Practical for "Causal Learning" on the UBRA Data Train – December 2025

Vanessa Didelez, Luca Bergen, Ronja Foraita

Part 2 (Thursday afternoon)

Consider the 'breastcancer' data on 2982 women diagnosed with primary breast cancer – this is file `bcrot`. This was an observational study and interest lies in estimating the average causal effect of hormonal therapy ('hormon' binary yes/no) on a synthetic outcome 'qol' (health related quality of life, continuous measure). The following covariates are included: 'age' = age at diagnosis; 'nodes' = number of positive lymph nodes (exp-transformed to 'enodes'); 'pr_1' Progestorone receptors (fmol/l) (log-transformed).

- Compare descriptively QOL for those who do and do not take hormonal therapy.
- Carry out some descriptive analyses (including overlap plots using a flexible model for the propensity score) to investigate the positivity assumption. Convince yourself that it makes sense to restrict the sample to those >= 40 years and nodes > 0.
- Use IPTW to estimate the average causal effect in the restricted population ("by hand" or with R-package `ipw`); investigate extreme weights and consider covariate balance of the re-weighted sample (R-package `cobalt`, function `love.plot`).
- Use regression standardisation to estimate the average causal effect in the restricted population (R-package `stdReg`) choosing a flexible outcome model.
- Finally, use AIPTW to estimate the average causal effect in the restricted population (R-package `AIPW` with default settings and super-learner). Investigate the covariance balance for the re-weighted sample.
- Compare all the above estimates including their precision (take the behaviour of the IPT-weights into account).

For your information, the true ACE on the restricted population is 2.07.

Note that the covariates are from a real dataset, but the outcome 'qol' in the above data is fictitious!

You find example R code by typing the following in R:

```
> library(DataTrainCausalLearning)
> openCode("Causal2_code.html")
```