Lab 6 - Latent Growth Models

Structural Equation Modeling - Instructor: Karen Nylund-Gibson

Adam Garber

May 07, 2020

Contents

`University of California, Santa Barbara`

---

1 Lab preparation

1.1 Creating a version-controlled R-Project with Github

Download repository here: https://github.com/garberadamc/SEM-Lab6

On the Github repository webpage:

a. `fork` your own `branch` of the lab repository
b. copy the repository web URL address from the `clone or download` menu

Within R-Studio:

c. click "NEW PROJECT"
d. choose option `Version Control`
e. choose option `Git`
f. paste the repository web URL path copied from the `clone or download` menu on Github page
g. choose location of the R-Project (too many nested folders will result in filepath error)

1.2 Data sources:

1. The first 3 models utilize a public use data subset the *Longitudinal Survey of American Youth (LSAY)* See documentation here
2. The 4th model utilizes a public use data subset the *High School Longitudinal Study (HSLS)* See documentation here

---

Load packages

```
library(gganimate)
library(hrbrthemes)
library(tidyverse)
library(haven)
library(janitor)
library(MplusAutomation)
library(rhdf5)
library(here)
library(kableExtra)
library(gtsummary)
library(semPlot)
```

1.3 `LSAY` data example - `Math Scores` across 6 timepoints

Read in data

```
lsay_data <- read_spss(here("data", "LSAY_Lab6.sav")) %>% select(-starts_with("AB"),
    ends_with("IMP"), -contains("BIO"), -contains("PHY"), -contains("SCI"), FATHED,
    MOTHED) %>% clean_names() %>% rename(math_07 = amthimp, math_08 = cmthimp, math_09 = emthimp,
    math_10 = gmthimp, math_11 = imthimp, math_12 = kmthimp)

lsay_data[lsay_data == 9999] <- NA
```

View metadeta

```
sjPlot::view_df(lsay_data)
```

Write a CSV file

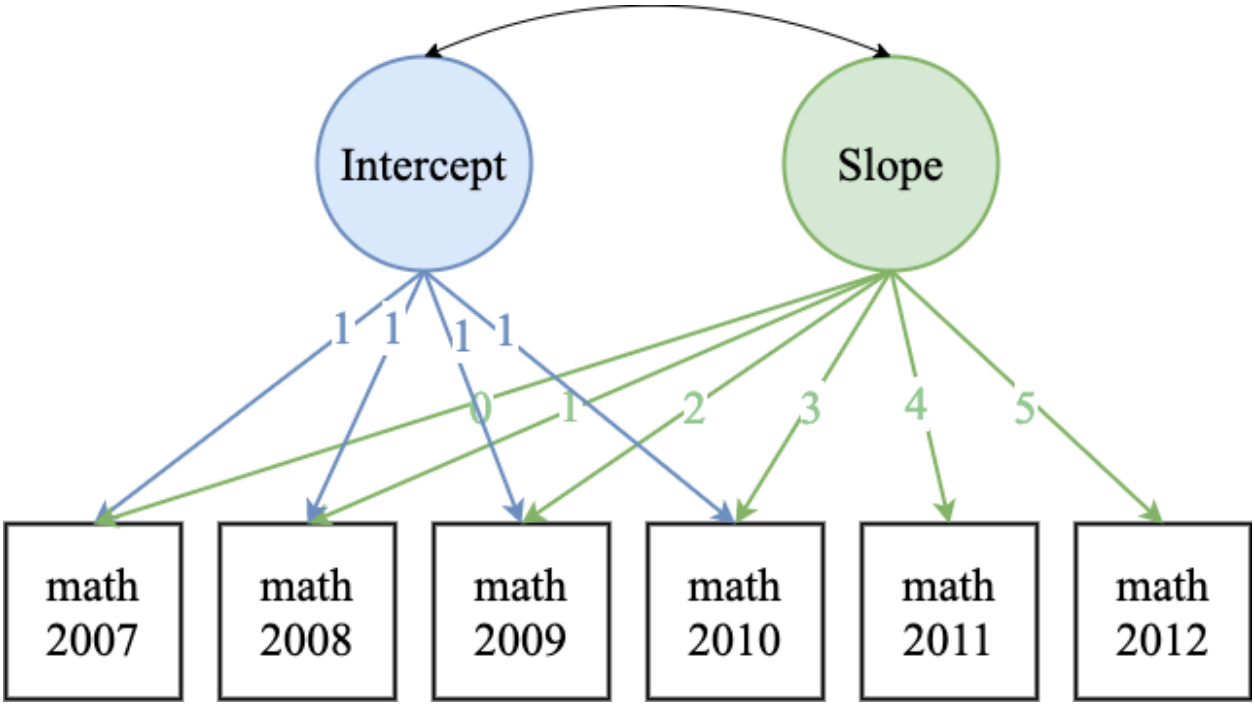
```
write_csv(lsay_data, here("data", "lsay_lab6_data.csv"))
```

Read in the CSV file (SPSS labels removed)

```
lsay_lab6 <- read_csv(here("data", "lsay_lab6_data.csv"))
```

Table. LSAY repeated measures

Name	Labels
math_07	7th grade math score (imputed)
math_08	8th grade math score (imputed)
math_09	9th grade math score (imputed)
math_10	10th grade math score (imputed)
math_11	11th grade math score (imputed)
math_12	12th grade math score (imputed)



1.4 Model 1 - Latent growth model with `fixed time effects` (equal intervals)

```
m1_growth  <- mplusObject(
  TITLE = "m1 growth model fixed time scores - Lab 6",
  VARIABLE =
    "usevar =
    math_07-math_12; ",

  ANALYSIS =
    "estimator = ML" ,

  MODEL =
   "i s | math_07@0 math_08@1 math_09@2 math_10@3 math_11@4 math_12@5; " ,

  OUTPUT = "sampstat standardized;",

  PLOT = "type=plot3;
          series = math_07-math_12(*)",

  usevariables = colnames(lsay_lab6),
  rdata = lsay_lab6)

m1_growth_fit <- mplusModeler(m1_growth,
                      dataout=here("mplus_files", "Lab6.dat"),
                      modelout=here("mplus_files", "m1_growth_Lab6.inp"),
                      check=TRUE, run = TRUE, hashfilename = FALSE)
```

Load in the `mplus.R` functions

```
source(here("mplus.R.txt"))
```

```
## [1] "Loaded rhdf5 package"
```

1.5 Plotting using `gh5` plot data generated by `Mplus`

1. View plots available for a given model
2. Generate plots using the `get.plot.___` function
3. Extract data and transform to tidy format
4. Plot with `ggplot`

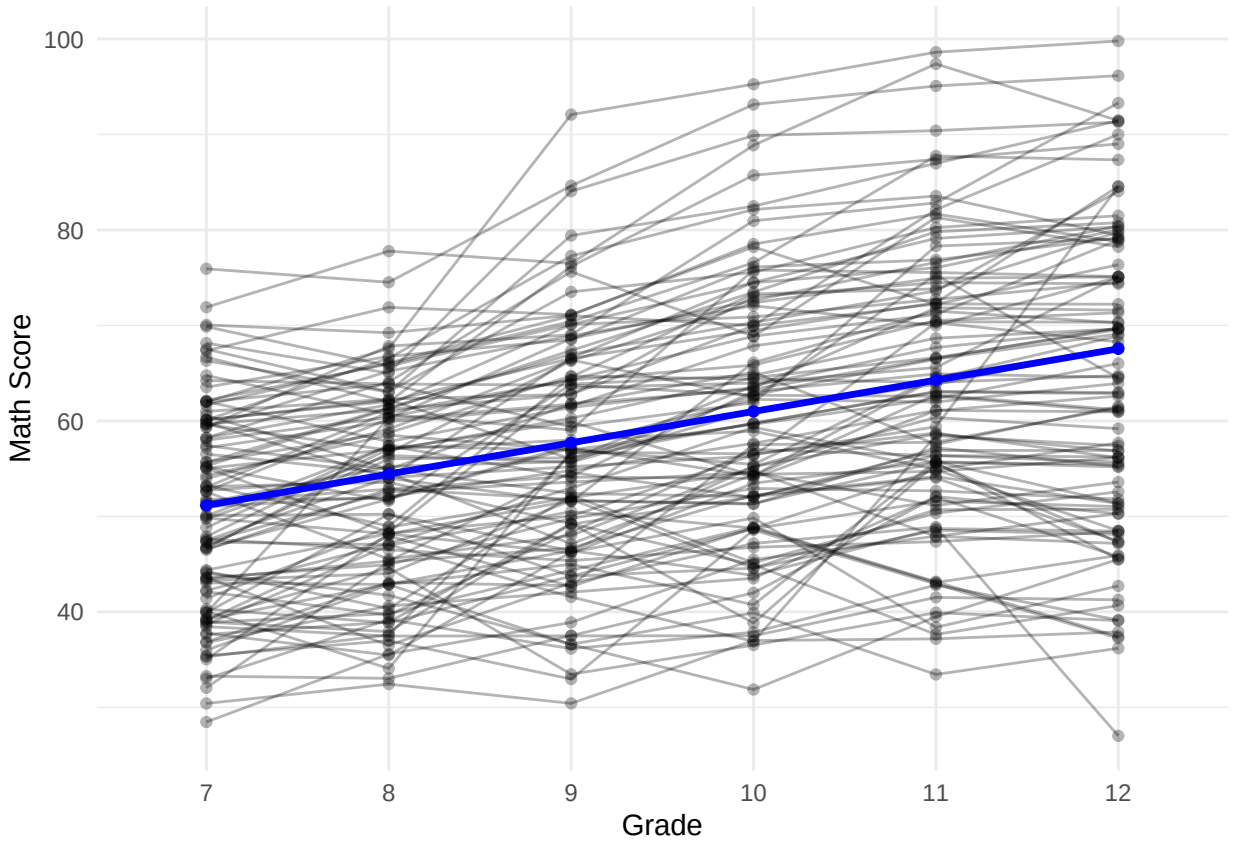
```
mplus.view.plots(here("mplus_files", "m1_growth_Lab6.gh5"))
```

Prepare plot data

```r
observed <- lsay_lab6 %>% select(starts_with("math")) %>%
  rownames_to_column() %>% drop_na()

obs100 <- observed[1:100,]

plot_obs <- obs100 %>%
  pivot_longer(`math_07`:`math_12`, # The columns I'm gathering together
               names_to = "grade", # new column name for existing names
              values_to = "value") # new column name to store values

gradelevels <- colnames(observed[,2:7])

mean_est <- as.data.frame(mplus.get.estimated_means(here("mplus_files", "m1_growth_Lab6.gh5"))) %>%
  mutate(grade = gradelevels)
```

Plot the model estimated means superimposted on the obserbed individual values

```r
growth_plot <- ggplot() +
  geom_point(data = plot_obs, aes(x = grade, y = value, group = rowname), alpha = .3) +     #
  geom_line(data = plot_obs, aes(x = grade, y = value, group = rowname), alpha = .3) +      #
  geom_point(data=mean_est, aes(x=grade, y = V1), color = "Blue", size = 1.5) +             #
  geom_line(data=mean_est, aes(x=grade, y = V1, group = 1), color = "Blue", size = 1.2) +   #
  scale_x_discrete(labels = c("7", "8", "9", "10", "11", "12")) +                           #
  labs(x="Grade", y="Math Score") +                                                         #
  theme_minimal()

growth_plot
```




```
ggsave(here("figures", "spaghetti_p1.png"), height = 6, width = 8, dpi = "retina")
```

Animate the plot with {gganimate}

```
growth_plot + transition_states(rowname, transition_length = 1, state_length = 1) +      #
  shadow_mark(color = "Magenta", alpha = .3)                                             #
```

```
anim_save(here("figures", "spaghetti_plot.gif"), height = 6, width = 8, dpi = "retina")
```

1.6 Model 2 - Latent growth model with `freely estimated time scores` (level-shape model or latent basis model)

```
m2_growth  <- mplusObject(
  TITLE = "m2 growth model freely estimated time scores - Lab 6",
  VARIABLE =
    "usevar =
    math_07-math_12; ",

  ANALYSIS =
    "estimator = ML" ,
```

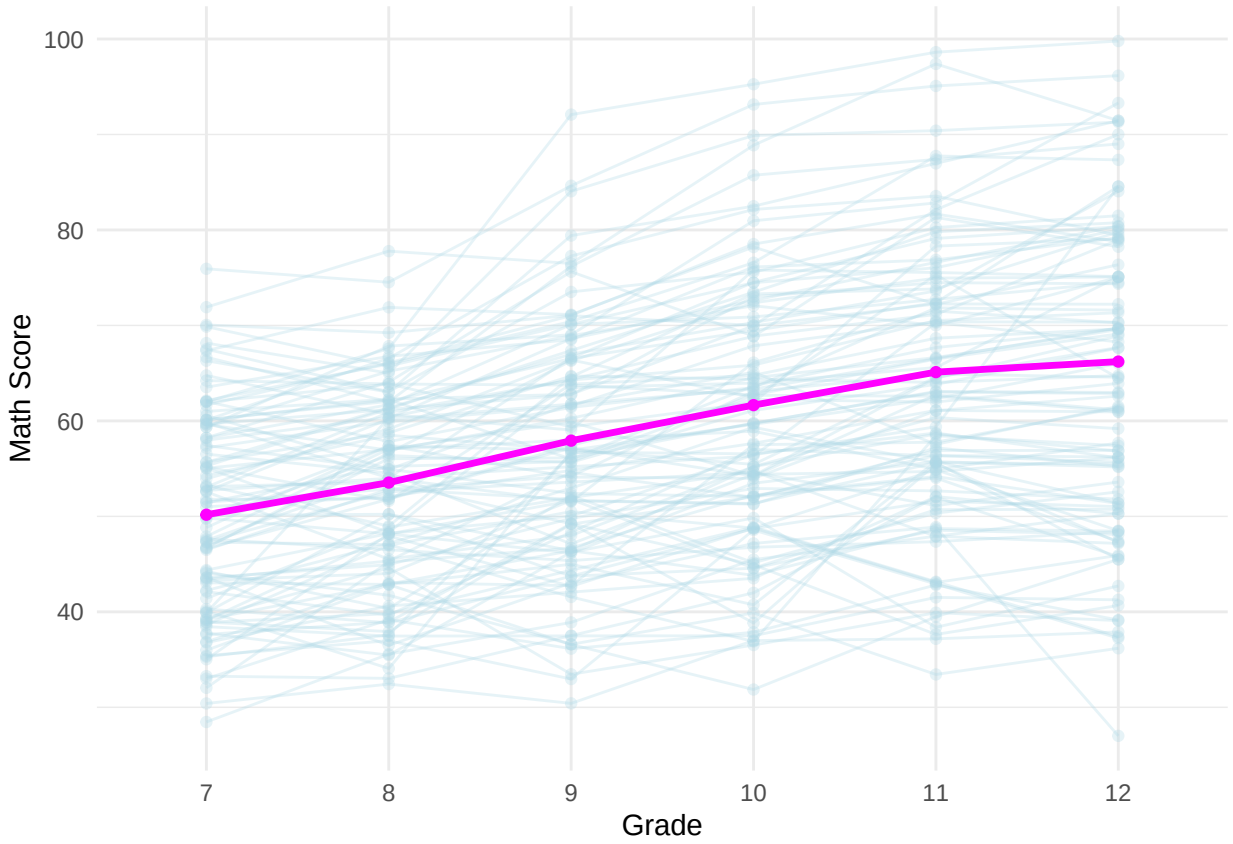
```
  MODEL =
   "i s | math_07@0 math_08@1 math_09* math_10* math_11* math_12*; " ,

  OUTPUT = "sampstat standardized;",

  PLOT = "type=plot3;
          series = math_07-math_12(*)",

  usevariables = colnames(lsay_lab6),
  rdata = lsay_lab6)

m2_growth_fit <- mplusModeler(m2_growth,
                      dataout=here("mplus_files", "Lab6.dat"),
                      modelout=here("mplus_files", "m2_growth_Lab6.inp"),
                      check=TRUE, run = TRUE, hashfilename = FALSE)
```

Prepare plot data

```
mean_est2 <- as.data.frame(mplus.get.estimated_means(here("mplus_files", "m2_growth_Lab6.gh5"))) %>%
    mutate(grade = gradelevels)
```

Plot the model estimated means superimposted on the obserbed individual values

```
growth_plot <- ggplot() +
  geom_point(data = plot_obs, aes(x = grade, y = value, group = rowname), color = "lightblue", alpha =
  geom_line(data = plot_obs, aes(x = grade, y = value, group = rowname), color = "lightblue", alpha = .3
  geom_point(data=mean_est2, aes(x=grade, y = V1), color = "magenta", size = 1.5) +
  geom_line(data=mean_est2, aes(x=grade, y = V1, group = 1), color = "magenta", size = 1.2) +
  scale_x_discrete(labels = c("7", "8", "9", "10", "11", "12")) +
  labs(x="Grade", y="Math Score") +
  theme_minimal()

growth_plot
```



1.7 Model 3 - Latent growth model with covariate and freely estimated time scores

```
m3_growth  <- mplusObject(
  TITLE = "m3 growth model with covariate and freely estimated time scores - Lab 6",
  VARIABLE =
    "usevar =
    math_07-math_12 fathed; ",

  ANALYSIS =
    "estimator = ML" ,

  DEFINE = "center fathed (grandmean);",

  MODEL =
   "i s | math_07@0 math_08@1 math_09* math_10* math_11* math_12*;
    i s on fathed; " ,

  OUTPUT = "sampstat standardized;",

  PLOT = "type=plot3;
```

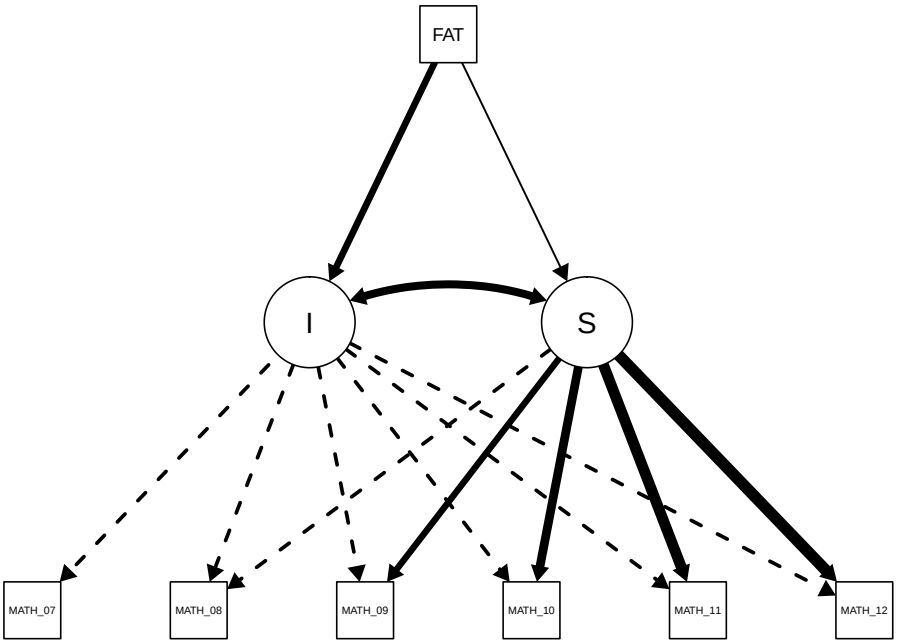
```
          series = math_07-math_12(*)",

  usevariables = colnames(lsay_lab6),
  rdata = lsay_lab6)

m3_growth_fit <- mplusModeler(m3_growth,
                      dataout=here("mplus_files", "Lab6.dat"),
                      modelout=here("mplus_files", "m3_growth_Lab6.inp"),
                      check=TRUE, run = TRUE, hashfilename = FALSE)
```

Check the path diagram of the model with {semPlot}

```
m3_output <- readModels(here("mplus_files", "m3_growth_Lab6.out"))
```

```
## Reading model:  /Users/agarber/Desktop/SEM_S20/Lab6_SEM/mplus_files/m3_growth_Lab6.out
```

```
semPaths(m3_output, "est", intercepts = FALSE, residuals = FALSE, fade = FALSE, edge.color = "black",
    edgeLabels = "")
```



1.8 HSLS data example - Academic expectations

```
hsls_rep <- read_csv(here("data", "hsls_rep_lab6.csv"))
```

Table. HSLS repeated measures

Question stem - `Highest level of education expected...`

Name	Labels	Levels
s1eduexp	9th grade (2009)	1 = less HS, 2 = HS, 3 = Bach, 5 = Master, 6 = Ph.D
s2eduexp	11th grade (2012)	1 = less HS, 2 = HS, 3 = Bach, 5 = Master, 6 = Ph.D
s4eduexp	3 years post high school (2016)	1 = less HS, 2 = HS, 3 = Bach, 5 = Master, 6 = Ph.D

1.9 Model 4 - Latent growth model with `categorical outcomes`

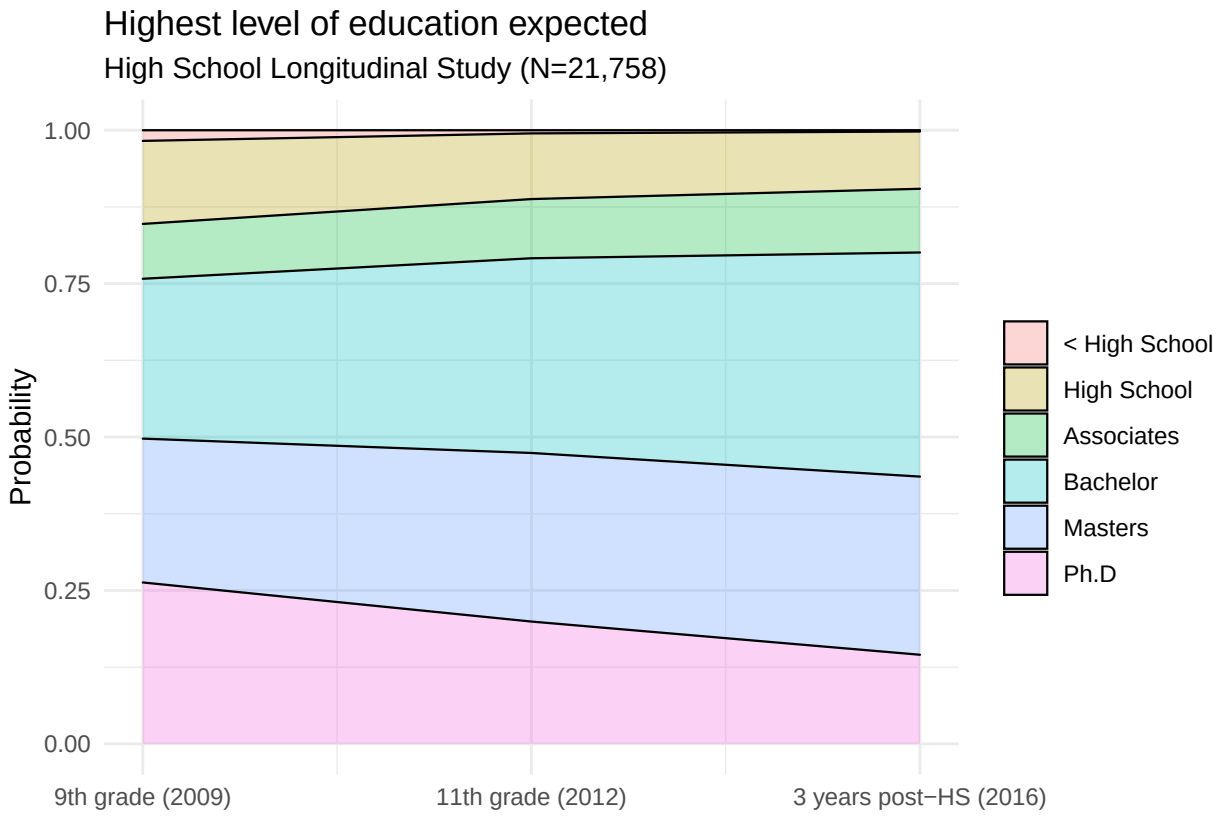
```
m4_growth  <- mplusObject(
  TITLE = "m4 growth model - HSLS  - Lab 6",

  VARIABLE =
    "usevar = s1eduexp-s4eduexp;
     categorical = s1eduexp-s4eduexp;",

  ANALYSIS = "" ,

  MODEL =
   "! 0=09 1=10 2=11 3=12 | 4=13 5=14 6=15 7=16

    i s | s1eduexp@0 s2eduexp@3 s4eduexp@7;  ",

  OUTPUT = "sampstat standardized;",

  PLOT = "type=plot3;
          series = s1eduexp-s4eduexp(*);",

  usevariables = colnames(hsls_rep),
  rdata = hsls_rep)

m4_growth_fit <- mplusModeler(m4_growth,
                      dataout=here("mplus_files", "Lab6.dat"),
                      modelout=here("mplus_files", "m4_growth_Lab6.inp"),
                      check=TRUE, run = TRUE, hashfilename = FALSE)
```

Prepare plot data

```r
loop_data <- lapply(1:6, function(k) {
  probs <- mplus.get.estimated_probabilities(here("mplus_files", "m4_growth_Lab6.gh5"),'process1',k,k)

  loop_data <- as.data.frame(probs) %>%
    mutate(cat = factor(k))
})

plot_data <- bind_rows(loop_data)

observed <- hsls_rep %>% select(contains("eduexp")) %>%
  rownames_to_column() %>% drop_na()

obs100 <- observed[1:100,]

plot_obs <- obs100 %>%
  pivot_longer(`s1eduexp`:`s4eduexp`,      # The columns I'm gathering together
               names_to = "year",          # new column name for existing names
              values_to = "value") %>%     # new column name to store values
  mutate(year = case_when(
         year == "s1eduexp" ~ 1,
         year == "s2eduexp" ~ 2,
         year == "s4eduexp" ~ 3,
  ))

yearlevels <- colnames(observed[,2:4])

prob_est <- plot_data %>%
  mutate(year = rep(1:3, 6))
```

Plot the model estimated probabilities (`categorical outcomes`)

```r
ggplot(data=prob_est, aes(x=year, y=V1, fill=cat)) +                                    #
  geom_area(alpha=0.3 , size=.4, colour="black") +                                      #
  scale_x_continuous(breaks = 1:3,                                                      #
    labels = c("9th grade (2009)","11th grade (2012)","3 years post-HS (2016)")) +      #
  scale_y_continuous("Probability") +                                                   #
  scale_fill_discrete("",                                                               #
    labels = c("< High School", "High School", "Associates", "Bachelor", "Masters", "Ph.D")) +  #
  labs(title="Highest level of education expected",                                     #
       subtitle = "High School Longitudinal Study (N=21,758)", y="Probability", x="") +  #
  theme_minimal()                                                                         #
```




```
ggsave(here("figures", "cat_growth_plot.png"), height = 6, width = 8, dpi = "retina")
```

Create an animated plot with {gganimate}

```
cat_plot <- ggplot(data = plot_obs, aes(x = year, y = value, group = rowname)) +                #
  geom_jitter(color = "black", alpha = 0, width = 0.1, height = .3) +                          #
  geom_line(color = "black") +                                                                 #
  scale_x_continuous(breaks = 1:3,                                                             #
    labels = c("9th grade (2009)","11th grade (2012)","3 years post-HS (2016)")) +             #
  scale_y_reverse(breaks = 1:6, labels = c("< HS", "HS", "AA", "BA/BS", "MA", "Ph.D")) +       #
  theme_minimal() + theme(panel.grid.minor = element_blank()) +                                  #
  labs(y="", title="Highest level of education expected",                                      #
       subtitle = "High School Longitudinal Study (N=100, sub-sample)")                        #

cat_plot + transition_states(rowname, transition_length = 3, state_length = 3) +               #
  shadow_mark(color = "blue", alpha = .15)                                                     #
```

```
anim_save(here("figures", "cat_growth_anim.gif"), height = 6, width = 8, dpi = "retina")
```

---

2 References

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. Structural equation modeling: a multidisciplinary journal, 25(4), 621-638.

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., ... & Leinwand, S. (2011). High School Longitudinal Study of 2009 (HSLS: 09): Base-Year Data File Documentation. NCES 2011-328. National Center for Education Statistics.

Miller, J. D., Hoffer, T., Suchner, R., Brown, K., & Nelson, C. (1992). LSAY codebook. Northern Illinois University.

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2017). Regression and mediation analysis using Mplus. Los Angeles, CA: Muthén & Muthén.

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

UC SANTA BARBARA