

ST2238
INTRODUCTORY BIOSTATISTICS

Prerequisite: ST1232

Workload: 39 lecture hours + 11 tutorial hours

This course introduces the basic principles and methods of biostatistics. It has been designed specifically for life sciences students who wish to learn modern research methods for analysing and extracting information from biological, biomedical and genomic data. Relevant examples will be used throughout the course to illustrate various techniques. The computer package Minitab will be used to enable students to put the techniques into practice. The course is split into four parts. The first will show how to demonstrate that two populations or treatments differ from each other in some way. The second part extends this to the case where there are multiple types of organisms or objects being considered. In the third part of the course, we cover regression, in which an explanatory variable x is used to predict or understand a response variable y . Finally we cover how to answer these same scientific questions when the data are not continuous measurements, but rather counts.

S/ N	Topics	Lecture hours
1.	Comparing two treatments or populations In this part of the course we cover the common situation in which you have samples from two populations, or from a single population to which two different treatments have been applied, and wish to assess whether the two populations differ or the treatment affects the outcome. There are four scenarios to consider depending on whether the observations look “normally” distributed or not, and whether the observations are paired (measurements taken on the same individual before and after treatment for example) or not. As well as recapping the one-sample t-test you learned in ST1232, we cover the paired and two-sample t-tests, and non-parametric tests: the sign test and the confusingly named Wilcoxon signed rank and Wilcoxon rank sum tests.	9
2.	Comparing two or more treatments or populations In this part of the course we extend the aim of the previous chapter to the situation in which there are many treatments or populations and we wish to investigate whether there are <i>any</i> differences between them. We do so using the technique called analysis of variance (ANOVA). We will cover one-way ANOVA, two-way ANOVA (which can be thought of as an extension of the paired t-test), experimental design and randomisation.	6
3.	Explaining outcomes via covariates In this part of the course we look at regression. In its simplest form, regression involves modelling a response y as a linear function of a predictor x plus additional, unexplained biological variability. We extend this to the scenario in which multiple predictors are used, and consider how to decide which predictors add light to darkness and which merely create more darkness. We also cover correlation and its non-parametric cousin, rank correlation.	6

4.	Categorical data analysis In this final part of the course, we consider the situation in which our observations are not continuous measurements, as are weight of dried plant mass, but discrete numbers, such as the number of birds' nests that are parasitised by cuckoos. Apart from the different nature of the data, the scientific questions are similar to those asked in the first three parts of the course: do two treatments differ in their effect?, do multiple treatments differ in their effect?, how do one or more predictors influence the outcome? In addition to answering these questions via the media of the binomial test, Fisher's exact test, contingency tables and McNemar's test, we address the issue of evaluating goodness of fit of a statistical model or assumption, epidemiological jargon such as odds ratios, and the problem of confounding. We conclude the course with lessons on logistic and Poisson regression, both methods of accounting for predictors in understanding count data.	12
	Total Lectures:	33h
	Tutorials:	11h
	Total hours:	50h

Note: some lecture hours will be used for other purposes, such as revision and the mid-term test. These hours are not included in the total lecture hour count above, but are in the total hours count.

REFERENCE BOOKS

Forthofer, Lee and Hernandez (2007), *Biostatistics: a guide to design, analysis, and discovery*, 2ed, Academic Press (Elsevier).

MODE OF ASSESSMENT: 40% CA, 60% Exam.

MODULE CO-ORDINATORS:

Dr Alex Cook

LECTURERS:

Dr Alex Cook