Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Mixed Effect Models in R

Giorgio Arcara

Neurophysiology Lab
IRCCS San Camillo Hospital, Venice, Italy

July 6, 2021

XPRAG.it

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

What are mixed effect models

Mixed effect models (or Mixed Models) are a powerful statistical tool that allows to deal with complex *dependencies* in the data, in a *modeling* framework.

Mixed effect models are an extension of linear models. As such, knowledge on linear model is required to properly use mixed models.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Why "Mixed" Models

They are called mixed models because they are able to handle both *fixed* and *random* effects.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Fixed and Random Effects

There are many ways to explain the difference between fixed and random effects:

- **fixed effects** reflects variables you are manipulating or that you are interested into (e.g., Priming condition, Stimulus type).

- **random effects** are related to observations sampled "randomly" from a larger population (e.g. Participants or Items). You are interested in taking into account their effects, but you are not really interested in them.

- . . .

See also here

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Pros of using Mixed Effect Models

Pros

With Mixed Effect models you can answer to more sophisticated scientific questions, and to handle more properly the data structure.

In many cases, Mixed Effect models have higher statistical *Power* than other analysis (e.g., ANOVA)

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Cons of Mixed Effect Models

Cons

The main problem of using mixed models is related to their flexibility.

Very often there are many ways to *fit* a mixed models and it is difficult to understand what is the "best" way to perform the analysis (the same is not true for ANOVA, in which typically, given a design, there is only one way to perform the appropriate ANOVA).

Hence, with mixed models it is easier to perform a suboptimal (or even wrong) analysis.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

When to use Mixed Effect Models

Mixed effect models are perfect for these cases:

- You want to take into account both Subjects and Item variability.

- You are interested in both covariates and categorical variables, and also in their interactions.

- You have dicothomous outcome variable (Generalized Linear Mixed Effect Models).

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

When the use Mixed Effect Models can be problematic

Mixed effect models may be difficult to handle in these cases:

- You have a complex factorial design.
- You have no idea on what you are doing.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

What to expect from this course

If you are not familiar with either R or Linear regression, it is very hard to properly understand mixed models.

However, you don't need to understand *everything* now, but you need to grasp some concepts and know where to retrieve additional information.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

What to expect from this course

This is a course on Mixed Effect Models in R

Working with R requires other additional knowledge (in **data wrangling**, basic input/output, plotting, etc.)

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

What to expect from this course

This course I will refer to some working[1] .R scripts that you could use as reference to develop your mixed effects models, or as starting point to elaborate some topics.

You can download the course materials from here
https://github.com/giorgioarcara/mixed-effect-models

---

[1]hopefully

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Some crucial points to keep in mind

Before starting, keep in mind these important points.

1. Understanding Mixed Models requires understanding (well) linear regressions and other more basic aspects of statistics.

2. Fitting a mixed model is almost never a one-shot process. It requires checking assumptions before, and then checking that the fit was good, adjusting if necessary.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Filling the gaps

There are many free resources online that will make you
independent with R and mixed Models

- **Analyzing Linguistic Data** Donwload. An excellent Book by
  H. Baayen. It contains all the information you need. The
  Mixed Model part is a bit outdated.
- **Tutorials on Mixed Models** by Bodo Winter [Link].
- textbfGLMM wiki. A comprehensive webpage maintained by
  Ben Bolker and world experts on Mixed Models.
- **Google for your problem!**. If the answer is on
  Cross-Validated you are on the right track.
- **help**. Don't forget to use the help function in R.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Filling the gaps

There are two other things you can do to improve your ability in deal with mixed models for your analysis:

- ask to colleagues with more experience.
- get ideas from published papers with similar design.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Learning mixed models

In general, consider learning mixed models as a neverending process, in which you can always improve and learn new things.

Few words on Linear Regressions

With linear regression you *model* a response variable as a function
of one predictor, using the formula of a straigth line.

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \qquad (1)$$

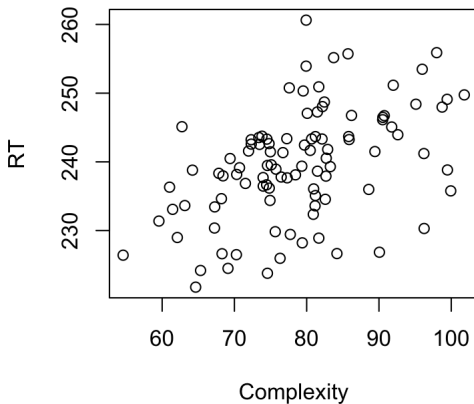the R syntax is a simplified version of the mathematical formula.

```
lm(y ~ x1)
lm(RT ~ Complexity, data=dat1)# a more realistic example
```

Introduction
Linear Models
Linear Mixed Models

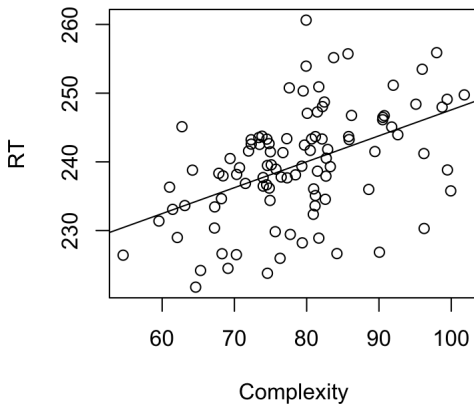Checking Assumptions
Model Selection
Advanced Topics

Some regression terminology

- **Dependent variable**: the variable to be predicted from other variables.
- **Predictors**: the variables that are supposed to predict the dependent variable.
- **Betas/Coefficients**: together with the Intercept, the core of the result of the regression.
- **Fitted values**: the results of the regression (they lie in the regression line).
- **Residuals**: the error of the models, i.e., the discrepancy between fitted and observed values.
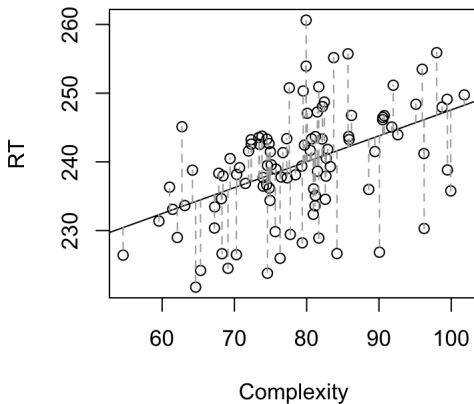
Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Predictor and Dependent Variable

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Regression Model Fit

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Residuals

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

One note about "Prediction "in regression

Beware that, strictly speaking, regression does not tell us nothing about prediction in the sense of "forecasting "or, even worse, about causality.

The term "prediction "is used because predictors is a common term in regression.

The more appropriate term would be **association**

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Fitting Regressions in R

```
set.seed(220) #  change the seed for different results

## SIMPLE LINEAR REGRESSION ##
Complexity = rnorm(100, mean = 80, sd = 10)
Error = rnorm(100, mean=0, sd = 8)
RT = 200 + 0.5*Complexity + Error
dat1 = data.frame(RT=RT, Complexity=Complexity)

mod1 = lm(RT~Complexity, dat1)
summary(mod1)
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Linear Regression Example (mod1)

```
Call:
lm(formula = RT ~ Complexity, data = dat1)

Residuals:
     Min       1Q   Median       3Q      Max
-16.9704  -4.0812   0.7874   4.4138  20.6332

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 209.83311    5.69333  36.856  < 2e-16 ***
Complexity    0.37735    0.07154   5.275 7.96e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.173 on 98 degrees of freedom
Multiple R-squared:  0.2211,Adjusted R-squared:  0.2132
F-statistic: 27.82 on 1 and 98 DF,  p-value: 7.959e-07
```

Introduction   Checking Assumptions
Linear Models   Model Selection
Linear Mixed Models   Advanced Topics

Interpreting results of regressions

Results show many things

- The Estimate column reports both the Intercept and the coefficient (*beta*) for our predictor
- The p-values obtained from `summary(mod1)` tells if the coefficient for the predictor is significantly different from zero. item
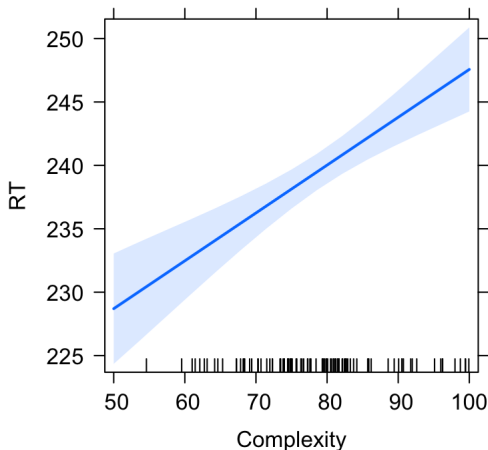- the R^2 indicate the proprtion of explained variance, which is an index of quality of fit for regressions.

The estimated regression equation is
$RT = 209.833 + 0.377 * Complexity$

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Simple Linear Regression

Complexity effect plot

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Multiple Linear Regressions

With **multiple** linear regression you can *model* a response variable as a function of more than one predictor

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \epsilon \tag{2}$$

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Multiple Linear Regression (mod2)

```
Call:
lm(formula = RT ~ Complexity + Imageability, data = dat2)

Residuals:
    Min      1Q  Median      3Q     Max
-20.478  -3.884   1.279   4.540  25.016

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 205.69685    7.03643  29.233  < 2e-16 ***
Complexity    0.35765    0.07629   4.688 9.01e-06 ***
Imageability -0.92280    0.14699  -6.278 9.66e-09 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.865 on 97 degrees of freedom
Multiple R-squared:  0.392,Adjusted R-squared:  0.3795
F-statistic: 31.27 on 2 and 97 DF,  p-value: 3.309e-11
```
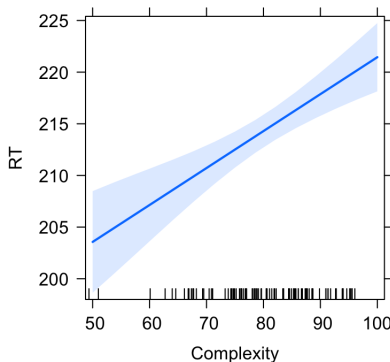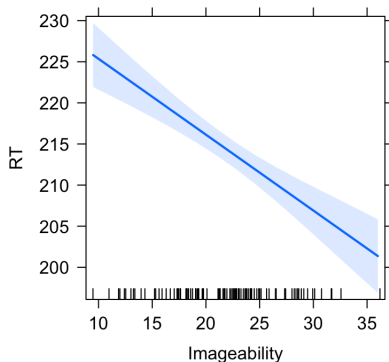
Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Multiple Linear Regression

Partial Effects of Complexity and Imageability on RT

Partial effects in Multiple Linear Regression

When plotting a partial effect, all other effects are kept to a given value (typically, the median).

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Effects in linear regressions

A peculiar aspect of regression is that the effect of each predictor is calculated *taking into account the effect of other predictors*.

(strictly speaking, each effect is estimated such to be *orthogonal*, that is uncorrelated, to all the others)

This is a very important aspect of regression also related to some problems (see Collinearity in these slides)

Introduction   Checking Assumptions
Linear Models   Model Selection
Linear Mixed Models   Advanced Topics

Other important aspects of regressions

Regression can handle not only continuous variables, but also categorical variables (in R, *factors*). In this case the factor must be transformed using *dummy coding* (<u>see here</u> or <u>here</u>).

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Factors in Linear Regression (mod3)

```
Call:
lm(formula = RT ~ PrimeType + Imageability, data = dat3)

Residuals:
    Min      1Q  Median      3Q     Max
-17.5095 -5.1034 -0.6084  5.1287 17.1625

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       201.3567     3.5101  57.365  < 2e-16 ***
PrimeTypePositive   6.0705     1.4776   4.108 8.33e-05 ***
Imageability       -2.0946     0.1734 -12.080  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.295 on 97 degrees of freedom
Multiple R-squared:  0.6087,Adjusted R-squared:  0.6006
F-statistic: 75.43 on 2 and 97 DF,  p-value: < 2.2e-16
```
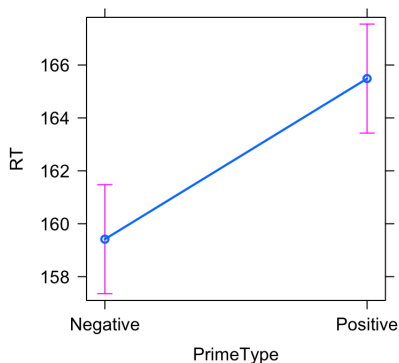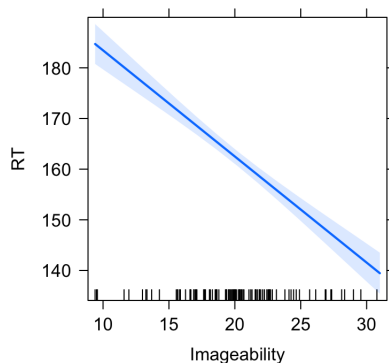
Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Factors in Linear Regression

Partial Effects of PrimeType and Imageability on RT

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interpreting results of regressions

Note that to properly check the significant factor you should use another function.

```
require(car)

Anova(mod3, type="III")

Anova Table (Type III tests)

Response: RT
              Sum Sq Df F value    Pr(>F)
(Intercept)   1970.0  1  37.019 2.327e-08 ***
PrimeType      898.2  1  16.879 8.328e-05 ***
Imageability  7765.1  1 145.919 < 2.2e-16 ***
Residuals     5161.9 97
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interpreting results of regressions

Broadly speaking, the `Anova` function actually tests whether the term contributes significantly to the model.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interaction between Factors (mod4)

```
Call:
lm(formula = RT ~ PrimeType * TaskDiff, data = dat4)

Residuals:
    Min      1Q  Median      3Q     Max
-20.0864 -4.9877 -0.8893  4.4921 21.4090

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                    202.152      1.574 128.419  < 2e-16 ***
PrimeTypePositive                9.949      2.226   4.469 2.15e-05 ***
TaskDiffEasy                     4.685      2.226   2.105   0.0379 *
PrimeTypePositive:TaskDiffEasy  -6.790      3.148  -2.157   0.0335 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.871 on 96 degrees of freedom
Multiple R-squared:  0.1909,Adjusted R-squared:  0.1656
F-statistic: 7.552 on 3 and 96 DF,  p-value: 0.0001374
```
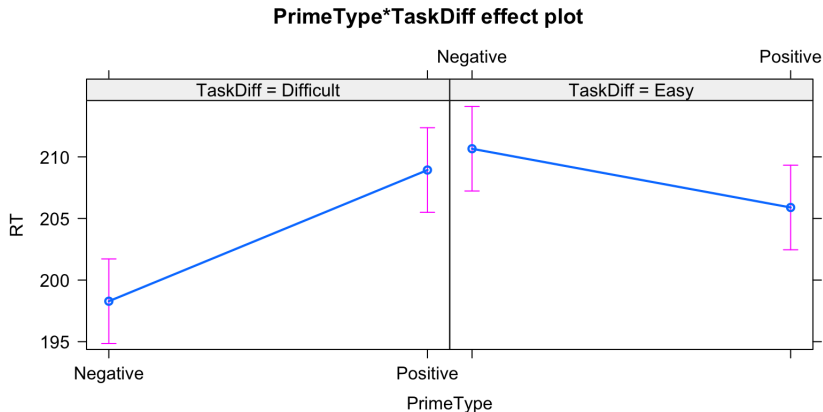
Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interaction between Factors

Partial Effects of interaction between PrimeType and Imageability on RT

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interaction between Factors

Beware!
To interpret significant results in the case of interactions of factors,
you should not use the significance of coefficients.

The equivalent of testing the significance of the overall interaction
is made again using the Anova function on the model, of the car
package.

For post-hocs you can use the emmeans package.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Post-hocs of mod4

```
> pairs(emmeans(mod4, ~PrimeType*TaskDiff))
 contrast                              estimate   SE df t.ratio p.value
 Negative,Difficult - Positive,Difficult  -10.65 2.45 96 -4.355  0.0002
 Negative,Difficult - Negative,Easy       -12.38 2.45 96 -5.061  <.0001
 Negative,Difficult - Positive,Easy        -7.61 2.45 96 -3.111  0.0129
 Positive,Difficult - Negative,Easy        -1.73 2.45 96 -0.707  0.8942
 Positive,Difficult - Positive,Easy         3.04 2.45 96  1.243  0.6010
 Negative,Easy - Positive,Easy              4.77 2.45 96  1.950  0.2144

P value adjustment: tukey method for comparing a family of 4 estimates
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interaction between a Factor and a continuous variable (mod5)

```
Call:
lm(formula = RT ~ PrimeType * Imageability, data = dat5)

Residuals:
    Min      1Q  Median      3Q     Max
-145.591 -33.166   3.205  32.017 121.188

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   218.992     35.925   6.096 2.27e-08 ***
PrimeTypePositive             -43.745     48.347  -0.905  0.36783
Imageability                   -5.492      1.731  -3.172  0.00203 **
PrimeTypePositive:Imageability  5.441      2.383   2.283  0.02461 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 56.44 on 96 degrees of freedom
Multiple R-squared:  0.3175,	Adjusted R-squared:  0.2961
F-statistic: 14.88 on 3 and 96 DF,  p-value: 4.951e-08
```
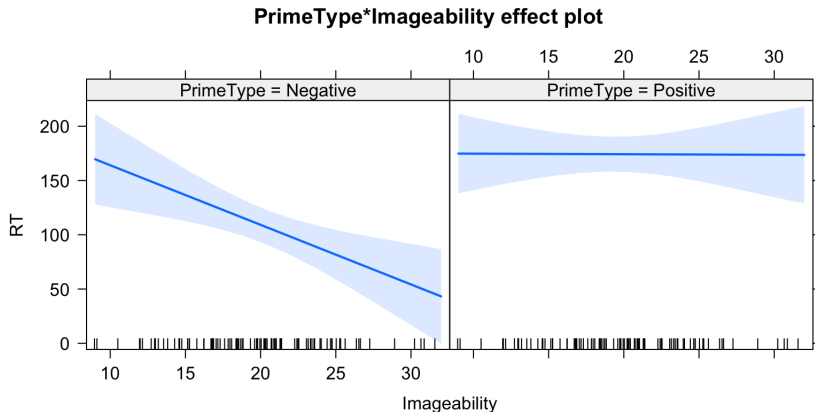
Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Interaction between a Factor and Continuous variable

Partial Effects of interaction of PrimeType and Imageability on RT

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Post hocs of mod5

Beware of interepretation of post-hocs in the case of interaction
between a factor and a continuous variable.

```
> pairs(emmeans(mod5, ~PrimeType)) #misleading
NOTE: Results may be misleading due to involvement in interactions
 contrast               estimate   SE df t.ratio p.value
 Negative - Positive      -63.4 11.4 96 -5.580  <.0001

> pairs(emmeans(mod5, ~PrimeType | Imageability))
Imageability = 19.7:
 contrast               estimate   SE df t.ratio p.value
 Negative - Positive      -63.4 11.4 96 -5.580  <.0001
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Post hocs of mod5

It could be very informative to test the difference in the marginal
means at several points of the covariate.

```
> Imag_values= quantile(dat5$Imageability, probs = c(0.05, 0.5, 0.90))
> pairs(emmeans(mod5, ~PrimeType | Imageability, at = list(Imageability = Imag_values)))
Imageability = 12.1:
 contrast              estimate   SE df t.ratio p.value
 Negative - Positive     -22.3 21.4 96  -1.043  0.2995

Imageability = 19.8:
 contrast              estimate   SE df t.ratio p.value
 Negative - Positive     -63.9 11.4 96  -5.628  <.0001

Imageability = 25.3:
 contrast              estimate   SE df t.ratio p.value
 Negative - Positive     -94.2 17.6 96  -5.359  <.0001
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Note about terms in lm

It is easy do guess that the higher the number of variables of interest (fixed effects), the higher the number of coefficients to be estimated. If you take into account all the possible interactions, they grows exponentially.

Usually, in linear model. The choice of what coefficients and interactions is part of **Model Selection** (see the section here)

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Results of Regressions

Checking regression results **not just about p-values**.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions

A regression analysis is meaningful only if some *assumptions* are met.

One crucial assumption (see also these slides) is the **independence** of observations. This means that each observation should be unrelated to the others.

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

more on dependency

Suppose you have an experiment in which you want to investigate the effect of frequency and Length of stimuli on Reaction times in two different experimental conditions (and each Participant is administered with both conditions).

can you use linear regressions?

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

more on dependency

Suppose you have an experiment in which you want to investigate
the effect of frequency and Length of stimuli on Reaction times in
two different experimental conditions (and each Participant is
administered with both conditions).

can you use linear regressions?

NO

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

from Linear regression to mixed effect models

In the example of the preceding slide there is a known dependency which is not properly taken into accoun in the model.

This lead to several problems: you cannot trust p-values, and you cannot trust the estimated parameters.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

from Linear regression to mixed effect models

To solve the problem you should specify some-way the dependencies. And you can do that with **Mixed Effect Models**

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Linear Mixed Models

The following is one of the simplest formula for a mixed model. In this simple case it is basically the same of a regression with the addition of a term.

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \mathbf{S_1} + \epsilon \tag{3}$$

see also here for further formulas

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

An example of Mixed effect model

In the next slides I will refer to an example, (with simulated data). You can download the data from <u>here</u>. together with an accompanying R scripts.

The data refer to an hypothetic esperiment:
Priming effects on Metaphors and Literal sentences

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

An example of Mixed effect model (example1)

This experiment[2] was conducted on 30 participants. The experiment includes **two factors**:

- Stimulus type (two levels: Metaphor vs Literal)
- Priming type (two levels: Positive vs Negative)

We are also interested on the potential role of **two continuous variables** associated with the stimuli that could have some role

- Complexity
- Imageability

Each stimulus type (Metaphor and Literal) has **50 stimuli**, administered twice (Postive and Negative prime)

---

[2]This is the example1 in the accompanying code

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

An example of Mixed effect model

Let's start from an hypotetical regression syntax including only predictors (i.e. *fixed effects*)

```
mod.lm = lm(RT ~ StimType*PrimeType + Imageability + Complexity,
data=dat)
# WRONG: dependencies are not considered.
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

An example of Mixed effect model

Let's start from an hypotetical regression syntax including only predictors (i.e. *fixed effects*)

```
mod.lm = lm(RT ~ StimType*PrimeType + Imageability + Complexity,
data=dat)
# WRONG: dependencies are not considered.
```

A mixed model

```
mod.lmer = lmer(RT ~ StimType*PrimeType + Imageability + Complexity +
(1 | Subject) + (1 | Item),  # random effects
data=dat)
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Mixed Effects Models

The fixed part of a Mixed model is basically the same as for a linear regression.

In addition, you specify the random effects, to specify dependencies in the data.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Results from LM

```
Call:
lm(formula = RT ~ StimType * PrimeType + Imageability + Complexity,
    data = dat)

Residuals:
    Min      1Q  Median      3Q     Max
-605.35 -123.86   -1.14  121.07  565.27

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                      493.9050    20.0185  24.672  < 2e-16 ***
StimTypeMetaphor                 -21.5665     5.8038  -3.716 0.000204 ***
PrimeTypePositive                -62.5149     5.7500 -10.872  < 2e-16 ***
Imageability                    -143.4412     2.1403 -67.020  < 2e-16 ***
Complexity                         0.7444     0.2335   3.188 0.001437 **
StimTypeMetaphor:PrimeTypePositive 105.4571   8.1318  12.968  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 181.8 on 7994 degrees of freedom
Multiple R-squared:  0.3857,Adjusted R-squared:  0.3853
F-statistic:  1004 on 5 and 7994 DF,  p-value: < 2.2e-16
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Results from LMM

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ StimType * PrimeType + Imageability + Complexity + (1 |
    Subj_ID) + (1 | Item_ID)
   Data: dat

REML criterion at convergence: 93168.7

Scaled residuals:
    Min      1Q   Median      3Q      Max
-2.75299 -0.76198 -0.04284  0.73742  3.01997

Random effects:
 Groups   Name        Variance Std.Dev.
 Item_ID  (Intercept)  9569     97.82
 Subj_ID  (Intercept) 18649    136.56
 Residual              5949     77.13
Number of obs: 8000, groups:  Item_ID, 200; Subj_ID, 20

Fixed effects:
                                  Estimate Std. Error        df t value Pr(>|t|)
(Intercept)                       493.9050    74.4849  199.2845   6.631 3.06e-10 ***
StimTypeMetaphor                  -21.5665    14.3047  201.8251  -1.508    0.133
PrimeTypePositive                 -62.5149     2.4390 7778.9995 -25.631  < 2e-16 ***
Imageability                     -143.4412     7.3385  196.0014 -19.546  < 2e-16 ***
Complexity                          0.7444     0.8006  196.0015   0.930    0.354
StimTypeMetaphor:PrimeTypePositive 105.4571    3.4493 7778.9995  30.574  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Correlation of Fixed Effects not shown)
```

Introduction
Linear Models
Linear Mixed Models
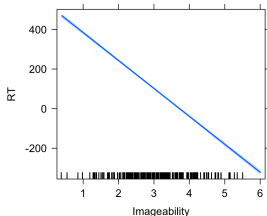Checking Assumptions
Model Selection
Advanced Topics

A first glance

At a first glance, there are some differences between lm and lmer, both in the output, and in the p-values.

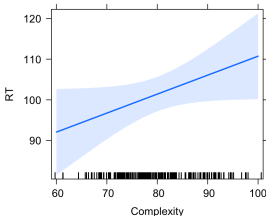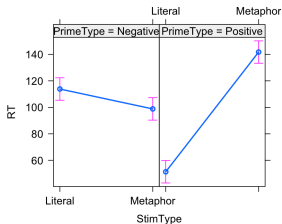This is related to the fact that some dependencies are taken into account.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Linear model on the data

partial effects of mod_lm

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Linear Mixed model on the data

partial effects of mod_lmer

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics
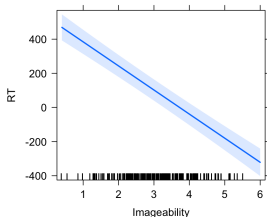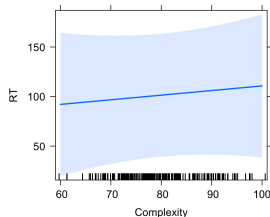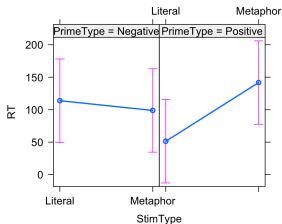
Linear model vs Linear Mixed Model

In the partial effects you can see how the error bands and bars are wider in the case of mixed model.

This is because in the linear model, each observation is treated basically as a separate Subject, and the model is fit like it is on thousands of separate (and independent) subjects (rather than on hundreds of Items adminsitered to dozens of subjects).

As such, there is an overestimation of the confidence in the estimated parameters.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Some terminology

Two crucial terms in Mixed Effect Models

- **Random Intercept**
- **Random Slope**

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Random Intercept

The **Random Intercept** represents an adjustment made to all observations that come from a specific level of a grouping variable. (e.g. the adjustment from all observation that come from the same participant)

In the mixed model estimation, it is associated to a standard deviation in the model.

The R syntax for specifying a random intercept is the following.

```
(1 | Subj_ID)
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Random Intercept in Psycholinguistic experiments

Typically in Psycholinguistic experiments there are two main random effects to be taken into account: participants, and stimuli (Clark, 1973).

This is easily taken into account in mixed effect models, specifying two crossed (i.e., independent) random effects:

```
(1 | Subj_ID)  + (1 | Item_ID).
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Random Intercept in Psycholinguistic experiments

In coarse words, adding Random intercepts in the models with the syntax:

```
(1|Subj_ID) + (1|Item_ID)
```

Indicates that the model should taken into account that there could be some dependencies in all observations that come from the same Subject and the same Item. In particular, some subjects could be in general faster then others or slower than others, or that responses to some Items could be faster the those to others.

This should be taken into account in assessing significance and in determining results.

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

A further note on Random Intercept

There is another way to interpret Random Intercepts and is about generalizability of results.

- (1 | Subj_ID) : will the results generalize to new Subjects?
- (1 | Item_ID): will the results generalize to new Items?

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Variability in Subjects

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics
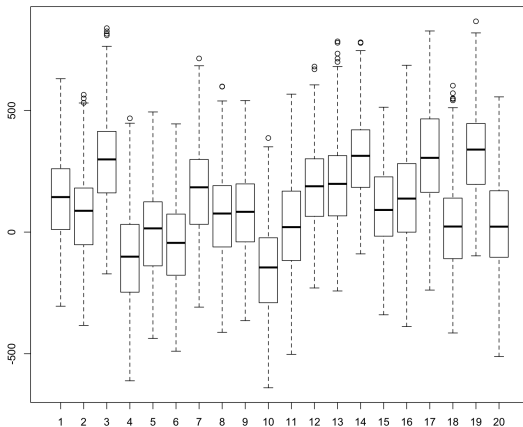
Variability in Items
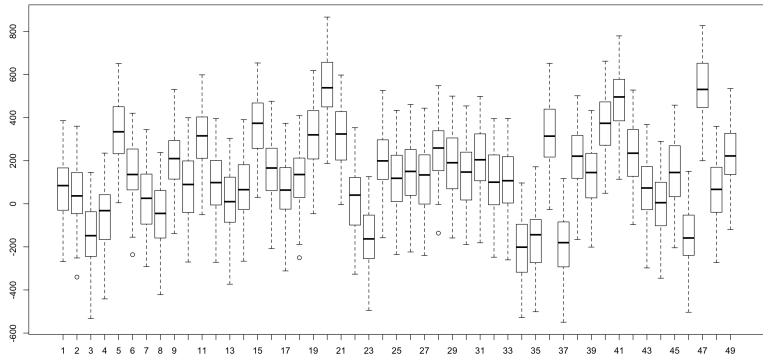
First 50 Items among 200 are shown

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Checking random effects

The adjustments of random effects can be inspected by using this command:

```
> ranef(mod.lmer1)
$Item_ID
    (Intercept)
1     18.748191
2   -167.967693
3   -172.824119
4     -3.032910
5     70.462635
```

The result indicates the adjustments to be made. For example, responses to Item 3 are particularly fast.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interim Note

When you use a simple ANOVA, actually the within-Subject dependancy is normally taken into account. To avoid problems, it is common to average all Items. This solution avoids dependency, but also prevents the possibility to generalize to other items.

In the past other analysis (like F_1 and F_2, F^1 or $MinF^1$, were used, see Baayen, 2008).

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Random Slope

The **Random Slope** is an adjustment in the model that try to take into account the fact that a fixed effect may be different across a grouping variable. For example, it may be the case that an experimental manipulation (e.g., Metaphor vs Literal) has a different effect in all subjects, and that some subjects may show a different effect.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics
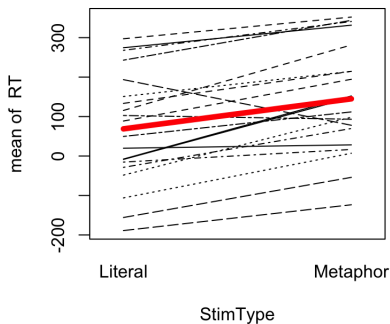
Random Slope

Black lines, are by subject slopes.
Red line is the group slope (averaging all subjects)

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Random Slope

In mixed models random slopes are described (and modeled) by a standard deviation and correlation.

The standard R syntax for specifying a random slope is the following.

```
(1 + StimType | Subj_ID)
```

Note that the syntax above, specify both a random slope, and a random intercept.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Random Slope

The **Random Slope** specification is normally associated to more than one parameter. One is the standard deviation of the slope, the other is the correlation between this slope with the Intercept.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model with random slope

A model with a Random Slope

```
lmer(RT ~ StimType * PrimeType + Imageability + Complexity +
 (1 + StimType | Subj_ID) + (1 | Item_ID), data = dat)
```

This model include the possibility that the slope associated to StimType (i.e., the effect of StimType) may be different across Subjects. We can include the information in the model as additional parameters.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model with random slope

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ StimType * PrimeType + Imageability + Complexity + (1 +
    StimType | Subj_ID) + (1 | Item_ID)
   Data: dat

REML criterion at convergence: 91745.2

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.60713 -0.74448  0.00371  0.72035  2.85084

Random effects:
 Groups   Name              Variance Std.Dev. Corr
 Item_ID  (Intercept)        9594     97.95
 Subj_ID  (Intercept)       20365    142.71
          StimTypeMetaphor   4292     65.51   -0.30
 Residual                    4900     70.00
Number of obs: 8000, groups:  Item_ID, 200; Subj_ID, 20

Fixed effects:
                                    Estimate Std. Error      df t value Pr(>|t|)
(Intercept)                         493.9050    75.0524 194.2459   6.581 4.25e-10 ***
StimTypeMetaphor                    -21.5665    20.4617  65.2559  -1.054    0.296
PrimeTypePositive                   -62.5149     2.2136 7759.9899 -28.241  < 2e-16 ***
Imageability                       -143.4412     7.3380 196.0445 -19.548  < 2e-16 ***
Complexity                            0.7444     0.8005 196.0442   0.930    0.354
StimTypeMetaphor:PrimeTypePositive  105.4571     3.1305 7759.9899  33.687  < 2e-16 ***
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Model with random slope

The random slope is associated with a standard deviation and a correlation. This last parameter is the correlation between the slope of the Subject and the Intercept.

This is taken into account because the slope and Intercept adjustments may be correlated, and this could be included in the model (to further consider dependencies in the data).

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Linear mixed model on the data (Random Slope)
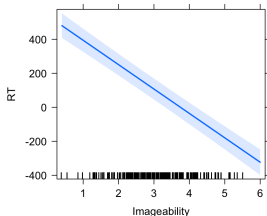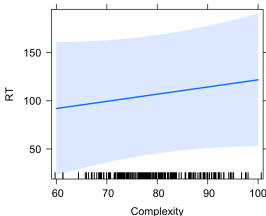
partial effects of mod_lmer2

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

Did you notice something about the way I modeled the Random Slope?

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics
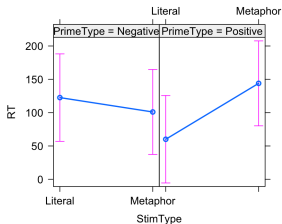
About Random Slope

Did you notice something about the way I modeled the Random Slope?

What about the `PrimeType`?

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

I could fit a more complex model

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

I could fit a more complex model

```
lmer(RT ~ StimType * PrimeType + Imageability + Complexity +
 (1 + StimType*PrimeType | Subj_ID) + (1 | Item_ID), data = dat)
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

I can also model the possibility that the effect of imageability
changes across participants!

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

I can also model the possibility that the effect of imageability
changes across participants!

```
lmer(RT ~ StimType * PrimeType + Imageability + Complexity +
 (1 + StimType*PrimeType | Subj_ID) + ( 1 + Imageability | Subject_ID) +
(1 | Item_ID)  , data = dat)
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

And what about an even more complex random structure?

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

And what about an even more complex random structure?

```
lmer(RT ~ StimType * PrimeType + Imageability + Complexity +
 (1 + StimType*PrimeType*Imageability | Subj_ID)  +
(1 | Item_ID)  , data = dat)
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

About Random Slope

And what about an even more complex random structure?

```
lmer(RT ~ StimType * PrimeType + Imageability + Complexity +
 (1 + StimType*PrimeType*Imageability | Subj_ID)  +
(1 | Item_ID)  , data = dat)
```

Although (in theory) it is possible to fit this kind of models, they typically don't work, leading to a problem in *convergence*

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Convergence issues

Fitting a mixed model is a process that can fail. The more common problems are issue in **convergence**

Broadly speaking, sometimes the algorithms beyond mixed models (or `lmer`) are not able to find a solution.

The large majority of the time, problems in convergence are related to random structure that are too much complicated.

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Convergence issues

Sometimes the model converge (so no Error), but a Warning about convergence appears.

In such cases, something did not go smooth in the fit, and you may want to check your model better. However this does not mean that the fit is wrong. Convergence issues is an advanced topic, beyond the scope of this slides (but see here)

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Other random structures

To improve possibility of convergence, one of the most common strategy is to remove the correlation between Intercept and slope. The syntax is the following.

```
(1  Subj_ID) + (0 + StimType  Subj_ID)
# uncorrelated random-intercepts and random slopes
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Other possibilities

There are other possibilities of random structure. See for example
here or here

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

How to choose the Random structure?

So how to choose the random structure?

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

How to choose the Random structure?

So how to choose the random structure?

1. Keep it Maximal"(Barr et al., 2013)
2. "Parsimonious Mixed Models "(Bates et al., 2015)

These are the two different main perspectives in choosing the random structure with pros and cons.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Maximal Random structure

The maximal structure is univocal, and according to the simulations by Barr et al. (2013), it is also the more robust choice. The downsides is that as the design become more and more complicated, the random structure will become *extremely* complicated, and it is very lilkely that convergence issue will occur.

The maximal random struture is basically the structure the include, for every intercept, all the possible interactions of fixed effects. In an example with StimType, and PrimeType as fixed effects, and Sub_ID and Item_ID as random variables the model would be:

```
RT ~ StimType*PrimeType +
(1+StimType*PrimeType | Subj_ID) + (1+StimType*PrimeType | Item_ID)
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Parsimonious Random structure

The parsimonious random structure is a reasonable random structure guided by the data, and that try to avoid overparametrizing the models.

It is a bit more difficult to determines a univocal parsimonious structure but typically it includes all relevant random intercepts, and also (if necessary) a random slope for some crucial variables guided from theory (for example, the main experimental variable) or from checks (see next slides).

```
# a potential parsimonious mixed model
RT ~ StimType*PrimeType +
(1 + StimType | Subj_ID) + (1 | Item_ID)
```

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interim Summary

So far we have discussed the main following points:

- Linear Mixed models are an extension of Linear regression
- Linear Mixed models have also this "random part" which may be a bit difficult to define.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Next steps

Have we discussed all topics related to linear regressions, relevant
for Mixed models?

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Next steps

Have we discussed all topics related to linear regressions, relevant for Mixed models?

Absolutely not. There are at least two major points remaining:

- checking the assumptions (and solving problem when assumptions are not met)
- model selection

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

A potential pipeline for mixed model analysis steps
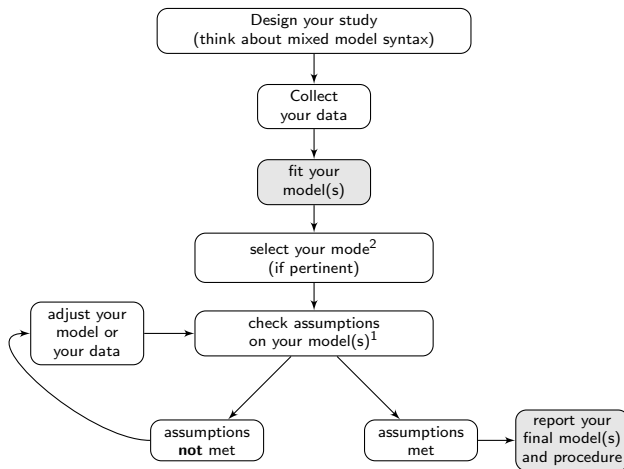
([1] note that some checks may be done even before the model is fit)
([2] assumption checks could be done also before model selection)

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

A potential pipeline for mixed model analysis steps

Check also the github repo associated with these slides for a
practical example of a pipeline.

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions

The assumptions to be checked for mixed effect models, are mostly the same to be checked for regression.

Often, checking assumptions implies performing some **diagnostics** on the fit.

Diagnostics may be numbers or (more often) plots.

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions

1 Linear relationship between predictors and dependent variable.

2 Homoscedasticity of the residuals.

3 Normality of the residuals.

4 No problems in multicollinearity of predictors.

5 Independence of residuals.
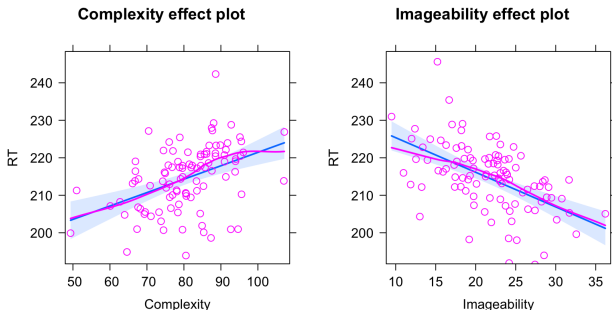
6 Independence of observations.

see also here
note that this list may be different across different sources.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (1)

1. Linear relationship between predictors and dependent variable.
   $\rightarrow$ plot partial residuals

```
plot(allEffects(mod2, partial.residuals=T))
```


Partial residuals of the two effects

Introduction
Linear Models
Linear Mixed Models
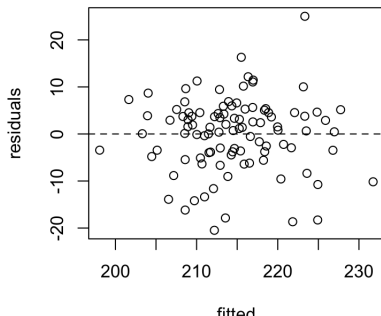Checking Assumptions
Model Selection
Advanced Topics

Solving issues (1)

If partial residuals suggest non linearity is present, you may add a transformation to the predictor to account for the non-linearity (see Baayen, 2008)

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (2)

2 Homoscedasticity of the residuals $\rightarrow$ plot residuals against fitted.

```
plot(fitted(mod2), resid(mod2))
```

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Solving issues (2)

Try to check if you are missing a relevant predictor or a random effect that can increase quality of fit (see Model Selection section)

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (3)

3 Normality of the residuals $\rightarrow$ plot residuals against fitted.

`hist(resid(mod2))`

Histogram of resid(mod2)

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (3)

Another common way to inspect normality of the residuals is to use a qqplot.

Normal Q-Q Plot

Introduction
Linear Models
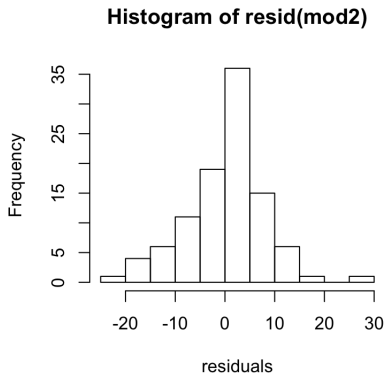Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics
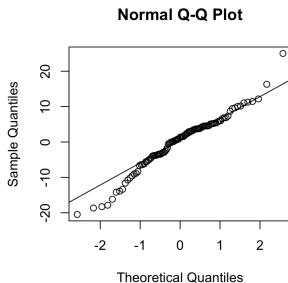
Solving issues (3)

Try to check if you are missing a predictor or a random effect that can increase quality of fit

Another commoon strategy is to refit the model excluding the outliers, to see if they influenced the results.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (4)

4 No problems with multicollinearity.

Multicollinearity is an index of the overall correlation between predictors that may affect parameter estimation. Remember, in regression each term is estimated to be uncorrelated from the others. What if they are highly correlated? See here

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (4)

4 No problems with multicollinearity.
   $\rightarrow$ check if Variance Inflation Factor < 10

```
vif(mod3) # note: I need at least two
# predictors to check collinearity

    PrimeType Imageability
    1.025689     1.025689
```

Note that in some cases GVIF is computed. See <u>here</u> for details or
`help(vif)`

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Solving issues (4)

You have several choices:

- remove one of the collinear variables.
- run a Principal Component Analysis (PCA) to solve the issue.
- you may residualize on variable on the other. (see here)

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (5)

5 Independence of residuals $\rightarrow$ use some measures or plot to check autocorrelation

Series resid(mod2)

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics
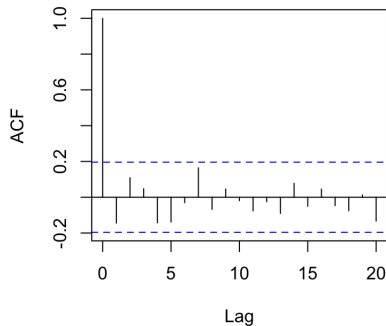
Solving issues (5)

Autocorrelation are typically related to trial order.

- you can add a covariate with Trial Number and a random slope by participant nTrial + (1 + nTrial | Subj_ID)
- you can add, as covariate, the value for the preceding trial. (see Baayen & Milin, 2010)
- you may want to use extenstion able to deal with authocorrelations as Generalized Additive MIxed Models (GAM) see this wonderful tutorial by Jacolien Van Rij.

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Assumptions in Regressions (6)

6 Independence of observations

$\rightarrow$ this is theoretically defined (e.s. observations come from the same participant)

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

Solving issues (6)

Use a mixed-effect model with appropriate random structure.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model Selection

When using mixed effect models (or in general when fitting models) there are many opportunities to perform the analysis.

This is different for example, from ANOVA, when given an experimental design, there is typically just one possible analysis.

We have seen a hint of this when seeing that several random structures can be used to analyze the same data. Hosever, the flexibility is not only for random part, but also for the fixed part.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection strategies

There are three main *model selection* strategies.

1 Fit just one model (theoretically driven), inspect the results.

2 compare few theoretical driven models.

3 stepwise selection of the best fitting model

The choice depends on the hypotheses or experimental aims.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (1)

Fit just one model

This is the more conservative and robust choice. Given some hytpotheses, you fit a model and check the significance of the relevant terms, using Anova p-values to test for significance.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (2)

Compare few theoretical driven models

With this choice you can compare few theoretically driven models
to test which one fit the data better.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (2) - LRT

To do this you, in the case of *nested* [3] models, it is normally used the Likelihood ratio Test (LRT). To perform LRT:

```
anova(mod1, mod2)
```

If the test is significant, the more complex model is selected, otherwise, the simpler model is selected.

---

[3]a smaller model is *nested* to a larger one if it contains all the terms and variables of the larger models

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Interim clarification: nested vs non-nestd

Consider the three models below.

```
mod.lmer8a = lmer(RT ~ StimType*PrimeType +
                  (1| Subj_ID) + (1 | Item_ID),
                  data=dat)


mod.lmer8b = lmer(RT ~ StimType*PrimeType + Imageability
                  (1| Subj_ID) + (1 | Item_ID),
                  data=dat)

mod.lmer8d = lmer(RT ~ StimType*PrimeType + Concreteness + Complexity +
                   (1| Subj_ID) + (1 | Item_ID),
                  data=dat)
```

mod.lmer8a is nested on mod.lmer8b. On the other side the
model mod.lmer8d has no nesting relationship with the others.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (2) - LRT

The LRT can be used also to select the more appropriate random
structure, with the same principle of before (if not significant,
select the simplest random structure).

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (2) - LRT

Some notes:

1. when doing an LRT, a warning is returned. Data are fitted with `ML`. By default, mixed model fitting occurs using `REML` method[4]. However LRT cannot be used if models ar fit with `REML`. For this reasons, when using `anova` to compare models, the models are refit with the `ML`.

2. it makes no sense to compare with LRT two models in which both Random and fixed parts are different.

---

[4]for difference between `ML` and `REML` see <u>here</u> or <u>here</u>

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (2) - AIC, BIC

For non-nested models you can use some measure as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

For these measures, the lower the measure the better the fit. Given the same data, they do not need nested models.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (3) - stepwise selection

With the stepwise selection, a series of model are fit on the same data, and predictors or random variables, are included (or excluded) in a progressive fashion, up to reach the best fit the model.

These are not recommended (Harrel, 2015), as many models are fit, typically without taking into account properly an inflated risk of Type 1 error.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Model selection (3) - stepwise selection

Model selection can be made by hand, starting from a NULL model and adding fixed (or random) variables, and then checking wether adding variables lead to better models (using LRT or AIC/BIC).

An alternative is to start from a full model and delete (one at the time) the variables that has the worse performance (for example the highest p-value).

Model selection (3) - stepwise selection

You can do model selection authomatically using the step function, loaded with the lmerTest package.

In this cases you can select the direction of selection (forward or backward).

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Where to go from here?

To understand mixed models, the best way is to use them on your data.

You can start from the examples in the github repository associated with these slides (see <u>here</u>)

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Advanced Topics

There are many developments of mixed effect models. Here some of the more common examples:

- Generalized Linear Mixed Effect Models (GLMM).
- accounting for non-linearities.
- Generalized Additive Mixed Models (GAMM).

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

GLMM

An assumpton of Linear Mixed Models (LMM) is that the dependent variables is continous.

GLMM allow to relax this assumption and use mixed models with dependent variables with different distributions (e.g. Binomial, Poisson, etc.) The more common model used for GLMM is with

binomial distribution. In these models the answer (at single trial level) can be either Correct (typically codified with 1 or wrong, codified with 0).

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

GLMM

In GLMM a common issue is that they are even more prone to have convergence issues.

For this reason it is better to avoid complicated design with GLMM.

Using GLMM in R typically requires minimal changes, that is to specifiy the family of the distribution of the data.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Non-linearities

So fare we discussed about linear relationships and we also have
seen that one assumption is about linear relationships between
predictors and dependent variable (see these slides).

However it is possible to also include non-linear relationships. See
book by Baayen (2008), freely available online Donwload.

Introduction
Linear Models
Linear Mixed Models
Checking Assumptions
Model Selection
Advanced Topics

GAMM

Generalized Additive Mixed Models (GAMM) represent a further development. In GAMM allows to flexibly model non-linearities in a bottom-up fashion. Basically the model itself try to understand if a non-linear function is better to explain the data.

GAMM are particularly powerful for handling interactions between continuous variables, which are quite an issue in linear models.

GAMM allows also flexible ways to model random effects, in particular, introducing the *random smooth* terms.

This tutorial by Jacolien Van Rij, provide a nice introduction to the topic

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Recommendations (before the experiment)

1. Use as many paticipants, and as many item as possible.
2. Avoid too much complicated design (especially if you have few participants and/or few items)
3. Make sure to have variability in the data.

Introduction
Linear Models
Linear Mixed Models

Checking Assumptions
Model Selection
Advanced Topics

Recommendations (after the experiment)

1. Spend some time to check the data, and that everything is good.
2. use plots to check the data.
3. start the analysis with specific hypotheses.