

Causal Modeling With GSS Data Using Multiple Approaches

Andy Grogan-Kaylor

4 Jul 2020 14:00:12

Setup

```
. clear all

. cd "/Users/agrogan/Desktop/newstuff/causal-modeling"
/Users/agrogan/Desktop/newstuff/causal-modeling
```

Get Data

```
. use "/Users/agrogan/Box Sync/DATA WAREHOUSE/General Social Survey Panel Data/GSS_panel2
> 010w123_R6 - stata.dta", clear
( )
```

ID Variable

```
. generate ID = id_1
```

Keep Only Relevant Variables

```
. keep ID satjob_? educ_? race_? incom16_?
```

Describe Data

```
. describe
Contains data from /Users/agrogan/Box Sync/DATA WAREHOUSE/General Social Survey Panel Dat
> a/GSS_panel2010w123_R6 - stata.dta
obs:                2,044
vars:                 13                      12 MAR 2018 16:24
size:                32,704
```

| variable name | storage type | display format | value label | variable label |
|---------------|-----------------|-------------------|----------------|---|
| educ_1 | byte | %8.0g | EDUC_1 | educ_1: HIGHEST YEAR OF SCHOOL COMPLETED |
| educ_2 | byte | %8.0g | EDUC_2 | educ_2: HIGHEST YEAR OF SCHOOL COMPLETED |
| educ_3 | byte | %8.0g | EDUC_3 | educ_3: HIGHEST YEAR OF SCHOOL COMPLETED |
| incom16_1 | byte | %8.0g | INCOM16 | incom16_1: RS FAMILY INCOME WHEN 16 YRS OLD |
| incom16_2 | byte | %8.0g | V1318_A | incom16_2: RS FAMILY INCOME WHEN 16 YRS OLD |
| incom16_3 | byte | %8.0g | V1319_A | incom16_3: RS FAMILY INCOME WHEN 16 YRS OLD |
| race_1 | byte | %8.0g | RACE_1 | race_1: RACE OF RESPONDENT |
| race_2 | byte | %8.0g | RACE_2 | race_2: RACE OF RESPONDENT |
| race_3 | byte | %8.0g | RACE_3 | race_3: RACE OF RESPONDENT |
| satjob_1 | byte | %8.0g | SATJOB_1 | satjob_1: JOB OR HOUSEWORK |

```

satjob_2      byte    %8.0g      SATJOB_2  satjob_2: JOB OR HOUSEWORK
satjob_3      byte    %8.0g      SATJOB_3  satjob_3: JOB OR HOUSEWORK
ID            float   %9.0g

```

Sorted by:

Note: Dataset has changed since last saved.

Codebook for Selected Variable(s)

```
. codebook satjob_3
```

```

satjob_3                                     satjob_3: JOB OR HOUSEWORK

```

```

      type: numeric (byte)
      label: SATJOB_3
      range: [1,4]
unique values: 4
unique mv codes: 3
      tabulation: Freq.  Numeric  Label
                  483      1  VERY SATISFIED
                  367      2  MOD. SATISFIED
                   69      3  A LITTLE DISSAT
                   39      4  VERY DISSATISFIED
                    4      .d  DK
                  1,073    .i  IAP
                    9      .n  NA

```

Analyses Relying On Wide Data

Correlation

```
. pwcorr satjob_3 educ_3, sig
```

| | satjob_3 | educ_3 |
|----------|-------------------|--------|
| satjob_3 | 1.0000 | |
| educ_3 | -0.0774 0.0166 | 1.0000 |

Regression with 1 Independent Variable

```
. regress satjob_3 educ_3
```

| Source | SS | df | MS | Number of obs | = | 957 |
|----------|------------|-----|------------|---------------|---|--------|
| Model | 3.53828635 | 1 | 3.53828635 | F(1, 955) | = | 5.76 |
| Residual | 586.493062 | 955 | .61412886 | Prob > F | = | 0.0166 |
| Total | 590.031348 | 956 | .617187602 | R-squared | = | 0.0060 |
| | | | | Adj R-squared | = | 0.0050 |
| | | | | Root MSE | = | .78366 |

| satjob_3 | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|----------|-----------|-----------|-------|-------|-----------------------|
| educ_3 | -.0216864 | .0090349 | -2.40 | 0.017 | -.0394169 -.003956 |
| _cons | 1.954439 | .1297867 | 15.06 | 0.000 | 1.699739 2.209139 |

Regression With Multiple Independent Variables

```
. regress satjob_3 educ_3 i.race_3 incom16_3
```

| Source | SS | df | MS | Number of obs | = | 951 |
|----------|------------|-----|------------|---------------|---|--------|
| Model | 5.81703392 | 4 | 1.45425848 | F(4, 946) | = | 2.36 |
| Residual | 582.580442 | 946 | .615835563 | Prob > F | = | 0.0517 |
| | | | | R-squared | = | 0.0099 |
| | | | | Adj R-squared | = | 0.0057 |
| Total | 588.397476 | 950 | .619365765 | Root MSE | = | .78475 |

| satjob_3 | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|-----------|-----------|-----------|-------|-------|----------------------|-----------|
| educ_3 | -.0215151 | .0092674 | -2.32 | 0.020 | -.0397021 | -.0033281 |
| race_3 | | | | | | |
| black | .1267666 | .0708898 | 1.79 | 0.074 | -.0123528 | .2658861 |
| other | .0677238 | .0985112 | 0.69 | 0.492 | -.1256019 | .2610495 |
| incom16_3 | .0115275 | .0280601 | 0.41 | 0.681 | -.0435398 | .0665947 |
| _cons | 1.89556 | .144649 | 13.10 | 0.000 | 1.61169 | 2.17943 |

Propensity Score

Data Wrangling

```
. generate twelve_years_3 = educ_3 >= 12 // 12 or more years of education
. generate twelve_years_2 = educ_2 >= 12 // 12 or more years of education
. generate twelve_years_1 = educ_1 >= 12 // 12 or more years of education
```

Propensity Score Analysis

```
. teffects psmatch (satjob_3) (twelve_years_3 incom16_3 i.race_3)
```

| | | | |
|---------------------------------------|--------------------|---|-----|
| Treatment-effects estimation | Number of obs | = | 952 |
| Estimator : propensity-score matching | Matches: requested | = | 1 |
| Outcome model : matching | min | = | 1 |
| Treatment model: logit | max | = | 296 |

| satjob_3 | Coef. | AI Robust Std. Err. | z | P> z | [95% Conf. Interval] | |
|----------------|-----------|---------------------|-------|-------|----------------------|----------|
| ATE | | | | | | |
| twelve_years_3 | | | | | | |
| (1 vs 0) | -.0410168 | .1083808 | -0.38 | 0.705 | -.2534393 | .1714057 |

Check Balance

Analyses Relying On Long Data

Reshape The Data

```
. reshape long satjob_ educ_ twelve_years_ incom16_ race_, i(ID) j(wave)
(note: j = 1 2 3)
```

| | | | |
|----------------|------|----|------|
| Data | wide | -> | long |
| Number of obs. | 2044 | -> | 6132 |

```

Number of variables      16  ->    7
j variable (3 values)    ->   wave
xij variables:
    satjob_1 satjob_2 satjob_3  ->  satjob_
      educ_1 educ_2 educ_3      ->   educ_
twelve_years_1 twelve_years_2 twelve_years_3-> twelve_years_
    incom16_1 incom16_2 incom16_3 ->  incom16_
      race_1 race_2 race_3      ->   race_

```

Clean Up Variable Names

```

. rename satjob_ satjob

. rename educ_ educ

. rename incom16_ incom16

. rename race_ race

. rename twelve_years_ twelve_years

```

Multilevel Model

```
. mixed satjob wave educ incom16 i.race || ID:
```

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log likelihood = -4161.775

Iteration 1: log likelihood = -4161.7476

Iteration 2: log likelihood = -4161.7476

Computing standard errors:

Mixed-effects ML regression

Group variable: ID

Number of obs = 3,595

Number of groups = 1,661

Obs per group:

min = 1

avg = 2.2

max = 3

Wald chi2(5) = 42.38

Prob > chi2 = 0.0000

Log likelihood = -4161.7476

| | satjob | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|--|---------|-----------|-----------|-------|-------|----------------------|-----------|
| | wave | -.018625 | .014015 | -1.33 | 0.184 | -.0460938 | .0088439 |
| | educ | -.018976 | .0054133 | -3.51 | 0.000 | -.0295859 | -.008366 |
| | incom16 | -.0350535 | .0154559 | -2.27 | 0.023 | -.0653465 | -.0047606 |
| | race | | | | | | |
| | black | .1695589 | .0451171 | 3.76 | 0.000 | .0811311 | .2579868 |
| | other | .035975 | .0543135 | 0.66 | 0.508 | -.0704776 | .1424276 |
| | _cons | 2.049073 | .0843019 | 24.31 | 0.000 | 1.883845 | 2.214302 |

| Random-effects Parameters | | Estimate | Std. Err. | [95% Conf. Interval] | |
|---------------------------|---------------|----------|-----------|----------------------|----------|
| ID: Identity | var(_cons) | .2305185 | .0161162 | .2009999 | .2643722 |
| | var(Residual) | .4174209 | .0131143 | .3924927 | .4439323 |

LR test vs. linear model: chibar2(01) = 322.95

Prob >= chibar2 = 0.0000

Fixed effects regression

```
. xtreg satjob wave educ incom16 i.race, i(ID) fe
Fixed-effects (within) regression      Number of obs   =       3,595
Group variable: ID                    Number of groups =       1,661
R-sq:                                Obs per group:
    within = 0.0052                      min =           1
    between = 0.0148                     avg =           2.2
    overall = 0.0122                     max =           3
                                         F(5,1929)        =       2.03
                                         Prob > F         =       0.0711
corr(u_i, Xb) = -0.0714
```

| satjob | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|---------|-----------|-----------------------------------|-------|-------|----------------------|-----------|
| wave | -.0237842 | .0152551 | -1.56 | 0.119 | -.0537023 | .006134 |
| educ | -.0087664 | .0158008 | -0.55 | 0.579 | -.0397548 | .022222 |
| incom16 | -.047186 | .0228265 | -2.07 | 0.039 | -.0919531 | -.0024189 |
| race | | | | | | |
| black | .3226033 | .2025604 | 1.59 | 0.111 | -.0746572 | .7198637 |
| other | .0383663 | .104807 | 0.37 | 0.714 | -.1671806 | .2439132 |
| _cons | 1.928458 | .227991 | 8.46 | 0.000 | 1.481323 | 2.375593 |
| sigma_u | .6861769 | | | | | |
| sigma_e | .64822634 | | | | | |
| rho | .52841711 | (fraction of variance due to u_i) | | | | |

F test that all u_i=0: F(1660, 1929) = 2.18 Prob > F = 0.0000

“Hybrid” Model

Generate Within And Between Variables

```
. bysort ID: egen educ_mean = mean(educ)
(6 missing values generated)

. generate educ_deviation = educ - educ_mean
(1,240 missing values generated)

. mixed satjob wave educ_mean educ_deviation incom16 i.race || ID:
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log likelihood = -4161.3224
Iteration 1:   log likelihood = -4161.2951
Iteration 2:   log likelihood = -4161.2951
```

```
Computing standard errors:
Mixed-effects ML regression      Number of obs   =       3,595
Group variable: ID              Number of groups =       1,661
                                Obs per group:
                                    min =           1
                                    avg =           2.2
                                    max =           3
                                Wald chi2(6)        =       43.30
                                Prob > chi2         =       0.0000
Log likelihood = -4161.2951
```

| satjob | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|----------------|-----------|-----------|-------|-------|----------------------|-----------|
| wave | -.0197009 | .0140588 | -1.40 | 0.161 | -.0472556 | .0078537 |
| educ_mean | -.0208983 | .0057775 | -3.62 | 0.000 | -.0322221 | -.0095745 |
| educ_deviation | -.0054971 | .0151667 | -0.36 | 0.717 | -.0352233 | .0242292 |
| incom16 | -.0343579 | .0154712 | -2.22 | 0.026 | -.0646809 | -.0040349 |
| race | | | | | | |

| | | | | | | |
|-------|----------|----------|-------|-------|-----------|----------|
| black | .1684699 | .0451261 | 3.73 | 0.000 | .0800245 | .2569154 |
| other | .0342568 | .0543368 | 0.63 | 0.528 | -.0722414 | .140755 |
| _cons | 2.075849 | .088866 | 23.36 | 0.000 | 1.901675 | 2.250023 |

| Random-effects Parameters | | Estimate | Std. Err. | [95% Conf. Interval] | |
|---------------------------|---------------|----------|-----------|----------------------|----------|
| ID: Identity | | | | | |
| | var(_cons) | .2304651 | .0161097 | .2009581 | .2643046 |
| | var(Residual) | .4173132 | .0131099 | .3923934 | .4438157 |

LR test vs. linear model: chibar2(01) = 323.08 Prob >= chibar2 = 0.0000