19-3

Nikhil Gopal

3/14/2022

19.4 *Pre-test and post-test:* 100 students are given a pre-test, then a treatment or control is randomly assigned to each, then they get a post-test. Given the following regression model:

```
post_test = a + b * pre_test + theta * z + error
```

where z = 1 for treated units and theta for controls. Further suppose that pre_test has mean 40 and standard deviation 15. Suppose b = 0.7 and theta = 10 and the mean for post_test is 50 for the students in the control group. Further suppose that the residual standard deviation of the regression is 10.

- a) Determine *a*
- b) What is the standard deviation of the post-test scores for the students in the control group?
- c) What are the mean and standard deviation of the post-test scores in the treatment group?

19.5

Causal inference using logistic regression: Suppose you have fit a model:

```
fit <- stan_glm(y ~ z + age + z:age, family=binomial(link="logit"), data=mydata)
```

with binary outcome y, treatment indicator z, and age measured in years. Give R code to produce an estimate and standard error of average treatment effect in a large population, given a vector n_pop of length 82 that has the number of people in the population at each age from 18 through 99:

```
predictions$effect <- predict(fit, n_pop)

avg_effect <- mean(exp(predictions$effect))

st_err <-sd(predictions$effect)/sqrt(82)
```