Tutorial: Exploring Random Effects

What Do Participants and Items Tell us Beyond the Fixed Effects?

Jalal Al-Tamimi

13 December 2024

Outline

- Fixed and random effects?
- Random Intercepts and Slopes?
- Why are they important in linguistics' research? And beyond?

Intro

- In linguistics (and other disciplines), we rarely use data coming from one participant and/or from one item/utterance (or corpora)
- Having multiple participants and/or items/utterances allows to reduce Type I error, controls for Type II error, Type S error and increases power.

Intro

- In linguistics (and other disciplines), we rarely use data coming from one participant and/or from one item/utterance (or corpora)
- Having multiple participants and/or items/utterances allows to reduce Type I error, controls for Type II error, Type S error and increases power.
- But choosing the right model is an *art* and depends heavily on the dataset you are working on.

Intro

- In linguistics (and other disciplines), we rarely use data coming from one participant and/or from one item/utterance (or corpora)
- Having multiple participants and/or items/utterances allows to reduce Type I error, controls for Type II error, Type S error and increases power.
- But choosing the right model is an *art* and depends heavily on the dataset you are working on.

1. $y = x\beta + \varepsilon$
    - $y \rightarrow$ outcome (DV) $\Rightarrow$ known
    - $x \rightarrow$ fixed effect (IV) $\Rightarrow$ known
    - $\beta \rightarrow$ coefficient of fixed effect $\Rightarrow$ unknown
    - $\varepsilon \rightarrow$ random error term $\Rightarrow$ unknown

Intro

- In linguistics (and other disciplines), we rarely use data coming from one participant and/or from one item/utterance (or corpora)
- Having multiple participants and/or items/utterances allows to reduce Type I error, controls for Type II error, Type S error and increases power.
- But choosing the right model is an *art* and depends heavily on the dataset you are working on.

1. $y = x\beta + \varepsilon$
    - $y \rightarrow$ outcome (DV) $\Rightarrow$ known
    - $x \rightarrow$ fixed effect (IV) $\Rightarrow$ known
    - $\beta \rightarrow$ coefficient of fixed effect $\Rightarrow$ unknown
    - $\varepsilon \rightarrow$ random error term $\Rightarrow$ unknown
2. $y = x\beta + \varepsilon + Zu$
    - $Z \rightarrow$ random effects term $\Rightarrow$ known
    - $u \rightarrow$ random effects coefficients $\Rightarrow$ unknown

Types of Errors?

1. Type I (or "false positive") $\Rightarrow$ falsely concluding there is an effect when none exists. (generally *rightarrow* inaccurate modelling strategies)
2. Type II (or "false negative") $\Rightarrow$ falsely concluding there is no effect when one in fact exists (generally *rightarrow* inaccurate modelling strategies)

		Statistical analysis result (**sample**)	
		Reject H_0	Don't reject H_0
Reality	H_0 is true	Type I error (α)	Correct decision (null result)
(**population**)	H_0 is false	Correct decision (significant)	Type II error (β)

3. Type S $\Rightarrow$ Inaccurate sign (generally due to *hidden* multicollinearity and low power = 1- β)
4. Type M $\Rightarrow$ Inaccurate magnitude (generally due to *hidden* multicollinearity and low power = 1- β)

Sonderegger, M. (2023). *Regression Modeling for Linguistic Data*. The MIT Press.

Fixed and random effects?

- In Linguistics (and beyond), we rarely use productions of one thing, from one speaker and from one item $\Rightarrow$ No ability to generalise and uncover language-specific patterns.

1. Multiple speakers
2. Multiple Items (words)
3. Multiple utterances where words are embedded
4. Multiple listeners in perception experiments

Fixed and random effects?

- In Linguistics (and beyond), we rarely use productions of one thing, from one speaker and from one item $\Rightarrow$ No ability to generalise and uncover language-specific patterns.

1. Multiple speakers
2. Multiple Items (words)
3. Multiple utterances where words are embedded
4. Multiple listeners in perception experiments

- Fixed effects :
  - Are those that are part of the experimental conditions, if you have exhausted all of its levels

Fixed and random effects?

- In Linguistics (and beyond), we rarely use productions of one thing, from one speaker and from one item $\Rightarrow$ No ability to generalise and uncover language-specific patterns.

1. Multiple speakers
2. Multiple Items (words)
3. Multiple utterances where words are embedded
4. Multiple listeners in perception experiments

- Fixed effects :
  - Are those that are part of the experimental conditions, if you have exhausted all of its levels
- Random effects:
  - Are random selections of the **population** you have and you want to generalise over them.
  - These are $\Rightarrow$ **subjects**; **Items**; **Utterances**; **corpora**; etc.
  - You are not using all the **population** of subjects, listeners, items, or utterances in your data!

Fixed and random effects?

- In Linguistics (and beyond), we rarely use productions of one thing, from one speaker and from one item $\Rightarrow$ No ability to generalise and uncover language-specific patterns.

1. Multiple speakers
2. Multiple Items (words)
3. Multiple utterances where words are embedded
4. Multiple listeners in perception experiments

- Fixed effects :
    - Are those that are part of the experimental conditions, if you have exhausted all of its levels
- Random effects:
    - Are random selections of the **population** you have and you want to generalise over them.
    - These are $\Rightarrow$ **subjects**; **Items**; **Utterances**; **corpora**; etc.
    - You are not using all the **population** of subjects, listeners, items, or utterances in your data!

Fixed and random effects?

- In Linguistics (and beyond), we rarely use productions of one thing, from one speaker and from one item $\Rightarrow$ No ability to generalise and uncover language-specific patterns.

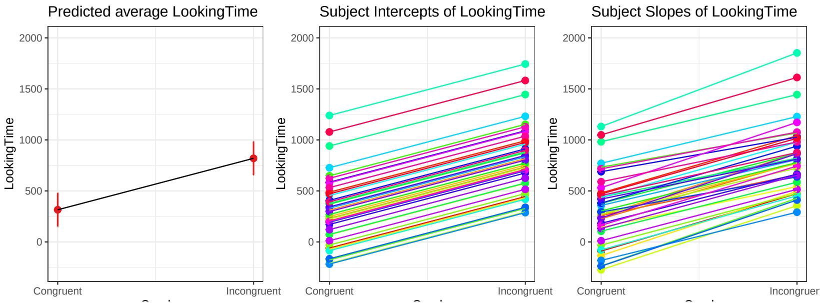
1. Multiple speakers
2. Multiple Items (words)
3. Multiple utterances where words are embedded
4. Multiple listeners in perception experiments

- Fixed effects :
  - Are those that are part of the experimental conditions, if you have exhausted all of its levels
- Random effects:
  - Are random selections of the **population** you have and you want to generalise over them.
  - These are $\Rightarrow$ **subjects**; **Items**; **Utterances**; **corpora**; etc.
  - You are not using all the **population** of subjects, listeners, items, or utterances in your data!

Random Intercepts and Random Slopes?

- Random Intercepts ⇒ averages of your **population**; and these are used in your statistical model to estimate the population-specific error term
- Random Slopes ⇒ adjustments to your **populations'** observations as a function of your fixed effects (within-subject or within-item)

```
plot_model(xmdl.Optimal, type = "pred", terms = "Cond", ci.lvl = 0.95, dodge = 0, show.legend = FALSE,
           title = "Predicted average LookingTime") + theme_bw() + geom_line() +
  coord_cartesian(ylim = c(-275, 2000)) +
plot_model(xmdl.rand.Interc, type = "pred", terms = c("Cond", "Subj"), pred.type = "re", ci.lvl = NA, dodg
           title = "Subject Intercepts of LookingTime") + theme_bw() + geom_line() +
  coord_cartesian(ylim = c(-275, 2000)) +
plot_model(xmdl.Optimal, type = "pred", terms = c("Cond", "Subj"), pred.type = "re", ci.lvl = NA, dodge =
           title = "Subject Slopes of LookingTime") + theme_bw() + geom_line() +
  coord_cartesian(ylim = c(-275, 2000))
```



A concrete example

We use a simulated dataset with ⇒ 40 subjects responded to a task involving 40 items in a fully crossed design, with two IVs: **Condition** with `congruent` and `incongruent` (within-subject and within-item) and **Age** with `young` and `old` (between-subject and within-item). The DV is LookingTime (in msec)

```
set.seed(42)
# define parameters
Subj_n = 40  # number of subjects
Item_n = 40  # number of items
b0 = 100       # intercept
b1 = 2.5 * b0   # fixed effect of condition
u0s_sd = 300   # random intercept SD for subjects
u0i_sd = 200   # random intercept SD for items
u1s_sd = 100   # random b1 slope SD for subjects
u1i_sd = 50   # random b1 slope SD for items
r01s = -0.3     # correlation between random effects 0 and 1 for subjects
r01i = 0.2     # correlation between random effects 0 and 1 for items
sigma_sd = 150 # error SD
# set up data structure
dataCong <- add_random(Subj = Subj_n, Item = Item_n) %>%
  # add within and then between categorical variable for subject
  add_within("Subj", Cond = c("Congruent", "Incongruent")) %>%
  add_recode("Cond", "Cond.Incongruent", Congruent = 0, Incongruent = 1) %>%
  add_between("Subj", Age = c("Young", "Old")) %>%
  add_recode("Age", "Age.Old", Young = 0, Old = 1) %>%
  # add random effects
  add_ranef("Subj", u0s = u0s_sd, u1s = u1s_sd, .cors = r01s) %>%
  add_ranef("Item", u0i = u0i_sd, u1i = u1i_sd, .cors = r01i) %>%
  add_ranef(sigma = sigma_sd) %>%
  # calculate DV
  mutate(LookingTime = b0 + b1 + u0s + u0i + #u0si + u1si +
          (((b1 + u1s) + 0.5) * Cond.Incongruent) + (((b1 + u1s) + 0.9) * Age.Old) +
          (((b1 + u1i) - 0.3) * Cond.Incongruent) + (((b1 + u1i) - 0.25) * Age.Old) + sigma)
```

RQ + Hypotheses

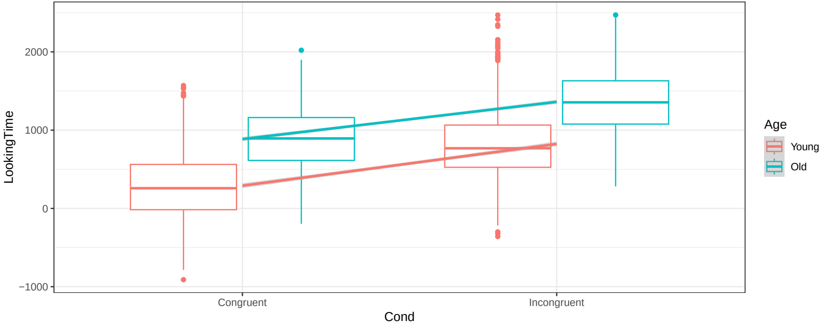
Our research question is as follows ⇒ `Age` of subject will impact the `Looking Time` in the two conditions.

Our hypothesis is ⇒ The older a subject is, the more the looking time it is to the incongruent condition

Visualisation I

An increase in LookingTime in the `incongruent` condition and overall, `older` participants show an increase in LookingTime. BUT there is no clear interaction

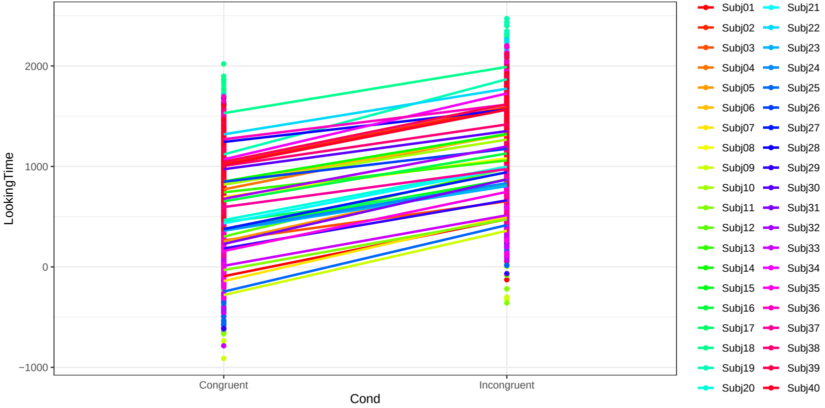
```
dataCong %>%
  ggplot(aes(x = Cond,
             y = LookingTime,
             colour = Age)) +
  theme_bw() +
  geom_boxplot() +
  geom_smooth(aes(as.numeric(Cond)), method = "lm")
```



Visualisation II

This figure shows that subjects are variable in how they responded to this task

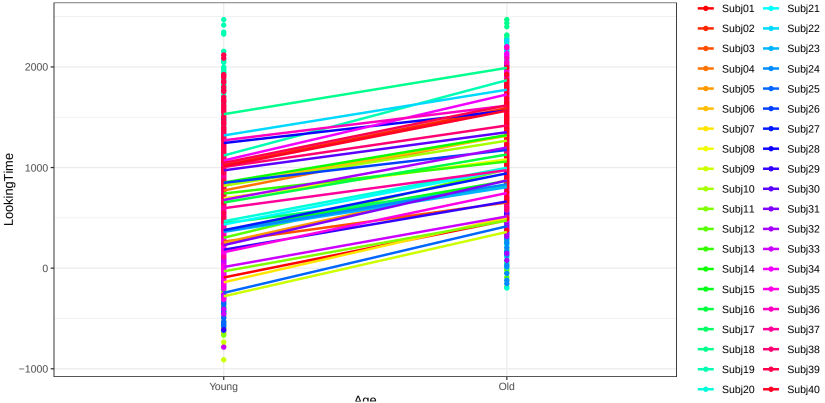
```
dataCong %>%
  ggplot(aes(x = Cond,
             y = LookingTime,
             colour = Subj)) +
  theme_bw() +
  geom_point() +
  geom_smooth(aes(as.numeric(Cond)), method = "lm", se = FALSE) +
  scale_colour_manual(values = paletteer_c("grDevices::rainbow", length(unique(dataCong$Subj))))
```



Visualisation III

This figure shows that subjects had an impact on the LookingTime in both age groups, simply due to their variable responses to the different items

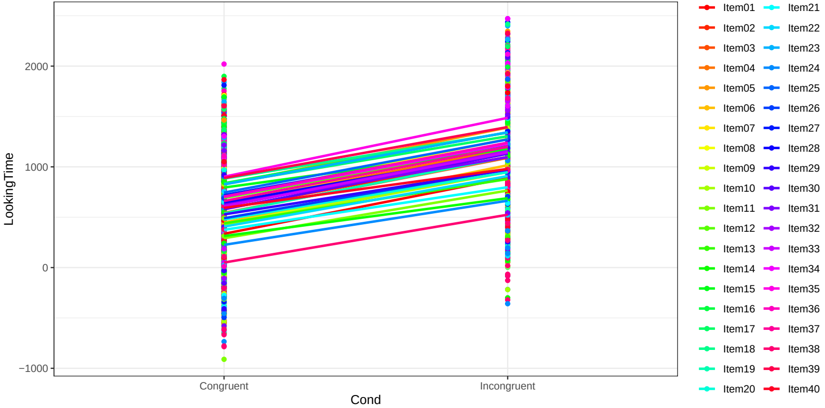
```
dataCong %>%
  ggplot(aes(x = Age,
             y = LookingTime,
             colour = Subj)) +
  theme_bw() +
  geom_point() +
  geom_smooth(aes(as.numeric(Cond)), method = "lm", se = FALSE) +
  scale_colour_manual(values = paletteer_c("grDevices::rainbow", length(unique(dataCong$Subj))))
```



Visualisation IV

This figure shows that items had an impact on the LookingTime in both conditions

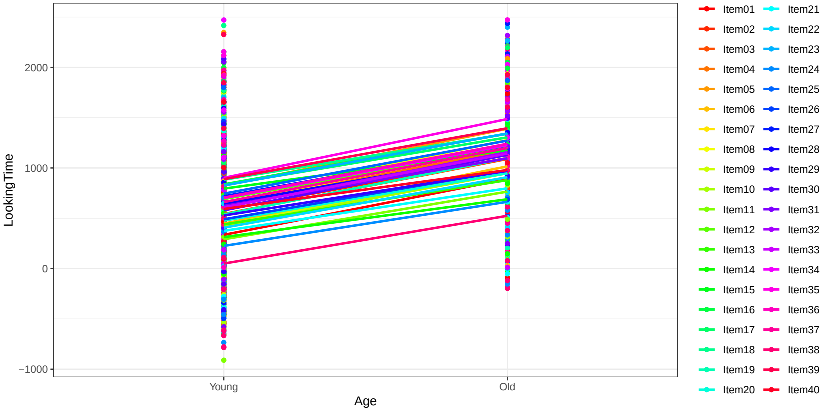
```
dataCong %>%
  ggplot(aes(x = Cond,
             y = LookingTime,
             colour = Item)) +
  theme_bw() +
  geom_point() +
  geom_smooth(aes(as.numeric(Cond)), method = "lm", se = FALSE) +
  scale_colour_manual(values = paletteer_c("grDevices::rainbow", length(unique(dataCong$Item))))
```



Visualisation V

This figure shows that items had an impact on the LookingTime in both age groups

```
dataCong %>%
  ggplot(aes(x = Age,
             y = LookingTime,
             colour = Item)) +
  theme_bw() +
  geom_point() +
  geom_smooth(aes(as.numeric(Cond)), method = "lm", se = FALSE) +
  scale_colour_manual(values = paletteer_c("grDevices::rainbow", length(unique(dataCong$Item))))
```



Modelling strategy I

Due to the variation observed in the data, one needs to model both random intercepts and random slopes.

```
## Crossed random intercepts
xmdl.rand.Interc <- dataCong %>%
  lmer(LookingTime ~ Cond + Age +
         (1 | Subj) +
         (1 | Item), data = ., REML = FALSE,
      control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
## Crossed random intercepts + By-speaker random slopes
xmdl.rand.Slope1 <- dataCong %>%
  lmer(LookingTime ~ Cond + Age +
         (1 + Cond | Subj) +
         (1 | Item), data = ., REML = FALSE,
      control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
## Crossed random intercepts + By-speaker and by-item random slopes
xmdl.rand.Slope2 <- dataCong %>%
  lmer(LookingTime ~ Cond + Age +
         (1 + Cond | Subj) +
         (1 + Cond | Item), data = ., REML = FALSE,
      control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
## Crossed random intercepts + By-speaker and by-item random slopes
xmdl.rand.Slope3 <- dataCong %>%
  lmer(LookingTime ~ Cond + Age +
         (1 + Cond | Subj) +
         (1 + Cond + Age| Item), data = ., REML = FALSE,
      control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
```

Modelling strategy II

We test with interactions

```
## Crossed random intercepts + Interaction
xmdl.rand.Interc.Int <- dataCong %>%
  lmer(LookingTime ~ Cond * Age +
         (1 | Subj) +
         (1 | Item), data = ., REML = FALSE,
       control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))## Crossed random interce
xmdl.rand.Slope1.Int <- dataCong %>%
  lmer(LookingTime ~ Cond * Age +
         (1 + Cond | Subj) +
         (1 | Item), data = ., REML = FALSE,
       control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
## Crossed random intercepts + By-speaker and by-item random slopes + Interaction
xmdl.rand.Slope2.Int <- dataCong %>%
  lmer(LookingTime ~ Cond * Age +
         (1 + Cond | Subj) +
         (1 + Cond | Item), data = ., REML = FALSE,
       control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
## Crossed random intercepts + By-speaker and by-item random slopes
xmdl.rand.Slope3.Int <- dataCong %>%
  lmer(LookingTime ~ Cond * Age +
         (1 + Cond | Subj) +
         (1 + Cond * Age| Item), data = ., REML = FALSE,
       control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
```

Model Comparison

We use a formal model comparison via a Maximum Likelihood χ^2 Test. Model `xmdl.rand.Slope3` is the optimal model as it improved the model fit over a simpler one

```
anova(xmdl.rand.Interc, xmdl.rand.Slope1, xmdl.rand.Slope2, xmdl.rand.Slope3, xmdl.rand.Interc.Int, xmdl.r
```

```
## Data: .
## Models:
## xmdl.rand.Interc: LookingTime ~ Cond + Age + (1 | Subj) + (1 | Item)
## xmdl.rand.Interc.Int: LookingTime ~ Cond * Age + (1 | Subj) + (1 | Item)
## xmdl.rand.Slope1: LookingTime ~ Cond + Age + (1 + Cond | Subj) + (1 | Item)
## xmdl.rand.Slope1.Int: LookingTime ~ Cond * Age + (1 + Cond | Subj) + (1 | Item)
## xmdl.rand.Slope2: LookingTime ~ Cond + Age + (1 + Cond | Subj) + (1 + Cond | Item)
## xmdl.rand.Slope2.Int: LookingTime ~ Cond * Age + (1 + Cond | Subj) + (1 + Cond | Item)
## xmdl.rand.Slope3: LookingTime ~ Cond + Age + (1 + Cond | Subj) + (1 + Cond + Age | Item)
## xmdl.rand.Slope3.Int: LookingTime ~ Cond * Age + (1 + Cond | Subj) + (1 + Cond * Age | Item)
##                      npar   AIC   BIC logLik deviance    Chisq Df Pr(>Chisq)
## xmdl.rand.Interc        6 42074 42110 -21031    42062
## xmdl.rand.Interc.Int    7 42050 42093 -21018    42036  25.8359  1  3.717e-07
## xmdl.rand.Slope1        8 41834 41883 -20909    41818 217.8699  1  < 2.2e-16
## xmdl.rand.Slope1.Int    9 41833 41888 -20908    41815   3.0247  1     0.0820
## xmdl.rand.Slope2       10 41808 41869 -20894    41788  27.3253  1  1.719e-07
## xmdl.rand.Slope2.Int   11 41807 41874 -20892    41785   3.0149  1     0.0825
## xmdl.rand.Slope3       13 41780 41858 -20877    41754  31.2599  2  1.629e-07
## xmdl.rand.Slope3.Int   18 41786 41895 -20875    41750   3.3401  5     0.6477
##
## xmdl.rand.Interc
## xmdl.rand.Interc.Int ***
## xmdl.rand.Slope1     ***
## xmdl.rand.Slope1.Int .
## xmdl.rand.Slope2     ***
## xmdl.rand.Slope2.Int .
## xmdl.rand.Slope3     ***
## xmdl.rand.Slope3.Int
```

Optimal model

We run the model via a REstricted Maximum Likelihood

```
xmdl.Optimal <- dataCong %>%
  lmer(LookingTime ~ Cond + Age +
         (1 + Cond | Subj) +
         (1 + Cond + Age | Item), data = ., REML = TRUE,
       control = lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e5)))
```

Summary

```r
summary(xmdl.Optimal)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: LookingTime ~ Cond + Age + (1 + Cond | Subj) + (1 + Cond + Age |
##     Item)
##    Data: .
## Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))
##
## REML criterion at convergence: 41724.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5337 -0.6485 -0.0054  0.6647  3.5358
##
## Random effects:
##  Groups   Name            Variance Std.Dev. Corr
##  Subj     (Intercept)     123480   351.40
##           CondIncongruent  10746   103.66   -0.26
##  Item     (Intercept)      38781   196.93
##           CondIncongruent   1872    43.27    0.31
##           AgeOld            1851    43.03   -0.13  0.69
##  Residual                  22613   150.38
## Number of obs: 3200, groups:  Subj, 40; Item, 40
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)       315.04      83.42   3.777
## CondIncongruent   504.35      18.54  27.204
## AgeOld            546.28     107.70   5.072
##
## Correlation of Fixed Effects:
##             (Intr) CndInc
## CndIncngrnt -0.120
## AgeOld      -0.646  0.016
```

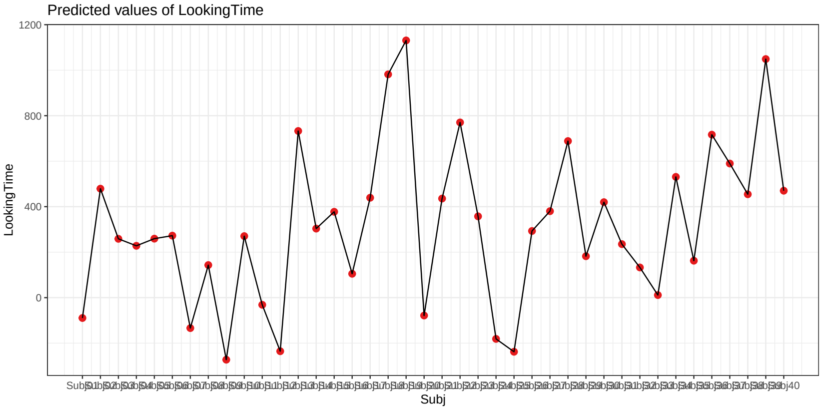
ANOVA

```r
Anova(xmdl.Optimal)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests
##
## Response: LookingTime
##         Chisq Df Pr(>Chisq)
## Cond 740.067  1  < 2.2e-16 ***
## Age   25.729  1  3.928e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.
```

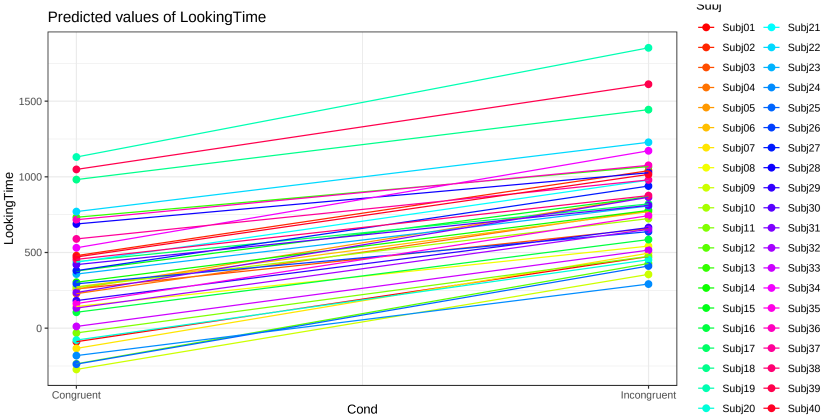
Subject specific-variation

```
plot_model(xmdl.Optimal, type = "pred", terms = "Subj", pred.type = "re",
           ci.lvl = NA, dodge = 0) + theme_bw() + geom_line()
```



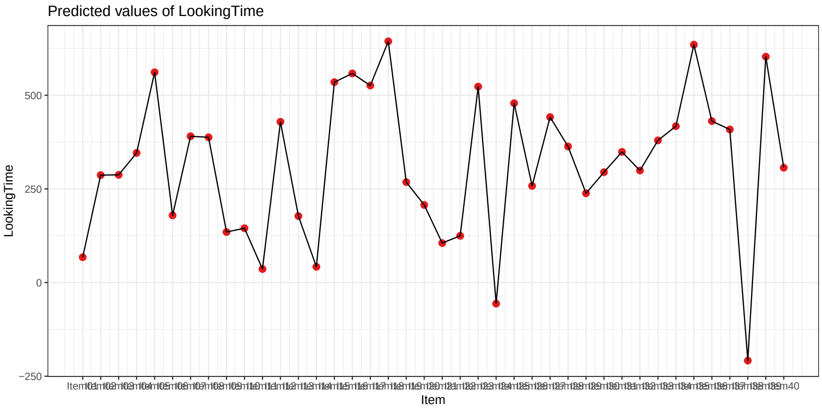
Subject specific-variation by Condition

```
plot_model(xmdl.Optimal, type = "pred", terms = c("Cond", "Subj"), pred.type = "re",
           ci.lvl = NA, dodge = 0, colors = paletteer_c("grDevices::rainbow", length(unique(dataCong$Subj)
  theme_bw() + geom_line()
```



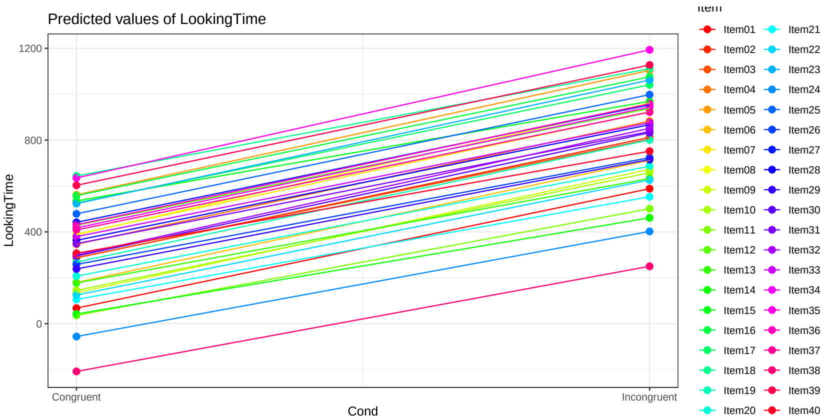
Item specific-variation

```
plot_model(xmdl.Optimal, type="pred", terms="Item", pred.type="re",
         ci.lvl = NA, dodge = 0) + theme_bw() + geom_line()
```



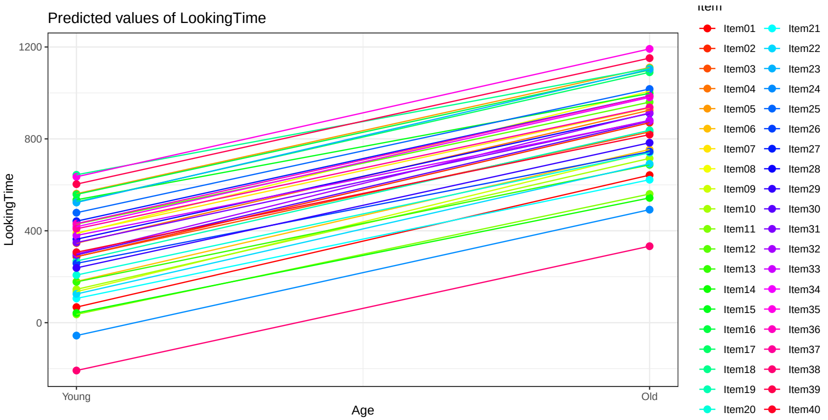
Item specific-variation by Condition

```
plot_model(xmdl.Optimal, type="pred", terms=c("Cond", "Item"), pred.type="re",
          ci.lvl = NA, dodge = 0, colors = paletteer_c("grDevices::rainbow", length(unique(dataCong$Subj)
  theme_bw() + geom_line()
```



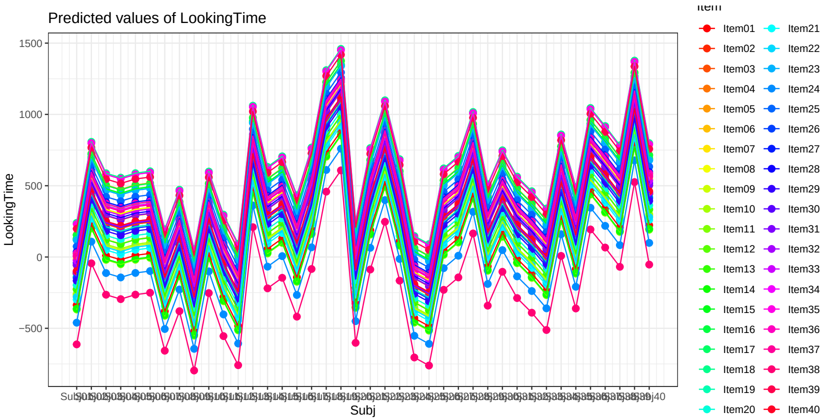
Item specific-variation by Age

```
plot_model(xmdl.Optimal, type="pred", terms=c("Age", "Item"), pred.type="re",
          ci.lvl = NA, dodge = 0, colors = paletteer_c("grDevices::rainbow", length(unique(dataCong$Item)
  theme_bw() + geom_line()
```



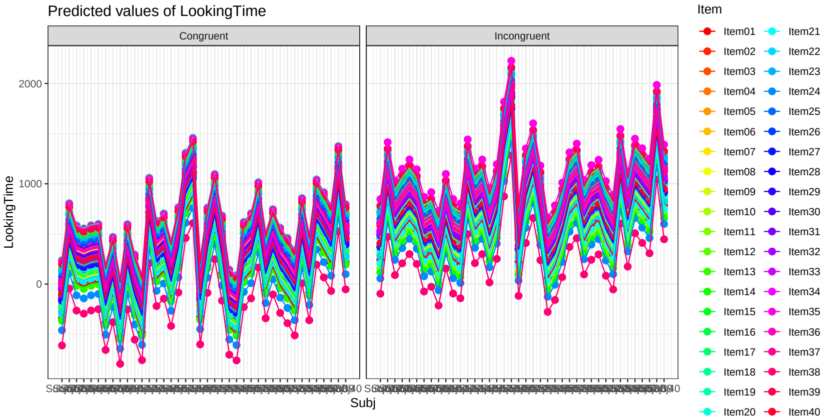
Item specific-variation by Subj

```
plot_model(xmdl.Optimal, type="pred", terms=c("Subj", "Item"), pred.type="re",
           ci.lvl = NA, dodge = 0, colors = paletteer_c("grDevices::rainbow", length(unique(dataCong$Item)
  theme_bw() + geom_line()
```



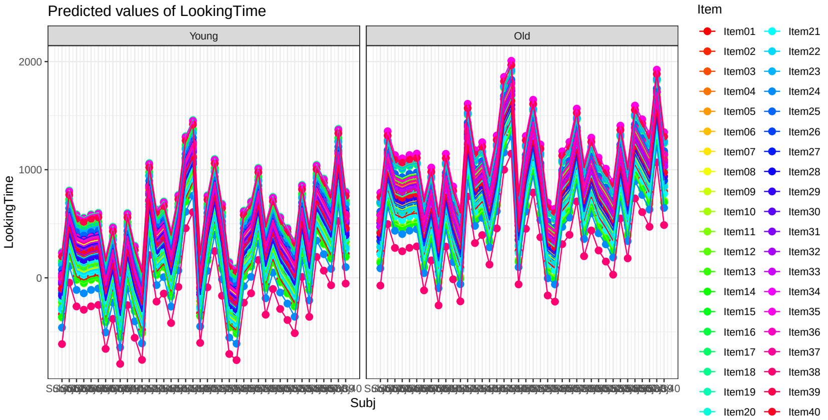
Item specific-variation by Subj by Cond

```r
plot_model(xmdl.Optimal, type="pred", terms=c("Subj", "Item", "Cond"), pred.type="re",
           ci.lvl = NA, dodge = 0, colors = paletteer_c("grDevices::rainbow", length(unique(dataCong$Subj)
  theme_bw() + geom_line()
```



Item specific-variation by Subj by Age

```
plot_model(xmdl.Optimal, type="pred", terms=c("Subj", "Item", "Age"), pred.type="re",
           ci.lvl = NA, dodge = 0, colors = paletteer_c("grDevices::rainbow", length(unique(dataCong$Subj)
  theme_bw() + geom_line()
```



Conclusion

- This tutorial showed how one can explore random effects and formally assess the need for Random slopes
- As a rule of thumb $\Rightarrow$ Any within-subject (or within-item) should be tested for a potential inclusion as a random slope
- Fixed effects provides averages over all observations, even when using mixed effects regressions; we need to explore what random effects (intercepts and slopes) tell us.
- In this example, we see that many subjects vary beyond the fixed effect; Standard Errors are not enough to quantify this type of variation. The same is true for items that are not explored routinely!

Conclusion

- This tutorial showed how one can explore random effects and formally assess the need for Random slopes
- As a rule of thumb $\Rightarrow$ Any within-subject (or within-item) should be tested for a potential inclusion as a random slope
- Fixed effects provides averages over all observations, even when using mixed effects regressions; we need to explore what random effects (intercepts and slopes) tell us.
- In this example, we see that many subjects vary beyond the fixed effect; Standard Errors are not enough to quantify this type of variation. The same is true for items that are not explored routinely!

I hope this tutorial helped you to uncover the role of participants and items and what they can tell us beyond the fixed effect!

Questions?