Causal Inference from Epidemiologic Data

Chapter 4. Ignorable assignment and propensity score.

1 Studies we address with this template.

We now expand interest to large population. We assume that variables used to decide assignment are measured, but we do not necessarily know the function of assignment.

Note. Appropriate for treatments that units adhere to recommendations, i.e., no non-compliance.

2 Role of models for potential outcomes.

cc. Context.

Patients: coronary artery disease.

Treatment Z: bypass ($Z = 1$) versus medical ($Z = 0$) therapy.

Outcome $Y_i(1), Y_i(0)$: months survived after the treatment.

Covariates X.

(1) Likelihood mode of inference: definition and elements (see suppl. notes).

(2) Small number of covariates.

(a) Suppose we have a single covariate X with k levels, e.g., X is a risk index and that X makes assignment ignorable, i.e.,

$$\text{pr}(Z_i = 1 \mid X_i, Y_i(0), Y_i(1)) = \text{pr}(Z_i = 1 \mid X_i).$$

How to estimate $Q = E(Y_i(1) - Y_i(0))$.

Fig. 1

We have

$$E(Y(1)) = \sum_k E(Y^{obs} \mid x_i = k, Z = 1)\text{pr}(X_i = k)$$

Denote n_k as the number of people in cell $X_i = k$; $\hat{y}_{k,z}$ as the sample average of Y^{obs} among people in the cell $X_i = k$ and $Z_i = z$.

Estimate $E(Y_i(1))$ by a consistent estimator $\sum_k \hat{y}_{k,1}\frac{n_k}{n}$. Then an estimator for Q is

$$\hat{Q}^{adj} = \sum_k (\hat{y}_{k,1} - \hat{y}_{k,0})\frac{n_k}{n} \qquad \text{(formula1)}$$

(direct adjustment by subclassification)

Note. Formula 1 is **not** generally applicable for a **non-linear** contrast of the potential outcomes.

(b) Suppose X is still scalar and still make the assignment ignorable, but X is continuous.

Fig. 2

Subclassification: split X into k classes. Then for class k, define $n_k, \hat{y}_{k,z}$ as before. An estimator for Q is,

$$\hat{Q}^{adj,k} = \sum_k (\overline{y}_{k,1} - \overline{y}_{k,0})\frac{n_k}{n}$$

Note. $\hat{Q}^{adj,k}$ is generally biased for Q.

(Cochran 1968)

$$R^k = 1 - \frac{E(\hat{Q}^{adj,k}) - Q}{E(\hat{Q}^{adj,1}) - Q}$$

Note. $R^k \approx 90\%$ for $k \geq 5$ for a large class of underlying models.

3 Propensity score.

(1) Definition and main properties of propensity scores (Rosenbaum and Rubin, 1983).

(a) Definition. $e(x) = \text{pr}(Z_i = 1 \mid X_i = x)$.

(b) Main properties.

Property 1. The propensity score $e(x)$ (calculated on X) balances the distribution of all X between the treatment groups, i.e.,

$$\text{pr}(Z_i = 1 \mid X_i, e(X_i)) = \text{pr}(Z_i = 1 \mid e(X_i)). \qquad (1)$$

Property 2. If Z is strongly ignorable given X, then Z is strongly ignorable given $e(x)$, i.e.,

$$(Y_i(1), Y_i(0)) \coprod Z_i \mid X_i \Longrightarrow (Y_i(1), Y_i(0)) \coprod Z_i \mid e(X_i) \qquad (2)$$

Note 1. The propensity score does balance the **observed** covariates, but **does not** generally balance **unobserved** covariates.

Note 2. The propensity score $e(x)$ needs to be estimated.

Note 3. Model on $e(x)$ is a tradeoff versus model on pr$(Y(z) \mid X)$.

Case study. Rosenbaum and Rubin (1984).

(2) Fitting and diagnostics.

(a) Fit a logistic regression:

$$\text{logitpr}(Z_i = 1 \mid X_i) = \beta' X_i$$

by stepwise selection to get a preliminary $e^o(X_i) = \exp(\hat{\beta} X_i)/(1 + \exp(\hat{\beta} X_i))$.

(b) Check the model in (a) to see whether

$$X_i \coprod Z_i \mid e^o(X_i) \qquad (ex.1)$$

Note 1. If (ex.1) holds, $e^o(X_i)$ is not necessary the propensity score, it is a "balancing score". (Rosenbaum and Rubin, 1983).

Note 2. We will create 5 classes on $e^o(X_i)$, $e_i^{o,*} = 1, ..., 5$.

- For every continuous covariate X, do an ANOVA of X on $Z \times e^{o,*}$. If (ex.1) holds, we should expect no main effects of Z or interaction of Z with $e^{o,*}$.
- For main effect of F, $F_{1,df}$.
- For interaction $Z \times e^{o,*}$, $F_{4,df}$.

(c) If a covariate is seriously unbalanced, either include it in the propensity score model or include some of its higher order terms or split the score into more classes.

(d) Before using the propensity score for outcome evaluation, check overlap between treatment groups. If necessary, discard those non-overlapped observations.

Fig. 3.

(3) Using the propensity score to estimate causal effect.

(a) Effect for all, $E(Y(0) - Y(1))$. Estimate using subclassification.
Within subclass of $e_k, (k = 1, ..., 5)$,

- Obtain $n_{k,1}, n_{k,0}$ – the number of units in class K respectively under different treatments.
- Estimate $E(Y(z))(z = 1, 0)$ by $\bar{y}_z^{adj} = \sum_{k=1}^{5} \bar{y}_{k,z} \frac{n_{k,1}+n_{k,0}}{n}$;

Note 1. If some survival statuses are censored randomly within classes k of the propensity score and treatment Z, then one can estimate $E(Y_i(1) \mid e_i = k)$ by a Kaplan-Meier estimator.

Note 2. A practical variance for $\bar{y}_z^{adj}$ is $\sum_{k=1}^{5} var(\bar{y}_{k,z})(\frac{n_{k,1}+n_{k,0}}{n})^2$.

(b) Effect for the "treated" units, $E(Y_i(0) - Y_i(1) \mid Z_i = 1)$.

Fig. 4

We have,

$$E(Y_i(1) \mid Z_i = 1) = E(Y_i^{obs} \mid Z_i = 1);$$

$$E(Y_i(0) \mid Z_i = 1) = E(E(Y_i(0) \mid Z_i = 1, e_i) \mid Z_i = 1). \qquad (ex.2)$$

Assuming ignorability on X_i, we have ignorability on $e(X_i)$, so RHS of (ex.2) becomes $E(E(Y_i(0) \mid Z_i = 0, e_i) \mid Z_i = 1)$. Therefore,

$$E(Y_i(0) \mid Z_i = 1) = \sum_k E(Y_i^{obs} \mid Z_i = 0, e_i = k)\text{pr}(e_i = k \mid Z_i = 1).$$

(c) Estimating effects of treated units using matching on this propensity score.

- For every treated unit i, get e_i and find a control unit $c(i)$ with $e_{c(i)}$ closely matching e_i.
- Then estimate $E(Y_i(1) - Y_i(0) \mid Z_i = 1)$ with average $y_i^{obs} - y_{c(i)}^{obs}$ over the treated units.

Note. Consider two versions of the above estimate,

- One using the true (unknown) propensity score e_i;
- One using the estimated propensity score $e(\hat{\beta}, X_i)$ based on a model

$$\text{logitpr}(Z_i = 1 \mid X_i, \beta) = \beta' X_i.$$

In general settings, estimates from the latter one are more precise than that from the former one (Rubin and Thomas, 1992,1996).

(d) Interaction of treatment with covariates.

(i)-(iii) See appendix.

(iv) Comparison of propensity score to other methods.

Case study. Lalonde (1986), Deheja and Wahba (1999).

- Interest lies in $E(Y_i(1) - Y_i(0) \mid Z = 1)$.
- Two approaches: (1) models for the potential outcomes **without** using propensity score; (2) models for the potential outcomes **with** using propensity score.

Note 1. Must set estimand as the effect for the treated.

Note 2. Approaches that combine propensity score with some regressions are generally better.

Appendix

Estimating treatment-covariate interaction.

We assume $(Y_i(0), Y_i(1)) \coprod Z_i \mid F_i, B_i$, and want to estimate

$$\text{pr}(Y(1) = 1 \mid F_i = 1) - \text{pr}(Y(0) = 1 \mid F_i = 1).$$

One right way of estimating:

(a) Model

$$\text{pr}(Y(z) = 1 \mid F_i = f, B_i = b) = f(z, f, b).$$

We can estimate this model by regression Y_i^{obs} on F_i, B_i, Z_i. We have,

$$\begin{aligned} & \text{pr}(Y_i(z) = 1 \mid F_i = 1) \\ = & \sum_b \text{pr}(Y_i(z) = 1 \mid F_i = 1, B_i = b)\text{pr}(B_i = b \mid F_i = 1) \\ = & \sum_b f(z, 1, b)\text{pr}(B_i = b \mid F_i = 1) \end{aligned}$$

Note. We can use (ex.3) to estimate $\text{pr}(Y_i(z) = 1 \mid F_i = 1)$, but sensitive to model specification.

(b) Using the propensity score.

If $e(B, F)$ is the propensity score given B, F, then $e(B, 1)$ is the propensity score given $B, F = 1$.

We can,

(1) estimate $e(B, F)$;

(2) stratify on $F = 1$;

(3) do subclassification on $e(B, 1)$, i.e., for units with $F = 1$;

(4) the formula for subclassification will estimate pr$(Y_i(z) = 1 \mid F_i = 1)$.

For another right way, see problem set 1.