UCL Network of Applied Statisticians in Health

Target Trial Emulation Short Course*
Practical 2 (Stata and R)
Guided Solutions

28 November 2022

Tasks and Comments

1. Examine the DAG: which arrow(s) would you remove if the data concerned an RCT?

   The arrow from L_0 to A_0 would have to be removed.

2. Read and summarise the data using either data format (you choose!). If using R you could read the data using the `haven` package (see separate R code).

   It is handier to read the data in long format to summarise the data because we want to count each individual only once

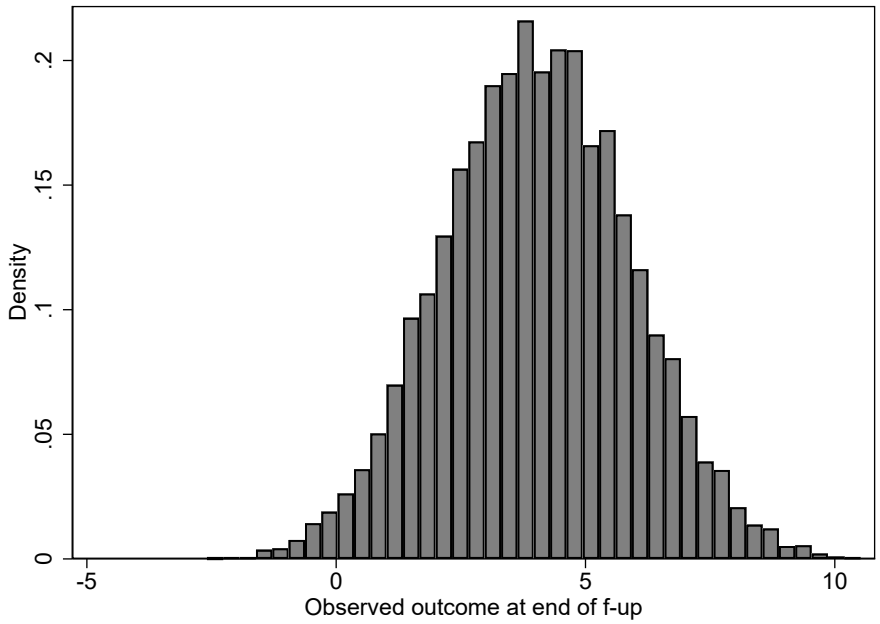
3. Examine the distribution of the outcome (remember that it is observed only at the end if follow-up. Note also that its value is repeated in each record (*i.e.* when t=0 and t=1) in the long format version).

   The mean of Y is 4.04 and its SD is 1.9. The range is from -2.6 to 10.5. The histogram shows that it is fairly symmetrically distributed, indicating that a linear regression model for the outcome would be adequate (*i.e.* without transformation).

4. How many patients initiate treatment at time 0? How many sustained treatment at time 1?

   Just over half (52.8%) of the patients were prescribed treatment at time 0. Of these only over half (56.8% were still on treatment at time 1. Of the 47.2% who

*Part of the work funded by MRC Methodology Grant: MR/R025215/1



had not started treatment at time 0, 44.6% started at time 1. Hence there is a substantial amount of treatment switching.

5. Estimate the conditional (confounder-adjusted) association between treatment initiation and the outcome using standard regression methods.

   The estimated regression coefficient for A_0, controlling for L_0, is 1.11 (95% CI: 1.05, 1.17) indicating a 1.1 increase in Y when exposed to A at time 0, conditioning on baseline confounder L_0.

6. Estimate the observational-analog of the ITT effect of the treatment using IPW estimation of a marginal structural model (MSM). Follow these steps:

   (a) Specify the MSM you are targeting.
   We are targeting the estimation of this model:

   $$E(Y^a) = \alpha_0 + \alpha_1 a.$$

   (b) Use unstandardised weights to estimate the ITT.
   The unstandardised weights have mean 2 because we are doubling up the population. The estimated ITT is 1.108 (1.038,1.178) which is close to the true ITT=1.10.

   (c) Use standardised weights to estimate the ITT.
   The standardised weights have mean 1 because now we are standardizing

the original weights to the original population size. The estimated ITT is again 1.108 (1.038,1.178). The reason why they are so close is that the logistic regression model for the denominator of the unstabilised weights produces very stable weights anyway.

7. Estimate the observational-analog of the ITT effect of the treatment using g-computation. If using Stata you can use the `teffects` command (see separate Stata code).

   The same estimate is found using g-computation, although the confidence interval is tighter: 1.108 (1.050, 1.165). This is because more parametric assumptions are made by g-computation and, if appropriate, they lead to greater precision.

8. Estimate the observational-analog of the PP effect of the treatment using IPW. This involves using the standardised weights of question 6c multiplied by the adherence weights described in the lecture (see separate Stata and R codes for guidance).

   The mean adherence weight is 1.8 which is the inverse of the prob(adherence) overall (=1/0.5611). The estimated PP is 1.986 (1.895, 2.076) which is close to the true PP=2.

9. Interpret all results.

   In this setting where there is a substantial lack of adherence (just over half of the patients stick to their original therapy) the ITT and the PP estimated effects are quite different. Note that the estimated coefficient from standard regression analysis is also an estimate of the ITT because the continuous outcome was modelled using linear regression and for these models conditional and marginal coefficient s are the same if there are no interaction between exposure and confounders (as is the case in this generated dataset.