

ALDA_HW3_EH

Elizabeth Hawkey

9/24/2017

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
```

```
## Loading tidyverse: tibble
```

```
## Loading tidyverse: tidyr
```

```
## Loading tidyverse: readr
```

```
## Loading tidyverse: purrr
```

```
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages -----
```

```
## expand(): tidyr, Matrix
```

```
## filter(): dplyr, stats
```

```
## lag(): dplyr, stats
```

```
library(broom)
```

```
library(tidyr)
```

```
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 /Users/elizabethhawkey/ejhawkey
```

```
PDS_data_2 <- read.csv(file = "~/Documents/PDS_project/Longitudinal_project/datasets/PDS_long_final.csv")
```

```
PDS_data_2$AGEMOSCAN <- as.numeric(PDS_data_2$AGEMOSCAN)
```

```
PDS_data_2$ADHDsum <- as.numeric(PDS_data_2$ADHDsum)
```

```
PDS_data_2$ADHD_INsum <- as.numeric(PDS_data_2$ADHD_INsum)
```

```
PDS_data_2$ADHD_HYIMsum <- as.numeric(PDS_data_2$ADHD_HYIMsum)
```

1. Run a series of models using a time-invariant nominal covariate.
a) where the covariate only predicts the intercept b) predicts both intercept and slope c) is rescaled eg centering. For all models, how does your model change from model to model. What is your final model?

```
#nominal covariate = sex
#a) where the covariate only predicts the intercept

mod.1 <- lmer(ADHDsum ~ agemo_converted + sex + (1|Subid_fMRI), data=PDS_data_2)
summary(mod.1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: ADHDsum ~ agemo_converted + sex + (1 | Subid_fMRI)
## Data: PDS_data_2
##
## REML criterion at convergence: 2732
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.4032 -0.4035 -0.1066  0.2421  4.6492
##
## Random effects:
## Groups Name Variance Std.Dev.
## Subid_fMRI (Intercept) 8.908 2.985
## Residual 3.959 1.990
## Number of obs: 554, groups: Subid_fMRI, 209
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      8.5212     1.0937  7.791
## agemo_converted -3.3747     0.5892 -5.727
## sex              -1.1163     0.4490 -2.486
##
## Correlation of Fixed Effects:
##              (Intr) agm_cn
## agem_cnvrted -0.766
## sex          -0.640  0.040

#b) predicts both intercept and slope
mod.2 <- lmer(ADHDsum ~ agemo_converted + sex + (agemo_converted|Subid_fMRI), data=PDS_data_2)
summary(mod.2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: ADHDsum ~ agemo_converted + sex + (agemo_converted | Subid_fMRI)
## Data: PDS_data_2
##
## REML criterion at convergence: 2702.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5203 -0.3163 -0.1150  0.2092  3.7891
```

```

##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   Subid_fMRI (Intercept)          65.158   8.072
##               agemo_converted 14.427   3.798   -0.98
##   Residual                        3.468   1.862
## Number of obs: 554, groups:  Subid_fMRI, 209
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      8.7777    1.1559   7.594
## agemo_converted  -3.4633    0.6146  -5.635
## sex              -1.1862    0.4201  -2.823
##
## Correlation of Fixed Effects:
##               (Intr) agm_cn
## agem_cnvrted -0.825
## sex          -0.558  0.027

```

```

# c) is rescaled eg centering
sex_c <- scale(PDS_data_2$sex, center = T, scale = F)
PDS_data_2$sex_c <- as.numeric(sex_c)
mod.3 <- lmer(ADHDsum ~ agemo_converted + sex_c + (agemo_converted|Subid_fMRI), data=PDS_data_2)
summary(mod.3)

```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ADHDsum ~ agemo_converted + sex_c + (agemo_converted | Subid_fMRI)
##   Data: PDS_data_2
##
## REML criterion at convergence: 2702.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5203 -0.3163 -0.1150  0.2092  3.7891
##
## Random effects:
##   Groups      Name                Variance Std.Dev. Corr
##   Subid_fMRI (Intercept)          65.158   8.072
##               agemo_converted 14.427   3.798   -0.98
##   Residual                        3.468   1.862
## Number of obs: 554, groups:  Subid_fMRI, 209
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      7.0189    0.9592   7.317
## agemo_converted  -3.4633    0.6146  -5.635
## sex_c            -1.1862    0.4201  -2.823
##
## Correlation of Fixed Effects:
##               (Intr) agm_cn
## agem_cnvrted -0.976
## sex_c        -0.023  0.027

```

The model fit improves (explains more of the variance) when a time variable is added, showing that sex is a more useful predictor of ADHD symptoms when the slopes are allowed to vary by subject. Centering sex does not improve the model.

2. Introduce a time-invariant continuous covariate and run models a-c from #1.

#continuous covariate = SES (Income_to_Need_Scan)
#a) where the covariate only predicts the intercept

```
mod.4 <- lmer(ADHDsum ~ Income_to_Need_Scan + (1|Subid_fmRI), data=PDS_data_2)
summary(mod.4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: ADHDsum ~ Income_to_Need_Scan + (1 | Subid_fmRI)
## Data: PDS_data_2
##
## REML criterion at convergence: 2758.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5855 -0.3508 -0.1653  0.2227  4.6363
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## Subid_fmRI (Intercept) 8.423    2.902
## Residual          4.314    2.077
## Number of obs: 554, groups: Subid_fmRI, 209
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      3.1860    0.3462   9.203
## Income_to_Need_Scan -0.5099    0.1407  -3.624
##
## Correlation of Fixed Effects:
##              (Intr)
## Incm_t_Nd_S -0.771
```

#b) predicts both intercept and slope

```
mod.5 <- lmer(ADHDsum ~ agemo_converted + Income_to_Need_Scan + (agemo_converted|Subid_fmRI), data=PDS_data_2)
summary(mod.5)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ADHDsum ~ agemo_converted + Income_to_Need_Scan + (agemo_converted |
## Subid_fmRI)
## Data: PDS_data_2
##
## REML criterion at convergence: 2709.9
##
## Scaled residuals:
```

```

##      Min      1Q  Median      3Q      Max
## -3.5957 -0.3202 -0.1196  0.2055  3.7449
##
## Random effects:
##   Groups      Name              Variance Std.Dev. Corr
##   Subid_fMRI (Intercept)      66.683   8.166
##               agemo_converted 16.112   4.014   -0.98
##   Residual                    3.476   1.864
## Number of obs: 554, groups:  Subid_fMRI, 209
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      7.0188    0.9670  7.258
## agemo_converted   -3.1322    0.6459 -4.849
## Income_to_Need_Scan -0.2420    0.1295 -1.869
##
## Correlation of Fixed Effects:
##           (Intr) agm_cn
## agem_cnvrted -0.939
## Incm_t_Nd_S -0.018 -0.254

```

#c) is rescaled eg centering

```

Income_to_Need_Scan_c <- scale(PDS_data_2$ Income_to_Need_Scan , center = T, scale = F)
PDS_data_2$ Income_to_Need_Scan_c <- as.numeric(Income_to_Need_Scan_c)
mod.6 <- lmer(ADHDsum ~ agemo_converted + Income_to_Need_Scan_c + (agemo_converted|Subid_fMRI), data=PDS_data_2)
summary(mod.6)

```

```

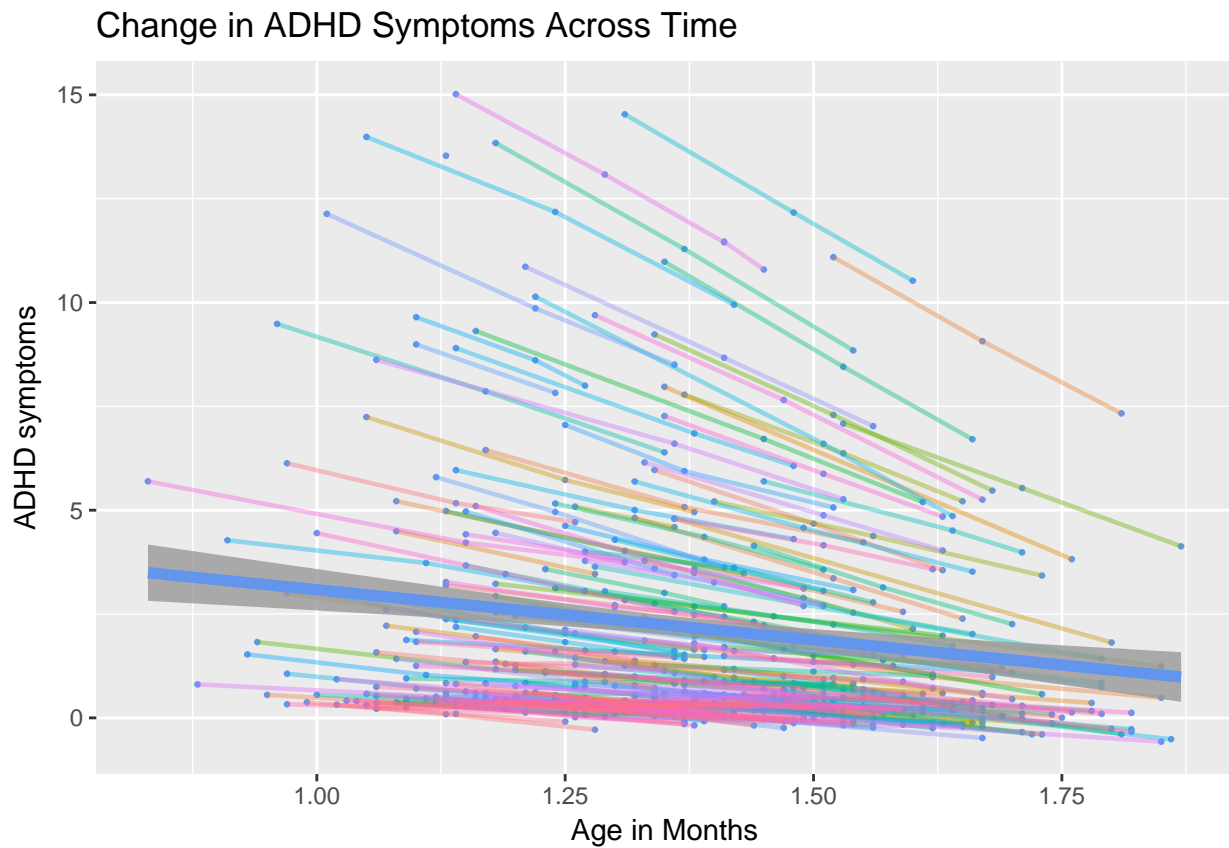
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ADHDsum ~ agemo_converted + Income_to_Need_Scan_c + (agemo_converted |
##   Subid_fMRI)
##   Data: PDS_data_2
##
## REML criterion at convergence: 2709.9
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -3.5957 -0.3202 -0.1196  0.2055  3.7449
##
## Random effects:
##   Groups      Name              Variance Std.Dev. Corr
##   Subid_fMRI (Intercept)      66.683   8.166
##               agemo_converted 16.112   4.014   -0.98
##   Residual                    3.476   1.864
## Number of obs: 554, groups:  Subid_fMRI, 209
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)      6.5596    0.9934  6.603
## agemo_converted   -3.1322    0.6459 -4.849
## Income_to_Need_Scan_c -0.2420    0.1295 -1.869
##
## Correlation of Fixed Effects:
##           (Intr) agm_cn
## agem_cnvrted -0.977

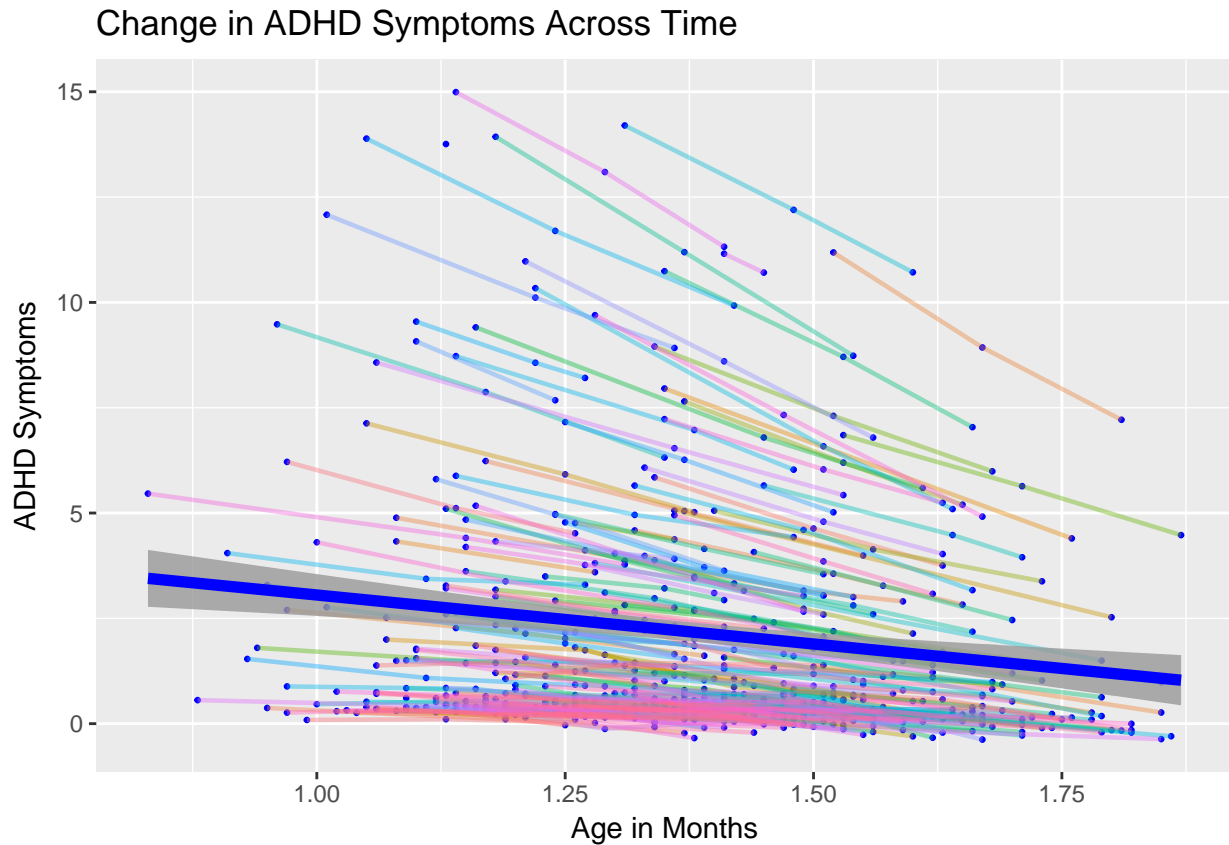
```

```
## Incm_t_N_S_ 0.230 -0.254
```

3. Graph both of your final models for the continuous and nominal models above.

Calculate confidence intervals around your estimates for your final models





```
## Computing profile confidence intervals ...
##           2.5 %    97.5 %
## .sig01      5.894002 10.6846893
## .sig02     -1.000000 -0.9462445
## .sig03           NA         NA
## .sigma           NA         NA
## (Intercept)  5.124051  8.9261166
## agemo_converted -4.695715 -2.2488878
## sex_c        -2.009822 -0.3610493

##Computing profile confidence intervals ...
##           2.5 %    97.5 %
## .sig01      5.6991115 10.79504166
## .sig02     -1.0000000 -0.93767917
## .sig03           NA         NA
## .sigma           NA         NA
## (Intercept)  4.5999620  8.52969672
## agemo_converted -4.4152518 -1.85632875
## Income_to_Need_Scan_c -0.5069412  0.02290092
```

4. Include both types of covariates in a single model. How does your interpretation of parameters change?

```
mod.7 <- lmer(ADHDsum ~ agemo_converted + Income_to_Need_Scan_c + sex_c + (agemo_converted|Subid_fmri),
summary(mod.7)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ADHDsum ~ agemo_converted + Income_to_Need_Scan_c + sex_c + (agemo_converted |
##      Subid_fmri)
##      Data: PDS_data_2
##
## REML criterion at convergence: 2701.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.4792 -0.3210 -0.1312  0.2327  3.8149
##
## Random effects:
##      Groups       Name                Variance Std.Dev. Corr
##      Subid_fmri (Intercept)          69.497   8.336
##                agemo_converted    17.398   4.171   -0.98
##      Residual                        3.435   1.853
## Number of obs: 554, groups:  Subid_fmri, 209
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      6.6245     0.9965   6.648
## agemo_converted  -3.1802     0.6476  -4.911
## Income_to_Need_Scan_c -0.2444     0.1281  -1.907
## sex_c            -1.1798     0.4150  -2.843
##
## Correlation of Fixed Effects:
##              (Intr) agm_cn I__N_S
## agem_cnvrted -0.978
## Incm_t_N_S_  0.226 -0.250
## sex_c        -0.025  0.029 -0.007
anova(mod.6, mod.7)
```

```
## refitting model(s) with ML (instead of REML)
## Data: PDS_data_2
## Models:
## mod.6: ADHDsum ~ agemo_converted + Income_to_Need_Scan_c + (agemo_converted |
## mod.6:      Subid_fmri)
## mod.7: ADHDsum ~ agemo_converted + Income_to_Need_Scan_c + sex_c + (agemo_converted |
## mod.7:      Subid_fmri)
##      Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mod.6  7 2721.2 2751.4 -1353.6  2707.2
## mod.7  8 2715.2 2749.8 -1349.6  2699.2 8.0128      1  0.004645 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Model 7 (including both covariates) appears to be a better fit since the AIC values are smaller.

5. If you have one available, introduce a time-varying covariate.

```
#time-varying covariate = MDD symptoms at each scan
mod.8 <- lmer(ADHDsum ~ agemo_converted + MDDCorescan + Income_to_Need_Scan_c + sex_c + (agemo_converted | Subid_fMRI), data=PDS_data_2)
summary(mod.8)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ADHDsum ~ agemo_converted + MDDCorescan + Income_to_Need_Scan_c +
##      sex_c + (agemo_converted | Subid_fMRI)
##      Data: PDS_data_2
##
## REML criterion at convergence: 2637.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2387 -0.4358 -0.0725  0.3033  3.4690
##
## Random effects:
##      Groups       Name                Variance Std.Dev. Corr
##      Subid_fMRI (Intercept)          34.80     5.899
##                  agemo_converted    6.54     2.557   -0.98
##      Residual                        3.51     1.874
## Number of obs: 554, groups:  Subid_fMRI, 209
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      4.18752    0.92356   4.534
## agemo_converted  -2.31166    0.59687  -3.873
## MDDCorescan       0.55689    0.06316   8.817
## Income_to_Need_Scan_c -0.25229    0.12145  -2.077
## sex_c            -1.04852    0.37086  -2.827
##
## Correlation of Fixed Effects:
##              (Intr) agm_cn MDDCrS I__N_S
## agem_cnvrted -0.969
## MDDCorescan  -0.271  0.134
## Incm_t_N_S_  0.213 -0.244  0.039
## sex_c        -0.038  0.038  0.029 -0.001
```

```
anova(mod.7, mod.8)
```

```
## refitting model(s) with ML (instead of REML)
## Data: PDS_data_2
## Models:
## mod.7: ADHDsum ~ agemo_converted + Income_to_Need_Scan_c + sex_c + (agemo_converted | Subid_fMRI)
## mod.7:      Subid_fMRI)
## mod.8: ADHDsum ~ agemo_converted + MDDCorescan + Income_to_Need_Scan_c + sex_c + (agemo_converted | Subid_fMRI)
## mod.8:      sex_c + (agemo_converted | Subid_fMRI)
##      Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
```

```
## mod.7  8 2715.2 2749.8 -1349.6 2699.2
## mod.8  9 2648.7 2687.6 -1315.4 2630.7 68.473      1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
library(merTools)
re.sim <- REsim(mod.7)
head(re.sim)

##   groupFctr groupID      term      mean    median      sd
## 1 Subid_fMRI   L025 (Intercept) -4.603313 -4.522919 6.634843
## 2 Subid_fMRI   L027 (Intercept) 11.400055 12.349789 7.932123
## 3 Subid_fMRI   L029 (Intercept) -3.422130 -3.925376 6.926335
## 4 Subid_fMRI   L032 (Intercept) -4.185303 -3.671231 6.703955
## 5 Subid_fMRI   L039 (Intercept)  6.956372  7.017352 6.375679
## 6 Subid_fMRI   L041 (Intercept)  8.989068  9.532653 7.421615

p.3 <- re.sim %>%
  filter(term == "(Intercept)")

ggplot(p.3, aes(mean)) +
  geom_density()
```

