Causal Inference from Epidemiologic Data

Chapter 3. Treatment assignment with known and varying probabilities.

1 Assumption of assignment: $q(z^*)$ are known or can be estimated, but $q(z^*)$ are not necessary to equal across units, where pr$(\underline{Z} = z^* \mid \underline{Y}(0), \underline{Y}(1)) = q(z^*)$.

2 Fisher's mode of inference, revisited.

(1) Procedure.

(a) Assign treatments.

(b) Assume null hypothesis, $H_0 : Y_i(1) = Y_i(0)$; and fill in the missing potential outcomes.

(c) Choose a statistic, e.g., $\overline{Y}_0 - \overline{Y}_1$ with observed value, $(\overline{y}_0 - \overline{y}_1)^{obs}$.

(d) List all possible assignment. The probability of assignment z^*, pr$(\underline{Z} = z^*)$ is $q(z^*)$.

(e) Get a p-value: $\sum q(z*)$ over all assignments z^* that are equally or more extreme than the observed one.

Note 1. The p-value defined in (e)is a function of assignment Z and observed outcomes $Y(Z)$, thus can be written as $\alpha(Y(Z), Z)$. It is a valid p-value in the sense that

$$\text{pr}(\alpha(Y(Z), Z) \leq 5\% \mid \text{if null is true}) = 5\%.$$

Note 2. If the null is rejected, the p-value $\alpha(Y(Z), Z)$ is not reasonable to decide what treatment is better and can lead to systematically wrong conclusions.

3 Neyman's mode of inference revisited.

(1) Procedure.

(a) Select an estimand of interest, e.g., $Q = \overline{Y}_0 - \overline{Y}_1$.

(b) Get an unbiased estimate of Q.

Note. $(\overline{y}_0 - \overline{y}_1)^{obs}$ is biased for $\overline{Y}_0 - \overline{Y}_1$, because here the assignment probabilities are not equal across units as in Fisher's mode.

(c) What is an unbiased estimator of Q? Horvitz-Thompson estimator.

Denote $e_i = \mathrm{pr}(Z_i | Y_i(0), Y_i(1))$. A Horvitz-Thompson estimator $\hat{Q}^{H-T}$ of Q is:

$$\frac{1}{2n} \sum_{i=1}^{2n} \frac{Y_i(1) Z_i}{e_i} - \frac{1}{2n} \sum_{i=1}^{2n} \frac{Y_i(0)(1 - Z_i)}{1 - e_i}$$

Note 1. H-T estimator is unbiased for causal effect Q.

Note 2. H-T estimator is generally associated with large variance.

Note 3. One can use covariates to increase precision of H-T estimator.