Causal Inference: Policy Evaluation — Assignment 1

Amit Aryal, Nicholas Putney, Dominik Schawrzkopf

2025-04-04

Question 1

```
## [1] "Difference in means by treatment status and standardized bias"
```

```
##               E(X|D=0) E(X|D=1) Difference  s.e. p-value Abs. SB
## under_40yo       0.617    0.629      0.013 0.007   0.073   2.601
## Observations 10336.000 8748.000         NA    NA      NA      NA
```

The dummy variable under_40yo seems well balanced between treatment and control groups, with low standardized bias (2.6%) and an insignificant difference in means at the 5% significance level.

We should still account for age differences in our analysis due to: 1. Heterogeneity: Age might moderate treatment effects (e.g., younger vs older participants respond differently), so including it enables subgroup or interaction analysis. 2. Robustness: Including covariates that are predictors of the outcome of interest protects against chance imbalances or unobserved heterogeneity in smaller subgroups. 3. Improved precision: if age group is a good predictor of employment status, then including it in the analysis will lead to a decrease mean squared error, leading to smaller standard errors for the ATE estimate.

Question 2

(a) see R script

(b) see R script

(c)

Monthly ATE Estimates: Under 40
Inverse Probability Weighting

Group	ATE	SE	Month	CI Low	CI High	Sig Eff
Under 40	−0.035	0.004	1	−0.043	−0.027	−0.035
Under 40	0.002	0.008	2	−0.015	0.018	NA
Under 40	0.077	0.009	3	0.060	0.095	0.077
Under 40	0.374	0.008	4	0.357	0.391	0.374
Under 40	0.375	0.007	5	0.362	0.388	0.375
Under 40	0.337	0.007	6	0.323	0.351	0.337
Under 40	0.302	0.006	7	0.290	0.315	0.302
Under 40	0.265	0.005	8	0.255	0.275	0.265
Under 40	0.228	0.005	9	0.218	0.238	0.228
Under 40	0.194	0.005	10	0.184	0.205	0.194
Under 40	0.170	0.005	11	0.161	0.179	0.170
Under 40	0.150	0.005	12	0.140	0.159	0.150
Under 40	0.132	0.004	13	0.125	0.140	0.132
Under 40	0.115	0.005	14	0.106	0.124	0.115
Under 40	0.102	0.004	15	0.094	0.109	0.102
Under 40	0.089	0.005	16	0.080	0.098	0.089
Under 40	0.077	0.004	17	0.070	0.085	0.077
Under 40	0.067	0.004	18	0.060	0.075	0.067
Under 40	0.056	0.003	19	0.050	0.062	0.056
Under 40	0.044	0.003	20	0.039	0.050	0.044
Under 40	0.036	0.003	21	0.030	0.042	0.036
Under 40	0.030	0.002	22	0.025	0.035	0.030
Under 40	0.026	0.002	23	0.022	0.031	0.026
Under 40	0.023	0.002	24	0.019	0.027	0.023

Monthly ATE Estimates: Over 40
Inverse Probability Weighting

Group	ATE	SE	Month	CI Low	CI High	Sig Eff
Over 40	−0.024	0.004	1	−0.032	−0.016	−0.024
Over 40	0.027	0.008	2	0.010	0.043	0.027
Over 40	0.109	0.011	3	0.088	0.130	0.109
Over 40	0.391	0.010	4	0.371	0.412	0.391
Over 40	0.432	0.009	5	0.414	0.450	0.432
Over 40	0.410	0.009	6	0.393	0.427	0.410
Over 40	0.393	0.009	7	0.375	0.410	0.393
Over 40	0.364	0.009	8	0.347	0.382	0.364
Over 40	0.326	0.008	9	0.310	0.342	0.326
Over 40	0.287	0.007	10	0.274	0.301	0.287
Over 40	0.259	0.008	11	0.244	0.275	0.259
Over 40	0.234	0.008	12	0.219	0.250	0.234
Over 40	0.211	0.007	13	0.197	0.225	0.211
Over 40	0.187	0.007	14	0.174	0.200	0.187
Over 40	0.164	0.007	15	0.150	0.178	0.164
Over 40	0.147	0.006	16	0.136	0.159	0.147
Over 40	0.129	0.006	17	0.117	0.141	0.129
Over 40	0.114	0.005	18	0.105	0.124	0.114
Over 40	0.097	0.006	19	0.086	0.108	0.097
Over 40	0.083	0.005	20	0.073	0.092	0.083
Over 40	0.073	0.005	21	0.063	0.083	0.073
Over 40	0.065	0.004	22	0.057	0.074	0.065
Over 40	0.056	0.005	23	0.047	0.065	0.056
Over 40	0.048	0.004	24	0.041	0.055	0.048

Question 3

Monthly ATEs on Employment Probability by Age Group
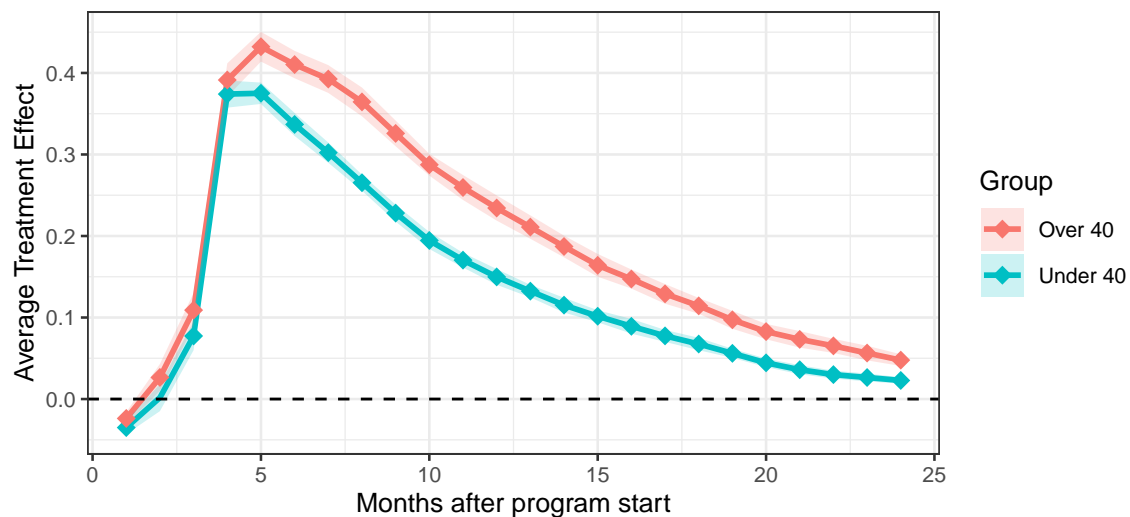
The figure would suggest that for the first month, participants in both age groups have a lower probability of employment than non-participants, perhaps due to their participation in the program in that month. The program had a large impact on both groups in the first five to six months, peaking at a change in employment probability by about 40 percentage points. After month 5, the effect of the program begins to decay. However, this decay is greater in the under 40 age group, reflected in a smaller average treatment effect.

The average treatment effect (ATE) is the effect of the program had everyone been treated, even including those who were not treated. The average treatment effect on the treated (ATET) is the effect of the program only on participants. This means that the ATET estimate is only relevant to participants and would not say something about the effect on the program on non-participants. This two quantities could differ if factors related to employment differed systematically between participants and non-participants. For example, the ATET might be higher if those who actually participated in the program are more responsive to the program compared to non-participants had they participated.

Question 4

(a)

In the case where jobseekers who participate in the online application program are more likely to get a job but at the expense of other unemployed workers, we should not compare program participants to non-participants to estimate the ATE(T), because these estimates would be biased. We would in this case overestimate the treatment effect of the online application program.

(b)

In this case, the Stable Unit Treatment Value Assumption (SUTVA) would be violated due to general equilibrium effects in the labor market.

(c)

Other violations of SUTVA could be spillover effects from treated to untreated (e.g. when participants would share their gained knowledge from the program with non-participants). Here the bias would be an underestimation of the treatment effect.

Question 5

The control and treatment groups are defined by the difference in distance between their location and the backbone network. Treatment groups are defined as individuals living less than 500m away from the network (T1), 500-1500m away (T2), 1500-2500m (T3) and 2500m-3000m (T4). The control group contains individuals located more than 3500m from the backbone network. The post-treatment period (t) is defined as at least one submarine cable has arrived in the country at time (t). The pre-treatment period as no submarine cable has arrived at time (t).

Question 6

(a)

The authors do not control for individual fixed effects due to the nature of the available data. DHS and Afrobarometer are cross-sectional and not longitudinal. Therefore it is not feasible to incorporate individual fixed effects.

(b)

It would not be a good idea to control for location-specific time period fixed effects, because the treatment varies across locations over time, which is exactly the variation those fixed effects would absorb.