

**Statistics for Graduate Biologists**  
**Dr Brian McCabe**  
tel: 41818/01, email [b.j.mccabe@zoo.cam.ac.uk](mailto:b.j.mccabe@zoo.cam.ac.uk)

**Aims**

- To help you to acquire the statistical skills necessary for research projects and evaluation of the literature.
- To provide practice in performing common statistical analyses using popular statistical packages.

**Learning outcomes**

- Detailed learning outcomes are given in the handout for each lecture. After each lecture, and the associated practical exercises and reading, you should be able to perform simple statistical analyses based on the ideas discussed in the lecture.
- At the end of the course, your understanding is expected to be sufficient for devising and analysing simple experimental designs independently. For more complex statistical problems, you should be able to design experiments and surveys, and analyse your data correctly, on the basis of specialist advice.

**Learning time**

It is suggested that, soon after each lecture, you spend one hour going over the material in it. The amount of subsequent private study will depend on your background and your field of interest.

Two hours' revision of the material in each practical is suggested. The practical exercises and computers are available throughout the year, so there is plenty of opportunity to acquire the biometrical skills that you need.

The course material should meet the requirements of most projects. Your supervisor will advise you if more specialised techniques are needed. For further advice, contact those who teach on this course, particularly Brian McCabe (e-mail [bjm1@cam.ac.uk](mailto:bjm1@cam.ac.uk)).

The practicals are designed to be self-paced and the software is available at all times on the PWF Managed Cluster, which is available at many sites across the University and also from college networks. Since the timetable is quite crowded at the beginning of term, you may find that you need to attend journal clubs, seminars, etc. that clash with the statistics practicals. Should such clashes occur, it will be possible to catch up with the statistics practical work in your own time. You may then obtain any necessary advice from demonstrators at the next practical that you are able to attend.

**REMEMBER - DECIDE ON YOUR STATISTICAL PROCEDURES WHEN DESIGNING  
YOUR PROJECT, BEFORE YOU COLLECT THE DATA**

## Books

The course material is covered by many books and you may wish to refer to several for alternative viewpoints. Two introductory books are especially recommended (asterisks): that by Hinton assumes no previous acquaintance with statistics; that by Dytham focuses on the practical aspects of computer data analysis.

### Introductory books

- Campbell, R.C. (1989) *Statistics for Biologists*, 3rd ed., Cambridge University Press, Cambridge.
- \* Dytham, C. (2003) *Choosing and Using Statistics – a Biologist's Guide* 2nd ed., Blackwell, Oxford.
- Grafen, A. and Hails, R. (2002) *Modern Statistics for the Life Sciences*, Oxford University Press, Oxford. (an excellent introduction to general linear models; more advanced than the other books in this section).
- Hawkins, D. (2005) *Biomeasurement*, Oxford University Press, Oxford. (good associated web site; mainly uses the SPSS statistical computing package).
- \* Hinton, P.R. (2004) *Statistics Explained*, 2nd ed., Routledge, London.
- Howell, D.C. (2004) *Fundamental Statistics for the Behavioral Sciences*, 5th ed., Cengage Learning.
- Moroney, M.J. (1968) *Facts from Figures*, 3rd ed., Penguin, London.
- Velleman, P.F. and Hoaglin, D.C. (1981) *Applications, Basics and Computing of Exploratory Data Analysis*, Duxbury Press, Boston.

### Standard general texts

- Sokal, R.R. and Rohlf, F.J. (1994), *Biometry*, 3rd ed., Freeman, New York.
- Zar, J.H. (1999), *Biostatistical Analysis*, 4th ed., Prentice-Hall, New Jersey.

### Non-parametric statistics

- Conover, W.J. (1999), *Practical Non-parametric Statistics*, 3rd ed., Wiley, New York.
- Siegel, S. (1956), *Non-parametric Statistics for the Behavioural Sciences*, 1st ed. McGraw-Hill, New York.
- Siegel, S. and Castellan, N.J. (1988), *Non-parametric Statistics for the Behavioural Sciences*, 2nd ed. McGraw-Hill, New York.

### Multivariate statistics (if you're interested)

- Manly, B. F. J. (2004), *Multivariate Statistical Methods: a Primer*, 3<sup>rd</sup> ed. Chapman & Hall/CRC, [www.crcpress.com](http://www.crcpress.com).

## Statistical tables

Lindley, D.V. and Scott, W.F. (1995), *New Cambridge Elementary Statistical Tables*, Cambridge University Press, Cambridge.

Rohlf, F.J. and Sokal, R.R. (1994), *Statistical Tables*, 3rd ed., Freeman, New York.

## Statistical computing

The *R* statistical package (see <http://www.r-project.org>) is powerful and comprehensive and will be used in the practicals on the University Managed Cluster Service (MCS). *R* may be downloaded free of charge to your own computer via the above web site. For instructions in the use of *R* see:

- <http://www.bioinformatics.babraham.ac.uk/training.html#rintro>
- Crawley, M. J. (2005), *Statistics: an Introduction Using R*. Wiley, Chichester, UK.
- Crawley, M. J. (2007), *The R Book*. Wiley, Chichester, UK.
- The comprehensive *GENSTAT*, *MINITAB* and *SPSS* packages are also available on the MCS.
- *Statistics for the Terrified*, which you might enjoy, is an interactive package on the MCS.

## Statistical e-books

The CAST series of e-books is work in progress, but covers a fairly wide range of topics, at different levels of sophistication. See [http://cast.massey.ac.nz/collection\\_public.html](http://cast.massey.ac.nz/collection_public.html)

## Lectures

Main Lecture Theatre, Department of Zoology, 2 pm

EXCEPT MONDAY 28<sup>TH</sup> JANUARY (5 PM)

From Friday 18 Jan 2013

Lecturer: Dr. Brian McCabe

1. *Samples and populations; the normal distribution. Fri 18 Jan*  
Measures of location and dispersion; samples and populations; unbiased estimators; the normal distribution; the standard normal curve; probability.
2. *The binomial distribution; testing hypotheses. Mon 21 Jan*  
The binomial distribution; testing hypotheses; the null hypothesis; statistical significance; one- and two-tailed tests; the effect of sample size; statistical power.
3. *The t distribution. Tue 22 Jan*  
Standard error of the mean; comparison of a sample mean with a hypothetical population mean; tables of critical values of  $t$ ; confidence limits; matched pairs  $t$ -test.
4. *Comparison of two independent, approximately normally distributed samples. Wed 23 Jan*  
Sums of squares; two-sample  $t$  test; confidence limits of the difference between two means; one- and two-tailed tests.
5. *One-way analysis of variance (ANOVA). Thu 24 Jan*  
Reasons for using ANOVA; the basic idea of an ANOVA; assumptions of an ANOVA; partition of total sum of squares into between-groups and error sums of squares; calculation of  $F$ ; tables of  $F$ ;  $t$ -tests derived from an ANOVA.
6. *Analysis of variance (continued). Fri 25 Jan*  
Experimental design; planned and unplanned comparisons; checking on the assumptions of an ANOVA; transformations; randomised blocks ANOVA; power analysis.
7. *Association. Mon 28 Jan NB – THIS LECTURE IS AT 5 PM*  
Correlation as a measure of association; Pearson and Spearman correlation coefficients; the basic idea of linear regression; linear regression as an analysis of variance; curvilinear and multiple regression.

8. *Analysis of variance; models and factorial designs. Tue 29 Jan*  
The model of an analysis of variance; factorial analysis of variance; the idea of interaction; general linear models; multivariate ANOVA.
  
9. *Techniques for non-normal data; non-parametric statistics. Wed 30 Jan*  
The three main levels of measurement; Wilcoxon test; Mann-Whitney  $U$  test;  $\chi^2$  goodness-of-fit test; contingency tables ( $\chi^2$  and Fisher exact probability tests).
  
10. *Further aspects of regression and multivariate methods. Thu 31 Jan*  
Prediction from a regression line; linear regression applied to grouped data; analysis of covariance; principal components analysis; discriminant analysis.

### **Practicals**

**(Titan Teaching Room 2, New Museums Site)**

**Mondays 10 am - 12 noon, Wednesdays and Fridays 3 – 5 pm**

**From Mon 21 Jan**

1. *Introduction to the software. Mon 21 Jan (10:00 – 12:00)*
  2. *Distributions. Wed 23 Jan (3:00 – 5:00)*
  3. *Comparison of two samples. Fri 25 Jan (3:00 – 5:00)*
  4. *Analysis of variance. Mon 28 Jan (10:00 – 12:00)*
  5. *Association. Wed 30 Jan (3:00 – 5:00)*
  6. *Regression models. Fri 1 Feb (3:00 – 5:00)*
  7. *Factorial and nested analyses of variance. Mon 4 Feb (10:00 – 12:00)*
- Plus extra exercises using multivariate methods (if you're interested).*

