

STA 320: Design and Analysis of Causal Studies
Homework 5. Due: Wednesday 9, April 7th, 2014

In this problem set, analyze the dataset `brscn320.txt` (available on Sakai under the directory of Resources/Data) on racial disparity in breast cancer screening. (1 point each)

1. *Propensity score approach with random effects models.*

- (a) Estimate the propensity score ($\Pr(W = 1|X)$, $W = 1$ indicates Black) using the a random effects model with all the covariates. Check the overlap of covariates. That is, draw the histograms or density plots of the propensity score for Blacks and Whites in one graph, and describe the overlap between the two groups.
- (b) Estimate the racial disparity between Blacks and Whites using the cluster-weighted estimator, given the estimated propensity score from 1(a).
- (c) Estimate the propensity score ($\Pr(W = 1|X)$, $W = 1$ indicates Black) using the a random effects model with only the individual level covariates, and use the estimated propensity score to estimate the racial disparity between Blacks and Whites using the cluster-weighted estimator.

2. *Propensity score approach with fixed effects models.*

- (a) Repeat Problem 1(a), using instead a fixed effects model.
- (b) Repeat Problem 1(b), using the propensity score estimated from 2(a).
- (c) Can we (or do we need to) repeat Problem 1(c) here using a fixed effects model? Why or Why not?

3. *Direct outcome regression approach:*

- (a) Fit the following fixed effects model for the outcome with all the covariates and report the estimate of γ , which can be interpreted as estimate of “disparity”. Do we need the cluster-level covariates here? (Note: Strictly a linear model is not proper here because the outcome is binary; a logistic model is more suitable. Nevertheless, the linear model is adequate here given the large sample size, and also for the convenience of comparison.)

$$Y_{hk} = \eta_h + W_{hk}\gamma + U_{hk}\beta + \epsilon_{hk} \tag{1}$$

- (b) Fit the following random effects model for the outcome with all the covariates and report the estimate of γ .

$$Y_{hk} = \eta_h + W_{hk}\gamma + X_{hk}\beta + \epsilon_{hk}, \quad \eta_h \sim N(0, \sigma_\eta^2) \quad (2)$$

- (c) Fit the random effects model in 3(b) with only the individual level covariates and report the estimate of γ .
4. Compare the results from all the above methods and comment.
5. (Extra credit) How would you calculate the standard errors of the estimates in the propensity score approach? You do not need to perform the analysis, just state what method you would use.