Assignment 8

EDH7916 | Summer C 2020

Benjamin Skinner

Using the `els_plans.dta` data set and the abbreviated code book from the lesson, answer the following questions. You do not need to save the final output as a data file: just having the final result print to the console is fine. You can account for missing values by dropping them.

For each question, show your data work and then answer the question in a short (1-2 sentence(s)) comment.

Questions

1. Are reading scores significantly different between students who plan to attend college and those who don't?
   1. Compute the correlation between reading scores and college plans variable, unweighted
   2. Compute using a `t.test()`, unweighted
   3. Recompute using the survey weight (`bystuwt`) and `svyttest()` (don't forget to set up your `svydesign()` first)
   4. Compute using a linear model, `lm()`, unweighted
   5. Recompute using the survey weight (`bystuwt`) and `svyglm()`
2. Regress student's college plans, `plan_col_grad`, on base year socioeconomic status (`byses1`), gender (`female`), mother's BA attainment (`moth_ba`), father's BA attainment (`fath_ba`), and low income status (`lowinc`). Store the fit in an object called `fit` and show the full results using `summary()`.
3. Using the fit object, generate predicted values. If you used `glm()` to account for the binary response, then use the argument `type = "response"` in `predict()` to get predicted probabilities. Store these values in an object and make a histogram of the range of predicted responses.
4. **BONUS** What's the marginal difference in plans to attend college due to low income status (holding all other values at their means)?

Submission details

- Upload your script (`<lastname>_assignment_8.R`) to our CANVAS course site no later than 11:59 p.m. EDT on the due date.