PSC7475: Varying Effects by Group

Week 8: Lecture 13

Prof. Weldzius

Villanova University

Slides Updated: 2025-03-10

Week 8

Week 8

- QSS Tidyverse Tutorial 6 due tomorrow

Week 8

- QSS Tidyverse Tutorial 6 due tomorrow
- Proposal for final project due Wednesday by midnight (upload to Blackboard)

Week 8

- QSS Tidyverse Tutorial 6 due tomorrow
- Proposal for final project due Wednesday by midnight (upload to Blackboard)
  - What is your tentative research question?
  - What data is available to answer this question?
  - Why are you interested in this question?

Week 8

- QSS Tidyverse Tutorial 6 due tomorrow
- Proposal for final project due Wednesday by midnight (upload to Blackboard)
  - What is your tentative research question?
  - What data is available to answer this question?
  - Why are you interested in this question?
  - How long? Be brief! Use bullet points. But you must upload a document
- Midterm: will discuss next week

Week 8

- QSS Tidyverse Tutorial 6 due tomorrow
- Proposal for final project due Wednesday by midnight (upload to Blackboard)
  - What is your tentative research question?
  - What data is available to answer this question?
  - Why are you interested in this question?
  - How long? Be brief! Use bullet points. But you must upload a document
- Midterm: will discuss next week
- This week: finishing up regression!

Heterogeneous treatment effects

Heterogeneous treatment effects

- **Heterogeneous treatment effects**: effect varies across groups.

Heterogeneous treatment effects

- **Heterogeneous treatment effects**: effect varies across groups.
  - Average effect of a drug is 0, but $+$ for men and $-$ for women.

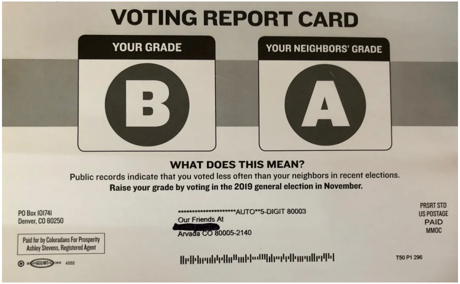
Heterogeneous treatment effects

- **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.

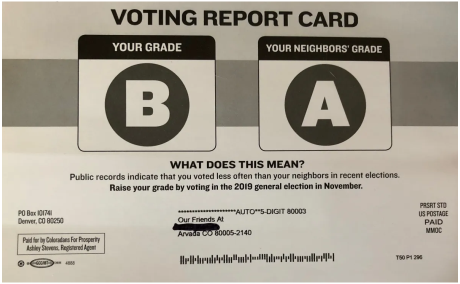
Heterogeneous treatment effects

- **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

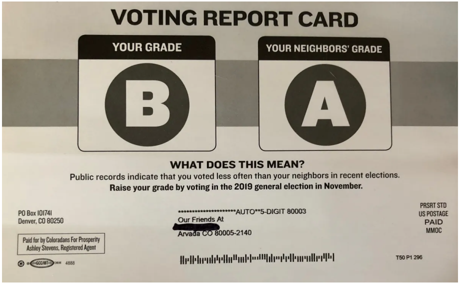


Social pressure experiment



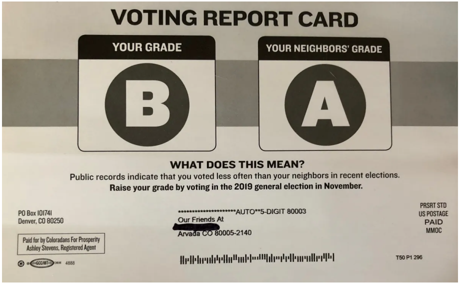
- `primary2004` whether the person voted in 2004, before the experiment.

Social pressure experiment



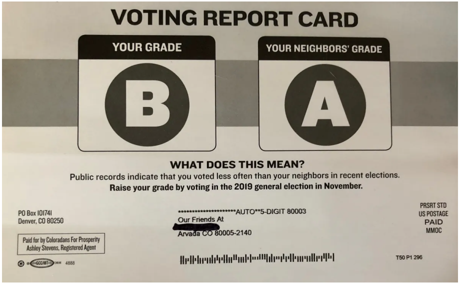
- `primary2004` whether the person voted in 2004, before the experiment.
- Do 2004 voters react differently to social pressure mailer than nonvoters?

Social pressure experiment



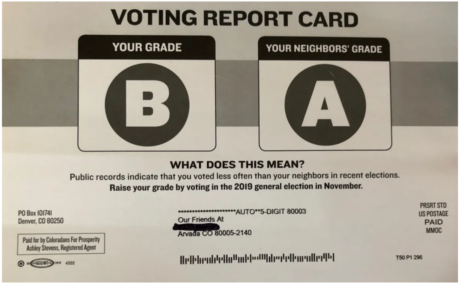
- `primary2004` whether the person voted in 2004, before the experiment.
- Do 2004 voters react differently to social pressure mailer than nonvoters?
- Two approaches:

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.

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

Subset approach

- Easy way to estimate heterogeneous effects: our old friend, `filter()`, `group_by()`, and `summarize()`. Woo!

Subset approach

- Easy way to estimate heterogeneous effects: our old friend, `filter()`, `group_by()`, and `summarize()`. Woo!
  - First, get the data

```
data(social, package="qss")
```

Subset approach

- Now, estimate the ATE for the **voters**:

Subset approach

- Now, estimate the ATE for the **voters**:

```
VotersATE <- social %>%
  filter(primary2004 == 1,
         messages %in% c("Control","Neighbors")) %>%
  group_by(messages) %>%
  summarize(primary2006_mean = mean(primary2006)) %>%
  pivot_wider(names_from = "messages",
              values_from = "primary2006_mean") %>%
  mutate(ate_v = Neighbors - Control) %>%
  select(ate_v)
VotersATE
```

```
## # A tibble: 1 x 1
##    ate_v
##    <dbl>
## 1 0.0965
```

Filter approach

- Now, estimate the ATE for the **nonvoters**:

Filter approach

- Now, estimate the ATE for the **nonvoters**:

```
NonvotersATE <- social %>%
  filter(primary2004 == 0,
         messages %in% c("Control","Neighbors")) %>%
  group_by(messages) %>%
  summarize(primary2006_mean = mean(primary2006)) %>%
  pivot_wider(names_from = "messages",
              values_from = "primary2006_mean") %>%
  mutate(ate_nv = Neighbors - Control) %>%
  select(ate_nv)
NonvotersATE
```

```
## # A tibble: 1 x 1
##   ate_nv
##    <dbl>
## 1 0.0693
```

Difference in effects

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

Difference in effects

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

```
VotersATE$ate_v - NonvotersATE$ate_nv
```

```
## [1] 0.02722908
```

Difference in effects

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

```
VotersATE$ate_v - NonvotersATE$ate_nv
```

```
## [1] 0.02722908
```

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

Interaction terms

Interaction terms

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

Interaction terms

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

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

Interaction terms

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

$$\begin{aligned} \text{turnout}_i = \alpha + \beta_1 \text{primary2004}_i + \beta_2 \text{neighbors}_i + \\ \beta_3(\text{primary2004}_i \times \text{neighbors}_i) + \varepsilon_i \end{aligned}$$

- Primary 2004 variable multiplied by the neighbors variable.
  - Equal to 1 if voted in 2004 (primary2004 $==1$) and received neighbors mailer (neighbors $==1$)

Interaction terms

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

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

- 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

Predicted values from non-interacted model

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

Predicted values from non-interacted model

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

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

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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)		

-
-

Predicted values from non-interacted model

- Let X_i = primary2004$_i$ and Z_i = 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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	

-
-

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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2$

-
-

Predicted values from non-interacted model

- Let X_i = primary2004$_i$ and Z_i = 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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2$

- Effect of Neighbors for non-voters:
-

Predicted values from non-interacted model

- Let X_i = primary2004$_i$ and Z_i = 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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2$

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

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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
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:

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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
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:

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

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 ($X_i = 0$)	$\hat{\alpha}$	
voter ($X_i = 1$)		

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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)		

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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	

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 ($X_i = 0$)	$\hat{\alpha}$	$\hat{\alpha} + \hat{\beta}_2$
voter ($X_i = 1$)	$\hat{\alpha} + \hat{\beta}_1$	$\hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$

Interpreting coefficients

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 \text{primary2004}_i + \hat{\beta}_2 \text{neighbors}_i$$

$$+\hat{\beta}_3(\text{primary2004}_i \times \text{neighbors}_i)$$

Interpreting coefficients

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 \text{primary2004}_i + \hat{\beta}_2 \text{neighbors}_i$$

$$+\hat{\beta}_3(\text{primary2004}_i \times \text{neighbors}_i)$$

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

Interpreting coefficients

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 \text{primary2004}_i + \hat{\beta}_2 \text{neighbors}_i$$

$$+\hat{\beta}_3(\text{primary2004}_i \times \text{neighbors}_i)$$

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

- $\hat{\alpha}$: turnout rate for 2004 nonvoters in control group.

Interpreting coefficients

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 \text{primary2004}_i + \hat{\beta}_2 \text{neighbors}_i$$

$$+\hat{\beta}_3(\text{primary2004}_i \times \text{neighbors}_i)$$

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

- $\hat{\alpha}$: turnout rate for 2004 nonvoters in control group.
- $\hat{\beta}_1$: avg difference in turnout between 2004 voters and nonvoters.

Interpreting coefficients

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 \text{primary2004}_i + \hat{\beta}_2 \text{neighbors}_i$$

$$+\hat{\beta}_3(\text{primary2004}_i \times \text{neighbors}_i)$$

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

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

Interpreting coefficients

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 \text{primary2004}_i + \hat{\beta}_2 \text{neighbors}_i$$

$$+\hat{\beta}_3(\text{primary2004}_i \times \text{neighbors}_i)$$

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

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

Interactions in R

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

```
social.neighbor <- social %>%
  mutate(neighbors = ifelse(messages=="Neighbors",1,
                            ifelse(messages=="Control",0,NA))
  select(primary2006,primary2004,neighbors) %>%
  drop_na()

fit <- lm(primary2006 ~ primary2004 + neighbors +
          primary2004:neighbors, data = social.neighbor)

coef(fit)
```

```
##           (Intercept)           primary2004
##            0.23710990            0.14869507
##             neighbors primary2004:neighbors
##            0.06929617            0.02722908
```

Interactions in R

```r
coef(fit)
```

```
##              (Intercept)             primary2004
##               0.23710990              0.14869507
##                neighbors primary2004:neighbors
##               0.06929617              0.02722908
```

Interactions in R

```r
coef(fit)
```

```
##          (Intercept)          primary2004
##           0.23710990           0.14869507
##            neighbors primary2004:neighbors
##           0.06929617           0.02722908
```

- Compare coefficients to earlier approach:

```r
NonvotersATE$ate_nv
```

```
## [1] 0.06929617
```

Interactions in R

```r
coef(fit)
```

```
##             (Intercept)            primary2004
##              0.23710990             0.14869507
##               neighbors primary2004:neighbors
##              0.06929617             0.02722908
```

- Compare coefficients to earlier approach:

```r
NonvotersATE$ate_nv
```

```
## [1] 0.06929617
```

```r
VotersATE$ate_v - NonvotersATE$ate_nv
```

```
## [1] 0.02722908
```