Course Description for Stat 541: Introduction to Biostatistics

Lecturer: Ismor Fischer, UW Department of Statistics

Objective: The overall goal of this course is to provide to students in the Summer Institute for Training in Biostatistics (SIBS) program an overview of fundamental statistical concepts, and a practical working knowledge of the basic statistical techniques they are likely to encounter in applied research and literature review contexts.

I. **Introduction.** General ideas, interpretation, and terminology: population, random sample, random variable, empirical data, etc. Describing the formal steps of the classical scientific method – hypothesis, experiment, observation, analysis and conclusion – to determine if sources of variation in a system are genuinely significant or due to random chance effects. General study design considerations: prospective (e.g., randomized clinical trials, cohort studies) versus retrospective (e.g., case-control studies).

II. **Exploratory Data Analysis and Descriptive Statistics.** Classification of data: numerical (continuous, discrete) and categorical (nominal – including binary – and ordinal). Graphical displays of data: tables, histograms, stemplots, boxplots, etc. Summary Statistics: measures of center (sample mean, median, mode), measures of spread (sample range, variance, standard deviation, quantiles), etc., of grouped and ungrouped data. Distributional summary using Chebyshev's Inequality.

III. **Probability Theory.** Basic definitions: experiment, outcomes, sample space, events, probability. Basic operations on events and their probabilities, including conditional probability, independent events. Specialized concepts include diagnostic tests (sensitivity and specificity, Bayes' Theorem, ROC curves), relative risk and odds ratios in case-control studies.

IV. **Probability Distributions and Densities.** Probability tables, probability histograms and probability distributions corresponding to discrete random variables, with emphasis on the classical Binomial and Poisson models. Probability densities and probability distributions corresponding to continuous random variables, with emphasis on the classical Normal (a.k.a. Gaussian) model.

V. **Statistical Inference.** Background: Sampling distributions and the Central Limit Theorem. Formulation of null and alternative hypotheses, and associated Type I and Type II errors. One- and two-sided hypothesis testing methods for population parameters – mostly, means and proportions – for one sample or two samples (independent or dependent), large (Z-test) or small (t-test). Light treatment of hypothesis testing for population variances (χ^2-test for one, F-test for two). Specifically, for a specified significance level, calculation of confidence intervals, acceptance/rejection regions, and p-values, and their application and interpretation. Power and sample size calculations. Brief discussion of nonparametric (Wilcoxon) tests. Multiple comparisons: ANOVA tables for means, χ^2 and McNemar tests on contingency tables for proportions. Mantel-Haenszel Method for multiple 2×2 tables (i.e., Test of Homogeneity $\rightarrow$ Summary Odds Ratio $\rightarrow$ Test of Association).

VI. **Survival Analysis.** Survival curves, hazard functions, Kaplan-Meier Product-Limit Estimator, Log-Rank Test, Cox Proportional Hazards Regression Model.

VII. **Linear Regression.** Plots of scattergrams of bivariate numerical data, computation of sample correlation coefficient r, and associated inference. Calculation and applications of corresponding least squares regression line, and associated inferences. Evaluation of fit via coefficient of determination r^2 and residual plot. Additional topics include: transformations (logarithmic and others), logistic regression (e.g., dose-response curves), and multilinear regression (including a brief discussion of drug-drug interaction, ANOVA formulation and model selection techniques).