Statistisk Analys av Komplexa Data
Longitudinella Data

Mattias Villani

Statistik
Institutionen för Datavetenskap
Linköpings Universitet

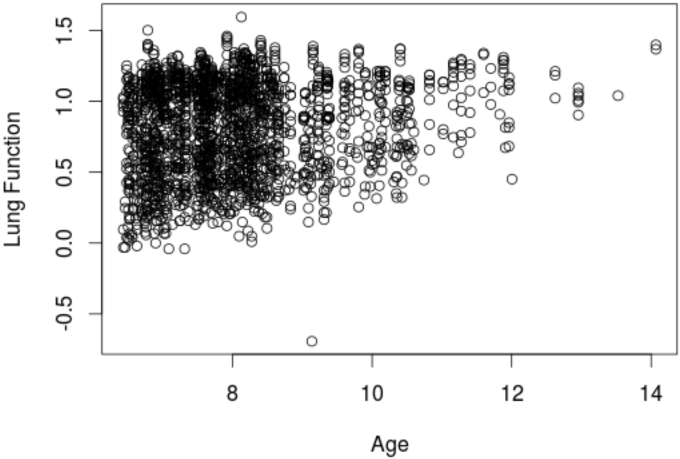
Momentets innehåll

- Introduktion till longitudinella data
- Modeller för väntevärdesprofiler
- Modeller för kovariansmatriser
- Modeller med fixed och random effects
- R-paket för analys av longitudinella data

Tvärsnittsdata (Cross-sectional data)

- **En** mätning per individ/subjekt (blodtryck för individ i)
- Mätningen **kan vara fler-dimensionell** (blodtryck och kroppstemperatur för individ i)
- De olika mätvariablerna (blocktryck och temp) kan vara beroende/korrelerade.
- **Ingen tidsdimension**
- Kan **jämföra olika delpopulationer** som råkar ha skilda åldrar, men ingen info om hur en given individ utvecklas över tiden.
- Mellan-individ effekter, men inga inom-individ effekter.
- Oparat t-test.
- Vanlig modell: **regression**

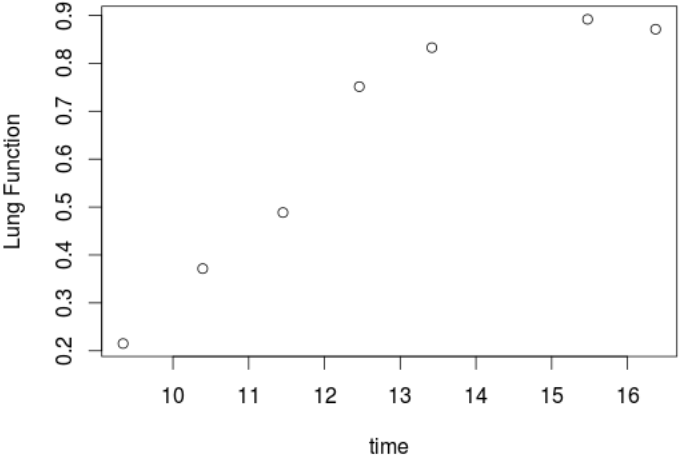
Exempel lungfunktion



TIDSSERIEDATA

- En mätvariabel som observeras **över tid**.
- Oftast många mätningar över tiden (lång tidsserie med > 100 observationer)
- **Beroende** mellan mätningar vid olika tidpunkter
- Starkast beroende mellan närliggande tidpunkter
- Mätvariabeln **kan vara fler-dimensionell**
- Vanlig modell: **ARIMA** eller **state-space** modeller

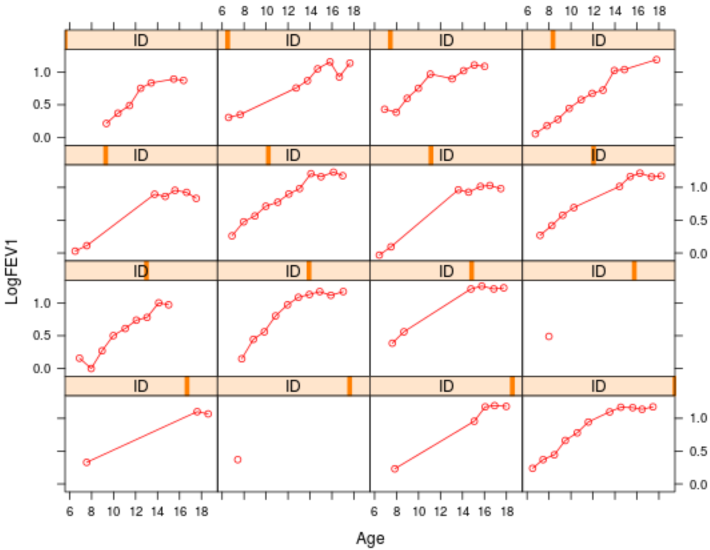
Exempel lungfunktion



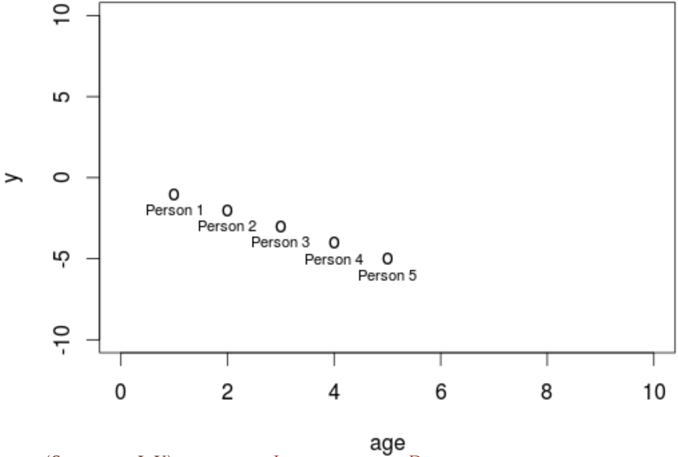
Longitudinella data

- **Samma individer** observeras vid **flera olika tidpunkter**.
- Ger information om **en individs förändring över tiden**.
- Tänk parat t-test.
- Kombo av tvärsnitts- och tidsseriedata.
- Ofta få mätningar per individ (5-20 st)
- Longitudinella data är i princip korta tidsserier, men har egna modeller och metoder.
- Mätningar mellan olika individer antas ofta vara oberoende
- Mätningarna för en individ tenderar att vara beroende. Autokorrelation.

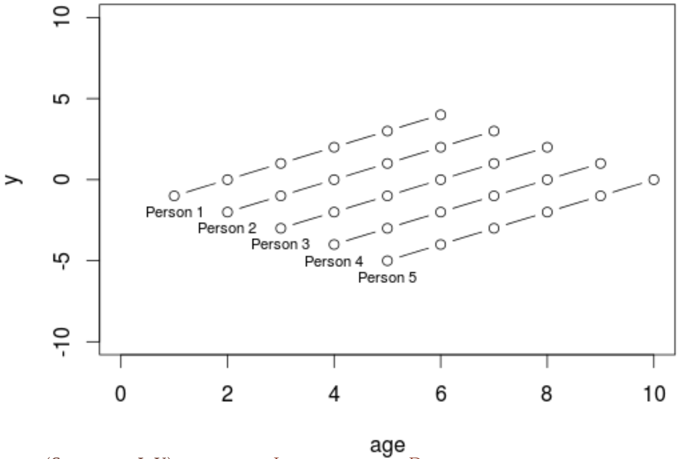
EXEMPEL LUNGFUNKTION



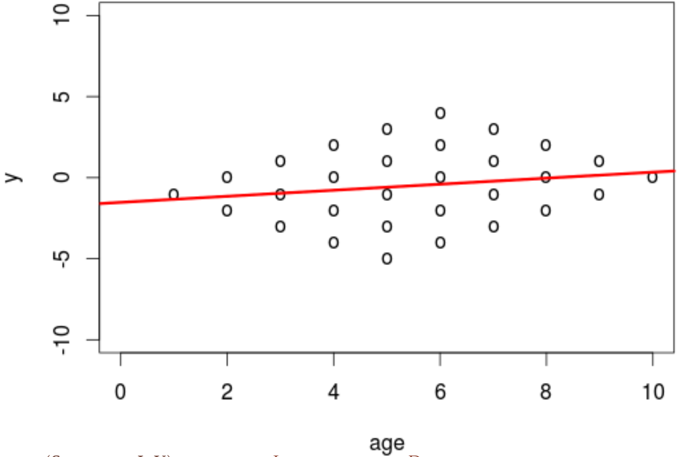
Varför är den longitudinella aspekten viktig?



Varför är den longitudinella aspekten viktig?



Varför är den longitudinella aspekten viktig?



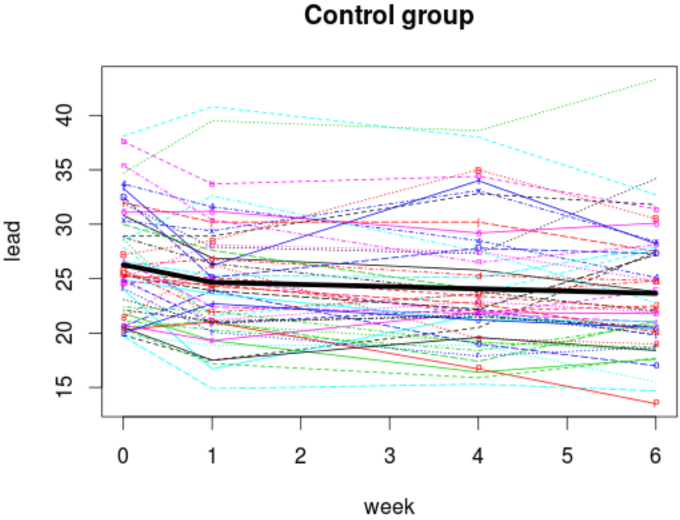
Longitudinella data, forts

- Egenskaper för autokorrelation i longitudinella data:
  - Positiv
  - Minskar med tidsavståndet mellan två observationer
  - Korrelation mellan mycket långa tidsavstånd är ofta skild från noll
  - Korrelation mellan mycket korta tidsavstånd är sällan nära ett
- Vanligt med missing data:
  - Saknade mättillfällen
  - Drop-outs
  - Överlevare
- Besläktade datatyper:
  - hierarkiska data (skolor med skolklasser med elever)
  - spatiala (rumsliga) data (huspriser i olika städer, miljödata)
  - tempo-spatiala data (månatliga mätningar av huspriser i olika städer)

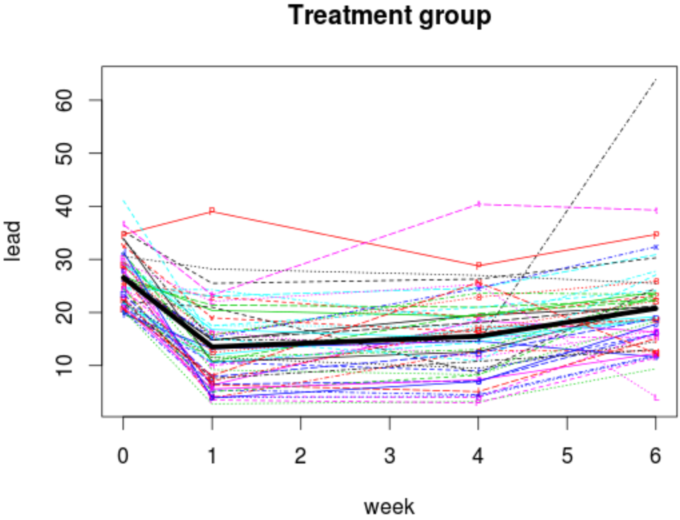
Hur man organiserar longitudinella data

```
fev <- read.table("../Data/LungFunctionGrowth.dat", header = TRUE)
fev[1:18, ]

   ID Height    Age InitialHeight InitialAge LogFEV1
1   1   1.20  9.341          1.20      9.341  0.2151
2   1   1.28 10.393          1.20      9.341  0.3716
3   1   1.33 11.452          1.20      9.341  0.4886
4   1   1.42 12.460          1.20      9.341  0.7514
5   1   1.48 13.418          1.20      9.341  0.8329
6   1   1.50 15.474          1.20      9.341  0.8920
7   1   1.52 16.372          1.20      9.341  0.8713
8   2   1.13  6.587          1.13      6.587  0.3075
9   2   1.19  7.650          1.13      6.587  0.3507
10  2   1.49 12.739          1.13      6.587  0.7561
11  2   1.53 13.774          1.13      6.587  0.8671
12  2   1.55 14.694          1.13      6.587  1.0473
13  2   1.56 15.822          1.13      6.587  1.1537
14  2   1.57 16.668          1.13      6.587  0.9243
15  2   1.57 17.632          1.13      6.587  1.1346
16  3   1.18  6.913          1.18      6.913  0.4318
17  3   1.23  7.975          1.18      6.913  0.3853
18  3   1.30  8.966          1.18      6.913  0.5988
```

Blymängder hos små barn - Kontrollgrupp



Blymängder hos små barn - Behandlingsgrupp



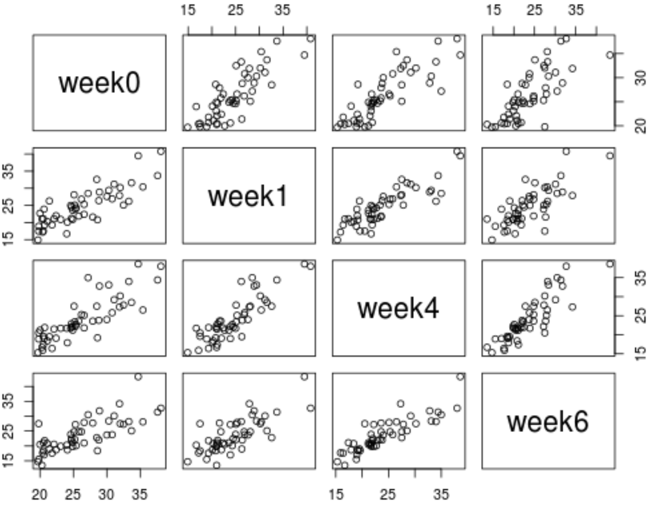
LONGITUDINELLA DATA ÄR MULTIVARIATA DATA

$$\underset{n_i \times 1}{Y_i} = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix}, \ i = 1, 2, ..., N.$$

- Kan modelleras med multivariate normalfördelning, och multivariat regression.
- Blymängder, kontrollgrupp:

$$Corr(Y) = \begin{pmatrix} 1 & 0.829 & 0.839 & 0.755 \\ \cdot & 1 & 0.860 & 0.759 \\ \cdot & \cdot & 1 & 0.869 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

LONGITUDINELLA DATA ÄR MULTIVARIATA DATA



Problem med direkt multivariat analys

- $Cov(Y)$ innehåller $T(T+1)/2$ fria parametrar, dvs många parametrar när T är stort.
- Missing data och drop-outs

$$Y_1 = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \end{pmatrix}, Y_2 = \begin{pmatrix} Y_{21} \\ NA \\ NA \\ NA \end{pmatrix}, Y_3 = \begin{pmatrix} Y_{31} \\ Y_{32} \\ NA \\ Y_{34} \end{pmatrix}$$

- Olika individer kan observeras vid olika tidpunkter.

Autokorrelerade mätningar är bra

- Intresse: förändringen mellan två tidpunkter. $\theta = \mu_2 - \mu_1$.
- Modell för tidpunkt 1 och 2

$$Y_1 \sim N(\mu_1, \sigma_1^2)$$
$$Y_2 \sim N(\mu_2, \sigma_2^2)$$

- Notera:

$$E(Y_2 - Y_1) = \mu_2 - \mu_1$$
$$Var(Y_2 - Y_1) = \sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2$$

- Estimator av förändringen mellan tidpunkterna:

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}(Y_{i2} - Y_{i1})$$

- Samplingvarians för $\hat{\theta}$

$$Var(\hat{\theta}) = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2}{N}$$

vilken är liten vid stark positiv (auto)korrelation.

Modell för väntevärdesprofiler

- **Väntevärdesprofilen**, **mean response profile**, över tiden för individ i:

$$E(Y_{ij}) = \beta_0 + \beta_1 \cdot t_{ij} + \beta_2 \cdot t_{ij}^2, \; i = 1, ..., N \text{ och } j = 1, ..., n_i$$

- Andra parametriska kurvor går också bra, t ex splines.
- Vi kan även ha en annan förklarande variabel X_1som är konstant över tiden (tidsinvariant):

$$E(Y_{ij}|X_i) = \beta_0 + \beta_1 \cdot t_{ij} + \beta_2 \cdot t_{ij}^2 + \beta_3 \cdot X_{i,1}$$

- Och vi kan ha en förklarande variabel som varierar över tid (tidsvariant)

$$E(Y_{ij}|X_i) = \beta_0 + \beta_1 \cdot t_{ij} + \beta_2 \cdot t_{ij}^2 + \beta_3 \cdot X_{i,1} + \beta_4 \cdot X_{ij,2}$$

Hur man organiserar longitudinella data

```
fev <- read.table("../Data/LungFunctionGrowth.dat", header = TRUE)
fev[1:18, ]

   ID Height    Age InitialHeight InitialAge LogFEV1
1   1   1.20  9.341          1.20      9.341  0.2151
2   1   1.28 10.393          1.20      9.341  0.3716
3   1   1.33 11.452          1.20      9.341  0.4886
4   1   1.42 12.460          1.20      9.341  0.7514
5   1   1.48 13.418          1.20      9.341  0.8329
6   1   1.50 15.474          1.20      9.341  0.8920
7   1   1.52 16.372          1.20      9.341  0.8713
8   2   1.13  6.587          1.13      6.587  0.3075
9   2   1.19  7.650          1.13      6.587  0.3507
10  2   1.49 12.739          1.13      6.587  0.7561
11  2   1.53 13.774          1.13      6.587  0.8671
12  2   1.55 14.694          1.13      6.587  1.0473
13  2   1.56 15.822          1.13      6.587  1.1537
14  2   1.57 16.668          1.13      6.587  0.9243
15  2   1.57 17.632          1.13      6.587  1.1346
16  3   1.18  6.913          1.18      6.913  0.4318
17  3   1.23  7.975          1.18      6.913  0.3853
18  3   1.30  8.966          1.18      6.913  0.5988
```

Modeller för väntevärdesprofiler, forts

- Vi kan skriva modellen in matrisform för $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{in})'$

$$E(Y_i|X_i) = \mu_i = X_i\beta$$

- Exempel: för modellen

$$E(y_{ij}|x) = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot x_{1,i} + \beta_4 \cdot x_{2,ij}$$

har vi

$$X_i = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 & x_{1,i} & x_{2,i1} \\ 1 & t_{i2} & t_{i2}^2 & x_{1,i} & x_{2,i1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & x_{1,i} & x_{2,in_i} \end{pmatrix}$$

- Notera: **multivariat regression** (multivariat respons) $\neq$ Multipel regression (en respons, flera förklarande variabler).

Väntevärdesprofiler med två grupper

- Kontrollgruppens väntevärdesprofil

$$E(y_{ij}|x_i) = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2$$

- Behandlingsgruppens väntevärdesprofil:

$$E(y_{ij}|x_i) = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) \cdot t + (\beta_2 + \beta_5) \cdot t^2$$

- Testa om behandlingen har någon som helst effekt: H_0: $\beta_3 = \beta_4 = \beta_5 = 0$. Vanligt F-test.

Väntevärdesprofiler med två grupper, forts

- Datamatris. Första personen kontroll, andra personen behandlad

$$X_i = \begin{pmatrix} 1 & t_{11} & t_{11}^2 & 0 & 0 & 0 \\ 1 & t_{12} & t_{12}^2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{1n_1} & t_{i1n_1}^2 & 0 & 0 & 0 \\ - & - & - & - & - & - \\ 1 & t_{21} & t_{21}^2 & 1 & t_{21} & t_{21}^2 \\ 1 & t_{22} & t_{22}^2 & 1 & t_{22} & t_{22}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{2n_2} & t_{2n_2}^2 & 1 & t_{2n_2} & t_{2n_2}^2 \\ - & - & - & - & - & - \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Väntevärdesprofiler med två grupper, forts

- R sätter upp X_i åt oss utifrån följande datamatris

$$Data = \begin{pmatrix} 1 & Y_{11} & t_{11} & T \\ 1 & Y_{12} & t_{12} & T \\ \vdots & \vdots & \vdots & \vdots \\ 1 & Y_{1n_1} & t_{1n_1} & T \\ - & - & - & - \\ 2 & Y_{21} & t_{21} & C \\ 2 & Y_{22} & t_{22} & C \\ \vdots & \vdots & \vdots & \vdots \\ 2 & Y_{2n_2} & t_{2n_2} & C \\ - & - & - & - \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

där den första kolumnen indikerar individ och sista kolumnen är en faktor-variabel som indikerar behandling (T) eller kontroll (C).

Estimation

- Modell
$$\underset{n_i \times 1}{Y_i} = \underset{n_i \times p}{X_i} \underset{p \times 1}{\beta} + \underset{n_i \times 1}{\epsilon_i}$$
där $\epsilon_i \overset{iid}{\sim} N(0, R_i)$
- Antag att R_i är kända.
- **Generalized Least Squares (GLS)**
$$\hat{\beta} = \left[\sum_{i=1}^{N} X_i' R_i^{-1} X_i\right]^{-1} \sum_{i=1}^{N} (X_i' R_i^{-1} y_i)$$
- Notera: att n_i kan variera över individerna. $X_i' R_i^{-1} X_i$ är alltid en $p \times p$ matris ocjh $X_i' R_i^{-1} y_i$ är en $p-$dimensional vektor.
- När R_i är okänd kan den ersättas med en skattning. Fortfarande konsistent skattning av β.
- Vi kan faktiskt sätta $R_i = \sigma^2 I$ och ändå få konsistenta skattningar. Men standardfelen för $\hat{\beta}$ blir inte rätt. Sandwich.

Funktionen gls i R-paketet nlme - Allmän

```
# Fitting quadratic mean response profiles with GLS model using different
# covariance structures. Child lead data - GLS
library(nlme)

# Reading data from file
leadData <- read.table("../Data/leadDataPP")

# General symmetric covariance structure
modelSym <- gls(lead ~ 1 + time * group + I(time^2) * group, data = leadData,
    correlation = corSymm(form = ~1 | id))
summary(modelSym)
```

```
Generalized least squares fit by REML
  Model: lead ~ 1 + time * group + I(time^2) * group
  Data: leadData
   AIC  BIC logLik
  2562 2614  -1268

Correlation Structure: General
 Formula: ~1 | id
 Parameter estimate(s):
 Correlation:
  1     2     3
2 0.236
3 0.592 0.615
4 0.427 0.529 0.526

Coefficients:
                  Value Std.Error t-value p-value
(Intercept)      22.458    0.7831  28.677  0.0000
time             -6.194    0.5610 -11.040  0.0000
groupP            3.333    1.1075   3.009  0.0028
I(time^2)         1.017    0.0965  10.536  0.0000
```

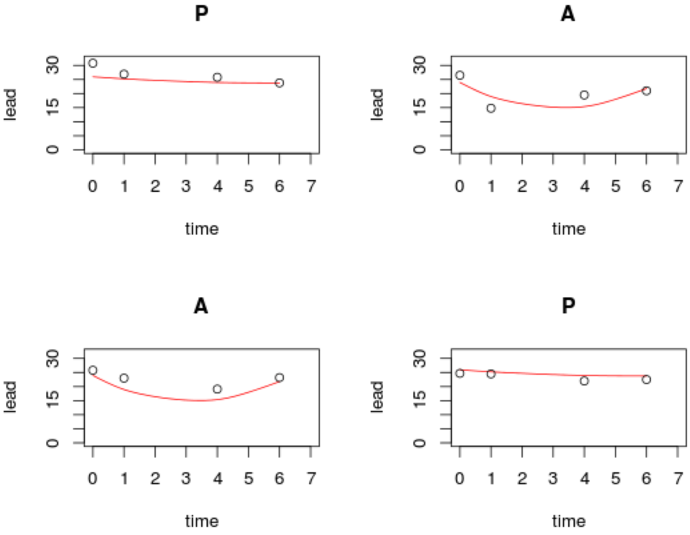
Funktionen gls i R-paketet nlme - Equi

```
# Fitting quadratic mean response profiles with GLS model using different
# covariance structures. Child lead data - GLS
library(nlme)

# Reading data from file
leadData <- read.table("../Data/leadDataPP")

# General symmetric covariance structure
modelSym <- gls(lead ~ 1 + time * group + I(time^2) * group, data = leadData,
    correlation = corCompSymm(form = ~1 | id))
summary(modelSym)


Generalized least squares fit by REML
  Model: lead ~ 1 + time * group + I(time^2) * group
  Data: leadData
   AIC  BIC logLik
  2573 2605  -1279

Correlation Structure: Compound symmetry
 Formula: ~1 | id
 Parameter estimate(s):
   Rho
0.5196

Coefficients:
                  Value Std.Error t-value p-value
(Intercept)      23.973    0.9267  25.870  0.0000
time             -7.541    0.6066 -12.432  0.0000
groupP            1.996    1.3105   1.523  0.1285
I(time^2)         1.196    0.0992  12.059  0.0000
time:groupP       6.624    0.8578   7.722  0.0000
groupP:I(time^2) -1.104    0.1403  -7.873  0.0000
```

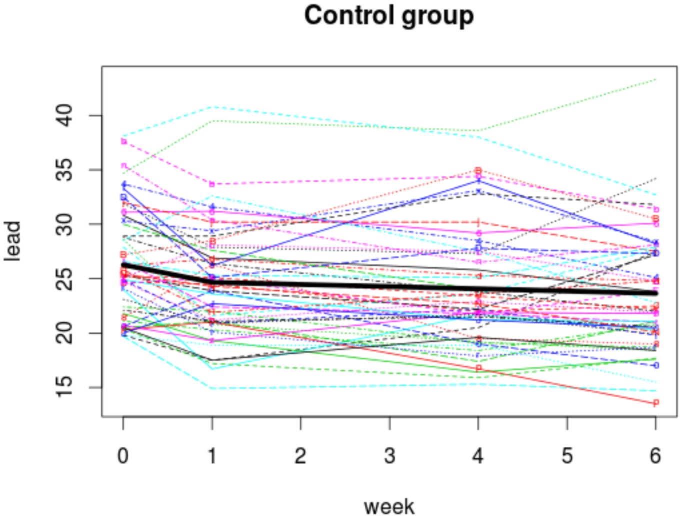
FUNKTIONEN GLS I R-PAKETET NLME - AR1

```
# Fitting quadratic mean response profiles with GLS model using different
# covariance structures. Child lead data - GLS
library(nlme)

# Reading data from file
leadData <- read.table("../Data/leadDataPP")

# General symmetric covariance structure
modelSym <- gls(lead ~ 1 + time * group + I(time^2) * group, data = leadData,
    correlation = corAR1(form = ~1 | id))
summary(modelSym)


Generalized least squares fit by REML
  Model: lead ~ 1 + time * group + I(time^2) * group
  Data: leadData
   AIC  BIC logLik
  2612 2643  -1298

Correlation Structure: AR(1)
 Formula: ~1 | id
 Parameter estimate(s):
   Phi
0.4914

Coefficients:
                  Value Std.Error t-value p-value
(Intercept)      23.911    0.9297  25.718  0.0000
time             -6.878    0.6746 -10.195  0.0000
groupP            2.051    1.3149   1.560  0.1196
I(time^2)         1.090    0.1062  10.266  0.0000
time:groupP       6.039    0.9540   6.330  0.0000
groupP:I(time^2) -1.011    0.1502  -6.732  0.0000
```

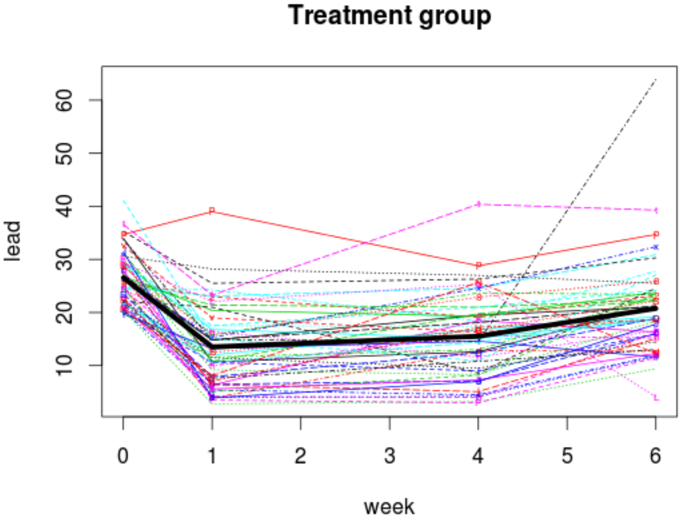
BLYMÄNGDER - FITTED VALUES FRÅN GLS EQUICORR



Blymängder hos små barn - Kontrollgrupp



Blymängder hos små barn - Behandlingsgrupp



Modeller för kovariansmatrisen

- $R_i = \sigma_i^2 I_n$. Oberoende observationer med olika varians för varje individ. Specialfall: $\sigma_i = \sigma^2$ för alla i.
- Equikorrelationsmodell

$$R = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots & \rho\sigma_1\sigma_n \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_1\sigma_n & \rho\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}$$

med korrelationsmatris

$$P = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

Modeller för kovariansmatrisen, forts

- Autoregressiv struktur

$$P = \begin{pmatrix} 1 & \rho^1 & \cdots & \rho^{n-1} \\ \rho^1 & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{pmatrix}$$

  där autokorrelationen avtar med tidsavståndet, t ex

$$Corr(Y_{i1}, Y_{i4}) = \rho^3$$

- Autoregressiv struktur för data med olika tid mellan observationstillfällen:

$$Corr(Y_{ij}, Y_{ik}) = \rho^{|t_{ij} - t_{ik}|}$$

Fixed effects och random effects

- Den vanliga linjära modellen (Fixed effects)

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \epsilon_{ij}$$

- Random intercept model:

$$Y_{ij} = (\beta_0 + b_i) + \beta_1 X_{ij} + \epsilon_{ij}$$

där $b_i \sim N(0, \sigma_b^2)$ är den individ-specifika delen av interceptet. Slumpmässigt.

- Marginell väntevärdesprofil

$$E(Y_{ij}|X_{ij}) = \beta_0 + \beta_1 X_{ij}$$

- Betingad väntevärdesprofil

$$E(Y_{ij}|X_{ij}, b_i) = (\beta_0 + b_i) + \beta_1 X_{ij}$$

Kovariansstruktur från random intercept

- Kovariansen för en random intercept modell är

$$Cov(Y_i) = \begin{pmatrix} \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \cdots & \sigma_b^2 \end{pmatrix} + R_i$$

- Om $R_i = \sigma^2 I_n$ ger detta en ekvi-korrelationsmatris med korrelationskoefficienten $\rho = \frac{\sigma_b^2}{\sigma^2 + \sigma_b^2}$.
- Ett slumpmässigt intercept ger varje individ dess eget intercept innebär:
  - observationerna är oberoende kring den betingade väntevärdesprofilen $(\beta_0 + b_i) + \beta_1 X_{ij}$
  - observationerna för en individ är beroende kring det marginella väntevärdet $\beta_0 + \beta_1 X_{ij}$. Autokorrelation genom random intercept.

Fixed effects och random effects

- Random slope

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{ij} + \epsilon_{ij}$$

  där $(b_{0i}, b_{1i}) \sim N_2(0, D)$ är den individ-specifika delen. Slumpmässigt.

- **General Linear Mixed Model** (GLMM)

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i$$

$$b_i \overset{iid}{\sim} N_q(0, D)$$

$$\varepsilon_i \overset{iid}{\sim} N_p(0, R_i)$$

VÄNTEVÄRDE OCH COVARIANS - GLMM

- Marginell väntevärdesprofil

$$E(Y_i|X_{ij}) = X_i\beta$$

- Betingad väntevärdesprofil

$$E(Y_i|X_i, b_i) = X_i\beta + Z_i b_i$$

- Kovariansmatris

$$\Sigma^{-1} = Cov(Y_i) = Z_i G Z_i' + R_i$$

- Notera 1: $Cov(Y_i)$ visar tydligt att variationen i data kan delas upp i
  - mellan-individsvariation $(Z_i G Z_i')$ och
  - inom-individsvariation (R_i).
- Notera 2: varianser och kovarianser för Y_i kan nu bero på förklarande variabler (Z_i). Linjär tidstrend i Z_i ger kvadratisk tidstrend i variansen.
- Notera 3: variablerna i Z_i bör även ingå i X_i.
- Notera 4: R_i kan parametriseras som tidigare, t ex via equi-korrelationmatris eller autoregressive struktur.

Krympning - Låna styrka

- GLMM modellen:
$$Y_i = X_i\beta + Z_i b_i + \epsilon_i$$
- Responsprofilen för den $i$te individen är
$$\hat{Y}_i = W_i \cdot (X_i\hat{\beta}) + (I - W_i) \cdot Y_i$$
där $\hat{\beta}$ är GLS skattning på populationsnivå och viktmatrisen W_i är
$$\underset{p\times p}{W_i} = \hat{R}_i(Z_i\hat{G}Z_i^{'} + \hat{R}_i)^{-1}$$
- Intuition:
$$\hat{Y}_i = \text{Vikt} \cdot \text{Populationsprofil} + \text{(1-Vikt)} \cdot \text{Observerad profil}$$
$$\text{Vikt} = \frac{\text{Variation inom individ}}{\text{Variation mellan individer} + \text{Variation inom individ}}$$
- Individer med svaga observationer (stort R_i) **lånar styrka** från populationen.

Modellval

- Hur väljer man bland modeller?
  - vilken modell för vänteväntevärdesprofilen?
  - vilken struktur på kovariansmatrisen?
  - Fixed eller random effects?
  - Vilka förklarande variabler?
  - etc etc
- Strategi: välj den modell som minimerar ett **informationskriterium**.
- AIC:

$$-2 \cdot \text{MaxLogLik} + 2 \cdot (\#\text{antal parametrar i modellen})$$

- BIC

$$-2 \cdot \text{MaxLogLik} + \ln N \cdot (\#\text{antal parametrar i modellen})$$

där MaxLogLik är log-likelihoodfunktionens maximum ($\ln L(\hat{\theta})$).

Longitudinell analys i R

- Flera paket att välja på, framförallt: **nlme** och **lme4**.
- nlme kan skatta linear mixed models med normalfördelade störningar (ϵ) och normalfördelade random effects (b_i).
- Många options, t ex olika strukturer på $R_i = Cov(\epsilon_{ij})$ och $D = Cov(b_i)$. T o m heteroscedastiska modeller för variansen är möjliga.
- lme4 liknar nlme, men har inte lika många kovarianstrukturer att välja på. Men lme4 kan skatta linear mixed models för data där responsen t ex är räknedata eller binär.
- lme4 kan alltså skatta en logistisk regression med fixed och random effects.
- SAS har PROC MIXED. Se Lindas föreläsningar om hierarkiska data.

Exempel - Longitudinell analys med nlme

```
# install.packages('nlme') # install package. uncomment if not installed.
library(nlme)  # load package
fev <- read.table("../Data/LungFunctionGrowth.dat", header = TRUE)
modelRandomSlopeAR1 <- lme(fixed = LogFEV1 ~ 1 + Age + log(Height) + InitialAge +
    log(InitialHeight), random = ~1 + Age | ID, data = fev, correlation = corAR1())
summary(modelRandomSlopeAR1)


Linear mixed-effects model fit by REML
 Data: fev
    AIC   BIC logLik
  -4588 -4532   2304

Random effects:
 Formula: ~1 + Age | ID
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev   Corr
(Intercept) 0.102919 (Intr)
Age         0.004731 -0.294
Residual    0.067148

Correlation Structure: AR(1)
 Formula: ~1 | ID
 Parameter estimate(s):
   Phi
0.3473
Fixed effects: LogFEV1 ~ 1 + Age + log(Height) + InitialAge + log(InitialHeight)
                     Value Std.Error   DF t-value p-value
(Intercept)        -0.2719   0.04248 1692   -6.40  0.0000
Age                 0.0225   0.00164 1692   13.71  0.0000
log(Height)         2.2546   0.05328 1692   42.31  0.0000
InitialAge         -0.0217   0.00819  297   -2.66  0.0083
log(InitialHeight)  0.3350   0.16040  297    2.09  0.0376
 Correlation:
```

BLYMÄNGD - RANDOM INTERCEPT MODELL

```r
# install.packages('nlme') # install package. uncomment if not installed.
library(nlme)  # load package

leadData <- read.table("../Data/leadDataPP")

# Fitting a random intercept model
modelRandomIntercept <- lme(fixed = lead ~ 1 + time * group + I(time^2) * group,
    random = ~1 | id, data = leadData, correlation = NULL)

modelRandomIntercept$coef$fixed


     (Intercept)             time           groupP       I(time^2)
          23.973           -7.541            1.996           1.196
     time:groupP groupP:I(time^2)
           6.624           -1.104


var(modelRandomIntercept$coef$random[[1]])


            (Intercept)
(Intercept)       19.97
```