Pre/post measurements

Nora Wickelmaier

Last modified: November 22, 2024

Analysis of longitudinal data

Advantages

- More power than cross sectional studies
- Each subject is its own control person
- Information about individual changes

Challenges

- Missing values
- Predictors can change over time
- In cross over designs: carry over effects

Analysis of longitudinal data

Advantages

- More power than cross sectional studies
- Each subject is its own control person
- Information about individual changes

Challenges

- Missing values
- Predictors can change over time
- In cross over designs: carry over effects

Today we will consider the simplest case with only two time points: pre and post score

Outline

- 1 t test
- 2 Change score analysis
- 3 Analysis of Covariance
- 4 Example: Acupuncture for shoulder pain
- 5 Linear mixed-effects model



• Straightforward analysis for two time points n=2

- Straightforward analysis for two time points n = 2
- Consider for each person the change score $d_i = y_{i2} y_{i1}$ and the linear model

$$d_i = \beta_0 + \varepsilon_i$$

$$y_{i2} = \beta_0 + y_{i1} + \varepsilon_i$$

- Straightforward analysis for two time points n = 2
- Consider for each person the change score $d_i = y_{i2} y_{i1}$ and the linear model

$$d_i = \beta_0 + \varepsilon_i$$

$$y_{i2} = \beta_0 + y_{i1} + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

• Interpretation:

 β_0 average change score

- Straightforward analysis for two time points n = 2
- Consider for each person the change score $d_i = y_{i2} y_{i1}$ and the linear model

$$d_i = \beta_0 + \varepsilon_i$$

$$y_{i2} = \beta_0 + y_{i1} + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

• Interpretation:

$$\beta_0$$
 average change score

• Hypothesis: Is there change between the first and second time point $(H_0: \beta_0 = 0)$?

- Straightforward analysis for two time points n = 2
- Consider for each person the change score $d_i = y_{i2} y_{i1}$ and the linear model

$$d_i = \beta_0 + \varepsilon_i$$

$$y_{i2} = \beta_0 + y_{i1} + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

• Interpretation:

 β_0 average change score

- Hypothesis: Is there change between the first and second time point $(H_0: \beta_0 = 0)$?
- Which effects are not considered by this?

- Straightforward analysis for two time points n = 2
- Consider for each person the change score $d_i = y_{i2} y_{i1}$ and the linear model

$$d_i = \beta_0 + \varepsilon_i$$

$$y_{i2} = \beta_0 + y_{i1} + \varepsilon_i$$

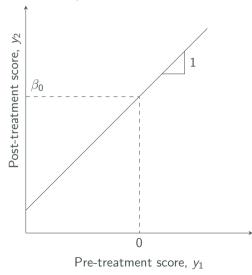
with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

• Interpretation:

$$\beta_0$$
 average change score

- Hypothesis: Is there change between the first and second time point $(H_0: \beta_0 = 0)$?
- Which effects are not considered by this?
- It is unclear if change is due to the treatment or just appeared over time

t test for comparing two time points



t test for comparing two time points

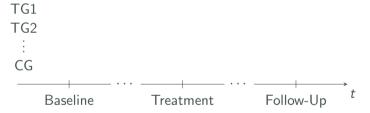
```
# simulate two dependent time points
cg \leftarrow MASS::mvrnorm(n = 50, mu = c(10, 10),
                     Sigma = matrix(c(1, .9, .9, 1), nrow = 2))
tg \leftarrow MASS::mvrnorm(n = 50, mu = c(10, 15),
                     Sigma = matrix(c(1, .9, .9, 1), nrow = 2))
sim <- as.data.frame(rbind(cg, tg))</pre>
names(sim) <- c("t1", "t2")
# add group variable
sim$group <- factor(rep(c("CG", "TG"), each = 50))</pre>
# group means
aggregate(cbind(t1, t2) ~ group, data = sim, FUN = mean)
```

t test for comparing two time points

```
# t. t.est.
t.test(sim$t2, sim$t1, paired = TRUE)
# add change score
sim$d <- sim$t2 - sim$t1
# linear model
lm1 \leftarrow lm(d \sim 1, data = sim)
lm2 \leftarrow lm(t2 \sim offset(t1), data = sim)
# visualization
plot(t2 ~ t1, data = sim)
abline(coef(lm2), 1)
```

Better design

• At least two groups are observed before (y_{i1}) and after (y_{i2}) a treatment



Research question:
 Do groups differ in the strength of their change?

• In regression notation with $x_i = 1$, if *i*th person is part of the treatment group, and else 0:

$$y_{i2} = \beta_0 + \beta_1 x_i + \varepsilon_i$$

• In regression notation with $x_i = 1$, if *i*th person is part of the treatment group, and else 0:

$$y_{i2} = \beta_0 + \beta_1 x_i + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

• Interpretation:

 eta_0 average follow-up score of reference group eta_1 effect of treatment group

• In regression notation with $x_i = 1$, if *i*th person is part of the treatment group, and else 0:

$$y_{i2} = \beta_0 + \beta_1 x_i + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

• Interpretation:

$$\beta_0$$
 average follow-up score of reference group β_1 effect of treatment group

• Hypothesis: Do groups differ at the second time point $(H_0: \beta_1 = 0)$?

• In regression notation with $x_i = 1$, if *i*th person is part of the treatment group, and else 0:

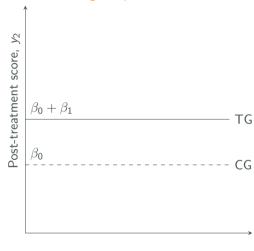
$$y_{i2} = \beta_0 + \beta_1 x_i + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

• Interpretation:

 β_0 average follow-up score of reference group β_1 effect of treatment group

- Hypothesis: Do groups differ at the second time point $(H_0: \beta_1 = 0)$?
- Since the baseline score is not considered, estimate of the change is biased



Pre-treatment score, y_1

```
# t test
t.test(t2 ~ group, data = sim, var.equal = TRUE)

# linear model
lm3 <- lm(t2 ~ group, data = sim)

# visualization
plot(t2 ~ t1, sim)
abline(h = cumsum(coef(lm3)))</pre>
```

Regression model

$$d_{i} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} - y_{i1} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} = \beta_{0} + y_{i1} + \beta_{1} x_{i} + \varepsilon_{i}$$

• Regression model

$$d_{i} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} - y_{i1} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} = \beta_{0} + y_{i1} + \beta_{1} x_{i} + \varepsilon_{i}$$

- Interpretation:
 - β_0 average change score for reference group
 - β_1 difference to β_0 in treatment group

• Regression model

$$d_{i} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} - y_{i1} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} = \beta_{0} + y_{i1} + \beta_{1} x_{i} + \varepsilon_{i}$$

- Interpretation:
 - β_0 average change score for reference group
 - β_1 difference to β_0 in treatment group
- Hypotheses:

• Regression model

$$d_{i} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} - y_{i1} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} = \beta_{0} + y_{i1} + \beta_{1} x_{i} + \varepsilon_{i}$$

- Interpretation:
 - β_0 average change score for reference group
 - β_1 difference to β_0 in treatment group
- Hypotheses:
 - Is there change in the reference group $(H_0: \beta_0 = 0)$?

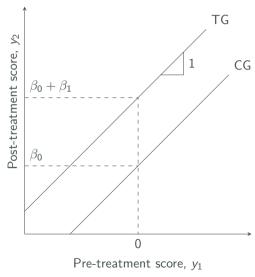
• Regression model

$$d_{i} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} - y_{i1} = \beta_{0} + \beta_{1} x_{i} + \varepsilon_{i}$$

$$y_{i2} = \beta_{0} + y_{i1} + \beta_{1} x_{i} + \varepsilon_{i}$$

- Interpretation:
 - β_0 average change score for reference group β_1 difference to β_0 in treatment group
- Hypotheses:
 - Is there change in the reference group $(H_0: \beta_0 = 0)$?
 - Does the change differ between groups $(H_0: \beta_1 = 0)$?



```
# regression model
lm4 <- lm(t2 ~ offset(t1) + group, data = sim)

# visualization
plot(t2 ~ t1, data = sim)
abline(coef(lm4)[1], 1)
abline(coef(lm4)[1] + coef(lm4)[2], 1)</pre>
```

Regression to the mean

- Change score analysis is based on the often too restrictive assumption that the follow-up score depends on the baseline score with a slope of 1
- Often baseline scores are negatively correlated with the change scores:
 Persons with low (bad) scores improve more than persons with high scores
- ullet This regression to the mean lets us expect a slope of <1 which has to be estimated from the data

3 Analysis of Covariance

Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

- Interpretation:
 - eta_0 average follow-up score in the reference group for $y_{i1}=0$
 - β_1 effect of baseline score
 - β_2 difference to follow-up score in the treatment group

Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

- Interpretation:
 - eta_0 average follow-up score in the reference group for $y_{i1}=0$
 - β_1 effect of baseline score
 - β_2 difference to follow-up score in the treatment group
- Hypotheses:

Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

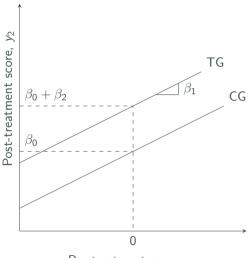
- Interpretation:
 - eta_0 average follow-up score in the reference group for $y_{i1}=0$
 - β_1 effect of baseline score
 - β_2 difference to follow-up score in the treatment group
- Hypotheses:
 - Is there a relationship between baseline and follow-up score $(H_0: \beta_1 = 0)$?

Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

- Interpretation:
 - β_0 average follow-up score in the reference group for $y_{i1} = 0$
 - β_1 effect of baseline score
 - β_2 difference to follow-up score in the treatment group
- Hypotheses:
 - Is there a relationship between baseline and follow-up score $(H_0: \beta_1 = 0)$?
 - Do the follow-up scores in the groups differ for persons with identical baseline scores $(H_0: \beta_2 = 0)$?

Analysis of Covariance (ANCOVA)



Pre-treatment score, y_1

Adjusted means

- With an ANCOVA model we can predict average follow-up scores for persons with the same baseline score
- For example, we get

$$\hat{y}_{i2} = \hat{\beta}_0 + \hat{\beta}_1 \, \bar{y}_1 + \hat{\beta}_2 \, x_i$$

for an average baseline score \bar{y}_1

• These conditional means are sometimes called (baseline) adjusted means

Analysis of Covariance (ANCOVA)

```
# ancova
lm5 \leftarrow lm(t2 \sim t1 + group, data = sim)
# adjusted means
predict(lm5, newdata = data.frame(t1 = mean(sim$t1),
                                     group = c("CG", "TG"))
# ancova with change score
lm5a \leftarrow lm(d \sim t1 + group, data = sim)
# visualization
plot(t2 ~ t1, data = sim)
abline(coef(lm5)[1], coef(lm5)[2])
abline(coef(lm5)[1] + coef(lm5)[3], coef(lm5)[2])
```

Analysis of Covariance with change score

Regression model

$$d_{i} = \beta_{0} + \beta_{1} y_{i1} + \beta_{2} x_{i} + \varepsilon_{i}$$

$$y_{i2} - y_{i1} = \beta_{0} + \beta_{1} y_{i1} + \beta_{2} x_{i} + \varepsilon_{i}$$

$$y_{i2} = \beta_{0} + (1 + \beta_{1}) y_{i1} + \beta_{2} x_{i} + \varepsilon_{i}$$

• For testing the difference of change in the groups (β_2) it is irrelevant if the dependent variable is the follow-up score or the change score

• Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \beta_3 (y_{i1} \cdot x_i) + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \beta_3 (y_{i1} \cdot x_i) + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

- Interpretation:
 - eta_0 average follow-up score in reference group for $y_{i1}=0$
 - β_1 effect of baseline score
 - β_2 difference to follow-up score in treatment group
 - β_3 difference between slopes in reference and treatment groups

Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \beta_3 (y_{i1} \cdot x_i) + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

- Interpretation:
 - β_0 average follow-up score in reference group for $y_{i1} = 0$
 - β_1 effect of baseline score
 - β_2 difference to follow-up score in treatment group
 - β_3 difference between slopes in reference and treatment groups
- Hypothesis:

Does effect of baseline score depend on group ($H_0: \beta_3 = 0$)?

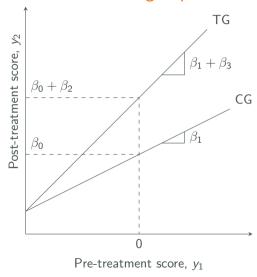
Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \beta_3 (y_{i1} \cdot x_i) + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.

- Interpretation:
 - β_0 average follow-up score in reference group for $y_{i1}=0$
 - eta_1 effect of baseline score
 - β_2 difference to follow-up score in treatment group
 - β_3 difference between slopes in reference and treatment groups
- Hypothesis:
 - Does effect of baseline score depend on group (H_0 : $\beta_3 = 0$)?
- Interpretation of adjusted means independently of baseline score implies $\beta_3 = 0$

Interaction between baseline score and group



Interaction between baseline score and group

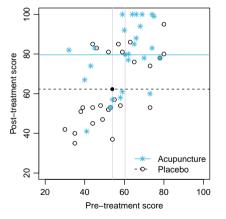
```
# regression model with interaction
lm6 \leftarrow lm(t2 \sim t1 * group, data = sim)
# visualization
plot(t2 ~ t1, data = sim)
abline(coef(lm6)[1], coef(lm6)[2])
abline(coef(lm6)[1] + coef(lm6)[3], coef(lm6)[2] + coef(lm6)[4])
# models are nested
anova (1m4, 1m5, 1m6)
```

4 Example: Acupuncture for shoulder pain

Example: Acupuncture for shoulder pain

- Kleinhenz et al. (1999) investigate the effect of acupuncture on the improvement of mobility for 52 patients with shoulder pain
- Patients are randomly assigned to two groups (placebo vs. acupuncture)
- Before and after the treatment a mobility score is measured
- Vickers and Altman (2001) show advantages of an analysis of covariance compared to other methods based on these data

Acupuncture: Follow-up analysis

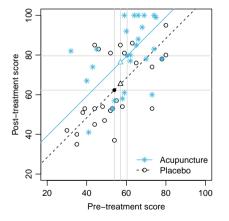


	Pla	Acu	Diff
Baseline	53.9	60.4	6.5
Follow-up	62.3	79.6	17.3
Change sc.	8.4	19.2	10.8
ANCOVA			12.7

$$y_{i2} = \beta_0 + \beta_1 x_i + \varepsilon_i$$

 $\hat{\beta}_1 = 17.3, 0.95-CI: (7.5, 27.1)$

Acupuncture: Change score analysis

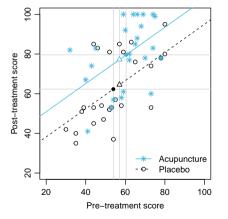


	Pla	Acu	Diff
Baseline	53.9	60.4	6.5
Follow-up	62.3	79.6	17.3
Change sc.	8.4	19.2	10.8
ANCOVA			12.7

$$y_{i2} = \beta_0 + y_{i1} + \beta_1 x_i + \varepsilon_i$$

 $\hat{\beta}_1 = 10.8, 0.95-CI: (2.3, 19.4)$

Acupuncture: ANCOVA



	Pla	Acu	Diff
Baseline	53.9	60.4	6.5
Follow-up	62.3	79.6	17.3
Change sc.	8.4	19.2	10.8
ANCOVA			12.7

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

 $\hat{\beta}_2 = 12.7, 0.95$ -CI: (4.1, 21.3)

Example: Acupuncture for shoulder pain

```
# read data
dat <- read.table("kleinhenz.txt", header = TRUE)</pre>
dat$grp <- factor(dat$grp, levels = c("plac", "acu"))</pre>
# follow-up analysis
m1 <- lm(post ~ grp, data = dat)
summary(m1)
confint(m1)
# change score analysis
m2 <- lm(post ~ offset(pre) + grp, data = dat)</pre>
# ancova
m3 <- lm(post ~ pre + grp, data = dat)
```

Example: Acupuncture for shoulder pain

```
# testing if slopes differ
m4 <- lm(post ~ pre * grp, data = dat)
anova(m3, m4)
# adjusted means
predict(m3, newdata = data.frame(pre = mean(dat$pre),
                                 grp = c("plac", "acu")))
# visualization
plot(post ~ pre, data = dat, xlim = c(20,100), ylim = c(20,100))
abline(coef(m3)[1], coef(m3)[2])
abline(coef(m3)[1] + coef(m3)[3], coef(m3)[2])
```

 We considered the basic case when several groups are observed at two time points (baseline and follow-up)

- We considered the basic case when several groups are observed at two time points (baseline and follow-up)
- Analysis of these kind of data considers the difference

$$d_i = y_{i2} - y_{i1}$$

or the adjusted follow-up scores which are considered to be independent

- We considered the basic case when several groups are observed at two time points (baseline and follow-up)
- Analysis of these kind of data considers the difference

$$d_i = y_{i2} - y_{i1}$$

or the adjusted follow-up scores which are considered to be independent

Analysis of Covariance

- We considered the basic case when several groups are observed at two time points (baseline and follow-up)
- Analysis of these kind of data considers the difference

$$d_i = y_{i2} - y_{i1}$$

or the adjusted follow-up scores which are considered to be independent

- Analysis of Covariance
 - has the highest power to detect differences for average change compared to the other methods

- We considered the basic case when several groups are observed at two time points (baseline and follow-up)
- Analysis of these kind of data considers the difference

$$d_i = y_{i2} - y_{i1}$$

or the adjusted follow-up scores which are considered to be independent

- Analysis of Covariance
 - has the highest power to detect differences for average change compared to the other methods
 - must be cautiously interpreted when groups are not randomly assigned

• Often people prefer the view of comparing the time effects within the groups instead of the adjusted follow-up scores

- Often people prefer the view of comparing the time effects within the groups instead of the adjusted follow-up scores
- Regression notation with indicator variable for time $(t_{i1}=0,\ t_{i2}=1)$ and group (x_i)

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + \beta_3 (t_{ij} \cdot x_i) + v_i + \varepsilon_{ij}$$

with $v_i \sim N(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$

- Often people prefer the view of comparing the time effects within the groups instead of the adjusted follow-up scores
- Regression notation with indicator variable for time $(t_{i1}=0,\ t_{i2}=1)$ and group (x_i)

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + \beta_3 (t_{ij} \cdot x_i) + v_i + \varepsilon_{ij}$$

with $v_i \sim N(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$

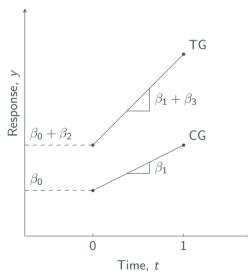
• Interpretation:

 β_0 mean baseline value in reference group

 β_1 time effect (slope) in reference group

 β_2 effect of treatment group

general effect on the slope of treatment group



t test

• For the two groups with j = 1, 2, we get

$$y_{i1} = \beta_0 + \beta_2 x_i + v_i + \varepsilon_{i1}$$

$$y_{i2} = \beta_0 + \beta_1 + \beta_2 x_i + \beta_3 x_i + v_i + \varepsilon_{i2}$$

• For the two groups with j = 1, 2, we get

$$y_{i1} = \beta_0 + \beta_2 x_i + v_i + \varepsilon_{i1}$$

$$y_{i2} = \beta_0 + \beta_1 + \beta_2 x_i + \beta_3 x_i + v_i + \varepsilon_{i2}$$

For the change score, we then get

$$y_{i2} - y_{i1} = \beta_1 + \beta_3 x_i + (\varepsilon_{i2} - \varepsilon_{i1})$$

• For the two groups with j = 1, 2, we get

$$y_{i1} = \beta_0 + \beta_2 x_i + v_i + \varepsilon_{i1}$$

$$y_{i2} = \beta_0 + \beta_1 + \beta_2 x_i + \beta_3 x_i + v_i + \varepsilon_{i2}$$

• For the change score, we then get

$$y_{i2} - y_{i1} = \beta_1 + \beta_3 x_i + (\varepsilon_{i2} - \varepsilon_{i1})$$

• Since ε_{ij} are independent, this results in the equation for the change score analysis

$$y_{i2} - y_{i1} = \beta_1 + \beta_3 x_i + \varepsilon_i$$

with

$$\varepsilon_i = \varepsilon_{i2} - \varepsilon_{i1} \sim N(0, \sigma_d^2 = 2\sigma^2)$$

• For the two groups with i = 1, 2, we get

$$y_{i1} = \beta_0 + \beta_2 x_i + v_i + \varepsilon_{i1}$$

$$y_{i2} = \beta_0 + \beta_1 + \beta_2 x_i + \beta_3 x_i + v_i + \varepsilon_{i2}$$

• For the change score, we then get

$$y_{i2} - y_{i1} = \beta_1 + \beta_3 x_i + (\varepsilon_{i2} - \varepsilon_{i1})$$

• Since ε_{ij} are independent, this results in the equation for the change score analysis

$$y_{i2} - y_{i1} = \beta_1 + \beta_3 x_i + \varepsilon_i$$

with

$$\varepsilon_i = \varepsilon_{i2} - \varepsilon_{i1} \sim N(0, \sigma_d^2 = 2\sigma^2)$$

→ LMM for two time points is equivalent to change score analysis!

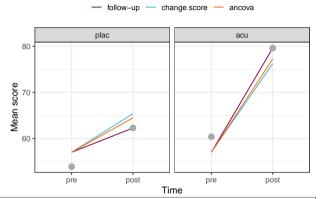
Linear mixed-effects model

Example: Acupuncture for shoulder pain

```
# read data
dat <- read.table("kleinhenz.txt", header = TRUE)</pre>
dat$grp <- factor(dat$grp, levels = c("plac", "acu"))</pre>
# change score analysis
m1 <- lm(post ~ offset(pre) + grp, data = dat)
# T.MM
datl <- reshape(dat, direction = "long",</pre>
                 varying = list(1:2), v.names = "score")
datl$time <- factor(datl$time, levels = 1:2, labels = c("pre", "post"))
m2 <- lmer(score ~ grp * time + (1 | id), data = datl)</pre>
# compare residual variances
sigma(m1)^2
2 * sigma(m2)^2
```

"Adjusted" time effects

 The ANCOVA model will give you adjusted time effects within the groups with a little reparametrization



```
m5 <- lm(post - pre ~ scale(pre) + grp, data = dat)
predict(m5, data.frame(pre = mean(dat$pre), grp = c("plac", "acu"))))</pre>
```

References

Kleinhenz, J., Streitberger, K., Windeler, J., Güßbacher, A., Mavridis, G., & Martin, E. (1999).Randomised clinical trial comparing the effects of acupuncture and a newly designed placebo needle in rotator cuff tendinitis. *Pain*, *83*(2), 235–241.

Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ*, 323(7321), 1123–1124.