Mixed-effect (or multilevel) linear models
With an introduction to general linear models

Çağrı Çöltekin

University of Groningen
Information Science

April 23, 2014

Mixed-effect models: what is it about?

- In all statistical analyses we studied so far (except repeated measures ANOVA) one of the assumptions is 'independent observations'.

Mixed-effect models: what is it about?

- In all statistical analyses we studied so far (except repeated measures ANOVA) one of the assumptions is 'independent observations'.
- Mixed-effect (or multilevel) models are one of the solutions to this problem.

Mixed-effect models: what is it about?

- In all statistical analyses we studied so far (except repeated measures ANOVA) one of the assumptions is ‘independent observations’.
- Mixed-effect (or multilevel) models are one of the solutions to this problem.
- The main difference is that we model some of the coefficients being random, having a distribution of their own.

Why not repeated measures ANOVA

Repeated measures ANOVA

- cannot handle 'crossed' random effects (a well-known example is language-as-a-fixed-effect fallacy, Clark 1973).
- is sensitive to missing data, unbalanced designs ...
- cannot handle numeric predictors.
- cannot handle non-normal response variables (no GLMs).

Mixed-effect models provide a general solution to these problems.

An introduction to example data/problem

We want to compare parenthetical expressions (par), like this one, to two other types of phrases: *phonological phrases* (PP) and *intonational phrases* (IP).

- **rate** the response variable: speech rate in syllables/s.
- **context** whether phrase is uttered as a *par*, *PP* or *IP*.
- **length** length of the phrase.
- **age** age of the participant
- **gender** gender of the participant.
- **sID** ID of the participant.
- **pID** ID of the phrase.

The question is real (from Güneș and Çöltekin 2014), but the data is overly simplified and fabricated for the purpose of demonstration.

Overview

The simple linear model

$$y_i = a + bx_i + e_i$$

- y is the *outcome* (or response, or dependent) variable. The index i represent each unit observation/measurement (sometimes called a 'case').
- x is the *predictor* (or explanatory, or independent) variable.
- a is the intercept.
- b is the slope of the regression line.
- a and b are called *coefficients*.
- $a + bx$ is the *deterministic* part of the model. It is the model's prediction of y ($\hat{y}$), given x.
- e is the residual, error, or the variation that is not accounted for by the model. Assumed to be (approximately) normally distributed with 0 mean (e_i are assumed to be i.i.d).

Interpreting a regression model fit

a the expected value of the response, when predictor is zero.

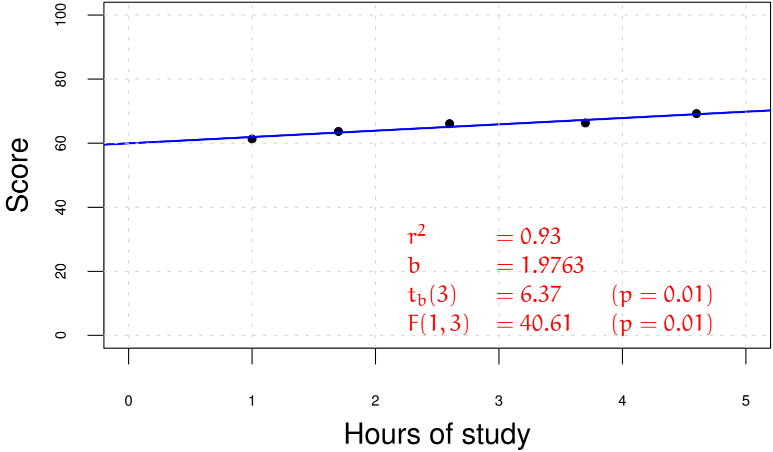
b the expected difference in the response for one-unit difference in the predictor.

r^2 variation explained by the model.

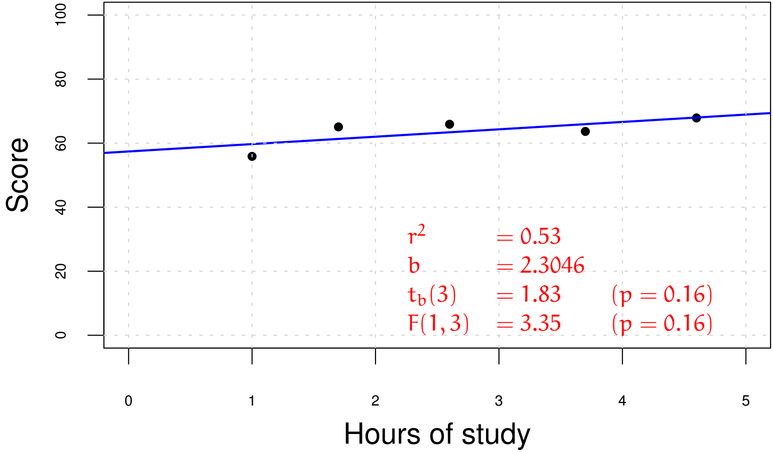
t-test for coefficients: is my coefficient significantly different than 0?

F-test for model fit: does the variation explained by the model larger than the unexplained variation?

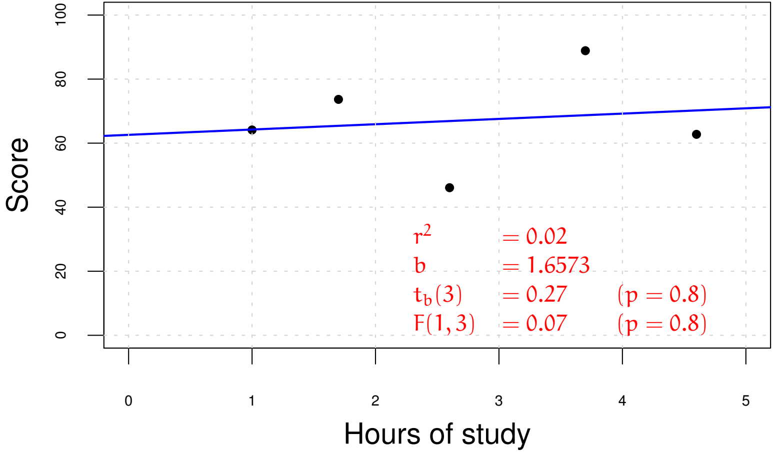
Visualizing regression results



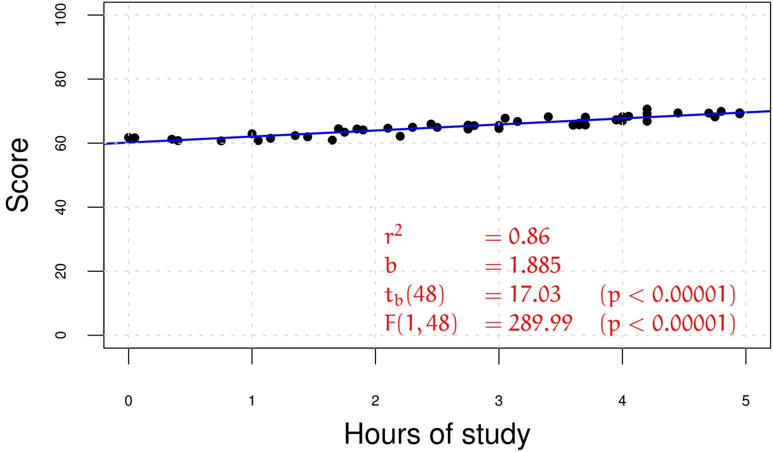
Visualizing regression results



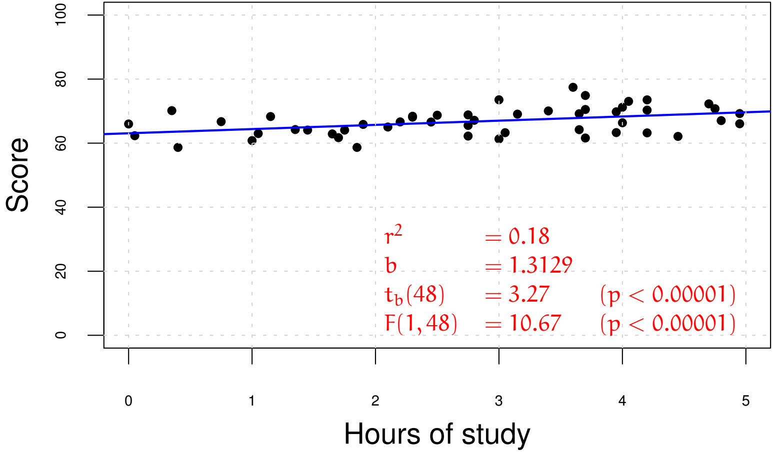
Visualizing regression results



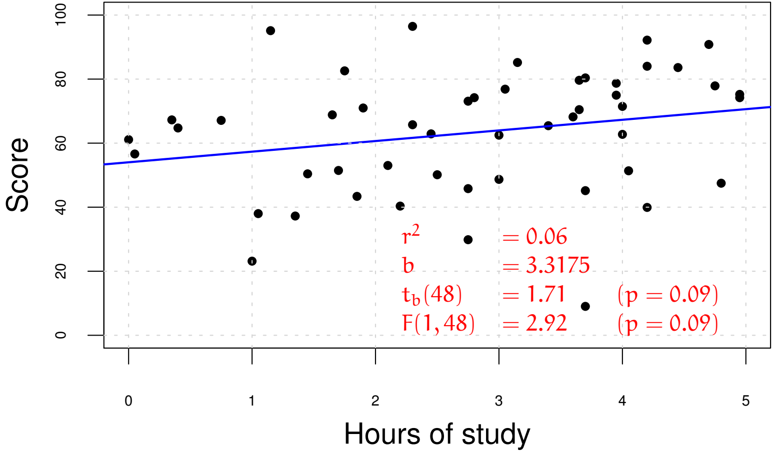
Visualizing regression results



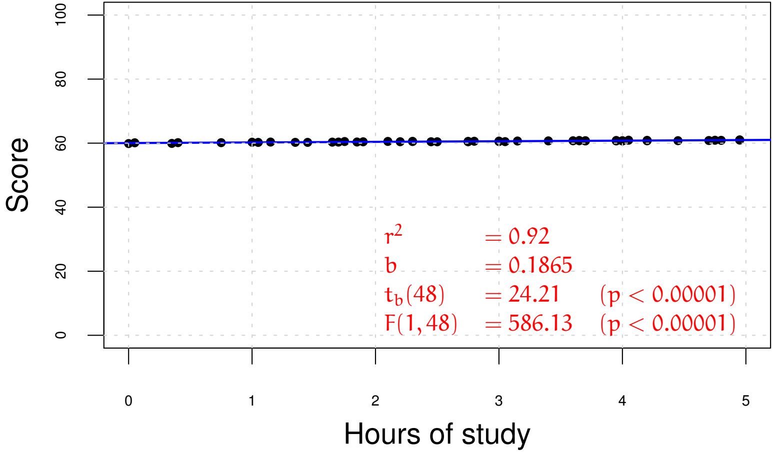
Visualizing regression results



Visualizing regression results



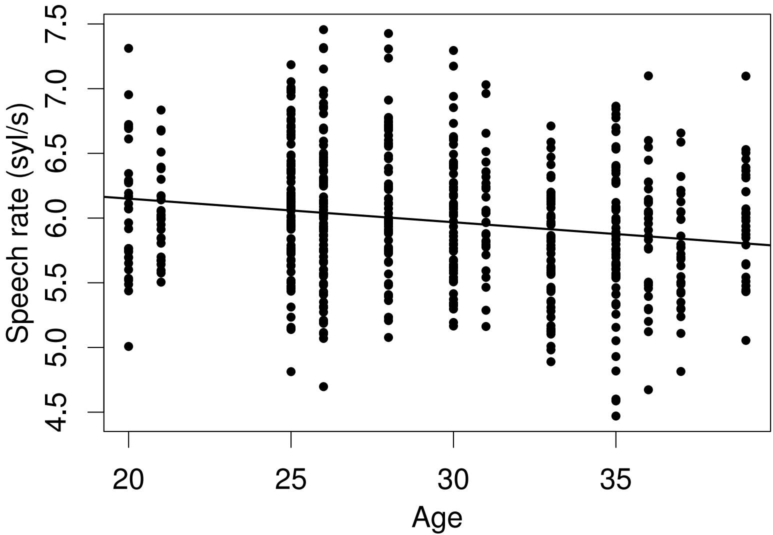
Visualizing regression results



Back to speech rate data: guessing rate from age

```
> summary(lm(rate ~ age, data=par))
lm(formula = rate ~ age, data = par)
Residuals:
    Min      1Q  Median     3Q     Max
-1.40639 -0.34965 -0.02414 0.31929 1.42219
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.514604 0.119122 54.688 < 2e-16
age       -0.018227 0.003983 -4.576 5.77e-06
---
Residual standard error: 0.5074 on 598 degrees of freedom
Multiple R-squared: 0.03383, Adjusted R-squared: 0.03222
F-statistic: 20.94 on 1 and 598 DF, p-value: 5.767e-06
```

And the graph...



Multiple regression

Regression analysis can be extended to multiple predictors.

$$y_i = a + b_1 x_{1i} + \ldots + b_k x_{ki} + e_i$$

- a is the intercept (as before).
- $b_{1..k}$ are the coefficients of the respective predictors.
- e is the error term (residual).

It is a generalization of simple regression with some additional power and complexity.

Regression with categorical predictors

- A categorical variable with N levels converted to $N - 1$ 'indicator' (or dummy) variables.
- Consider the 'context' variable with three levels ('pp', 'par', 'ip'), we can code it as two variables, 'context.par', 'context.ip' :

level	context.par	context.ip
pp	0	0
par	1	0
ip	0	1

- Other coding options (contrasts) are possible. The inferences will not change.
- If the levels are ordered, transforming the categorical variable into a numeric variable may be more appropriate.

Regression with categorical predictors

level	context.par	context.ip
pp	0	0
par	1	0
ip	0	1

In this case estimated regression equation is:

$$rate = a + b_{par} \times context.par + b_{ip} \times context.ip$$

If the value of the categorical predictor is:

pp $rate = a + b_{par} \times 0 + b_{ip} \times 0$

par $rate = a + b_{par} \times 1 + b_{ip} \times 0$

ip $rate = a + b_{par} \times 0 + b_{ip} \times 1$

Example: predicting speech rate from gender

We want to know whether there is a gender effect in speech rate. Normally, we would do a t-test:

```
> t.test(rate ~ gender, var.equal=T, data=par)
       Two Sample t-test
data: rate by gender
t = 2.0302, df = 598, p-value = 0.04278
alternative hypothesis: true difference in means is not equal to
    0
95 percent confidence interval:
 0.002782932 0.167778450
sample estimates:
mean in group F mean in group M
      6.020452        5.935172
```

Doing t-test with regression

- We have two levels of the predictor ('F' and 'M').
- We code 'F' as 0 and 'M' as 1.

$$y_i = a + b \times genderM_i + e_i$$

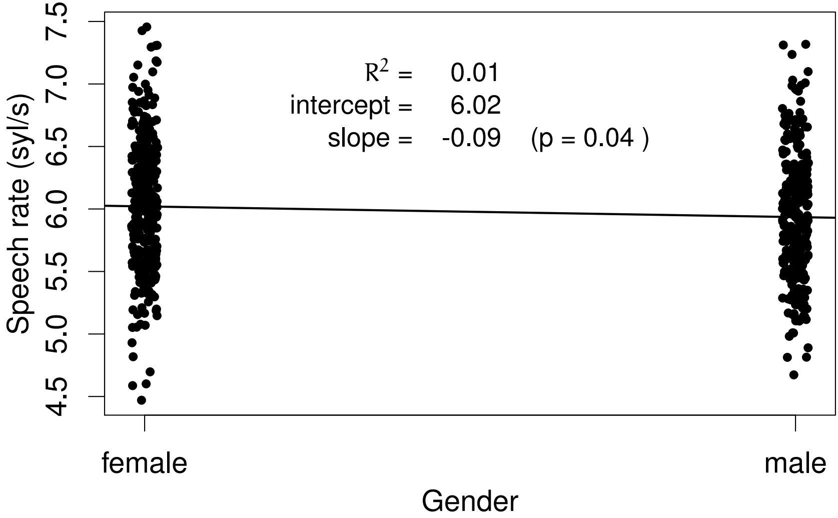
a (intercept) is the mean of level 'F'.

b (slope) is the mean difference between 'M' and 'F'.

Example: t-test with regression

```
> summary(lm(rate ~ gender, data=par))
Call: lm(formula = rate ~ gender, data = par)
Residuals:
    Min       1Q   Median      3Q     Max
-1.55020 -0.36688 -0.00631 0.34235 1.43561
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.02045  0.02970 202.69  <2e-16
genderM   -0.08528  0.04201  -2.03  0.0428
---
Residual standard error: 0.5145 on 598 degrees of freedom
Multiple R-squared: 0.006845, Adjusted R-squared: 0.005184
F-statistic: 4.122 on 1 and 598 DF, p-value: 0.04278
```

The graph: t-test with regression



ANOVA as regression

Now, we try to answer the main question: effect of context on speech rate (but in a wrong way).
Remembering that we code three levels as two indicator (dummy) variables:

$$rate_i = a + b_1 \times context.par_i + b_2 \times context.ip_i + e_i$$

a (intercept) is the mean of context 'pp'.

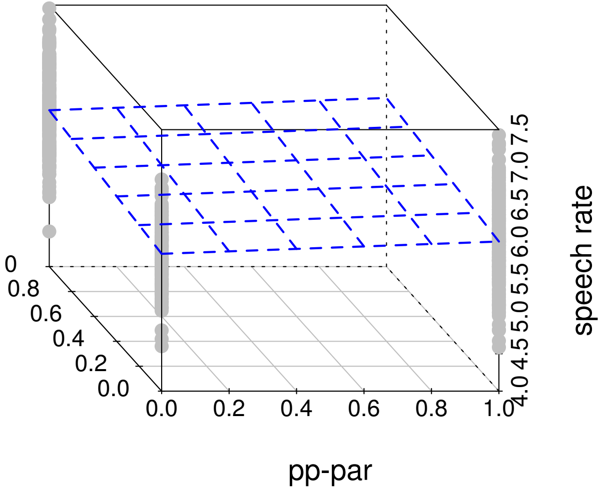
b_1 (slope of context.par) is the mean difference between 'pp' and 'par'.

b_2 (slope of context.pp) is the mean difference between 'pp' and 'ip'.

Example: ANOVA with regression

```
> summary(lm(rate ~ context, data=par))
Call: lm(formula = rate ~ context, data = par)
Residuals:
     Min       1Q   Median       3Q      Max
-1.62276 -0.33728 -0.01849 0.33646 1.42393
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.83792  0.03576 163.257 < 2e-16 ***
contextpar 0.16458   0.05057  3.254  0.0012 **
contextip  0.25510   0.05057  5.044 6.05e-07 ***
---
Residual standard error: 0.5057 on 597 degrees of freedom
Multiple R-squared: 0.04198, Adjusted R-squared: 0.03877
F-statistic: 13.08 on 2 and 597 DF, p-value: 2.756e-06
```

ANOVA as regression: the picture



The ANOVA summary

```
> summary.aov(lm(rate ~ context, data=par))
           Df Sum Sq Mean Sq F value Pr(>F)
context     2   6.69   3.345   13.08 2.76e-06 ***
Residuals 597 152.68   0.256
```

Note that the fitted model is the same, we only summarize the results differently.

Factorial ANOVA

```
> summary(lm(rate ~ context + gender, data=par))
Call: lm(formula = rate ~ context + gender, data = par)
Residuals:
     Min       1Q   Median       3Q      Max
-1.66540 -0.33977 -0.02321 0.34723 1.38129
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.88056  0.04118 142.809 < 2e-16 ***
contextpar 0.16458   0.05043  3.263 0.00116 **
contextip  0.25510   0.05043  5.058 5.64e-07 ***
genderM   -0.08528   0.04118 -2.071 0.03879 *
---
Residual standard error: 0.5043 on 596 degrees of freedom
Multiple R-squared: 0.04883, Adjusted R-squared: 0.04404
F-statistic: 10.2 on 3 and 596 DF, p-value: 1.477e-06
```

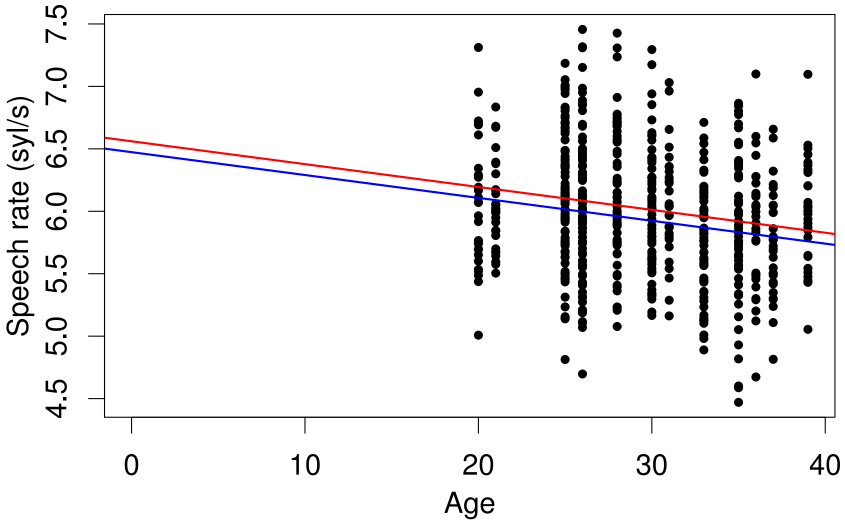
Factorial ANOVA with interactions

```
> summary(lm(rate ~ context * gender, data=par))
Call: lm(formula = rate ~ context * gender, data = par)
Residuals:
    Min      1Q  Median      3Q     Max
-1.69784 -0.34763 -0.02523 0.33703 1.39576
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.86258   0.05046 116.173 < 2e-16 ***
contextpar       0.16809   0.07137  2.355  0.0188 *
contextip        0.30551   0.07137  4.281 2.17e-05 ***
genderM         -0.04933   0.07137 -0.691  0.4897
contextpar:genderM -0.00702 0.10093 -0.070 0.9446
contextip:genderM -0.10083 0.10093 -0.999 0.3182
---
Residual standard error: 0.5046 on 594 degrees of freedom
Multiple R-squared: 0.05081, Adjusted R-squared: 0.04282
F-statistic: 6.36 on 5 and 594 DF, p-value: 9.134e-06
```

Mixing numeric and categorical variables

```
> summary(lm(rate ~ age + gender, data=par))
Call: lm(formula = rate ~ age + gender, data = par)
Residuals:
    Min       1Q   Median      3Q      Max
-1.44950 -0.36720 -0.01878 0.33254 1.37852
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.560531 0.120761 54.327 < 2e-16 ***
age       -0.018308 0.003972 -4.609 4.95e-06 ***
genderM   -0.087111 0.041315 -2.108 0.0354 *
---
Residual standard error: 0.506 on 597 degrees of freedom
Multiple R-squared: 0.04097, Adjusted R-squared: 0.03776
F-statistic: 12.75 on 2 and 597 DF, p-value: 3.772e-06
```

The picture

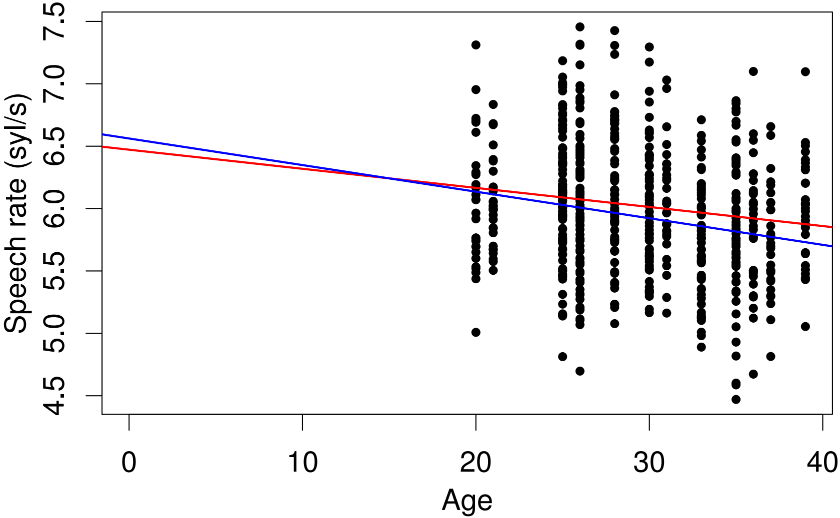


Mixing numeric and categorical variables with interactions

```
> summary(lm(rate ~ gender * age, data=par))
Call: lm(formula = rate ~ gender * age, data = par)
Residuals:
    Min       1Q   Median      3Q     Max
-1.46601 -0.37159 -0.01213 0.33294 1.38302

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.471979 0.168310 38.453 < 2e-16
genderM     0.089725  0.237657  0.378 0.70591
age        -0.015306  0.005619 -2.724 0.00664
genderM:age -0.006005 0.007947 -0.756 0.45019
---
Residual standard error: 0.5062 on 596 degrees of freedom
Multiple R-squared: 0.04189, Adjusted R-squared: 0.03707
F-statistic: 8.686 on 3 and 596 DF, p-value: 1.198e-05
```

The picture



What is a multilevel model?

$$y_i = a + b_1 x_{1i} + \ldots + b_k x_{ki} + e_i$$

- In a classical regression model, the parameters (a and b_i) are 'fixed'.
- In multilevel models, we model one or more parameters as 'random', being drawn from a distribution.
- Multilevel modeling is about estimating the main model, and the random parameters simultaneously.

A simple example: accounting for between-subject variation

$$y_i = a_{j[i]} + bx_i + e_i$$

$$a_j = \mu_a + \epsilon_j$$

A simple example: accounting for between-subject variation

$$y_i = a_{j[i]} + bx_i + e_i$$

$$a_j = \mu_a + \epsilon_j$$

For our example,

- y (response variable) is the speech rate, indexed by each observation (speaker × item × context).
- a_j is intercept for subject j, notation $j[i]$ indicates subject j associated with i^{th} observation.
- e_i are the error for each observation i.
- μ_a is the mean intercept for all subjects.
- ϵ_j is the error (variation) associated with the intercept due to the each subject j.

Equivalently...

$$y_i = a + \epsilon_{j[i]} + bx_i + e_i$$

Equivalently...

$$y_i = a + \epsilon_{j[i]} + bx_i + e_i$$

For our example,

- y (response variable) is the speech rate, indexed by each measurement (speaker × item × context).
- a is the common intercept (mean intercept for all subjects).
- ϵ_j the error (deviation from mean intercept) for each subject.
- e is the error for each observation.

A simple example with R

```
> library(lme4)
> m1 <- lmer(rate ~ length + (1|sID), data=par, REML=F)
> summary(m1)
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: rate ~ length + (1 | sID)
     AIC      BIC   logLik deviance df.resid
   821.2    838.8   -406.6    813.2      596
Random effects:
 Groups   Name        Variance Std.Dev.
 sID      (Intercept) 0.01788  0.1337
 Residual             0.21786  0.4668
Number of obs: 600, groups: sID, 20
Fixed effects:
            Estimate Std. Error t value
(Intercept)  4.76804    0.13802   34.55
length       0.21603    0.02382    9.07
Correlation of Fixed Effects:
       (Intr)
length -0.966
```

R output: fixed effects

```
            Estimate Std. Error t value
(Intercept) 4.76804  0.13802  34.55
length      0.21603  0.02382   9.07
```

$$y = a + bx$$

- y is the speech rate.
- $a = 4.77$ is the intercept for an average participant (speech rate at $length = 0$).
- $b = 0.22$ is the expected change in speech rate for one syllable increase in phrase length for all participants.

R output: fixed effects

```
            Estimate Std. Error t value
(Intercept) 4.76804  0.13802  34.55
length      0.21603  0.02382   9.07
```

$$y = a + bx$$

- y is the speech rate.
- $a = 4.77$ is the intercept for an average participant (speech rate at $length = 0$).
- $b = 0.22$ is the expected change in speech rate for one syllable increase in phrase length for all participants.

Our best estimate of speech rate and phrase length relation is:

$$rate = 4.77 + 0.22 \times length$$

R output: random effects

```
Random effects:
 Groups   Name        Variance Std.Dev.
 sID      (Intercept) 0.01788  0.1337
 Residual             0.21786  0.4668
Number of obs: 600, groups: sID, 20
```

- Intercept varies by participant (`sID`) according to a normal distribution with $mean = 0$ and $sd = 0.018$, $N(0, 0.018)$.
- Residuals are distributed with $N(0, 0.218)$.

R output: random effects

```
Random effects:
 Groups   Name        Variance Std.Dev.
 sID      (Intercept) 0.01788  0.1337
 Residual             0.21786  0.4668
Number of obs: 600, groups: sID, 20
```

- Intercept varies by participant (`sID`) according to a normal distribution with $mean = 0$ and $sd = 0.018$, $N(0, 0.018)$.
- Residuals are distributed with $N(0, 0.218)$.

Estimated main model is:

$$rate = 4.77 + \epsilon_{subj} + 0.22 \times length + e_r$$

where $\epsilon_{subj} \sim N(0, 0.018)$ and $e_r \sim N(0, 0.218)$.

Random effects

For our example model, `lmer()` fits a random intercept for each participant. We do not see this in the summary of the main model. Here is how to obtain them:

```
> ranef(m1)
$sID
   (Intercept)
1  -0.036494898
2  -0.137225794
3  -0.090392453
4   0.222219430
...
```

Random effects

For our example model, `lmer()` fits a random intercept for each participant. We do not see this in the summary of the main model. Here is how to obtain them:

```
> ranef(m1)
$sID
   (Intercept)
1  -0.036494898
2  -0.137225794
3  -0.090392453
4   0.222219430
...
```

Model for subject 1: $rate = 4.77 + \epsilon_{subj} + 0.22 \times length + e_r$

Random effects

For our example model, `lmer()` fits a random intercept for each participant. We do not see this in the summary of the main model. Here is how to obtain them:

```
> ranef(m1)
$sID
   (Intercept)
1  -0.036494898
2  -0.137225794
3  -0.090392453
4   0.222219430
...
```

Model for subject 1: $rate = 4.77 - 0.037 + 0.22 \times length + e_r$

Random effects

For our example model, `lmer()` fits a random intercept for each participant. We do not see this in the summary of the main model. Here is how to obtain them:

```
> ranef(m1)
$sID
   (Intercept)
1  -0.036494898
2  -0.137225794
3  -0.090392453
4   0.222219430
...
```

Model for subject 1: $rate = 4.77 - 0.037 + 0.22 \times length + e_r$
Model for subject 4: $rate = 4.77 + \epsilon_{subj} + 0.22 \times length + e_r$

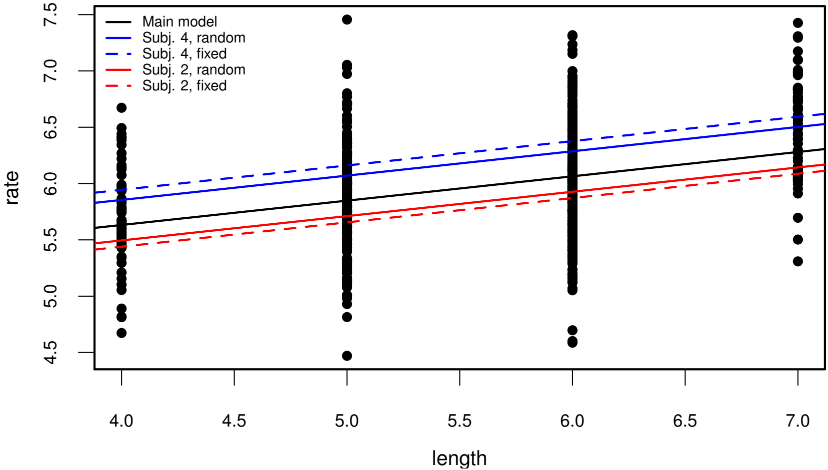
Random effects

For our example model, `lmer()` fits a random intercept for each participant. We do not see this in the summary of the main model. Here is how to obtain them:

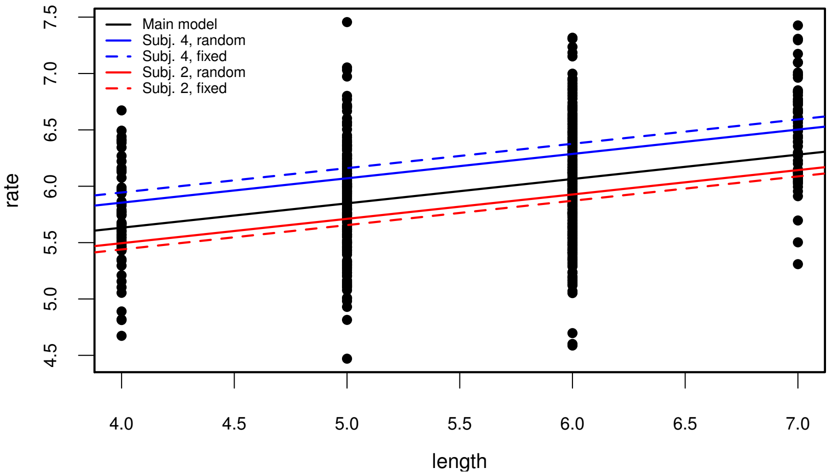
```
> ranef(m1)
$sID
   (Intercept)
1  -0.036494898
2  -0.137225794
3  -0.090392453
4   0.222219430
...
```

Model for subject 1: $rate = 4.77 - 0.037 + 0.22 \times length + e_r$
Model for subject 4: $rate = 4.77 + 0.222 + 0.22 \times length + e_r$

Random effect estimates



Random effect estimates



Estimate for each subject is affected by the complete population (pulled towards the average).

Where are my p-values?

The calculation of the p-values in mixed-effect models is not straightforward.

There are a few options:

- Model comparison with likelihood ratio test.
- Profile-based confidence intervals.
- Sampling based p-values.

Model comparison: is the predictor justified?

```
> m1 <- lmer(rate ~ length + (1|sID), data=par, REML=F)
> m0 <- lmer(rate ~ 1 + (1|sID), data=par, REML=F)
> anova(m0, m1)
Models:
m0: rate ~ 1 + (1 | sID)
m1: rate ~ length + (1 | sID)
   Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
m0  3 896.16 909.35 -445.08 890.16
m1  4 821.24 838.82 -406.62 813.24 76.924  1  < 2.2e-16 ***
---
Signif. codes: 0 ‘’*** 0.001 ‘’** 0.01 ‘’* 0.05 ‘’. 0.1 ‘’1
```

This is general test that you can also apply for justifying other sorts of model complexity.

Profile-based confidence intervals

```
> m1.profile <- profile(m1)
> confint(m1.profile)
                 2.5 %    97.5 %
.sig01      0.08381928 0.2071916
.sigma      0.44113431 0.4949640
(Intercept) 4.49707507 5.0389991
length      0.16926906 0.2627934
```

- `.sig01` is the error due to subjects (estimate was 0.1337)
- `.sigma` is the residual error (0.4668)
- The estimates for intercept and slope were 4.768 and 0.216.
- Note that the intervals (esp. for standard deviations) are not symmetric.

Random intercepts and random slopes

$$y_i = a + \epsilon^{(a)}_{j[i]} + (b + \epsilon^{(b)}_{j[i]})x_i + e_i$$

Random intercepts and random slopes

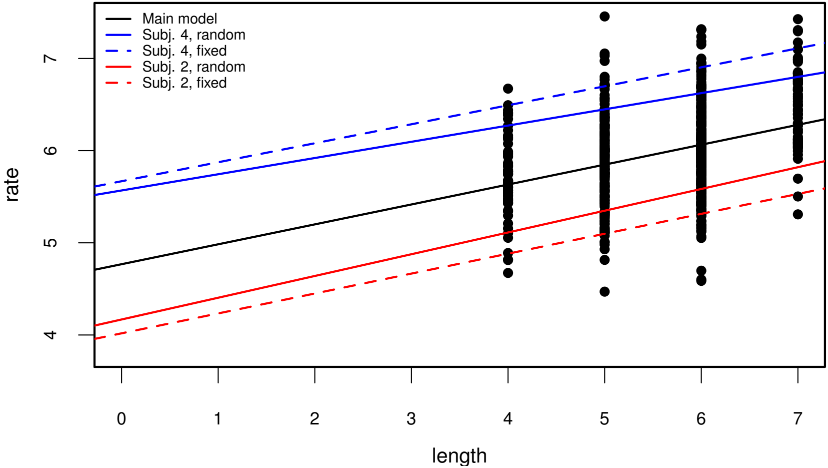
$$y_i = a + \epsilon^{(a)}_{j[i]} + (b + \epsilon^{(b)}_{j[i]})x_i + e_i$$

$\epsilon^{(a)}_{j[i]}$ Random variation of intercept due to subjects.

$\epsilon^{(b)}_{j[i]}$ Random variation of slope due to subjects.

e_i Random variation that we cannot account for.

Random intercepts and random slopes: visualization



Random intercepts and random slopes: example

```
> m2 <- lmer(rate ~ scale(length, scale=F) + (length|sID),
       data=par, REML=F)
> summary(m2)
    AIC     BIC  logLik deviance df.resid
  824.6   851.0  -406.3   812.6      594
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 sID      (Intercept) 0.0001942 0.01393
          length      0.0004582 0.02140 1.00
 Residual             0.2175505 0.46642
Number of obs: 600, groups: sID, 20
Fixed effects:
                      Estimate Std. Error t value
(Intercept)            5.97781   0.03546  168.6
scale(length, scale = F) 0.21603 0.02428   8.9
Correlation of Fixed Effects:
           (Intr)
scl(ln,s=F) 0.166
```

Correlated random effects

```
> m2.u <- lmer(rate ~ scale(length, scale=F)
             + (1|sID) + (0+length|sID),
             data=par, REML=F)
> summary(m2.u)
    AIC     BIC  logLik deviance df.resid
  822.6   844.6  -406.3   812.6     595
Random effects:
 Groups   Name        Variance  Std.Dev.
 sID      (Intercept) 0.0009367 0.03061
 sID.1    length      0.0005397 0.02323
 Residual             0.2175370 0.46641
Number of obs: 600, groups: sID, 20
Fixed effects:
                      Estimate Std. Error t value
(Intercept)            5.97781   0.03543 168.70
scale(length, scale = F) 0.21603 0.02436 8.87
```

Do we need the estimate the correlation parameter(s)?

```
> anova(m2.u, m2)
Data: par
Models:
m2.u: rate ~ scale(length, scale = F) + (1 | sID) + (0 + length |
     sID)
m2: rate ~ scale(length, scale = F) + (length | sID)
    Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
m2.u 5 822.65 844.63 -406.32 812.65
m2   6 824.64 851.02 -406.32 812.64 0.0062  1    0.9375
```

Are random slopes justified?

```
> anova(m1, m2.u)
Data: par
Models:
m1: rate ~ length + (1 | sID)
m2.u: rate ~ scale(length, scale = F) + (1 | sID) + (0 + length |
     sID)
    Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
m1    4 821.24 838.82 -406.62 813.24
m2.u 5 822.65 844.63 -406.32 812.65 0.5893  1     0.4427
```

More sources of variation

- In most (psycho)linguistics experiments/data, we also have variation due to items (words, phrases, sentences).
- Typically, the items are crossed with the subjects.
- Repeated measures ANOVA cannot deal with such crossed designs. (see, Clark 1973; Raaijmakers, Schrijnemakers, and Gremmen 1999 and Baayen 2008, chapter 7).
- Mixed-effect can deal with this type of data well.

More sources of variation: example formulation

$$y_i = a + \epsilon_{j[i]} + \delta_{k[i]} + bx_i + e_i$$

- New source of error, $\delta_{k[i]}$ represents the error due to item (phrase) k.
- We can also include random variation of slope due to items, subjects or both.

Example: the effect of context

```
> m3 <- lmer(rate ~ context + (1+context|sID) + (1+context|pID),
    data=par, REML=F)
> summary(m3)
     AIC      BIC   logLik deviance df.resid
   725.0    795.4   -346.5    693.0      584
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 sID      (Intercept) 0.02353  0.1534
          contextpar  0.01224  0.1106   -0.30
          contextip   0.01286  0.1134   -0.08 -0.38
 pID      (Intercept) 0.02756  0.1660
          contextpar  0.10709  0.3272   -0.21
          contextip   0.06139  0.2478   -0.05  0.37
 Residual             0.15430  0.3928
Number of obs: 600, groups: sID, 20; pID, 10
Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.83792    0.06859   85.12
contextpar   0.16458    0.11342    1.45
contextip    0.25510    0.09124    2.80
```

Example: do we need random slopes?

```
> m3 <- lmer(rate ~ context + (1+context|sID) + (1+context|pID),
    data=par, REML=F)
> m4 <- lmer(rate ~ context + (1|sID) + (1+context|pID), data=par
    , REML=F)
> m5 <- lmer(rate ~ context + (1|sID) + (1|pID), data=par, REML=F
    )
> anova(m5, m4, m3)
Models:
m5: rate ~ context + (1 | sID) + (1 | pID)
m4: rate ~ context + (1 | sID) + (1 + context | pID)
m3: rate ~ context + (1 + context | sID) + (1 + context | pID)
   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
m5  6 781.66 808.04 -384.83   769.66
m4 11 722.02 770.39 -350.01   700.02 69.639      5  1.218e-13 ***
m3 16 725.01 795.36 -346.50   693.01  7.016      5     0.2195
```

Example: do we need random slopes?

```
> m3 <- lmer(rate ~ context + (1+context|sID) + (1+context|pID),
    data=par, REML=F)
> m4 <- lmer(rate ~ context + (1|sID) + (1+context|pID), data=par
    , REML=F)
> m5 <- lmer(rate ~ context + (1|sID) + (1|pID), data=par, REML=F
    )
> anova(m5, m4, m3)
Models:
m5: rate ~ context + (1 | sID) + (1 | pID)
m4: rate ~ context + (1 | sID) + (1 + context | pID)
m3: rate ~ context + (1 + context | sID) + (1 + context | pID)
   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
m5  6 781.66 808.04 -384.83   769.66
m4 11 722.02 770.39 -350.01   700.02 69.639      5  1.218e-13 ***
m3 16 725.01 795.36 -346.50   693.01  7.016      5     0.2195
```

- The random slopes for item is justified, but no need for random slopes for each subject.

Example: summary of our best model

```
> summary(m4)
    AIC     BIC   logLik deviance df.resid
  722.0   770.4  -350.0   700.0      589
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 sID      (Intercept) 0.02067 0.1438
 pID      (Intercept) 0.02712 0.1647
          contextpar  0.10583 0.3253 -0.20
          contextip  0.06012 0.2452  -0.05 0.37
 Residual             0.16129 0.4016
Number of obs: 600, groups: sID, 20; pID, 10
Fixed effects:
            Estimate Std. Error t value
(Intercept) 5.83792  0.06746  86.53
contextpar  0.16458  0.11044   1.49
contextip   0.25510  0.08732   2.92
```

Predictors at different levels

```
> m6 <- lmer(rate ~ context + length + age
           + (1|sID) + (context|pID), data=par, REML=F)
> summary(m6)
     AIC      BIC   logLik deviance df.resid
   702.9    760.0   -338.4    676.9      587
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 sID      (Intercept) 0.01107  0.1052
 pID      (Intercept) 0.03356  0.1832
          contextpar  0.10583  0.3253   -0.95
          contextip   0.06011  0.2452   -0.22 0.37
 Residual             0.16134  0.4017
Number of obs: 600, groups: sID, 20; pID, 10
Fixed effects:
             Estimate Std. Error t value
(Intercept)  4.940482  0.257982  19.150
contextpar   0.164582  0.110437   1.490
contextip    0.255099  0.087319   2.921
length       0.256112  0.033627   7.616
age         -0.018227  0.005515  -3.305
```

More multilevel scenarios

Our example was a typical application of mixed-effect models: accounting for crossed random effects. But mixed effect models can also be very useful if your data is organized hierarchically. For example in a second language acquisition study:

Level	Predictors
Students in classrooms	age, gender …
Classrooms in schools	size, teacher …
Schools in cities/districts	average income, city size …
Cities in countries	native language(s), SLA policies …

Summary

- Multilevel (or mixed-effect) models are a generalization of linear models where some of the coefficients are considered random variables.
- We model parameters as 'random' where we have a systematic variation that we can explain by some of the variation in the data.
- Mixed-effect models are useful in modeling both hierarchical and non-hierarchical random-effects structure in the data.

Where to go from here

- We have only covered examples of models with normal errors, multilevel models can also be fitted for any 'generalized linear model'.
- Once you have taken the path of specifying 'random' parameters, you can embrace it, and use Bayesian inference.

Recommended reading:

- Gelman and Hill (2007): if you are serious about mixed-effect models, worth the time.
- Baayen (2008): takes a linguistic perspective, also includes discussion of different solutions to language-as-a-fixed-effect fallacy.
- Bates (2010): written by the author of `lme4`, authoritative and rather accessible .

Data generation

```
e.intercept <- 6
e.subj.a.gender <- 0.1
e.subj.a.age <- - 0.1
e.subj.a.resid <- 0.1

e.phr.a.length <- 0.1
e.phr.a.resid <- 0.06

e.phr.ip.length <- 0.2
e.phr.ip.resid <- 0.2
e.phr.par.length <- 0.1
e.phr.par.resid <- 0.2

e.pp <- -0.04
e.par <- 0
e.ip <-  0.06

e.resid <- 0.4
```

Data generation (cont.)

```
par.subj <- data.frame(sID=factor(1:20),
                       gender=factor(rep(c('F', 'M'), each=10)),
                       age=round(runif(20, 20,40)))
par.phrase <- data.frame(pID=factor(1:10), length=round(runif(10,
     4, 7)))
par.subj$a.subj <- c(e.subj.a.gender,0)[par.subj$gender] +
                    e.subj.a.age * scale(par.subj$age)
par.subj$a.subj.e <- par.subj$a.subj + rnorm(20, 0, e.subj.a.
    resid)

par.phrase$a.phr <- e.phr.a.length * scale(par.phrase$length)
par.phrase$a.phr.e <- par.phrase$a.phr + rnorm(10,0, e.phr.a.
    resid)

par.phrase$b.ip <- e.phr.ip.length * scale(par.phrase$length,
    scale=F)
par.phrase$b.ip.e <- par.phrase$b.ip + rnorm(10, 0, e.phr.ip.
    resid)
```

Data generation (cont.)

```
par.phrase$b.par <- e.phr.par.length * scale(par.phrase$length,
    scale=F)
par.phrase$b.par.e <- par.phrase$b.par + rnorm(10, 0, e.phr.par.
    resid)

par.phrase$b.pp.e <- rnorm(10, 0, 0.2)

par.data <- data.frame(sID=rep(par.subj$sID, each=10*3), pID=rep(
    par.phrase$pID, each=3, times=10))
par.data$context <- factor(c("pp", "par", "ip"), levels=c("pp", "
    par", "ip"))

par.data$rate <- e.intercept
par.data$rate <- par.data$rate + par.subj$a.subj.e[par.data$sID]
par.data$rate <- par.data$rate + par.phrase$a.phr.e[par.data$pID]

par.data$rate <- par.data$rate + (par.phrase$b.pp.e[par.data$pID]
     + e.pp) * (par.data$context == 'pp')
```

Data generation (cont.)

```
par.data$rate <- par.data$rate + (par.phrase$b.par.e[par.data$pID
    ] + e.par) * (par.data$context == 'par')
par.data$rate <- par.data$rate + (par.phrase$b.ip.e[par.data$pID]
     + e.ip) * (par.data$context == 'ip')
par.data$rate <- par.data$rate + rnorm(600, 0, e.resid)
par <- merge(merge(par.data, par.subj), par.phrase)
```

References

Baayen, R. Harald (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge University Press.

Bates, Douglas M. (2010). *lme4: Mixed-Effects Modeling with R.* Prepublication version at: http://lme4.r-forge.r-project.org/book/. New York: Springer (in preparation). URL: http://lme4.r-forge.r-project.org/book/.

Clark, Herbert H (1973). "The language-as-fixed-effect fallacy: A critique of language statistics in psychological research". In: *Journal of Verbal Learning and Verbal Behavior* 12, pp. 335–359.

Gelman, Andrew and Jennifer Hill (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Güneş, Güliz and Çağrı Çöltekin (2014). "Mapping to prosody: not all parentheticals are alike". In: *Parenthetical Verbs*. Ed. by S. Schneider, J. Glikman, and M. Avanzi. Berlin: De Gruyter, (to appear).

Raaijmakers, Jeroen G. W., Joseph M. C. Schrijnemakers, and Frans Gremmen (1999). "How to Deal with "The Language-as-Fixed-Effect Fallacy": Common Misconceptions and Alternative Solutions". In: *Journal of Memory and Language* 41, pp. 416–426.