

## 19-3

Nikhil Gopal

3/14/2022

```
rm(list = ls())
library(rstanarm)
library(rosdata)
library(tidyverse)
```

### 19.3

Gain scores: In the discussion of gain-score models in Section 19.3, we noted that if we include the pre-treatment measure of the outcome in a gain score model, the coefficient on the treatment indicator will be the same as if we had just run a standard regression of the outcome on the treatment indicator and the pre-treatment measure. Show why this is true.

This data set comprises data from classrooms, that measure effect of the Electric Company educational television program on child reading scores. The treatment column indicates whether students were assigned to a class that watched the show, and the pre\_test and post\_test columns give the score of each class room at the beginning and end of the study.

To investigate the claim in the question, I made a gain\_score column by subtracting the pre treatment test score from the post treatment test score.

```
df <- rosdata::electric

df$gain_score <- df$post_test - df$pre_test

#gain score model with pre treatment indicator as the outcome
fit_gain <- stan_glm(gain_score ~ pre_test + treatment, data = df, refresh = 0)

#standard regression of outcome onto the treatment indicator and pre-treatment measure
fit_standard <- stan_glm(post_test ~ pre_test + treatment, data = df, refresh = 0)
```

```
df %>% group_by(treatment) %>%
  summarise(improvement = mean(gain_score))
```

```
## # A tibble: 2 x 2
##   treatment improvement
##       <int>       <dbl>
## 1         0        23.1
## 2         1        26.8
```

I fit two regression models, one where I regressed the pre treatment score and the treatment variable onto the gain score, and one standard regression where I regressed the pre treatment score and the treatment variable onto the post treatment score:

```
print(fit_gain)
```

```
## stan_glm
## family:      gaussian [identity]
## formula:     gain_score ~ pre_test + treatment
## observations: 192
## predictors:   3
## -----
##              Median MAD_SD
## (Intercept) 61.5      1.5
## pre_test    -0.5      0.0
## treatment    4.7      1.2
##
## Auxiliary parameter(s):
##           Median MAD_SD
## sigma 8.1      0.4
##
## -----
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

```
print(fit_standard)
```

```
## stan_glm
## family:      gaussian [identity]
## formula:     post_test ~ pre_test + treatment
## observations: 192
## predictors:   3
## -----
##              Median MAD_SD
## (Intercept) 61.6      1.5
## pre_test     0.5      0.0
## treatment    4.7      1.1
##
## Auxiliary parameter(s):
##           Median MAD_SD
## sigma 8.1      0.4
##
## -----
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

We see that in both models, the coefficient on the treatment indicator is 4.7. Thus, we can confirm that the claim in the question is true.