Problem Set 2: Interaction

Nora Nickels

4/26/2018

Problem Set 2

Problem 2: Interactions Terms

```r
# Load libraries

library(tidyverse)
library(forcats)
library(broom)
library(modelr)
library(stringr)
library(titanic)
library(rcfss)
library(car)
library(haven)

options(digits = 3)
set.seed(1234)
theme_set(theme_minimal())


# Load data

biden_df = read.csv("biden.csv") %>%
  na.omit() %>%
  mutate(dem = factor(dem),
         rep = factor(rep))
```

Linear Regression Model

```r
# Fit linear regression model of biden score on age, education and the interaction b/t age and educ.
lm_biden <- biden_df %>%
  lm(biden ~ age + educ + age * educ, data = .)

# Report regression coefficients
tidy(lm_biden)
```

```
##          term estimate std.error statistic  p.value
## 1 (Intercept)   38.374    9.5636      4.01 6.25e-05
## 2         age    0.672    0.1705      3.94 8.43e-05
## 3        educ    1.657    0.7140      2.32 2.04e-02
## 4    age:educ   -0.048    0.0129     -3.72 2.03e-04
```

```r
biden_coef <- tidy(lm_biden)
```

```r
# Report goodness of fit
glance(lm_biden)
```

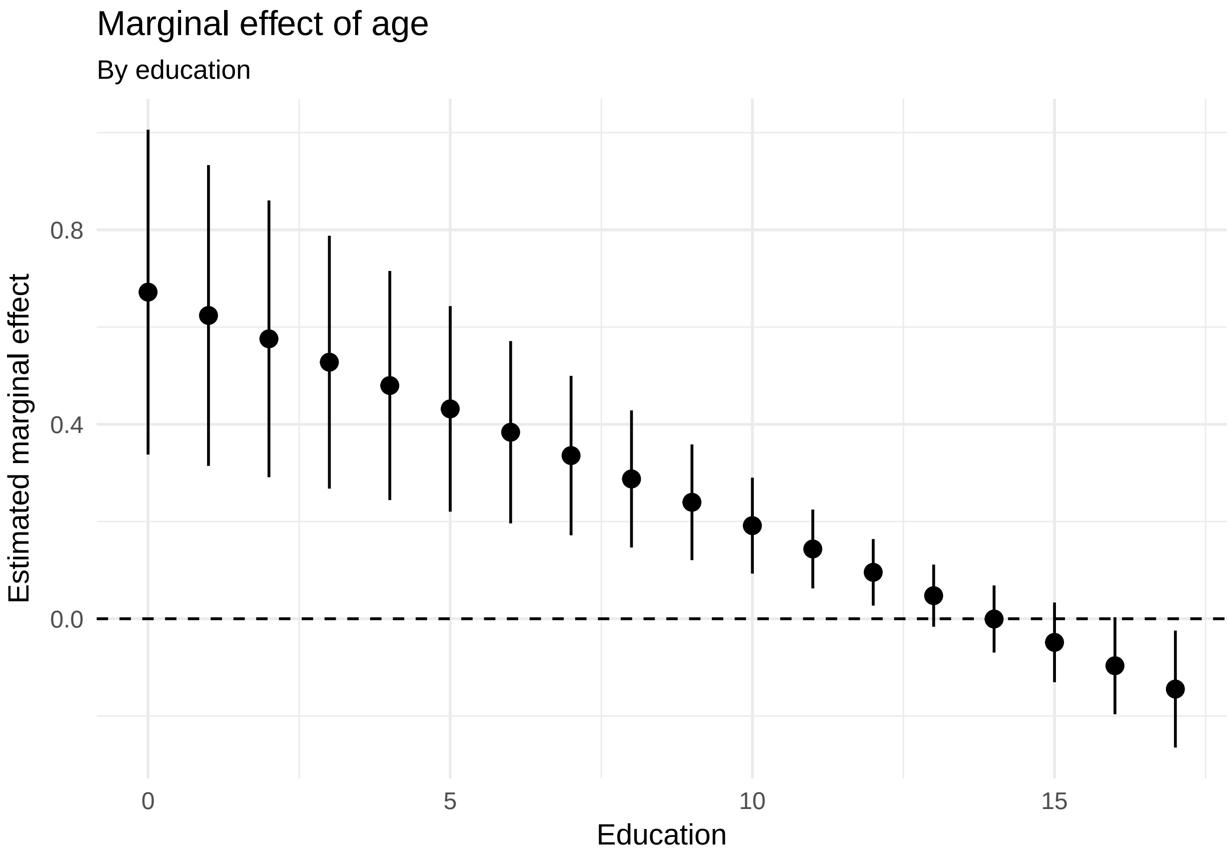
```
##   r.squared adj.r.squared sigma statistic  p.value df logLik   AIC   BIC
## 1    0.0176        0.0159  23.3      10.7 5.37e-07  4  -8249 16509 16536
##   deviance df.residual
## 1   976688        1803
```

```r
biden_fit <- glance(lm_biden)
```

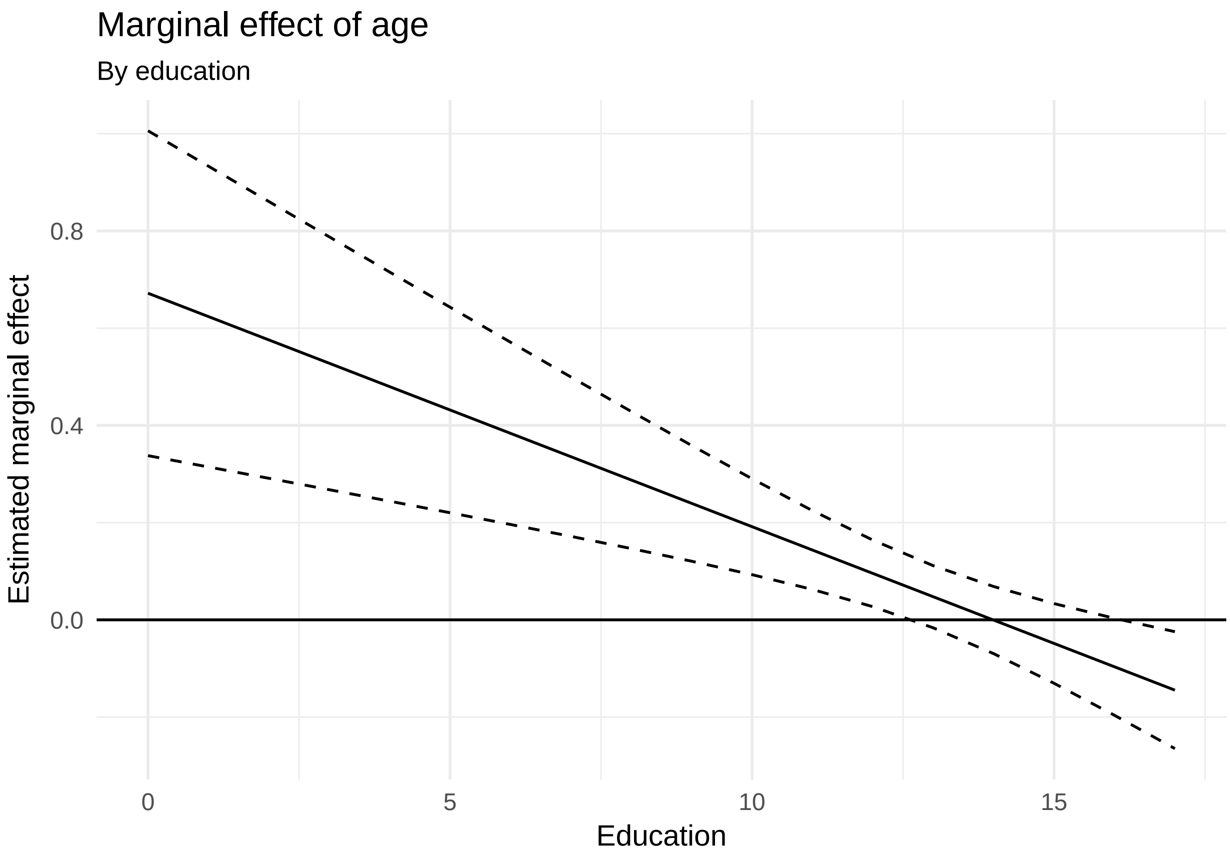
2A: Evaluate the marginal effect of age on Joe Biden thermometer rating, conditional on education

```r
# function to get point estimates and standard errors
# model - lm object
# mod_var - name of moderating variable in the interaction
instant_effect <- function(model, mod_var){
  # get interaction term name
  int.name <- names(model$coefficients)[[which(str_detect(names(model$coefficients), ":"))]]

  marg_var <- str_split(int.name, ":")[[1]][[which(str_split(int.name, ":")[[1]] != mod_var)]]

  # store coefficients and covariance matrix
  beta.hat <- coef(model)
  cov <- vcov(model)

  # possible set of values for mod_var
  if(class(model)[[1]] == "lm"){
    z <- seq(min(model$model[[mod_var]]), max(model$model[[mod_var]]))
  } else {
    z <- seq(min(model$data[[mod_var]]), max(model$data[[mod_var]]))
  }

  # calculate instantaneous effect
  dy.dx <- beta.hat[[marg_var]] + beta.hat[[int.name]] * z

  # calculate standard errors for instantaeous effect
  se.dy.dx <- sqrt(cov[marg_var, marg_var] +
                     z^2 * cov[int.name, int.name] +
                     2 * z * cov[marg_var, int.name])

  # combine into data frame
  data_frame(z = z,
             dy.dx = dy.dx,
             se = se.dy.dx)
}

# Plot point range plot conditional on education

instant_effect(lm_biden, "educ") %>%
  ggplot(aes(z, dy.dx,
             ymin = dy.dx - 1.96 * se,
             ymax = dy.dx + 1.96 * se)) +
```

```r
  geom_pointrange() +
  geom_hline(yintercept = 0, linetype = 2) +
  labs(title = "Marginal effect of age",
       subtitle = "By education",
       x = "Education",
       y = "Estimated marginal effect")
```




```r
# Plot line plot conditional on education

instant_effect(lm_biden, "educ") %>%
  ggplot(aes(z, dy.dx)) +
  geom_line() +
  geom_line(aes(y = dy.dx - 1.96 * se), linetype = 2) +
  geom_line(aes(y = dy.dx + 1.96 * se), linetype = 2) +
  geom_hline(yintercept = 0) +
  labs(title = "Marginal effect of age",
       subtitle = "By education",
       x = "Education",
       y = "Estimated marginal effect")
```



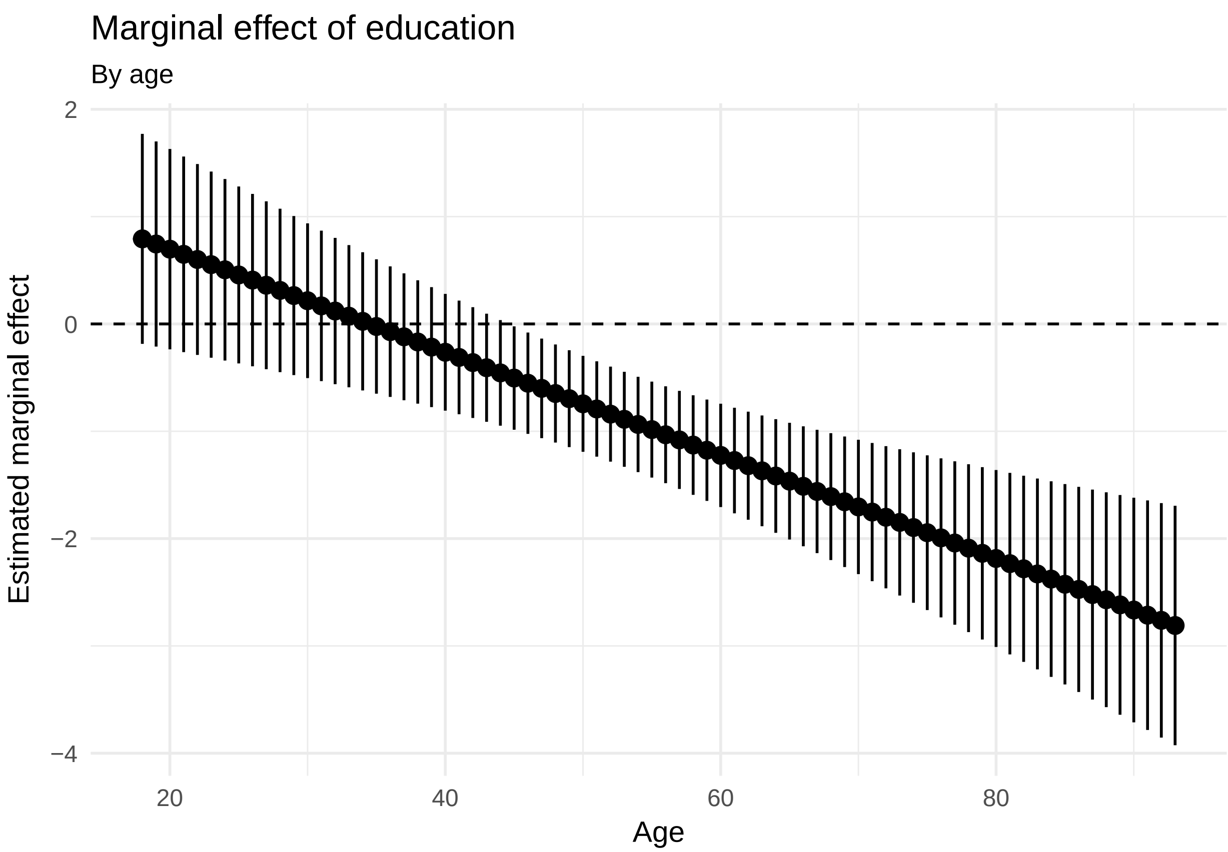

```r
# Test statistical significance conditional on education

linearHypothesis(lm_biden, "age + age:educ")
```

```
## Linear hypothesis test
##
## Hypothesis:
## age  + age:educ = 0
##
## Model 1: restricted model
## Model 2: biden ~ age + educ + age * educ
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1   1804 985149
## 2   1803 976688  1      8461 15.6  8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the above plots and linear hypothesis testing, we can conclude that there is a significant marginal effect of age on biden rating, conditional on education. The marginal effects is positive for individuals over 14 years of education and negative for individuals under 14 years of education For individuals who have 13 - 16 years of education roughly, we can't confidently say that the marginal effect differs from zero.

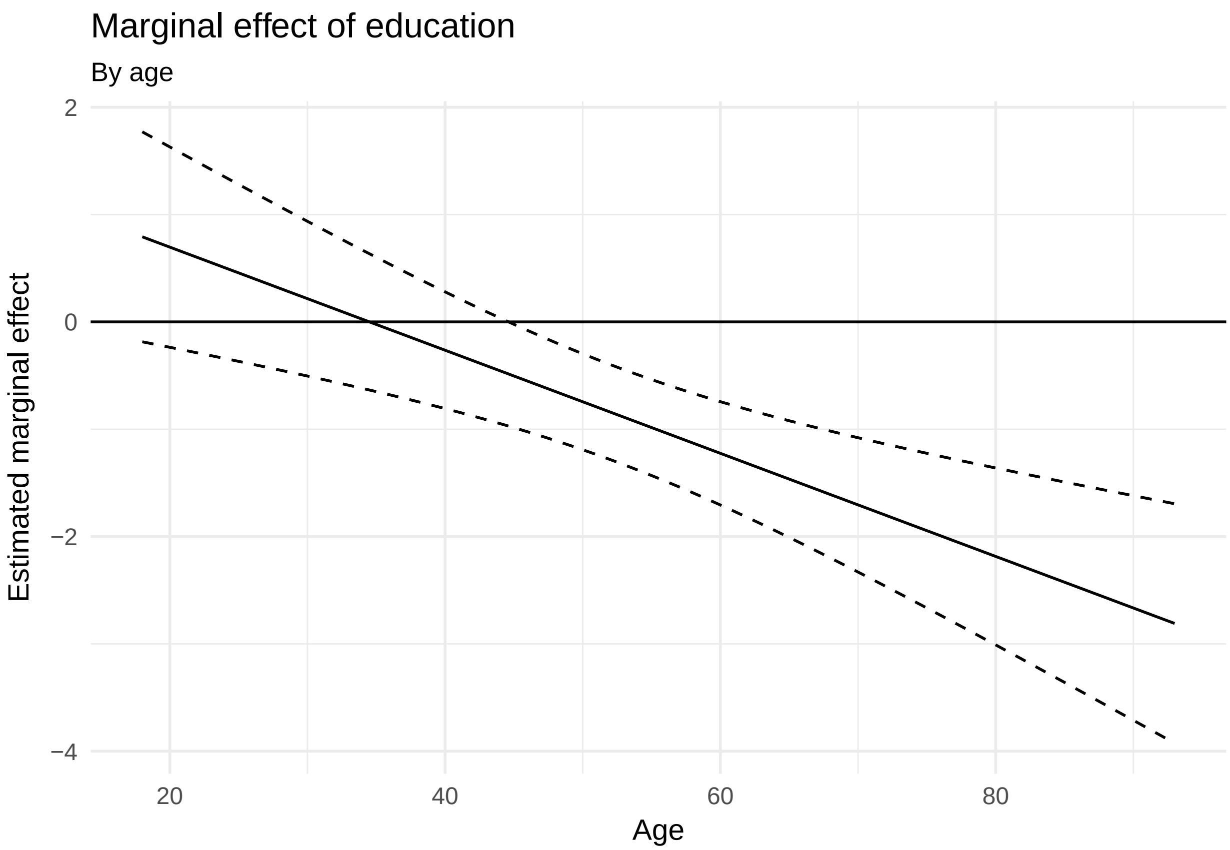
2B: Evaluate the marginal effect of education on Joe Biden thermometer rating, conditional on age

```
# marginal effect

# Plot point range plot conditional on age

instant_effect(lm_biden, "age") %>%
  ggplot(aes(z, dy.dx,
             ymin = dy.dx - 1.96 * se,
             ymax = dy.dx + 1.96 * se)) +
  geom_pointrange() +
  geom_hline(yintercept = 0, linetype = 2) +
  labs(title = "Marginal effect of education",
       subtitle = "By age",
       x = "Age",
       y = "Estimated marginal effect")
```




```
# Plot line plot conditional on age

instant_effect(lm_biden, "age") %>%
  ggplot(aes(z, dy.dx)) +
  geom_line() +
  geom_line(aes(y = dy.dx - 1.96 * se), linetype = 2) +
  geom_line(aes(y = dy.dx + 1.96 * se), linetype = 2) +
  geom_hline(yintercept = 0) +
  labs(title = "Marginal effect of education",
```

```
        subtitle = "By age",
        x = "Age",
        y = "Estimated marginal effect")
```




```
# Test statistical significance conditional on education

linearHypothesis(lm_biden, "educ + age:educ")
```

```
## Linear hypothesis test
##
## Hypothesis:
## educ  + age:educ = 0
##
## Model 1: restricted model
## Model 2: biden ~ age + educ + age * educ
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1   1804 979537
## 2   1803 976688  1      2849 5.26  0.022 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the above plots and linear hypothesis testing, we can conclude that there is a significant marginal effect of education on biden rating, conditional on age. However, the marginal effect is only significant for individuals who are over 45 years of age; for these individuals, the marginal effect of education on biden rating is negative.