Problem Set #2 Question 2

Xi Chen

April 29, 2018

2. Interaction terms

Estimate the following linear regression model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$ where Y is the Joe Biden feeling thermometer, X_1 is age, and X_2 is education. Report the parameters and standard errors.

```
library(tidyverse)
library(forcats)
library(broom)
library(modelr)
library(stringr)
library(titanic)
library(coefplot)
library(car)
library(plotly)
library(haven)

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

```
# Load biden data
data <- read_csv("data/biden.csv") %>%
    mutate_each(funs(as.factor(.)), female) %>%
    na.omit
attach(data)

m0 <- lm(biden ~ age + educ + age*educ)
tidy(m0)
```

```
##          term   estimate  std.error statistic      p.value
## 1 (Intercept) 38.3735103 9.56356681  4.012468 6.254443e-05
## 2         age  0.6718750 0.17049152  3.940812 8.430505e-05
## 3        educ  1.6574253 0.71399213  2.321350 2.037897e-02
## 4    age:educ -0.0480341 0.01290186 -3.723037 2.028851e-04
```

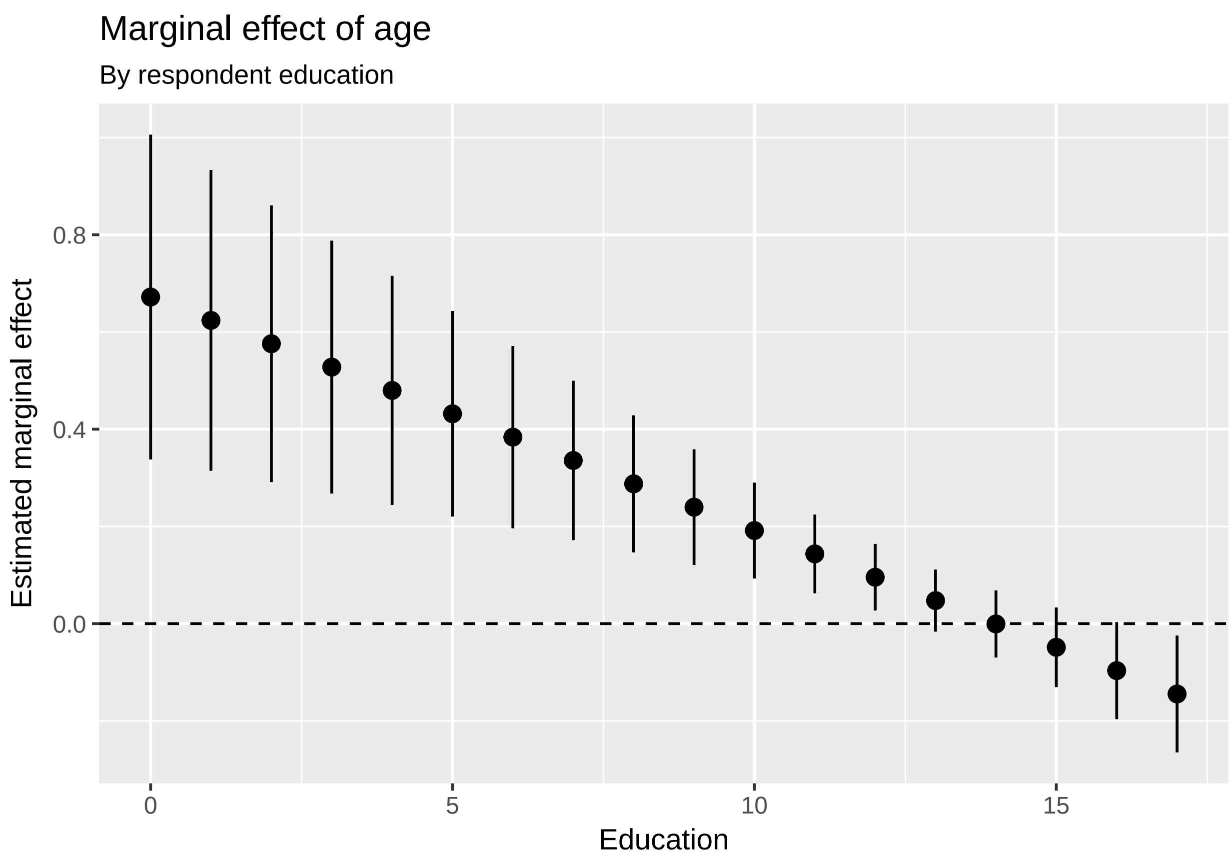
Comments:

The estimate of parameter β_0 is 38.3735 with standard error of 9.5636, β_1 is 0.6719 with standard error of 0.1705, β_2 is 1.6574 with standard error of 0.7140, β_3 is -0.0480 with standard error of 0.0129.

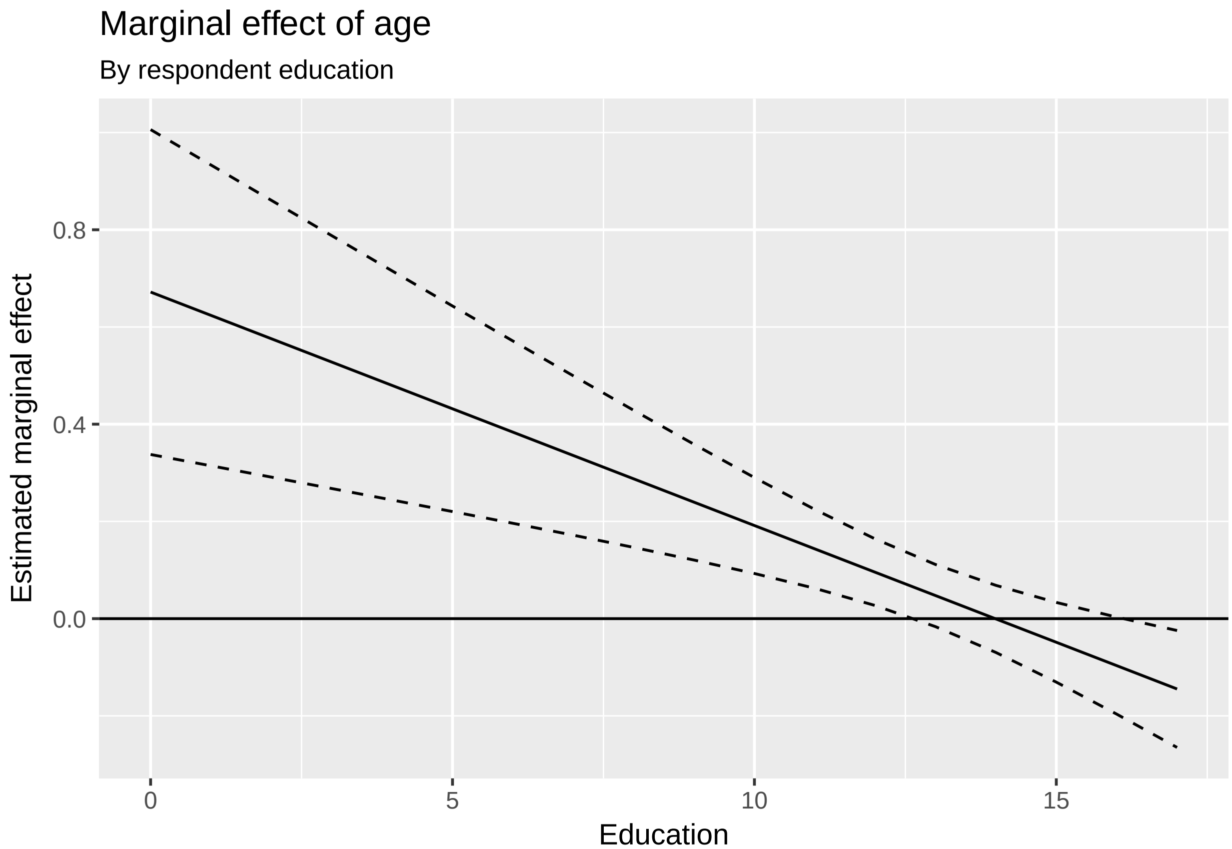
(a) Evaluate the marginal effect of age on Joe Biden thermometer rating, conditional on education. Consider the magnitude and direction of the marginal effect, as well as its statistical significance.

```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)
}
```

```r
# point range plot
instant_effect(m0, "educ") %>%
  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 age",
       subtitle = "By respondent education",
       x = "Education",
       y = "Estimated marginal effect")
```




```r
# line plot
instant_effect(m0, "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 respondent education",
       x = "Education",
       y = "Estimated marginal effect")
```




```
# Hypothesis testing
linearHypothesis(m0, "age + age:educ")
```

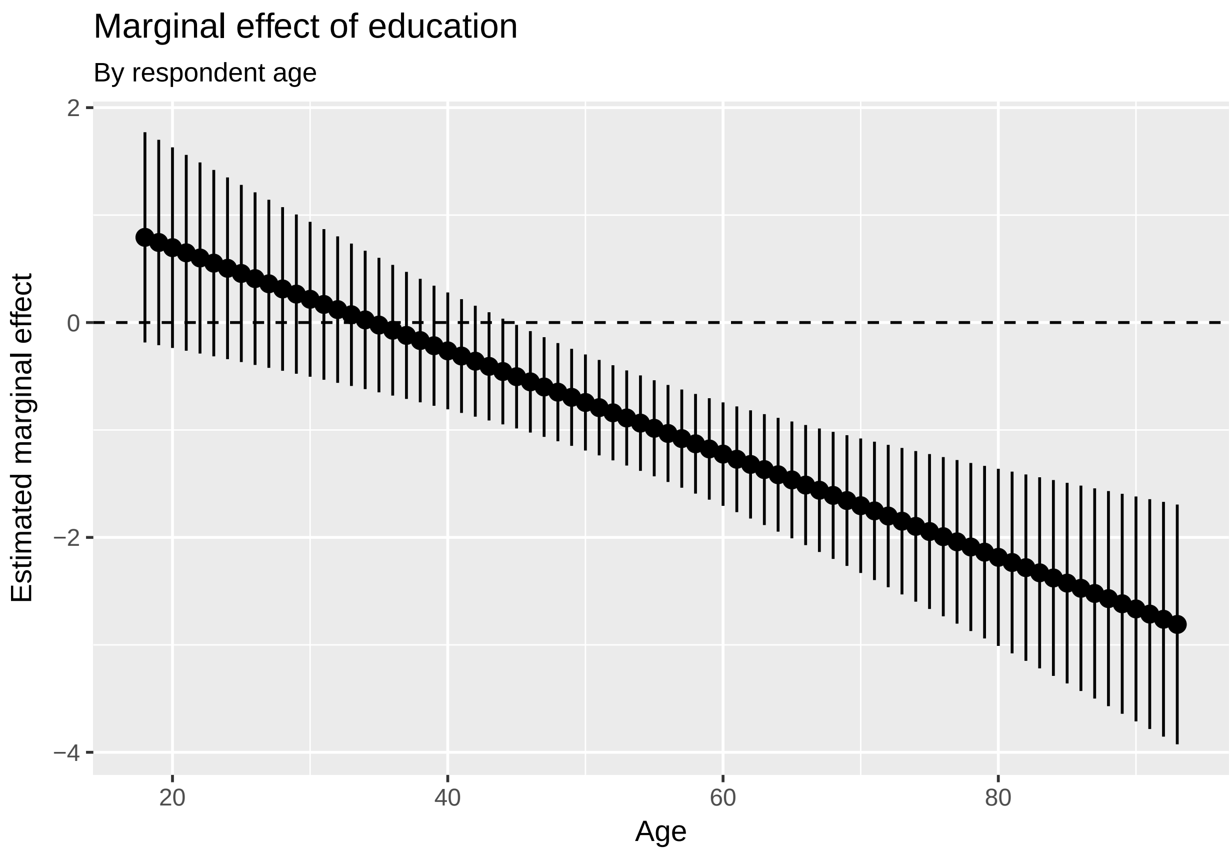
```
## Linear hypothesis test
##
## Hypothesis:
## age  + age:educ = 0
##
## Model 1: restricted model
## Model 2: biden ~ age + educ + age * educ
##
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1   1804 985149
## 2   1803 976688  1    8461.2 15.62 8.043e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comments:

The magnitude and direction of the marginal effect can be observed from the plots; the hypothesis testing is to evaluate its statistical significance. The above results show that the marginal effect of age on Joe Biden thermometer rating, conditional on education, whose direction is nagative (age decreases as education increases), is statistically significant (p-value < 0.001).

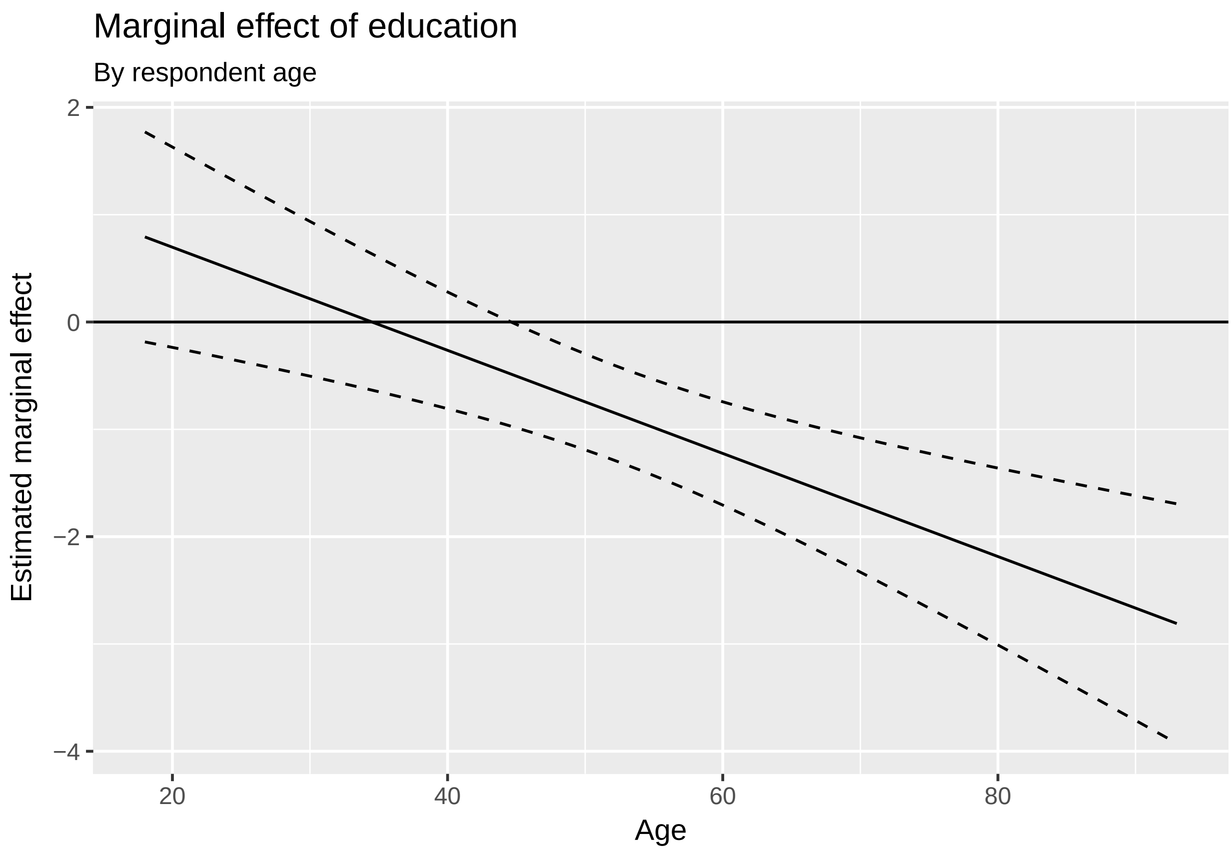
(b) Evaluate the marginal effect of education on Joe Biden thermometer rating, conditional on age. Consider the magnitude and direction of the marginal effect, as well as its statistical significance.

```
# point range plot
instant_effect(m0, "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 respondent age",
       x = "Age",
       y = "Estimated marginal effect")
```




```
# line plot
instant_effect(m0, "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 respondent age",
       x = "Age",
```

```r
       y = "Estimated marginal effect")
```




```r
# Hypothesis testing
linearHypothesis(m0, "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    2849.1 5.2595 0.02194 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comments:

The magnitude and direction of the marginal effect can be observed from the plots; the hypothesis testing is to evaluate its statistical significance. The above results show that the marginal effect of education on Joe Biden thermometer rating, conditional on age, whose direction is negative (education decreases as age increases), is statistically significant at the 0.05 level (p-value < 0.05).