Biostatistics course
2023/2024

Syllabus

1. General review (1 day)
   a. What is Biostatistics?
   b. Population/Sample/Sample size
   c. Type of Data – quantitative and qualitative variables
   d. Common probability distributions
   e. Work example – Malaria in Tanzania
2. Applications in Medicine (7 days)
   a. Construction and analysis of diagnostic tools – Binomial distribution, ROC curve, sensitivity, specificity, Rogal-Gladen estimator (2 days)
   b. Estimation of treatment effects - generalized linear models (2 days)
   c. Survival analysis - Kaplan-Meier curve, log-rank test, Cox's proportional hazards model (3 days)
3. Applications in Genetics, Genomics, and other 'omics data (3 days)
   a. Genetic association studies – Hardy-Weinberg test, homozygosity, minor allele frequencies, additive model, multiple testing correction (1 day)
   b. Methylation association studies – M versus beta values, estimation of biological age (1 day)
   c. Gene expression studies based on RNA-seq experiments – Tests based on Poisson and Negative-Binomial (1 day)
4. Other Topics (3 day)
   a. Estimation of Species diversity – Diversity indexes, Poisson mixture models (1 day)
   b. Serological analysis – Gaussian (skew-normal) mixture models (1 day)
   c. Advanced sample size and power calculations(1 day)

1 day – Presentations

Evaluation

40% - Presentation – 60% Oral exam (individual).

READING MATERIAL

Applications in Medicine

1. Caroni C. Regression Models for Lifetime Data: An Overview. Stats 2022; 5: 1294-1304. doi: 10.3390/stats5040078
2. Longley RJ, White MT, Takashima E, Brewster J, Morita M, Harbers M, Obadia T, Robinson LJ, Matsuura F, Liu ZSJ, Li-Wai-Suen CSN, Tham WH, Healer J, Huon C, Chitnis CE, Nguitragool W, Monteiro W, Proietti C, Doolan DL, Siqueira AM, Ding XC, Gonzalez IJ, Kazura J, Lacerda M, Sattabongkot J, Tsuboi T, Mueller I. Development and validation of serological markers for detecting recent Plasmodium vivax infection. Nat Med. 2020;26(5):741-749. doi: 10.1038/s41591-020-0841-4.
3. McCullaugh P, Nelder JA. Generalized Linear Models. Chapman & Hall CRC, New York.
4. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Phys Med Biol. 2018;63(7):07TR01. doi: 10.1088/1361-6560/aab4b1.
5. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. Am J Epidemiol. 1978; 107(1):71-6. doi: 10.1093/oxfordjournals.aje.a112510.

Applications in Genetics, Genomics, and other 'omics data

1. Campagna MP, Xavier A, Lechner-Scott J, Maltby V, Scott RJ, Butzkueven H, Jokubaitis VG, Lea RA. Epigenome-wide association studies: current knowledge, strategies and recommendations. Clin Epigenetics. 2021;13(1):214. doi: 10.1186/s13148-021-01200-8.
2. Fang Z, Martin J, Wang Z. Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. Cell Biosci. 2012; 2(1):26. doi: 10.1186/2045-3701-2-26.
3. Grabowska AD, Lacerda EM, Nacul L, Sepúlveda N. Review of the Quality Control Checks Performed by Current Genome-Wide and Targeted-Genome Association Studies on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Front Pediatr. 2020;8:293. doi: 10.3389/fped.2020.00293.
4. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608. doi: 10.1002/mpr.1608.
5. Uffelmann E, Huang QQ Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. Genome-wide association studies. Nat Rev Methods Primers 2021; 1:59. doi: 10.1038/s43586-021-00056-9

Other Topics

1. de Oliveira MR, Subtil A, Gonçalves L. Common Medical and Statistical Problems: The Dilemma of the Sample Size Calculation for Sensitivity and Specificity Estimation. Mathematics 2020; 8:1258. doi: 10.3390/math8081258
2. Domingues TD, Mouriño H, Sepúlveda N. Analysis of antibody data using Finite Mixture Models based on Scale Mixtures of Skew-Normal distributions. medRxiv 2002; 2021.03.08.21252807; doi: 10.1101/2021.03.08.21252807

3. Gonçalves L, de Oliveira MR, Pascoal C, Pires A. Sample size for estimating a binomial proportion: comparison of different methods. J. Appl. Stat. 2012; 39, 2453-2473. doi: 10.1080/02664763.2012.713919
4. Malato, J.; Graça, L.; Sepúlveda, N. Impact of Misdiagnosis in Case-Control Studies of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Diagnostics 2023; *13*: 531. doi: 10.3390/diagnostics13030531
5. Sepúlveda N, Drakeley C. Sample size determination for estimating antibody seroconversion rate under stable malaria transmission intensity. Malar J. 2015;14:141. doi: 10.1186/s12936-015-0661-z.
6. Sepúlveda N, Paulino CD, Carneiro J. Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. J Immunol Methods. 2010;353(1-2):124-37. doi: 10.1016/j.jim.2009.11.009.
7. Sepúlveda N, Paulino CD, Drakeley C. Sample size and power calculations for detecting changes in malaria transmission using antibody seroconversion rate. Malar J. 2015;14:529. doi: 10.1186/s12936-015-1050-3.
8. Sepúlveda N, Stresman G, White MT, Drakeley CJ. Current Mathematical Models for Analyzing Anti-Malarial Antibody Data with an Eye to Malaria Elimination and Eradication. J Immunol Res. 2015;2015:738030. doi: 10.1155/2015/738030.