Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Mixed Effects Models

Applied Multivariate Statistics – Spring 2012

Overview

- Repeated Measures: Correlated samples
- Random Intercept Model
- Random Intercept and Random Slope Model

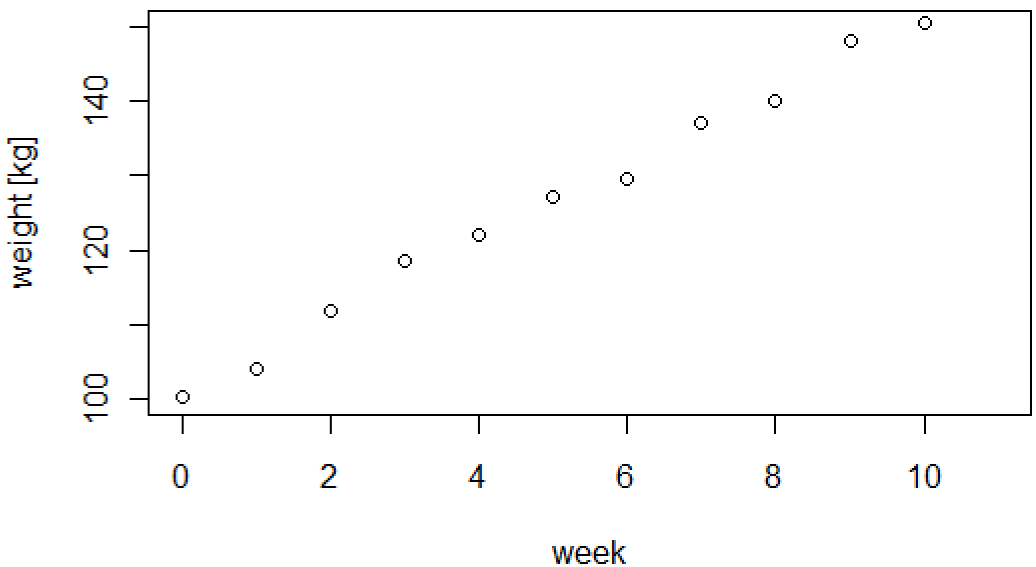
- Case studies

Revision: Linear Regression

- Example: Strength gain by weight training
- For one person:

$$y_j = \beta_0 + \beta_1 x_j + \epsilon_j \quad \epsilon_j \sim N(0, \sigma^2)\ i.i.d$$

"fixed" effects



Several Persons: Repeated Measures

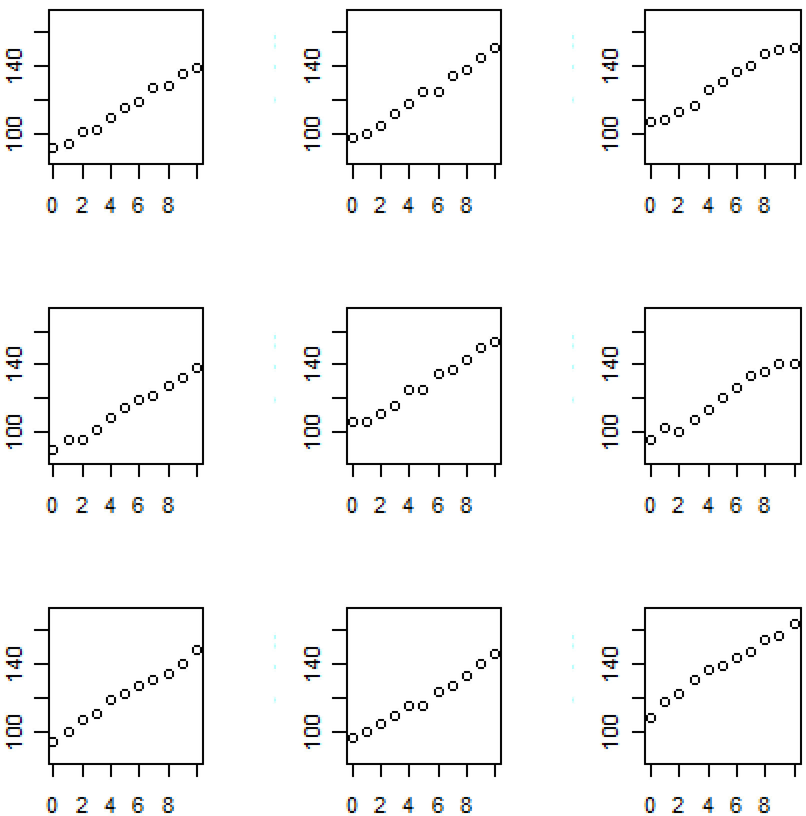
- Problem 1:
  Observations within persons are more correlated than observations between persons

- Problem 2:
  The parameters of each person might be slightly different

Weight Training revisited

Each person has

individual starting strength

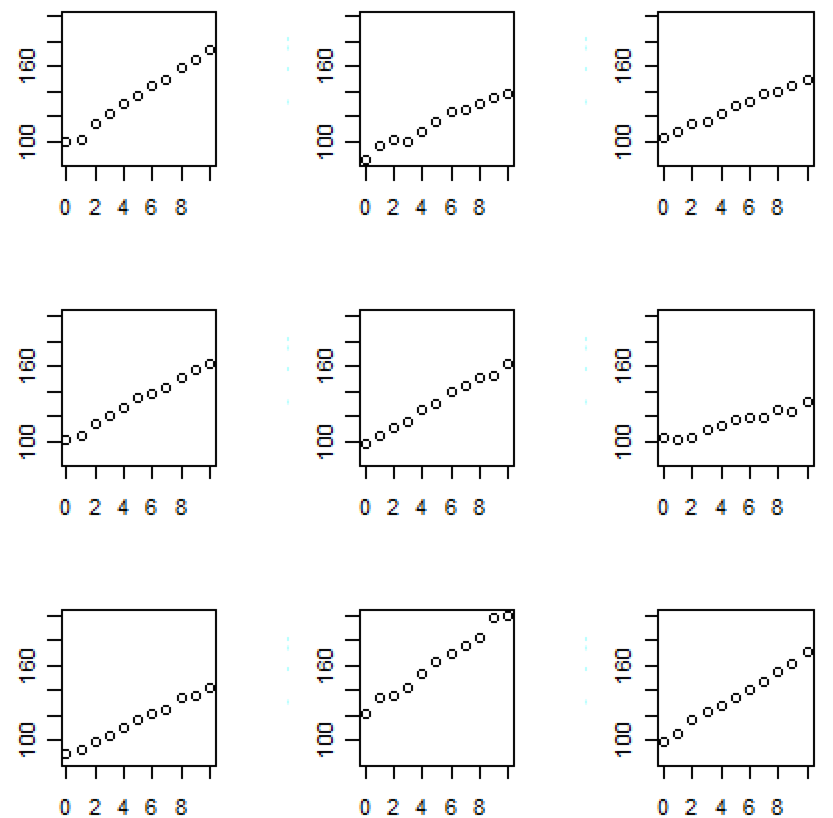


Each person has

individual starting strength

&

response to training



Dealing with repeated measures

- Alternative 1: Block effects

$$y_{ij} = (\beta_0 + \beta_{0,i}) + \beta_1 x_j + \epsilon_j \quad \epsilon_j \sim N(0, \sigma^2)\ i.i.d$$

"fixed" effects

Estimate: $\beta_0, \beta_{0,i}, \beta_1, \sigma$

Allows inference on individuals but not on population

- Alternative 2: Mixed effects (contains "fixed" and "random" effects)

E.g.: Random Intercept model

$$y_{ij} = (\beta_0 + u_i) + \beta_1 x_j + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma^2),\ u_i \sim N(0, \sigma_u^2)\ i.i.d$$

$$u_i, \epsilon_{ij}\ indep.$$

"random" effects

"fixed" effects

Fixed + Random

=

Mixed

Estimate: $\beta_0, \beta_1, \sigma, \sigma_u$

Allows inference on populations but not on individuals

Several Persons: Repeated Measures

- Problem 1:
  Observations within persons are more correlated than observations between persons

- Problem 2:
  The parameters of each person might be slightly different

Solved

Random Intercept Model implies correlated samples

- In Random Intercept Model, we do not explicitly model correlation of samples
- However, this is already implicitly captured in the model:

$$Var(Y_{ij}) = \sigma^2 + \sigma_u^2$$

$$Cov(Y_{ij}, Y_{ik}) = \sigma_u^2$$

$$Cov(Y_{ij}, Y_{lk}) = 0$$

- Within person, samples are correlated, between persons samples are uncorrelated

- Restriction: Correlation within person is the same for samples close or distant in time

Extending the Random Intercept Model: Random Intercept and Random Slope Model

$$y_{ij} = (\beta_0 + u_{i1}) + (\beta_1 + u_{i2})x_j + \epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0, \sigma^2),\ u_i \sim MVN(0, \Sigma)\ i.i.d$$

Estimate: $\beta_0, \beta_1, \sigma, \Sigma$

Similar calculations as before:

$$Var(Y_{ij}) = \sigma_1^2 + 2\sigma_{12}x_j + \sigma_2^2 x_j^2 + \sigma^2$$

$$Cov(Y_{ij}, Y_{ik}) = \sigma_1^2 + \sigma_{12}(x_j + x_k) + \sigma_2^2 x_j x_k$$

$$Cov(Y_{ij}, Y_{lk}) = 0$$

More complex correlations within person is possible

Several Persons: Repeated Measures

- Problem 1:
  Observations within persons are more correlated than observations between persons

  “Solved”

- Problem 2:
  The parameters of each person might be slightly different

  Solved

Summary of models for repeated measures

- Block effect (using fixed effects):
  Allows inference on individuals but not on population

- Mixed effects:
  Allows inference on population but not on individuals
  - Random Intercept:
    Individually varying intercept
    Models constant correlation within person
  - Random Intercept and Random Slope:
    Individually varying intercept and slope
    Models varying correlation within person

  More complex models possible, but harder to fit

Estimation of mixed effects models

- Maximum Likelihood (ML):
  - Variance estimates are biased
  + Tests between two models with differing fixed and random effects are possible

- Restricted Maximum Likelihood (REML):
  + Variance estimates are unbiased
  - Can only test between two models that have same fixed effects

Recommended

for

final model fit

(default in R)

- P-values etc. using asymptotic theory

Model diagnostics

- Residual analysis as in linear regression:
  - Tukey-Anscombe Plot
  - QQ-Plot of residuals

- Additionally: Predicted random effects must be normally distributed, therefore
  - QQ-Plots for random effects

Mixed effects models in R

- Function “lme” in package “nlme”
- Package “lme4” is a newer, improved version of package “nlme”, but to me, it still seems to be under construction and therefore is not so reliable

Interpretation of output 1/2

```
> fmw <- lme(weight ~ week, data = w, random = ~ 1 + week | pers)
> summary(fmw)
Linear mixed-effects model fit by REML
 Data: w
       AIC      BIC    logLik
  507.0283 522.4766 -247.5142

Random effects:
 Formula: ~1 + week | pers
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev   Corr
(Intercept) 9.725198 (Intr)
week        1.536847 0.426
Residual    1.965135

Fixed effects: weight ~ week
               Value Std.Error DF  t-value p-value
(Intercept) 99.86966  3.262722 89 30.60930       0
week         5.90099  0.516076 89 11.43435       0
 Correlation:
     (Intr)
week 0.408

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-2.653728335 -0.521019073 -0.008623998  0.591299144  2.577181144

Number of Observations: 99
Number of Groups: 9
```

$$y_{ij} = (\boxed{99.9} + u_{i1}) + (\boxed{5.9} + u_{i2})x_j + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \boxed{1.97^2}),\ \ u_i \sim MVN(0, \Sigma)\ i.i.d$$

with $\Sigma = \begin{pmatrix} 9.72^2 & 0.43 * 1.54^2 * 9.72^2 \\ 0.43 * 1.54^2 * 9.72^2 & 1.54^2 \end{pmatrix}$

Interpretation of output 2/2

- Using the function “intervals” for 95% confidence intervals:

At first meeting, people lift on ave. 100 kg (95%-CI: 93-106)

```
> intervals(fmw) ## fixed parameters of modes
Approximate 95% confidence intervals

 Fixed effects:
                lower      est.      upper
(Intercept) 93.386703 99.869663 106.352622
week         4.875554  5.900986   6.926417
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: pers
                           lower      est.      upper
sd((Intercept))        5.9201094 9.7251978 15.9759670
sd(week)               0.9346872 1.5368470  2.5269402
cor((Intercept),week) -0.2489383 0.4257207  0.8222167

 Within-group standard error:
   lower     est.    upper
1.684676 1.965135 2.292284
```

Per week people can lift 6 kg more (4.9-6.9)

The stand.dev. of weights in first week is 10 (6-16) kg

The stand.dev. in training progress is 1.5 (0.9-2.5) kg/week

There is no clear connection btw. weight in first week and training progress, since CI of correlation covers 0.

Typical deviation from fitted line is 2.0 (1.7-2.3) kg

Concepts to know

- Form of RI and RI&RS model and interpretation
- Model diagnostics

R functions to know

- Function “lme” in package “nlme”
  Functions:
  - “groupedData”, “lmList”
  - “intervals”, “coef”, “ranef”, “fixef”