Mandy Vogel

University Leipzig

October 4, 2015

Overview

Mixed or Multilevel Models

Table of Contents I

load the data

- load the data using `load()`, the file name is gewichte.rdata

Data

- person id (personennr)
- sex (geschlecht: 0 - male, 1 - female)
- weight in kg (wert)
- gestational age (ssw)
- age in years (age)

```
> head(gewichte)
  personennr geschlecht wert ssw        age
1      37766          1 4.40  37 0.13141684
2        808          1 3.01  39 0.03011636
3        810          0 7.80  40 0.52566735
5      37766          1 5.40  37 0.25462012
7        927          1 5.32  41 0.24366872
8        908          1 6.20  37 0.32306639
```

summarize the data

- how many rows, how many columns?
- how many persons?
- what is the min/max/mean age/weight/gestational age?

Data

- we want to model the weight dependent on... (fixed effects)
- looking at the data frame and the question: what could be a random effect variable in the model

Fitting with `lmer`

- first we fit a varying intercept model with no predictors (a null model) - so the only fixed effect estimate will be the overall mean
- is there any meaningful interpretation to the model?

Fitting with lmer

- first we fit a varying intercept model with no predictors (a null model) - so the only fixed effect estimate will be the overall mean

```
> m0 <- lmer(wert ~ 1 + (1 | personennr), data = gewichte)
> m0
Linear mixed model fit by REML ['lmerMod']
Formula: wert ~ 1 + (1 | personennr)
   Data: gewichte
REML criterion at convergence: 6070.36
Random effects:
 Groups     Name        Std.Dev.
 personennr (Intercept) 0.4468
 Residual               2.3678
Number of obs: 1321, groups:  personennr, 199
Fixed Effects:
(Intercept)
      5.894
```

Fitting with lmer

- the display() function (arm package) provides a more convenient way to display the more important information

```
> display(m0)
lmer(formula = wert ~ 1 + (1 | personennr), data = gewichte)
coef.est  coef.se
    5.89      0.07

Error terms:
 Groups     Name        Std.Dev.
 personennr (Intercept) 0.45
 Residual               2.37
---
number of obs: 1321, groups: personennr, 199
AIC = 6076.4, DIC = 6063.6
deviance = 6067.0
```

Fitting with `lmer`

- now we add a fixed effect: `age`
- the first part is the fixed effect (age) and after that we add `(1 | personennr)` which allows the intercept to vary by child.

```
> m1 <- lmer(wert ~ age + (1 | personennr), data = gewichte)
```

Fitting with lmer

```
> display(m1)
lmer(formula = wert ~ age + (1 | personennr), data = gew
            coef.est coef.se
(Intercept) 3.63     0.06
age         6.81     0.06

Error terms:
 Groups     Name        Std.Dev.
 personennr (Intercept) 0.82
 Residual               0.69
---
number of obs: 1321, groups: personennr, 199
AIC = 3246, DIC = 3223
deviance = 3230.5
```

- interpret the fixed effects!

Compare to `lm()`

- just to see the difference to the linear model:

```
> summary(m.lm)

Call:
lm(formula = wert ~ age, data = gewichte)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8861 -0.6445 -0.0534  0.6112  4.7646

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.60999    0.04366   82.68   <2e-16 ***
age          6.80105    0.09469   71.82   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.088 on 1319 degrees of freedom
Multiple R-squared:  0.7964, Adjusted R-squared:  0.7962
F-statistic:  5159 on 1 and 1319 DF,  p-value: < 2.2e-16
```

Mixed Models with R

- now we take a closer look at the mixed model:

```
> coef(m1)
$personennr
    (Intercept)      age
1      3.319122 6.811165
2      4.462539 6.811165
3      5.129777 6.811165
4      3.761104 6.811165
5      3.906263 6.811165
6      3.840744 6.811165
7      3.787864 6.811165
8      3.263762 6.811165
....
```

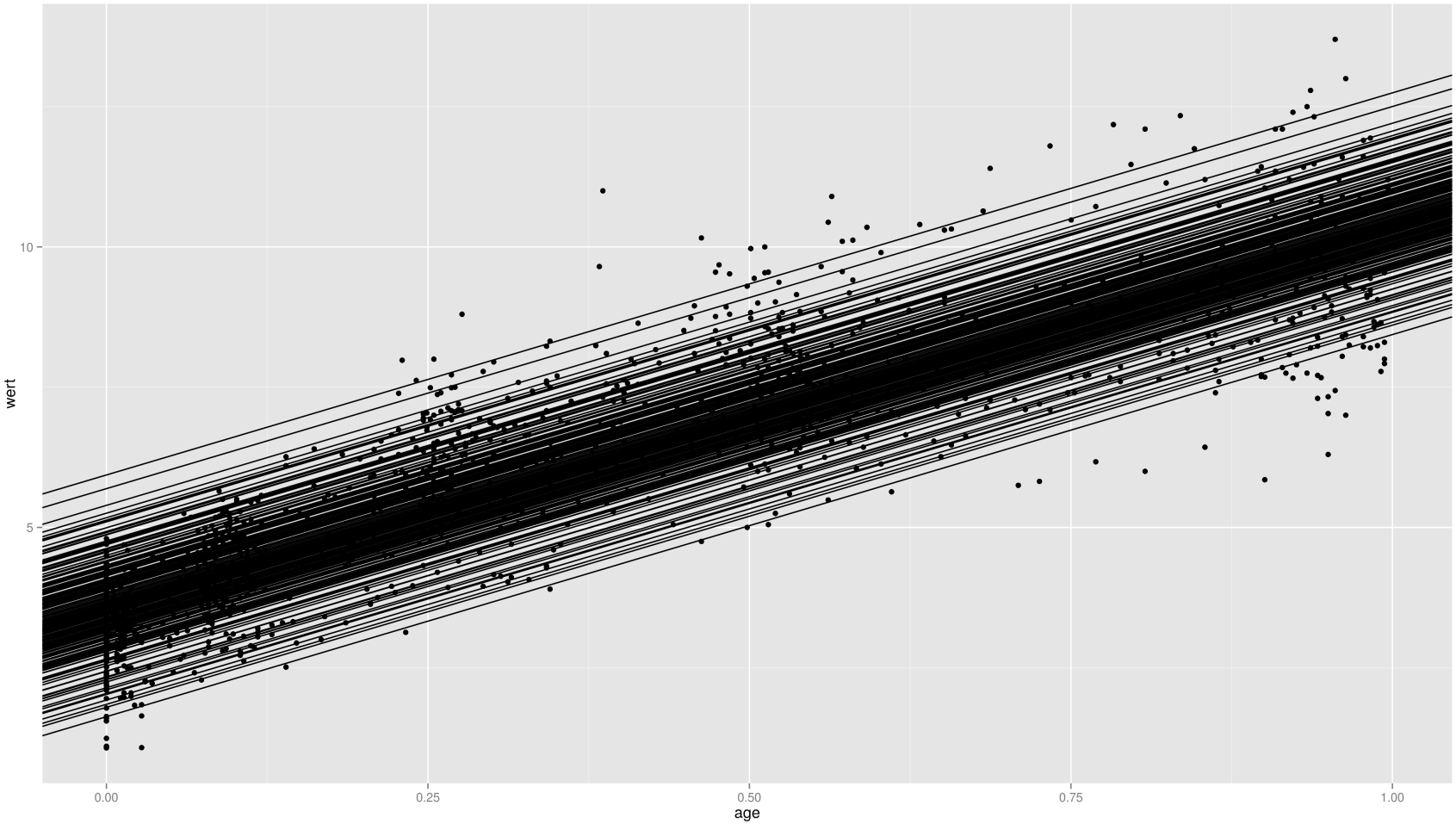
- so the regression line for child 1 is gewicht $= 3.32 + 6.81 \cdot$ age, for child 2: gewicht $= 4.46 + 6.81 \cdot$ age
- the slopes are all identical because they were specified thus in the model (only the intercept was defined varying by person)

Visualizing the Model

- we get back a intercept and a slope for every child, the appropriate layer in ggplot is the `geom_abline()`
- so first we extract to cœfficients into a data frame and rename the columns
- then we set up a ggplot object, because we want to add the data points as well we give the data as well as the æsthetics directly to the layers

```
> ggplot() +
+     geom_point(data = gewichte,aes( x = age, y = wert)) +
+     geom_abline(data = coefficienten,
+                 aes( intercept = intercept, slope = age))
```

Visualizing the Model



Exercises

- what are the assumption behind the model (varying intercept/fixed slope)?
- now let age also be varying per person (1 age | personennr)
- extract the fixed effects. Do they differ from the model with fixed slope?
- visualize it!
- is it more reasonable to have also a random slope?
- add sex as a fixed effect (first without interaction with age), convert it to a factor with the labels male and female; is there a mentionable effect?
- do a normal linear model and compare the cœfficients (for the fixed effects)
- do the same adding gestational age