Random Coefficient Models for longitudinal Data: Autism Data

Longitudinal Data

- The defining characteristic of a longitudinal data is that individuals are measured repeatedly over time
- Longitudinal data can be considered as a special case of repeated measures data, where measures are repeated over time instead of over regions or under various conditions
- Longitudinal studies are in contrast to cross-sectional studies, in which a single outcome is measured for each individual

Examples of Repeated-Measures data in different disciplines

Substance Abuse

- Study Design: Prospective observational study
- Unit of analysis/Level 2: Subject (random factor) : College
- Subject Level covariates: Geographic region, public/private, rural/urban
- Level 1 / Time variable: Year
- Dependent variable : Percentage of students who use marijuana each academic year
- Level 1 covariates: School ranking, cost of tuition

Business

- Study Design: Panel study
- Unit of analysis/Level 2: Subject (random factor) : Company
- Subject Level covariates: Industry, geographic region
- Level 1/Time variable: Quarter
- Dependent variable : Stock value in each quarter
- Level 1 covariates: Quarterly sales, workforce size

The Autism case study

- The data were collected by the researchers at the University of Michigan (Anderson et al. 2009)
- The study design is a prospective longitudinal study of 214 children
- Children were divided into three diagnostic groups at years of age: autism, pervasive developmental disorder (PDD), and nonspectrum children
- A subset of 158 autism spectrum disorder (ASD) children, including autistic and PDD children is considered for this case study
- Information on each child at ages 2,3,5,9, and 13 years
- However, not all children were measured at each age

- One of the objetcives was to determine the relative influence of the initial diagnostic category (autism or PDD), language proficiency at age 2, and other covariates on the developmental trajectories of the socialization (response) of these children

Exploratory Data Analysis (EDA)

```
# Read in Rat Brain data in long format.
autism <- read.csv("/Users/munnibegum/Library/CloudStorage/OneDrive-BallStateUniversity/Mydocs_21/Teach
head(autism)
```

```
##   age vsae sicdegp childid
## 1   2    6       3       1
## 2   3    7       3       1
## 3   5   18       3       1
## 4   9   25       3       1
## 5  13   27       3       1
## 6   2   17       3       3
```

```
summary(autism)
```

```
##       age             vsae           sicdegp         childid
##  Min.   : 2.000   Min.   :  1.00   Min.   :1.000   Min.   :  1.00
##  1st Qu.: 2.000   1st Qu.: 10.00   1st Qu.:1.000   1st Qu.: 48.75
##  Median : 4.000   Median : 14.00   Median :2.000   Median :107.50
##  Mean   : 5.771   Mean   : 26.41   Mean   :1.956   Mean   :105.38
##  3rd Qu.: 9.000   3rd Qu.: 27.00   3rd Qu.:3.000   3rd Qu.:158.00
##  Max.   :13.000   Max.   :198.00   Max.   :3.000   Max.   :212.00
##                   NA's   :2
```

- Study participants are children having referrals to one of the two autism clinic before the age of 3
- Social development was assessed at each age using Vineland Adaptive Behavior Interview survey
- Children's socialization measures were obtained from parent's report through this survey
- Vineland Socialization Age Equivalent (VSAE), a combined socialization score
- VSAE included assessment of interpersonal relationships, play/leisure time activities, and coping skill
- Initial language development was assessed using the Sequenced Inventory of Communication Development (SICD)
- Children were placed into one of three groupd (SICDEGP) based on their initial SICD scores on the expressive language subsclae at age 2

Specifications of Autism Data

- Study Design: Prospective observational study
- Unit of analysis/Level 2: Subject (random factor) : Child
- Subject Level covariates: Sex, baseline language level
- Level 1/Time variable: Age
- Dependent variable : Socialization score at each age
- Level 1 covariates: Amount of therapy recieved

Data Summary

Number of observations at each level of Age:

```
summary(factor(autism$age))
```

```
##   2   3   5   9  13
## 156 150  91 120  95
```

- The number of children examined at each age differs due to attrition over time
- There are fewer children at age 5 years because one of the clinics did not schedule children to be examined at that age

Number of observations at each level of Age within each group defined by the SICDEGP factor

```
table(factor(autism$sicdegp), factor(autism$age))
```

```
##
##      2  3  5  9 13
##   1 50 48 29 37 28
##   2 66 64 36 48 41
##   3 40 38 26 35 26
```

Overall descriptive statistics for socilization score (VSAE) and at each age

```
summary(autism$vsae)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.00   10.00   14.00   26.41   27.00  198.00       2
```

```
tapply(autism$vsae, factor(autism$age), summary, na.rm = TRUE)
```

```
## $`2`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    6.00    9.00    9.09   11.25   20.00
##
## $`3`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    4.00   10.00   13.00   15.26   19.00   63.00       1
##
## $`5`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.00   12.00   18.00   21.48   28.50   77.00
##
## $`9`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    3.00   15.00   26.00   39.55   63.00  171.00       1
##
## $`13`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     7.0    17.5    45.0    60.6    85.5   198.0
```

- Overall VSAE scores range from 1 to 198 with a mean of 26.41
- The minimum values do not differ much at each age but the means and maximum VSAE scores increased markedly at later ages
- Growth models or longitudunal data models are suitable for analysis of such data

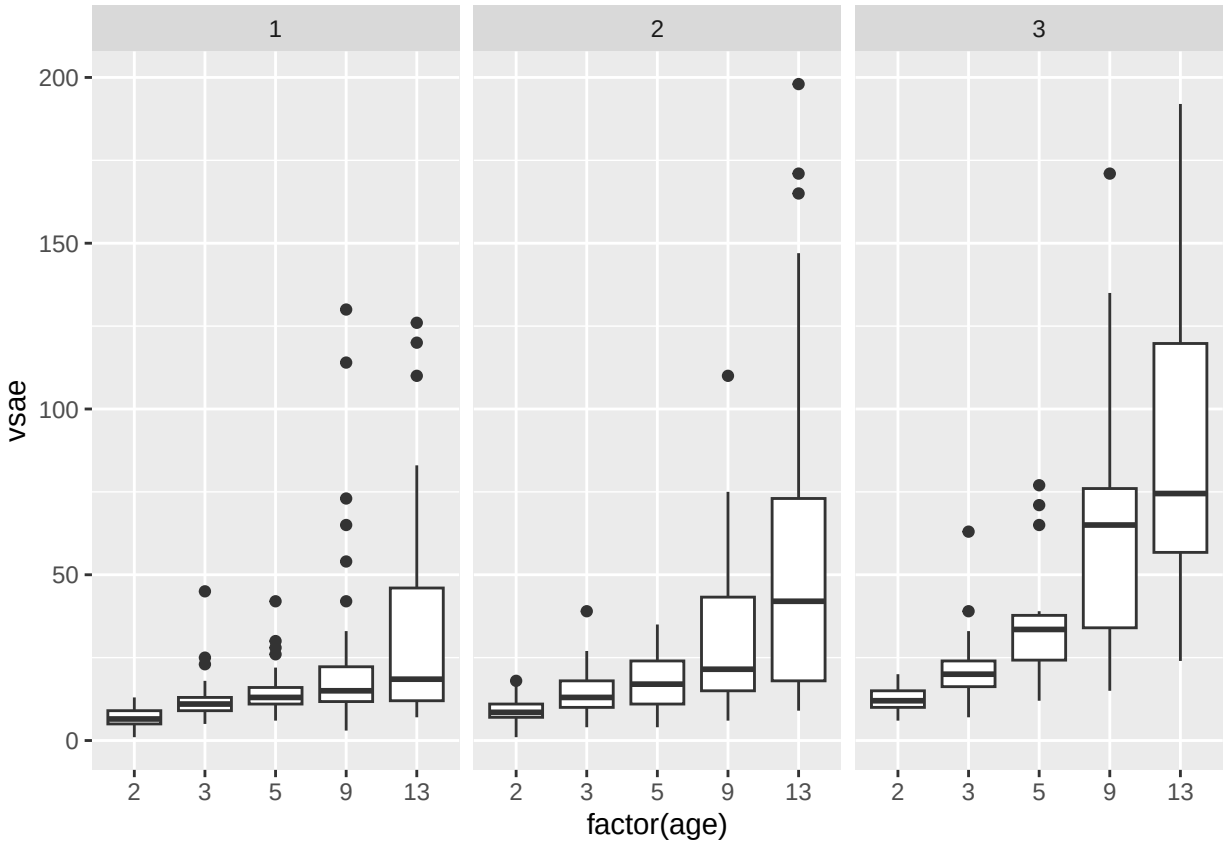
Boxplots of Activation levels by treatment and brain region

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
ggplot(data=autism, mapping = aes(x =factor(age), y = vsae))+
  geom_boxplot()+
  facet_wrap(~factor(sicdegp), nrow=1)
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_boxplot()`).
```

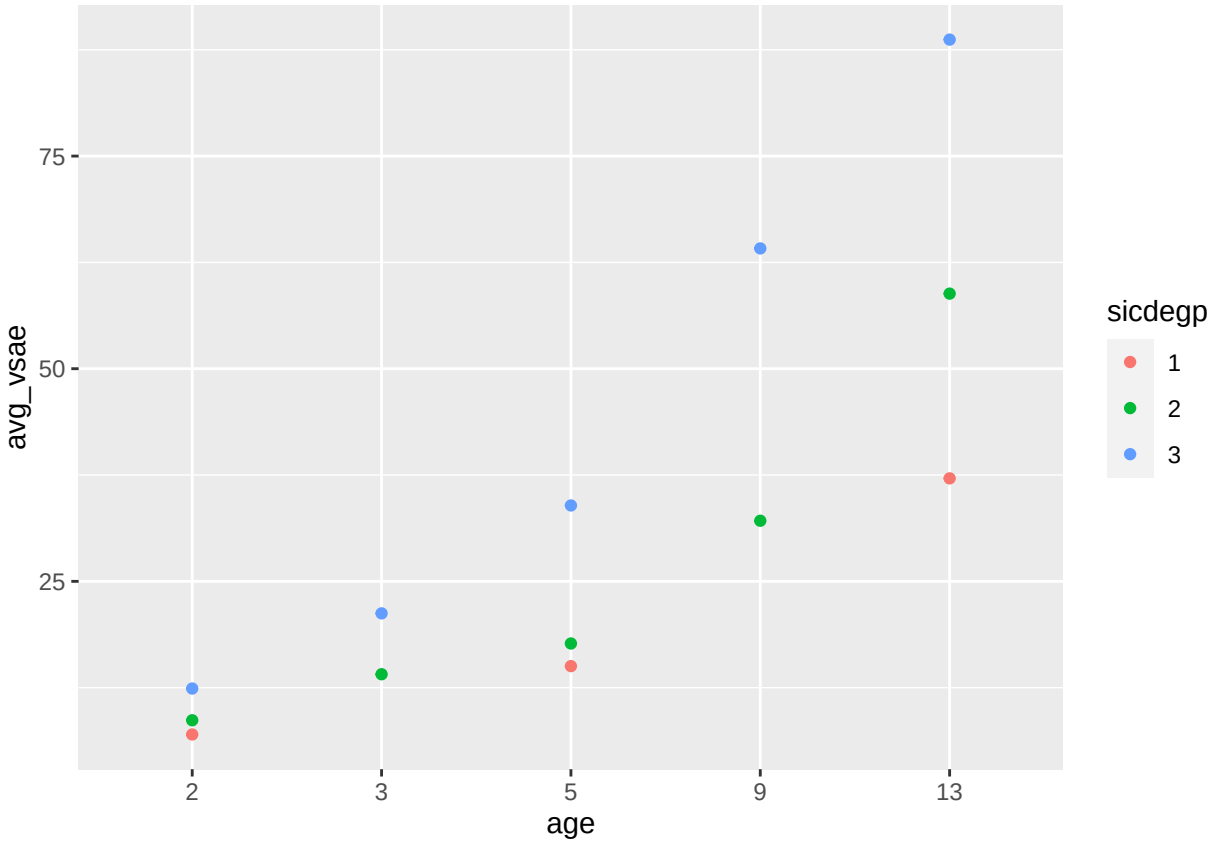


Mean VSAE profiles by SICD Group

```
autism %>%
  mutate(sicdegp= factor(sicdegp)) %>%
  mutate(age= factor(age)) %>%
  group_by(sicdegp, age) %>%
  summarize(avg_vsae = mean(vsae)) %>%
  ggplot()+
  geom_point(aes(y = avg_vsae, x = age, col = sicdegp))
```

```
## `summarise()` has grouped output by 'sicdegp'. You can override using the
## `.groups` argument.
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



- The mean profiles show that on average VSAE scores increase with age
- There may be a quadratic trend in VSAE scores, in particular in SICD group 2
- Both linear and quadratic fixed effects of age should be added to the model
- Possible interactions between the linear and quadratic fixed effects of age and SICD group should be considered as well

Individual child VSAE trajectory by SICD Group

```
library(lcsm)
autism_1 <- filter(autism, sicdegp == 1)
autism_2 <- filter(autism, sicdegp == 2)
autism_3 <- filter(autism, sicdegp == 3)



df_W1 <- autism_1 %>% pivot_wider(names_from = age, values_from = vsae)

colnames(df_W1) <- c("grp", "cid", "Age2", "Age3", "Age5", "Age9", "Age13" )
```

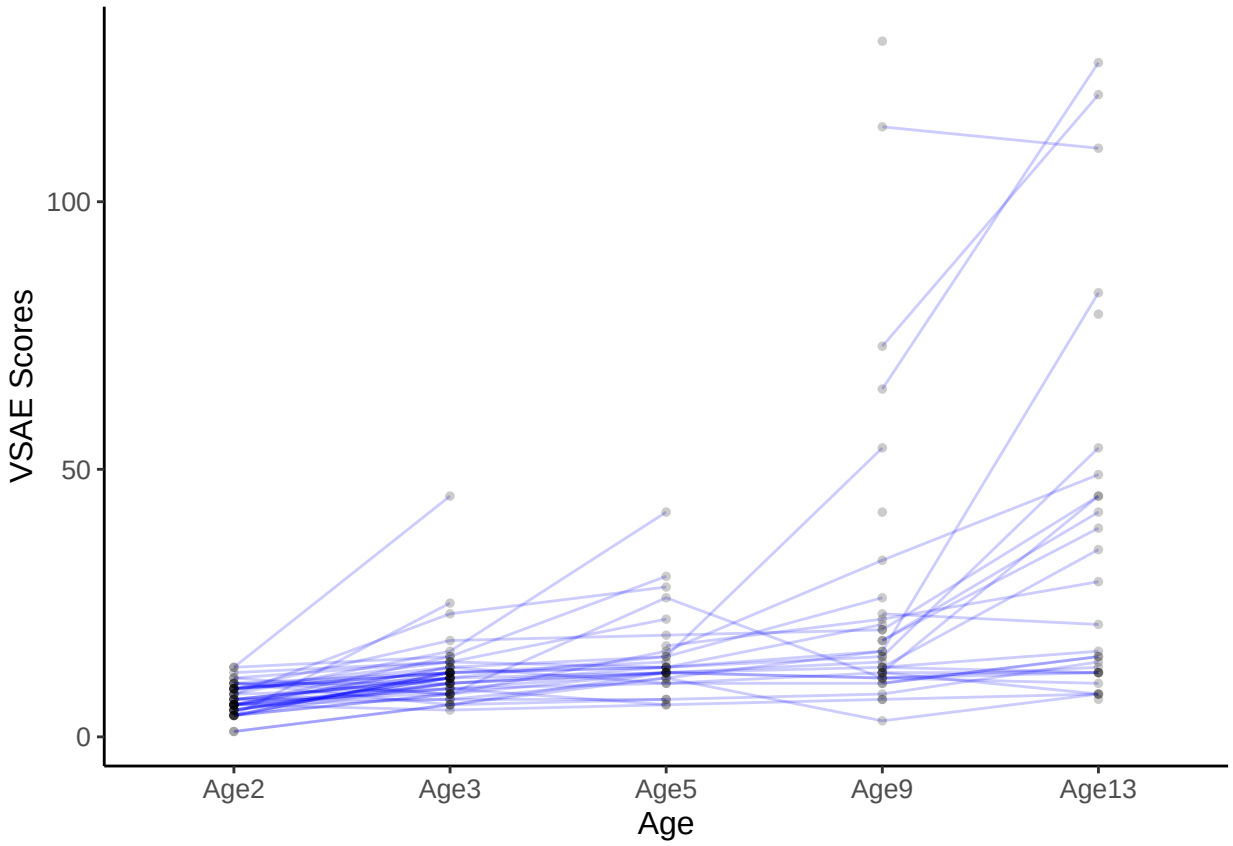
```
df_tr<- select(df_W1, cid, Age2, Age3, Age5, Age9, Age13)

x_var_list <- c("Age2", "Age3", "Age5", "Age9", "Age13")

plot_trajectories(data = df_tr,
                  id_var = "cid",
                  var_list = x_var_list,
                  xlab = "Age", ylab = "VSAE Scores",
                  connect_missing = FALSE,
                  title_n = "sicdegp =1")
```

```
## Warning: Removed 38 rows containing missing values (`geom_line()`).
```

```
## Warning: Removed 60 rows containing missing values (`geom_point()`).
```




```
## Similarly you can create trajectory plots for sicdegp = 2 and 3
```

- A substantial variation in observed VSAE scores from child to child is visible within each SICD group
- The VSAE scores for some children tend to increase as the children get older
- The scores for the other children remain relatively constant
- There is not much variability in the initial VSAE scores at age 2 year for any level of SICD group
- Overall there is increasing between-child variability in the VSAE scores at each successive years of age
- A random coefficient model is suitable to address these features of data

Overview of the Autism Data Analysis

- The individual trajectory plots call for a quadratic regression model for each child, which describes their VSAE scores as a function of age
- The initial model includes the fixed effects of age, age-squared, SICD group, and SICD group by age interaction and the the SICD by age-squared interaction
- Three random effects associated with each child: a random intercept, a random age effect, and a random age-squared effect are considered
- These random effects allow each child to have a unique parabolic trajectory
- The coefficients defining each trajectory vary randomly around fixed effects defining the mean VSAE growth curve for each SICD group

Final Model selection

Final model is chosen based on the likelihood ratio tests for various models considered under different hypotheses to address mean and covariance structure of the measurements