Attitudinal aims

In addition to specific learning outcomes, the course aims to shape the attitudes of learners regarding the field of biostatistics. Specifically, the course aims to:

1. Instill an appreciation for how statistics is important to a variety of biomedical disciplines.
2. Provide a foundation that enables critical evaluation of modern statistical approaches applied to biomedical research.
3. Provide a foundation that enables active participation in the application of statistics to biomedical research.

Learning outcomes

Each numbered item states a learning aim for the course (with the number of lectures devoted to each), and the items that follow indicate the respective learning outcomes (or objectives).

1. Demonstrate the ability to generate appropriate descriptive statistics (2 lectures, 1 lab).
   a. Identify nominal, ordinal, discrete and continuous data.
   b. Interpret frequency distributions for summarizing data sets (such as relative and cumulative frequency).
   c. Interpret graphical methods for summarizing data sets (such as boxplots, histograms and scatterplots).
   d. Interpret summary statistics for summarizing data sets including measures of central tendency (such as mean, median and mode) and measures of dispersion (such as range, variance and interquartile range).
   e. Assess which methods for summarizing a data set are most appropriate to highlight interesting features of the data and identify outlying values.
   f. Identify the features that describe a data distribution.
   g. Use an appropriate software tool for data summary and exploratory data analysis.

2. Demonstrate an understanding of the basic concepts of probability (1 lecture, 0.5 lab).
   a. Define rudimentary mathematical properties of probability.
   b. Explain the sample space for random experiments.
   c. Describe probability in terms of long-term relative frequencies in repetitions of experiments.
   d. Define the terms independent, disjoint and complementary events.
   e. Calculate probabilities of single events, complementary events and the unions and intersections of collections of events.

3. Demonstrate an understanding of the concepts of diagnostic testing and screening (1 lecture, 0.5 lab).
   a. Recognize Bayes' Rule and how it applies to diagnostic testing.
   b. Define the terms sensitivity, specificity, predictive value positive and predictive value negative.
   c. Identify the features that contribute to an optimal screening test.
   d. Explain the role of receiver operating characteristic (ROC) curves in assessing a screening test.

4. Demonstrate an understanding of the basic concepts of probability distributions (2 lectures, 1 labs).
   a. Explain the difference between discrete and continuous random variables.
   b. Calculate the mean and variance of a discrete random variable.
   c. Apply general properties of the expectation and variance operators.
   d. Identify the key properties of the Normal distribution.
   e. Find the following for a Normal distribution: (i) the probability over a set of values, (ii) a percentile.
   f. Identify the key properties of the Binomial distribution.

g. Compute probabilities for a Binomial distribution.
h. Identify the key properties of the Poisson distribution.
i. Compute probabilities for a Poisson distribution.
j. Explain the difference between the Normal, Binomial and Poisson distributions.

5. Understand the concept of the sampling distribution of a statistic (2 lectures, 1 lab).
   a. Explain the difference between a population and a sample, and between parameters and statistics.
   b. Describe properties of the sampling distribution of the sample mean and relate this to the Central Limit Theorem.
   c. Interpret a confidence interval and identify features that determine the width of a confidence interval.

6. Design studies for obtaining data while minimizing bias and inefficiency (0.5 lecture, 0.5 lab).
   a. Identify methodological differences between observational studies and experimental studies.
   b. Identify features common in experimental design (such as experimental units, treatments, factors)
   c. Identify features common in observational design (such as cases, controls, cohort)
   d. Identify possible sources of bias in observational and experimental studies.

7. Understand the foundations for classical inference involving hypothesis testing (0.5 lecture, 0.5 lab).
   a. Identify the components of a classical hypothesis test, including the parameter of interest, the null and alternative hypotheses and the test statistic.
   b. Define the P-value of a test statistic.
   c. Compute, or approximate, the P-value of a test statistic.
   d. Explain the difference between type I and type II errors.
   e. Calculate statistical power using statistical software and identify the parameters that contribute to statistical power (such as sample size, effect size, standard deviation, alpha-level).
   f. Calculate the required sample size to achieve the desired statistical power using statistical software.

8. Apply inferential methods relating to the means of Normal distributions (1 lecture, 1 lab).
   a. Identify the difference between the t distribution and the Normal distribution.
   b. Construct one- and two-sided hypothesis tests and confidence intervals for the mean of a Normal distribution where the underlying variance is either known or unknown.
   c. Construct one- and two-sided hypothesis tests and confidence intervals for the mean difference from a matched-pair design where appropriate.
   d. Conduct one- and two-sided hypothesis tests and confidence intervals for the difference in the means of two Normal distributions where the underlying variance is either known or unknown.

9. Apply and interpret the methods of one-way analysis of variance (ANOVA) (1.5 lectures, 1 lab).
   a. Identify situations where one-way ANOVA is and is not appropriate.
   b. State the modeling assumptions underlying ANOVA.
   c. State the null and alternative hypotheses for the ANOVA test.
   d. Explain the partitioning of the total sum of squares into the within- and between-group components.
   e. Identify the degrees of freedom associated with each sum of squares.
   f. Perform the F test in ANOVA, evaluating or approximating the P-value of the test statistic.
   g. Interpret an ANOVA table.
   h. Explain ANOVA output from statistical software.

10. Demonstrate an understanding of, and corrections for, multiple comparisons (0.5 lecture, 0.5 lab).
    a. Explain the difference and appropriate use of planned versus post-hoc comparisons.
    b. Conduct an appropriate correction for multiple comparisions.

11. Interpret and analyze categorical data (2 lectures, 1 lab).
    a. Recognize when an RxC contingency table is an appropriate way to summarize a data set.
    b. Compute and interpret marginal and conditional distributions from a contingency table.
    c. Perform the Chi-squared test for association for an RxC contingency table.
    d. Describe the possible relationships between two categorical variables in a two-way table (including risk ratios and odds ratios and respective confidence intervals).
    e. Identify the differences between the risk ratio and odds ratio in terms of how they are calculated and how they are interpreted.

12. Apply and interpret basic summary and modeling techniques in the context of simple and multiple linear models with Normally distributed errors (2 lectures, 1 lab).
    a. Identify a possible relationship between two continuous variables from a scatterplot.
    b. Interpret a sample correlation (such as the Pearson correlation coefficient).
    c. Define the concept of least squares estimation in linear regression.
    d. Identify which variable is most naturally the response variable in a regression analysis.
    e. Fit a linear model to a data set via software.
    f. Evaluate the fit of a linear model by consideration of the residuals and R^2.
    g. Conduct the appropriate data transformation to improve model fit.
    h. Interpret the parameter estimates in a fitted linear model including interaction terms.
    i. Conduct inference for the slope and intercept parameters, including construction of confidence intervals and hypothesis tests.
    j. Calculate interval estimates for the mean and predicted responses at a given value of the explanatory variable.
    k. Explain the linear regression output from statistical software.

13. Apply and interpret basic summary and modeling techniques in the context of logistic regression (2 lectures, 1 lab).
    a. Identify situations where a logistic regression model would be appropriate.
    b. Fit a logistic regression model to a data set via software.
    c. Interpret the parameter estimates in a logistic regression model including interaction terms.
    d. Conduct inference for the parameter estimates, including construction of confidence intervals and hypothesis tests.
    e. Explain the logistic regression out from statistical software.

14. Apply and interpret basic summary and modeling techniques in the context of person-time and survival data (2 lectures, 1 lab).
    a. Identify the structure and content of person-time data.
    b. Compute an incidence rates and incidence rate ratios, and construct confidence intervals for incidence rate ratios.
    c. Identify the structure and content of survival (or time-to-onset) data.
    d. Interpret graphical tools to summarize survival data (such as Kaplan-Meier plots).
    e. Fit a Cox proportional hazards model to a data set via software.
    f. Interpret the parameter estimates in a Cox proportional hazards model including interaction terms.
    g. Conduct inference for the parameter estimates, including construction of confidence intervals and hypothesis tests.

h. Explain the Cox proportional hazards output from statistical software.

15. Recognize and apply modern approaches to random effects and repeated measures/longitudinal data (4 lectures, 1 lab).

a. Identify the structure of random effects, repeated measures/longitudinal data and identify the features common in repeated measures/longitudinal study designs.
b. Identify the difference between single-group and parallel group designs.
c. Identify the difference between independent, unstructured and compound symmetry covariance structures.
d. Conduct random-effects ANOVA using statistical software and interpret the output.
e. Conduct repeated-measures ANOVA using statistical software and interpret the output.
f. Understand the limitations of repeated-measures ANOVA including covariance assumptions and missing data issues.
g. Identify situations where a linear mixed-effects model would be appropriate.
h. Interpret graphical tools to summarize repeated measures/longitudinal data (such as interaction plots).
i. Interpret the parameter estimates from a linear-mixed effects model and explain linear mixed model output from statistical software.
j. Identify situations where a generalized estimating equation (GEE) model would be appropriate.
k. Interpret the parameter estimates from a GEE model and explain GEE model output from statistical software.