Teaching portfolio

Leslie Foldager[1,2]

[1]Management and Modelling, Department of Animal and Veterinary Sciences, Faculty of Technical Sciences, Aarhus University, AU Viborg – Research Centre Foulum, Blichers Allé 20, DK8830 Tjele, Denmark. Email: leslie@anivet.au.dk

[2]Bioinformatics Research Centre, Department of Molecular Biology and Genetics, Faculty of Natural Sciences, Aarhus University, Aarhus, Denmark

24th October 2024

Contents

1 Description and documentation

1.1 Teaching carried out

- 2023, 16 Nov: Guest lecture (1 hour): Power and sample size by simulations using R, as part of Mixed modelling in R workshop (14–16 Nov 2023) held by Gavin Simpson, Dept. of Animal and Veterinary Sciences, Aarhus University, Tjele, Denmark. Level: PhD.
- 2019, 23 Sep: Guest lecture (1 hour) in statistics for bachelor students in agriculture from Dalum Landbrugsskole, Denmark. Held by Ib Sillebak Kristensen, Dept. of Agroecology, Aarhus University, Tjele, Denmark.
- 2017, fall (2 Oct – 18 Dec): Lecturer on ad–hoc PhD course: Selected Topics from Generalised Linear Models and Survival Analysis, 5 ETCS, Department of Animal Science, Aarhus University, Tjele, Denmark. Level: PhD course.
- 2012, 19–30 Nov: Guest lecturer of 2 weeks lectures (4 hours/week) and exercises (3 hours/week) on Association Mapping in the master course Statistical Methods in Bioinformatics held by associate professor Asger Hobolt, Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. Level: MSc course.
- 2008, Jan–Feb: Internal course: Statistics for biomedical laboratory technician: a brush-up (Danish title: Statistik for Bioanalytikere: Et Brush-up), 3 half-hour lectures, Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark.
- 2006, fall: Internal course: Crash course on R, fall 2006, Centre for Basic Psychiatric Research, Aarhus University Hospital, Risskov, Denmark. Participants were post. docs and PhD students from the fields of biology and molecular biology.
- 2002, fall (10 Sep – 8 Nov): Lecturer on Biostatistics Course - Fall 2002: Statistical Analysis of Biological Problems using Linear and Nonlinear Models, 8 full days, Biometry Research Unit, Danish Institute of Agricultural Sciences, Foulum, Denmark. Level: postgraduate. Participants: research assistants, PhD students, scientists and senior scientists with various MSc backgrounds (veterinary, biology, agriculture, chemistry-biotechnology, biochemical engineering).
- 2001, fall (5 Sep – 5 Dec): Lecturer on Biostatistics Course - Fall 2001: Statistical analysis of biological problems using non-linear models (Danish title: Kursus i Biostatistik - Efterår 2001: Statistisk Analyse af Biologiske Problemstillinger med Ikke-lineære Modeller), 8 full days, Biometry Research Unit, Danish Institute of Agricultural Sciences, Foulum, Denmark. Level: postgraduate. Participants: scientists and senior scientists with MSc background from biology and agriculture.
- 1995–1997: Teaching assistant on various undergraduate statistical and probability theoretical courses, in total six one-semester courses, Department of Theoretical Statistics, University of Aarhus, Aarhus, Denmark.

1.2 Examinations carried out

- The ad–hoc PhD course in 2017 was completed by the participant handing in a "ready for submission" manuscript and a short report with the R and SAS scripts used for analyses presented in the manuscript, followed by a short oral examination based on questions/remarks to the manuscript.
- The two courses held at the Danish Institute of Agricultural Sciences, Foulum, were completed by an exercise handed in for approval and presented orally but without grading.

1.3 Experience of supervision

- 2015–2024: Involved in supervision of students (PhD, master, bachelor) and researchers from the Faculty of Technical Sciences (2015–2020: Faculty of Science and Technology), Aarhus University, as part of the statistical consultancy function at Dept. of Animal and Veterinary Sciences (2015–2022: Dept. of Animal Science).
- 2006–2024 Supervision of students (PhD) and researchers doing somatic research projects as part of sideline statistical consultancy.
- 2004–2012: Involved in supervision of a number of students (PhD, master, bachelor, research sabbatical) attached to from the Unit for Psychiatric Research, Aalborg Psychiatric Hosptial, Aarhus University Hospital, Aalborg.
- 2002–2014: Involved in supervision of a number of students (PhD, master, bachelor, research sabbatical) and researchers from the Faculty of Health, Aarhus University, attached to the Centre for Psychiatric Research (2013–2014: Translational Neuropsychiatry Unit), Aarhus University Hospital, Risskov.

1.4 Courses completed in university pedagogics or other education courses

In 1995 I followed a short course in teaching performance and pedagogics for teaching assistants.

1.5 Experience of teaching teams, supervision by colleagues etc.

The course held in 2002 at the Danish Institute of Agricultural Sciences, Foulum, was in collaboration with a senior colleague.

1.6 Experience of direction of studies and development of degree programmes, including postgraduate teaching and continuing and further education

Nothing to report.

1.7 Contributions to the development of subject areas, subjects or disciplines

Nothing to report.

1.8 Contributions to textbooks or teaching material

Nothing except from exercises and slide handouts, see below.

1.9 Other experience of teaching and university pedagogics

I have 10 years of experience (2004–2014) teaching/instructing gymnastics, football and handball - both children, adolescents and adults.

1.10 Examples of teaching plans, teaching material used and guidelines

As examples of teaching/course plans I have attached the course web-page from the course that I held in 2002 at the Danish Institute of Agricultural Sciences, Foulum.

Lecture notes from one of the lectures in 2002 (a lecture on growth curve models) are given as an example of teaching material. Moreover, two exercises (a PC exercise and a compulsory exercise) are attached and exemplifies the material used in the 2002 course.

2 Evaluations

2.1 Evaluations by students

The evaluation of the ad–hoc course in 2017, and two courses that I held at the Danish Institute of Agricultural Sciences, Foulum, in 2001 and 2002 are attached. They are mainly written in Danish though.

2.2 Statements by directors of studies, heads of institutedepartment or course managers, e.g. in connection with educational development

Nothing to report.

3 Attached documents

3.1 Example of course plan: Biostatistics Course - Fall 2002

Ministry of Food, Agriculture and Fisheries
Danish Institute of Agricultural Sciences
Biometry Research Unit

Homepage of Biostatistics Course - Fall 2002

The Biometry Research Unit gives a course in biostatistics in the fall 2002. The subtitle of the course is Statistical Analysis of Biological Problems using Linear and Nonlinear Models. The target group is scientists, PhD-students and others using statistics in their everyday life.

The pages was updated: Nov 8, 2002

The contents of this Homepage:

Aim

In agricultural research, the most frequently applied statistical tool is linear models. Moreover, when we start to work with other statistical tools, linear models serves as the natural basis. In practice, however, linear models are often inadequate for the problems on hand - an example (in Danish) is given here.

During the course we will go through (or repeat) linear models in a way that puts regression, variance, covariance analysis, and a number of other frequently applied models into a coherent framework. In connection with this, we will also repeat a number of elementary statistical concepts. The practical applications of these topics are illustrated using *SAS*. As illustrated by the example an attractive alternative to the linear models are the nonlinear models. With the appearance of faster and faster computers this is becoming even more obvious.

Afterwards the participant have gained

- a thorough review of elementary statistical concepts including confidence with the Linear Normal Models.
- familiarity with the concept and practical use of nonlinear models, e.g. models for growth curves, logistic regression, dose-response experiments, and models for enzymatic kinetic.
- experience with practical application of these models during exercises.

Furthermore, the course serves as a good basis for participation in courses dealing with the more advanced statistical methods, e.g. the Biometry Research Units PhD-courses.

Prerequisites, workload, and evaluation

- The participants are expected to have a knowledge of statistics corresponding to Skovgaard, Stryhn & Rudemo, *Basal Biostatistik*, Del 1, KVL 1998, DSR Forlag.
- The typical workload is 5-6 hours of preparation for each day plus a bit more for each of the two compulsory exercises. In addition, of course, the time spend on the concluding exercise.
- The course is evaluated by a slightly larger concluding exercise that have to be returned for approval to the lectures and presented orally to the other participants. Afterwards, on the condition that the compulsory and concluding exercises are approved, the Biometry Research Unit issues a course certificate.
- It is assumed that the participants agree that exercises must be returned and approved.

Registration and course fee

There are still vacant seats - here is the application form (in Danish). Due to the ordering of course material registration should be done as soon as possible. The course fee is kr. 9775,- and includes the course material.

Time and place

The course is held at Danish Institute of Agricultural Sciences, Research Centre Foulum, and consist of 8 full days from 9:00 until 15:45 - the dates are given below. The first two days, however, will be held in 'Mødelokale 1' at Agro Business Park (nearby Research Centre Foulum) while the other days will be in the 'Miniauditorium'.

List of topics

- Elementary statistical concepts and linear normal models
  - Exploratory analysis
  - The normal distribution
  - Mean value, variance, confidence limits, basic calculations
  - Model building, estimation, hypothesis testing, simplifying models, model assessment
  - Analysis of variance, regression analysis, comparing regression lines, contrasts
  - Introduction to the statistical software package R.
  - proc univariate, proc summary, proc gplot, proc gchart, proc means, proc ttest
- Nonlinear models
  - Models derived from biological /physical concepts or relationships
  - Growth curves, logistic regression, dose-response experiments, enzymatic kinetic
  - Comparing models / selection of the model
  - Correlated observations
  - proc nlin

Course description

An outline of the course is given below. It may be split into three blocks (but there may be some overlaps in the timetable). Exercises are given throughout the course, partly as hands-on PC exercises, to help the theory fall into place. We will use *SAS* for many of the analyses but we like to point out that this is not a *SAS* course.

We will show how another statistical software (freeware) package, namely *R*, may be used as a plotting tool. Actually, we might have used *R* for all the analyses too! is worth knowing without applications.

Block 1 [10.-11. September]

The primary aim is to **refresh theory** and ensure a common point of reference. Most of these topics are assumed to be well-known, see Prerequisites, workload, and evaluation.

- Refresh theory and ensure a common point of reference
- Exploratory analysis / Preliminary Investigations
- Normal Data
- Mean value, variance, confidence limits, basic calculations
- Setting up a model, estimation, hypothesis testing, simplifying models
- The method of least squares
- The Likelihood Method
- Transformation of data
- Linear Regression
- Model Assessment Using Residuals
- Software: *SAS* and *R*
- `proc univariate`, `proc summary`, `proc gplot`, `proc gchart`, `proc means`, `proc ttest`

Block 2 [26.-27. September and 10. October]

The Linear Normal Model. Many of the frequently applied statistical models may be put into a coherent framework - the linear normal models. In linear models the response and the explanatory variables are, in some sense, connected linearly. This, on the other hand, does not mean that the response curve necessarily is linear, see the example.

- The Linear Normal Model.
- Linear Algebra (matrix algebra)
- Regression Analysis
- One-way Analysis of Variance
- Comparison of Regression Lines
- Two-way Analysis of Variance
- Estimability and Contrasts
- Correlation and Covariance
- Analysis of covariance
- Designing experiments
- `proc glm`

Block 3 [10.-11. October and 7.-8. November]

Nonlinear models. The linear models does not always give an adequate description - consider e.g. growth or decay curves where we often observe exponential growth or decay, maybe followed by a levelling out. Thus, it is apparent that other kinds of models are needed. In this context, the nonlinear models serves as an attractive alternative to the linear models.

- Nonlinear models.
- Models derived from biological /physical concepts or relationships
- Growth curves, logistic regression, dose-response experiments, enzymatic kinetic
- Comparing models / selection of the model
- Correlated observations

- Convergence, initial values, assessing the fit, estimation, inference
- Review of the course topics
- Concluding exercise
- `proc nlin`

Day 1: 10. September
Elementary statistical concepts, exploratory analysis, normal data, linear regression, 1. compulsory exercise. [Details]

Day 2: 11. September
Linear regression in detail; estimation, model assessment, confidence limits, hypothesis testing. More than two samples. [Details]

Day 3: 26. September
Linear normal models, linear algebra, designing experiments. [Details]

Day 4: 27. September
Linear normal models, contrasts, correlation, two-way ANOVA, 2. compulsory exercise. [Details]

Day 5: 10. October
Linear normal models, analysis of covariance, statistical concept revisited, introduction to nonlinear models. [Details]

Day 6: 11. October
Nonlinear models; introduction, examples, practical considerations. [Details]

Day 7: 7. November
Nonlinear models; correlated observations, inference, comparing models, logistic regression. [Details]

Day 8: 8. November
Presentation of the participants concluding exercise, review of the course topics, evaluation. [Details]

Course material

The course material will consist of:

1. Preben Blæsild and Jørgen Granfeldt (2002), Statistics with Applications in Biology and Geology. Referred to as BG, included in the course charge, and handed out the first day.
2. Lecture notes prepared by the lectures. All files (will later appear in a table below) ending with ".pdf" are to be read using Adobe Acrobat Reader. This programme may, if necessary, be downloaded from www.adobe.com.
3. The lecture notes will be available both as 'full screen' version (the version used for the presentation) and as 'compact' versions containing 8 full screen pages per page (the version you probably like to print).
4. These lecture notes will not be handed out but you are very welcome to print you own 'hard-copy'. We intend to have these notes ready no later than the day before the relevant lectures takes place. We retain the rights of modifying these lecture notes even after the first appearance at the course homepage. To ease your ability of seeing whether you have the right file or not a date and time is given after the 'File' column.
5. We are going to use the statistical software package *R* and *SAS*. *R* is available from www.r-project.org as Free Software under the terms of the Free Software Foundation's GNU General Public License.
6. To install packages for *R* (e.g. the `ash` package, see below) do the following:
   - Start the *R* program.
   - Choose the menu `Packages` → Install package from CRAN …
   - Select `ash` and press the OK button.

Note this only has to be done once (unless you install a new version of *R* or re-install the old version).

7. The folder structure with the data, *SAS* and *R* files referred to in the exercises can be accessed here: BiostatCourse/

   You may set the folder that *R* uses by the function `setwd('folder specifikation')` as can be seen from the SetFolder.R file (date/time: 13-09-02 10.40). If you save SetFolder.R to the folder of the *R* program - typically C:\Programmer\R\rw1051 - you can just write `source('SetFolder.R')` in the *R* console.

Day	Topic	File	Date and time
1	Introduction to *R* (Venables & Smith, 2002)	[pdf]	04-09-02 8.28
	1. compulsory exercise	[pdf]	09-09-02 9.37
	• *SAS* program with data	oyster.sas	13-09-02 9.37
	• data for import in *R*	oyster.txt	13-09-02 9.31
	• *SAS* - solution	oysterSol.sas	10-10-02 00.26
	• *R* program for making some plots	oyster.R	10-10-02 00.31
	PC exercise 1	[pdf]	09-09-02 9.38
	• *SAS* program with data	PCexercise1b.sas	13-09-02 9.38
	• *R* program that read data	PCexercise1b.R	13-09-02 10.42
	• data for import in *R*	PCexercise1b.txt	13-09-02 9.34
	Lecture notes: Introduction to Linear and Non-linear Models	[compact] [full screen]	09-09-02 12.43
	• *SAS* program for the examples	CourseIntro.sas	11-09-02 20.27
	• *R* program for the examples	CourseIntro.R	11-09-02 20.27
	• text file with the tree data	tree.txt	02-09-02

			14.56
	• text file with the carcass grade data	CarcassGrade.txt	04-09-02 12.30
	• text file with the tooth growth data	ToothGrowth.txt	03-09-02 16.16
	• text file with the concrete data	concrete.txt	05-09-02 9.41
	• the DoBy.R file - `ash` package required	DoBy.R	11-09-02 20.27
	• the ash.zip file	ash.zip	13-09-02 09.10
	• the DoByExample.R file - try `source("DoByExample.R")`	DoByExample.R	04-09-02 11.35
	• text file with the data used by DoByExample.R	DietOxData.txt	29-05-02 08.24
	Lecture notes: Introduction to *R*	[compact] [full screen]	11-09-02 20.27
	Lecture notes: The One Sample Problem	[compact] [full screen]	13-09-02 0.36
	Lecture notes: Introduction to 1. compulsory exercise		
	and Writing statistical reports - an example	[compact] [full screen]	13-09-02 1.02
	Example: Writing statistical reports - an example	[pdf]	09-09-02 9.36
	• *SAS* program with data	spinach.sas	13-09-02 9.37
	• *SAS* program giving the results reported	spinachSol.sas	13-09-02 9.39
	• text file with spinach data for import in R	spinach.txt	13-09-02 9.44
	• *R* program (not very well structured) giving the results reported	spinach.R	13-09-02 10.46
2	PC exercise 2	[pdf]	10-09-02 22.45

	• *SAS* program that load data	PCexercise2a.sas	13-09-02 9.35
	• *SAS* program - solution	PCexercise2aSol.sas	13-09-02 9.40
	• *R* program - solution	PCexercise2aSol.R	13-09-02 10.41
	• data from BG Example 3.4 for import in *SAS* or *R*	BGexample3.4.dat	16-07-01 14.36
	• *SAS* program with data	PCexercise2b.sas	13-09-02 9.41
	• *SAS* program - solution	PCexercise2bSol.sas	13-09-02 9.42
	• *R* program - solution	PCexercise2bSol.R	13-09-02 10.41
	• data for import in R	PCexercise2b.txt	13-09-02 9.41
	PC exercises 3 and 4	[pdf]	10-09-02 23.31
	• *SAS* program with spinach data	PCexercise3.sas	13-09-02 9.44
	• spinach data for import in R	spinach.txt	13-09-02 9.44
	• *SAS* program - solutions	spinachSol.sas	13-09-02 9.39
	• *R* program - solutions	spinach.R	13-09-02 10.46
	• *SAS* program with milk production data	PCexercise4.sas	13-09-02 9.44
	• milk production data for import in R	milk.txt	13-09-02 9.44
	Lecture notes: An Introduction: Regression Analysis	[compact] [full screen]	09-09-02 9.12
	Lecture: The 'history file' from the session with *R* examples	MyRsession.txt	11-09-02 20.27
	Lecture notes: Two or more samples	[compact] [full screen]	13-09-02 0.43
			13-09-

	Lecture notes: Another look at the regression problem	[compact] [full screen]	02 1.24
3	Basic concepts from Linear Algebra	[compact] [full screen]	01-10-02 12:45e
	PC exercises in linear algebra	[pdf]	23-09-02 14:30
	• Solutions	[pdf]	01-10-02 12:45
	• *R* program - solutions	LinAlgExerciseSol.R	25-09-02 23.44
	The Linear Normal Model (Leslie)	[compact] [full screen]	25-09-02 10:19
	• *SAS* programs	LNMinPractice.sas	30-09-02 00.16
	• *R* programs	LNMinPractice.R	30-09-02 00.17
	PC exercise 5	[pdf]	26-09-02 00:21
	• text file with the tree data	tree.txt	02-09-02 14.56
	• *R* program - solutions	PCexercise5Sol.R	30-09-02 00.19
4	Hypothesis testing	[compact] [full screen]	23-09-02 14:30
	PC exercises on hypothesis testing/power calculation	[pdf]	26-09-02 00:17
	• *R* programs	Powersim.R	26-09-02 00.07
	Analysis of variance (ANOVA)	[compact] [full screen]	27-09-02 01:10

	• *SAS* programs	ANOVAinPractice.sas	30-09-02 00.17
	• Data from BG example 3.3 (soya beans)	BGexample3.3.dat	24-09-02 00.33
	Linear Normal Models (Søren) - Parameterisations, estimability and contrasts (REVISED VERSION)	[compact] [full screen]	01-10-02 12:45
	2. compulsory exercise	[pdf]	30-09-02 12.48
	• *SAS* data set for 2. compulsory exercise	evitbiop.sas7bdat	25-09-02 12.19
	• *SAS* program for loading data	CompExer2.sas	30-09-02 13.11
	• *SAS* program - solution	CompExer2Sol.sas	25-10-02 09.50
	• *R* program - solution	CompExer2Sol.R	25-10-02 09.48
	Basic Statistical Concepts (additional notes supplied by Søren)	[compact] [full screen]	02-10-02 07:09
5	Introduction to Nonlinear Models	[compact] [full screen]	08-10-02 10:32
	• *R* programs for the examples	NonLinExamples.R	09-10-02 21.39
	• *SAS* programs for the examples	[NonLinEx1.sas] [NonLinEx2.sas]	08-10-02 10.31
	• Data for the examples	[growth.txt] [fertilizer.txt]	08-10-02 11.02
	PC exercise 6	[pdf]	09-10-02 23:26
	• *SAS* program with the beans data	PCexercise6.sas	09-10-02

			21.23
	• text file with the beans data	beans.txt	09-10-02 15.52
	• *R* program - solutions	PCexercise6Sol.R	13-10-02 21.12
	• *SAS* program - solutions	PCexercise6Sol.sas	13-10-02 21.13
	Vitamin E Case Study	[compact] [full screen]	10-10-02 08:30 10:32
	• Programs for the case study	[EvitCaseStydy.sas] [EvitCaseStudy.R]	10-10-02 08:38
	• The DoBy R-package	[doby_1.000.zip]	10-10-02 08:30
	• Data for the Vitamin E case study	evitbiop.csv	08-10-02 22.07
6	Growth Curve Models	[compact] [full screen]	09-10-02 00:52
	• *R* programs for the example	GrowthExample.R	13-10-02 21.11
	• Data for the example	boars.txt	08-10-02 22.07
	R and *SAS* - how to do session	Rhistory.txt	11-10-02 13.09
	PC exercise 7	[pdf]	13-10-02 20:25
	• *SAS* program with the boars protein data	PCexercise7.sas	13-10-02 21.13
	• text file with the boars protein data	PCexercise7.txt	13-10-02

			21.09
	• *R* program - solutions	PCexercise7Sol.R	13-10-02 21.11
	• *SAS* program - solutions	PCexercise7Sol.sas	13-10-02 21.12
	Variance homogeneity, transformation and confidence intervals	[compact] [full screen]	11-10-02 02:10
7	On covariance and correlation	[compact] [full screen]	05-11-02 14:05
	Other Aspects of Nonlinear Models	[compact] [full screen]	08-11-02 01:16
	• *SAS* program for the examples	OtherAspectsExample.sas	08-11-02 01:09
	• *R* program for the examples	OtherAspectsExample.R	08-11-02 01:09
	• Data for the sheep example	sheep.txt	06-10-02 20.52
	• Data for the ytterbium example	yb.txt	04-11-02 20.46
	• Data for the ytterbium example (in *R*)	yb.R.txt	04-11-02 21.06
	PC exercise 8	[pdf]	07-11-02 00:28
	• *SAS* program for Exercise 8a	PCexercise8a.sas	07-11-02 00.14
	• *SAS* program for Exercise 8b	PCexercise8b.sas	07-11-02 00.14
	• text file with the NDF data	NDF.txt	05-11-02 13.02
			05-11-

	• *SAS* data file with the NDF data	ndf.sas7bdat	02 13.31
	• *SAS* program - solutions for Exercise 8a	PCexercise8aSol.sas	08-11-02 01.16
	• *SAS* program - solutions for Exercise 8b	PCexercise8bSol.sas	08-11-02 01.21

Complementary literature

- Douglas M. Bates & Donald G. Watts (1988). *Nonlinear regression analysis & its applications*. Wiley, New York.
- Norman Draper & Harry Smith (1981). *Applied regression analysis*, 2nd edition. Wiley, New York.
- John B. Fraleigh & Raymond A. Beauregard (1990). *Linear Algebra*, 2nd edition. Addison-Wesley.
- Søren Højsgaard (2001). Grundkursus i SAS. Biometry Research Unit, Danish Institute of Agricultural Sciences. (in Danish)
- Bent Jørgensen (1993). *The Theory of Linear Models*. Chapman & Hall, New York.
- R. Mead, R.N. Curnow & A.M. Hasted (1993). *Statistical methods in agriculture and experimental biology*, 2nd edition. Chapman & Hall, London.

SAS Online documentation

The participants from DIAS may find a rather comprehensive online documentation for *SAS* at

- http://dokumentation.agrsci.dk.

Lectures

Research assistant Leslie Foldager, tel. +45 8999 1231, email: Leslie.Foldager@agrsci.dk

and Senior scientist Søren Højsgaard, tel. +45 8999 1703, email: Soren.Hojsgaard@agrsci.dk

from the Biometry Research Unit.

Detailed schedules

Day 1 Day 2 Day 3 Day 4 Day 5 Day 6 Day 7 Day 8

The initials in parenthesis (SHD=Søren Højsgaard, LFO=Leslie Foldager) is just a help to the lectures so that we know who is responsible for preparing the exercise or lecture.

Day 1 (10. September)

Topics:

- Brush-up of some elementary statistical concepts and tools
- The regression problem

- The one sample problem - Normal data
- Using *SAS* and *R*

Literature:

- BG ch. 2
- BG sec. 3.1
- BG sec. 3.3
- BG sec. 6.1-6.3
- Lecture notes

PC Exercise:
Introduction to *R*. (LFO)
Exercise:
1. compulsory exercise. (LFO)
Notes:
Søren will leave at lunchtime.
Schedule:

9:00-9:30
Welcome, introduction, outline of the course. (SHD/LFO)
09:30-10:55
Lecture: Introduction to Linear and Non-linear Models; Introduction to *R*. (SHD)
11:00-11:50
PC-Exercise 1: Introduction to *R*. (LFO)
12:30-13:15
Lecture: The one sample problem. (LFO)
13:15-14:45
PC Exercise: Continuing PC-Exercise 1. (LFO)
14:45-15:00
Coffee and cake.
15:00-15:45
Lecture: Introduction to 1. compulsory exercise and Writing statistical reports - an example. (LFO)

Day 2 (11. September)

Topics:

- Linear regression in detail.
- Estimation, test, confidence limits, model assessment, transformation
- Two or more samples - Normal data

Literature:

- BG sec. 3.2
- BG sec. 3.3
- Lecture notes

PC Exercise:
Exploratory analysis using *SAS* and *R*. Model assessment using *SAS* and *R*, `ODS`, grouped data. (LFO)

Schedule:

9:00-9:20
Picking up on day 1.
9:20-10:55
Lecture: An Introduction: Regression Analysis. (SHD)
11:00-11:50
PC Exercise 2: Do some exploratory analysis using *SAS* and *R*, save graphs, one sample problem. (LFO)
12:30-13:15
Lecture: Sørens *R* session and a little bit on *SAS* output delivery system (`ODS`).(SHD)
13:15-14:00
Lecture: Two or more samples. (LFO)

(Note: we had to skip 'Testing the hypothesis about the variances' and 'One-way analysis of variance')
14:15-15:00
PC Exercise 3 & 4: Model assessment using *SAS* and *R*, using `ODS`, regression. Grouped data (two or more samples). (LFO)
15:00-15:45
Coffee, cake and lecture: Another look at the regression problem. (LFO)

Day 3 (26. September)

Topics:

- Linear Algebra
- The Linear Normal Model
- The Likelihood Method

Literature:

- Lecture notes
- BG sec. 4.1
- BG sec. 3.1.4

Notes:
1. compulsory exercise returned for approval.

Søren will leave at lunchtime.
PC exercises:
Exercises in linear algebra. LNMs for the tree data using linear algebra.
Schedule:

9:00-9:20
Picking up on day 2
9:20-10:55
Lecture: Basic Concepts from Linear Algebra for use in connection with linear models. (SHD)

(Note: there will be a break during this lecture)
11:00-11:50
PC Exercise: Linear algebra with *R*. (SHD)

12:30-13:15
Lecture: The Linear Normal Model (LNM) in general. (LFO)

Formulation of LNM with terms from linear algebra (the formulation with linear spaces and projections).

13:15-14:45
PC Exercise 5: Fit LNMs to the tree data by using matrix formulation and *R*. (LFO)

14:45-15:00
Coffee and cake.

15:00-15:45
Lecture: The likelihood method. (LFO)

Day 4 (27. September)

Topics:

- Hypothesis testing, error types, power analysis
- Analysis of variance
- Different parameterisations, estimability and contrasts, least squares means
- Treating a sequence of models

Literature:

- Lecture notes
- BG sec. 4.2
- BG ch. 11
- BG ch. 5
- BG sec. 3.2.2
- BG sec. 4.3

Notes:
2. compulsory exercise is to appear at the homepage Monday 30. September.

PC exercises:
Exercise on hypothesis testing / power calculation. (SHD)

Schedule:

9:00-9:20
Picking up on day 3. Some comments on the concluding exercise.

9:20-10:00
Lecture: Hypothesis testing. (SHD)

10:10-10:55
PC Exercise: Power calculation using *R*. (SHD)

11:00-11:50
Lecture: Analysis of variance (soya beans data). Test for variance homogeneity. (LFO)

12:30-12:45
Lecture: Analysis of variance (soya beans data). Test for variance homogeneity. (cont.)

12:45-14:00
Lecture: Different parameterisations, parameterisations in *SAS* and *R*, estimability and contrasts, least squares means. (SHD)

Note: Sørens lecture notes 'Linear Normal Models' are used for this lecture.

14:00-14:15
Introduction to 2. compulsory exercise. (SHD)
14:15-15:15
Lecture: Treating a sequence of models with comparison of regression lines as an example (the blood pressure data). (LFO)

Day 5 (10. October)

Topics:

- Introduction to Nonlinear Models
- Case Study: Vitamin E data

Literature:

- Lecture notes

Notes:
2. compulsory exercise returned for approval.
PC exercises:
Fitting nonlinear models.
Schedule:

9:00-9:20
Picking up on day 4
9:20-10:00
Picking up on 1. compulsory exercise. (LFO)
10:10-10:55
Lecture: Introduction to Nonlinear Models. (LFO)
11:00-11:50
PC Exercise 6: Fitting nonlinear models. (LFO)
12:30-13:15
Lecture: Introduction to Nonlinear Models - continued. (LFO)
13:15-14:30
PC Exercise 6: continued. (LFO)
14:30-15:45
Lecture: Case Study: Vitamin E data. (SHD)

Day 6 (11. October)

Topics:

- Growth curve models
- Variance heterogeneity, Transformations and Confidence intervals

Literature:

- Lecture notes

Notes:

Subject of the concluding exercise have to be defined.

PC exercises:

Fitting nonlinear models.

Schedule:

9:00-9:20

Picking up on day 5

9:20-10:30

Lecture: Growth curve models. (LFO)

10:45-11:50

R and *SAS* - how to do session. (SHD)

12:30-14:00

PC-Exercise 7: Fitting nonlinear models. (LFO)

14:00-14:45

Variance heterogeneity, Transformation and Confidence intervals. (SHD)

14:45

Autumn holidays (for the lucky ones)

Day 7 (7. November)

Topics:

- Other Aspects of Nonlinear Models
  - inference and comparing models
  - comparison between groups or individuals
  - correlated observations
- On covariance and correlation

Literature:

- Lecture notes

Notes:

Evaluation scheme

PC exercises:

Comparing nonlinear models and comparison between groups.

Schedule:

9:00-9:45

Picking up on day 6. Evaluation schemes handed out (return at the end of the day). Picking up on 2. compulsory exercise.

9:50-11:00

Lecture: Other Aspects of Nonlinear Models (inference and comparing models, correlated observations, comparison between groups or individuals). (LFO)

11:00-11:50

PC Exercise 8: Exercise 8a. (LFO)

12:30-13:15

Lecture: Other aspects of Nonlinear Models - continued. (LFO)

13:15-14:45

PC Exercise 8 (cont.): Exercise 8b. (LFO)

15:00-15:45

Lecture: On covariance and correlation. (SHD)

Day 8 (8. November)

Topics:

Concluding exercises presented by the participants.

Literature:

Notes:

PC exercises:

None

Schedule:

9:00-9:20

Picking up on day 7

9:20-11:50

Participants presentation of the concluding exercise: approx. 20 min. each incl. questions.

12:30-13:00

Continued.

13:00-14:00

Concluding remarks on linear and nonlinear models, review of the course topics, and evaluation. (SHD / LFO)

File translated from TEX by TTH, version 3.12.
On 8 Nov 2002, 08:37.

3.2 Example of lecture notes: Biostatistics Course - Fall 2002

Growth Curve Models

Lecture notes

Leslie Foldager
Biometry Research Unit
Danish Institute of Agricultural Sciences

October 8, 2002

Contents

Introduction

- In the lectures and exercises yesterday we saw a few examples of growth curves. In the lecture today the growth will be on the weight of an animal as function of time, but this is just for convenience.
- For a somewhat mathematical entrance to this subject we refer to France and Thornley (1984) chapter 5.
- Moreover, Draper and Smith (1981) ch. 10, Lindsey (2001) ch. 10, Mead *et al.* (1993) ch. 12, and many other textbooks considers growth models.

A growth example

- We consider a serial slaughter experiment (Tullis, 1982) in which 45 pigs (boars, gilts and male castrates) were feeded *ad lib.* and slaughtered between 50 and 330 days of age.
- After slaughter the pigs were dissected, grinded and sampled for determination of chemical composition.
- Table 1 shows total masses of fat (in kg) retained in the body of boars slaughtered from 20 to 200 kg live weight. In Figure 1 we have shown a scatter plot of retained fat versus time.

Time	Fat (kg)	Time	Fat (kg)
52	2.6	126	17.9
55	0.9	165	18.2
91	5.3	195	30.0
91	3.3	231	46.4
123	16.2	285	56.1
123	17.5	330	61.2
124	11.3	332	61.1

Table 1: Time in days and the retained fat (in kg) in slaughtered boars.

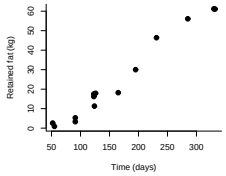


Figure 1: Scatter plot of retained fat (in kg) versus time for the data from the example.

- We will consider a few details on the following growth models:
  - Exponential growth
  - Monomolecular growth
  - Sigmoidal curves
    - Logistic growth
    - Gompertz growth
    - Richards curve

- The form of a growth curve is sometimes chosen by simply looking at some plots of the data. However, it is better to select or construct a function that have biological interpretation and meaningful parameters.
- The functional relationship in the growth models is often derived from knowledge on the rates of growth dy/dt typically as a solution of a differential equation.

- Observations from experiments concerning growth will generally have the following characteristics:
  - The variance increases with size, i.e. it is not constant over time.
  - The growth is a nonlinear function of time, often reaching a limiting value, see the example.
  - A series of observations on a given individual (repeated measurements) will usually be correlated – we will consider this in one of the lectures next time.
  - If several series are observed, they may have different growth either inherently or due to environmental effects.

Exponential growth

- Suppose we have the following rate of growth

$$\frac{dy}{dt} = \beta \cdot y \, .$$

- Note that even though we do not write this explicitly, y is a function of time (we could have used y_t or $y(t)$ but we will not).
- The solution of this differential equation is the exponential growth curve

$$y = \alpha \cdot e^{\beta \cdot t} \, .$$

- We can check that this is a solution to the differential equation by differentiating y with respect to t:

$$\frac{dy}{dt} = \beta \cdot \alpha \cdot e^{\beta \cdot t} = \beta \cdot y \, .$$

- In the exponential growth model the parameter α is the initial size (at time zero), i.e. $\alpha = y_0$.
- For $\beta > 0$ this function will usually only be applicable for the early growth stage (see also the section concerning logistic growth) since nothing will continue growing exponentially forever.

- For $\beta < 0$ this may be a good model of exponential decline, e.g. decaying activity.
- A variant of the exponential growth curve which allows for a gradual approach to an upper limit is the growth curve given by

$$y = \gamma(1 - e^{-\beta \cdot t}) \, , \, \beta > 0 \, ,$$

  where γ is the parameter determining the limit.
- Examples of these simple exponential models are shown in Figure 2-4. We will consider a practical example of the last variant next time.

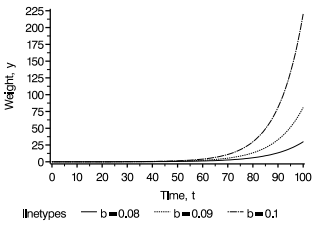


Figure 2: Example of an exponential growth curve of the type $y = \alpha \cdot \exp(\beta \cdot t)$ for $\alpha = 0.01$ and varying values of $b = \beta > 0$.

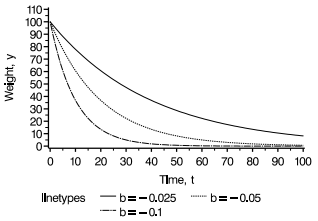


Figure 3: Example of an exponential growth curve of the type $y = \alpha \cdot \exp(\beta \cdot t)$ for $\alpha = 100$ and varying values of $b = \beta < 0$.

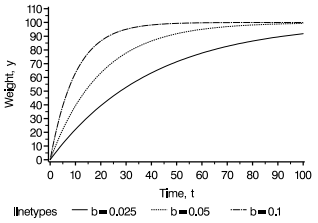


Figure 4: Example of an exponential growth curve of the type $y = \gamma(1 - \exp(-\beta \cdot t))$ for $\gamma = 100$ and varying values of $b = \beta > 0$.

Monomolecular growth

- One of the simplest assumptions leading to a growth curve approaching a level γ (a limiting value) is that the growth rate is proportional to the difference between the level and the actual size, i.e.

$$\frac{dy}{dt} = \beta(\gamma - y) \, , \, \beta > 0 \, .$$

- The solution to this differential equation is

$$y = \gamma - (\gamma - \alpha)e^{-\beta t} = \gamma - \delta e^{-\beta t}\,,$$

the monomolecular growth function, which you tried to fit to the bean plant data in Exercise 1 yesterday.

- Here α is the initial size $\alpha = y_0$ and if $\alpha = 0$ the solution reduces to

$$y = \gamma(1 - e^{-\beta t})\,,$$

von Bertalanffy growth curve, which have been used in ecology to describe animal growth (cf. Lindsey, 2001).

- The monomolecular curve have a rapid initial growth followed by a levelling off.
- Note that von Bertalanffy growth curve was the last example in the exponential growth section and that we, as noted there, will consider a practical example next time.

Sigmoidal curves

- Suppose that the growth rate increases to a maximum before steadily declining to zero.
- This can be accomplished by the following differential equation:

$$\frac{dy}{dt} = f(y) \cdot (\gamma - g(y))\,,$$

where $f(\cdot)$ and $g(\cdot)$ are some functions of the current size and γ is the level that the growth curve will approach.

- This will lead to sigmoidal (S-shaped) growth curves.

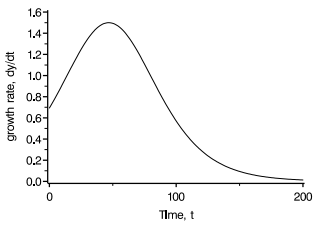


Figure 5: Example of the rate of growth from a logistic growth curve $y = (\alpha\gamma)/(\alpha + (\gamma - \alpha)\exp(-\beta t))$ with $\alpha = 20$, $\gamma = 150$ and $\beta = 0.04$.

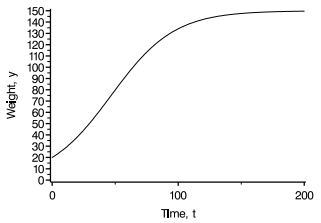


Figure 6: Example of a logistic growth curve $y = (\alpha\gamma)/(\alpha + (\gamma - \alpha)\exp(-\beta t))$ with $\alpha = 20$, $\gamma = 150$ and $\beta = 0.04$.

Logistic growth

- Qualitatively, we can divide the growth of an animal in four stages:
  1. Early exponential growth where the rate of growth is proportional to weight.
  2. Linear growth where more and more energy is devoted to maintenance.
  3. Diminishing growth as a maintenance balance is approached.
  4. Antithesis through senescence. This part is often disregarded since no observations are made at this stage (or it is irrelevant to consider this stage).

- The growth rate at the first stage is proportional to the weight of the animal:

$$\frac{dy}{dt} = \beta \cdot y\,.$$

- The solution of this differential equation is (as we have already mentioned)

$$y = \alpha e^{\beta t}\,,$$

where α is the initial weight, i.e. the growth is exponential.

- The growth at the second stage is linear in time, i.e.

$$y = \alpha + \beta t\,.$$

- The third stage is a limiting stage where the growth rate approaches zero and the weight approaches a limiting level γ.
- The fourth stage will not be considered.
- When we only considers the first three stages we may describe the growth by the following differential equation:

$$\frac{dy}{dt} = \beta y(\frac{\gamma - y}{\gamma})\,.$$

- The solution of this differential equation, known as the logistic curve, is

$$y = \frac{\gamma}{1 + e^{\eta - \beta t}}.$$

- We see that this curve approaches the upper limit γ as time tends to infinity since $\exp(\eta - \beta t)$ tends to zero.
- The parameter η have no direct interpretation but may be seen as a measure of the difference in weight from birth to mature since isolating η at time $t = 0$ and letting $\alpha = y_0$ we find that

$$\eta = \ln(\frac{\gamma}{\alpha} - 1)\,.$$

- By plugging this result into the equation of the logistic curve we find the following re-parameterised version:

$$y = \frac{\alpha\gamma}{\alpha + (\gamma - \alpha)e^{-\beta t}}\,,$$

which you tried to fit in practice yesterday.

- This way of expressing the logistic curve have the advantage that the initial weight is a parameter in the model.
- The rate of growth as a function of time and the form of the logistic growth curve is shown in Figure 5 and 6.

- The inflection point (that is the point on the response curve where the absolute growth rate is greatest) is the time point where $y = \gamma/2$.
- Note that this corresponds to the derived parameter LD_{50} (Lethal Dose 50) in dose-response experiments, that is the dose where half of the individuals dies (or responds to the treatment).

A growth example (continued)

- In practice we first have to determine the initial values of the parameters for the optimisation procedure (`proc nlin` or `nls()`).
- From the scatter plot in Figure 1 we decide to use $\gamma_0 = 65$ as initial value for the parameter γ which determines the limiting value at mature.

- We note that

$$\ln(\frac{y}{\gamma - y}) = \ln(\frac{\alpha}{\gamma - \alpha}) + \beta t\,,$$

i.e. fitting a linear regression of

$$\ln(\frac{y}{\gamma_0 - y})$$

versus time t we can use the estimated intercept $\tilde{\alpha} = -4.43$ and slope $\tilde{\beta} = 0.022$ to determine initial values of α and β.

- We find that

$$\alpha_0 = \frac{\gamma_0 e^{\tilde{\alpha}}}{1 + e^{\tilde{\alpha}}} = 0.76$$

and $\beta_0 = \tilde{\beta} = 0.022$.

- The predicted growth curve is shown in Figure 7 together with a two plots for model assessment. From this we see that the curve tends to be above the observed values when time is less than about 200 days.
- The estimated parameters is $\alpha = 1.29$, $\beta = 0.020$ and $\gamma = 65.5$, i.e. nearby our initial guesses.

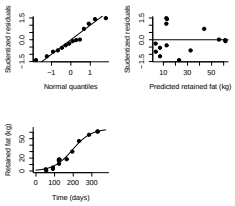


Figure 7: Model assessment of the growth model for the data from the example.

Some practical considerations

- It may be difficult to obtain data over a sufficient range of time containing information on both the initial growth and the third stage.
- The estimation of β is based on the first stage, thus we need observations from this early stage.
- At the other end of the timescale we need observations from the third stage to be able to estimate γ, the weight when a maintenance balance has been reached.

- For the estimation of α we need information on both the first and the third stage.
- If we have repeated weight measurements of the same animal, curves may be fitted separately for each animal and the fitted growth curves for different animals compared through the fitted parameters.
- In other experiments the measuring procedure may require destruction of the organism and repeated measurements is not obtainable.
- In such situations the fitted growth curve is then appropriate to a population of animals rather than individuals.

- It may then be advisable to take larger samples at a relatively small number of well chosen time points rather than smaller samples more frequently.
- It is possible, as we just saw, to find initial values via the following transformation (known as the logit link function from generalised linear models):

$$\begin{aligned}\ln(\frac{y}{\gamma-y}) &= \ln(\frac{y/\gamma}{1-y/\gamma}) \\ &= \ln(\frac{\alpha}{\gamma-\alpha})+\beta t \\ &= -\eta+\beta t\,.\end{aligned}$$

Gompertz growth

- One of the drawbacks of the logistic growth curve is that it is symmetric (time ranging from $-\infty$ to ∞) around its point of inflection.
- We can extend the logistic growth curve to allow for asymmetry around the inflection point.
- There are various ways of writing the Gompertz growth function, see e.g. France and Thornley (1984).

- We will just show one of them:

$$y=\gamma e^{-\delta e^{-\beta t}}\,,$$

where $\delta=\ln(\frac{\gamma}{\alpha})$.

- Here α is the initial weight y_0 and γ is the final weight, i.e. the approached level.
- We see that δ is a pseudo-parameter which measures the difference between the initial and final weight.
- The last parameter β describes the decay in the specific growth rate.

- The point of inflection is the time point where $y=\gamma/e$ (with $e=\exp(1)\approx 2.71828$).
- Compared to the logistic growth, the Gompertz curve shows faster early growth, but a slower approach to the asymptote, with a longer linear period about the inflection point.
- It is possible to find initial values via the following transformation (known as the log-log link function from generalised linear models):

$$\ln(-\ln(\frac{y}{\gamma}))=\ln(\delta)-\beta t\,.$$

- We can invert the asymmetry by:

$$y=\gamma(1-e^{-\delta e^{-\beta t}})\,,$$

- This is known as the complementary Gompertz curve.
- Again, it is possible to find initial values by a transformation (known as the C-log-log link function from generalised linear models):

$$\ln(-\ln(1-\frac{y}{\gamma}))=\ln(\delta)-\beta t\,.$$

A growth example (continued again)

- There is of course no reason to change the initial value $\gamma_0=65$ for the parameter γ determining the limiting value at mature.
- From fitting a linear regression of

$$\ln(-\ln(\frac{y}{\gamma_0}))$$

versus time t we obtain the estimated intercept $\tilde{\alpha}=2.27$ and slope $\tilde{\beta}=-0.015$ from which we determine initial values of δ and β.

- We find that

$$\delta_0=e^{\tilde{\alpha}}=9.68$$

and $\beta_0=-\tilde{\beta}=0.015$.

- The predicted growth curve is shown in Figure 8 together with a two plots for model assessment. The fit seems to be better than that obtained from the logistic growth model, see Figure 7. For comparison we have also shown a scatter plot in Figure 8 with the predicted growth curves from both the logistic and Gompertz model.
- The estimated parameters is $\delta=6.31$, $\beta=0.010$ and $\gamma=75.8$, i.e. the initial value for γ_0 was a bit off.

- We may also calculate the estimated value of α:

$$\alpha = \frac{\gamma}{\exp(\delta)} = 0.14 \, .$$

- We note that the estimated limiting value at mature is quite a bit higher in Gompertz growth model than in the logistic growth model.

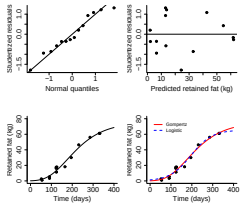


Figure 8: Model assessment of the growth model for the data from the example. In the last plot we compare the fit obtained using Gompertz growth model with that obtained from the logistic growth model.

Richards curve

- In the plant sciences, Richards was the first to apply a growth equation developed by von Bertalanffy to describe the growth of animals (cf. France and Thornley, 1984).
- Richards curve is very general and have the monomolecular ($\nu = -1$), the logistic ($\nu = 1$) and the Gompertz ($\nu = 0$) as special cases, where ν is a parameter in Richards equation.
- As with the other growth curves there are various ways of writing Richards curve.

- We will give the one that was used in Labouriau *et al.* (2000):

$$y = \alpha\{1 + \text{sign}(\nu)e^{\beta - \kappa t}\}^{-1/\nu} \, .$$

- Here $\alpha, \beta > 0$ and $\nu \geq -1$ but $\nu \neq 0$ (for $\nu = 0$ use the Gompertz equation given in the last section).
- The sign function is $\text{sign}(\nu) = 1$ if $\nu > 0$ and $\text{sign}(\nu) = -1$ if $\nu < 0$.
- The inflection point is at the time point

$$t = \frac{\beta - \ln(-\text{sign}(\nu)\nu)}{\kappa} \, .$$

- The expected response at the inflection point is given by

$$\mu = \alpha(\nu + 1)^{-1/\nu} \, .$$

- We see that the point of inflection now is able to occur at any fraction of the final weight, as ν varies over the range $-1 \leq \nu < \infty$.
- The parameter κ controls the position of the inflection point.
- The intercept (or initial value) of Richards curve is

$$y_0 = \alpha\{1 + \text{sign}(\nu)e^{\beta}\}^{-1/\nu} \, .$$

- The parameter α is the limiting size (the asymptote of the curve).
- The parameter ν determines the relative value (compared to the limiting size) of the Richards function at the inflection point:

$$\frac{y}{\alpha} = (\nu + 1)^{-1/\nu} \, .$$

- The parameter β controls the initial size.
- In practice, due to numerical problems, Richards curve is rather difficult to fit. It may be that this model is just too flexible and that it for many practical situations really have too many parameters, i.e. is an example of over-parameterisation.

*

References

Draper, N. and Smith, H. (1981). *Applied Regression Analysis*. John Wiley & Sons, New York, 2nd edition.

France, J. and Thornley, J. (1984). *Mathematical Models in Agriculture*. Butterworth.

Labouriau, R., Schulin-Zeuthen, M. and Danfær, A. (2000). Statistical Analysis of Pigs Development: An application of Richards regression models. *Internal Report 14*, Biometry Research Unit, Danish Institute of Agrucultural Sciences.

Lindsey, J. (2001). The statistical analysis of stochastic processes in time. http://www.luc.ac.be/∼jlindsey/manuscripts.html.

Mead, R., Curnow, R. and Hasted, A. (1993). *Statistical Methods in Agriculture and Experimental Biology*. Chapman & Hall, London, 2nd edition.

Tullis, J. (1982). *Protein growth in pigs*. Phd thesis, University of Edinburgh.

Index

3.3 Example of PC exercises: Biostatistics Course - Fall 2002

Biostatistics Course – Fall 2002
PC Exercises 7

Leslie Foldager
Biometry Research Unit
Danish Institute of Agricultural Sciences

October 10, 2002

1 Exercise 7

The data is from a serial slaughter experiment (Tullis, 1982) in which 45 pigs (boars, gilts and male castrates) were fed *ad lib.* and slaughtered between 50 and 330 days of age. After slaughter the pigs were dissected, grinded and sampled for determination of chemical composition. Table 1 shows total mass of protein (in kg) retained in the body of boars slaughtered from 20 to 200 kg live weight. The data is available from the files \SAS\PCEXERCISE7.SAS and \DATA\PCEXERCISE7.TXT. In this exercise you are free to choose between *SAS*, *R* or a combination.

Time (days)	Protein (kg)
52	3.4
55	2.1
91	7.7
91	7.4
123	11.2
123	14.1
124	10.7
126	10.5
165	16.2
195	26.5
231	24.9
285	27.4
330	29.6
332	29.8

Table 1: Data for Exercise 7. Time in days and the retained protein in kg in slaughtered boars.

1. Make a scatter plot of protein versus time and propose some (nonlinear) models.
   Hint: try to look at the nonlinear models that were used in exercise 6 and in Example 1 in the lecture notes *Introduction to Nonlinear Models* from yesterday.

2. Fit a monomolecular growth model of the following type:

$$y = \gamma(1 - e^{-\beta t}) \ , \ \beta > 0 \, .$$

Hint 1: Look at the scatter plot to determine a reasonable starting value for γ and let us denote it γ_0.
Hint 2: To determine the initial value of β look at the following derived model:

$$\ln(\frac{\gamma_0 - y}{\gamma_0}) = -\beta t \ ,$$

fit the regression line with zero intercept (`noint` option in *SAS* and `lm(z -1+t)` in *R*), and derive β_0 from the estimated slope.

3. How well does this model fit the data and are there obvious drawbacks?
Hint 1: Plots of residuals and predicted values.
Hint 2: What is the predicted value at time zero?

4. Is it possible to modify this model to take the drawbacks into account?
Hint 1: Try to fit the following model:

$$y = \gamma - (\gamma - \alpha)e^{-\beta t} \ , \ \beta > 0 \, .$$

Hint 2: To determine the initial values of α and β look at the following derived model:

$$\ln(\gamma_0 - y) = \ln(\alpha - \gamma_0) - \beta t \ ,$$

fit the regression $z = \tilde{\alpha} + \tilde{\beta} t$ with $z = \ln(\gamma_0 - y)$, $\tilde{\alpha} = \ln(\alpha - \gamma_0)$ and $\tilde{\beta} = -\beta$, and derive α_0 and β_0 from the estimates for $\tilde{\alpha}$ and $\tilde{\beta}$, i.e. by

$$\alpha_0 = \gamma_0 - e^{\tilde{\alpha}}$$

and

$$\beta_0 = -\tilde{\beta} \ .$$

5. How well does this model fit the data and are the drawbacks solved?
Hint 1: What is the estimate of α and is this estimate realistic?

6. Now try to fit a logistic growth curve, i.e.

$$y = \frac{\alpha\gamma}{\alpha + (\gamma - \alpha)e^{-\beta t}} ,$$

where α is the initial amount of protein and γ is the limiting amount.
Hint 1: Determine a starting value of γ as before.
Hint 2: Fit the following linear regression to determine initial values α_0 and β_0:

$$\ln(\frac{y}{\gamma - y}) = \ln(\frac{\alpha}{\gamma - \alpha}) + \beta t .$$

7. Check the logistic growth model by various plots of the residuals, and by adding the predicted curves to the scatter plots. How well does this model fit the data?

8. What do the parameters tell us about the growth of these pigs?

9. Look at the correlation matrix for the parameters. Are there any problems? If there are can we then do something to solve these problems?

10. If we let

$$\eta = \ln(\frac{\gamma - \alpha}{\alpha})$$

then the logistic growth model can be written as (re-parameterised):

$$y = \frac{\gamma}{1 + e^{\eta - \beta t}} .$$

Try to fit this model and see if the correlations are lower.
Hint: Note that η is minus the intercept in the linear regression from Hint 2 in question 6 while the other parameters are unchanged.

References

Tullis, J. (1982). *Protein growth in pigs*. Phd thesis, University of Edinburgh.

3.4 Example of compulsory exercise: Biostatistics Course - Fall 2002

Biostatistics Course – Fall 2002
Compulsory exercise 1

Leslie Foldager and Søren Højsgaard
Biometry Research Unit
Danish Institute of Agricultural Sciences

September 8, 2002

During the course two compulsory exercises and a final project have to be solved. These exercises must be prepared individually, returned, and approved by the lectures. Data will be available via the course homepage:

http://www.jbs.agrsci.dk/Biometri/Courses/Kursusforloeb/BiostatKursus-E02/

1 Writing a statistical report

A report should generally consist of the following three parts which may be divided into subsections, if necessary:

1. **Presentation of the problem and the data.** Present the problem, the data, and the main questions that the statistical analysis is intended to answer.

2. **Statistical analysis.** Describe the application of the statistical procedures and methods. That is, present the questions in statistical terms. A precise mathematical formulation of the statistical models, and of the assumptions behind these models, belongs to this section.

3. Conclusion. Discuss the results of the statistical analysis. Answer the questions posed in part 1.

The statistical terms, such as *test*, *estimator*, *likelihood* should primarily enter into part 2 whereas parts 1 and 3 should be written in a language directed at the researcher who obtained the data. A statistical report should be a clear and fluently written text, such that it can be understood by readers who know the basics of statistical data analysis, but not necessarily the conventions of the statistics course you are currently following. Only the most relevant tables and graphs should be included in part 2, whereas computer programs should be put in an appendix or let out of the report.

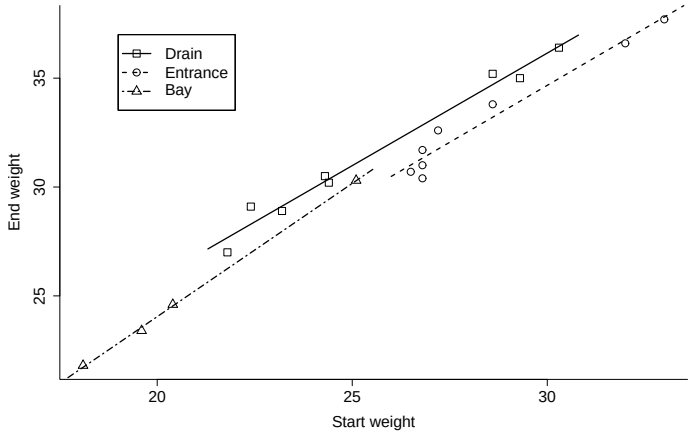
Note Raw output from the programs are not acceptable, not even in an appendix. You should produce some nice tables containing the relevant parts of the output. Output from *SAS* programs may be arranged in tables and exported to e.g. *Word* by using *SAS* Output Delivery System, `ODS`.

2 Data

The exercise concerns the impact of heated water on the growth of oysters.

A number of net were placed on three different locations in a cooling water system of a power station. Eight sets of net were placed at the entrance channel, eight sets at the drain channel, and four sets were placed in a bay nearby the power station. The three placements are regarded as a treatment factor (`treat`).

The oysters are cleaned and weighed at the start of the experiment and one month later. The start weight (`startwgt`) and end weight (`endwgt`) are registered individually for each set of net. The data, presented in Figure 1 and Table 1, is from the final scientific university examination, part 1, Biostatistics, University of Aarhus, Winter 1997/97, Exercise 3.



Figur 1: End weight plotted against start weight of 20 sets of net with oysters. The sets were place on three different locations in a cooling water system of a power station. The lines represents least squares fit.

3 The exercise

Write a statistical report based on your analysis of the oyster data. Some of the questions one could ask are:

- Does heating water have any impact on the growth of oysters?
  (The reason why the experiment are made)
- Are there any relationship between start weight and end weight?
  (Set up a model)
- How well does the suggested model fit to the data?
  (Model assessment)

An appropriate length of the report will typically be about 5 pages plus tables and graphs. The report may be written in English or Danish.

4

Location	Start weight	End weight
drain	28.6	35.2
drain	22.4	29.1
drain	23.2	28.9
drain	24.4	30.2
drain	29.3	35.0
drain	21.8	27.0
drain	30.3	36.4
drain	24.3	30.5
entrance	27.2	32.6
entrance	32.0	36.6
entrance	33.0	37.7
entrance	26.8	31.0
entrance	28.6	33.8
entrance	26.8	31.7
entrance	26.5	30.7
entrance	26.8	30.4
bay	20.4	24.6
bay	19.6	23.4
bay	25.1	30.3
bay	18.1	21.8

Tabel 1: Start weight and end weight of 20 sets of net with oysters. The sets were place on three different locations in a cooling water system of a power station.

3.5 Evaluation of ad-hoc Course - Fall 2017

Fra: Sarah-Lina Aagaard Schild
Til: Jan Tind Sørensen
Cc: Leslie Foldager
Emne: evaluering af ad hoc kursus
Dato: 28. februar 2018 12:04:40

Hej Jan

Jeg er ikke klar over, om det er nødvendigt med en evaluering af det ad hoc statistikkursus, som jeg har haft ved Leslie, men nu får du en ☺

VH
SL

Evaluering af kurset: "Selected Topics from Generalized Linear Models and Survival Analysis"

Kurset blev indledt med en lektion om simpel logistisk regression og afsluttet med lektioner om multipel poisson regression, håndteringen af overdispersion og zero inflation, vejledning i håndteringen af egne data og endelig en skriftelig afrapportering samt en mundtlig eksamen. Opbygningen af kurset fungerede rigtig godt og den faste gennemgang af læsestoffet var med til at afklare mange af de spørgsmål, jeg havde forud for lektionerne. Øvelserne, en del af hjemmearbejdet, var udbytterige og gav mig værktøjer jeg vil kunne bruge til håndteringen af forskellige typer af fremtidige data. Alt i alt et rigtig godt kursus.

3.6 Evaluation of Biostatistics Course - Fall 2001

5. december 2001, LFO

Kursusevaluering af

Kursus i Biostatistik – Efterår 2001
Statistisk Analyse af Biologiske Problemstillinger
med Ikke-lineære Modeller

Underviser: Leslie Foldager

I alt 5 kursusdeltagere, hvoraf 4 gennemførte på tilfredsstillende vis, og fik udleveret kursusbevis, mens 1 deltager kun afleverede den første obligatoriske opgave (der blev stillet tre) og afstod fra at lave et mini-projekt (afsluttende opgave).

Mundtlig evaluering:
De fire deltagere, som fremlagde mini-projekter, var nogenlunde enige om følgende kommentarer:

Generelt var de tilfredse med kursets indhold og gennemførelse. God vægtning mellem emnerne, godt at der var "så meget" af det mere grundlæggende (een og flere normalfordelte observationsrækker, lineær regression, variansanalyse og lineære normale modeller generelt) og det som egentlig var kursets overskrift "ikke-lineære modeller".

De syntes godt om kursus hjemmesiden, hvorfra bl.a. forelæsningsnoter og eksempler på SAS programmer til eksempler og øvelsesopgaver kunne downloades.

Der var et ønske om mere vægt på anvendelse af SAS og forståelse af SAS-udskrifter og mindre vægt på teorien. Dog ikke sådan at forstå, at der overhovedet ikke skal være noget teori, men det skal holdes på så lavt niveau som muligt. Måske inddrage SAS/Insight eller lignende.

Kurset har givet en bedre forståelse af såvel teorien som SAS-udskrifterne. En del løse ender er blevet bundet sammen.

Vedr. øvelserne (der blev stillet ca. 2 opgaver pr. gang + 3 obligatoriske opgaver): nok lidt for mange, svær at nå, evt. regne nogle af øvelserne i forbindelse med kursusdagene.

Ønske: en samlet oversigt over de modeller som er blevet anvendt i løbet af kurset med en kort beskrivelse og tilhørende SAS procedure. Gerne med et plot af en typisk respons kurve (ikke mindst vedr. de ikke-lineære modeller, bl.a. de forskellige modeller for vækst). Formål: at undgå at skulle bladre hele kursusmaterialet igennem, når man senere har et eksempel, der ligner et eller andet fra kurset.

Knapt så godt: lineære algebra (synes de var mindre relevant) og til tider for meget teori.

Skriftlig evaluering:

Alle har enten et godt (2) eller meget godt (2) helhedsindtryk af kurset.

Egen arbejdsindsats betegnes som god for 3 og mindre god for 1. (dette er i tråd med mit indtryk)

Det generelle indtryk af svarene på relevans, behandling og tidsforbrug er: *relevant eller meget relevant, ok behandling eller god behandling, ok tidsforbrug*.

2 deltagere synes, at tiden anvendt på *sandsynlighedsteori* har været for lang. Deltagerne har vidt forskellig opfattelse af relevansen af dette emne (1 i hver kategori og 1 ikke besvaret).

1 deltager har mange krydser i "Var tiden for lang".

2 deltagere synes, at emnet *lineær algebra* var mindre relevant, og at der blev brugt for lang tid på dette emne. 1 deltager syntes endvidere, at behandlingen af emnet var mindre god.

2 deltagere synes, at tiden anvendt på *bootstrap metoder* var for lang, 1 at det var mindre relevant. Der var dog også 1 deltager, som syntes, at der blev brugt for kort tid på dette emne.

3 deltagere (dvs. næsten alle) syntes at tiden brugt på *ikke-lineære modeller* var for kort.

Underviserens kommentarer:

Vedr. *ikke-lineære modeller*: den mundtlige evaluering peger på, at det ikke er de lineære normale modeller, man skal skære i for at få mere tid til de ikke-lineære modeller. Den mundtlige evaluering gav heller indtryk af, at problemet er helt så stort, som spørgeskemaet måske antyder.

Det er måske en idé at lade *bootstrap metoderne* indgå mere integreret, eller evt. helt droppe emnet. Jeg gennemgik (pga. mangel på tid) kun eksempler, som man ikke behøvede bootstrapping til, så et par mere relevante eksempler bør medtages, hvis emnet skal gentages.

Jeg synes, at forslaget om at lave en liste over de anvendte modeller er udmærket og værd at overveje nærmere.

Det afsluttende mini-projekt blev kun fremlagt mundtligt (dog har 2 deltagere også afleveret lidt skriftligt). Måske skal man skære antallet af obligatoriske opgaver ned til 2 og så kræve en skriftlig rapportering af mini-projektet. Een deltager foreslår at undlade den mundtlige præsentation – det synes jeg dog ikke, man skal gøre.

Man kan overveje, om nogle af øvelsesopgaverne skal regnes i forbindelse med kursusdagene evt. i et computerrum (men har vi det?).

Jeg havde planer om, at deltagerne skulle gennemgå øvelsesopgaverne, men generelt havde de ikke nået at lave dem, eller ikke nået at lave så meget at de ville til tavlen. Resultatet blev, at de fleste øvelsesopgaver blev gennemgået på tavlen/projektoren af undertegnede, bl.a. med en gennemgang af tilhørende SAS programmer og udskrifter. Til gengæld gav dette mulighed for at bruge noget tid på SAS programmer og udskrifter, som deltagerne jo er meget interesserede i.

Deltagerne har tidligere udtrykt stor tilfredshed med forelæsningsnoterne (dvs. kopier af slides), som de har modtaget i forbindelse med (dvs. umiddelbart inden) gennemgangen af de enkelte emner. Det er ikke noget stort ekstra arbejde at give dem disse kopier, men det har taget en hel del tid at producere disse slides, som blev vist vha. en "lyskanon". Til gengæld er de forhåbentlig lette at genbruge eller genbruge dele af, da selve teksten ligger i TeX-format.

3.7 Evaluation of Biostatistics Course - Fall 2002

6th November 2002, LFO

Evaluation scheme for

Biostatistics Course – Fall 2002
Statistical Analysis of Biological Problems
using Linear and Nonlinear Models

Dear participant

With this scheme we ask you to comment (anonymous if you like) on the course.

Thanking you in anticipation.

What is your overall impression of the course?

Excellent ___ Good _5_ Less good ___ Poorly ___

How will you describe your own effort?

Excellent ___ Good _5_ Less good ___ Poorly ___

Evaluation of course topics (tick off)	Relevance			Handling			Time spent was		
Emner:	+	+/-	-	+	+/-	-	too short	OK	too long
Introduction to linear and nonlinear models	4	1	0	4	1	0	0	4	0
Introduction to *R*	2	3	0	1	3	1	1	3	1
The one sample problem: statistical model, random variables, estimation, precision, hypothesis, confidence limits	4	1	0	2	3	0	0	4	0
An Introduction: Regression Analysis: assumptions, estimation, model check, variance, hypothesis test, conf. int.	4	1	0	3	2	0	0	3	0
Two or more samples: models, assumptions, hypotheses	4	1	0	1	4	0	0	3	0
Another look at the regression problem: regression parameters, comparison of regression lines, using *SAS* and *R*	4	1	0	2	3	0	0	3	0
Basic concepts from linear algebra: vectors, matrices, lin. alg. in LNMs, linear combinations, spaces, projections	3	2	0	2	3	0	0	4	0
The linear normal model in general: LNM in general terms, likelihood method, estimation, test, sequence of models	2	3	0	1	4	0	1	2	1
Hypothesis testing: various aspects of hypothesis testing, significance level, power of the test, size of an experiment	4	1	0	2	2	0	0	4	0
Analysis of variance: one-way anova, Bartlett's test for variance homogeneity, two-way anova, additivity model	4	1	0	2	2	0	0	4	0
Linear normal models (Sørens lecture notes): Different parametrisations, estimability and contrasts, least sq. means	3	2	0	1	2	1	1	2	0
Introduction to nonlinear models: examples, in *SAS* and *R*, numerical optimisation, initial values, model assessment	4	1	0	2	3	0	0	4	0
Case study: Vitamin E data: using *SAS* and *R*, checking variance assumptions, transformations, qq-plots, analyses	4	1	0	3	2	0	1	3	0
Growth curve models: exponential, monomolecular, logistic and Gompertz growth	2	2	0	1	3	0	0	2	1
Variance heterogeneity, transformation and confidence intervals	3	1	0	2	2	0	0	3	0
Other aspects of nonlinear models: Inference, comparing models, comp. between groups, correlated observ.	3	2	0	1	4	0	0	3	1
Gennemsnit:	**3.2**	**1.4**	**0.0**	**1.8**	**2.5**	**0.1**	**0.2**	**3.0**	**0.2**

Kommentarer fra Leslie:

Ovenstående opgørelse er på basis af 5 returnerede skemaer. Der var 9 deltagere på kurset, hvoraf 2 ikke afleverede nogen obligatoriske eller afsluttende opgaver, 2 afleverede mere eller mindre konsekvent for sent (den ene af disse har i skrivende stund endnu ikke afleveret den afsluttende opgave), mens de sidste 5 var mere fokuserede på at få afleveret tingene til tiden.

Man kunne overveje et introducerende kursus i *R* forud for kurset og så forudsætte kendskab til *R* svarende til dette. Der er flere kursister der synes, at der har været brugt for meget tid på *R*. Der er helt klart ting, der er smartere at gøre i f.eks. *R* end i *SAS*, og nogle af kursisterne vil givetvis også benytte *R* som supplement i fremtiden. Det introducerende notesæt *R for Beginners* af Emmanuel Paradis er værd at have i tankerne!

Der har været en del kritiske kommentarer omkring tidspunktet hvor forelæsningsnoterne var færdige og tilgængelige – altså at det var for sent. Måske er det noget, man skal prøve at gøre noget ved, men det er lidt svært, når vi først umiddelbart før kurset skal starte ved, at det bliver til noget. Derved kommer man fra starten bagud, og kan så bare halse efter de resterende 2 måneder.

Det er en ganske glimrende idé at have 2 undervisere, men det kræver en del koordinering, og igen er det en fordel, hvis man i lidt bedre tid end 2-3 uger før ved, at kurset bliver til noget. Jeg synes, at man bør overveje kraftigt, at det er "senioren", der har hovedansvaret for kurset, når der er en "yngre" og en "ældre" medarbejder på kurset. Derfor kan det jo godt være "junioren", der laver det meste fodarbejde.

Strukturen med blokke af 2 hele kursusdage med 2-3 uger mellem hver blok synes at fungere fint.

Ideen med at benytte egne data til den afsluttende opgave synes også at falde i god jord hos kursisterne.

Det har virket godt med de praktiske computerøvelser, men det kræver lidt benarbejde at installere software, og sørge for at relevante datasæt m.v. er tilgængelige. Man kan godt forestille sig at det bliver lettere, når IT-funktionen får tingene i Miniauditoriet sat lidt mere i system (til gengæld vil der så være en udgift, men det er jo heller ikke gratis i timeløn at have os til at gøre det). Vi betalte nogle lærepenge den første dag, hvor computerne (i Forskerparken) ikke var gjort ordentligt parate. Man skal simpelthen sørge for i rigtig god tid at få tjekket hvordan systemerne fungere, om relevant software er installeret, samt overveje nøje hvordan det hele rent praktisk skal gribes an, når øvelserne går i gang.

Man kan altid diskutere på hvilket fagligt niveau, der skal undervises på. En del af mine forelæsningsnoter og forelæsninger var nok lagt lidt for højt op, men det var med udgangspunkt i forløbet sidste år, hvor opbygningen til at forstå de lidt sværere afsnit var mere grundig. Pointen er, at det er bedst at have nummer 2 underviser med allerede første gang, man underviser i kurset.

Skive, 6. november 2002

Leslie Foldager