Bayesian Causal Inference
Assignment

15 novembre 2024

You will examine data from the Job Search Intervention Study (JOBS II), a completely randomized experiment of an intervention for unemployed individuals to prevent poor mental health and to promote high-quality reemployment. The control condition consisted of a booklet briefly describing job search methods and tips. The intervention condition consisted of five active training sessions.
The aim here is to assess causal effects of the program on high-risk unemployed people, namely, study participants with values of a risk score depending on financial strain, assertiveness, and depression score greater than a pre-fixed threshold (1.38).
The sample consists of 398 high risk individuals with non-missing values on the relevant variables. The primary outcomes of interest are depression, measured with a sub-scale of 11 items based on the Hopkins Symptom Checklist, and re-employment, a binary variable taking on value 1 if a subject works for 20 hours or more per week, measured six months after the intervention assignment. A set of covariates are observed for each subject. Table 1 lists the variables we observed.
In JOBS II, individuals assigned to the control condition were prohibited from attending intervention sessions, but individuals assigned to the treatment condition may not show up to the intervention: only 54% of individuals who were assigned to the intervention actually received the treatment. In the literature, this problem is known as noncompliance. Here we ignore noncompliance issues focusing on assessing causal effects of the assignment to the treatment (Intention-to-treat analysis).

1. Load the dataset in R (Filename: JOBSII_HR.dta)

2. For each variable, calculate the mean for the whole sample and within each treatment group. For continuous covariates, also report the medians, standard deviation and ranges within each treatment group. Record your results in a table. In a few sentences, comment on what you see and whether it is expected.

3. Let π denote the probability of re-employment, n_c the number of individuals assigned to the control condition, and n_t the number of individuals assigned to the treatment condition. Consider the following Bayesian model:

   - $X_i(0) \sim \text{Bernoulli}(\pi)$ for each $i \in \{1, \ldots, n_c\}$;
   - $X_i(1) \sim \text{Bernoulli}(\pi)$ for each $i \in \{1, \ldots, n_t\}$;
   - The prior distribution for π is $\text{Beta}(1, 1)$.

   Compute the two posterior distributions, $p(\pi \mid x_1(0), \ldots, x_{n_c}(0))$ and $p(\pi \mid x_1(1), \ldots, x_{n_t}(1))$, under the exchangeability assumption. Represent and compare these posterior distributions, including the posterior means and 90% credible intervals. Comment on the results in a few sentences.

4. **Bayesian model-based analysis.** Bayesian model-based analysis for the outcome variable "depression six months after the intervention assignment." Assume that $Y_i(0)$ and $Y_i(1)$ are independent and are both log-normally distributed:

$$\begin{bmatrix} \log(Y_i(0)) \\ \log(Y_i(1)) \end{bmatrix} \mid \theta \sim \mathcal{N}\left(\begin{bmatrix} \beta_c \\ \beta_t \end{bmatrix}, \begin{bmatrix} \sigma_c^2 & 0 \\ 0 & \sigma_t^2 \end{bmatrix} \right)$$

   with $\theta = (\beta_c, \beta_t, \sigma_c^2, \sigma_t^2)$. Consider the following prior distributions

$$\begin{aligned} \beta_c &\sim \mathcal{N}(\nu_c, \tau_c^2) & \beta_t &\sim \mathcal{N}(\nu_t, \tau_t^2) \\ \sigma_c^2 &\sim \text{Scaled Inverse-}\chi^2(a_c, b_c^2) & \sigma_t^2 &\sim \text{Scaled Inverse-}\chi^2(a_t, b_t^2) \end{aligned}$$

Table 1: JOBSII data: Pre-treatment variables

Variable	Description	Label
Pre-treatment variables		
Gender	0 = Male; 1 = Female	sex
Age	Age in years	age
Race	0 = White; 1 = Non-white	race
Education	School grade completed	educ
Marital status	0 = Married; 1 = Other	nonmarried
Motivation	Motivation level before treatment	motivation
Pre-treatment Depression	Depression level before treatment	depress0
Economic hardship	Economic hardship before treatment	EconHard
Assertiveness	Assertiveness before treatment	assertive
Assignment variable	0 = Control; 1 = Treatment	Z
Treatment variable	0 = Control; 1 = Treatment	W
Outcome variables six months after the intervention assignment		
Depression	Depression level after treatment	depress6
Re-employment	0 = unemployed; 1 = employed	employ6

Derive the posterior distributions of the finite sample average causal effect and the super-population average causal effect. Plot the resulting posterior distributions in a histogram and report the following summary statistics of the resulting posterior distributions: mean, standard deviation, median, 2.5% and 97.5% percentiles.

Hint. For the finite sample average causal effect, you need to apply to all the imputed values the inverse transformation to obtain imputations of depression. For the super-population average causal effect, recall that if $\log(Y) \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}[Y] = \exp\left\{\mu + 0.5 \cdot \sigma^2\right\}$