

# Week 2 Homework

*Molly Farry-Thorn*

*September 7, 2017*

1. Run linear models on all of your subjects (a basic regression). What is the average intercept, the average slope?

```
model.1p <- lm(Utterances.with.Letters ~ Session, data=parents)
summary(model.1p)

##
## Call:
## lm(formula = Utterances.with.Letters ~ Session, data = parents)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.510  -7.849  -4.817  -1.329  197.490
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.5854     1.5239   1.697 0.090257 .
## Session       0.7437     0.2077   3.580 0.000369 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.22 on 646 degrees of freedom
## Multiple R-squared:  0.01945,    Adjusted R-squared:  0.01794
## F-statistic: 12.82 on 1 and 646 DF,  p-value: 0.0003694

(summary(model.1p)$sigma)**2

## [1] 332.1347

model.1c <- lm(Utterances.with.Letters ~ Session, data=children)
summary(model.1c)

##
## Call:
## lm(formula = Utterances.with.Letters ~ Session, data = children)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.634  -6.752  -3.343  -0.933  165.129
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.05067     1.33293   0.038   0.97
## Session       0.88198     0.18118   4.868 1.42e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.96 on 650 degrees of freedom
```

```
## Multiple R-squared:  0.03518,    Adjusted R-squared:  0.03369
## F-statistic: 23.7 on 1 and 650 DF,  p-value: 1.417e-06
```

```
(summary(model.1c)$sigma)**2
```

```
## [1] 254.5669
```

### For Parents:

Average intercept = 2.29 utterances with letters at Session 0 (around child age 10 months)

Average slope = .74 utterances with letters every 4 months

Residual variance = 332.2693

### For Children:

Average intercept = 0.05 utterances with letters at Session 0 (around child age 10 months)

Average slope = .88 utterances with letters every 4 months

Residual variance = 254.5669

## 2. Now run a mlm/lmer model with only a random intercept. What is the ICC?

```
model.2p <- lmer(Utterances.with.Letters ~ 1 + (1|Subject), data = parents)
summary(model.2p)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Utterances.with.Letters ~ 1 + (1 | Subject)
## Data: parents
##
## REML criterion at convergence: 5575.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.1017 -0.3594 -0.2152 -0.0640 10.0856
##
## Random effects:
##  Groups   Name      Variance Std.Dev.
## Subject (Intercept)  44.5      6.671
## Residual              294.1     17.149
## Number of obs: 648, groups: Subject, 55
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    7.368      1.124    6.555
```

```
44.5/(44.5+294.1)
```

```
## [1] 0.1314235
```

```
model.2c <- lmer(Utterances.with.Letters ~ 1 + (1|Subject), data = children)
summary(model.2c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Utterances.with.Letters ~ 1 + (1 | Subject)
## Data: children
##
```

```
## REML criterion at convergence: 5465.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.1005 -0.3171 -0.2153 -0.0986  10.3950
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   Subject (Intercept) 22.43    4.736
##   Residual             241.36   15.536
## Number of obs: 652, groups: Subject, 55
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   5.7741    0.8822    6.545
22.43/(22.43+241.36)
## [1] 0.08502976
```

### For Parents:

Average intercept = 7.37 utterances with letters

Slope = we set it at 0

Residual variance = 294.1

ICC = 0.1314235

### For Children:

Average intercept = 0.05 utterances with letters

Slope = we set it at 0

Residual variance = 241.36

ICC = 0.08502976

**What does residual variance look like compared to linear model? Create a graph to show this effect.**

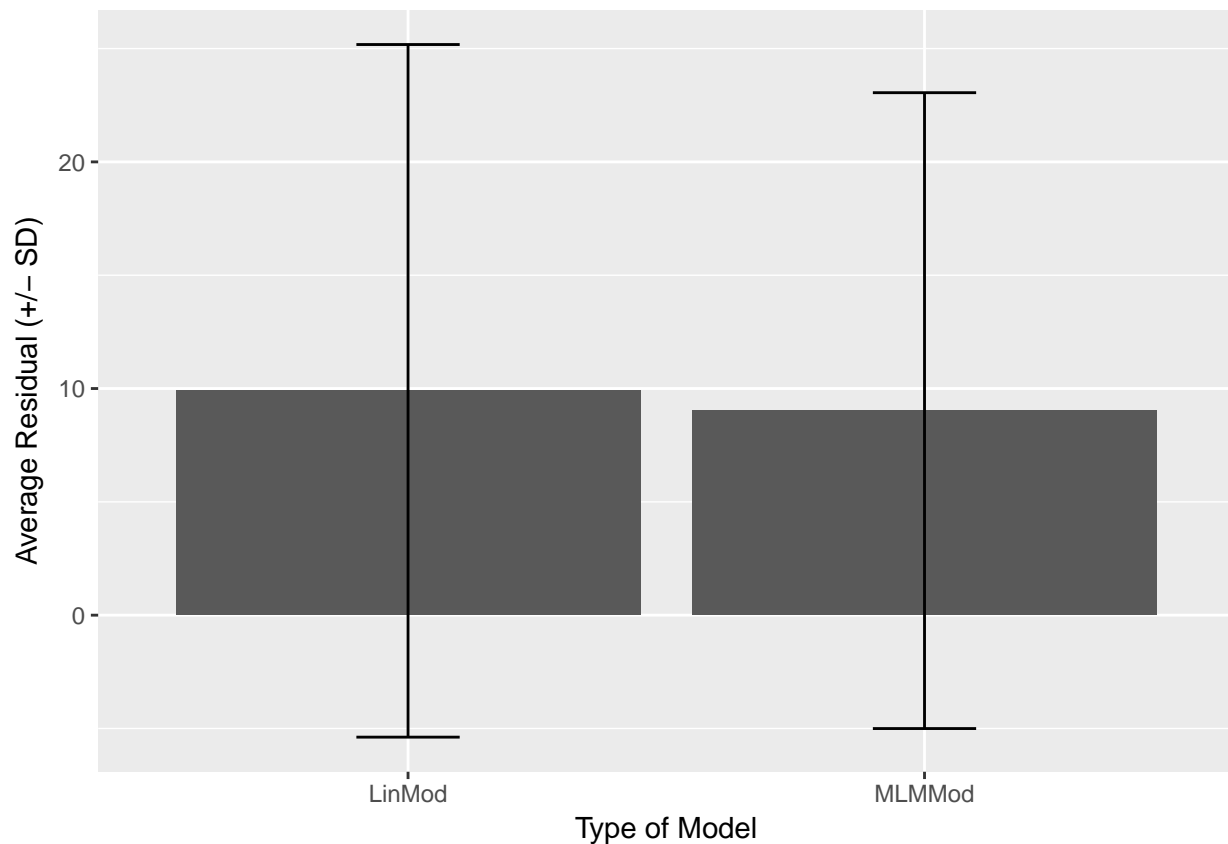
```
#plotting residuals
library(broom)
lin.resid <- abs(augment(model.1p)[,c(5)])
mean.lin.resid <- mean(lin.resid)
sd.lin.resid <- sd(lin.resid)

mlm.resid <- abs(augment(model.2p)[,c(4)])
mean.mlm.resid <- mean(mlm.resid)
sd.mlm.resid <- sd(mlm.resid)

dataframe <- data.frame("Type" = c("LinMod", "MLMod"),
                        "Mean" = c(mean.lin.resid, mean.mlm.resid), "SD" = c(sd.lin.resid, sd.mlm.resid))

ResidPlot <- ggplot(dataframe, aes(Type, Mean)) +
```

```
geom_col() +
  geom_errorbar(aes(ymin = Mean - SD, ymax = Mean + SD), width=0.2)
ResidPlot + labs(y="Average Residual (+/- SD)", x = "Type of Model")
```



Residual variance (the variance not accounted for by the model) is smaller in the in the mixed-effect model compared to the linear model, even though the linear model uses age as a predictor whereas the MLM model does not.

**3. Introduce a fixed slope term. What is the difference in terms of the fixed effects estimates between this estimate and the previous? Of the residual standard error?**

```
model.3p <- lmer(Utterances.with.Letters ~ 1 + Session + (1|Subject), data = parents)
summary(model.3p)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Utterances.with.Letters ~ 1 + Session + (1 | Subject)
## Data: parents
##
## REML criterion at convergence: 5562.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.2139 -0.4125 -0.1893  0.0343  9.9391
##
```

```
## Random effects:
##   Groups   Name              Variance Std.Dev.
##   Subject (Intercept)  45.08      6.714
##   Residual                287.38    16.952
## Number of obs: 648, groups:  Subject, 55
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)   2.5448     1.6824    1.513
## Session       0.7449     0.1933    3.852
##
## Correlation of Fixed Effects:
##           (Intr)
## Session -0.744
```

```
model.3c <- lmer(Utterances.with.Letters ~ 1 + Session + (1|Subject), data = children)
summary(model.3c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Utterances.with.Letters ~ 1 + Session + (1 | Subject)
##   Data: children
##
## REML criterion at convergence: 5441.5
##
## Scaled residuals:
##   Min       1Q   Median       3Q      Max
## -1.1762 -0.4023 -0.1680  0.0331 10.3798
##
## Random effects:
##   Groups   Name              Variance Std.Dev.
##   Subject (Intercept)  23.47      4.844
##   Residual                231.43    15.213
## Number of obs: 652, groups:  Subject, 55
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  0.008465    1.429321    0.006
## Session      0.887184    0.172811    5.134
##
## Correlation of Fixed Effects:
##           (Intr)
## Session -0.786
```

### For Parents:

Average intercept = 2.55 utterances with letters at Session 0 (around child age 10 months) (*7.37 in model without age*)

Average slope = .75 utterances with letters every 4 months (*.74 in linear model*)

Residual variance = 287.38 (*294.1 in model without age*)

### For Children:

Average intercept = 0.001 utterances with letters at Session 0 (around child age 10 months) (*0.05 in model without age*)

Average slope = .89 utterances with letters every 4 months (*.88 in linear model*)

Residual variance = 231.43 (*241.36 in model without age*)

The fixed effect estimates of the intercept have decreased. The estimates of the slopes are about equal to those of the linear model. The residual error has decreased.

Create a graph to show both fixed effects estimates and the CIs around them.

```
##FOR PARENTS
# Get fixed effect confidence intervals
confint(model.3p, level = .95)

## Computing profile confidence intervals ...

##           2.5 %    97.5 %
## .sig01      4.8464579  8.787091
## .sigma     16.0180269 17.948930
## (Intercept) -0.7510901  5.840187
## Session      0.3656021  1.124076

CIs <- data.frame(confint(model.3p, level = .95)[3:4,])

## Computing profile confidence intervals ...

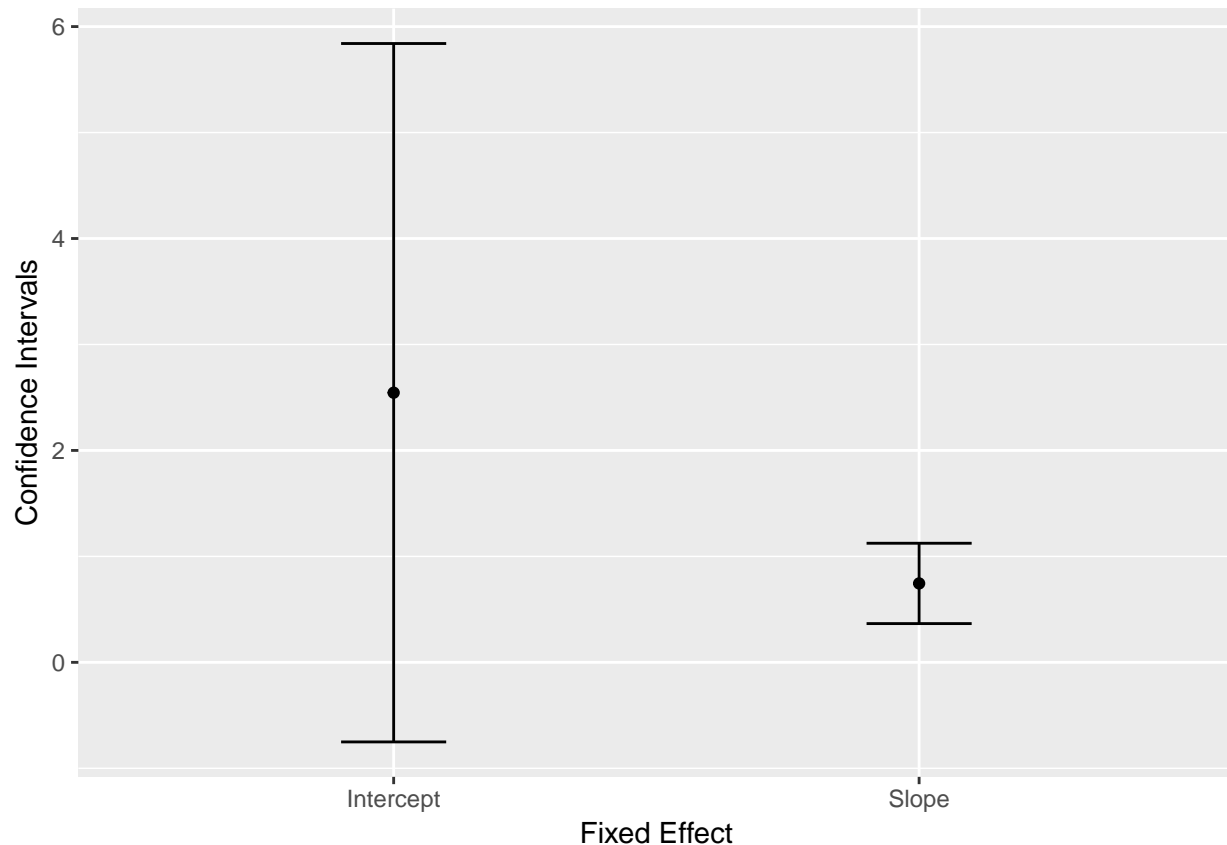
CIs$fixef <- fixef(model.3p)

summary1 <- tidy(model.3p)
summary1

##           term      estimate std.error statistic    group
## 1      (Intercept)  2.5448102  1.6824034   1.512604   fixed
## 2         Session   0.7448542  0.1933443   3.852475   fixed
## 3 sd_(Intercept).Subject  6.7143301      NA      NA Subject
## 4 sd_Observation.Residual 16.9521541      NA      NA Residual

# graph fixed effect CIs
dataframe <- data.frame("Fixed Effect" = c("Intercept", "Slope"),
                        "Mean" = c(CIs[1,3], CIs[2,3]), "LowerCI" = c(CIs[1,1], CIs[2,1]), "UpperCI" = c(CIs[1,2], CIs[2,2]))

CIPlot <- ggplot(dataframe, aes(Fixed.Effect, Mean)) +
  geom_point() +
  geom_errorbar(aes(ymin = LowerCI, ymax = UpperCI), width=0.2)
CIPlot + labs(y="Confidence Intervals", x = "Fixed Effect")
```



```
##FOR CHILDREN
```

```
# Get fixed effect confidence intervals
```

```
confint(model.3c, level = .95)
```

```
## Computing profile confidence intervals ...
```

```
##           2.5 %    97.5 %
```

```
## .sig01      3.1555429  6.592970
```

```
## .sigma      14.3772293 16.104843
```

```
## (Intercept) -2.7908665  2.807329
```

```
## Session      0.5481127  1.226046
```

```
CIs <- data.frame(confint(model.3c, level = .95)[3:4,])
```

```
##Computing profile confidence intervals ...
```

```
CIs$fixef <- fixef(model.3c)
```

```
# graph fixed effect CIs
```

```
dataframe <- data.frame("Fixed Effect" = c("Intercept", "Slope"),
```

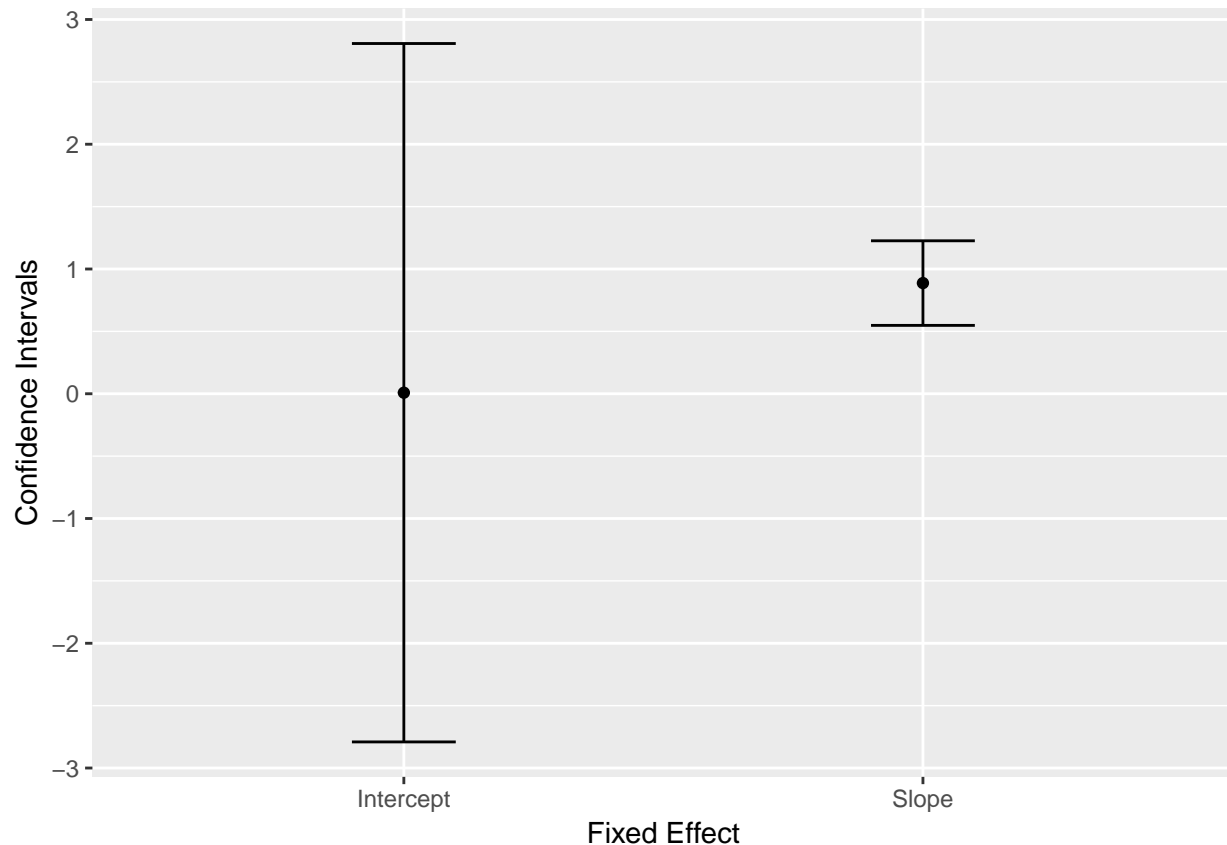
```
                        "Mean" = c(CIs[1,3], CIs[2,3]), "LowerCI" = c(CIs[1,1], CIs[2,1]), "UpperCI" = c(CIs[1,2], CIs[2,2]))
```

```
CIPlot <- ggplot(dataframe, aes(Fixed.Effect, Mean)) +
```

```
  geom_point() +
```

```
  geom_errorbar(aes(ymin = LowerCI, ymax = UpperCI), width=0.2)
```

```
CIPlot + labs(y="Confidence Intervals", x = "Fixed Effect")
```



4. Run an additional model with a random slope. How does this change compare to the previous model? Should you keep the random slope or not?

```
model.4p <- lmer(Utterances.with.Letters ~ 1 + Session + (1 + Session|Subject), data = parents)
summary(model.4p)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Utterances.with.Letters ~ 1 + Session + (1 + Session | Subject)
## Data: parents
##
## REML criterion at convergence: 5487.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7444 -0.3140 -0.2232 -0.0046  7.1589
##
## Random effects:
##  Groups   Name                Variance Std.Dev. Corr
##  Subject  (Intercept)    19.041     4.364
##           Session         3.163     1.778  -1.00
## Residual                    243.260   15.597
## Number of obs: 648, groups: Subject, 55
##
## Fixed effects:
```



```
##               Estimate Std. Error t value
## (Intercept)    2.5858     1.4313    1.807
## Session        0.7398     0.2987    2.477
##
## Correlation of Fixed Effects:
##      (Intr)
## Session -0.810

model.4c <- lmer(Utterances.with.Letters ~ 1 + Session + (1 + Session|Subject), data = children)
summary(model.4c)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Utterances.with.Letters ~ 1 + Session + (1 + Session | Subject)
## Data: children
##
## REML criterion at convergence: 5397.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6942 -0.2634 -0.1600 -0.0619  9.9342
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## Subject (Intercept) 9.170 3.028
## Session 1.682 1.297 -1.00
## Residual 206.758 14.379
## Number of obs: 652, groups: Subject, 55
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept) 0.001973 1.269116 0.002
## Session 0.892108 0.239371 3.727
##
## Correlation of Fixed Effects:
##      (Intr)
## Session -0.806
```

### For Parents:

Average intercept = 2.55 utterances with letters at Session 0 (around child age 10 months) (*2.55 in model without random slope*)

Average slope = .74 utterances with letters every 4 months (*.75 in model without random slope*)

Residual variance = 243.260 (*287.38 in model without random slope*)

### For Children:

Average intercept = 0.002 utterances with letters at Session 0 (around child age 10 months) (*0.001 in model without random slope*)

Average slope = .89 utterances with letters every 4 months (*.89 in model without random slope*)

Residual variance = 206.76 (*231.43 in model without random slope*)

The residual error has decreased in both models, suggesting I should keep the random slope.

## 5. Interpret the correlation between the slope and the intercept.

The fixed effects correlation is -0.81. The lower the intercept the greater the slope. There is a very strong relationship between the average intercept and the average slope.

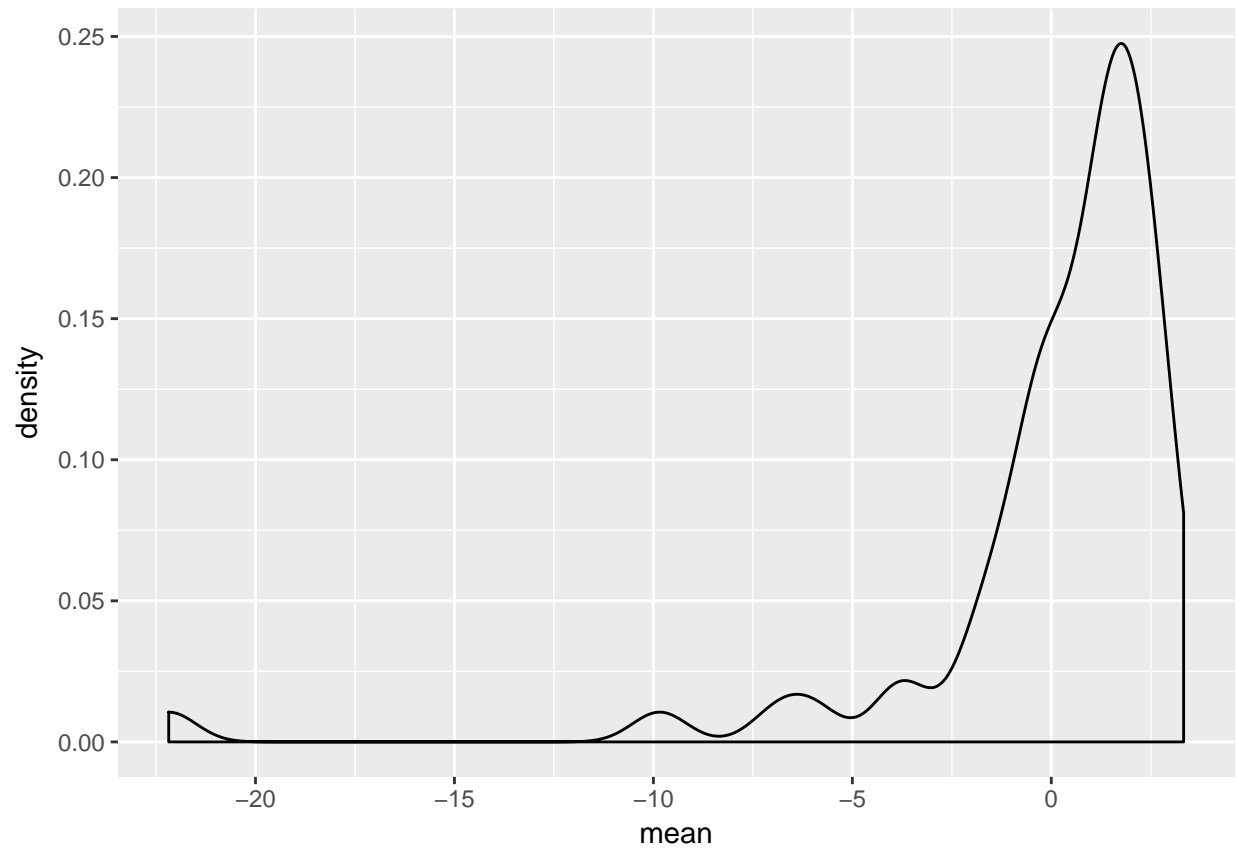
## 6. Create a density plot of the random effects from your final model.

```
library(merTools)

## Loading required package: arm
## Loading required package: MASS
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##   select
##
## arm (Version 1.9-3, built: 2016-11-21)
## Working directory is C:/Users/Molly/Documents/1-descriptives-and-graphs-mfarrythorn
##
## Attaching package: 'arm'
## The following objects are masked from 'package:psych':
##
##   logit, rescale, sim
##
## Attaching package: 'merTools'
## The following object is masked from 'package:psych':
##
##   ICC
## FOR PARENTS
re.sim.p <- REsim(model.4p)

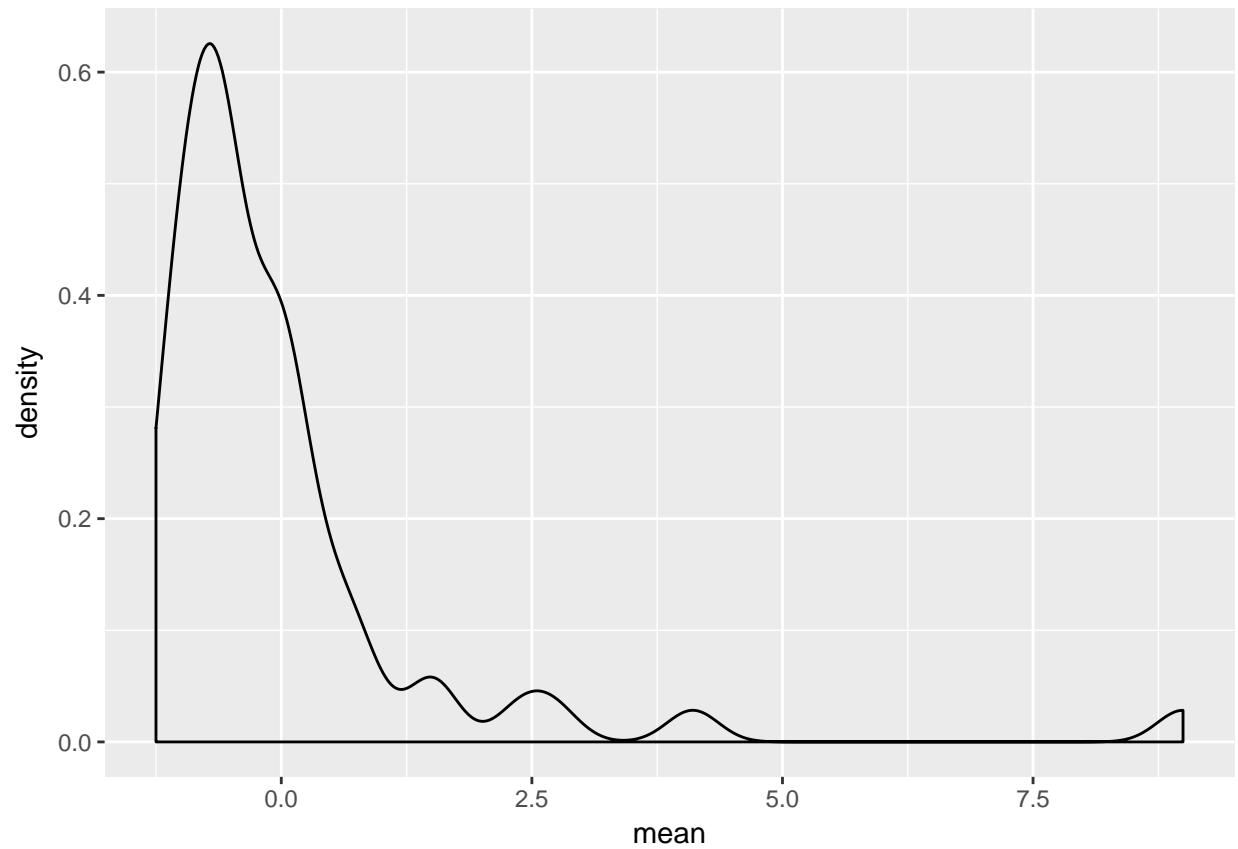
# Intercept random effects
p1.gg1 <- re.sim.p %>%
  filter(term == "(Intercept)")

ggplot(p1.gg1, aes(mean)) +
  geom_density()
```



```
# Slope random effects
p1.gg2 <- re.sim.p %>%
  filter(term == "Session")

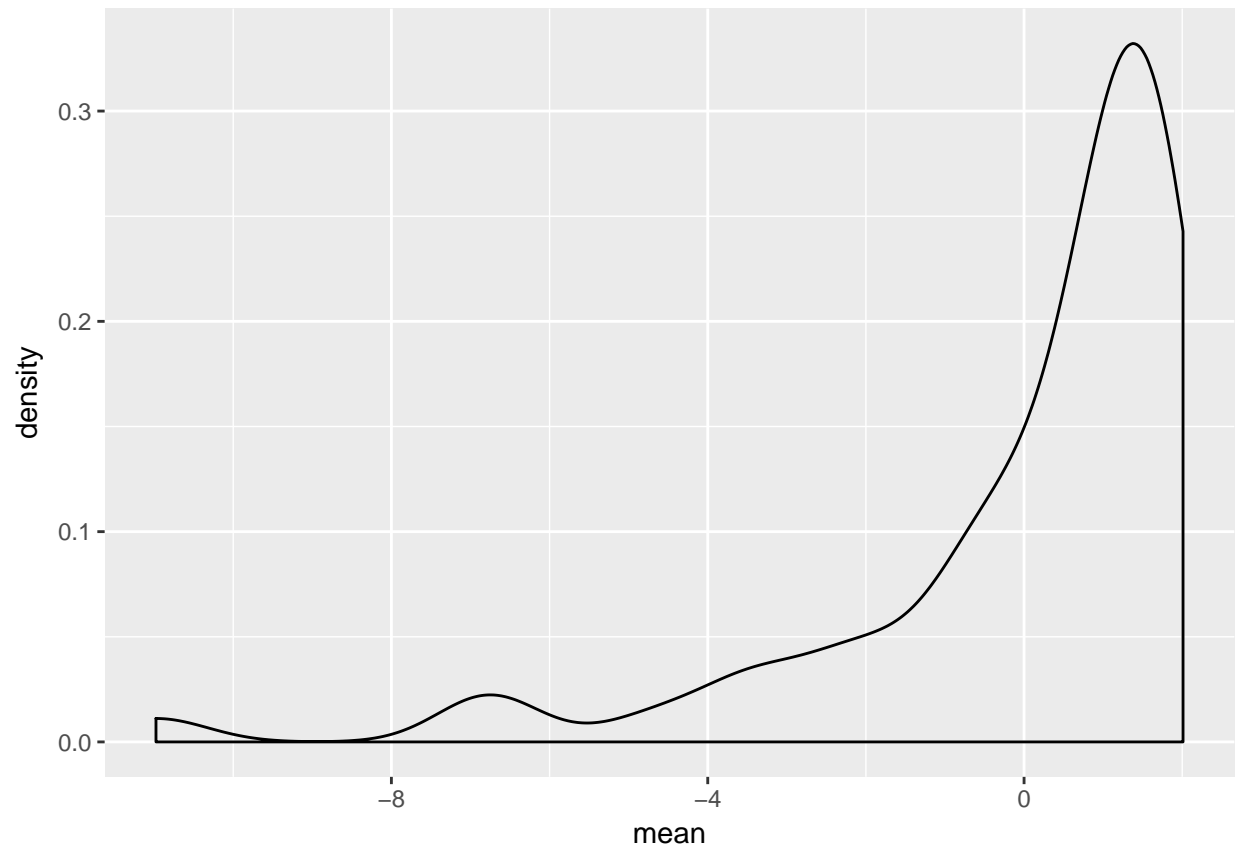
ggplot(p1.gg2, aes(mean)) +
  geom_density()
```



```
## FOR CHILDREN
re.sim.c <- REsim(model.4c)

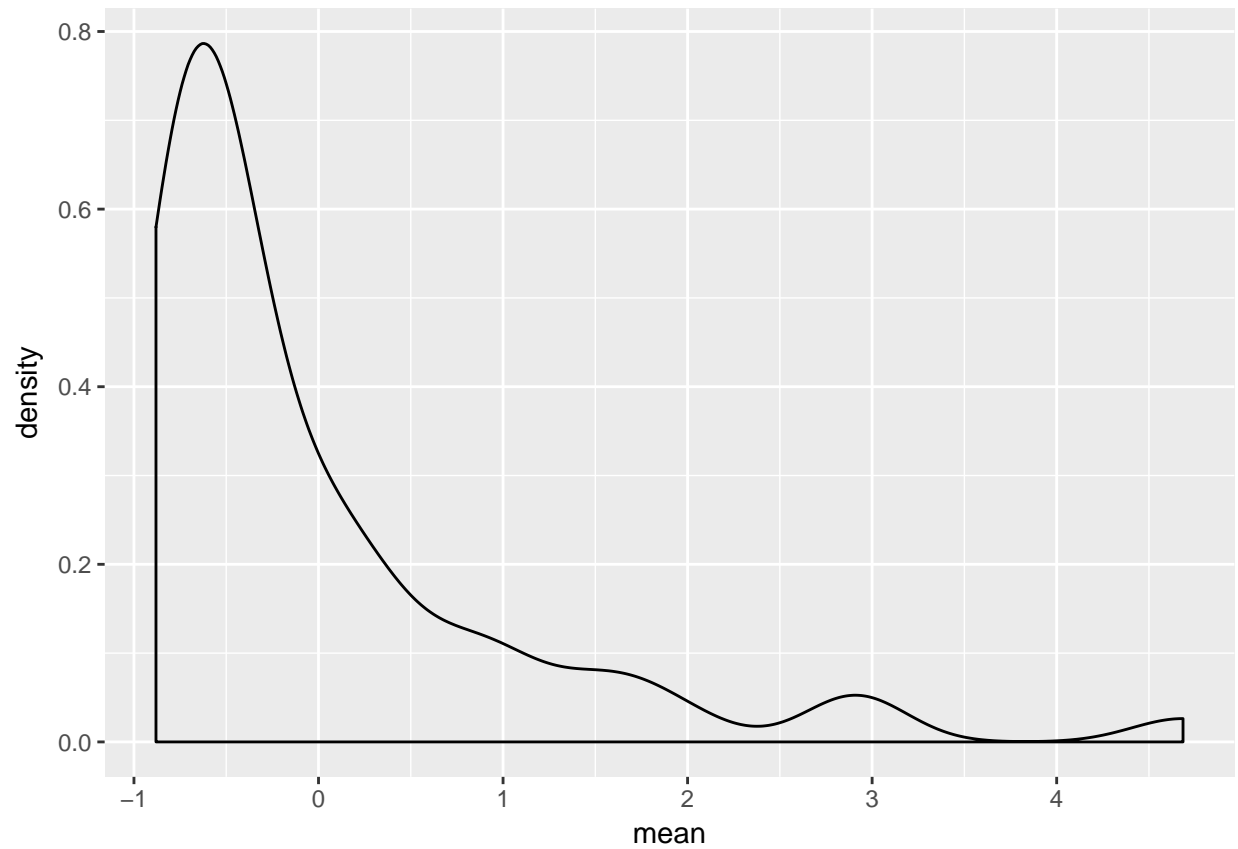
# Intercept random effects
p1.gg3 <- re.sim.c %>%
  filter(term == "(Intercept)")

ggplot(p1.gg3, aes(mean)) +
  geom_density()
```



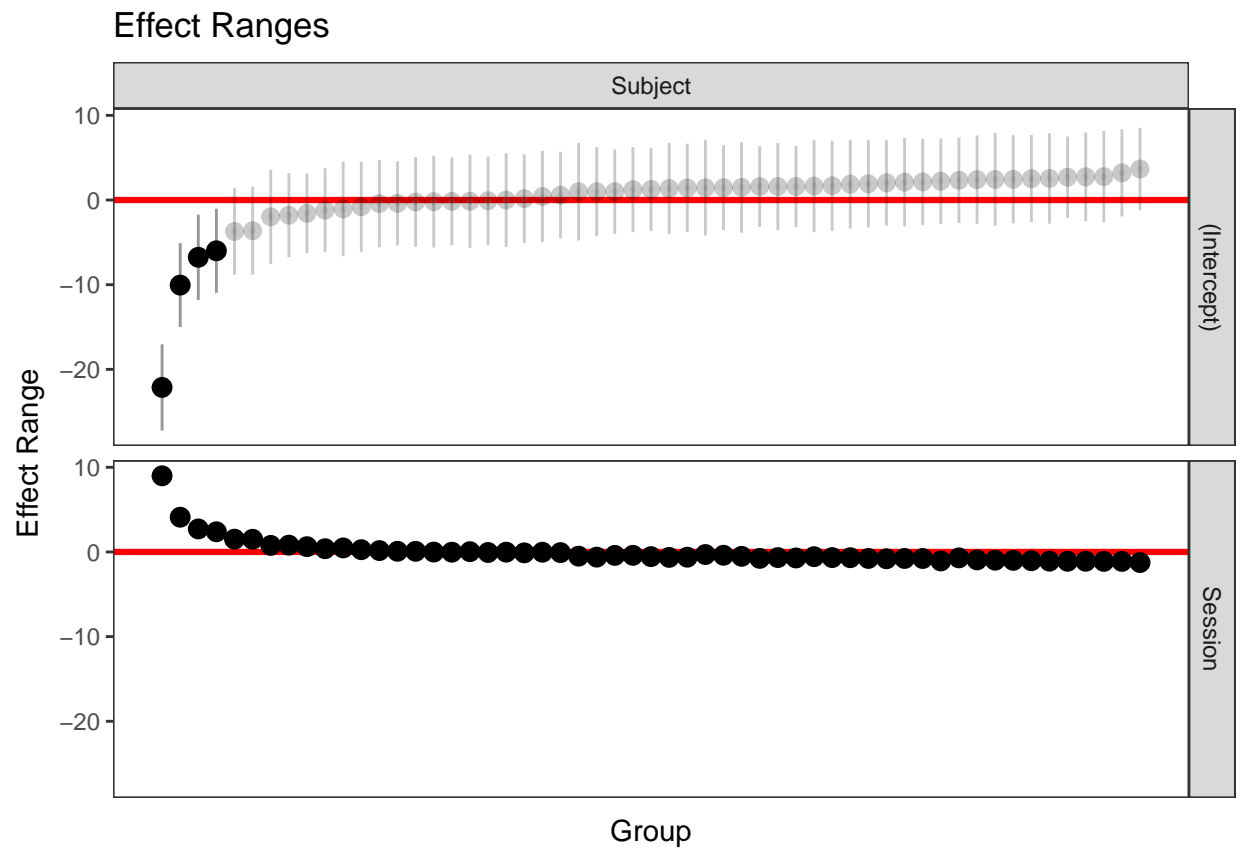
```
# Slope random effects
p1.gg4 <- re.sim.c %>%
  filter(term == "Session")

ggplot(p1.gg4, aes(mean)) +
  geom_density()
```



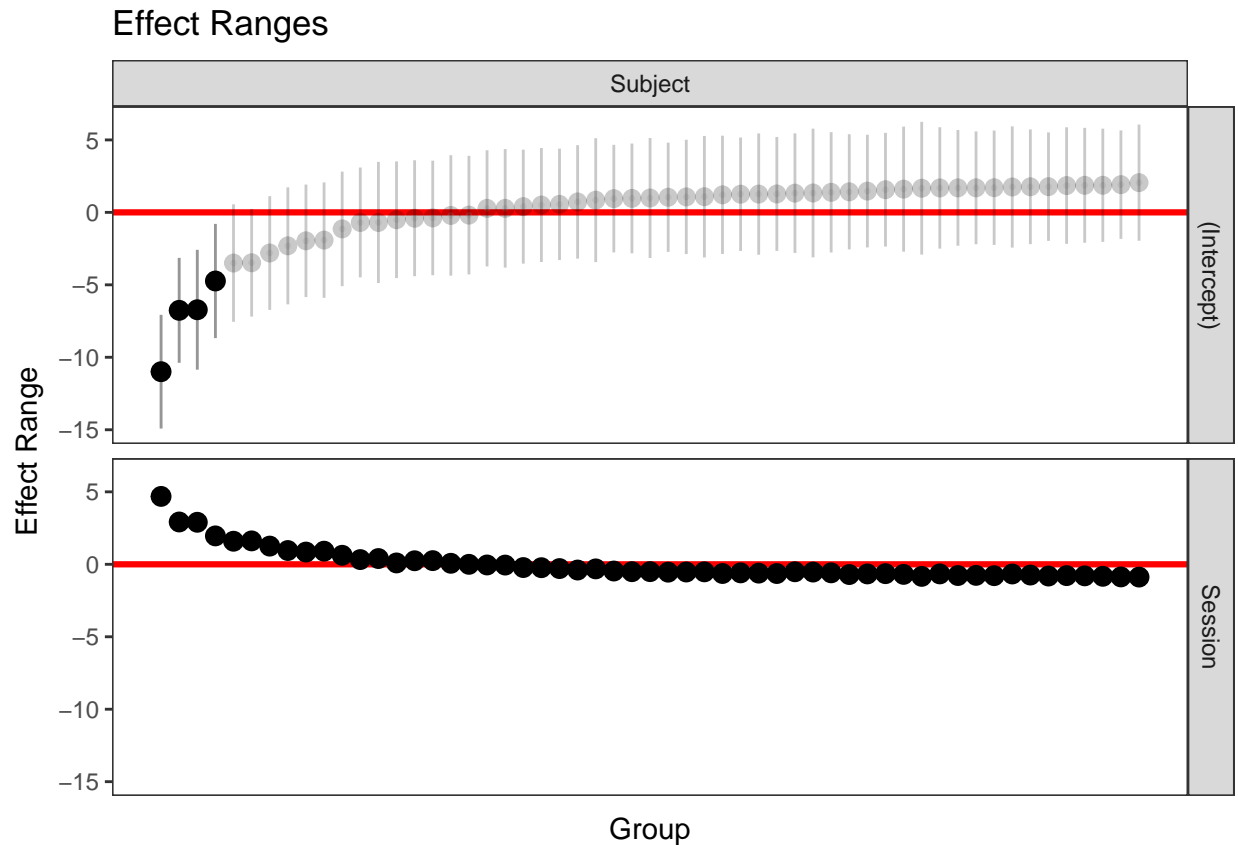
7. Create a caterpillar plot of the random effects. Is there any person that seems odd in terms of a large standard errors around intercept and slope estimates?

```
p1 <- plotREsim(re.sim.p)
p1
```



# 75, 27

```
p2 <- plotREsim(re.sim.c)
p2
```



#75, 25, 89

The listed participant numbers above have effect ranges for the intercept differ significantly from 0. However, by and large, standard errors around the intercept and slope values all seem to be within normal limits.

8. Create a plot of the trajectory, along with a spaghetti plot of each person's individual slope. Set the alpha level (transparency) on the individual slopes to make them easier to see.

```
coefs <- data.frame(coef(model.4p)[[1]])

ggplot(data = parents, aes(Session, Utterances.with.Letters)) +
  stat_smooth(aes(Session, Utterances.with.Letters), method = lm, se = F) +
  xlim(0,12) + ylim (-20,20) +
  geom_abline(data = coefs, aes(slope = Session, intercept = X.Intercept.), alpha = 0.2)
```

```
## Warning: Removed 69 rows containing non-finite values (stat_smooth).
```



