Index of Lecture 12: Repeated measures models (revised on p. 2)

Practical Information

Today's lecture: repeated measures data and modelling:

- introduction + "warnings" + simpler methods,
- 2 advanced methods (not part of course curriculum, but useful in practice!) — including some Stata code demonstrations.

Textbook reading — GO Section 16.6:

- brief discussion of repeated measures data with no worked examples, but good distinction between approaches,
- VER2/MER, Chapter 23: more details and (advanced) worked examples.

Supplementary material:

- notes on repeated measures: from courses taught in Denmark using SAS (lecture will demonstrate Minitab and Stata methods of analysis; SAS programming is, of course, not part of curriculum),
- lecture on repeated measures modelling (GSR Days 2013, at media page).

Other news / schedule:

- 5th home assignment due today; assignment # 6 posted (due April 7),
- you should be well into your projects by now...

Introduction to Repeated Measures

Longitudinal data or Repeated measures (that is, a series of measurements) on the same observational unit (subject):

- very common; usually with repeated measures over time (e.g., observational data),
- special case of multivariate data,
- fits roughly into hierarchical data framework when subjects are put at the lowest level but one (despite ignoring the time ordering),
- explanatory variables may be either time-varying or time-constant (i.e., at subject level); in the latter case, time-varying effects may be of primary interest,
- time points may be equidistant (designed experiments), non-equidistant but the same for all subjects, or variable between subjects in spacing and numbers,[1]
- modelling is not easy $\Rightarrow$ going from simple to complex modelling is recommended.

Contents of this course:

- simple univariate methods for repeated measures (separate lecture notes), except:
  - classical repeated measures ANOVA ("ϵ-correction") available in Stata and SAS only $\Rightarrow$ not core material in course,
- mixed models with complex correlation structures for repeated measures (available in Stata, SAS, R) $\Rightarrow$ introduction, but not core material in course.

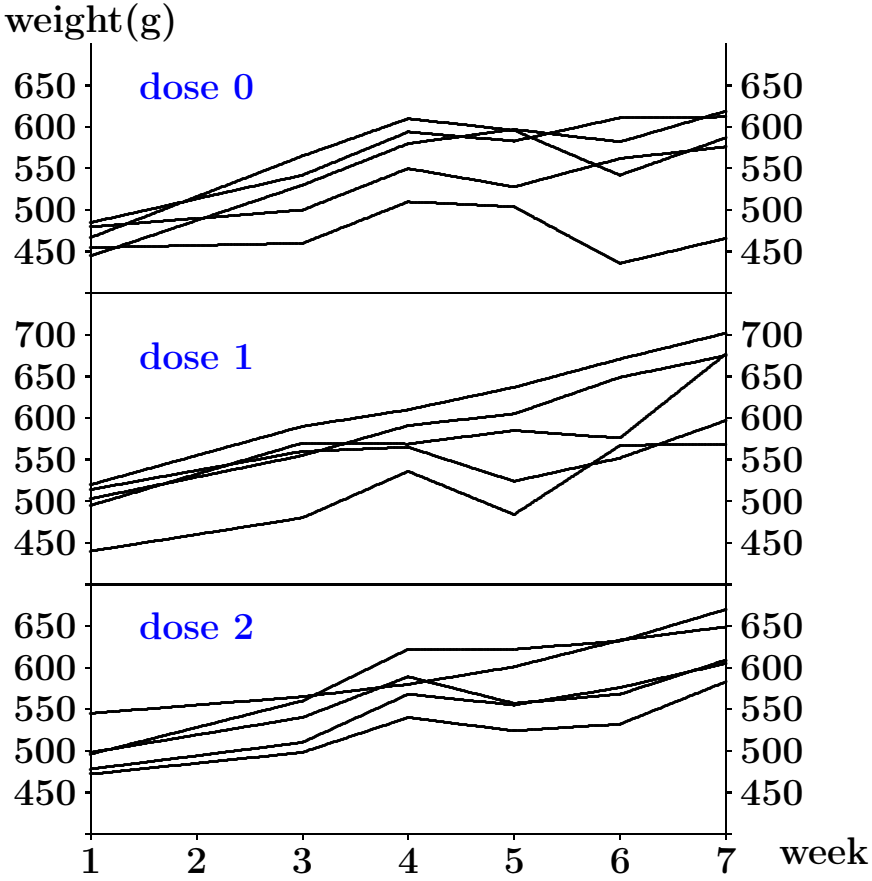
[1] Corresponding to increasing complexity for planning and data analysis.

Data Example: Growth of Guinea Pigs

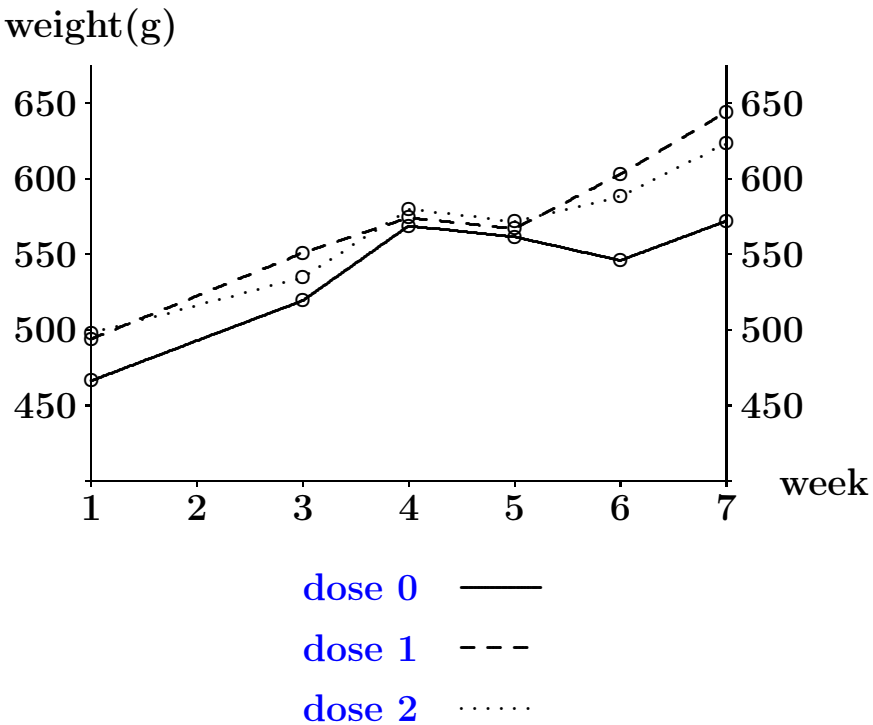
Literature data set (Crowder & Hand, 1990):

- weights of 15 guinea pigs over 7 weeks (excluding week 2),
- 3 treatment groups: 0, 1 and 2 doses of vitamin E; treatments applied at week 5.

profile plots



mean plot



Structuring and Plotting Repeated Measures Data

Same notation as for hierarchical data: Y_{ij} = observation of i^{th} subject at time j.

Handling of incomplete series:

- the observation times (present/missing) are important:
  e.g., the two series $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ and $(Y_{i1}, Y_{i2}, Y_{i4}, Y_{i7})$ are not the same,
- the statistical method/software should take care of this.

Two possible formats of datasets:

- long: usual format, one outcome record per observation,
- wide: one record per subject, multiple (outcome) columns: one per time point.

Two main types of plots:

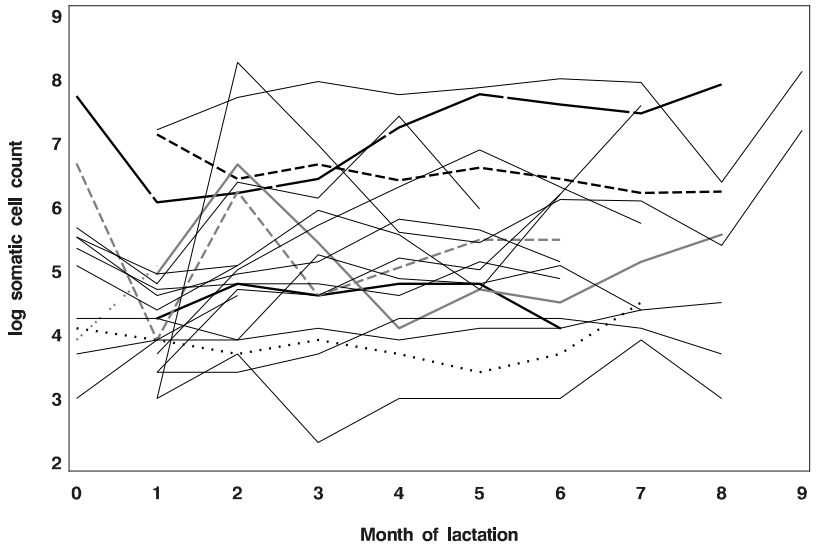
- Mean plot: average values across "groups" of subjects,
  - shows group trends over time (i.e., time interaction plots),
  - commonly used for presentation of results.
- Profile plot[2]: series of values over time for subjects,
  - shows variability between subjects and consistency of mean patterns,
  - pick suitable subsample(s) if dataset is too large to plot all profiles in one graph.

[2] The name "spaghetti plot" is also used (Hedeker and Gibbons, 2006).

Plots for Somatic Cell Count Data (scc_40)

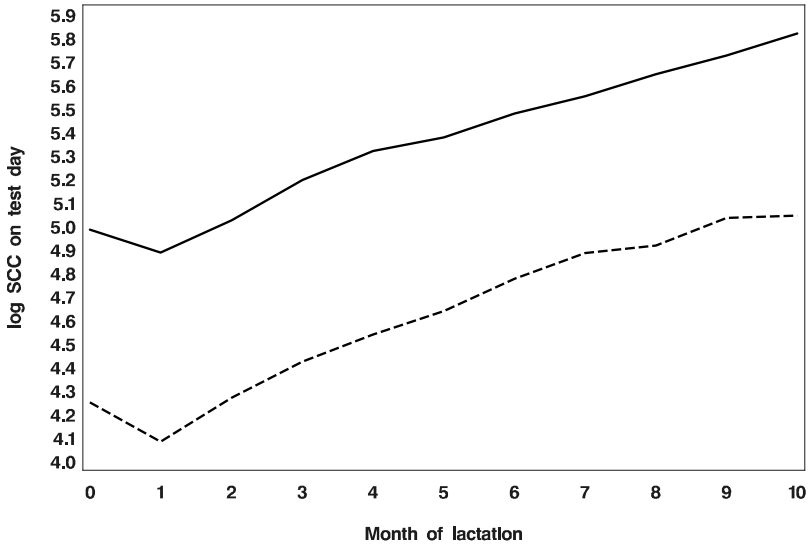
Profile plot for 1 herd (# 3) with 21 cows:

* highly variable patterns (both outcome and time points: highly unbalanced and incomplete dataset),
* no clear effect of time (test month) is seen: the appearance is very noisy,
* could also plot against actual time (days in milk).



Mean plot for heifers (dashed line) and older cows (solid):

* surprisingly regular and almost parallel curves,
* potentially misleading due to missing values.



Univariate Methods for Repeated Measures

Overview of simple/classical univariate methods: (primarily for balanced or close to balanced data)

0) ignore repeated measures = disaster; (effectively assuming all measures taken on different subjects),

i) separate analyses for different times; a valid method if correctly used (Bonferroni adjustment for multiple testing), but weak method (does not utilize full information),

ii) analysis of response feature(s) from $(\boldsymbol{Y}_{1j}, \ldots, \boldsymbol{Y}_{nj})$, e.g.,

- mean, slope, or curvature (or higher order polynomials, but difficult to interpret),
- gain = (end value − initial value), or area under curve (AUC);

may be a very effective method, for well chosen feature,

iii) hierarchical or split-plot model (treating the data as hierarchical without any special consideration for the repeated measures) — often problematic,

iv) repeated measures ANOVA = amendments to split-plot ANOVA when it is inadequate:

- only for balanced data (incomplete series excluded, and requires same observation times for all subjects),
- available in standard software (e.g., Stata, SAS, SPSS),
- was a standard method for repeated measures, but is little used today (and is only acceptable in special cases).

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 (e.g., 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,
- 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 each profile to a single feature,
- provides only limited information (e.g., no predictions).

Response Features for Guinea Pig Data

Examples of response features:

- regression coefficients: for each animal i, fit a regression on time (t_j's),

$$Y_{ij} = \beta_0^{(i)} + \beta_1^{(i)} t_j + \beta_2^{(i)} t_j^2 + \ldots + \varepsilon_{ij}, \; j = 1, \ldots, 6,$$

  and use $\hat{\beta}_0^{(i)}$, $\hat{\beta}_1^{(i)}$, $\hat{\beta}_2^{(i)}$, ... as features[3],

- gain after treatment onset (week 5), i.e, $D_i = Y_{i6} - Y_{i4}$.

Analysis of response features: (one-way ANOVA)

Statistic	Feature (of i^{th} animal) mean $\hat{\beta}_0^{(i)}$	slope $\hat{\beta}_1^{(i)}$	curvature $\hat{\beta}_2^{(i)}$	gain D_i
mean dose 0	539	16.0	-3.88	10.4
mean dose 1	572	22.6	0.49	77.0
mean dose 2	566	19.6	-0.40	51.4
SE	17	3.6	0.85	9.3
F-test	1.06	0.83	7.45	13.0

$F(2, 12, .95) = 3.89$

- clearly significant difference between group 0 and groups 1-2 for two response features: gain and curvature,
- both features reflect the same pattern in the data: the growth for dose 0 animals stagnates after week 5.

[3] Note: only highest order coefficient in the regression equation is useful, so multiple models may need to be fit.

Hierarchical (Split-Plot) Model

Repeated measures treated only as hierarchical structure (i.e., subject random effects):

- model for guinea pig data:

$$Y_{ij} = \mu + \alpha_{\text{dose}(i)} + \beta_j + (\alpha\beta)_{\text{dose}(i),j} + A_i + \varepsilon_{ij},$$

  where α's $\sim$ dose effects, β's $\sim$ week effects, $(\alpha\beta)$ their interaction, and A_i's $\sim$ animal random effects (from $\mathrm{N}(0, \sigma_A^2)$),

- split-plot interpretation:
  - subjects as "large units", as opposed to within-subject "units",[4]
  - balanced data: classical split-plot ANOVA analysis,
- assumes same correlation between all a subject's observations, that is, no matter their distance in time (compound symmetry or exchangeable correlations),
  - intuitively one would expect close observations to be more correlated than distant ones $\Rightarrow$ for long series of measures: unreasonable assumption,
  - for pairs ($n=2$): always ok!, short series ($n = 3, 4$): maybe not so bad,
  - fixed effects in model may add/modify correlation,
- REML variance estimates: in guinea pig example:
  $\hat{\sigma}_A^2 = 1374,\ \hat{\sigma}^2 = 543,\ \hat{\rho} = 1374/(543+1374) = 0.72,$
  — quite high correlation and hardly constant over time...

[4] Contrary to the split-plot design, times are not randomized to split plots (no randomization within subjects).

Repeated Measures ANOVA (ϵ-correction)

= amendments to split-plot ANOVA when it is inadequate:

- test for inherent assumption of equal correlations[5],
  - Mauchly's test (not in Stata!), with reference χ^2-distribution,
- estimate seriousness of violation of the assumption by "ϵ-statistics",
  - $\tilde{\epsilon}$: Huynh-Feldt statistic,
  - $\hat{\epsilon}$: Greenhouse-Geisser statistic, (typically: $\hat{\epsilon} < \tilde{\epsilon}$)
- possible to "correct" the F-tests in the split-plot ANOVA for violation of the assumption using the ϵ-statistics (for $\epsilon < 1$ only; $\epsilon \geq 1 \sim$ split-plot model ok):
  - F-statistics for main effect of time and interaction time $\times$ mainplot factor: $F(\mathrm{df}_1, \mathrm{df}_2) \to F(\epsilon\mathrm{df}_1, \epsilon\mathrm{df}_2)$,
  - F-statistic for main effect of mainplot factor: ok!

Guinea pig example:

- Mauchly's test: statistic = 29.4, df = 14, $P = 0.0093$ — clearly significant,
- adjusted P-values:

Source	F	split-plot	HF-corr.($\tilde{\epsilon}=0.72$)	GG-corr.($\hat{\epsilon}=0.49$)
Weeks	52.55	$< 0.001\,{}^{5}_{60}$	$< 0.001\,{}^{4}_{43}$	$< 0.001\,{}^{2}_{30}$
Weeks×Groups	1.80	$0.080\,{}^{10}_{60}$	$0.11\,{}^{7}_{43}$	$0.15\,{}^{5}_{30}$

- conclusion: clear violation, but apparently not very consequential.

[5] The split-plot ANOVA is valid under a less restrictive assumption of "sphericity", and the method really focuses on deviations from sphericity.

Mixed Models[6] for Repeated Measures

Idea: extend the linear (mixed) model to allow within-subject correlation (among ε's) in models of the (familiar) form,

$$\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}, \tag{1}$$

when the random effects $\boldsymbol{u}$ are still independent of the errors $\boldsymbol{\varepsilon}$,

- different correlation structures (CS; next slides) possible for $\boldsymbol{\varepsilon} \Rightarrow$ flexible models,
- hierarchical structure retained (both $\boldsymbol{u}$'s and $\boldsymbol{\varepsilon}$).

Statistical inference for (extended) mixed models:

- fixed and random effects inference unchanged,[7]
- nested CS may be tested by likelihood-ratio test,
- non-nested CS may be compared by AIC, balancing model fit vs. # parameters,
- practical advice: parsimonious modelling of CS:
  - account for correlations without overfitting the model to the data,
  - complex CS require good data,
  - possible to transfer explanatory power from fixed to random part of model (time-dependent $\boldsymbol{x}$'s may be correlated over time as well).

[6] Also called covariance pattern models when not truly "mixed" (i.e., without random effects).
[7] Choice of denominator degrees of freedom for Wald tests and CIs is an issue: Satterthwaite and Kenward-Roger methods preferable (Schaalje *et al.*, 2001).

Covariance and Correlation Matrices

For a series $\boldsymbol{Y} = (\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_n})$ of observations on a subject[8], the covariance matrix $\mathrm{Cov}(\boldsymbol{Y})$ and correlation matrix $\mathrm{Corr}(\boldsymbol{Y})$ are the $\boldsymbol{n \times n}-$matrices of all pairs of covariances[9] and correlations[10],

$\mathrm{Cov}(\boldsymbol{Y})$ $\qquad\qquad$ $\mathrm{Corr}(\boldsymbol{Y})$

$$\begin{pmatrix} \mathrm{Var}(\boldsymbol{Y_1}) & & & & \\ \mathrm{Cov}(\boldsymbol{Y_1},\boldsymbol{Y_2}) & \mathrm{Var}(\boldsymbol{Y_2}) & & & \\ \mathrm{Cov}(\boldsymbol{Y_1},\boldsymbol{Y_3}) & \mathrm{Cov}(\boldsymbol{Y_2},\boldsymbol{Y_3}) & \mathrm{Var}(\boldsymbol{Y_3}) & & \\ \vdots & \vdots & \vdots & \vdots & \\ \mathrm{Cov}(\boldsymbol{Y_1},\boldsymbol{Y_n}) & \mathrm{Cov}(\boldsymbol{Y_2},\boldsymbol{Y_n}) & \mathrm{Cov}(\boldsymbol{Y_3},\boldsymbol{Y_n}) & \cdots & \mathrm{Var}(\boldsymbol{Y_n}) \end{pmatrix} \quad \begin{pmatrix} 1 & & & & \\ \mathrm{Corr}(\boldsymbol{Y_1},\boldsymbol{Y_2}) & 1 & & & \\ \mathrm{Corr}(\boldsymbol{Y_1},\boldsymbol{Y_3}) & \mathrm{Corr}(\boldsymbol{Y_2},\boldsymbol{Y_3}) & 1 & & \\ \vdots & \vdots & \vdots & \vdots & \\ \mathrm{Corr}(\boldsymbol{Y_1},\boldsymbol{Y_n}) & \mathrm{Corr}(\boldsymbol{Y_2},\boldsymbol{Y_n}) & \mathrm{Corr}(\boldsymbol{Y_3},\boldsymbol{Y_n}) & \cdots & 1 \end{pmatrix},$$

note: the matrices are symmetric; for clarity, the values above the diagonal are left blank.

Simplest example: independent/uncorrelated observations with the same variance $\boldsymbol{\sigma^2}$ (shown for 4 observations):

$$\mathrm{Cov}(\boldsymbol{Y}) = \begin{pmatrix} \boldsymbol{\sigma^2} & & & \\ 0 & \boldsymbol{\sigma^2} & & \\ 0 & 0 & \boldsymbol{\sigma^2} & \\ 0 & 0 & 0 & \boldsymbol{\sigma^2} \end{pmatrix}, \quad \mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

[8] We suppress the subject's index i to ease the notation.
[9] Formally, $\mathrm{Cov}(\boldsymbol{Y_1},\boldsymbol{Y_2}) = \mathrm{E}(\boldsymbol{Y_1}\boldsymbol{Y_2}) - (\mathrm{E}\boldsymbol{Y_1})(\mathrm{E}\boldsymbol{Y_2})$, and $\mathrm{Cov}(\boldsymbol{Y_1},\boldsymbol{Y_1}) = \mathrm{Var}(\boldsymbol{Y_1})$.
[10] Correlations are computed as: $\mathrm{Corr}(\boldsymbol{Y_1},\boldsymbol{Y_2}) = \mathrm{Cov}(\boldsymbol{Y_1},\boldsymbol{Y_2})/\sqrt{\mathrm{Var}(\boldsymbol{Y_1})\mathrm{Var}(\boldsymbol{Y_2})}$.

Repeated Measures Correlation Structures

Examples of covariance/correlation structures ($4 \times 4-$matrices) for observations with homogeneous variances:

- **compound symmetry** or **exchangeable**, $\sim$ hierarchical model:

$$\mathrm{Cov}(\boldsymbol{Y}) = \begin{pmatrix} \sigma^2+\sigma_c^2 & & & \\ \sigma_c^2 & \sigma^2+\sigma_c^2 & & \\ \sigma_c^2 & \sigma_c^2 & \sigma^2+\sigma_c^2 & \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma^2+\sigma_c^2 \end{pmatrix}, \mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho & \rho & 1 & \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

where ρ (ICC) $= \sigma_c^2/(\sigma^2+\sigma_c^2)$,

- **autoregressive**: (first order, often ar(1))

$$\mathrm{Cov}(\boldsymbol{Y}) = \begin{pmatrix} \sigma^2 & & & \\ \rho\sigma^2 & \sigma^2 & & \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 & \\ \rho^3\sigma^2 & \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}, \mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & \rho & 1 & \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

- **arma(1,1)**: (ar(1) and ma(1), $\sim$ time series)

$$\mathrm{Cov}(\boldsymbol{Y}) = \begin{pmatrix} \sigma^2 & & & \\ \gamma\sigma^2 & \sigma^2 & & \\ \gamma\rho\sigma^2 & \gamma\sigma^2 & \sigma^2 & \\ \gamma\rho^2\sigma^2 & \gamma\rho\sigma^2 & \gamma\sigma^2 & \sigma^2 \end{pmatrix}, \mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ \gamma & 1 & & \\ \gamma\rho & \gamma & 1 & \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{pmatrix}$$

- **Toeplitz**: ("stationary")

$$\mathrm{Cov}(\boldsymbol{Y}) = \begin{pmatrix} \sigma^2 & & & \\ \rho_1\sigma^2 & \sigma^2 & & \\ \rho_2\sigma^2 & \rho_1\sigma^2 & \sigma^2 & \\ \rho_3\sigma^2 & \rho_2\sigma^2 & \rho_1\sigma^2 & \sigma^2 \end{pmatrix}, \mathrm{Corr}(\boldsymbol{Y}) = \begin{pmatrix} 1 & & & \\ \rho_1 & 1 & & \\ \rho_2 & \rho_1 & 1 & \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}$$

Mixed Models for Guinea Pig Data

- fairly long series (6 time points) with strong time effects ($\sim$ growth)
  $\Rightarrow$ autocorrelation likely to be present,
- same model equation as for hierarchical model (12L – 10), but with correlation structure within subjects (i.e., among $(\varepsilon_{ij})_{j=1,\ldots,6}$ for each animal i),
- key effect = interaction dose×weeks: impacts of dose should depend on time,
- results[11] based for some correlation structure models, with P-value for interaction:

Structure	# params.	$-2 \log L$	AIC	$\hat{\rho}(1)$	$\hat{\rho}(2)$	$\hat{\rho}(3)$	P-value[12]
compound symmetry	2	720.0	724.0	.72	.72	.72	.080
ar(1)[11]	2	708.6	712.6	.82	.67	.55	.154
RE and ar(1)[11]	3	708.1	714.1	.82	.70	.62	.143
Toeplitz[11]	6	698.4	710.4	.81	.75	.59	.027
unstructured	21	661.4	703.4	.84	.76	.69	$<.001$

  - substantial differences in inference for the interaction,
  - more complex correlation structures have better fit and $\Rightarrow$ significance for the interaction, which can explored by margins and pairwise comparisons.

[11] Results for time-dependent correlation structures based on equidistant times, moving week 1 to week 2.
[12] P-values from F-tests with df $\sim$ experimental design (`nlme` library in R); note that P-values in Stata are too liberal because based on a normal approximation for a not very large dataset.