Gov 51: Varying Effects by Groups

Matthew Blackwell

Harvard University

Heterogeneous treatment effects

- **Heterogeneous treatment effects**: effect varies across groups.
    - Average effect of a drug is 0, but + for men and − for women.
    - Important questions for determining who should receive treatment.
- Social pressure experiment:
    - `primary2004` whether the person voted in 2004, before the experiment.
    - Do 2004 voters react differently to social pressure mailer than nonvoters?
- Two approaches:
    - Subsets, subsets, subsets.
    - Interaction terms in regression.

Subset approach

- Easy way to estimate heterogeneous effects: our old friend, `subset()`.
- First, estimate the ATE for the voters:

```
social <- read.csv("data/social.csv")
voters.t <- subset(social, primary2004 == 1 & neighbors == 1)
voters.c <- subset(social, primary2004 == 1 & control == 1)
ate.v <- mean(voters.t$primary2006) - mean(voters.c$primary2006)
ate.v
```

```
## [1] 0.0965
```

- Now, estimate the ATE for the nonvoters:

```
nonvoters.t <- subset(social, primary2004 == 0 & neighbors == 1)
nonvoters.c <- subset(social, primary2004 == 0 & control == 1)
ate.nv <- mean(nonvoters.t$primary2006) - mean(nonvoters.c$primary2006)
ate.nv
```

```
## [1] 0.0693
```

Difference in effects

- How much does the estimated treatment effect differ between groups?

```
ate.v - ate.nv
```

```
## [1] 0.0272
```

- Any easier way to allow for different effects of treatment by groups?

Interaction terms

- Can allow for different effects of a variable with an **interaction term**:

$$\text{turnout}_i = \alpha + \beta_1 \text{primary2004}_i + \beta_2 \text{neighbors}_i + \beta_3 \left(\text{primary2004}_i \times \text{neighbors}_i\right) + \varepsilon_i$$

- Primary 2004 variable multiplied by the neighbors variable.
    - Equal to 1 if voted in 2004 (`primary2004 == 1`) and received neighbors mailer (`neighbors == 1`)
- Easiest to understand by investigating predicted values.

Predicted values from non-interacted model

- Let $X_i = \text{primary2004}_i$ and $Z_i = \text{neighbors}_i$:

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

	Control ($Z_i = 0$)	Neighbors ($Z_i = 1$)
non-voter ($X_i = 0$)	$\hat{\alpha} + \hat{\beta}_1 0 + \hat{\beta}_2 0 \hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_1 0 + \hat{\beta}_2 1 \hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2$

- Effect of Neighbors for non-voters: $(\hat{\alpha} + \hat{\beta}_2) - (\hat{\alpha}) = \hat{\beta}_2$
- Effect of Neighbors for voters: $(\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2) - (\hat{\alpha} + \hat{\beta}_1) = \hat{\beta}_2$

Predicted from interacted model

- Now for the interacted model:

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

	Control ($Z_i = 0$)	Neighbors ($Z_i = 1$)
non-voter ($X_i = 0$)	$\hat{\alpha} + \hat{\beta}_1 0 + \hat{\beta}_2 0 + \hat{\beta}_3 0 \cdot 0\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_1 0 + \hat{\beta}_2 1 + \hat{\beta}_3 0 \cdot$
voter($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$

- Effect of Neighbors for non-voters: $(\hat{\alpha} + \hat{\beta}_2) - \hat{\alpha} = \hat{\beta}_2$
- Effect of Neighbors for voters:
  $(\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3) - (\hat{\alpha} + \hat{\beta}_1) = \hat{\beta}_2 + \hat{\beta}_3$

Interpreting coefficients

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}_1 \text{primary2004}_i + \widehat{\beta}_2 \text{neighbors}_i + \widehat{\beta}_3 \left(\text{primary2004}_i \times \text{neighbors}_i \right)$$

	Control Group	Neighbors Group
2004 primary non-voter	$\widehat{\alpha}$	$\widehat{\alpha} + \widehat{\beta}_2$
2004 primary voter	$\widehat{\alpha} + \widehat{\beta}_1$	$\widehat{\alpha} + \widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_3$

- $\widehat{\alpha}$: turnout rate for 2004 nonvoters in control group.
- $\widehat{\beta}_1$: avg difference in turnout between 2004 voters and nonvoters.
- $\widehat{\beta}_2$: effect of neighbors for 2004 nonvoters.
- $\widehat{\beta}_3$: difference in the effect of neighbors mailer between 2004 voters & nonvoters.

Interactions in R

- You can include an interaction with `var1:var2`:

```
social.neighbor <- subset(social, neighbors == 1 | control == 1)
fit <- lm(primary2006 ~ primary2004 + neighbors +
            primary2004:neighbors, data = social.neighbor)
coef(fit)
```

```
##           (Intercept)             primary2004
##                0.2371                  0.1487
##             neighbors primary2004:neighbors
##                0.0693                  0.0272
```

- Compare coefficients to subset approach:

```
ate.nv
```

```
## [1] 0.0693
```

```
ate.v - ate.nv
```

```
## [1] 0.0272
```