Table of Contents

[All the code for easy cut and paste 1](#_Toc102569984)

[Examine model differences for nested and fixed effects 2](#_Toc102569985)

[When you add the effect of match advantage to the model - overall 3](#_Toc102569986)

[When you add the effect of match advantage to the model – coefficients 3](#_Toc102569987)

[If you add a random slope of match advantage to see differences in language – overall 4](#_Toc102569988)

[If you add a random slope of match advantage to see differences in language – coefficients 4](#_Toc102569989)

[If you make a graph of the raw differences 5](#_Toc102569990)

[If you graph the differences after controlling for other random effects 5](#_Toc102569991)

# All the code for easy cut and paste

```{r SP\_lme\_update, message = FALSE, warning = FALSE}

overall\_excluded\_lang <- SP\_V\_lme\_data %>%

group\_by(Language, Subject) %>%

summarise(N\_trials = n()) %>%

group\_by(Language) %>%

summarise(N = n()) %>%

filter(N < 25) %>%

pull(Language) ## Exclude the languages less than 25 participants

# just in case but doesn't appear to be excluding anyone

SP\_V\_lme\_data\_excluded <- subset(SP\_V\_lme\_data, !(Language %in% overall\_excluded\_lang))

intercept.model <- lm(response\_time ~ 1,

data = SP\_V\_lme\_data\_excluded)

subject.model <- lmer(response\_time ~ 1 + (1|Subject),

control = lmerControl(optimizer = "bobyqa",

optCtrl = list(maxfun = 1e6)),

data = SP\_V\_lme\_data\_excluded)

item.model <- lmer(response\_time ~ 1 + (1|Subject) + (1|Target),

control = lmerControl(optimizer = "bobyqa",

optCtrl = list(maxfun = 1e6)),

data = SP\_V\_lme\_data\_excluded)

language.model <- lmer(response\_time ~ 1 + (1|Subject) + (1|Target) + (1|Language),

control = lmerControl(optimizer = "bobyqa",

optCtrl = list(maxfun = 1e6)),

data = SP\_V\_lme\_data\_excluded)

# which is best

AIC(intercept.model)

AIC(subject.model)

AIC(item.model)

AIC(language.model)

AIC(subject.model) < AIC(intercept.model)

AIC(item.model) < AIC(subject.model)

AIC(language.model) < AIC(item.model)

fixed.model <- lmer(response\_time ~ Match + (1|Subject) + (1|Target) + (1|Language),

control = lmerControl(optimizer = "bobyqa",

optCtrl = list(maxfun = 1e6)),

data = SP\_V\_lme\_data\_excluded)

AIC(fixed.model)

AIC(fixed.model) < AIC(language.model)

summary(fixed.model)

fixed.randomslope.model <- lmer(response\_time ~ Match + (1|Subject) + (1|Target) + (1 + Match|Language),

control = lmerControl(optimizer = "bobyqa",

optCtrl = list(maxfun = 1e6)),

data = SP\_V\_lme\_data\_excluded)

AIC(fixed.randomslope.model)

AIC(fixed.randomslope.model) < AIC(fixed.model)

summary(fixed.randomslope.model)

# plot of the data for just raw score differences

ggplot(SP\_V\_lme\_data\_excluded, aes(Language, response\_time, color = Match)) +

theme\_classic() +

ylab("Response Latencies") +

xlab("Language") +

stat\_summary(fun = mean,

geom = "point") +

stat\_summary(fun.data = mean\_cl\_normal,

geom = "pointrange") +

theme(axis.text.x = element\_text(angle = 90))

# plot of data controlling for other factors

coef\_model <- coef(fixed.randomslope.model)$Language

coef\_model$MATCHING <- coef\_model$`(Intercept)`

coef\_model$MISMATCHING <- coef\_model$MATCHING + coef\_model$MatchMISMATCHING

coef\_model$Language <- rownames(coef\_model)

library(parameters)

se\_model <- as.data.frame(standard\_error(fixed.randomslope.model, effects = "random")$Language)

coef\_model$se <- se\_model$`(Intercept)`

coef\_data <- coef\_model %>%

pivot\_longer(cols = c("MATCHING", "MISMATCHING")) %>%

mutate(lower = value - 2\*se,

upper = value + 2\*se) %>%

rename(Match = name)

ggplot(coef\_data, aes(Language, value, color = Match)) +

theme\_classic() +

ylab("Response Latencies") +

xlab("Language") +

geom\_point() +

geom\_pointrange(data = coef\_data, aes(ymin = lower, ymax = upper)) +

theme(axis.text.x = element\_text(angle = 90))

```

# Examine model differences for nested and fixed effects

> AIC(intercept.model)

[1] 865581.7

> AIC(subject.model)

[1] 834003.2

> AIC(item.model)

[1] 832167.8

> AIC(language.model)

[1] 832044.8

> AIC(subject.model) < AIC(intercept.model)

[1] TRUE

> AIC(item.model) < AIC(subject.model)

[1] TRUE

> AIC(language.model) < AIC(item.model)

[1] TRUE

This section suggests that the addition of the subject variable improves model fit over the intercept only model. Then the addition of the item improves the model over subjects only. Last, the addition of a random effect of language intercept improves the model more than subject intercepts and target item intercepts.

# When you add the effect of match advantage to the model - overall

> AIC(fixed.model)

[1] 832044

> AIC(fixed.model) < AIC(language.model)

[1] TRUE

The fixed effects model is “better” than the random effects only model BUT please notice this difference is .8, which many would agree makes them approximately equivalent.

# When you add the effect of match advantage to the model – coefficients

> summary(fixed.model)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [

lmerModLmerTest]

Formula:

response\_time ~ Match + (1 | Subject) + (1 | Target) + (1 | Language)

Data: SP\_V\_lme\_data\_excluded

Control:

lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1000000))

REML criterion at convergence: 832032

Scaled residuals:

Min 1Q Median 3Q Max

-5.6380 -0.5659 -0.1003 0.4575 11.1471

Random effects:

Groups Name Variance Std.Dev.

Subject (Intercept) 26850.5 163.86

Target (Intercept) 964.6 31.06

Language (Intercept) 1680.5 40.99

Residual 28442.0 168.65

Number of obs: 62788, groups: Subject, 3317; Target, 48; Language, 18

Fixed effects:

Estimate Std. Error df t value Pr(>|t|)

(Intercept) 678.1793 11.3900 24.1523 59.542 <2e-16 \*\*\*

MatchMISMATCHING 0.8519 1.3545 59496.0349 0.629 0.529

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr)

MMISMATCHIN -0.060

Examining this model, we can see that there’s no effect of match … however, there is a lot of variance by language.

# If you add a random slope of match advantage to see differences in language – overall

> AIC(fixed.randomslope.model)

[1] 832048

> AIC(fixed.randomslope.model) < AIC(fixed.model)

[1] FALSE

This model is not better than a model without the slope. That implies that after controlling for all the other effects, basically all languages see the same small differences in match versus mismatch.

# If you add a random slope of match advantage to see differences in language – coefficients

> summary(fixed.randomslope.model)

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: response\_time ~ Match + (1 | Subject) + (1 | Target) + (1 + Match | Language)

Data: SP\_V\_lme\_data\_excluded

Control: lmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1000000))

REML criterion at convergence: 832032

Scaled residuals:

Min 1Q Median 3Q Max

-5.6381 -0.5660 -0.1003 0.4575 11.1470

Random effects:

Groups Name Variance Std.Dev. Corr

Subject (Intercept) 26850.475328 163.86115

Target (Intercept) 964.596023 31.05795

Language (Intercept) 1678.306720 40.96714

MatchMISMATCHING 0.003836 0.06194 1.00

Residual 28441.982144 168.64751

Number of obs: 62788, groups: Subject, 3317; Target, 48; Language, 18

Fixed effects:

Estimate Std. Error df t value Pr(>|t|)

(Intercept) 678.1646 11.3843 23.9917 59.57 <2e-16 \*\*\*

MatchMISMATCHING 0.8799 1.3545 28845.5610 0.65 0.516

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr)

MMISMATCHIN -0.050

optimizer (bobyqa) convergence code: 0 (OK)

boundary (singular) fit: see help('isSingular')

Same note as above – no differences in match after controlling for everything else.

# If you make a graph of the raw differences

Chart, scatter chart

Description automatically generated

Small differences but big variance in the overall average for the effects.

# If you graph the differences after controlling for other random effects

Chart, box and whisker chart

Description automatically generated

Same basic results – no differences between match and mismatch when you control for other sources of variance … but differences in intercept. I would probably suggest this picture with a variable that indicates sample size with the size of the dot.