HW5

Tyler Reed

11/9/2021

6.5

(a)

```
earnings <- read.csv("~/mySTA631/HW5_files/Earnings/data/earnings.csv")
(avg_heights <- earnings %>%
                        select(height, male) %>%
                        group_by(male) %>%
                        summarize(mean = mean(height)))
```

```
## # A tibble: 2 x 2
##    male  mean
##   <int> <dbl>
## 1     0  64.5
## 2     1  70.1
```

```
(avg_earnings_htmale <- -26.0 + (0.6 * avg_heights$mean[2]) + (10.6 * 1))
```

```
## [1] 26.65333
```

```
(avg_earnings_htfemale <- -26.0 + (0.6 * avg_heights$mean[1]) + (10.6 * 0))
```

```
## [1] 12.69185
```

(b)

```
(avg_earn_samplemale <- earnings %>%
                                select(earnk, male) %>%
                                filter(male == 1) %>%
                                summarise(mean = mean(earnk)))
```

```
##       mean
## 1 30.10518
```

```
(avg_earn_samplefemale <- earnings %>%
                                select(earnk, male) %>%
                                filter(male == 0) %>%
                                summarise(mean = mean(earnk)))
```

```
##       mean
## 1 15.84794
```

```
(avg_earn_adults_wt <- (0.52 * avg_earn_samplefemale) + (0.48 * avg_earn_samplemale))
```

```
##       mean
```

Parameters	Values	Parameters	Values
Avg Weighted Earnings by Sex	22.69141	Avg Earnings from Sample	21.15
Avg Earnings by Avg Female Height	12.69185	Avg Female Earnings from Sample	15.85
Avg Earnings by Avg Male Height	26.65333	Avg Male Earnings from Sample	30.11

```
## 1 22.69141
```

(c)

```
(avg_earn_samplepop <- mean(earnings$earnk))
```

```
## [1] 21.1473
```

```
(avg_earn_samplemale <- earnings %>%
                                select(earnk, male) %>%
                                filter(male == 1) %>%
                                summarise(mean = mean(earnk)))
```

```
##       mean
## 1 30.10518
```

```
(avg_earn_samplefemale <- earnings %>%
                                select(earnk, male) %>%
                                filter(male == 0) %>%
                                summarise(mean = mean(earnk)))
```

```
##       mean
## 1 15.84794
```

```
earn_table_ht <- tibble("Parameters" = c("Avg Weighted Earnings by Sex",
             "Avg Earnings by Avg Female Height",
             "Avg Earnings by Avg Male Height"),
       "Values" = c(avg_earn_adults_wt,
                    avg_earnings_htfemale,
                    avg_earnings_htmale))

earn_table_sample <- tibble("Parameters" = c("Avg Earnings from Sample",
             "Avg Female Earnings from Sample",
             "Avg Male Earnings from Sample"),
       "Values" = c(avg_earn_samplepop,
                    avg_earn_samplefemale[1,1],
                    avg_earn_samplemale[1,1]))
```

```
kable(list(earn_table_ht, earn_table_sample), digits = 2)
```

Interpretation Comparing the average earnings by average heights of both males and females and the sample averages of earnings of males and females shows that on average, if a male or female is of average height, both will earn less than the sample average which does not consider height.

Also, the average weighted earnings by sex compared to average earnings direct from the sample is higher on average when accounting for the given difference in population proportions of males and females. The sample weights of male and female are 37.1% and 62.8%, respectively. The lower weights of the sample account for the lower direct sample average earnings by sex than the weighted earnings by sex calculated from the given population parameters.

6.6

(a)

```
heights <-read.table("~/mySTA631/HW5_files/PearsonLee/data/Heights.txt", header=TRUE)
heights_ltmeanmother <- heights %>%
  filter(mother_height < mean(mother_height))
head(heights_ltmeanmother)
```

```
##   daughter_height mother_height
## 1            52.5          59.5
## 2            52.5          59.5
## 3            53.5          59.5
## 4            53.5          59.5
## 5            55.5          59.5
## 6            55.5          59.5
```

```
fit_1 <-stan_glm(daughter_height ~ mother_height, data=heights_ltmeanmother)
```

```
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 3.1e-05 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 0.31 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
## Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 0.033156 seconds (Warm-up)
## Chain 1:                0.170405 seconds (Sampling)
## Chain 1:                0.203561 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 1.2e-05 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 0.12 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2: Iteration:  400 / 2000 [ 20%]  (Warmup)
```

```
## Chain 2: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 2: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 0.028833 seconds (Warm-up)
## Chain 2:                0.170474 seconds (Sampling)
## Chain 2:                0.199307 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 1.1e-05 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 0.11 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 3: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 3: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 0.029666 seconds (Warm-up)
## Chain 3:                0.169961 seconds (Sampling)
## Chain 3:                0.199627 seconds (Total)
## Chain 3:
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 4).
## Chain 4:
## Chain 4: Gradient evaluation took 1.2e-05 seconds
## Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 0.12 seconds.
## Chain 4: Adjust your expectations accordingly!
## Chain 4:
## Chain 4:
## Chain 4: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4: Iteration: 1001 / 2000 [ 50%]  (Sampling)
```

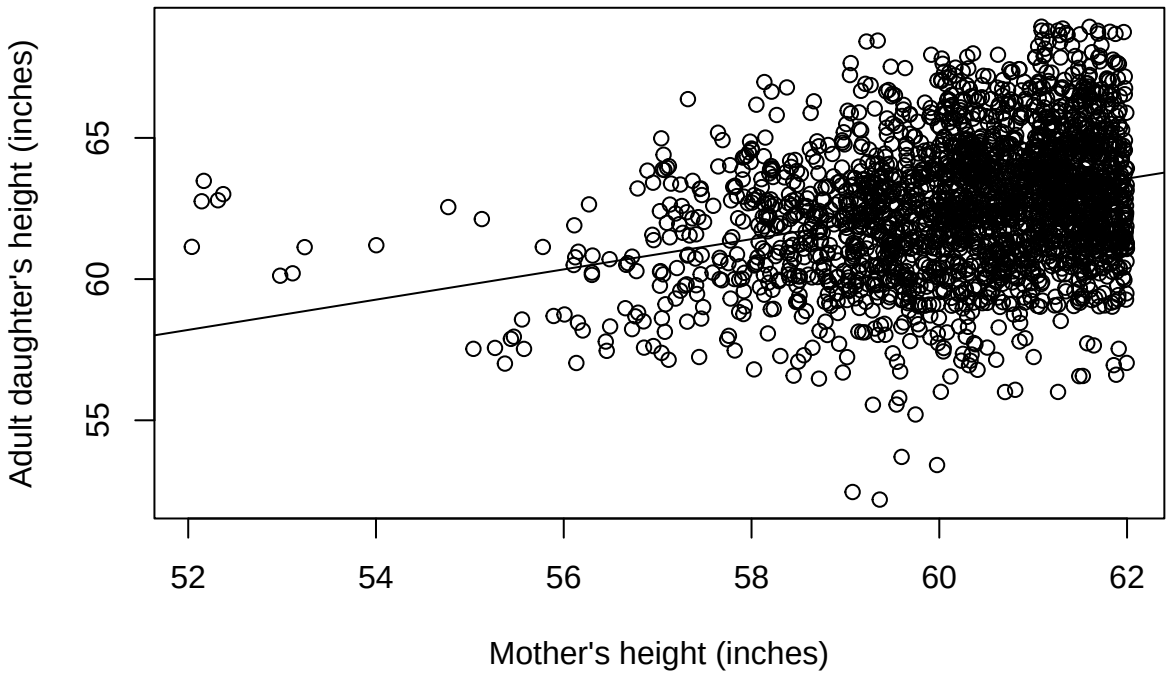
```
## Chain 4: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 4:
## Chain 4:  Elapsed Time: 0.030978 seconds (Warm-up)
## Chain 4:                0.170345 seconds (Sampling)
## Chain 4:                0.201323 seconds (Total)
## Chain 4:
```

```
print(fit_1)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      daughter_height ~ mother_height
##  observations: 2327
##  predictors:   2
## ------
##               Median MAD_SD
## (Intercept)   30.3    2.1
## mother_height  0.5    0.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 2.3    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

```
n <-nrow(heights_ltmeanmother)
mother_height_jitt <-heights_ltmeanmother$mother_height + runif(n, -0.5, 0.5)
daughter_height_jitt <-heights_ltmeanmother$daughter_height + runif(n, -0.5, 0.5)

a_hat <-coef(fit_1)[1]
b_hat <-coef(fit_1)[2]
plot(mother_height_jitt, daughter_height_jitt,
     xlab="Mother's height (inches)",
     ylab="Adult daughter's height (inches)",
     abline(a_hat, b_hat))
```



Observations less than mean `mother_height`

	Median	MAD_SD
(Intercept)	30.4	2.1
mother_height	0.5	0.0

Auxiliary parameter(s):

	Median	MAD_SD
sigma	2.3	0.0

All observations

	Median	MAD_SD
(Intercept)	29.8	0.8
mother_height	0.5	0.0

Auxiliary parameter(s):

	Median	MAD_SD
sigma	2.3	0.0

Interpretation Confirmation that the parameters are about the same.

(b)

```
heights <-read.table("~/mySTA631/HW5_files/PearsonLee/data/Heights.txt", header=TRUE)
heights_ltmeandaught <- heights %>%
  filter(daughter_height < mean(daughter_height))
head(heights_ltmeandaught)
```

```
##   daughter_height mother_height
## 1            52.5          59.5
## 2            52.5          59.5
## 3            53.5          59.5
## 4            53.5          59.5
## 5            55.5          59.5
## 6            55.5          59.5
```

```
fit_1 <-stan_glm(daughter_height ~ mother_height, data=heights_ltmeandaught)
```

```
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 2.2e-05 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 0.22 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
## Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 0.031774 seconds (Warm-up)
## Chain 1:                0.21006 seconds (Sampling)
## Chain 1:                0.241834 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 1.1e-05 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 0.11 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 2: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 2: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2: Iteration: 2000 / 2000 [100%]  (Sampling)
```

```
## Chain 2:
## Chain 2:  Elapsed Time: 0.025517 seconds (Warm-up)
## Chain 2:                0.209422 seconds (Sampling)
## Chain 2:                0.234939 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 1.2e-05 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 0.12 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 3: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 3: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 0.022462 seconds (Warm-up)
## Chain 3:                0.208778 seconds (Sampling)
## Chain 3:                0.23124 seconds (Total)
## Chain 3:
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 4).
## Chain 4:
## Chain 4: Gradient evaluation took 1.2e-05 seconds
## Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 0.12 seconds.
## Chain 4: Adjust your expectations accordingly!
## Chain 4:
## Chain 4:
## Chain 4: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 4: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 4:
## Chain 4:  Elapsed Time: 0.03084 seconds (Warm-up)
## Chain 4:                0.208814 seconds (Sampling)
## Chain 4:                0.239654 seconds (Total)
```

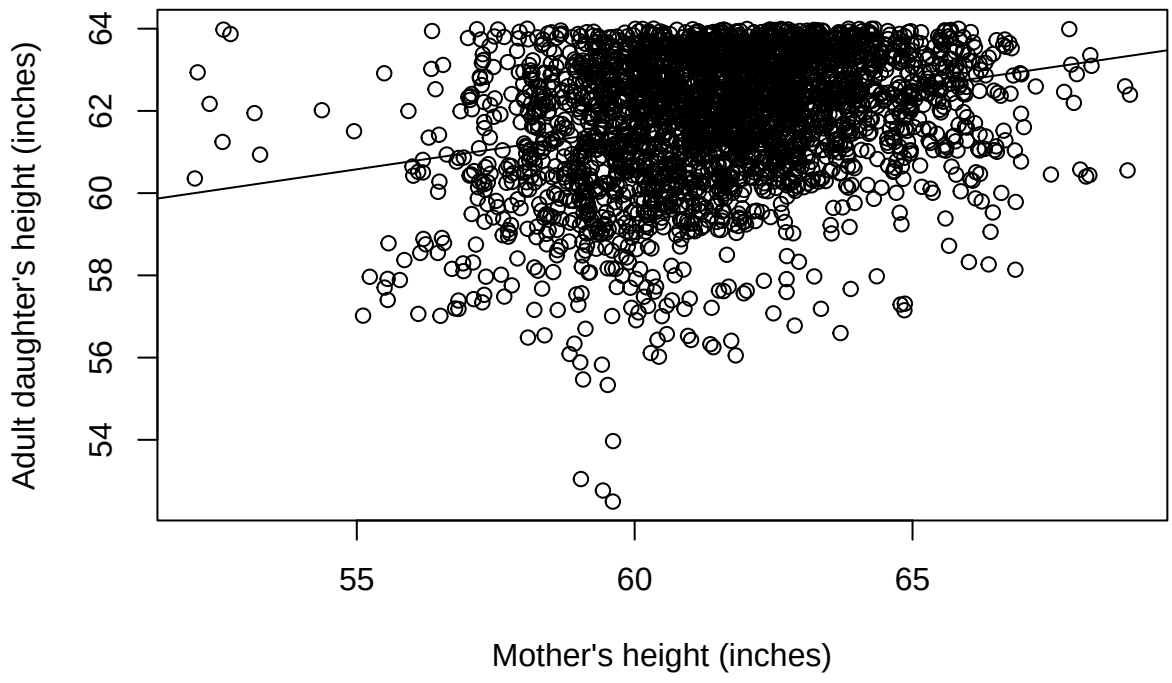
```
## Chain 4:
```

```
print(fit_1)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      daughter_height ~ mother_height
##  observations: 2924
##  predictors:   2
## ------
##               Median MAD_SD
## (Intercept)   49.7    0.8
## mother_height  0.2    0.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 1.5    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

```
n <-nrow(heights_ltmeandaught)
mother_height_jitt <-heights_ltmeandaught$mother_height + runif(n, -0.5, 0.5)
daughter_height_jitt <-heights_ltmeandaught$daughter_height + runif(n, -0.5, 0.5)

a_hat <-coef(fit_1)[1]
b_hat <-coef(fit_1)[2]
plot(mother_height_jitt, daughter_height_jitt,
     xlab="Mother's height (inches)",
     ylab="Adult daughter's height (inches)",
     abline(a_hat, b_hat))
```



7.2

7.3

8.5