

Course 3BA1

Statistics and Numerical Analysis

Dr. K. Mosurski
Wks 1-18

Dr A Butterfield
Wks 19-24

Examination: College 3 hour Exam

Format :

Section A 6 questions from Statistics

Section B 2 questions from Numerical Analysis

5 questions for full marks at least one from section B.

Continuous assessment:

Stats : Two sets of exercises

Ex 1 over the Xmas vacation,

Ex 2 over Easter vacation.

Mark = average mark over the two exercises

The *stats* CA is not compulsory it will only count (in June) if it improves the overall mark.

Final mark = $\max(EM, 0.85 \cdot EM + 0.15 \cdot CA)$

EM = exam mark (including NA cont ass.)

CA = stats cont ass.

CA not used in supplementals.

see Dr Butterfield for Numerical Analysis CA

Aims of the Course

- To understand the basic ideas of probability and statistics
- To learn some common statistical techniques
- To gain experience in using statistical computer packages (mainly DATADESK and MINITAB)

There are two parts to this course:

Part I: Probability theory. - Michaelmas

Probability theory is the mathematics of uncertainty and randomness

Tutorials to reinforce the concepts.

Part II: Statistics (Hilary term)

Statistics applies the mathematics of probability to data involving uncertainty.

Labs to teach statistical analysis using statistical packages.

Book:

Chatfield, C, Statistics for Technology, Chapman and Hall, 1970

Examples

Example 1: Measuring the velocity of Light

Michelson Morley Velocities of
Light km/sec – 299000. 100
measurements

850	800	880	720
740	850	910	840
900	880	850	850
1070	900	870	850
930	840	840	780
850	830	840	890
950	790	850	840
980	810	840	780
980	880	840	810
880	880	840	760
1000	830	890	810
980	800	810	790
930	790	810	810
650	760	820	820
760	800	800	850
810	880	770	870
1000	880	760	870
1000	880	740	810
960	860	750	740
960	720	760	810
960	720	910	940
940	620	920	950
960	860	890	800
940	970	860	810
880	950	880	870

Minimum = 620
Maximum = 1070

Here we are trying to find out the value of an unknown quantity – we have imperfect experimental measurements.

How do we “estimate” the true value?
When we get it how precise is the estimate?

Possible estimates:

- (1) Average of the 100 values.
- (2) Mid point between min and max
- (3) Take logs of (299000+reading) average them and transform back. (Geometric average.)

... other possibilities.

Which of these is “best” – closest to the true value
How repeatable is the experiment.

To answer these need a description of how the data vary. – A model for the variation.

A Probability model for the data :

Example 2: Software testing

A completed piece of software is being tested. The number of bugs discovered in each 10 man-hours of test is recorded.

man-hours	Bugs
0 - 10	4
10-20	2
20-30	8
30-40	6
40-50	4
50-60	6
60-70	6
70-80	3
80-90	3
90-100	2

Questions:

- How bug-free is the software? i.e. How many bugs to start with ?

How much more testing before it is cost effective to release the software?

The data here are not in a straight forward format. No intuitively obvious way to estimate “initial number of bugs” – Need to construct a probability model – one of the parameters of the model will be N the initial number of bugs.

Example 3 Comparisons

A college has 1475 PCs in its public labs.

- 733 of the PCs are Apple Mac OS, at the end of the term 41 needed system reinstalls.
- 742 are PCs running Windows XP, 60 of these died during the term.

	Mac OS	Windows
Died	41	60
Survived	692	682

Is there any difference in reliability?

Fewer Macs crashed and more survived but would this happen again in the second term?

Need probability model to assess the strength of the assertion that there is a difference.

Example 4 – Salaries.

An investigation is required of computer programmers,... sorry, software engineers prospects in employment.

Where/How do you start? Statistics provides guidance on how to collect data for such an investigation.

What variables you need to consider? :

Salary, age, education, etc.

Collecting information requires effort (money) how much do you need to collect to get reasonable accurate information?

How to sample – which do you select, how selection might affect precision/accuracy.

How to perform the data analysis?