

# Project 1

This project involves the menstruation project that we discussed in class. The paper can be downloaded [here](#). The project below uses a random sample of the whole experiment due to confidentiality issues with limited variables.

The variables are as follow:

1. “redeem\_status”: redemption status (outcome variable)
  2. “discount”: discount rates [10, 25, 50, 75, 90]
  3. “treatment”: awareness treatment intervention (0 for this project)
  4. “religion”: 1 for Hindu and 0 for others
  5. “brahmin”: caste 1
  6. “chhetri”: caste 2
  7. “janjati”: other castes
  8. “age”: respondent’s age
  9. “marital”: married 1 otherwise 0
  10. “girls\_number”: number of girls in the household
  11. “boys\_number”: number of boys in the household
  12. “father\_educa”: father’s education
  13. “mother\_educa”: mother’s education
  14. “ever\_school”: if ever gone to ever\_school 1 otherwise 0
  15. “highest\_educa”: highest level of schooling
  16. “kitchen\_mens”: disallowed in the kitchen during menstruation
  17. “holyplace\_mens”: disallowed in holy places during menstruation
  18. “shed\_mens”: kept in shed during menstruation
  19. “cause\_hormones”: 1 if a respondent correct picked cause as hormones
  20. “source\_uterus”: 1 if a respondent correct picked source as uterus
  21. “use\_sanitarypad”: 1 if ever used sanitary pads, 0 otherwise
  22. “sanitarypad”: have attended other programs involving sanitary pads
  23. “frequentuse”: frequently use sanitary pads
- a. Note that this experiment consist of multiple arms. In order to simplify things, we’ll focus on those who did not receive awareness treatment.

Given the data set, there should be five discount levels [10, 25, 50, 75, 90 %]. For this project, we’ll consider high versus low discount. Here, high discount would refer to those with dicount values 75% and 90% discount coupons; otherwise the respondent is considered in the low discount category. Please create this variable. Note that this is the “treatment” variable for your project.

- b. Perform a balance exercise. Simply put, you are comparing the means of the covariates between the treated versus the control groups and checking whether the means are statistically different. Note that you can do this using a regression.

$$X_i = \alpha + \kappa treat_i + \epsilon_i.$$

Here,  $X$  is the covariate and  $treat$  is the treatment (high vs low discount). The standard error will help infer whether the difference is statistically significant at the conventional levels.

- c. Comment on the balance exercise from part b.
- d. Evaluate the treatment effect. Comment on the nature of the estimated treatment effect (e.g., ATT, ATE) as well as its magnitude.

hint: `summary(lm(redeem_status ~ treat + sanitarypad, data))`

- e. Next, include the following covariates linearly on your regression specification in part d. `sanitarypad`, `+religion` + `brahmin` + `chhetri` + `use_sanitarypad` + `frequentuse` + `shed_mens` + `kitchen_mens` + `source_uterus`
- f. Consider a more elaborate specification with all the covariates in e plus `highest_educa`, `source_uterus`, `cause_hormones`. You are throwing in the knowledge-related variables in the specification.

Comment on the treatment effect. Is it drastically different from the one that you estimated in d?

- g. Build a probit model to predict the probability of treatment. Use all the covariates that you used to build the model specification in f. Note that you are trying to estimate the propensity scores.
- h. Plot this histogram of the estimated propensity score in g by the treatment status. Comment on the overlap.
- i. Estimate the treatment score using the IPW approach.