Andy's Modeling Updates

Andy Shen

2023-04-06

Agenda 4/6/2023

- Discuss R Project workflow and data pre-processing script
- Andy's updates on modeling
- "Evaluating situational decomposition" qualms
- Discuss structure of report

R Projects

- "Sandbox" workspace for a specific project
- Main advantage is there is no need to change working directories or paths
- There is only 1 R project: `code`
- To open the project, double click `code.Rproj` (should open a new RStudio window)
- You must have open the `.Rproj` file before opening and running any `.R` or `.Rmd` file

Data pre-processing

- There is an R script in the `code/` directory
- Judgment calls and decision rules

Section 1

Replicating Felson Study 1: Effect of intoxication on sexual intercourse

Judgment calls

Note: all of the code can be found in the accompanying `.Rmd` file for these slides

- Removing NAs in outcome (intercourse)
- People who refused to respond were categorized in the reference category (no for sex, never for alcohol)

Total association with logistic regression

Table 1: Total association logistic regression odds ratios (Andy).

Gender	Occasionally	Frequently	OR_diff
all_gender	4.4	8.8	4.3
male	4.0	8.1	4.2
female	4.9	9.3	4.4

Table 2: Total association logistic regression odds ratios (Felson et al.).

Gender	Occasionally	Frequently	OR_diff
all_gender	4.0	8.5	4.5
male	3.6	8.7	5.1
female	4.4	7.7	3.3

Takeaways

- Let's assume this subsample is representative of the larger sample used in the paper
- In this case, our analysis has "closed the gap" in terms of odds ratios
- Perhaps the tendency to have sex isn't *that* drastic for occasional vs frequent drinkers?
- How does this relate to spuriousness?

Spuriousness of intoxication on sober sex

Table 3: Our spuriousness values

	Occasionally	Frequently
all_gender	91.0	84.4
males	88.7	84.0
females	92.7	84.6

Table 4: Felson's spuriousness values

	Occasionally	Frequently
all_gender	95.7	91.6
males	95.3	91.2
females	97.3	93.6

Section 2

Replicating Felson Study 2: Effect of intoxication on contraceptive use

Data pre-processing

- Same R script: `data-preprocessing.R`
- Reduce cases down to respondents who have had sex ($n = 2565$)

Judgment Calls

- NAs in contraception use were not given benefit of the doubt

Total association using logistic regression

Table 5: Total association logistic regression odds ratios (Andy).

Gender	Occasionally	Frequently	OR_diff
all_gender	1.00	1.40	0.40
male	0.97	1.36	0.39
female	1.02	1.57	0.55

Table 6: Total association logistic regression odds ratios (Felson et al.).

Gender	Occasionally	Frequently	OR_diff
all_gender	1.05	1.38	0.33
male	1.00	1.40	0.40
female	1.07	1.51	0.44

Takeaways

- This seems pretty good
- Now let's discuss spuriousness

Spuriousness of intoxication on contraception use

Table 7: Our spuriousness values

	Occasionally	Frequently
all_gender	-3129.5	-3.3
males	646.4	-27.9
females	-460.2	36.1

Table 8: Felson's spuriousness values

	Occasionally	Frequently
all_gender	NA	46.9
males	NA	41.1
females	NA	68.3

Side note

- Both the coefficients from binary LR and multinomial LR are significant for frequent female drinkers only in Felson's study.
- This happens to be the only result with a reasonable spuriousness value!

Section 3

Critiquing situational decomposition

Questions from our proposal

- Does SD make sense?
- Under what assumptions would SD yield the correct result?
- Are these assumptions reasonable in this setting?
- Can we construct an example in which SD fails?

Issues with situational decomposition

- What happens when there is above 100% spuriousness?
- What happens when the coefficient is negative? (Felson seems to wave it away)

Next steps

- Andy
  - Write up today's results into a latex document called `final-report.tex` (formatted by TD)
  - Start preprocessing data for instrumental variables analysis
  - Look into doing IVLS with binary outcome (IV logistic regression)
- Tiffany
  - Look into a sensitivity analysis method (Rosenbaum p-value bound?)
  - Write up "Evaluating situational decomposition" after all the questions in the proposal (section 3.1) are discussed
- Andy and Tiffany
  - Come up with a mathematically driven causal inference model
  - Decide which covariates should be adjusted for in IVLS or a re-do of Felson