




1

Teaching Multivariate Statistics: What are Lectures For?

Nick Fieller

Department of Probability & Statistics
University of Sheffield, UK


19th May 2003

Teaching Multivariate Statistics: What Are Lectures For?
19/5/2003


2

■ Introductions

- ◆ Who?
- ◆ Where from?
- ◆ What past & future involvement in MVA teaching?
 - & to whom?
- ◆ Why?
 - What are you looking for in this workshop?




Teaching Multivariate Statistics: What Are Lectures For?
19/5/2003




3

■ Outline

- ◆ Background:
 - Sheffield's courses:— recent & past history
- ◆ Why such detailed notes?
- ◆ How is the material used?
 - What else do the students see?
- ◆ Development of content
- ◆ Current content
 - Linking themes and notable omissions




Teaching Multivariate Statistics: What Are Lectures For?
19/5/2003




4

■ Key fact:

- ◆ Sheffield's MVA course is to
 - Level 3 BSc
 - Level 3 or level 4 MMath
 - MSc
- ◆ MSc includes
 - residential students who attend lectures
 - non-residential distant learners who don't
- ◆ Single course to all groups
 - Or, rather, single *lecture series* to all




Teaching Multivariate Statistics: What Are Lectures For?
19/5/2003




5

■ Ideally should teach these groups separately

- ◆ Was the case until ~1983 (BSc vs MSc)
 - I took over the 2 courses from David Grey & Morris Walker & combined them
- ◆ & also ~1990 – 1995 when separate 'B'-Level for BSc taught by Gillian Constable
- ◆ BSc 'S'-Level & MSc continued together
 - 'B' & 'S' equivalent to BSc & MMath



Teaching Multivariate Statistics: What Are Lectures For?
19/5/2003




6


■ Some conflict

MSc ↔ MMath/BSc

- & MMath ↔ BSc
 - (level / mathematical sophistication)
- ◆ MSc take complete package of statistics courses: aim is to provide a thorough training as a professional statistician
- ◆ MMath/BSc take a pick&mix selection of all maths subjects (pure, applied & statistics)
 - Attendance is voluntary
 - & we need the student numbers




Teaching Multivariate Statistics: What Are Lectures For?
19/5/2003






7

- **Particular problem with MSc**
 - ◆ Residents attend lectures
 - ◆ Non-residents don't
 - ◆ Course should be equivalent for both
 - ◆ What are the lectures for?
 - Cannot disadvantage DLs
 - Should not waste time of RSs
 - Lectures are compulsory (Uni policy)
 - Cannot give material to only one group & not to the other
 - (close collaboration & communication between them)
- Inevitably this affects the way I teach




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




8

- How did we get into this position??

- Somebody thought it was a good idea at the time
 - ◆ But a brief history explains some things about the way this course is constructed




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




9

- **Background**
 - ◆ **The Sheffield MSc in Statistics**
 - Long established (mid 60s)
 - Broad based
 - leads to professional statistical career in any branch of application
 - » (pharmaceutical, government, industry,...)
 - Successful
 - Jobs
 - PhD




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




10

- **Why Distance Learning?**
 - ◆ New EPSRC scheme of funding prompted a forward look
 - ◆ Patterns of training changing
 - Career moves into Statistics
 - Skills updates
 - ◆ Student priorities changing
 - indebtedness



Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




11


- **The Aim:**

Set up a DL Arm to the MSc that

 - ◆ can be followed by part-time study
 - 2 or 3 years
 - ◆ will be of high quality –
 - **fully equivalent to existing MSc**
 - cannot afford to devalue our current MSc
 - ◆ can be developed and sustained with available resources
 - some help from MTP funding available




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




12

- **Organizational Consequences**
 - ◆ Adaptation of existing courses
 - Lecture courses
 - Practical/professional skills courses
 - ◆ Evolution not revolution
 - ◆ Basic delivery technology
 - WebCT
 - ◆ Distance Learning Manager
 - → Caitlin
 - ◆ Teaching & Learning Support Unit Advice
 - University expertise tapped & highly valuable




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003





13

- **Needs of Distant Learners**
 - ◆ Time, determination, equipment
 - ◆ Encouragement
 - contact with department
 - contact with other students
 - weekly reports to Caitlin
 - ◆ Route map
 - ◆ Supplements to lecture material
 - amplification
 - reinforcement: five-finger exercises
 - ◆ Feedback




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




14

- **Extras for Distant Learners**
 - ◆ Time, determination, equipment
 - ◆ Encouragement
 - contact with department
 - contact with other students
 - weekly reports to Caitlin
 - ◆ Route map
 - ◆ Supplements to lecture material
 - amplification
 - reinforcement: five-finger exercises
 - ◆ Feedback




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




15

- **Extras for Distant Learners**
 - ◆ Time, determination, equipment
 - ◆ Encouragement
 - contact with department
 - contact with other students
 - weekly reports to Caitlin
 - ◆ Route map
 - **Supplements to lecture material**
 - amplification
 - reinforcement: five-finger exercises
 - ◆ Feedback




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




16

- Obligated to extend this good teaching practice to all MSc students
 - ◆ & BSc

- Why had we not been doing this already?



Teaching Multivariate Statistics: What Are lectures For?
19/5/2003



17

- **Our Solution**
 - ◆ Residential weeks
 - ◆ Regular contact
 - ◆ Course guide
 - ◆ Discussion board
 - ◆ Extended course materials




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




18

- **Our Solution (ctd)**
 - ◆ Use web for delivery to DLs
 - Material can be corrected, updated, extended
 - ◆ Use technology supported by University
 - WebCT (Web Course Tools)
 - May not be ideal but technical support available
 - Allows us to concentrate on course content
 - ◆ Don't attempt to use Web as a teaching medium in its own right
 - Residential students taking course in parallel
 - RSs must have access to Web material (but maybe not BSc)




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003





19

- Clarification
 - ◆ **fully equivalent to existing MSc**
 - cannot afford to devalue our current MSc
- DL version of MSc course must be **same** as existing course to residents
 - ◆ Same material
 - ◆ Same exercises
 - ◆ **Same deadlines**
 - ◆ Same pace of material
 - Can't 'read notes when they like', keep in step
 - ◆ Same exams




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




20

- **Immediate Problems**
 - ◆ **Lecture notes not self-contained for DL**
 - ⇒ Expand notes to include verbal explanations & additional diagrams & computer transcripts given in lectures
 - ⇒ Provide simple *task sheets* each week
 - ◆ **Web pages for DLs useful for Residents but UGs have no access to these**
 - ⇒ I provided duplicate web pages for UGs, (data sets, solutions,)
 - some others did not do this




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




21

- **Introduced problems**
 - ◆ Danger of *over-teaching*
 - ◆ Too much expected by students
 - e.g. formal solutions expected for Task Sheets such as **verify result 2.3** (*'easy to shew that..'*)
 - ⇒ **refuse** firmly & resolutely
 - ◆ Students less tolerant of slips & misprints
 - & they can be detected!
 - no avoidance of public apologies




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




22

- **Introduced problems**
 - ◆ **What are lectures for??**
 - if web material is self-contained?
 - ⇒ Demonstrate more examples
 - ⇒ Provide summary notes for lecture display
 - *very substantial task*
 - ◆ Must keep students occupied in the lecture
 - Displaying copies of printed notes is **disaster**
 - PowerPoint slides must **add** value
 - Computer demonstrations must be extra
 However...




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




23

- **N.B.** lecture notes for complete course distributed at beginning
 - ◆ Alternative models:
 - Distribute lecture by lecture at start of lecture
 - Similarly for DLs (time release)
 - Fill-in-the-blanks notes
 - Students have to keep awake & fill in key numbers, proofs etc
 - ◆ Overall, variety in department is good
 - No single model universally preferred




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




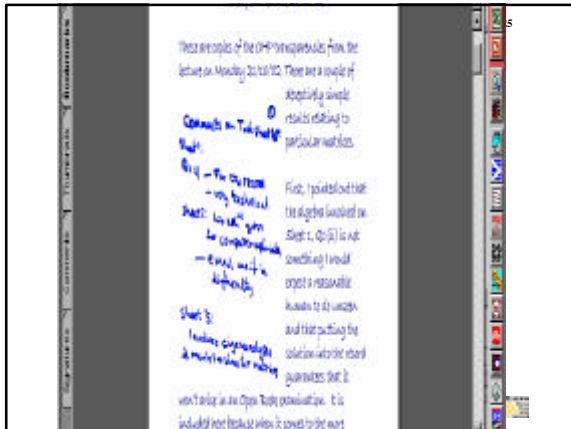
24

- **However**
 - ◆ Must not be too valuable since DLs are not at lectures
 - ◆ In lectures students will annotate notes from slides
 - ◆ Put copies of slides on web for DLs
 - but not distributed in advance
 - ◆ Sometimes write extra slides e.g. in answer to query – can scan and put into .pdf and onto web



Teaching Multivariate Statistics: What Are lectures For?
19/5/2003



26

- ◆ Tablet PC (with OCR)????
- ◆ Weekly bulletin on discussion board
 - Outline material covered
 - Examples covered
 - Key points made from examples
- ◆ Computer demonstrations ok but must be used sparingly
 - students find it easy to loose thread
 - keep computer sessions short
 - 5 to 10 minutes max,
 - perhaps 2 or 3 times per lecture
- ◆ PowerPoint animation can be used effectively – also should be used sparingly

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003

27

Discriminant Analysis

■ Results:

Eigenvalue	32.4704	0.2883	0.0000	0.0000
Proportion	0.9912	0.0088	0.0000	0.0000
Cumulative	0.9912	1.0000	1.0000	1.0000

Eigenvector	1	2	3	4
sepal-l	0.0671	-0.0005	0.2173	0.1481
sepal-w	0.1284	-0.1779	-0.2171	0.0154
petal-l	-0.1814	0.0758	-0.2461	0.0329
petal-w	-0.2331	-0.2331	0.3170	-0.2033

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003

28

Discriminant Analysis

■ Results:

Eigenvalue	32.4704	0.2883	0.0000	0.0000
Proportion	0.9912	0.0088	0.0000	0.0000
Cumulative	0.9912	1.0000	1.0000	1.0000

Eigenvector	1	2	3	4
sepal-l	0.0671	-0.0005	0.2173	0.1481
sepal-w	0.1284	-0.1779	-0.2171	0.0154
petal-l	-0.1814	0.0758	-0.2461	0.0329
petal-w	-0.2331	-0.2331	0.3170	-0.2033

Zero eigenvalues mean eigenvectors arbitrary, non-interpretable, irrelevant

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003

29

Discriminant Analysis

■ Results:

Eigenvalue	32.4704	0.2883	0.0000	0.0000
Proportion	0.9912	0.0088	0.0000	0.0000
Cumulative	0.9912	1.0000	1.0000	1.0000

Eigenvector	1	2	3	4
sepal-l	0.0671	-0.0005	0.2173	0.1481
sepal-w	0.1284	-0.1779	-0.2171	0.0154
petal-l	-0.1814	0.0758	-0.2461	0.0329
petal-w	-0.2331	-0.2331	0.3170	-0.2033

Zero eigenvalues mean eigenvectors arbitrary, non-interpretable, irrelevant

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003

Statistical Analysis of MV Data

■ Computational Note:

- ◆ Quadratic forms such as those above can be easily calculated at sight:

$$(x,y) \begin{pmatrix} a & c \\ c & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2cxy + by^2$$

$$(x,y,z) \begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} =$$

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003



Statistical Analysis of MV Data

- Examples of calculations
 - 1-sample T²-test
 - Data on head length of 1st and 2nd sons in 25 families (x₁ and x₂ respectively)
 - n=25
 - H₀: both mean head lengths are = 182
 - i.e. μ₀ = (182, 182)' and x = (x₁, x₂)'

$$T^2 = n(\bar{x} - \mu_0)'S^{-1}(\bar{x} - \mu_0)$$

Statistical Analysis of MV Data

- Data:

$$\bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = \begin{pmatrix} 185.72 \\ 183.84 \end{pmatrix}$$

$$S = \begin{pmatrix} 91.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix}$$

$$T^2 = n(\bar{x} - \mu_0)'S^{-1}(\bar{x} - \mu_0)$$

Statistical Analysis of MV Data

$$\bar{x} - \mu_0 = \begin{pmatrix} \bar{x}_1 - \mu_{01} \\ \bar{x}_2 - \mu_{02} \end{pmatrix} = \begin{pmatrix} 185.72 - 182 \\ 183.84 - 182 \end{pmatrix} = \begin{pmatrix} 3.72 \\ 1.84 \end{pmatrix}$$

n = 25, p = 2

$$T^2 = 25(3.72, 1.84)S^{-1}(3.72, 1.84)'$$

Statistical Analysis of MV Data

$$S = \begin{pmatrix} 91.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix}$$

$$|S| = 91.481 \times 96.775 - 66.875^2 = 4380.81$$

$$S^{-1} = \frac{1}{4380.81} \begin{pmatrix} 96.775 & -66.875 \\ -66.875 & 91.481 \end{pmatrix}$$

$$T^2 = 25(3.72, 1.84)S^{-1}(3.72, 1.84)'$$

Statistical Analysis of MV Data

$$T^2 = \frac{25}{4380.81} (3.72, 1.84)S^{-1}(3.72, 1.84)'$$

$$= \frac{25}{4380.81} (3.72, 1.84) \begin{pmatrix} 96.775 & -66.875 \\ -66.875 & 91.481 \end{pmatrix} \begin{pmatrix} 3.72 \\ 1.84 \end{pmatrix}$$

$$= \frac{25}{4380.81} (3.72, 1.84) \begin{pmatrix} 96.775 \times 3.72 - 66.875 \times 1.84 \\ 66.875 \times 3.72 - 91.481 \times 1.84 \end{pmatrix}$$

Statistical Analysis of MV Data

$$= \frac{25}{4380.81} (3.72, 1.84) \begin{pmatrix} 360.00 - 123.85 \\ -248.78 + 168.33 \end{pmatrix}$$

$$= \frac{25}{4380.81} (3.72, 1.84) \begin{pmatrix} 236.15 \\ -80.45 \end{pmatrix}$$

$$= \frac{25}{4380.81} (3.72 \times 236.15 - 1.84 \times 80.45)$$

$$= \frac{25}{4380.81} (878.4675 - 148.828)$$


$$= \frac{25}{4380.81} (729.6395) = 4.17$$


Statistical Analysis of MV Data


$T^2(p,n) \equiv \frac{np}{n-p+1} F_{p,n-p+1}$

$F_{p,n-p+1} = \frac{n-p+1}{np} T^2(p,n) = \frac{25-2+1}{50} 4.17 = 2.00$

$2.00 < F_{2,24}(95\%) = 3.4$ so little evidence that the mean head lengths are not both equal to 182.0




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003




38

- **Extra teaching features: Discussion Lists**
 - ◆ more open discussion available to all
 - necessary to discourage direct emails & personal queries
 - reply **only** to discussion lists
 - ◆ weekly news summaries for each course
 - material covered, queries raised
 - sometimes repeated verbally in lectures




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003


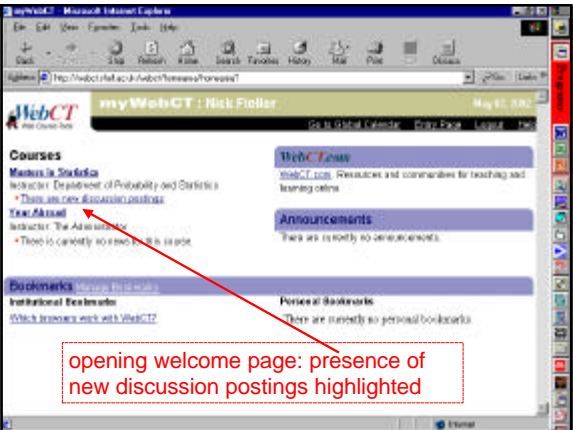


39

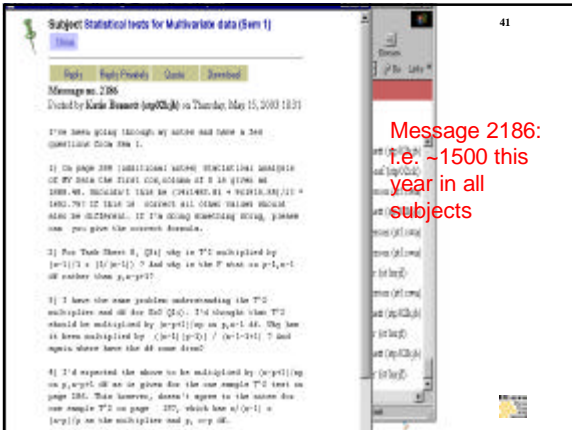
- **WebCT**
 - ◆ Decision made early to use **WebCT** as medium for course delivery to DLs
- Web based system for
 - ◆ organizing/storing files e.g.
 - web pages
 - postscript or pdf documents
 - data sets
 - ◆ discussion lists, email,.....
 - ◆ submitting assignments and maintaining records
 - ◆ &c., &c.
- Teaching & Learning Support Unit can provide technical help on WebCT
 - We discovered it is not bug free




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003


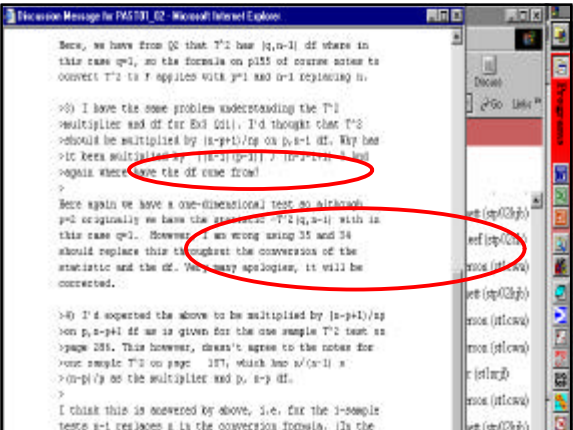
41



Message 2186:
i.e. ~1500 this
year in all
subjects




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003






49

- **Current content**
- NB 20 contact hours for lectures & examples**
- ◆ Guidelines are to spend further 180 hours self-study
 - task sheets
 - exercises
 - lecture notes
 - exam revision
- ◆ Includes 1 reading week with no lectures
- ◆ i.e. ~4 hours extra per week + revision
 - rarely done




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




50

- **General themes:**
 - ◆ Dimensionality reduction
 - ◆ Understanding variability
 - ◆ Interpretation of data
 - ◆ Communication
- i.e. same as in all applied statistics
 - (except the first, perhaps)
- **No substitute for experience**
 - Perhaps more so than in other areas of applied statistics




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




51

- **Dimensionality reduction**
- ◆ **Data display (1 lecture)**
 - Most 'fancy' techniques are generally useless
 - No substitute for scatterplots
 - Need to choose a few dimensions to display
- ◆ **Principal component analysis (~4+)**
 - Introduces elegant maths
 - Optimization via eigenanalysis
 - Proportional amounts of variation 'explained'
 - Interpretation of loadings & hence of PCs
 - no substitute for experience
 - More examples the better, explain in detail




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




52

- ◆ **Understanding PCA is central to course**
 - Many other techniques are 'by analogy'
 - Interpretations of loadings central to explaining results of statistical tests T^2 -tests, MANOVA etc.
- ◆ **Themes picked up later in discriminant analysis and union-intersection tests**
 - i.e. interpretations of loadings, scree-graphs

But meanwhile an interlude on




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




53

- **Multidimensional scaling**
- ◆ 'sell' as the reverse/dual of PCA
- ◆ nice examples
- ◆ unusual problems
- ◆ develops
 - interpretation
 - explanation
 - communication
- ◆ **Key advantage in placing here is allows time for assimilation of PCA**
 - exercises
 - discussion etc.
 - in fact reinforcement returned to several times




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




54

- **Linear Discriminant Analysis**
- ◆ Initially as informal dimensionality reduction technique and informal classification
 - Allows use of eigenanalysis for optimization techniques again
 - Reinforces interpretations of loadings
 - Informal use of scree-graphs
 - Examples to compare PCA & LDA
 - Highlight extra insight gained from LDA
 - Extra information on group structure gives extra interpretation of analysis




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003





55

- **Statistical Analysis**
 - ◆ Basic theory of multivariate normal
 - ◆ Construction of LRTs
 - ◆ Application of T^2 -tests (1 & 2 sample)
 - Do they need to know that T^2 is a LRT??
- **Doubts here**
 - Shortened considerably over recent years
 - more emphasis on computer implementation and interpretation
 - ◆ Do we still need ANY background theory?




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




56

- **Statistical Analysis (c^{td})**
 - ◆ Union intersection tests
 - (i.e. project data into one-dimension & maximize value of test-statistic over choice of projection, c.f. PCA & maximizing variance over projection)
 - ◆ Follows spirit of dimensionality reduction
 - uses eigen – optimization technique again
 - uses interpretation of loadings
 - allows interpretation of MANOVA and so ties in with LDA
 - has drawback of not being universally applicable, unlike LRTs




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




57

- **Students dislike this section most**
 - ◆ Especially LRTs
 - Especially since most do not see LRTs in **any** other context
 - ◆ However, LRTs only way of constructing certain types of test
 - e.g. $H_0: \Sigma = \lambda \Sigma_0$ when $x \sim N_p(\mu, \Sigma)$ for some λ and specified Σ_0
 - ◆ Dislike idea of mere recipes for T^2 -tests
 - ◆ Techniques definitely needed in research environments
 - both academic and industrial research




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




58

- **Statistical Discriminant Analysis**
 - ◆ Attempt to bring together informal LDA section and statistical ideas
 - ◆ Introduces and uses more modern computer intensive resampling methods
 - jackknife
 - bootstrapping
 - permutation tests
 - ◆ This will be extended with full adoption of S+
 - (but some illustrations already in appendices)




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




59

- **Notable Omissions**
 - ◆ Bayesian methods (e.g. in discrimination)
 - Most UGs have no Bayesian upbringing (even though all MSc do)
 - ◆ Correspondence Analysis
 - ◆ Canonical Correlation Analysis
 - ◆ Cluster Analysis
 - ◆ Tree-based methods
 - Classification trees and regression trees
 - ◆ Neural networks & SVMs
- **Life is short!**




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




60

- ◆ Some of these topics are provided briefly in appendices with R code
 - Cannibalised from courses given in University of Tampere, Finland (and to industry under SSU)
- ◆ Aim is to provide pointers for MSc students who need to use them
 - e.g. coursework assessed project typically requires use of at least one of these methods
 - most students use clustering and scaling methods for language similarities project
 - Many will use these in their dissertations




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003





61

Practicalities & Technicalities

- **Agenda**
 - ◆ Working Group Session
 - ◆ Discussion
 - ◆ Unexpected difficulties in teaching
 - practicalities
 - technicalities




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003




62

Practicalities & Technicalities

- **Working Group Session**
 - What split between math theory and computer-package based instruction?
 - answers dependent on target audience so identify different targets
 - What theoretical topics to include?
 - What practical topics to include?
 - Should any MVA be taught on service courses for non-specialists?
 - A O B?




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003




63

Unexpected difficulties in teaching practicalities

- ◆ Interpreting scatterplots
 - Relating points to numbers
 - Relating points to objects
 - Describing characteristics of objects giving rise to points in different parts of the plots
- ◆ Result of never drawing a scatterplot by hand
 - (c.f. Normal Q-Q plots. Histograms,




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003




64

Solutions??

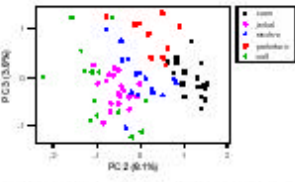
- ◆ Unrealistic to require drawing by hand
 - (we need the student numbers)
- ◆ Constant and repetitive emphasis
 - e.g. P10 of Exercises & Solutions
 - Bryan Manly's Thai Dogs data
 - Prehistoric and Cuons in top right hand corner of plot on PCs 2 & 3:
 - What will these dogs look like in relation to the others???




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003




65



The plot on PCs 2 & 3 displays about 12% of the information. It separates the prehistoric (& the cuons) in the top right hand corner. To appear in the top r.h. corner cases have to have large values for those variables with positive coefficients on **both** PCs 2 & 3 and small values for those with negative coefficients on PCs 2 & 3, i.e. large values for X_2 and X_4 (ignoring any variable with a very small coefficient, even if positive) and small values for X_1 and X_6 , i.e. prehistoric dogs and cuons have short 'chunky' mandibles.




Teaching Multivariate Statistics: What Are lectures For? 19/5/2003




66

Interpretations of loadings

- ◆ Surprisingly difficult idea for some students
- ◆ Many will memorize that first PC 'always' reflects overall size of objects
 - even when data are on questionnaire analysis and then difficulty in 2nd & 3rd PCs
- ◆ White leghorn chicken data is best example I have
 - but don't have raw data on file to produce scatterplots
 - Google??



Teaching Multivariate Statistics: What Are lectures For? 19/5/2003




Dimensionality Reduction

• Values rounded to 2 digits

Original variable	Principal Components					
	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆
x ₁ skull l.	0.35	0.53	0.76	-0.04	0.02	0.00
x ₂ skull b.	0.33	0.70	-0.64	0.00	0.00	0.03
x ₃ humerus	0.44	-0.19	-0.05	0.53	0.18	0.67
x ₄ ulna	0.44	-0.25	0.02	0.48	-0.15	-0.71
x ₅ femur	0.44	-0.28	-0.06	-0.50	0.65	-0.13
x ₆ tibia	0.44	-0.22	-0.05	-0.48	-0.69	0.17

All positive and approx equal

Dimensionality Reduction

• Values rounded to 1 digit

Original variable	Principal Components						
	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	
x ₁ skull l.	0.4	0.6	0.7	0	0	0	skull
x ₂ skull b.	0.4	0.6	-0.7	0	0	0	
x ₃ humerus	0.4	-0.2	0	0.5	0	0.7	wing
x ₄ ulna	0.4	-0.2	0	0.5	0	-0.7	
x ₅ femur	0.4	-0.2	0	-0.5	0.6	0	leg
x ₆ tibia	0.4	-0.2	0	-0.5	-0.6	0	

Dimensionality Reduction

• Values rounded to 0 digits

Original variable	Principal Components						
	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	
x ₁ skull l.	+	+	+				skull
x ₂ skull b.	+	+	-				
x ₃ humerus	+	-		+		+	wing
x ₄ ulna	+	-		+		-	
x ₅ femur	+	-		-	+		leg
x ₆ tibia	+	-		-	-		

Dimensionality Reduction

- ◆ 1st PC — sum of all components
 - Measures **size**
- ◆ 2nd PC — (skull) vs (wing & leg)
 - Measures **overall body shape**
 - (big heads & little bodies) vs (little heads & big bodies)
- ◆ 3rd PC — skull length vs width
 - Measures **shape of skull**
- ◆ 4th PC — wings vs legs
 - Measures **body shape**
- ◆ 5th & 6th measure wing & leg shape

PCA of leghorn on correlation matrix

Original variable	Principal Components					
	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆
x ₁ skull l.	-0.35	-0.40	-0.85	-0.05	0.01	0.03
x ₂ skull b.	-0.29	-0.81	0.50	-0.02	0.01	0.04
x ₃ humerus	-0.44	0.26	0.11	-0.50	0.60	0.33
x ₄ ulna	-0.45	0.20	0.10	-0.47	-0.60	-0.41
x ₅ femur	-0.45	0.16	0.10	0.50	0.37	-0.58
x ₆ tibia	-0.45	0.21	0.07	0.47	-0.38	0.62
eigenvalue	4.46	0.78	0.46	0.17	0.08	0.05

Dimensionality Reduction

• Correlation matrix: values rounded to 0 digits

Original variable	Principal Components						
	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	
x ₁ skull l.	-	-	-				skull
x ₂ skull b.	-	-	+				
x ₃ humerus	-	+		-	-	+	wing
x ₄ ulna	-	+		-	+	-	
x ₅ femur	-	+		+	+	-	leg
x ₆ tibia	-	+		+	-	+	



Dimensionality Reduction 73

- Covariance matrix: values rounded to **0 digits**

Original	Principal Components						
variable	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	
x ₁ skull l.	+	+	+				skull
x ₂ skull b.	+	+	-				
x ₃ humerus	+	-		+		+	wing
x ₄ ulna	+			+		-	
x ₅ femur	+	-		-	+		leg
x ₆ tibia	+	-		-	-		

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003

Dimensionality Reduction 74

- ◆ Most PCs same in structure
 - Except low order ones
- ◆ Some have changed sign
 - signs are arbitrary
 - can multiply all coefficients by -1
 - relative +/- signs are important for contrasts
- ◆ Sum of all eigenvalues is 6 (= p)
 - diagonal terms of correlation matrix are all = 1

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003

Difficulties in Technicalities 75

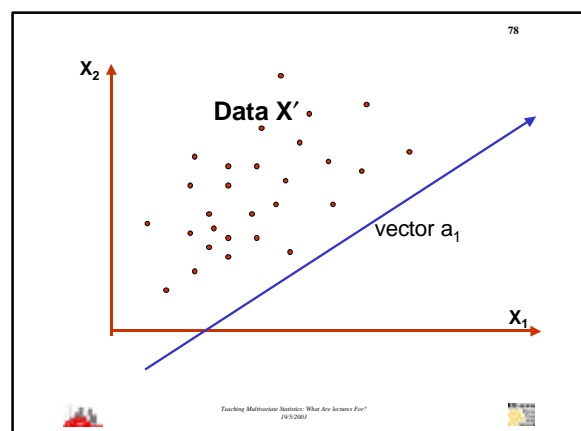
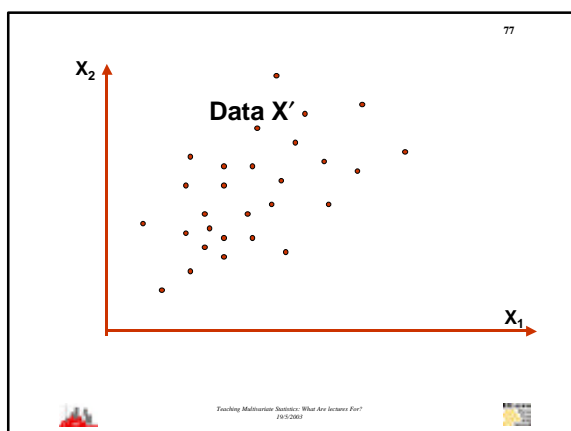
- ◆ Anything with likelihoods
 - Not included in exam for past 3 years
 - only Erasmus students answered question in previous couple of years
 - Statistics exams should not be the vehicle for discovering that students cannot do mathematics
 - Not sustainable to carry on teaching this for ever

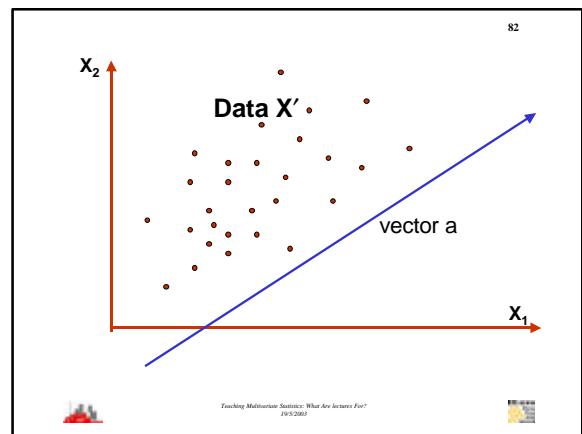
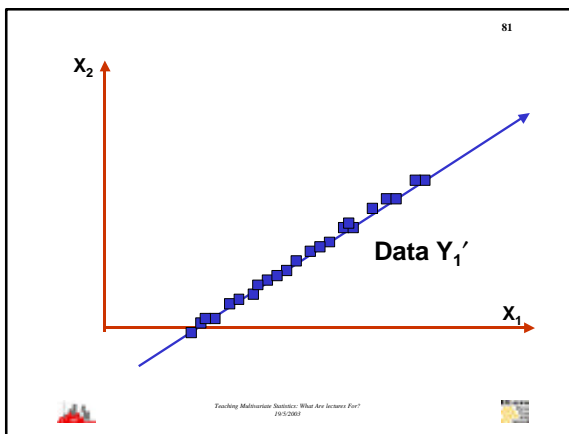
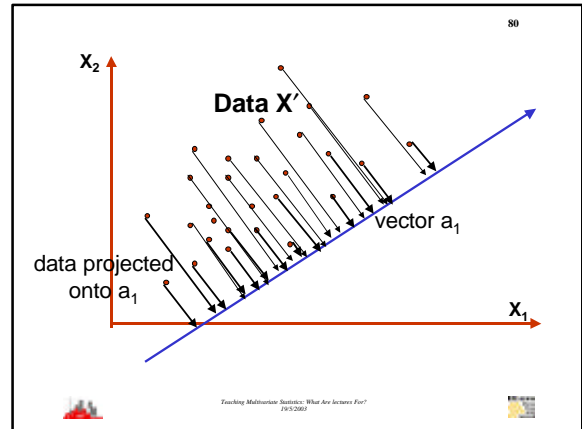
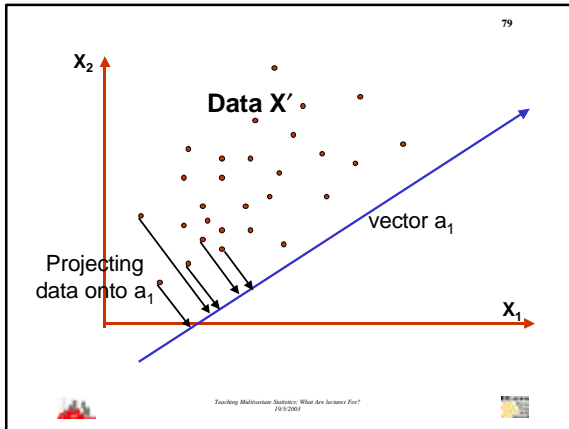
Teaching Multivariate Statistics: What Are lectures For? 19/5/2003

Geometrical idea of projections 76

- ◆ i.e. if X' is an $n \times p$ data matrix and β is a p -vector then $Y' = X'\beta$ is the projection of the data onto the vector β
- ◆ ok at recognizing Y' is $n \times 1$
 - idea that Y' is a view of the same data in 1-dim

Teaching Multivariate Statistics: What Are lectures For? 19/5/2003







85

Assessment

- **Agenda:**
 - ◆ Working group session
 - ◆ Discussion
 - ◆ A O B




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




86

- **Working Groups: Assessment**
 - What are the objectives of assessment?
 - What ways can these be achieved?
 - Does it depend on target group?
 - Should account be taken of custom & practice in other modules?
 - How can assessment be fair to all?
 - A O B??




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




87

Summary & Conclusions & Future

- **Topics for discussion**
 -
 -
 -
- ◆ **Future:**
 - What do you think will be key changes?
 - By when?
 - How can we improve knowledge & teaching?
 -




Teaching Multivariate Statistics: What Are lectures For?
19/5/2003




88

- **Aside:**
Teaching of Statistic in the Computer Age
 - Proceedings of 6th ISI Round Table in Canberra
 - ◆ Only 1 paper on teaching statistics at University level (Terry Speed)
 - ◆ Reviews computer-intensive methods
 - including cross-validation, jackknifing
 - (e.g. in estimating discrimination rates)
 - Advocates teaching these at University level

“University statistics teachers should rise to the new challenge.....” **1984**



Teaching Multivariate Statistics: What Are lectures For?
19/5/2003



89



Teaching Multivariate Statistics: What Are lectures For?
19/5/2003



90

<http://www.shef.ac.uk/nickfieller>



Teaching Multivariate Statistics: What Are lectures For?
19/5/2003