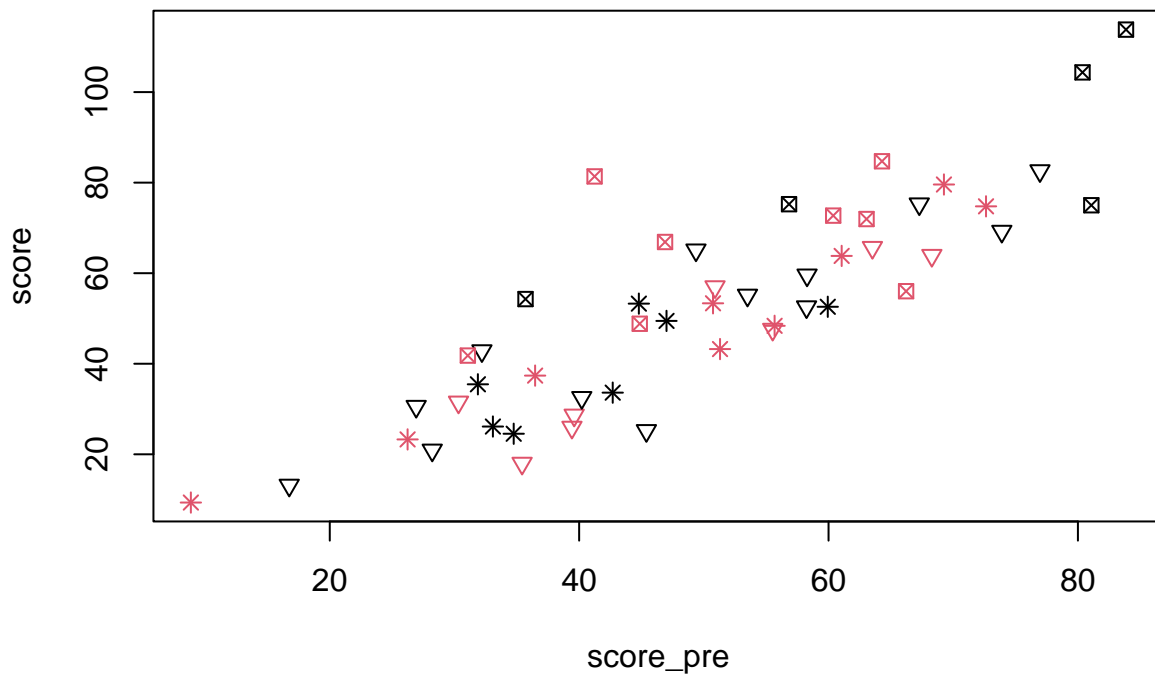# Assignment 6a

Satvik Saha

2024-10-08

Answer 1

```r
library(rstanarm)

n <- 50
score_pre <- rnorm(n, mean = 50, sd = 20)
sex <- sample(c(0, 1), n, replace = TRUE)
grade <- sample(c(6, 7, 8), n, replace = TRUE)
score <- 5 + score_pre + 0.0 * sex + ifelse(grade == 7, 5, -5) + 10 * rnorm(n)
df <- data.frame(
  score_pre = score_pre,
  sex = sex,
  grade = grade,
  score = score
)
plot(score_pre, score, pch = grade, col = sex + 1)
```


(a) With the encoding 0 for women and 1 for men, it is easy to read the model without the 'sex' variable as the model for women, and the model without the 'sex' variable but with the coefficient for 'sex' added to the intercept as the model for men. An encoding with 1, 2 for women and men respectively is only slightly more awkward to read off the two models.

1

(b)

```
stan_glm(
  score ~ score_pre + sex + grade,
  data = df,
  refresh = 0
)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      score ~ score_pre + sex + grade
##  observations: 50
##  predictors:   4
## ------
##             Median MAD_SD
## (Intercept) -14.3   14.9
## score_pre     1.2    0.1
## sex          -0.2    3.5
## grade         1.3    2.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 12.3    1.3
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

(c)

```
stan_glm(
  score ~ score_pre + sex + factor(grade),
  data = df,
  refresh = 0
)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      score ~ score_pre + sex + factor(grade)
##  observations: 50
##  predictors:   5
## ------
##                 Median MAD_SD
## (Intercept)     -3.2    5.2
## score_pre        1.0    0.1
## sex             -2.3    3.0
## factor(grade)7  17.2    3.8
## factor(grade)8   1.7    3.5
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 10.1    1.1
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

Answer 2

Consider the model
$$y_i = a + bh_i + cm_i + dh_im_i + \epsilon_i,$$

with $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, and y, h, m denoting the variables _earnings, *height*, and *male*.

We have our estimates $\hat{a} = -9.3, \hat{b} = 0.4, \hat{c} = -29.3, \hat{d} = 0.6$.

(a) Setting $m'_i = m_i - 0.5$, we may rewrite our model as

$$y_i = (a + 0.5c) + (b + 0.5d)h_i + c(m_i - 0.5) + dh_i(m_i - 0.5) + \epsilon_i$$
$$= a' + b'h_i + c'm'_i + d'h_im'_i + \epsilon_i,$$

so our new coefficients are $\hat{a}' = \hat{a} + 0.5\hat{c} = -23.95$, $\hat{b}' = \hat{b} + 0.5\hat{d} = 0.7$, $\hat{c}' = \hat{c} = -29.3$, $\hat{d}' = \hat{d} = 0.6$.

(b) Setting $h'_i = h_i - 66$, we may rewrite our model as

$$y_i = (a' + 66b') + b'(h_i - 66) + (c' + 66d')m'_i + d'(h_i - 66)m'_i + \epsilon_i$$
$$= a'' + b''h_i + c''m'_i + d''h'_im'_i + \epsilon_i,$$

so our new coefficients are $\hat{a}'' = \hat{a}' + 66\hat{d}' = 15.65$, $\hat{b}'' = \hat{b}' = 0.7$, $\hat{c}'' = \hat{c}' + 66\hat{d}' = 10.3$, $\hat{d}'' = \hat{d}' = 0.6$.

(c) Identical to (b)

Answer 3

(a) The estimated regression line for the control group ($z = 0$) can be read as

$$y = 1.2 + 1.6x,$$

and the model for the treatment group ($z = 1$) can be read as
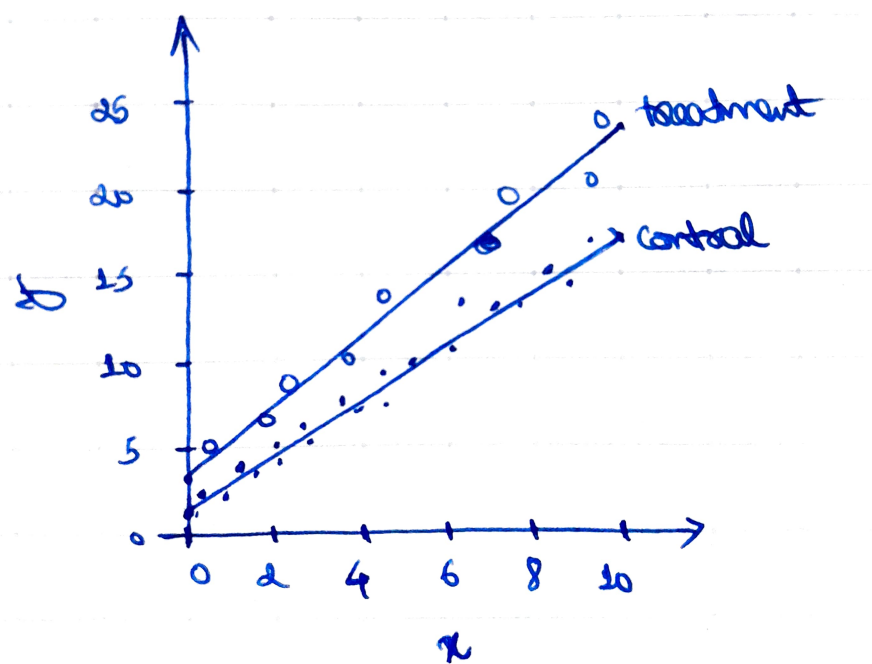
$$y = 1.2 + 1.6x + 2.7 + 0.7x = 3.9 + 2.3x.$$

(b)

Figure 1: Control and treatment regression lines