TIGP Bio 2020 Fall Syllabus

Fundamental Statistical Methods in Bioinformatics (S1)

Place: Room 107, New Building of the Institute of Information Science, Academia Sinica.

Time: Thursday 09:00-12:00

Chair: Dr. Shinsheng Yuan (shinshengyuan@gmail.com)

Outline: This course covers the fundamentals of statistics and basic tools for bioinformatics analysis. In the first part students will learn basic statistical concepts and methods, including probability, random variables and distributions, parameter estimation, hypothesis testing, regression analysis, and categorical data analysis. In the second part several commonly used methods in bioinformatics will be introduced, including statistical meta analysis, survival analysis, clustering, classification, and nonparametric statistics.

Textbook: Fundamentals of Biostatistics (author: Bernard Rosner), Cengage Learning.

Reference book: Pattern Recognition (author: Richard O. Duda, Peter E. Hart, and David G. Stork), Wiley.

Lecturers: Chen-Hsiang Yeang, Hsin-Chou Yang, Shinsheng Yuan, Grace Shieh, Hsuan-Yu Chen, and Wei-Chung Liu

TA: Erickson Erigio Fajiculary 艾瑞克 **Email**: efajiculay@yahoo.com

TA Office hours: Tuesday 16:00 pm-18:00 pm

TA Office location: Room 304, Old Building of the Institute of Information Science, Academia Sinica

Grades: Midterm exam 50%. Final exam 50%.

For the most up-to-date syllabus, please visit https://tigpbp.iis.sinica.edu.tw/tigpbio/index.html

Week	Date	Topics/Brief Description	Sub-topics/ Detailed Description	Lecturers
1	2020/09/16(Wed)	Descriptive Statistics, Genomic Data Analysis	(1) Introduction to statistics (2) Descriptive statistics (3) Fundamental of molecular biology (4) Genomic data analysis	Dr. Chen-Hsiang Yeang
1	2020/09/17 @N106	Probability	(1) Applications in statistical genetics (2) Combinatorial analysis (3) Axioms of probabilities (4) Conditional probability and independence (5) Random variable and distribution function	Dr. Hsin-Chou Yang
2	2020/09/24	Discrete Distributions and Contingency Tables	(1) An application in pharmacogenetic study (2) Discrete/continuous/mixed distributions (3) Joint/marginal/conditional distributions (4) Special discrete distributions (5) Introduction to contingency table	Dr. Hsin-Chou Yang
4	2020/10/01(Holiday) -> 10/08	Continuous Distributions, Expectations and Basic Statistics	(1) Continuous random variable (2) Expectation (3) Basic statistics (4) Limit theorems (optional)	Dr. Hsin-Chou Yang
5	2020/10/15	Parameter Estimation and Confidence Interval	(1) Unbiasedness (2) Point estimation (substitution principles, least square estimate, maximum likelihood	Dr. Shinsheng Yuan

			estimate) (3) Interval estimation	
6	2020/10/22	Hypothesis Testing, P-value and False Discovery Rate	(1) Hypothesis testing (2) Type I error and type II error (3) P-value (4) One-sample and two-sample z-tests (5) One-sample, two-sample, and paired t-tests (6) Bonferroni adjustment, false discovery rate, and q value	Dr. Grace Shieh
7	2020/10/29 @O108 ->Moved to 12/16(Wed)	--	--	--
8	2020/11/05	Review Week (no class)		
9	2020/11/12	Midterm Exam		
10	2020/11/19	Correlation, Regression Analysis and ANOVA	(1) Linear regression with one independent variable (2) Inference in regression analysis (3) Diagnostic and remedial measures (4) Simultaneous inferences and other topics (5) Matrix approach to simple linear regression (6) Multiple linear regression (7) Building the regression model (8) Qualitative predictor variables (9) Analysis of variance	Dr. Grace Shieh
11	2020/11/25(Wed) ->Moved to 12/01 (Tues)	--	--	--
11	2020/11/26	Survival Data Analysis	(1) Mantel-Haenszel test (2) Survival and hazard functions (3) Kaplan Meier estimate (4) Log-rank test (5) Proportional-hazards model (6) Lung cancer study	Dr. Hsuan-Yu Chen
12	2020/12/01(Tues) 14:00-17:00 (Moved from 11/25)	Logistic Regression and Statistical Meta Analysis	(1) Logistic regression (2) Meta analysis (effect size, precision, study weights, summary effect, heterogeneity, fixed-effect model, random-effect model, software)	Dr. Shinsheng Yuan
12	2020/12/02(Wed) ->12/03	Clustering	(1) Clustering by geometry (K-means, EM algorithm, hierarchical clustering, self-organizing map, principal component analysis, independent component analysis) (2) Clustering on graphs (Basic concepts, max flow – min	Dr. Chen-Hsiang Yeang

			cut, normal cuts, spectral clustering, and community detection) (3) Advanced topics (Chinese restaurant process and affinity propagation)	
12	2020/12/03 ->Moved to 12/24	--	--	--
13	2020/12/10	Nonparametric Statistics (I)	(1) Bootstrap (2) One-sample sign test (3) One-sample Wilcoxon signed-rank test (4) Wilcoxon rank-sum test (Mann-Whitney U test) (5) Sign test for paired data (6) Wilcoxon signed-rank test for paired data	Dr. Wei-Chung Liu
14	2020/12/16(Wed) (Moved from 10/29)	Analysis of Categorical Data, Chisq and Multinomial	(1) Incidence proportion and odds ratio (2) Two-sample test for binomial proportions (3) Contingency-table approach (4) Fisher's exact test (5) McNemar's test (6) Kappa statistic	Dr. Hsuan-Yu Chen
14	2020/12/17	Nonparametric Statistics (II)	(1) Kruskal-Wallis test (2) Randomization/permutation test for two-way ANOVA (3) The product-moment correlation coefficient (4) Spearman rank correlation (5) Kendall's coefficient of rank correlation	Dr. Wei-Chung Liu
15	2020/12/24 (Moved from 12/03)	~~Review Week (no class)~~ -> Classification	(1) Binary classifications (Bayesian detection theory, naïve Bayes and other generative models, non-parametric methods, Fisher linear discriminant, support vector machines and kernel techniques, artificial neural networks, and boosting) (1) Advanced topics in supervised learning (multi-class classification, feature selection, semi-supervised learning, active learning)	Dr. Chen-Hsiang Yeang
16	2020/12/31	Final Exam		