Advanced Data Analysis

Overview of Longitudinal Data Analysis

Outline

- Introduction
- Simplest models
- Estimation
- More advanced models
- Inference
- Computing

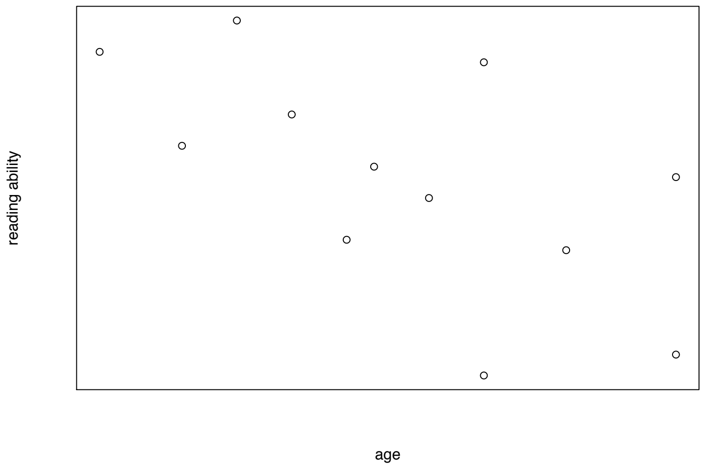
Outline

- Introduction
- Simplest models
- Estimation
- More advanced models
- Inference
- Computing

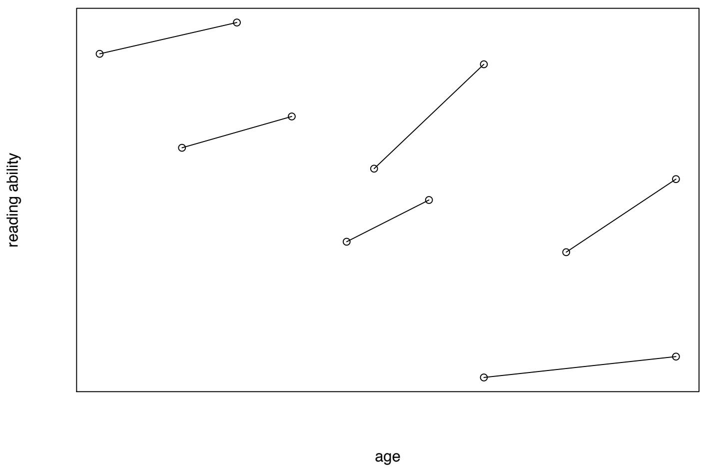
Longitudinal Studies

The defining characteristic of longitudinal studies is that subjects are measured repeatedly over time.

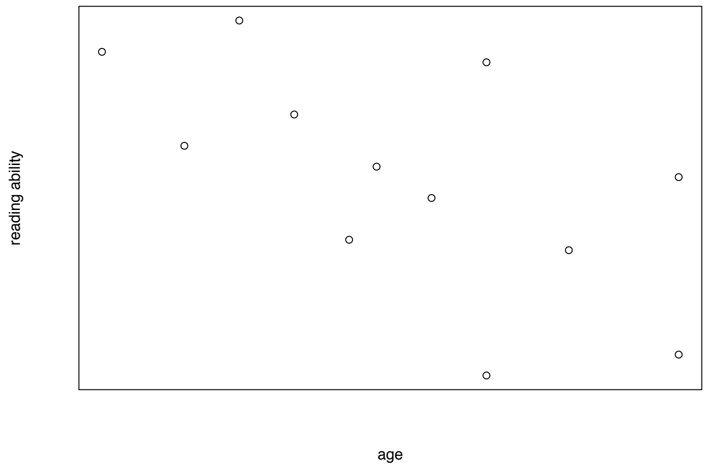
This is in contrast to cross-sectional studies where a single outcome is measured for each individual.


reading ability
age

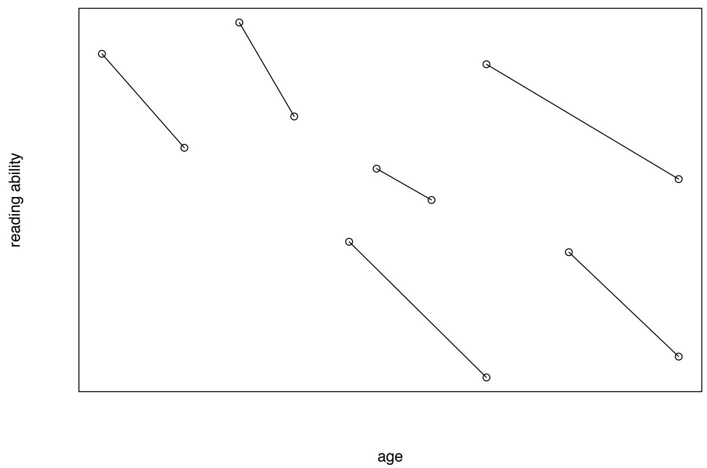
5


reading ability
age

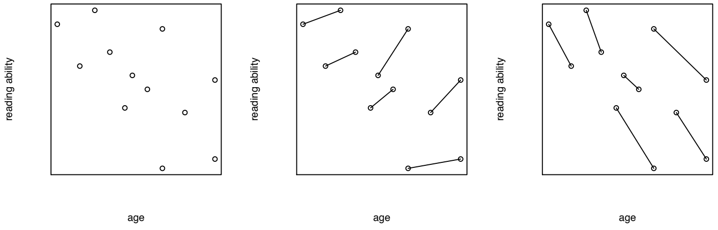
6


reading ability
age

7


reading ability
age

8



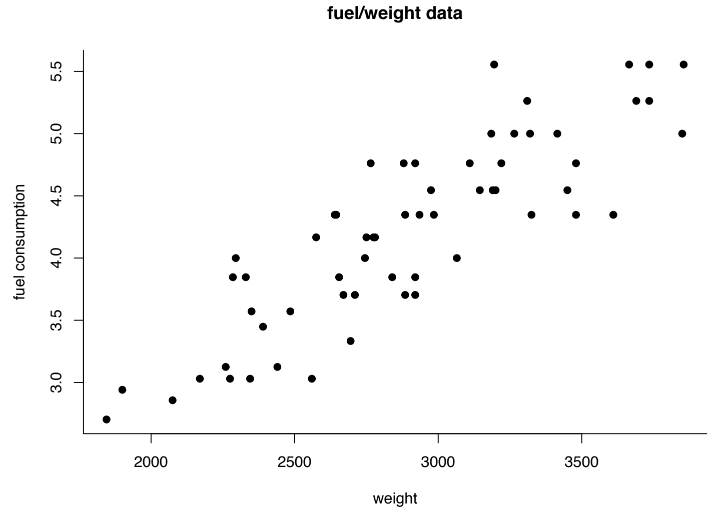
Cross-sectional versus Longitudinal

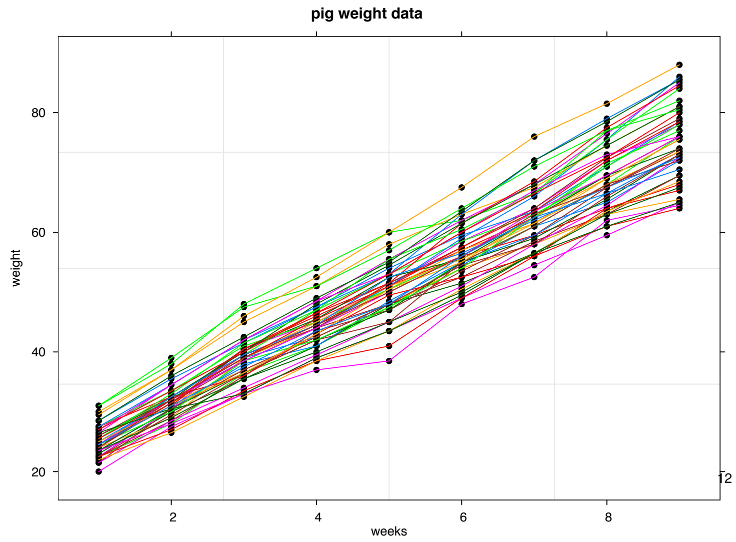
The next three slides show real data examples of

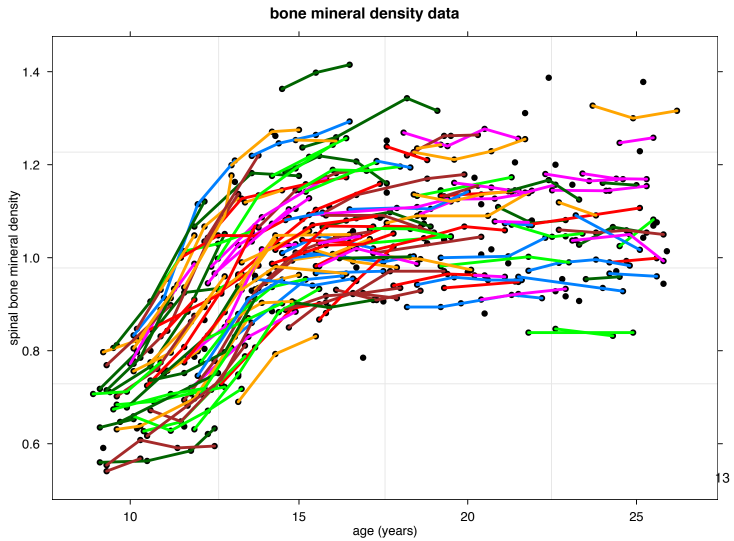
Cross-sectional data (1st slide)

and

Longitudinal data (2nd & 3rd slides)







Goals for This Course Segment

- What is a longitudinal design?
- Some of the basic models and their properties.
- Estimation of parameters in basic models.
- Some exposure to inference.
- Some exposure to computing.

Observations on Longitudinal Data Analysis

- Complicated due to within-subject correlation.
- Emerging area (most research post-1980; most widespread software post-1990).
- Software still fairly new and not extremely stable.
- Lot of dust still to settle on 'best practice'.

What Won't Be Achieved

Typical graduate courses on longitudinal data analysis have 45-60 lectures, several assignments, computer labs, projects etc.

These lectures (alone) will not train you comprehensively in the various nuances required for good applied longitudinal data analysis.

Rather, you will be exposed (to varying degrees) to some of the basic ideas and principles of longitudinal data analysis.

Books

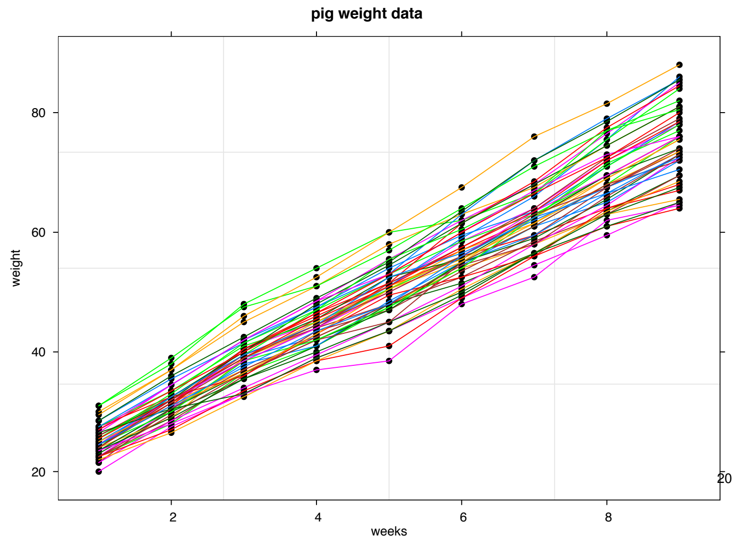
- Diggle, Heagerty, Liang & Zeger (applied)
- Fitzmaurice, Laird & Ware (applied)
- Pinhiero & Bates (applied; for `S-PLUS` & `R` users)
- Verbeke & Molenberghs (applied; for `SAS` users)
- McCulloch & Searle (theory; for algebra lovers)

Style of Delivery

- These slides ('big picture' stuff).
- Whiteboard interludes ('small picture' stuff).

Outline

- Introduction
- Simplest models
- Estimation
- More advanced models
- Inference
- Computing



Double Subscript Notation

The letter y is usually used for a generic response or outcome variable.

Let

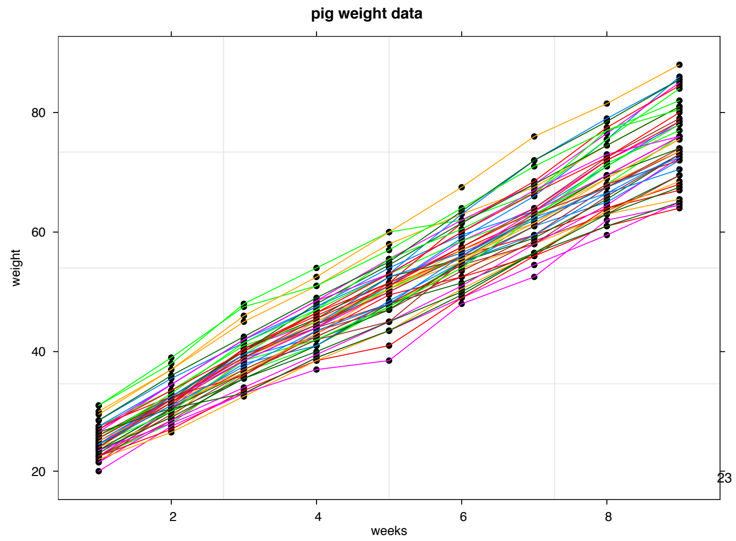
$$m = \text{number of subjects}$$
$$n_i = \text{number of measurements on subject } i \ (1 \le i \le m)$$

Then

$$y_{ij} = \text{response for } j\text{th measurement on subject } i.$$

Whiteboard Interlude I

This is to illustrate the ideas of double subscripting.



Naïve Model for Pig Weights

The slopes look about the same.

But each pig seems to have his/her own intercept

$$\Longrightarrow \quad y_{ij} = \beta_{0i} + \beta_1 x_{ij} + \varepsilon_{ij}$$

for $1 \le i \le 48$, with $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

But this model has 50 parameters!:

$$\beta_{01}, \beta_{02}, \ldots, \beta_{0,48}, \beta_1 \text{ and } \sigma_\varepsilon^2.$$

And only the last 2 are interpretable.

Random Intercept Model

An better model is:

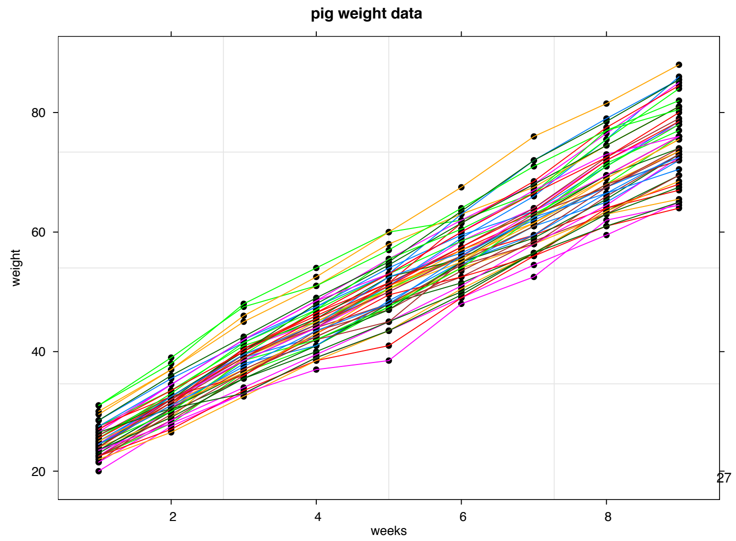
$$y_{ij} = U_i + \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

where

$$U_i \overset{\text{ind.}}{\sim} N(0, \sigma_U^2)$$

are independent of the

$$\varepsilon_{ij} \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon^2).$$



Random Effects

The U_i are called

random effects

By design they are centred around zero and correspond to the ith pig's devation from the 'average' intercept β_0.

Mixed Model Terminology

$$y_{ij} = \underbrace{U_i}_{\text{random effect}} + \underbrace{\beta_0 + \beta_1 x_{ij}}_{\text{fixed effects}} + \varepsilon_{ij}$$

The right-hand side has a mixture of random effects and fixed effects and so is called a

(LINEAR) MIXED MODEL

Whiteboard Interlude II

This is to describe the essential properties of the

random intercept model.

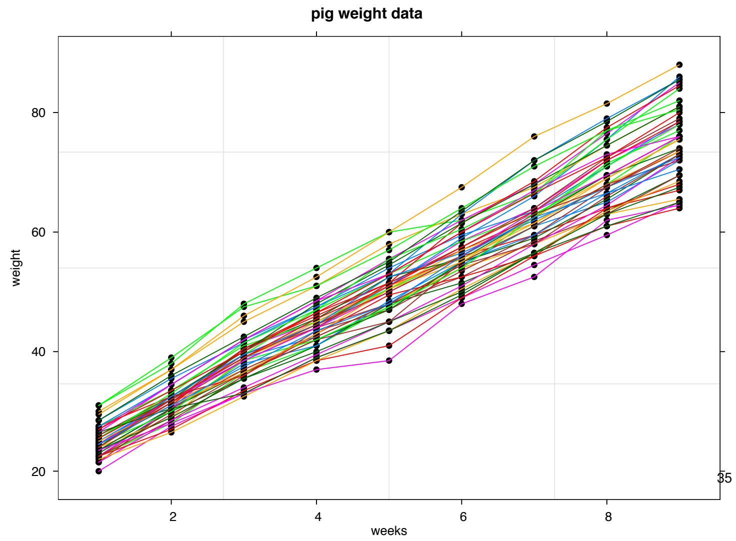
Remarks on Random Intercept Model

- Invokes a positive within-subject correlation $\rho = \sigma_U^2/(\sigma_U^2 + \sigma_\varepsilon^2)$.
- Correlation is same for all subjects and regardless of distance apart in time.
- This type of correlation structure is known as exchangeable correlation or compound symmetry.
- The β_0 and β_1 are called fixed effects. The U_i are called random effects.
- Since the model contains both fixed and random effects it is a special case of a mixed effects model; or mixed model for short.

- The parameters σ_U^2 and σ_ε^2 are often referred to as variance components.

Outline

- Introduction
- Simplest models
- Estimation
- More advanced models
- Inference
- Computing



Estimation of Model Parameters

The random intercept model:

$$y_{ij} = U_i + \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

has 4 parameters:

$$\beta_0, \beta_1, \sigma_u^2 \quad \text{and} \quad \sigma_\varepsilon^2.$$

What do computers do to find 'good values' for these parameters when fed a new data set?

Pig Weights Example

$$\begin{aligned} \widehat{\beta}_0 &= 19.36 \\ \widehat{\beta}_1 &= 6.21 \\ \widehat{\sigma}_u^2 &= 15.14 \\ \widehat{\sigma_\varepsilon}^2 &= 4.39 \\ \Longrightarrow \widehat{\rho} &= 15.14/(4.39+15.14) = 0.775 \end{aligned}$$

But how do computers get these numbers?

Answer

Essentially, they apply the principle of

Maximum Likelihood

The reason for the word 'essentially' is that often they use

Restricted Maximum Likelihood

Common abbreviations are:

ML and REML

Whiteboard Interlude III

To get warmed up for ML and REML for the random intercept model, we will first apply it to the simpler (more familiar?)

simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{\text{ind.}}{\sim} N(0, \sigma_\varepsilon^2).$$

Summary of ML and REML for Simple Linear Regression

ML for β_0 and β_1 leads to

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}_{\text{ML}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

ML for σ_ε^2 maximises

$$\ell_P(\sigma_\varepsilon^2) = -\tfrac{1}{2}\left[n\log(\sigma_\varepsilon^2) + (1/\sigma_\varepsilon^2)\text{RSS}\right] - \frac{n}{2}\log(2\pi).$$

$$\Longrightarrow \widehat{\sigma}_{\varepsilon\,\text{ML}}^{2} = \frac{1}{n}\text{RSS}$$

REML for σ_ε^2 maximises

$$\ell_P(\sigma_\varepsilon^2) - \tfrac{1}{2}\log\left|\frac{1}{\sigma_\varepsilon^2}\boldsymbol{X}^T\boldsymbol{X}\right|.$$

$$\Longrightarrow \widehat{\sigma}_{\varepsilon\,\text{REML}}^{2} = \frac{1}{n-2}\text{RSS}.$$

Whiteboard Interlude IV

This will discuss ML and REML estimation for the random intercept model

with parameters:

$$\beta_0,\ \beta_1,\ \sigma_U^2 \quad \text{and} \quad \sigma_\varepsilon^2.$$

Summary of ML/REML for Random Intercept Models

1. Set up the $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices.
2. Get REML estimates of σ_U^2 and σ_ε^2 by maximising

$$\ell_R(\sigma_U^2, \sigma_\varepsilon^2) = \ell_P(\sigma_U^2, \sigma_\varepsilon^2) - \tfrac{1}{2}\log|\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}|$$

where

$$\ell_P(\sigma_U^2, \sigma_\varepsilon^2) = -\tfrac{1}{2}\left[\log|\boldsymbol{V}| + \boldsymbol{y}^T\boldsymbol{V}^{-1}\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\}\boldsymbol{y}\right] - \frac{n}{2}\log(2\pi)$$

and

$$\boldsymbol{V} = \sigma_U^2\boldsymbol{Z}\boldsymbol{Z}^T + \sigma_\varepsilon^2\boldsymbol{I}$$

3. Set $\widehat{\boldsymbol{V}} = \widehat{\sigma}_u^2\boldsymbol{Z}\boldsymbol{Z}^T + \widehat{\sigma}_\varepsilon{}^2\boldsymbol{I}$ using the REML estimates of $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_\varepsilon{}^2$.
4.
$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = (\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{y}$$

This is exactly what the computer does to fit a random intercept model!

More General Models

Other longitudinal data analysis models just have different $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices and covariance matrices for $\boldsymbol{u}$.

This means that the algorithm on the previous slides requires only slight modification for other models.

Outline

- Introduction
- Simplest models
- Estimation
- More advanced models
- Inference
- Computing

More Advanced Models

We will spend a little bit of time on:

- Random slopes extension.
- AR(1) correlation.

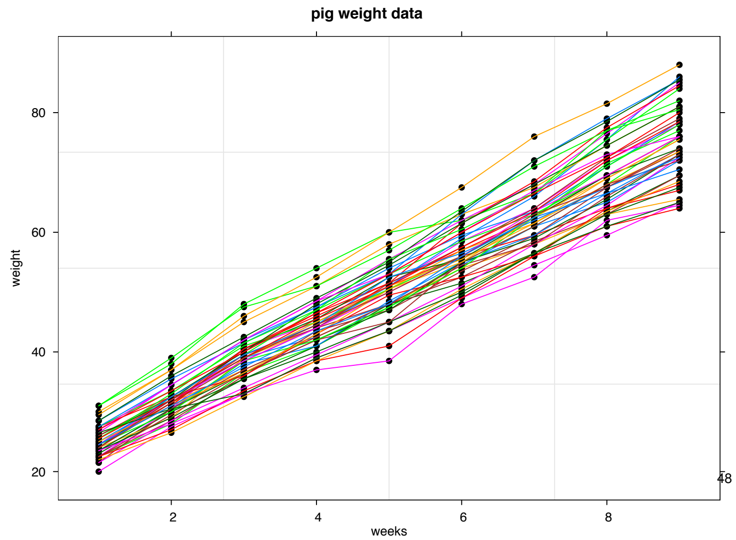
Random Slopes Extension

For the pig weight data we have so far considered the random intercept model:

$$y_{ij} = U_i + \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}, \quad U_i \overset{\text{ind.}}{\sim} N(0, \sigma_U^2)$$

This assumes that all pigs have the same slope β_1.

Is this reasonable?



Random Slopes Extension

A common extension is to allow for each pig to have his/her own slope:

$$y_{ij} = U_i + \beta_0 + (\beta_1 + V_i)\, x_{ij} + \varepsilon_{ij}.$$

$$\begin{bmatrix} U_i \\ V_i \end{bmatrix} \overset{\text{ind.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_U^2 & \rho_{uv}\sigma_U\sigma_V \\ \rho_{uv}\sigma_U\sigma_V & \sigma_V^2 \end{bmatrix} \right).$$

Pig Weights Example

$$\begin{aligned} \widehat{\beta}_0 &= 18.56 \\ \widehat{\beta}_1 &= 6.21 \\ \widehat{\sigma}_u^2 &= 7.00 \\ \widehat{\sigma}_v^2 &= 0.38 \\ \widehat{\sigma_\varepsilon}^2 &= 1.60 \end{aligned}$$

AR(1) Correlation

Anne has her systolic blood pressure (SBP) measured every Thursday morning for 8 consecutive weeks. If a random intercept model is imposed then the correlation matrix for Anne's responses is of the form:

	Wk 1	Wk 2	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8
Wk 1	1	ρ	ρ	ρ	ρ	ρ	ρ	ρ
Wk 2	ρ	1	ρ	ρ	ρ	ρ	ρ	ρ
Wk 3	ρ	ρ	1	ρ	ρ	ρ	ρ	ρ
Wk 4	ρ	ρ	ρ	1	ρ	ρ	ρ	ρ
Wk 5	ρ	ρ	ρ	ρ	1	ρ	ρ	ρ
Wk 6	ρ	ρ	ρ	ρ	ρ	1	ρ	ρ
Wk 7	ρ	ρ	ρ	ρ	ρ	ρ	1	ρ
Wk 8	ρ	ρ	ρ	ρ	ρ	ρ	ρ	1

for some $0 < \rho < 1$. As mentioned before this is called exchangeable correlation.

Limitation of Exchangeable Correlation

Exchangeable correlation says that:

Corr. between 1st & 2nd week

$=$ Corr. between 1st & 7th week

$=$ Corr. between 2nd & 5th week

$= \rho$

There is no time effect in the correlation.

AR(1) Correlation

The first-order autoregressive or AR(1) correlation structure leads to

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

AR(1) Correlation Structure

The AR(1) correlation structure is such that

Corr. of Anne's SBP k weeks apart $= \rho^k$.

This means that the correlation weakens for measurements further apart in time.

The first-order autoregression title comes from conditions such as:

$$\varepsilon_{ij} = \rho\varepsilon_{i,j-1} + \xi_{ij}.$$

Inference

- Standard deviations for the $\widehat{\beta}_i$s.
- Likelihood ratio tests.

Standard Deviations for the $\widehat{\beta}_i$s

Let $\widehat{\boldsymbol{\beta}}$ denote the vector of β_is $(1 \leq i \leq p)$ and

Cov$(\widehat{\boldsymbol{\beta}})$ = covariance matrix of $\widehat{\boldsymbol{\beta}}$.

Then

standard deviation of $\widehat{\beta}_i = \sqrt{i\text{th diagonal entry of Cov}(\widehat{\beta})}$.

Cov($\widehat{\beta}$) Results for Regression Models

For ordinary regression models the standard result is

$$\text{Cov}(\widehat{\beta}) = \sigma_\varepsilon^2 (\boldsymbol{X}^T\boldsymbol{X})^{-1}.$$

So

$$\widehat{\text{st.dev.}}(\widehat{\beta}_i) = \widehat{\sigma}_\varepsilon \sqrt{i\text{th diagonal entry of } (\boldsymbol{X}^T\boldsymbol{X})^{-1}}.$$

Cov($\widehat{\beta}$) Results for Longitudinal Models

For longitudinal models we have

$$\text{Cov}(\widehat{\beta}) = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}.$$

So

$$\widehat{\text{st.dev.}}(\widehat{\beta}_i) = \sqrt{i\text{th diagonal entry of } (\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}}.$$

The computer uses this for finding confidence intervals and p-values.

Likelihood Ratio Tests for Nested Models

Estimation for Normal response longitudinal models is simply (restricted) maximum liklihood.

Therefore the likelihood ratio paradigm may be used for hypothesis testing.

Likelihood Ratio Paradigm

H_0 : smaller model
versus
H_1 : larger model

Test statistic:

$$\lambda = -2\{\text{maximum log-likelihood under smaller model} \\ -\text{maximum log-likelihood under larger model}\}$$

Distribution under H_0

Classical theory says that under H_0

$$\lambda \sim \chi^2_k$$

where k is the difference in number of parameters between H_0 and H_1.

The assumptions of the classical theory don't hold in the longitudinal world due to lack of dependence and parameters lying on boundaries of parameters spaces. Current ongoing research is confronting these issues.

Books

- Diggle, Heagerty, Liang & Zeger (applied)
- Fitzmaurice, Laird & Ware (applied)
- Pinhiero & Bates (applied; for `S-PLUS` users)
- Verbeke & Molenberghs (applied; for `SAS` users)
- McCulloch & Searle (theory; for algebra lovers)

Outline

- Introduction
- Simplest models
- Estimation
- More advanced models
- Inference
- Computing

Computing

Since the early 1990s software packages have started to support longitudinal data analysis. Examples are:

- The `MIXED` and `NLMIXED` procedures in `SAS`.
- The `lme()` function in `R` and `S-PLUS`.
- The `Stata` package.
- The `SUDAAN` package.
- The `Genstat` package.

Live R Demonstration

Let's look at fitting of the random intercept in

the R function `lme()`