Gov 51: Interactions with Continuous Variables

Matthew Blackwell

Harvard University

Social pressure experiment

- We'll look at the Michigan social pressure get-out-the-vote experiment.
- Load the data and create an age variable:

```
social <- read.csv("data/social.csv")
social$age <- 2006 - social$yearofbirth
summary(social$age)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     20.0    41.0    50.0    49.8    59.0   106.0
```

```
social.neighbors <- subset(social,
                           neighbors == 1 | control == 1)
```

Heterogeneous effects

- Last time:
  - Effect of the Neighbors mailer differ for previous voters vs nonvoters?
  - Used an interaction term to assess **effect heterogeneity** between groups.
- How does the effect of the Neighbors mailer varies by age?
  - Not just two groups, but a continuum of possible age values.
- Remarkably, the same **interaction term** will work here too!

$$Y_i = \alpha + \beta_1 \text{age}_i + \beta_2 \text{neighbors}_i + \beta_3 (\text{age}_i \times \text{neighbors}_i) + \epsilon_i$$

Predicted values from non-interacted model

- Let X_i = age$_i$ and Z_i = neighbors$_i$:

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}_1 X_i + \widehat{\beta}_2 Z_i$$

	Control ($Z_i = 0$)	Neighbors ($Z_i = 1$)
25 year-old ($X_i = 25$)	$\widehat{\alpha} + \widehat{\beta}_1 25$	$\widehat{\alpha} + \widehat{\beta}_1 25 + \widehat{\beta}_2$
26 year-old ($X_i = 26$)	$\widehat{\alpha} + \widehat{\beta}_1 26$	$\widehat{\alpha} + \widehat{\beta}_1 26 + \widehat{\beta}_2$

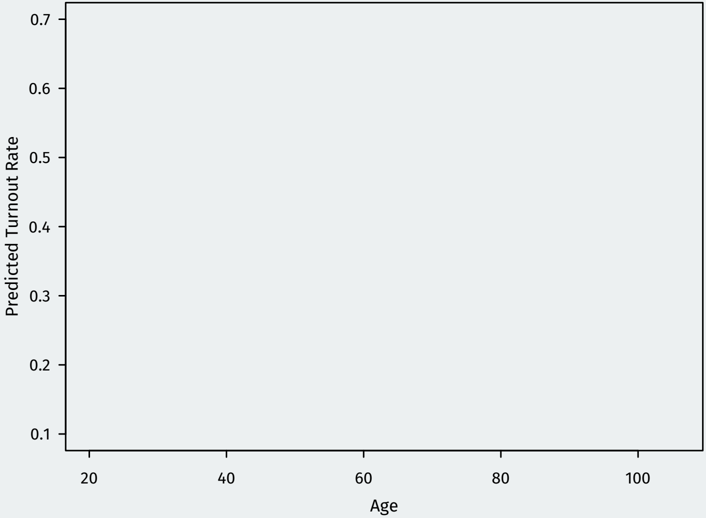
- Effect of Neighbors for a 25 year-old:

$$(\widehat{\alpha} + \widehat{\beta}_1 25 + \widehat{\beta}_2) - (\widehat{\alpha} + \widehat{\beta}_1 25) = \widehat{\beta}_2$$

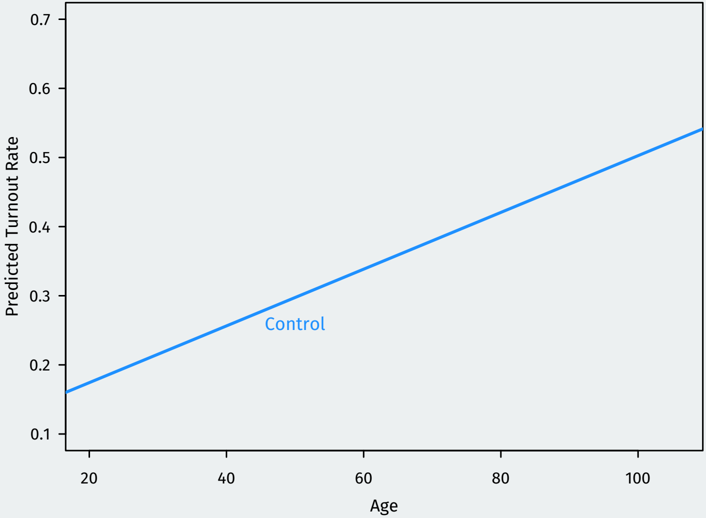
- Effect of Neighbors for a 26 year-old:

$$(\widehat{\alpha} + \widehat{\beta}_1 26 + \widehat{\beta}_2) - (\widehat{\alpha} + \widehat{\beta}_1 26) = \widehat{\beta}_2$$

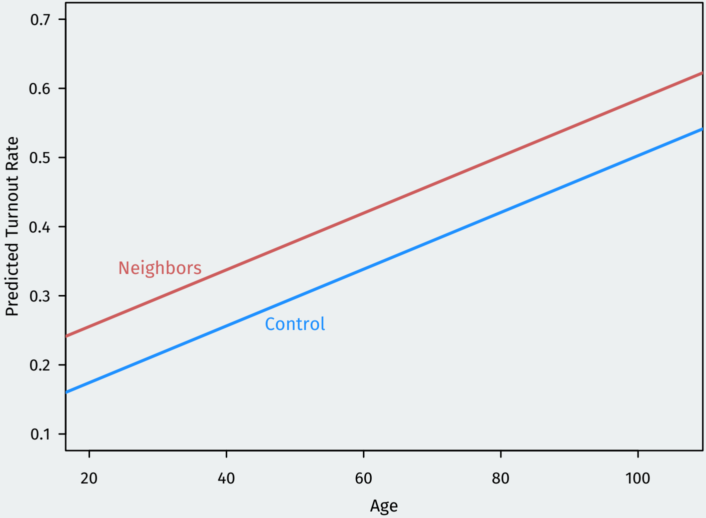
Visualizing the regression



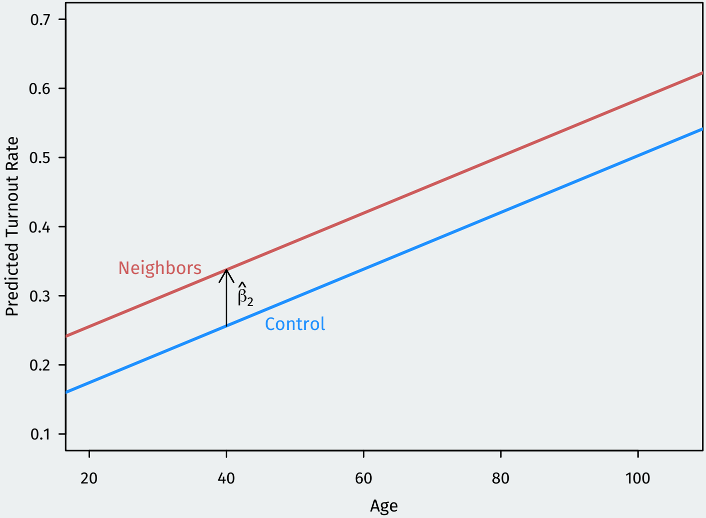
Visualizing the regression



Visualizing the regression



Visualizing the regression



Predicted values from interacted model

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}_1 X_i + \widehat{\beta}_2 Z_i + \widehat{\beta}_3 X_i Z_i$$

	Control ($Z_i = 0$)	Neighbors ($Z_i = 1$)
25 year-old ($X_i = 25$)	$\widehat{\alpha} + \widehat{\beta}_1 25$	$\widehat{\alpha} + \widehat{\beta}_1 25 + \widehat{\beta}_2 + \widehat{\beta}_3 25$
26 year-old ($X_i = 26$)	$\widehat{\alpha} + \widehat{\beta}_1 26$	$\widehat{\alpha} + \widehat{\beta}_1 26 + \widehat{\beta}_2 + \widehat{\beta}_3 26$

- Effect of Neighbors for a 25 year-old:

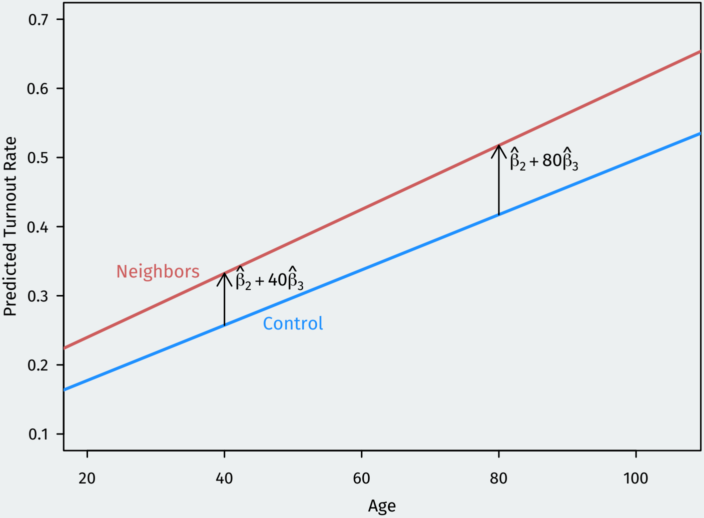
$$(\widehat{\alpha} + \widehat{\beta}_1 \times 25 + \widehat{\beta}_2 + \widehat{\beta}_3 25) - (\widehat{\alpha} + \widehat{\beta}_1 25) = \widehat{\beta}_2 + \widehat{\beta}_3 25$$

- Effect of Neighbors for a 26 year-old:

$$(\widehat{\alpha} + \widehat{\beta}_1 26 + \widehat{\beta}_2 + \widehat{\beta}_3 26) - (\widehat{\alpha} + \widehat{\beta}_1 26) = \widehat{\beta}_2 + \widehat{\beta}_3 26$$

- Effect of Neighbors for a x year-old: $\widehat{\beta}_2 + \widehat{\beta}_3 x$

Visualizing the interaction



Interpreting coefficients

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}_1 \text{age}_i + \widehat{\beta}_2 \text{neighbors}_i + \widehat{\beta}_3 (\text{age}_i \times \text{neighbors}_i)$$

- $\widehat{\alpha}$: average turnout for 0 year-olds in the control group.
- $\widehat{\beta}_1$: slope of regression line for age in the control group.
- $\widehat{\beta}_2$: average effect of Neighbors mailer for 0 year-olds.
- $\widehat{\beta}_3$: change in the **effect** of the Neighbors mailer for a 1-year $\uparrow$ in age.
    - Effect for x year-olds: $\widehat{\beta}_2 + \widehat{\beta}_3 x$
    - Effect for $(x+1)$ year-olds: $\widehat{\beta}_2 + \widehat{\beta}_3 (x+1)$
    - Change in effect: $\widehat{\beta}_3$

Interactions in R

- You can use the `:` way to create interaction terms like last time:

```
int.fit <- lm(primary2006 ~ age + neighbors + age:neighbors,
              data = social.neighbors)
coef(int.fit)
```

```
##   (Intercept)           age     neighbors
##      0.097473      0.003998      0.049829
## age:neighbors
##      0.000628
```

- Or you can use the `var1 * var2` shortcut, which will add both variable and their interaction:

```
int.fit2 <- lm(primary2006 ~ age * neighbors, data = social.neighbors)
coef(int.fit2)
```

```
##   (Intercept)           age     neighbors
##      0.097473      0.003998      0.049829
## age:neighbors
##      0.000628
```

General interpretation of interactions

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}_1 X_i + \widehat{\beta}_2 Z_i + \widehat{\beta}_3 X_i Z_i$$

- $\widehat{\alpha}$: average outcome when X_i and Z_i are 0.
- $\widehat{\beta}_1$: average change in Y_i of a one-unit change in X_i when $Z_i = 0$
- $\widehat{\beta}_2$: average change in Y_i of a one-unit change in Z_i when $X_i = 0$
- $\widehat{\beta}_3$ has two equivalent interpretations:
    - Change in the effect/slope of X_i for a one-unit change in Z_i
    - Change in the effect/slope of Z_i for a one-unit change in X_i
- These hold no matter what types of variables they are!