R Coding Demonstration Week 7: Interactions and Nonlinearities in the Transphobia Experiment (Tidy)

Matthew Blackwell

Gov 51 (Harvard)

Introduction

- Today we're going to cover some tools for exploring bivariate relationships.
- We'll use the data from the Broockman & Kalla (2016) transphobia study.
- Basic summary of experiment:
    - Randomly assigned door-to-door canvassers to two conditions
    - Conditions: perspective-taking script (treatment) or recycling script (placebo)
    - Follow up surveys at 3 days, 3 weeks, 6 weeks, and 3 months.

Data

```
library(tidyverse)
phobia <- read.csv("data/transphobia_all.csv")
```

Variable Name	Description
age	Age of the respondent in years
female	1=respondent marked “Female” on voter registration, 0 otherwise
voted_gen_14	1 if respondent voted in the 2014 general election
voted_gen_12	1 if respondent voted in the 2012 general election
treat_ind	1 if respondent was assigned to treatment, 0 for control
racename	character name of racial identity indicated on voter file
democrat	1 if respondent is a registered Democrat
therm_trans_t0	0-100 feeling therm. about transgender people at baseline
therm_trans_tX	0-100 feeling therm. about transgender people in Wave X after treatment
therm_obama_t0	0-100 feeling therm. about Barack Obama at baseline
therm_obama_tX	0-100 feeling therm. about Barack Obama in Wave X after treatment

Question 1

Run a regression of thermometer scores for transgender people in wave 1 on the treatment indicator (`treat_ind`), the indicator for if the respondent is a Democrat (`democrat`), and the interaction between the two variables.

Interpret each of the coefficients in terms of the effects of the intervention.

Answer 1

```
int_dem_fit <- lm(therm_trans_t1 ~ treat_ind * democrat, data = phobia)
int_dem_fit
```

```
##
## Call:
## lm(formula = therm_trans_t1 ~ treat_ind * democrat, data = phobia)
##
## Coefficients:
##       (Intercept)          treat_ind           democrat  treat_ind:democrat
##             52.48               5.69               3.45                1.75
```

```
## effect of treatment for Republicans
coef(int_dem_fit)["treat_ind"]
```

```
## treat_ind
##      5.69
```

```
## effect of treatment for Democrats
coef(int_dem_fit)["treat_ind"] + coef(int_dem_fit)["treat_ind:democrat"]
```

```
## treat_ind
##      7.44
```

Question 2

Run a regression of thermometer scores for transgender people in wave 1 on the treatment indicator (`treat_ind`), the indicator for if the respondent is a woman (`female`), and the interaction between the two variables.

Interpret each of the coefficients in terms of the effects of the intervention. If you have time, compare the estimated effects here to the estimated difference in means of `therm_trans_t1` between treated and control within levels of `female`.

Answer 2

```
int_fem_fit <- lm(therm_trans_t1 ~ treat_ind * female, data = phobia)
int_fem_fit
```

```
##
## Call:
## lm(formula = therm_trans_t1 ~ treat_ind * female, data = phobia)
##
## Coefficients:
##     (Intercept)        treat_ind           female  treat_ind:female
##           52.14             1.74             3.20             8.71
```

```
## effect of treatment for Republicans
coef(int_fem_fit)["treat_ind"]
```

```
## treat_ind
##      1.74
```

```
## effect of treatment for Democrats
coef(int_fem_fit)["treat_ind"] + coef(int_fem_fit)["treat_ind:female"]
```

```
## treat_ind
##      10.4
```

Answer 2 (cont'd)

```
phobia %>%
  group_by(female, treat_ind) %>%
  summarize(across(therm_trans_t1, mean, na.rm = TRUE)) %>%
  pivot_wider(names_from = treat_ind, values_from = therm_tran
  mutate(diff_in_means = `1` - `0`)
```

```
## # A tibble: 2 x 4
## # Groups:   female [2]
##   female   `0`   `1` diff_in_means
##    <int> <dbl> <dbl>          <dbl>
## 1      0  52.1  53.9           1.74
## 2      1  55.3  65.8          10.4
```

Question 3

Run a regression of thermometer scores for transgender people in wave 1 on the treatment indicator (`treat_ind`), age (`age`), and the interaction between the two variables.

What is the estimated effect for a 25 year old? For a 50 year old?

Answer 3

```
int_age_fit <- lm(therm_trans_t1 ~ treat_ind * age, data = phobia)
int_age_fit
```

```
##
## Call:
## lm(formula = therm_trans_t1 ~ treat_ind * age, data = phobia)
##
## Coefficients:
##   (Intercept)      treat_ind            age  treat_ind:age
##      67.85075        6.34083       -0.28738        0.00915
```

Answer 3 (cont'd)

```
pred_data <- data.frame(treat_ind = c(1, 0, 1, 0),
                        age = c(25, 25, 50, 50))

## Respondents who rate Obama at 0
pred_data$preds <- predict(int_age_fit, newdata = pred_data)

pred_data %>%
  mutate(treat_ind = ifelse(treat_ind == 1, "Treated", "Control")) %>%
  pivot_wider(names_from = treat_ind, values_from = preds) %>%
  mutate(effects = Treated - Control)
```

```
## # A tibble: 2 x 4
##     age Treated Control effects
##   <dbl>   <dbl>   <dbl>   <dbl>
## 1    25    67.2    60.7    6.57
## 2    50    60.3    53.5    6.80
```

Question 4

Run a regression of thermometer scores for transgender people in wave 1 on the treatment indicator (`treat_ind`), Obama thermometer scores at baseline (`therm_obama_t0`), and the interaction between the two variables.

What is the estimated effect of the intervention for someone who rated Obama at 0? For someone who rated Obama at 100?

Answer 4

```
int_obama_fit <- lm(therm_trans_t1 ~ treat_ind * therm_obama_t0,
                    data = phobia)
int_obama_fit
```

```
##
## Call:
## lm(formula = therm_trans_t1 ~ treat_ind * therm_obama_t0, data = phobia)
##
## Coefficients:
##           (Intercept)             treat_ind         therm_obama_t0
##                46.214                -6.465                 0.133
## treat_ind:therm_obama_t0
##                   0.181
```

Answer 4 (cont'd)

```
pred_data <- data.frame(treat_ind = c(1, 0, 1, 0),
                        therm_obama_t0 = c(0, 0, 100, 100))

## Respondents who rate Obama at 0
pred_data$preds <- predict(int_obama_fit, newdata = pred_data)

pred_data %>%
  mutate(treat_ind = ifelse(treat_ind == 1, "Treated", "Control")) %>%
  pivot_wider(names_from = treat_ind, values_from = preds) %>%
  mutate(effects = Treated - Control)
```

```
## # A tibble: 2 x 4
##   therm_obama_t0 Treated Control effects
##            <dbl>   <dbl>   <dbl>   <dbl>
## 1              0    39.7    46.2   -6.47
## 2            100    71.2    59.5   11.7
```

Question 5

Run a regression of baseline transgender thermometer scores (`therm_trans_t0`) on age and the square of age to assess the nonlinear relationship between them.

Calculate predicted values from the model for ages 18 to 90 and plot these as a line.

Answer 5

```
agesq_fit <- lm(therm_trans_t0 ~ age + I(age ^ 2), data = phobia)
agesq_fit
```

```
##
## Call:
## lm(formula = therm_trans_t0 ~ age + I(age^2), data = phobia)
##
## Coefficients:
## (Intercept)          age     I(age^2)
##    66.80112     -0.39864      0.00206
```

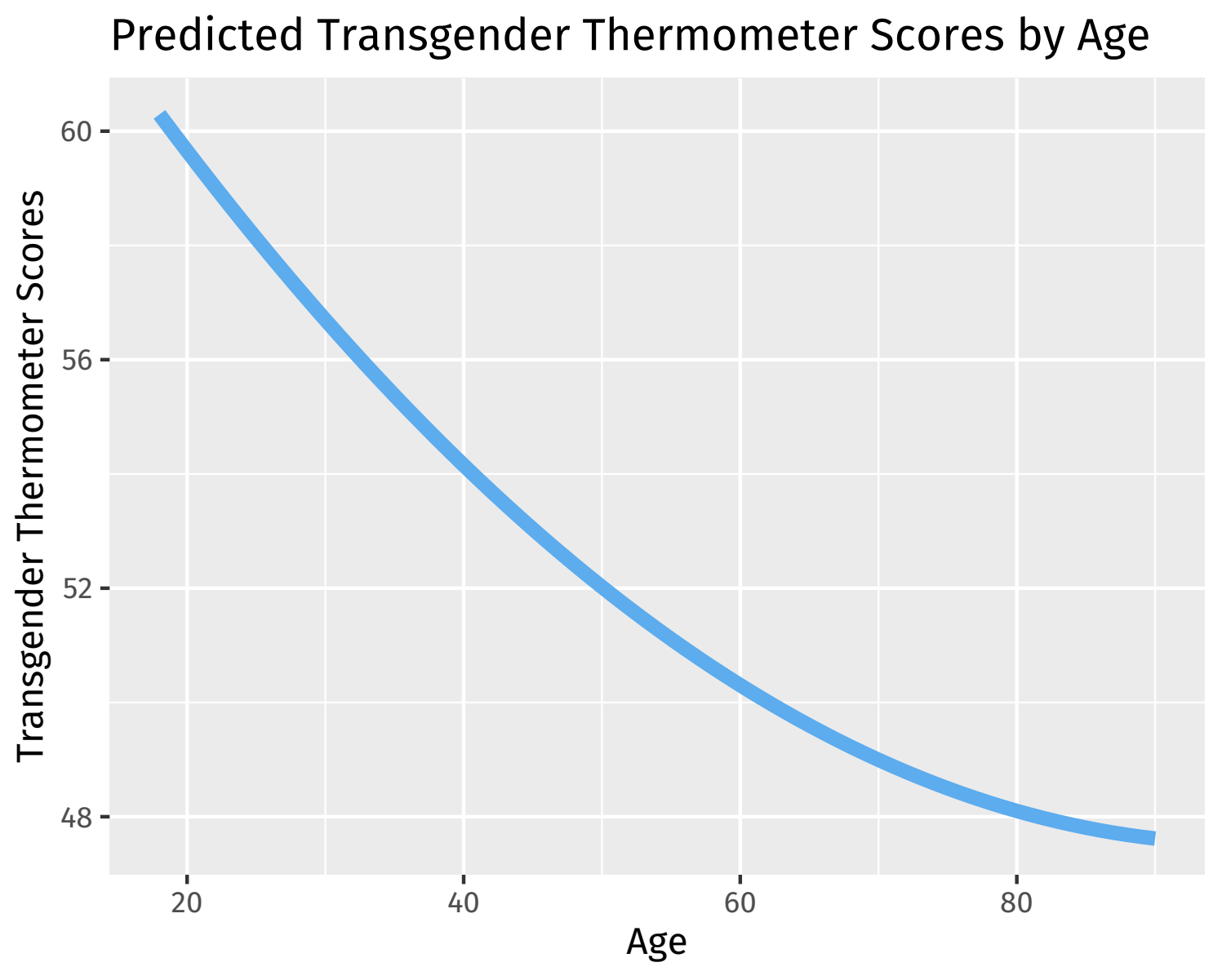
```
pred_data <- tibble(age = 18:90)

pred_data$pred_therm <- predict(agesq_fit, newdata = pred_data)
head(pred_data, n = 3)
```

```
## # A tibble: 3 x 2
##     age pred_therm
##   <int>      <dbl>
## 1    18       60.3
## 2    19       60.0
## 3    20       59.7
```

Answer 5 (cont'd)

```
ggplot(pred_data, aes(x = age, y = pred_therm)) +
  geom_line(size = 2, color = "steelblue2") +
  labs(x = "Age", y = "Transgender Thermometer Scores",
       title = "Predicted Transgender Thermometer Scores by Age")
```

Answer 5 (cont'd)



Question 6

Run a regression of baseline Obama thermometer scores (`therm_obama_t0`) on age and the square of age to assess the nonlinear relationship between them.

Calculate predicted values from the model for ages 18 to 90 and plot these as a line.

Answer 6

```
agesq_obama_fit <- lm(therm_obama_t0 ~ age + I(age ^ 2), data = phobia)
agesq_obama_fit
```

```
##
## Call:
## lm(formula = therm_obama_t0 ~ age + I(age^2), data = phobia)
##
## Coefficients:
## (Intercept)          age     I(age^2)
##    82.04797     -0.66011      0.00464
```

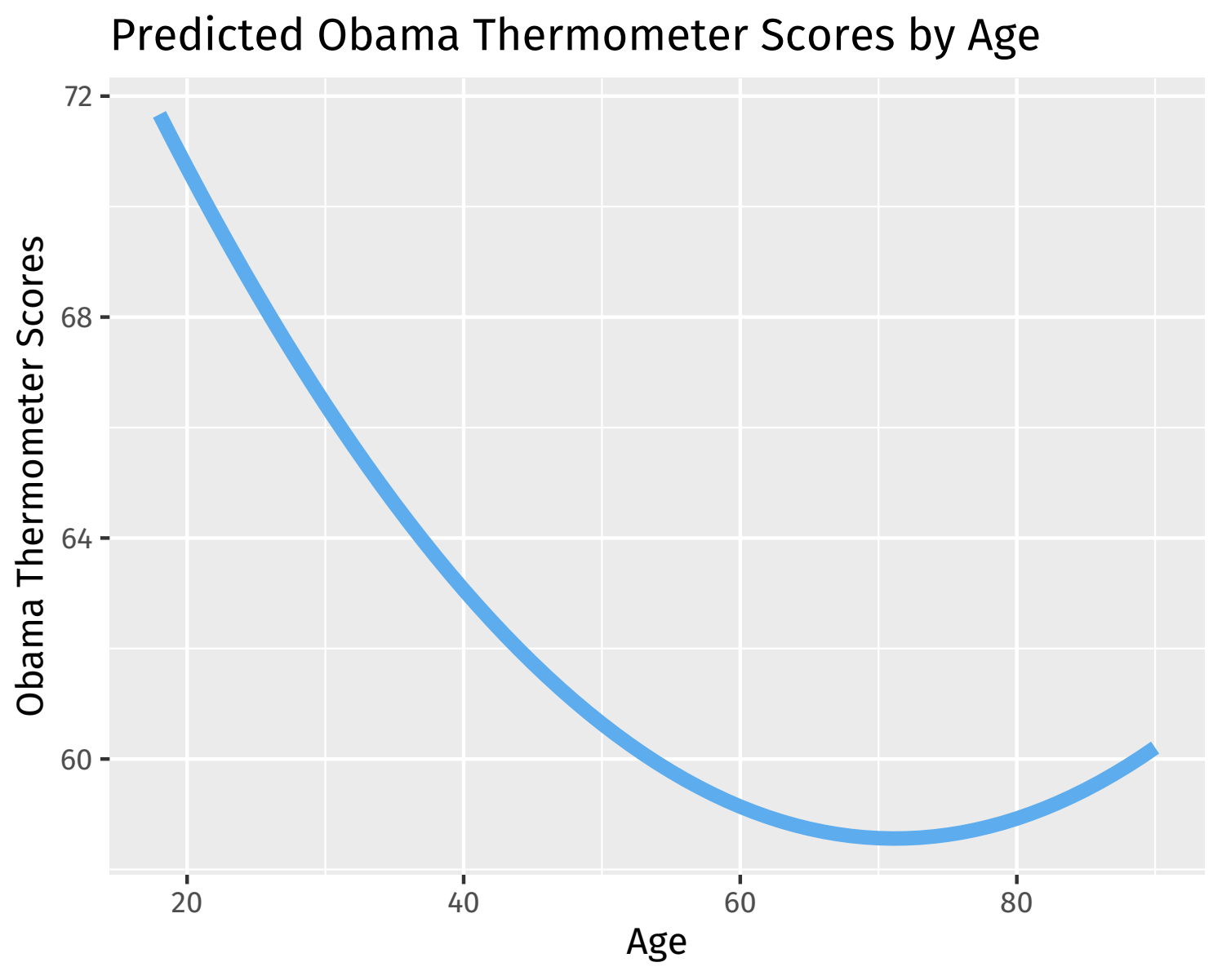
```
pred_data$pred_obama_therm <- predict(agesq_obama_fit, newdata = pred_data)
head(pred_data, n = 3)
```

```
## # A tibble: 3 x 3
##     age pred_therm pred_obama_therm
##   <int>      <dbl>            <dbl>
## 1    18       60.3             71.7
## 2    19       60.0             71.2
## 3    20       59.7             70.7
```

Answer 6 (cont'd)

```
ggplot(pred_data, aes(x = age, y = pred_obama_therm)) +
  geom_line(size = 2, color = "steelblue2") +
  labs(x = "Age", y = "Obama Thermometer Scores",
       title = "Predicted Obama Thermometer Scores by Age")
```

Answer 6


Predicted Obama Thermometer Scores by Age
Obama Thermometer Scores
72
68
64
60
20
40
60
80
Age

Question 7 (challenge question)

Run a regression of wave 1 transgender thermometer scores on the following: treatment indicator, age, age squared, the interaction between treatment and age, and the interaction between treatment and age squared. Create a plot of the predicted curves of the treated and control groups as a function of age.

Answer 7

```
nonlin_int_fit <- lm(therm_trans_t1 ~ treat_ind * age + treat_ind * I(age ^ 2),
                     data = phobia)

tr_pred_data <- pred_data %>% mutate(treat_ind = 1)
ct_pred_data <- pred_data %>% mutate(treat_ind = 0)
tr_pred_data$pred_therm <- predict(nonlin_int_fit, newdata = tr_pred_data)
ct_pred_data$pred_therm <- predict(nonlin_int_fit, newdata = ct_pred_data)
all_preds <- bind_rows(tr_pred_data, ct_pred_data) %>%
  mutate(treat_ind = ifelse(treat_ind == 1, "Treated", "Control"))

ggplot(all_preds, aes(x = age, y = pred_therm, group = treat_ind)) +
  geom_line(aes(color = treat_ind), size = 2) +
  ylim(c(20, 80)) +
  labs(x = "Age", y = "Transgender Thermometer Scores")
```

Answer 7 (cont'd)