Inverse Probability Weighting for Dealing with Missing Outcomes

Yuqing Qiu yuq2002@med.cornell.edu

Introduction

- IPW is a method to adjust for selection bias in observational studies
- Selection bias results in unrepresentative samples

Selection bias

- **Characteristics can be different between participants and non-participants** (bad eligibility criteria, non-response in survey data, rare characteristics)
- **between treatment and control groups** – Ideal case is randomized groups
- **between participants who stayed and those who dropped out** – Ideal case for complete case analysis is MCAR or MAR

Inverse Probability Weighting

Creates a pseudo-population by assigning to each participant a weight that is proportional to the participant's probability

- of being selected in the sample
- of receiving treatment/exposure
- of being a complete case
- …

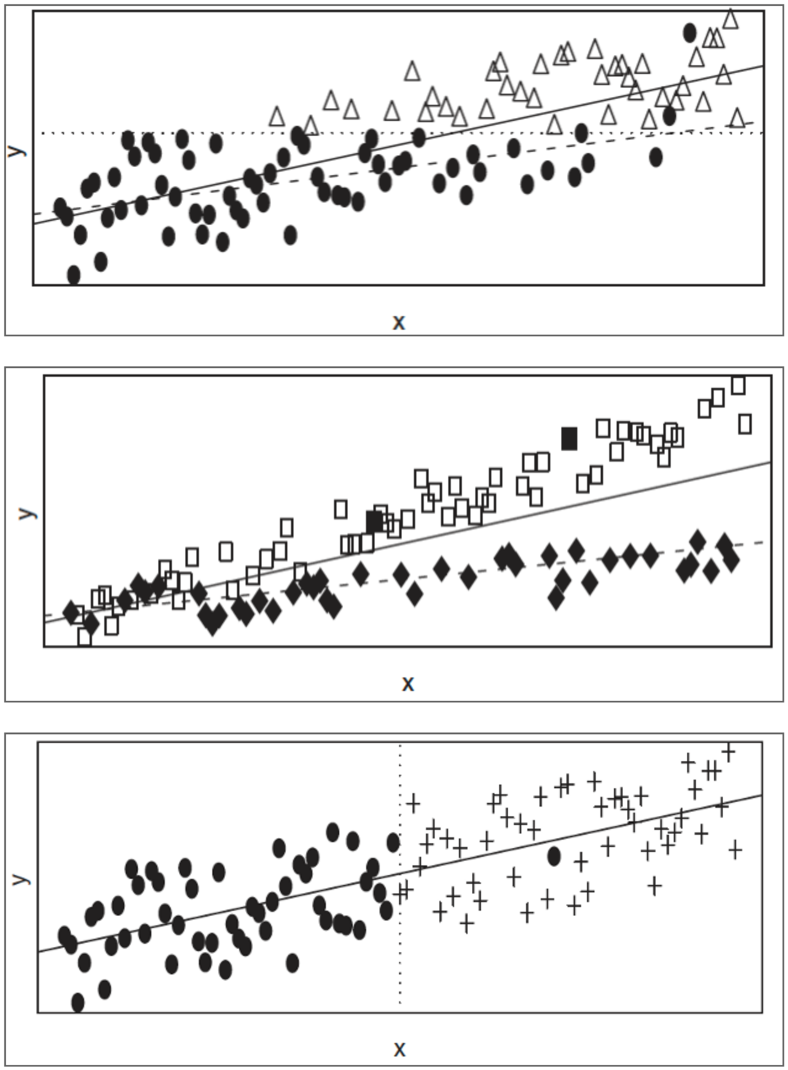
$$\boxed{w_i = \frac{1}{P(A_i = a_i | \boldsymbol{C}_i = \boldsymbol{c}_i)}}$$

Process of using IPW for Missing Data

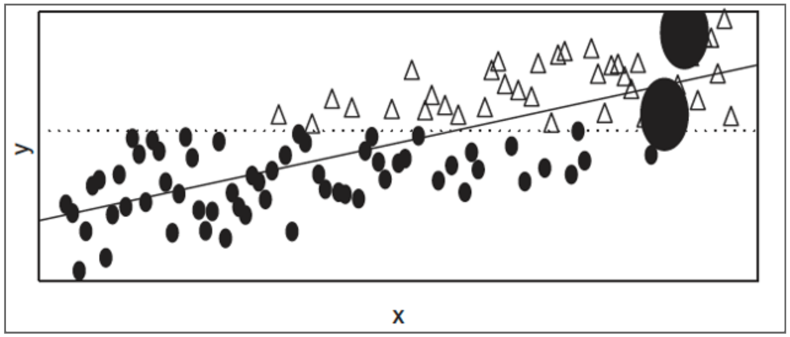
- Assess whether complete case analysis is biased
- Build missingness model to predict the **probability of being a complete case**
- Calculate inverse probability weight
- Build analysis model, weighting for IPW

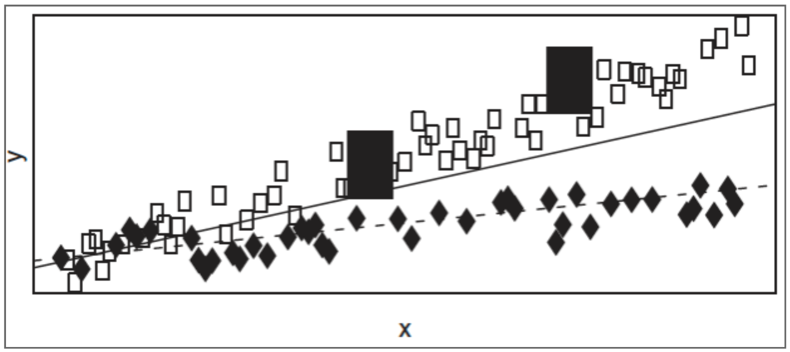
Assess whether complete case analysis is biased

Effect of missingness on estimate



How IPW solves it



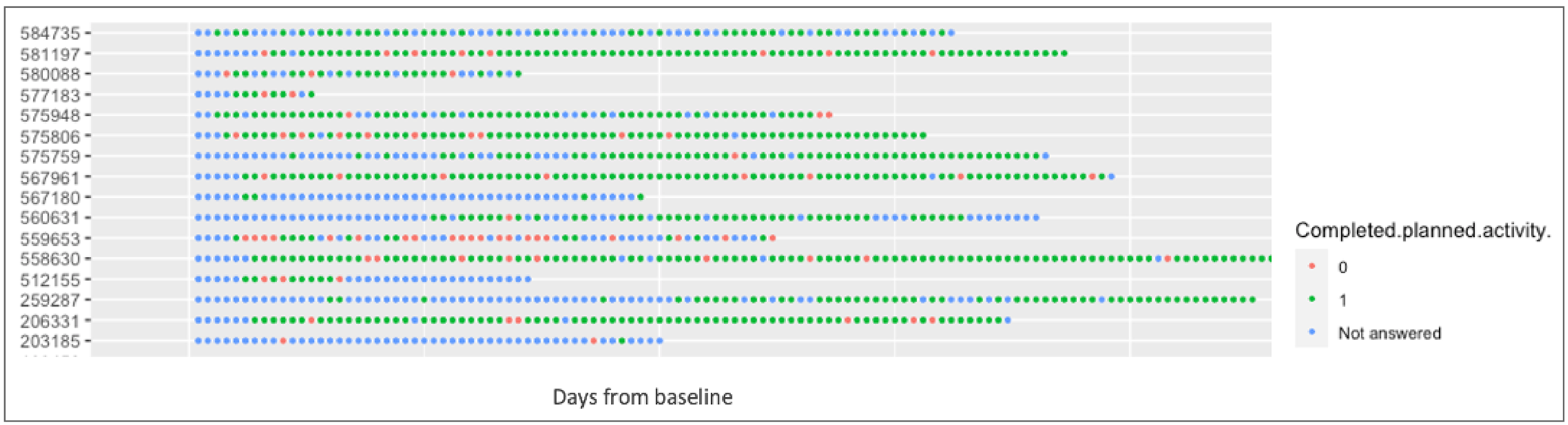


Choice of predictors in missingness model

- Identify a priori candidate predictors for missingness
- Remove those are independent of X and Y
- Add variables associated with Y

Example study

Topic: Gamification via mHealth to Improve Adherence to Psychotherapy in Depressed Older Adults
Outcome: Self-reported between-session homework (HW) completion
Intervention: Phase II with gamification vs. Phase I without
Confounders: Age, gender, ethnicity, insurance, education, baseline BISBAS, baseline MADRS
Problem: Not all patients reported HW completion at a daily bases, and we assume not reporting is associated with not completing



Example - Build missingness model

Identify a priori candidate predictors for missingness:

Patient level: age, education, motivation level, depression severity

Daily level: days from baseline, pain level, stress level, anhedonia, pedometers, travel diameters, active time, time at home, daily talking time, sleep hours, sleep scores

Remove those are independent of X and Y

No in our case

Add variables associated with Y

No additional

Example - Build missingness model

```
## Build missing model
missing_model <- glmer(
  response ~
    ## patient level demo
    age_consent + education+ finres_4 +

  ## patient level clinical
  bisbas_drive + bisbas_funtot + baseline_madrs +

  ## daily level active measures
  pain + stress+ morning_anhedonia + sadness +

  ## daily level passive measures
  days_from_start_date + steps + time_at_home + conversation_percent + tic_voiced_tim
  travel_diameter + active_hours + total_location_duration + radius_of_gyration + awa
  (1 | Participant), data = adherence, family = binomial)

hoslem.test(adherence$response, predict(missing_model, type = "response"))

## Store weight in data
adherence <- adherence %>%
  mutate(ipw = 1 / predict(missing_model, type = "response"))
```

Example - Check weight

Weight for each non-missing data point

Example - Build analysis model with weighting

```
## analysis model
md <- glmer(`HW_complete` ~

    ## Intervention
    gamification +

    ## Confounders
    age_consent + gender + ethnicity + finres_4  + educa

    ## random effect
    (Arm|Participant),
    # (1|access_groups/Participant)
    # (1|access_groups) + (1|Participant:access_groups)

    ## add weight
    weights = ipw,
    data = adherence, family = binomial)
```

Further improvement

Weight stabilization - narrower range of ipw

Truncation - reduce the bias introduced by extremely large weight for a participant that is caused by the misspecification of the missingness model

Package ipw

```r
library(ipw)

ipwpoint(
  ## exposure group
  exposure = response,

  ## model setting
  family = "binomial", link = "logit",

  ## ~ 1 for unstabilized, ~ exposure prevalence for stabilized
  numerator = ~ 1,

  ## predictors
  denominator = ~ predictors_in_missingness_model,

  ## truncation
  trunc = 0.05,

  data = adherence)
```

Other common application

- Inverse Probability Treatment Weighting (IPTW)
- Fit a marginal structural model (MSM) with IPW

Reference

Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res. 2013 Jun;22(3):278-95. doi: 10.1177/0962280210395740. Epub 2011 Jan 10. PMID: 21220355.

Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008 Sep 15;168(6):656-64. doi: 10.1093/aje/kwn164. Epub 2008 Aug 5. PMID: 18682488; PMCID: PMC2732954.

van der Wal, W. M., & Geskus, R. B. (2011). ipw: An R Package for Inverse Probability Weighting. *Journal of Statistical Software*, *43*(13), 1–23.

Thank You!