

# STATS 211: Statistical Methods II

## Lecture 8: Longitudinal data analysis

Babak Shahbaba

UCI, Winter 2012

## Background

- Early biomedical studies focused on modeling the occurrence of a specific outcome (i.e., disease, death, ...).
- Later, survival analysis models (started by D.R. Cox's paper in 1972) emphasized the importance of evaluating time-to-event as opposed to whether events occur or not.
- Investigators become interested in learning how the levels and patterns of risk factors change over time. Therefore, many biomedical studies started to follow the population of interest over time in order to understand the disease process and how this process is affected by temporal changes in the corresponding risk factors.
- This led to the development of longitudinal data analysis methods.

## Longitudinal data

- In longitudinal studies, measurements of the same individuals are taken repeatedly over time.
- Longitudinal data are basically clustered data.
- Each cluster includes a set of repeated measures for a single individuals.
- Therefore, within each cluster, observations have temporal order.
- In that sense, longitudinal data are similar to time series.
- However, they are different from time series in that longitudinal data include many observations with few repeated measures; whereas, time series data usually include many repeated measures over a long time period for few (usually  $N = 1$ ) subjects.

## Objectives of longitudinal analysis

- Longitudinal analysis allows us model within individual changes.
- The main objective of longitudinal studies is to assess the within-individual changes in the response variable and identify how these changes are related to a set of covariates.
- In these studies, each individual acts as its own control.
- The correlation among measurements on the same individual allows us to estimate the rate of change in the response variable and the effect of covariates on the rate of change more precisely.
- However, to take advantage of the underlying correlation structure to improve the precision and efficiency of our estimates, we need to account for this structure in our statistical models.
- Ignoring the correlation structure, on the other hand, results in incorrect estimation of sampling variability, which could lead to wrong inference.

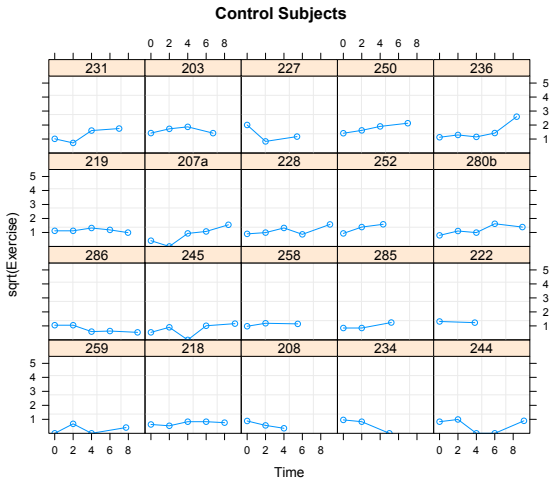
## Objectives of longitudinal analysis

- One approach to account for the dependencies among measurements on the same individual is to use mixed effects models discussed in the previous lecture.
- Alternatively, we can use linear regression models with non-diagonal covariance structure. That is, we model the covariance structure as well as the mean.
- We can also use mixed effects models with non-diagonal covariance structure for measurements on the same subject.
- In what follows, we start with some simple exploration techniques for longitudinal data. Next, we discuss mixed effect models for such data. Finally, we discuss modeling the covariance structure.

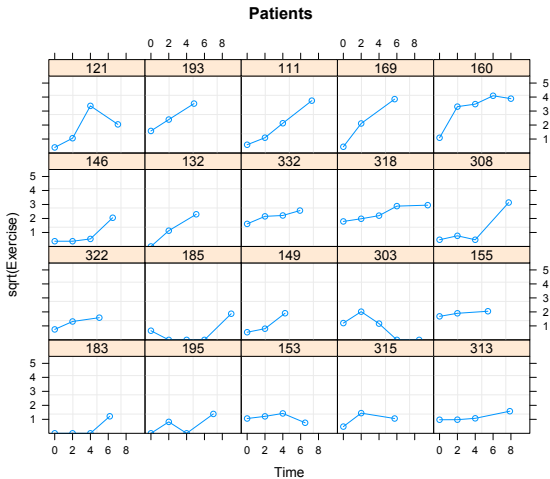
## Exploring longitudinal data

- In this lecture, we discuss longitudinal data analysis methods. However, we first start with exploring longitudinal data.
- As an example, we look at the Blackmoor and Davis's data on exercise histories of 138 teenage girls hospitalized for eating disorders and 98 control subjects.
- The data set is available from the *car* package.
- We start by plotting the response profile (trajectory, time plot) for a sample of subjects

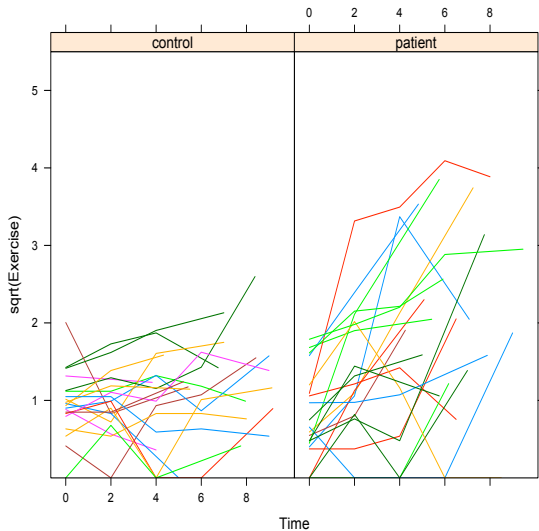
# Time plot



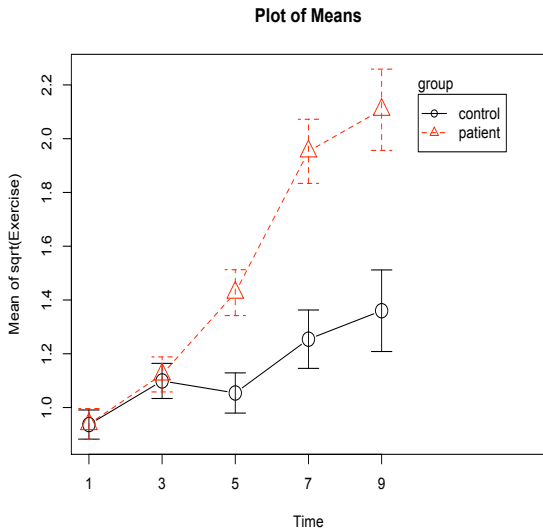
# Time plot



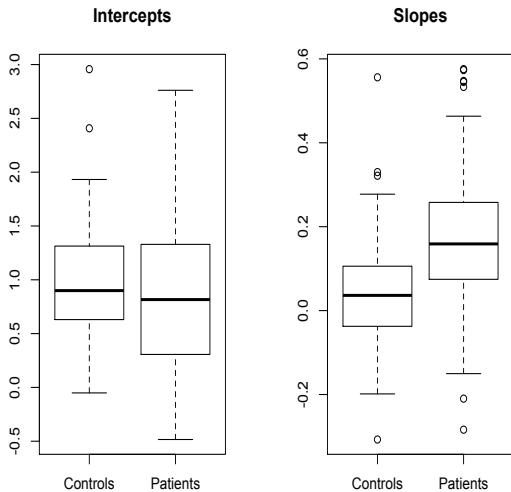
# Spaghetti plot



# Plot of means



# Fitting individual linear models



## Linear mixed effects models

- Since longitudinal data are in fact clustered data, we can model them using linear mixed effects (LME) models.
- This is based on the idea that the natural heterogeneity in the population is the result of random variation (from one subject to another) in some model parameters.
- Therefore, we modify the linear regression model such that some of the regression parameters become subject specific.
- This way, we explicitly distinguish between-subject variability from within-subject variability.
- In general, the LME models provide a very flexible framework for modeling longitudinal data.

## Linear mixed effects models

- The random intercept model for longitudinal data can be written as

$$y_{ij} = \alpha_i + x_{ij}\beta + \epsilon_{ij}$$

$$\alpha_i \sim N(0, \sigma_\alpha^2)$$

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

- The *conditional model* given  $\alpha_i$  is

$$E(y_{ij}|\alpha_i) = \alpha_i + x_{ij}\beta$$

- By integrating over random intercepts, we obtain the *marginal model*

$$E(y_{ij}) = x_{ij}\beta$$

- While  $\beta$  captures the overall mean response in the population,  $\alpha_i$  captures how the mean response varies from one subject to another.

## Linear mixed effects models

- In general, some coefficients as well as the intercept can be subject specific

$$y_i = x_i\beta + z_i\gamma_i + \epsilon_i$$

$$\gamma_i \sim N(0, \Lambda)$$

$$\epsilon_i \sim N(0, \sigma^2 I_n)$$

- For the *conditional model* given  $\gamma_i$ , we have

$$E(y_i | \gamma_i) = x_i\beta + z_i\gamma_i$$

$$\text{Cov}(y_i | \gamma_i) = \text{Cov}(\epsilon_i) = \sigma^2 I_n$$

- For the error term we could use a non-diagonal covariance matrix.
- Here, given the random effect parameters, the observation for the same subject become independent.
- In some situations, this is not a realistic assumption.

## Linear mixed effects models

- For the *marginal model* we have

$$\begin{aligned}E(y_i) &= x_i\beta \\ \text{Cov}(y_i) &= \text{Cov}(z_i\gamma_i) + \text{Cov}(\epsilon_i) \\ &= \sigma^2 I_n + z_i\Lambda z_i^T\end{aligned}$$

- As we can see, the repeated measurements for the same subject are not independent marginally.
- That is, by using random effect parameters we introduce a correlation structure among the repeated observations for the same subject.

# Assignment

- Consider the following random intercept model,

$$\begin{aligned}y_{ij}|\alpha_i &\sim N(\alpha_i + \beta x_{ij}, \sigma_y^2) \\ \alpha_i &\sim N(\mu_\alpha, \sigma_\alpha^2)\end{aligned}$$

1. write down the marginal distribution of  $y_i$ , where  $y_i = (y_{i1}, \dots, y_{in_i})$ .
2. Write down  $\text{Cov}(y_i)$  and  $\text{Cor}(y_{ij}, y_{ik})$ .

## The LME model for the Blackmoor data

- We use the `lme` function from the `nlme` package to model the Blackmoor data.
- The model includes an interaction term between time and group along with their main effect.

```
> lme.full <- lme(sqrt.ex ~ time * group, random = ~time |
+   subject, method = "ML", data = Blackmoor)
> summary(lme.full)
```

Linear mixed-effects model fit by maximum likelihood

Data: Blackmoor

AIC	BIC	logLik
1994.449	2033.259	-989.2247

Random effects:

Formula: ~time | subject

Structure: General positive-definite, Log-Cholesky parametrization

StdDev	Corr
--------	------

(Intercept) 0.49528619 (Intr)

time 0.08740435 0.199

Residual 0.51787325

Fixed effects: sqrt.ex ~ time \* group

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.9589194	0.06743817	712	14.219239	0.0000
time	0.0370754	0.01441755	712	2.571547	0.0103
grouppatient	-0.1230540	0.08692496	229	-1.415635	0.1582
time:grouppatient	0.1311980	0.01816761	712	7.221535	0.0000

Correlation:

	(Intr) time	grptn
time	-0.284	
grouppatient	-0.776	0.220
time:grouppatient	0.225	-0.794 -0.275

Standardized Within-Group Residuals:

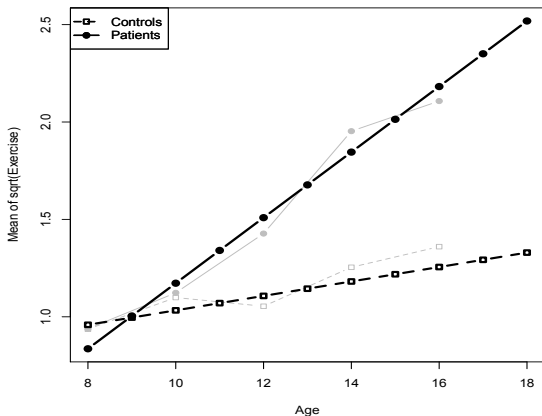
Min	Q1	Med	Q3	Max
-3.00532523	-0.50698420	0.03126225	0.49627872	4.06030233

Number of Observations: 945

Number of Groups: 231

## Blackmoor data

- The following figure shows the fitted model (black lines) along with the plot of means of the observed data (gray lines).



## Modeling the correlation structure

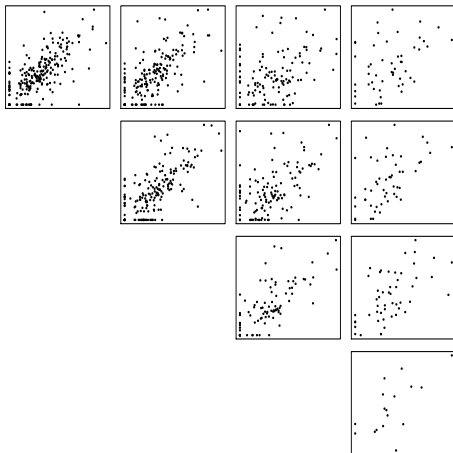
- As mentioned above, by using random effect parameters we introduce a correlation structure among the repeated observations for the same subject even if we assume the observations are conditionally independent given the random effect parameters.
- We could, of course, model the correlation structure directly.
- In what follows, we discuss some common correlation structures.
- To see which one is more appropriate, we could start by exploring the correlation structure of the observed data.

## Exploring the correlation structure

- We can use the observed data to visualize the correlation structure in the response variable.
- To remove the effect of covariates, we first regress  $y_{ij}$  on  $x_{ij}$  and obtain the residuals  $r_{ij} = y_{ij} - x_{ij}\beta$ .
- We then plot  $r_{ij}$  against  $r_{ik}$ .
- If observations are not collected at common time points, we have to put them in a common set of time intervals.

# Exploring the correlation structure

Rows and columns present the ordered time points



## Linear models for longitudinal data

- Suppose our data include  $N$  subjects, where there are  $n_i$  observations for subject  $i$ :  $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$ .
- Denote the covariance between  $y_{ij}$  and  $y_{ik}$  as  $\sigma_{jk}$

$$\sigma_{jk} = E[(y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})]$$

where  $\mu_{ij} = E(y_{ij})$  and  $\mu_{ik} = E(y_{ik})$ .

- The correlation between two measurements  $y_{ij}$  and  $y_{ik}$  is  $\rho_{jk} = \sigma_{jk} / \sigma_j \sigma_k$ , where  $\sigma_j$  and  $\sigma_k$  are standard deviations of  $y_{ij}$  and  $y_{ik}$  respectively.
- Note that, we still assume observations from different subjects are independent.

## Linear models for longitudinal data

- We can write the covariance matrix for repeated measures as follows:

$$\text{Cov}(y_i) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \dots & \sigma_n^2 \end{pmatrix}$$

- We can write the corresponding correlation matrix for  $n$  repeated measures as follows:

$$\text{Corr}(y_i) = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \dots & \dots & 1 \end{pmatrix}$$

## Covariance patterns

- The covariance matrix for  $y_i$  has  $n(n + 1)/2$  unique parameters.
- If we do not impose any structure on the covariance matrix, we have to estimate many parameters with a very limited amount of data.
- Additionally, this unstructured form of covariance causes difficulties when observations are collected at different time intervals, and the number of observations varies from one subject to another.

## Covariance patterns

- The correlation between repeated measures is usually positive.
- The correlation usually decreases with increasing time difference.
- We could use this information to impose structure on the covariance matrix in order to reduce the number of parameters.
- We need of course to keep a balance since too much constraint could result in gross misspecification.
- By imposing a structure on the covariance matrix, we make it a function of smaller number of parameters,  $\theta$ . Therefore, we write the covariance matrix as  $\Sigma(\theta)$ .

## Examples of covariance patterns

- Compound symmetry: constant variance and correlation.

$$\text{Cov}(y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \dots & \dots & 1 \end{pmatrix} \quad \rho \geq 0$$

- Toeplitz: observations equally separated in time have the same correlation.

$$\text{Cov}(y_i) = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \vdots \\ \rho_2 & \rho_1 & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{n-1} & \dots & \dots & \dots & 1 \end{pmatrix} \quad \rho \geq 0$$

## Examples of covariance patterns

- Autoregressive: a special case of Toeplitz.

$$\text{Cov}(y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \dots & \vdots \\ \rho^2 & \dots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \dots & \dots & \dots & 1 \end{pmatrix} \quad \rho \geq 0$$

- Banded: the correlation becomes zero beyond a certain time.
- Exponential: Generalization of autoregressive when observations are not equally spaced;  $\text{Corr}(y_{ij}, y_{ik}) = \rho^{|t_{ij} - t_{ik}|}$ .

## Linear models for longitudinal data

- Corresponding to each observation  $y_{ij}$ , we have a set of observed values on  $p$  covariates (the elements of the first column are all 1's)

$$x_i = \begin{pmatrix} x_{i11} & \dots & x_{i1p} \\ \vdots & \dots & \vdots \\ x_{in_i1} & \dots & x_{in_ip} \end{pmatrix}$$

- The covariates may change (e.g., blood pressure) or may remain the same (e.g., gender) over time.
- As before, the regression model can be written as

$$\begin{aligned} y_i &= x_i\beta + \epsilon_i \\ \epsilon_i &\sim N(0, \Sigma_i) \end{aligned}$$

- Or alternatively

$$y_i \sim N(x_i\beta, \Sigma_i)$$

## Estimation and statistical inference

- We can estimate the model parameters,  $\beta$  and  $\Sigma_i$  (or  $\Sigma_i(\theta)$ ), by maximizing the likelihood function

$$\begin{aligned}\ell(\beta, \Sigma) = & -\frac{\sum_{i=1}^N n_i}{2} \log(2\pi) - 1/2 \sum_{i=1}^N \log |\Sigma_i| - \\ & 1/2 \left[ \sum_{i=1}^N (y_i - x_i \beta)^T \Sigma_i^{-1} (y_i - x_i \beta) \right]\end{aligned}$$

- For a given  $\Sigma_i$ , we have the following *generalized least squares* (GLS) estimate for  $\beta$ :

$$\hat{\beta} = \left[ \sum_{i=1}^N (x_i^T \Sigma_i^{-1} x_i) \right]^{-1} \sum_{i=1}^N (x_i^T \Sigma_i^{-1} y_i)$$

## Estimation and statistical inference

- For simplicity, assume the observations are collected at a common set of times,  $t_j$ . Therefore, we can write the estimate for  $\beta$  as follows:

$$\hat{\beta} = (x^T \Sigma^{-1} x)^{-1} x^T \Sigma^{-1} y$$

where  $x$  is the matrix of covariates for all observations,  $y$  is the corresponding response variable, and  $\Sigma$  is a block-diagonal matrix.

- The above estimate is known as *weighted least squares* estimate, which has the following properties:

$$E(\hat{\beta}) = \beta, \quad \text{Cov}(\hat{\beta}) = (x^T \Sigma^{-1} x)^{-1}$$

- Note that the weight matrix is the inverse of covariance matrix.

## Estimation and statistical inference

- So far, we have estimated  $\beta$  for a given value of  $\Sigma$  (or  $\theta$ ).
- In general, we have to estimate these parameters, for example, by maximizing the likelihood function with respect to  $\Sigma$  (or  $\theta$ ).
- Then,

$$\hat{\beta} = (x^T \hat{\Sigma}^{-1} x)^{-1} x^T \hat{\Sigma}^{-1} y$$

- As before, inference regarding the significance of model parameters  $\beta$  can be done using Wald and likelihood ratio statistics.

## ML vs. REML

- We could use Wald and likelihood ratio tests for inference regarding the  $\beta$  as long as we use the ML estimate for  $\Sigma$ .
- However, as discussed before, the MLE for  $\Sigma$  is biased the same way that the MLE for  $\sigma^2$  is biased in linear regression models.
- Patterson and Thompson (1971) proposed the method of *restricted maximum likelihood* (REML) to address this issue.
- One definition of REML is as follows. Instead of the likelihood function  $L(\beta, \Sigma)$ , we assume a uniform prior distribution over  $\beta$  and use the marginalized likelihood function as follows:

$$L_{RE}(\Sigma) = \int L(\beta, \Sigma) d\beta$$

- This is a function of  $\Sigma$  only. The value of  $\Sigma$  that maximizes this function is  $\tilde{\Sigma}$ .
- Using this estimate instead of  $\hat{\Sigma}$ , we obtain REML for  $\beta$ .

## ML vs. REML

- Note that we cannot use the likelihood ratio test for inference regarding  $\beta$  anymore if we use REML estimates since the commonly used distributional assumption for this test does not hold.
- We could however use the Wald test.

## Selecting the covariance structure

- When comparing two models with the same fixed effects but different covariance structures, we can use the likelihood ratio test as long as the models are nested: one model is a special case of the other one.
- For example, autoregressive is a special case of Toeplitz.
- When this is not the case, we can use other methods such as AIC or BIC.
- In this case, it is common to use REML estimates to calculate log-likelihoods, and we calculate the penalty based on the number of covariance parameters (the number of fixed parameters are the same between the two models).

## Residual analyses and diagnostics

- Similar to standard linear regression models, we examine the residuals obtained from a LME model to assess the adequacy of the fitted model and identify potential outliers.

$$r_{ij} = y_{ij} - \mathbf{x}_{ij}^T \hat{\beta}$$

- Recall that for linear regression models, we examine the plot of residuals to make sure there is no trend, and the residuals are randomly scattered around zero.
- For longitudinal data, however, the residuals are correlated and, in general, they do not have constant variance

$$\text{Cov}(r_i) \approx \Sigma_i$$

- Also, the residuals may be correlated with the covariates.

## Residual analyses and diagnostics

- To address these issues, we need to transform residuals so they resemble the residuals we obtain from linear regression models: with constant variance and zero correlation.
- For this, we first find the Cholesky decomposition of  $\hat{\Sigma}_i = \hat{\sigma}^2 I_n + z_i \hat{\Lambda} z_i^T$

$$\hat{\Sigma}_i = L_i L_i^T$$

- Then, we transform the residuals as follows

$$r_i^* = L_i^{-1} r_i$$

- Now, we can treat  $r_i^*$  as if they are obtained from standard linear regression models. For example, we can plot them against transformed predicted values,  $L_i^{-1} x_i \hat{\beta}$ , and transformed covariates,  $L_i x_i$ .