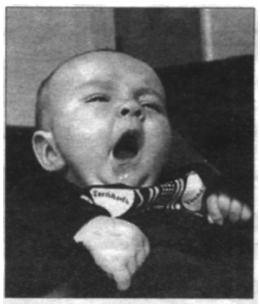
Data analysis methods in weather and climate research

Dr. David B. Stephenson
Department of Meteorology
University of Reading
www.met.rdg.ac.uk/cag

1. Introduction
- Aims of this course
- History and definition of statistics
- Key concepts
- Statistical software

This course aims to develop your

- appreciation of what statistics is really about
- skills at analyzing data using statistical software
- ability to choose and apply appropriate methods
- ability to correctly interpret the analysis results
- capability to learn more about statistics …



Statistics: a great cure for insomnia

This course will not …

- teach you EVERYTHING about statistics!
- give you cookbook recipes for success
- show you how to perform miracles with data
- be the last thing you learn in statistics …

Prerequisites

- Some knowledge of mathematics
- Some experience of computing
- … and lots of curiosity!

Learning activities

- Lectures Mon 11-12 GL61
- Practical classes Tue 14-16 GL20
- Self-study of notes/books/web

Course plan

1. Introduction

Getting started with the R statistical language

2. Exploratory Data Analysis (EDA)

3. Basic probability concepts

4. Probability distributions

5. Parameter estimation

6. Statistical hypothesis testing

7. Basic linear regression

8. Multiple and non-linear regression

9. Introduction to time series

1. Introduction

- Brief history of statistics and probability
- Why "statistics" is more than just "data analysis"
- Descriptive versus inferential statistics
- Statistical software

1. Definition of statistics

Statistics – first applied to the political science concerned with the facts of a state or a community XVIII; all derived immediately from German *statistisch* adj., *statistik sb.;* whence *statistician XIX.*

1. Brief history of Statistics and Probability

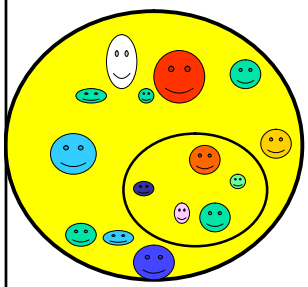
Period	Statistics	Probability
1500-1599	Sir W. Petty	
1600-1699	J. Graunt E. Halley	B. Pascal P. de Fermat
1700-1799		J. Bernoulli C. Huygens A. DeMoivre Rev. Bayes
1800-1899	C.F. Gauss F. Galton	P.S. Laplace
1900-1999	K. Pearson Student aka W. Gosset R. Fisher + many others	Kolmogorov + many others

A generall Bill for this present year, ending the 19 of December 1665. according to the Report made to the KINGS most Excellent Majesty. By the Company of Parish Clerks of London, &c.

1. What exactly is statistics ?

- Exploration and description of sample data
- Statistical inference (e.g. hypothesis testing)
- Stochastic modelling of uncertainty
- Design of experiments and surveys

1. Some fundamental concepts



- Descriptive statistics – exploration and summary of a sample of data
- Inferential statistics – use of sample data to infer properties of the whole population

Nothing is certain in this world …

For example – what will the weather be like in Reading tomorrow?

Some uncertainty is due to random variation between samples (*aleatoric*) and some is due to inherent ignorance about the world (*epistemic*).

Statistical methods can be used to quantify and model this uncertainty.

1. Sources of uncertainty

Observational uncertainty
- sampling error in measurements **A**
- systematic error in measurements (e.g. instrumental biases) **E**
- Inherent uncertainty in statistics caused by sampling of natural variability over finite periods (e.g. the historical record). **A**

Model data sampling uncertainty
- finite length model simulations (e.g. climate time-slices) **A**
- finite length sample of future boundary conditions (e.g. SSTs) **A**
- finite ensemble of all possible model runs **A**

Model parametric uncertainty (known unknowns)
- physical uncertainty in model parameters (e.g. cloud physics) **E**
- sampling uncertainty in statistical estimates of parameters **A**
- non-uniqueness in model parameters (parameter degeneracy) **E**

Model structural uncertainty (unknown unknowns)
- incomplete knowledge of factors (e.g. future emissions). **E**
- misidentification of physical and/or statistical models. **E**
- representation error (e.g. grid boxes rather than point locations). **E**

1. Statistical software

- Statistical language-based software
  e.g. Splus, R, SAS
- Interactive Spreadsheet-like packages e.g. Minitab, SPSS, Excel
- Data analysis software with stat routines e.g. MATLAB, PV-Wave, IDL
- Home made subroutines
  e.g. numerical recipes, friend's code, etc.

1. Summary

- Statistical thinking is a crucial skill for scientific research
- Statistics is more than just mindless data analysis
- Careful choice of methods and good interpretation of the results are skills that need to be learnt
- Key concepts: **sample** and **population**
- Exploratory data analysis (descriptive) and inferential (model-based) approaches
- Powerful statistical software is available to help you!

Further reading …

1. Hand out notes – please spend a couple of hours going through each chapter. Also web based material: www.met.rdg.ac.uk/courses/stats
2. M.R. Spiegel (1992) "Theory and problems of statistics", Schaum ouline series.
3. M.H. DeGroot and M.J. Schervish (2002) Probability and Statistics, (3rd edition), Addison-Wesley, 816 pp.
4. D. Wilks (2005) "Statistical Methods in the Atmospheric Sciences", Academic Press, 2nd edition.
5. H. von Storch and F. Zwiers (1999) "Statistical analysis in climate research", Cambridge University Press.