

Index of Lecture 1–L

Page	Title
1	Welcome
2	Course outline
3	Plotting repeated measures data
4	Data example: growth of guinea pigs
5	Guinea pig data and mean plot
6	Separate time points analyses
7	Response feature analyses
8	Graph for Ventilation volumes
9	Graph for Rat body weights
10	Graph for Seizure counts

— in addition to overheads by Davis (2002):
1–7 + 17–20 + 24–32.

WELCOME !

VHM 882: Statistical analysis of repeated measures models
— a course with two faces:

- 3-credit directed studies course
 - * set up as part of the program of Elmabrok Masaoud,
 - * more students welcome (contact me if you are interested),
- weekly 1-hours overview lectures of the methods covered in the course,
 - * open to audit, partly or fully, for anybody interested,
 - * primary focus will be on explaining the methods by examples.

I hope this construction will work out well.

Course material:

- textbooks: primarily Davis (2002), Diggle et al. (2002),
- webpage (<http://www.upei.ca/~hstryhn/vhm882>) has links to overheads and programs (+ schedule),
- paper handouts only provided in limited numbers (contact me to reserve yourself a copy).

Today's lecture: Chapters 1–2 in Davis (2002),

- introduction to repeated measures data and analysis,
- univariate methods, specifically: separate time points analysis and response features (or summary statistics).

COURSE OUTLINE

Topics covered:

- types of response variables:
 - * continuous and normally-distributed,
 - * categorical (dichotomous, ordinal, count),
 - * continuous and not normally distributed,
- statistical models:
 - * normal distribution based: linear mixed models, multivariate analysis, repeated measures ANOVA,
 - * for categorical and non-normal data: primarily generalized linear (mixed) models,
- other topics:
 - * design and sample size,
 - * time-varying predictors,
 - * missing data.

Statistical software: primarily R (version 2.0.1),

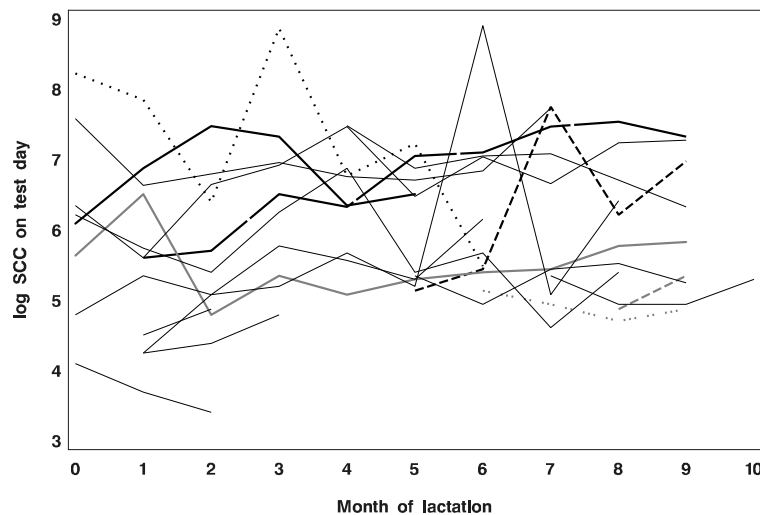
- free software (<http://www.r-project.org>),
- developed and maintained by statisticians for statisticians \Rightarrow no user-friendly point-and-click environment,
- the computing platform where most statistical innovations become accessible,
- *not* the focus of overview lectures (except 2 sessions).

PLOTTING REPEATED MEASURES DATA

Two main types of plots:

- Mean plots: average values across “groups” subjects,
 - * group trends over time,
 - * commonly used for presentation of results.
- Profile plots: series of values over time for subjects,
 - * shows variability between subjects and consistency of mean patterns,
 - * pick suitable subsample(s) if the dataset is too large to plot all profiles in one graph.

Profile plot for milk test data (1 herd, 18 cows):

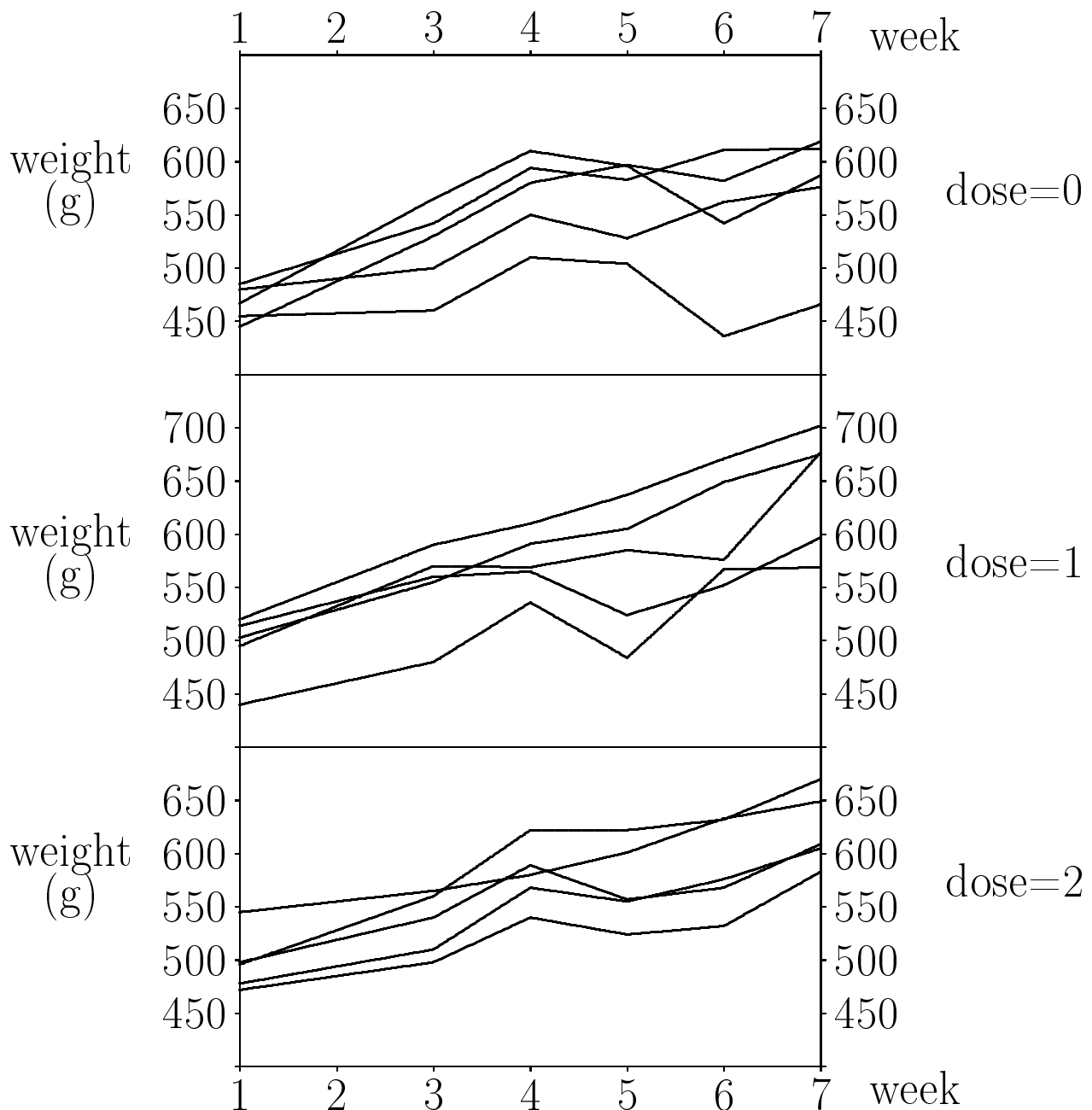


- highly variable patterns (both outcome and times),
- no clear effect of time is seen.

DATA EXAMPLE: GROWTH OF GUINEA PIGS

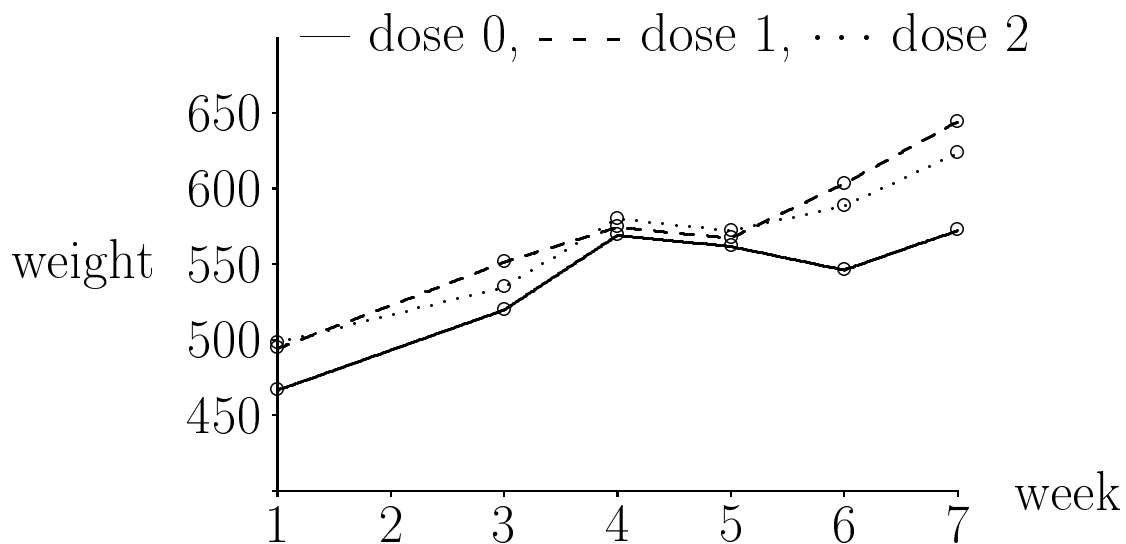
Literature data set (Exercise 5.12 in Davis, 2002):

- weight of 15 guinea pigs for 7 weeks (excl. week 2),
- 3 treatment groups: 0, 1 and 2 doses of vitamin E,
- treatments applied at week 5.



GUINEA PIG DATA AND MEAN PLOT

Dose	Animal	Weeks					
		1	3	4	5	6	7
0	1	455	460	510	504	436	466
	2	467	565	610	596	542	587
	3	445	530	580	597	582	619
	4	485	542	594	583	611	612
	5	480	500	550	528	562	576
1	6	514	560	565	524	552	597
	7	440	480	536	484	567	569
	8	495	570	569	585	576	677
	9	520	590	610	637	671	702
	10	503	555	591	605	649	675
2	11	496	560	622	622	632	670
	12	498	540	589	557	568	609
	13	478	510	568	555	576	605
	14	545	565	580	601	633	649
	15	472	498	540	524	532	583



SEPARATE TIME POINTS ANALYSES

Idea I: analyse for some or all time points with respect to the design of the subjects (Guinea pig data: 3 groups).

Advantages:

- eliminates the “problem” of repeated measures
⇒ simple analysis (Guinea pig data: 1-way ANOVA),
- corresponds to information shown in the mean plot.

Drawbacks:

- does not include development over time in statistical analysis (⇒ substantial information unutilized),
- does not separate within and between-subject variability ⇒ loss of power,
- involves multiple tests (correlated because on the same subjects) ⇒ correction for multiple testing necessary, e.g. a Bonferroni correction multiplying all P -values by the number of time points analysed
⇒ weak analysis (but valid).

Idea II: assess time effects by pairwise comparisons of times within subjects (e.g., paired t -tests),

- involves large number of correlated tests ⇒ unwieldy, and low power after correction for multiple tests,
- not recommended.

RESPONSE FEATURE ANALYSES

Idea: 1) compute single statistic (response feature, summary statistic) from each subject's profile, and 2) analyse these statistics with respect to the design of the subjects.

Examples of response features:

mean, median, sum, standard deviation, gain (last – initial), slope, curvature, correlation coefficient, area under curve. . .
— feature should be of scientific interest and represent profile reasonably well (judgement call; statistical assessment of e.g. linearity not necessary).

Advantages:

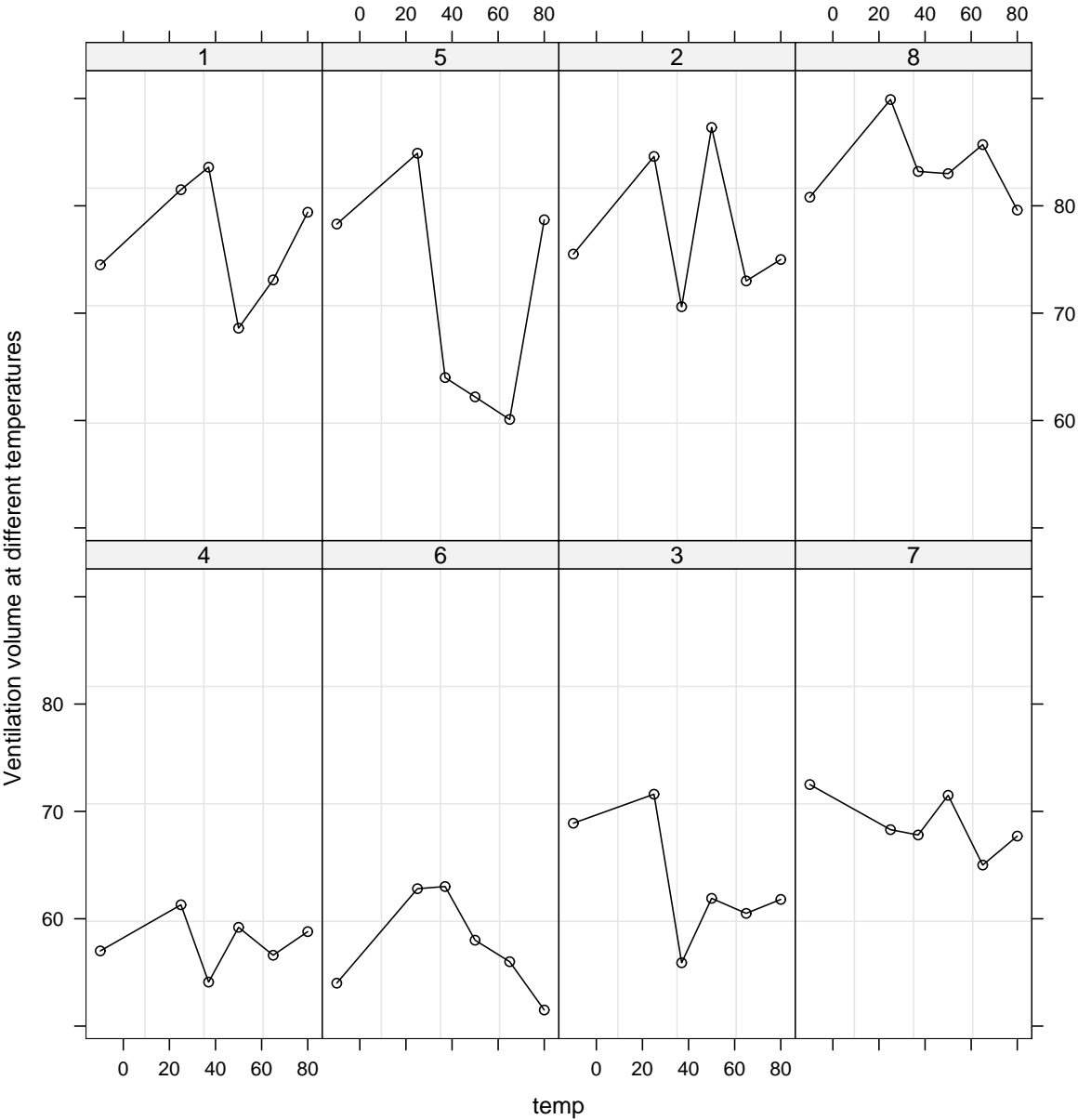
- eliminates the “problem” of repeated measures
⇒ simple analysis (Guinea pig data: 1-way ANOVA),
- gives direct access to feature of interest that may be difficult to extract from complex analysis,
- suitably chosen features can be both powerful and robust towards model assumptions and data irregularities (Everitt, 1995).

Drawbacks:

- no objective way of selecting a feature,
- loss of information by reducing profile to single feature,
- provides only limited information (e.g., no predictions).

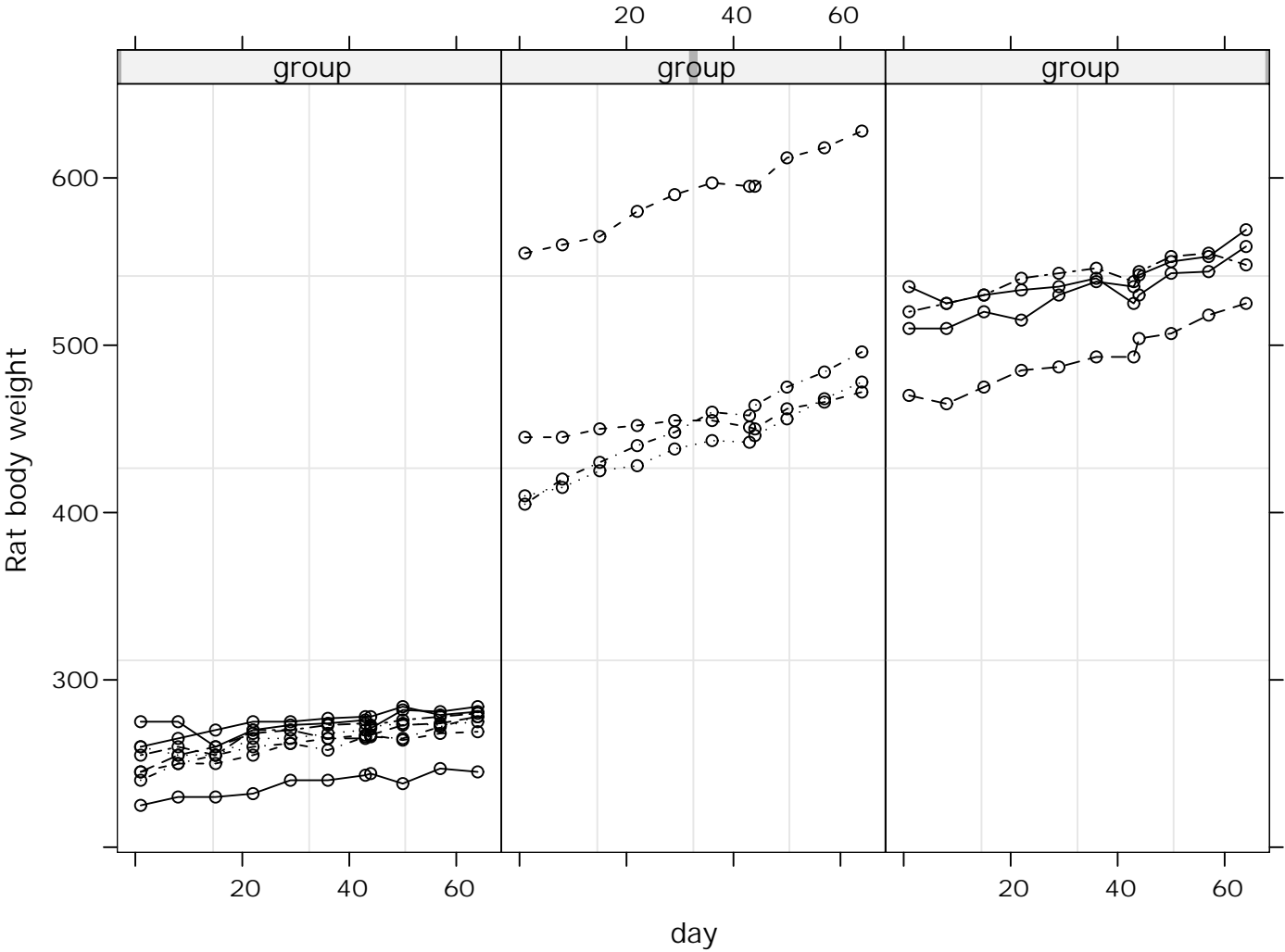
GRAPH FOR VENTILATION VOLUMES

Profile plot for 8 subjects:



GRAPH FOR RAT BODY WEIGHTS

Profile plot for 16 rats in 3 groups:



Rats differ much in initial size (no further explanation given in literature reference).

GRAPH FOR SEIZURE COUNTS

Profile plot for 59 subjects in 2 groups:

