Causal Inference Methods and Case Studies

STAT24630

Jingshu Wang

Lecture 9

Topic: Two case studies

- Case study 1: Evaluation of SWIM program
- Case study 2: Analysis of HOMEFOOD randomized trial

Case study 1: evaluation of the SWIM program

Background

- SWIM (Saturation Work Initiative Model) was operated by the County of San Diego, California, from 1985 to 1987
- Targeted to individuals applying for or receiving benefits under the Aid to Families with Dependent Children (AFDC) Program, aim to maximize participation in employment-promoting activities among **heads of single-parent families without preschool-age children** (mostly women) and heads of two-parent families (mostly men).

- SWIM provided job search and unpaid work experience, and education and training to those who still did not find regular employment

- Possible goals including: increasing overall employment and earnings levels among AFDC recipients; reducing the level of AFDC receipt among long-term or potential long-term AFDC recipients; saving money for government budgets by reducing AFDC and other welfare expenditures; and reducing poverty

Case study 1: evaluation of the SWIM program

- Samples: $N = 3211$ individuals, who are head of single-parent families

- Randomization and treatment assignment:
  - Samples were randomly assigned to either an experimental or control group
  - No further details about the randomization mechanism, we treat it as a completely randomized experiment with $N_t = 1604$ and $N_c = 1607$
  - Individuals in the experiment group were required to participate in SWIM
  - Individuals in the control group were not eligible for SWIM activities but could, on their own initiative, enroll in community education and training programs.

  - We should interpret the treatment effect as the effect of participating in the program versus being denied access to this particular program, rather than as the effect of participating versus not participating in any job-training program
    Non-compliance

Pre-treatment covariates

Pre-treatment covariates

- Include individual-level background characteristics and records of earning prior to experiment
- Pre-treatment covariates are all well balanced

Variable		All ($N = 3211$)		Controls ($N_c = 1607$)		Treated ($N_t = 1604$)	
		Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
Pre-treatment variables							
`female`	female	0.91	(0.28)	0.92	(0.28)	0.91	(0.28)
`agege35`	(age $\geq$ 35)	0.46	(0.50)	0.46	(0.50)	0.46	(0.50)
`hsdip`	(high school diploma)	0.56	(0.50)	0.56	(0.50)	0.56	(0.50)
`nevmar`	(never married)	0.30	(0.46)	0.30	(0.46)	0.30	(0.46)
`divwid`	(divorced or widowed)	0.37	(0.48)	0.37	(0.48)	0.36	(0.48)
`numchild`	(number of children)	1.76	(1.08)	1.76	(1.07)	1.76	(1.10)
`chldlt6`	(children younger than 6)	0.10	(0.30)	0.10	(0.31)	0.10	(0.29)
`af-amer`	(African-American)	0.42	(0.49)	0.43	(0.49)	0.42	(0.49)
`hisp`	(Hispanic)	0.25	(0.44)	0.25	(0.43)	0.26	(0.44)
`earnyrm1`	(earnings year minus 1)	1.57	(3.54)	1.60	(3.56)	1.53	(3.51)
`empyrm1`	(positive earnings year minus 1)	0.39	(0.49)	0.40	(0.49)	0.39	(0.49)

Outcome variables

- The experiment had recorded many outcomes, including annual earnings, employed or not in each year, annual AFDC payments for 5 years
- Here, we focus on earnings of the first two years post-randomization
  - Annual earnings increase compared to the pre-randomization year even for the control group

Outcomes variables							
earnyr1	(earnings year 1)	1.85	(3.78)	1.69	(3.76)	2.02	(3.80)
empyr1	(positive earnings year 1)	0.46	(0.50)	0.40	(0.49)	0.52	(0.50)
earnyr2	(earnings year 2)	2.57	(5.08)	2.26	(4.68)	2.89	(5.44)
empyr2	(positive earnings year 2)	0.45	(0.50)	0.40	(0.49)	0.49	(0.50)

Fisher's exact p-values

(based on 1,000,000 draws from randomization distribution)

Post-Program Earnings	Statistic	All (3,211)	No High School (1,409)	High School (1,802)
Year 1	T^{rank}	< 0.0001	< 0.0001	0.0014
	$T^{\text{rank-gain}}$	< 0.0001	< 0.0001	0.0001
	T^{dif}	0.0131	0.0051	0.1967
Year 2	T^{rank}	< 0.0001	0.0017	< 0.0001
	$T^{\text{rank-gain}}$	< 0.0001	0.0020	0.0002
	T^{dif}	0.0004	0.0980	0.0018

$$R_i = \sum_{i'=1}^{N} \mathbf{1}_{Y_{i'}^{\text{obs}} < Y_i^{\text{obs}}} + \frac{1}{2}\left(1 + \sum_{i'=1}^{N} \mathbf{1}_{Y_{i'}^{\text{obs}} = Y_i^{\text{obs}}}\right) - \frac{N+1}{2}$$

$$T^{\text{rank}} = \left|\overline{R}_{\text{t}} - \overline{R}_{\text{c}}\right|$$

$$R'_i = \sum_{i'=1}^{N} \mathbf{1}_{Y_{i'}^{\text{obs}} - X_{i'} < Y_i^{\text{obs}} - X_i} + \frac{1}{2}\left(1 + \sum_{i'=1}^{N} \mathbf{1}_{Y_{i'}^{\text{obs}} - X_{i'} = Y_i^{\text{obs}} - X_i}\right) - \frac{N+1}{2}$$

$$T^{\text{rank,gain}} = \left|\overline{R'}_{\text{t}} - \overline{R'}_{\text{c}}\right|$$

$$T^{\text{dif}} = \left|\overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}}\right|$$

Fisher’s exact p-values

- Why is the mean difference statistics less powerful than the rank-based statistics?
  - Rank-based statistics are more sensitive if many individuals have a non-zero treat effect but the effects are small

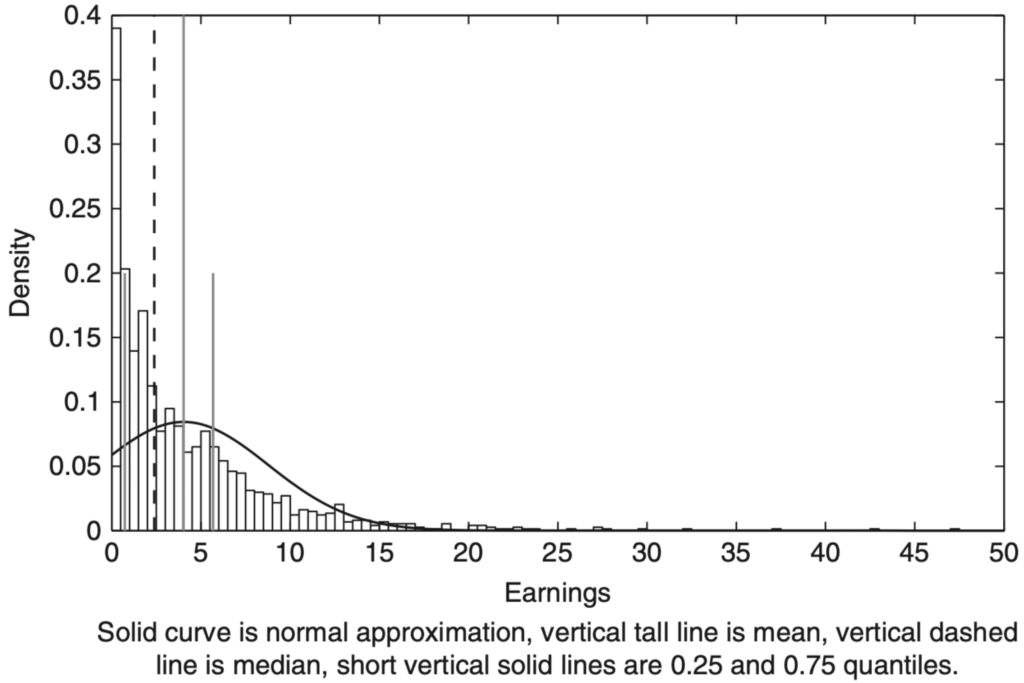


Figure 11.1. Histogram-based estimate of the distribution of Year 1 earnings, for those with positive earnings, San Diego SWIM program data

- We do not construct CI using Fisher’s approach as here we do not believe in the constant treatment effect assumption

Neyman's repeated sampling approach

- We can apply Neyman's approach to either the whole population or any subgroup

Post-Program Earnings		All (3,211)	Young (1,738)	Old (1,473)	Unemployed (1,949)	Employed (1,262)	No HS (1,409)	HS (1,802)
Year 1	Est	0.33	0.19	0.50	0.34	0.38	0.41	0.27
	$\widehat{\text{(s. e.)}}$	(0.13)	(0.17)	(0.21)	(0.13)	(0.25)	(0.15)	(0.21)
Year 2	Est	0.63	0.52	0.76	0.58	0.77	0.31	0.87
	$\widehat{\text{(s. e.)}}$	(0.18)	(0.24)	(0.27)	(0.19)	(0.33)	(0.19)	(0.28)

- Post-stratification

$$\begin{aligned}\hat{\tau}^{\text{strat}} &= \frac{N(\text{empl})}{N(\text{empl}) + N(\text{unempl})} \cdot \hat{\tau}^{\text{dif}}(\text{empl}) + \frac{N(\text{unempl})}{N(\text{empl}) + N(\text{unempl})} \cdot \hat{\tau}^{\text{dif}}(\text{unempl}) \\ &= \frac{1262}{1262 + 1949} \cdot 0.38 + \frac{1949}{1262 + 1949} \cdot 0.34 = 0.36 \;\; (\widehat{\text{s. e.}}\ 0.15),\end{aligned}$$

Regression analysis

- We can incorporate all 11 pre-treatment covariates and include interactions in the linear regression model to allow heterogeneity of conditional average treatment effects across $\boldsymbol{X}$

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + (X_i - \overline{X})\beta + W_i \cdot (X_i - \overline{X})\gamma + \varepsilon_i$$

  - Compare linear regression without / with covariates

Covariates	Earnings Year 1				Earnings Year 2			
	Est	$(\widehat{\text{s. e.}})$	Est	$(\widehat{\text{s. e.}})$	Est	$(\widehat{\text{s. e.}})$	Est	$(\widehat{\text{s. e.}})$
`Treat`	0.33	(0.13)	0.36	(0.12)	0.63	(0.18)	0.66	(0.17)
`Intercept`	1.69	(0.09)	1.68	(0.09)	2.26	(0.12)	2.25	(0.11)

  - We only see a moderate reduction of the standard errors

Regression analysis

Hypothesis testing

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + (X_i - \overline{X})\beta + W_i \cdot (X_i - \overline{X})\gamma + \varepsilon_i$$

- Test whether average treatment effect is 0. H_0: $\mathbb{E}\big(Y_i(1) - Y_i(0)\big) = 0$
  - For the regression model, we test H_0: $\tau = 0$
  - Z-value $\hat{\tau}^{\text{ols}}/\sqrt{\widehat{\mathbb{V}}_\tau}$ compare with $N(0, 1)$ to obtain a p-value
- Test whether the conditional treatment effect is 0 for every level of $\boldsymbol{X}$. H_0: $\mathbb{E}(Y_i(1) - Y_i(0)|\ \boldsymbol{X}_i = \boldsymbol{x}) = 0$ for all $\boldsymbol{x}$
  - For the regression model, we test H_0: $\tau = 0$ and $\gamma = 0$
  - Test statistics $\begin{pmatrix} \hat{\tau}^{\text{ols}} \\ \hat{\gamma}^{\text{ols}} \end{pmatrix}^T \hat{\mathbb{V}}_{\tau,\gamma}^{-1} \begin{pmatrix} \hat{\tau}^{\text{ols}} \\ \hat{\gamma}^{\text{ols}} \end{pmatrix}$ compared with $\chi^2(\dim(\boldsymbol{X}) + 1)$
- Test whether treatment effect is heterogenous across covariates. H_0: $\mathbb{E}(Y_i(1) - Y_i(0)|\ \boldsymbol{X}_i = \boldsymbol{x}) \equiv \tau$ for all $\boldsymbol{x}$
  - For the regression model, we test H_0: $\gamma = 0$
  - Test statistics $(\hat{\gamma}^{\text{ols}})^T \widehat{\mathbb{V}}_\gamma^{-1} \hat{\gamma}^{\text{ols}}$ compared with $\chi^2(\dim(\boldsymbol{X}))$ to obtain a p-value

Regression analysis

Table 11.5. ***P-Values for Tests of Constant and Zero Treatment Effects Assumptions, for San Diego SWIM Data***

Null Hypothesis		Earnings Year 1	Earnings Year 2
Zero effect	$\mathcal{X}^2(12)$ approximation	0.018	<0.001
	Fisher exact p-value	0.157	0.014
Constant effect	$\mathcal{X}^2(11)$ approximation	0.122	0.002

- Little evidence for heterogenous effect across X for the first-year earnings, but clear evidence of heterogenous effect for the second year
- The fisher’s exact p-value are computed using the same test statistics but under Fisher’s sharp null and use Fisher’s randomization framework to obtain the reference distribution of the test statistics

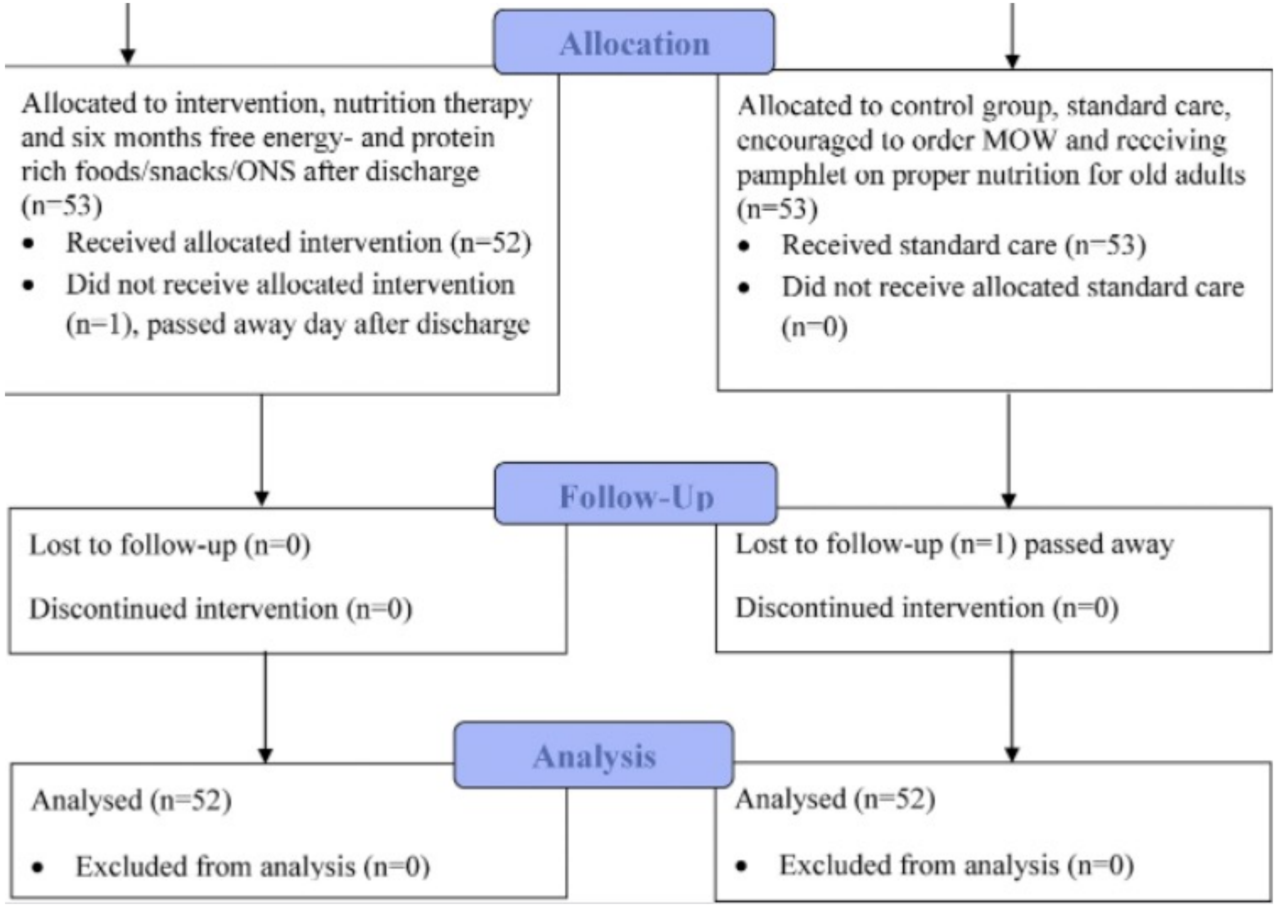
Case study 2: Analysis of HOMEFOOD randomized trial

[HOMEFOOD randomised trial–Six-month nutrition therapy improves quality of life, self-rated health, cognitive function, and depression in older adults after hospital discharge. *Clinical Nutrition ESPEN (2022)*.]

- Goal: investigate the effect of nutrition therapy on health-related quality of life
- Participants: Eligible participants were community dwelling patients discharging home from the hospital within 24 h, aged ≥65 years, and at risk for malnutrition
- Randomization: participants were randomly allocated (ratio = 1:1) to either the intervention or the control group by using a random number generated by the principal investigator
- Intervention: nutrition therapy from a clinical nutritionist consists of 5 home visits, 3 telephone calls, free supplemental energy- and protein-rich foods

Case study 2: Analysis of HOMEFOOD randomized trial

[HOMEFOOD randomised trial–Six-month nutrition therapy improves quality of life, self-rated health, cognitive function, and depression in older adults after hospital discharge. *Clinical Nutrition ESPEN (2022)*.]



- Non-compliance is a common issue in randomized experiments
- In this example, reasons that patients dropout are likely unrelated to the treatment
- Our analysis will be based on the N = 104 individuals

Variables	Control(n = 53)			intervention(n = 53)			P-value[a]
	mean	±	SD	mean	±	SD	
Age (years)	81.8	±	6.0	83.3	±	6.7	0.228
Female (%)		52.8			71.7		0.045
Higher education (in %)		66.0			69.8		0.677
Lives alone (%)		66.0			66.0		0.999
Alcohol (yes in %)		45.3			37.7		0.430
Smoking (yes in %)		9.4			3.8		0.241
Height (m)	1.7	±	0.1	1.7	±	0.1	0.326
Weight (kg)	76.5	±	19.1	78.3	±	18.3	0.615
BMI (kg/m^2)	26.9	±	5.3	28.5	±	6.5	0.188
SPPB (score)	2.4	±	2	2.5	±	1.8	0.839
ICD-10 diagnoses (no.)	10.5	±	3.8	10.3	±	4.9	0.877
Medications (no.)	12.4	±	4.2	12.2	±	5.8	0.893
MMSE (score)	25.9	±	2.9	26.1	±	2.8	0.702
EQ-5D (index)	0.688	±	0.193	0.694	±	0.146	0.852
Self-rated health (scale)	61.3	±	18.1	58.8	±	19.9	0.493
CES - D (score)	5.6	±	4.7	5.4	±	4.2	0.861

- We still want to check for covariates balancing even in randomized experiment
- If some covariates are not balanced, our analysis is still valid, but our conclusion can be very inaccurate
- Here sex is not balanced well, one solution is to use post-stratification and estimate causal effect on female and male groups separately
- Equivalently, we may also want to add sex as a covariate in linear regression
- Check R example 4 for data analysis