Moderation

A Course in `MplusAutomation`

Adam Garber

Outline

1. Conduct two moderation analyses using different applied examples.
2. Create simple slope plots from new parameters calculated using the `MODELCONSTRAINT=` option in Mplus.
3. Evaluate moderation results using the `MODELCONSTRAINT=` option in Mplus to conduct difference tests between point estimates.

Data source for example 1:

The first example utilizes the *Vocabulary and Education* dataset from the National Opinion Research Center General Social Survey. GSS Cumulative Datafile 1972-2016 (Fox, 2008) See documentation here

This data is available via the R-package {`carData`} and can be directly loaded into the R environment.

Note: All models specified in the following exercise are for demonstration only and are **not** theoretically justified or valid.

Load packages

```
library(tidyverse)
library(MplusAutomation)
library(here)
library(gt)
library(gtsummary)
library(carData)  # contains data for example 1 (`Vocab`)
library(Ecdat)    # contains data for example 2 (`Star`)
```

Read the `data.frame` into your R-environment directly from the {`carData`} package

```
data(Vocab)
```

Prepare data for analysis (`filter` rows & convert `year` to a factor)

```r
vocab2 <-  Vocab %>%
  filter(year %in% c(1974, 2016)) %>%
  mutate(year = droplevels(factor(year)))
```

Take a look at focal variables

Name	Labels
year	Year of the survey (1974, 2016)
education	Education, in years
vocabulary	Vocabulary test score: number correct on a 10-word test

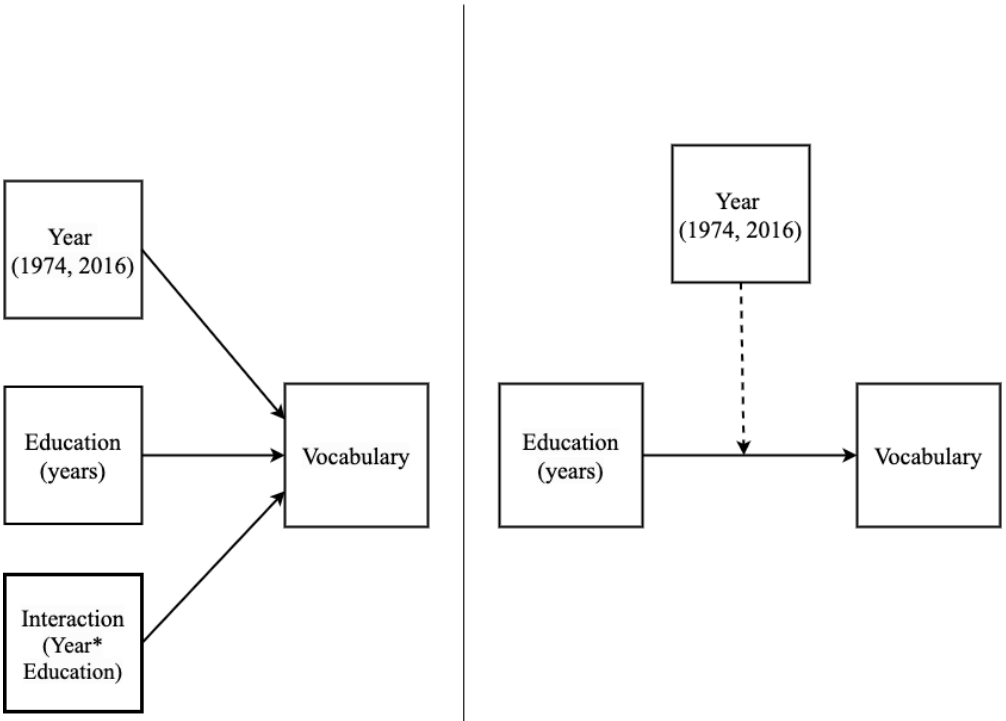
Check some basic descriptives with the {gtsummary} package

```r
table1 <- tbl_summary(vocab2,
                      statistic = list(all_continuous() ~ "{mean} ({sd})"),
                      missing = "no" ) %>%
  bold_labels()

table1
```

Characteristic	**N = 3307**
year	
1974	1446 (44%)
2016	1861 (56%)
sex	
Female	1812 (55%)
Male	1495 (45%)
education	12.9 (3.1)
vocabulary	6.02 (2.05)

Estimate moderation example 1

1. Covariate: Years of education (`education`)
2. Moderator: Year of the survey with 2-levels 1974 and 2016 (`year`)
3. Outcome: Vocabulary test score - number correct on a 10-word test (`vocabulary`)




```
m1_model  <- mplusObject(
  TITLE = "m1 model moderation",

  VARIABLE =
   "usevar =
    year              !!! covariate/moderator
    education         !!! covariate
    vocabulary        !!! outcome
    int_yred;         !!! interaction of year and education",

  DEFINE =
    "center education (grandmean);
     int_yred = year*education;  !!! create interaction term !!!",

  ANALYSIS =
    "estimator = MLR" ,

  MODEL =
   "[vocabulary](b0); !!! intercept !!!

    vocabulary on     !!! outcome !!!
```

```
    year(b1)          !!! covariate/moderator !!!
    education(b2)     !!! covariate !!!
    int_yred(b3);     !!! interaction of year and education !!! " ,

  MODELCONSTRAINT =
  "LOOP(x,-1,1,0.01);
   PLOT(y1974 y2016);
   y1974 = b0 + b2*x;
   y2016 = b0 + b1 + (b2 + b3)*x;

   new(hi_y1974 lo_y1974 hi_y2016 lo_y2016 diff_hi);

   hi_y1974 = b0 + b2*(6.28);  !!! value (6.28) used is plus-minus 2 SD's on X !!!
   lo_y1974 = b0 + b2*(-6.28);
   hi_y2016 = b0 + b1 + (b2 + b3)*(6.28);
   lo_y2016 = b0 + b1 + (b2 + b3)*(-6.28);

   diff_hi = hi_y2016 - hi_y1974; !!! test the significance of distance between 2 points !!! ",

  OUTPUT = "sampstat standardized modindices (3.84)",

  PLOT = "type=plot3;",

  usevariables = colnames(vocab2),
  rdata = vocab2)

m1_model_fit <- mplusModeler(m1_model,
                    dataout=here("08-moderation", "mplus_files", "model1_vocab.dat"),
                    modelout=here("08-moderation", "mplus_files", "model1_vocab.inp"),
                    check=TRUE, run = TRUE, hashfilename = FALSE)
```

Create the simple slope plot from `Mplus` model output

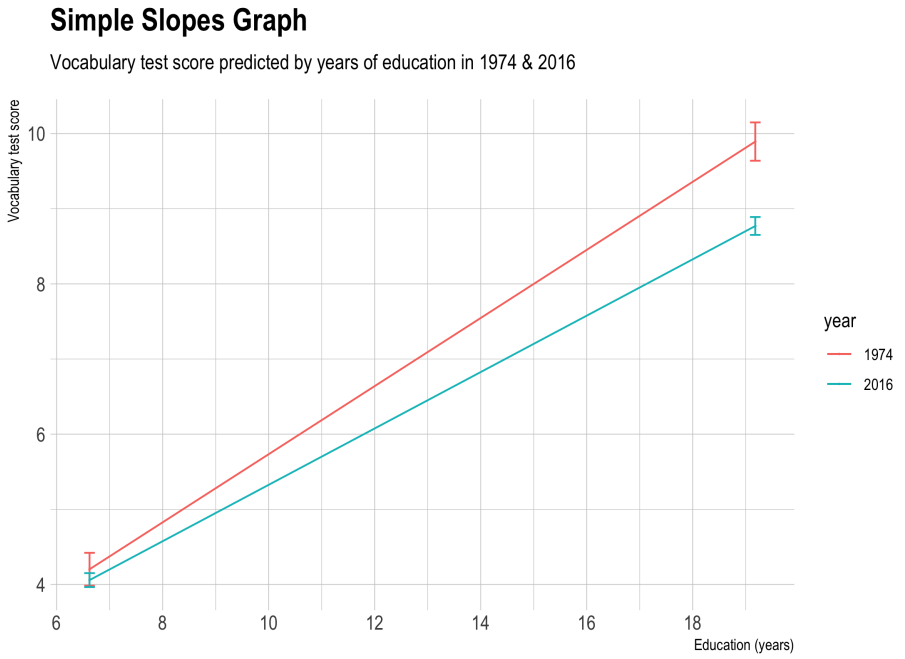
Extract the output parameters generated using the `model constraint`

```
simp_slope <- data.frame(m1_model_fit[["results"]][["parameters"]][["unstandardized"]]) %>%
  filter(paramHeader == "New.Additional.Parameters")  %>%
  filter(param != "DIFF_HI") %>%
  select(param, est, se) %>%
  mutate(year = case_when(
     param %in% c("HI_Y1974", "LO_Y1974") ~ "1974",
     param %in% c("HI_Y2016", "LO_Y2016") ~ "2016")) %>%
  mutate(education = case_when(
     param %in% c("HI_Y1974", "HI_Y2016") ~ 6.28,
     param %in% c("LO_Y1974", "LO_Y2016") ~ -6.28))
```

Plot the interaction effect with `ggplot`

```r
# un-center `education` so values on x-axis are on the original scale
plot_data <- simp_slope %>% mutate(education = education + 12.9)

ggplot(plot_data, aes(x=education, y=est, color=year, group=year)) +
  geom_point(size=0) +
  geom_line() +
  geom_errorbar(aes(ymin=est-se, ymax=est+se),
                width=.2) +
  scale_x_continuous(breaks = c(seq(6,20,2))) +
  labs(title = "Simple Slopes Graph",
       subtitle = "Vocabulary test score predicted by years of education in 1974 & 2016",
           x = "Education (years)",
           y = "Vocabulary Test Score") +
  theme_minimal()
```

```r
ggsave(here("08-moderation", "figures", "m1_simple_slope.png"), height = 6, width = 8)
```



Data source for example 2:

The next example utilizes the **Effects on Learning of Small Class Sizes (Star)** dataset from the *Introduction to Econometrics* textbook. (Stock et al., 2003) See documentation here

This data is available via the R-package {`Ecdat`} and can be directly loaded into the `R` environment.

Read the dataframe into your R-environment from package {Ecdat}

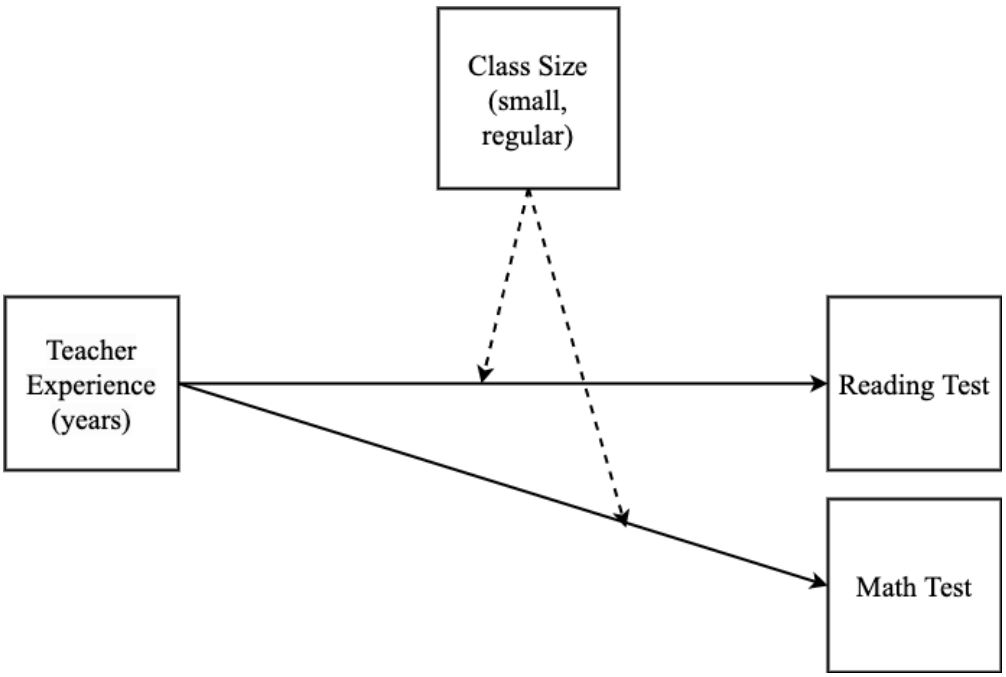
```
data(Star)

star_data <- as.data.frame(Star)
```

Take a look at the variables in the Star dataset

Name	Labels
tmathssk	total math scaled score
treadssk	total reading scaled score
classk	type of class (small, regular, regular with aide)
totexpk	years of total teaching experience

Subset and recode variables to use in moderation model with select, mutate, and case_when

```
mod_data <- star_data %>%
  select(totexpk, # years of total teaching experience
         classk,  # type of class, a factor with levels (regular,small.class,regular.with.aide)
         tmathssk, treadssk)  %>%
  mutate(classk = case_when(
         classk == "small.class" ~ "small.class",
         classk %in% c("regular.with.aide", "regular") ~ "regular")) %>%
  mutate(classk = fct_rev(classk))
```



Estimate moderation example 2

1. covariate: Years of education (`totexpk`)
2. moderator: type of class (small, regular) (`classk`)
3. outcome 1: total math scaled score (`tmathssk`)
4. outcome 2: total reading scaled score (`treadssk`)

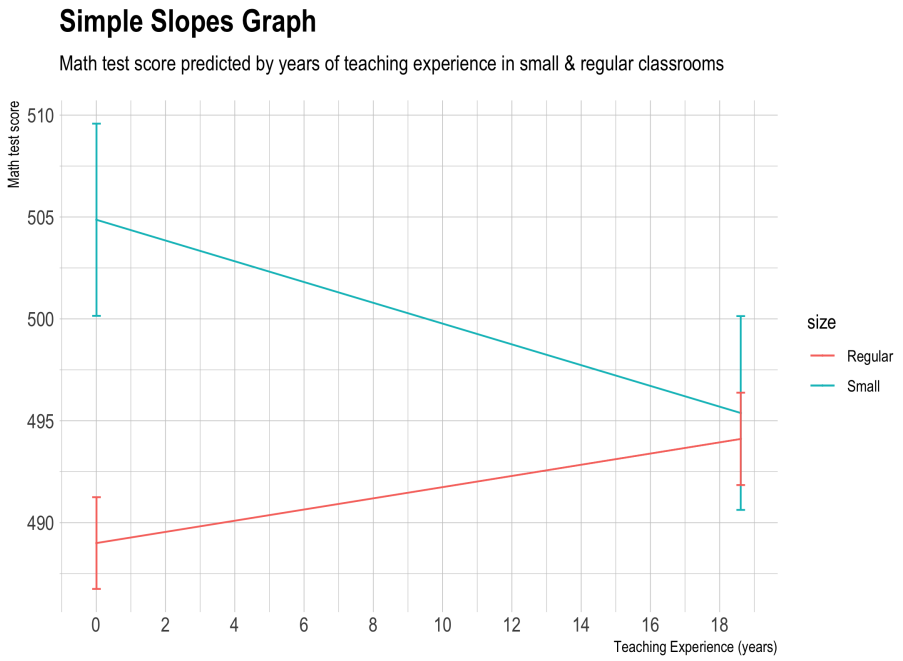
```
m2_model  <- mplusObject(
  TITLE = "m2 model indirect",

  VARIABLE =
   "usevar =
    totexpk classk
    tmathssk, treadssk
    tchXclas;   ",

  DEFINE =
    "center totexpk (grandmean);
     tchXclas = totexpk*classk;  ! create interaction term" ,

  ANALYSIS =
    "estimator = mlr; ",

  MODEL =
   "treadssk on classk totexpk tchXclas;

    [tmathssk](b0);

    tmathssk on
    classk (b1)
    totexpk (b2)
    tchXclas (b3); ",

  MODELCONSTRAINT =

  "LOOP(x,-1,1,0.01);
   PLOT(small regular);

   small = b0 + b2*x;
   regular = b0 + b1 + (b2+b3)*x;

   new(hi_small lo_small hi_regular lo_regular diff_hi);

   hi_small = b0 + b2*(9.3);
   lo_small = b0 + b2*(-9.3);
   hi_regular = b0 + b1 + (b2 + b3)*(9.3);
   lo_regular = b0 + b1 + (b2 + b3)*(-9.3);
   diff_hi = hi_small - hi_regular; ",
```

```
  OUTPUT = "sampstat standardized modindices (3.84)",

  PLOT = "type=plot3;",

  usevariables = colnames(mod_data),
  rdata = mod_data)

m2_model_fit <- mplusModeler(m2_model,
                    dataout=here("08-moderation", "mplus_files", "model2_caschools.dat"),
                    modelout=here("08-moderation", "mplus_files", "model2_caschools.inp"),
                    check=TRUE, run = TRUE, hashfilename = FALSE)
```

Create the simple slope plot from Mplus model output

```
 simp_slope2 <- data.frame(m2_model_fit[["results"]][["parameters"]][["unstandardized"]]) %>%
   filter(paramHeader == "New.Additional.Parameters")  %>%
   filter(param!= "DIFF_HI") %>%
   select(param, est, se) %>%
   mutate(size = case_when(
      param %in% c("HI_SMALL", "LO_SMALL") ~ "Small",
      param %in% c("HI_REGUL", "LO_REGUL") ~ "Regular")) %>%
   mutate(experience = case_when(
      param %in% c("HI_SMALL", "HI_REGUL") ~ 9.3,
      param %in% c("LO_SMALL", "LO_REGUL") ~ -9.3))
```

```
# un-center 'experience' so values on x-axis are on the original scale
mean_exp <- mean(mod_data$totexpk)
plot_data2 <- simp_slope2 %>% mutate(experience = experience + mean_exp)

ggplot(plot_data2, aes(x=experience, y=est, color=size, group=size)) +
  geom_point(size=0) +
  geom_line() +
  geom_errorbar(aes(ymin=est-se, ymax=est+se), width=.25) +
  scale_x_continuous( breaks = c(seq(0,18,2))) +
  labs(title = "Simple Slopes Graph",
       subtitle = "Math test score predicted by years of teaching experience in small & regular classroo
           x = "Teaching Experience (years)",
           y = "Math Test Score") +
  theme_minimal()

ggsave(here("08-moderation", "figures", "m2_simple_slope.png"), height = 6, width = 8)
```



End

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.

Horst, A. (2020). Course & Workshop Materials. GitHub Repositories, https://https://allisonhorst.github.io/

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

Further resources & examples here:

https://garberadamc.github.io/project-site/

https://www.adam-garber.com/
