STAT 242b/542b THEORY OF STATISTICS: COURSE SYLLABUS

GRADING: Weekly homework 40%, First test 30%, Second test 30%
DISCUSSION SESSION: Monday 8:00pm-9:00pm, 24 Hillhouse Ave.
OFFICE HOURS: MWF: In Classroom 10:15 to 10:25am (after class), then Stat Dept, Dana House (10:30am to 11am)
TFs: Xinyuan Chen, Xin Xu, Mingrui Zhang. Basement of Dana House.

WEEK 1: January 18,20. PROBABILITY REVIEW

* Probability models and the situations in which they arise (Ch. 1,2,3).
* Expected values and variances of sample means. Use of the Central Limit Theorem. Normal approximation. Delta method. (Ch. 4,5).

WEEK 2: Jan 23,25,27. DISTRIBUTIONS BASED ON NORMAL

* SOME MORE PROBABILITY FOR STATISTICS
  Normal, Chi-square, t, and F distributions for statistics based on the normal. (Ch. 6).
* METHOD OF MOMENTS and the Delta method (Ch. 8.3, 8.4)

WEEK 3: Jan 30. Feb 1,3. ESTIMATION (Ch. 8)

* METHOD OF MAXIMUM LIKELIHOOD (Sections 8.3, 8.4, 8.5)

WEEK 4: February 6, 8,10. SAMPLING DISTRIBUTIONS, LARGE SAMPLE APPROXIMATIONS, RELATIVE EFFICIENCY (Sections 8.3, 8.4, 8.5).

* Comparison of sampling distributions of MLE and other estimators. Relative efficiency.
* Notions of consistency, asymptotic normality, asymptotic variance and standard error.
* Consistency of the MLE.
* Asymptotic Normality of the MLE.

WEEK 5: February 13,15,17,20. SUFFICIENCY, RISK, AND BAYES PROCEDURES (Sec. 8.6-8.8).

* Sufficiency and Likelihood factorization
* Mean Squared error. Bias and variance tradeoff.
* Unbiased estimators. Cramer-Rao inequality.
* Bayes estimators.

WEEK 6: February 22,24. TESTING STATISTICAL HYPOTHESES (Ch.9).

* *Bayes' Formulation, Significance and Power, Neyman-Pearson Formulation.*
* Notions of simple and composite hypotheses concerning distributions and their parameters.
* Neyman-Pearson Lemma for optimal tests in simple versus simple cases.
* Generalized likelihood ratios for composite cases (not necessarily optimal). Simplifying the form of generalized likelihood ratios.
* Choosing the threshold of a test statistic according to the desired significance level.

WEEK 7: February 27, March 1,3. MORE ON TESTING HYPOTHESES AND REVIEW

* Tests for a specific value of a parameter.
  -- one-sided alternatives. Uniformly most powerful tests made easy.
  -- two-sided alternatives. Relationship with confidence intervals.
  -- examples.
* Tests for goodness of fit of an estimated model.
  -- Chi-square test statistic
  -- Generalized likelihood ratio test statistic
  (both approximately Chi-square distributed under the null hypothesis.)
  -- Accounting for degrees of freedom.
  -- Example.
* Review of Inference
  -- Selecting probability models.
  -- Estimating parameters in the model.
  -- Testing statistical hypotheses.

WEEK 8: March 6,8,10. FIRST EXAM AND SUMMARIZING DATA

* **March 6. <u>FIRST TEST</u>**
  Covering chapters 8, 9 and any additional material presented.
* March 8,10. SUMMARIZING DATA (a sampling of ch. 10).
  Plotting and interpreting empirical distributions,
  quantile-quantile plots, histograms, densities, stem & leaf, and box-plots.

March 11-26. SPRING BREAK

WEEK 9: March 27,29,31. HYPOTHESIS TESTS: LOCATIONS IN 2 SAMPLES (Ch. 11).

* Tests for location in independent samples
  -- various tests do about the same thing.
  -- specifics for tests of location depend on assumptions about the variances.
* Tests for paired samples
  -- Reduces to a one-sample test regarding differences.
  -- Easier and more accurate than with independent samples.
* Issues in the design of experiments.

WEEK 10: April 3,5,7. HYPOTHESIS TESTS: LOCATIONS IN MANY SAMPLES (Ch. 12).

* One-way layout
  -- plots of means or medians (with box-plots).
  -- multiple comparisons.
  -- analysis of variance.
* Two-way layout
  -- superimposed plots of mean response for each value of a second factor.
  -- the additive model and its graphical interpretation.
  -- tests for additive and interaction models.
  -- analysis of variance.

WEEK 11: April 10,12,14. LEAST SQUARES FITS OF LINES/CURVES (Ch. 14.1,14.2).

* Solving and interpreting the least squares problem in linear regression.
  -- Graphical interpretation (best fitting line)
  -- Solution in terms of standardized variables (slope = correlation coefficient r)
  -- Solution in terms of original variables
  -- Pythagorean identity for sum of squared errors and its geometrical interpretation.
  -- least squares projection yields residual vector which is
  orthogonal to the vector of the explanatory variable.
  -- residual plots.
* Statistical Properties of the estimated parameters.
  -- standard linear model with homoscedastic errors.
  -- means, variances, and distribution of estimated parameters.
  -- estimated variances, standard errors, confidence intervals.
  -- heteroscedastic errors and non-linearities revealed through residual plots.
  -- uses of transformations. Transforming inputs to correct for non-linearities.
  Transforming outputs to correct for heteroscedasticity.
* Galton and the ``regression'' interpretation of least squares.

WEEK 12: April 17,19,21. LINEAR LEAST SQUARES WITH MANY VARIABLES.

* Iterative Projection Interpretation
* Matrix Interpretation
* Statistical Properties
  -- Standard statistical model
  -- Mean and covariance matrix of least squares estimates.
  -- Estimation of sigma^2.
  -- Standard errors of coefficients.
  -- Confidence intervals for coefficients.
  -- Hypothesis tests concerning values of coefficients.
  -- t-values and interpreting regression output
  (testing whether a coefficient is zero, when the others are not).
  -- standard errors and confidence intervals for regression fits.
* Practice with multiple regression, tests and confidence intervals.

WEEK 13: April 24,26. MODEL SELECTION.

* Step-size selection by examination of largest magnitude t-value (merits and difficulties).
* Prediction error criterion (merits and difficulties).
* Practice with model selection by either method.

* **April 28: <u>SECOND TEST</u>**