Statistics and biological data analysis

This course aims at bridging the gap between classical statistical courses and real life biological data analysis. Students will be introduced to basic concepts in statistics and probability distributions that underlie most of the statistical methods employed in testing biological hypotheses. The course is structured to give a natural flow to a data analysis question starting with formulation of hypothesis, experimental design, data collection, and analysis. Here is a brief breakdown of the course content:

1. **Biological data and measures of central tendency and dispersion:** Categories of biological data will be discussed. Relevant basic data summarization (mean, mode, median, standard deviation, Skewness, kurtosis …) will also be addressed.
2. **Overview of Probability distributions relevant to biological data:** Quick introduction to probability and probability distributions will be given with emphasis on their application to biological studies
3. **Hypothesis, Experimental design and Data collection:** Hypothesis and experimental design are usually overlooked or are not given sufficient attention in biological experiments. This section will stress on how to formulate a hypothesis, design a relevant experimental design, and finally address the point on how to gather data in an efficient and organized manner.
4. **Graphical exploratory data analysis:** Covers different ways of presenting biological data graphically while simultaneously focusing on the pertinent information. Points on how to embed data summarization to graphics will also be addressed.
5. **Low level data analysis:** Introduction will be given to basic parametric and non-parametric statistical methods that will test the hypothesis at hand. Focus will also be given on how one decides to use a test based on the

Jun.-Prof. Dr. Medhanie Mulaw
Comprehensive Cancer Center Ulm
Institute of Experimental Cancer Research
University Hospital Ulm
Albert-Einstein-Allee 11, 89081 Ulm
Tel.- Nr.: 0731 - 500 65812
FAX - Nr: 0731-500 65802
E-Mail: medhanie.mulaw@uni-ulm.de

experimental design and the nature of the data collected (see points 2 and 3).

6. **Advanced data analysis:** Regression, analysis of variance (ANOVA), and generalized linear models will be covered in this section. Depending on the availability of time, quick introduction to multivariate analysis (principal components (PCA), linear discriminant analysis (LDA), clustering and multi-dimensional scaling…) will be given along with some applied examples from biological studies (microarrays, next generation sequencing, cytometry…).

The course will primarily focus on biological examples and exercises and how one would select an appropriate statistical method to extract meaningful information. Relevant statistical packages will also be suggested but it is not the scope of the course to have a hands on training for any of these applications. The course is also not meant as a replacement to the basic introductory course on statistics. Students who have taken at least one introductory course on statistics at either undergraduate or graduate level will benefit most out of this course. Students who never took any course on statistics and would like to join the course are highly encouraged to read any of the plethora of excellent text books on statistics. The online freely available OpenIntro Statistics textbook would be a good starting point (http://www.openintro.org/stat/textbook.php).

At the end of the course, students will be able to:

1. Understand the principles of hypothesis formulation and have basic grasp on how to design an experiment and collect data in a scientific manner
2. Be able to make an informed decision on selecting a relevant statistical analysis for a given dataset
3. Know how to present their results in a clear and informative way
4. Have a hint about some of the advanced statistical methods used in complex biological datasets

Jun.-Prof. Dr. Medhanie Mulaw
Comprehensive Cancer Center Ulm
Institute of Experimental Cancer Research
University Hospital Ulm
Albert-Einstein-Allee 11, 89081 Ulm
Tel.- Nr.: 0731 - 500 65812
FAX - Nr: 0731-500 65802
E-Mail: medhanie.mulaw@uni-ulm.de

References:

Myra L. Samuels, Jeffrey A.Witmer, Andrew A. Schaffner. **Statistics for the life sciences**.**4th ed.** Pearson Education, Inc.

Gerry P. Quinn, Michael J. Keough. **Experimental Design and Data Analysis for Biologists**. Cambridge.

Helle Sørensen and Claus Thorn Ekstrom. **Introduction to Statistical Data Analysis for the Life Sciences**. CRC Press.

Jun.-Prof. Dr. Medhanie Mulaw
Comprehensive Cancer Center Ulm
Institute of Experimental Cancer Research
University Hospital Ulm
Albert-Einstein-Allee 11, 89081 Ulm
Tel.- Nr.: 0731 - 500 65812
FAX - Nr: 0731-500 65802
E-Mail: medhanie.mulaw@uni-ulm.de