

36-402 Homework 10

Eu Jing Chua

eujingc

April 14, 2019

Question 1

We can check if the deltas are correct by adding them back to the pre-test values and seeing if they match the post-test values.

Table 1: Check for Pre-test + Delta = Post-test

x	
let	TRUE
body	TRUE
form	TRUE
numb	TRUE
relat	TRUE
clasf	TRUE

Question 2

Q2 a)

Table 2: Estimates for deltalet

	Estimate	SE
Regular watchers mean	13.220	0.810
Irregular watchers mean	2.481	0.918
Difference in means	10.739	1.224

Q2 b)

In order for this difference in means to be a sound estimate of the causal effect of switching from, there must be no other confounding sources that affect the subjects' knowledge of letters and whether they are regular watchers or not. This may not be realistic, as other variables such as age and social background could affect their knowledge of letters.

We could test this by using a linear regression model of `deltalet` against `regular`, TODO

Question 3

Q3 a)

Table 3: Coefficients and SE of linear regression

	Coefficient	SE
(Intercept)	-5.3100	5.5100
factor(regular)1	8.0500	1.8100
factor(site)2	7.4900	2.0600
factor(site)3	-4.0200	1.7200
factor(site)4	-1.1900	1.7200
factor(site)5	1.4200	2.4500
factor(sex)2	1.0700	1.2000
age	0.1830	0.1170
factor(setting)2	0.2070	1.3300
factor(encour)1	0.9680	1.7000
peabody	-0.0145	0.0473
prelet	-0.5360	0.1060
prebody	0.0524	0.1380
preform	0.3870	0.2660
prenumb	0.1860	0.1240
prerelat	-0.0123	0.3040
preclasf	-0.0707	0.1930

Q3 b)

`id` should not be included in the regression as it is simply the ID number of the subject, having no relationship at all to the study besides identifying subjects.

`viewcat` should not be included too as the other covariate `regular` is a direct indicator of `viewcat`. Since `regular` is directly derived from `viewcat`, including both would be redundant and introduce problems with highly correlated covariates in linear regression.

Similarly, we exclude all the `post` variables as it is essentially the same as what we want to predict, as the `post` variables are the result of `pre` variables added with the `delta` variables. If we already knew the `post` variables, we would not be predicting anything useful or new.

Q3 c)

Someone who only took 401 might report that the average effect of making a child become a regular watcher of Sesame Street is an increase of 8.05 in score of the letter test.

Q3 d)

To infer the causal effect of becoming a regular watcher of Sesame Street on the change in score of the letter test based on the above model, we would first need to assume there are no other confounding sources between the two variables. Additionally, we also need to assume that all the additional covariates we are controlling do not create new confounding sources by controlling for them. This is plausible but highly unlikely as including everything blindly increases the chances of creating new confounding sources.

Question 4

Q4 a)

The set of variables are **setting** and **site**.

Q4 b) Using a kernel regression with cross-validated bandwidths,

Table 4: Average treatment effect of regular watching

	x
Average treatment effect	8.6600
SE	0.0334