analyses, and dimension reduction techniques (e.g., PCA). Examples and homework will be given in MATLAB, which will be introduced at the beginning of the course. Previous experience in programming and basic statistics is desirable but not required.

Requirements & Grading

1. Class attendance (30%)
2. Homework (50%)
3. Final report (20%)

Week 1 (Monday 1/23): Matlab-I (Lee)

1. Matrix operations
2. Import/exporting data
3. Plotting: plot, image, hist, bar

Week 1+ (Thursday 1/26): Matlab-II (Lee): SHM L101A (4-6 pm)

1. Control statements: if, else, break, for while
2. m files and function: mean, sum, sort, max/min, etc.

Week 2 (2/1): Probability (Lee): SHM C428

1. Axioms of probability theory
2. Joint/conditional probability, independence
3. Probability distribution/cumulative distributions
4. Expected values - linear operator
5. Multivariate probability distribution / marginal distribution
6. Univartiate/multivariate Gaussian distribution

Week 3 (2/8): Probability - II (Murray)

1. Bernoulli processes
2. Poisson processes
3. Spike train analysis
4. Markov processes

Week 4 (2/15): Hypothesis testing (Chang)

1. Generating testable hypothesis (concept)

2. False negatives and false positives (concept)
2. Generating / simulating data (with practical component)
3. Performing permutation and bootstrap type tests using the simulated data
4. Plotting your results following permutation-type tests (determining noise level)

Week 5 (2/22): Maximum likelihood estimation (Lee)

1. Probability distribution vs. likelihood
2. Example 1 - binomial distribution
3. Example 2 - Gaussian dsitribution
4. Example 3 - Reinforcement learning (fminsearch)

Week 6 (3/1): Regression / linear regression (Kwan)

1. Two-factor ANOVA, main and interaction effects
2. Correlation
3. Simple linear regression
4. Multiple linear regression

Week 7 (3/8): GLM (Levy): SHM C428

1. Regressions of binary and count data
2. Logistic regression
3. Probit regression
4. Poisson regression

Week 8 (3/29): Model comparisons (Chang)

1. Reasons for model-fitting your data and selecting an appropriate model (concept)
2. Comparing models with different number of parameters (concept)
2. Dealing with multiple comparisons (with practical component)
3. Fitting your data to different models and evaluating individual model fits
4. Comparing the model fits to find the best model (AIC, BIC, cross-validation)

Week 9 (4/5): Time series / spectral analysis (Kwan)

1. Sum of cosines, Fourier series
2. Discrete Fourier transform
3. Power spectrum
4. Sampling frequency, aliasing, filters
5. Cross-correlation, coherence

Week 10 (4/12): Dimensionality Reduction and PCA (Clark)

1. Two examples: Family of extracted filters / NIST handwritten digits
2. Dimensions and dimensionality in data
3. Variance and signal
4. Graphically what PCA does
5. Covariance matrix and eigenvectors
6. Projecting data points or dimensions
7. Failures
8. Linear discriminant analysis, ICA, t-SNE, and others, with caveats

Week 11 (4/19): classification & decoding (Murray)

1. Linear discriminant analysis
2. Perceptrons
3. Support vector machine
4. Neural networks

Week 12 (4/26): Time series analysis II – analysis of neuroimaging data (Levy)

1. Characteristics of the BOLD signal and neuroimaging data
2. Univariate analysis: GLM
3. Solutions for multiple comparisons in neuroimaging data
4. Introduction to multivariate analyses

Week 13 (5/3): Image analysis (Clark)

1. Two examples: Catherine dataset from Mi1 & Old Liana dataset (orBrian dataset)
2. What is an image/movie: properties (pixels, levels, bit-depth, etc.)
3. Image manipulation in Matlab is all about playing with pixel values to extract the data that you want
4. Threshholding, contiguous areas, properties
5. Filtering: median, mean, other shapes
6. Computing intensities and dF/F
7. PCA, ICA, other methods applied to image sequences