exercises_week12

Lindsey Greenhill

4/21/2021

Question 7.1

Part a

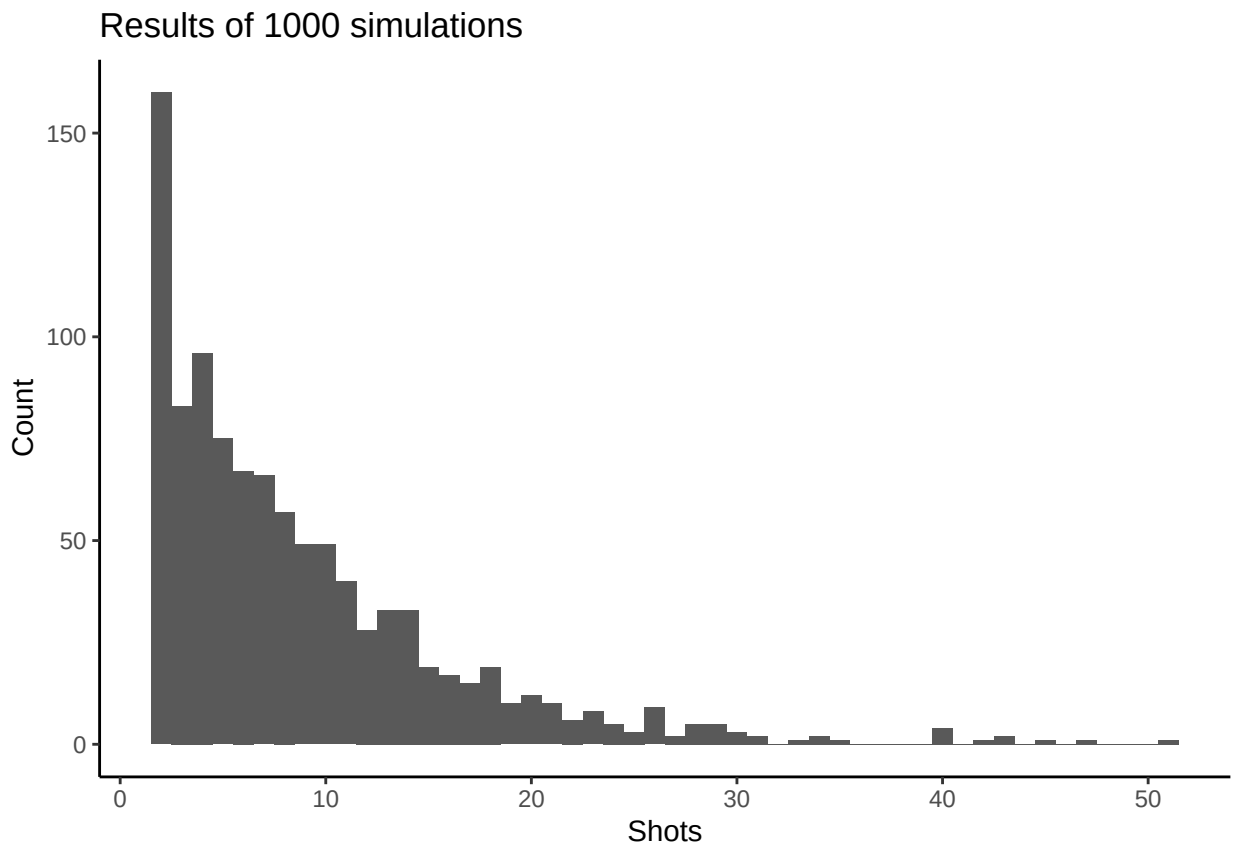
In the code below, I write a function that simulates a person taking a shots until he misses 2 in a row with a 60% probability of making the shot.

```
# write function for taking shots

shoot <- function(){

 # first shot

 shot_1 <- rbinom(n = 1, size = 1, prob = .6)
 bool <- FALSE
 count <- 1
 success <- if_else(shot_1 == 1, 1, 0)
 while(bool == FALSE){

   # next shot

   shot_2 <- rbinom(n = 1, size = 1, prob = .6)
   if((shot_2 == 0) & (shot_1 == 0)){
     count <- count + 1
     bool <- TRUE
   }
   else{
     success <- if_else(shot_2 == 1, success + 1, success)
     count <- count + 1
     shot_1 <- shot_2
   }
 }

 # have to return both so I can calculate prop success for next q

 return(c(count,success))
}
```

Part b

In the code below, I simulate the shoot function written in part a 1000 times. I then plot the results of the simulation and present summary statistics of the distribution.

```
# creating vector to store shot results

results <- tibble()

# simulating the shots 1000 times

for(i in 1:1000){
  r <- shoot()
  vec <- tibble(count = r[1],
                success = r[2])
  results <- results %>%
    rbind(vec)
}

# plotting the results

plot_1_b <- ggplot(results, aes(x = count)) +
  geom_histogram(binwidth = 1) +
  labs(title = "Results of 1000 simulations",
       x = "Shots",
       y = "Count") +
  theme_classic()

# summary table of mean and sd

sum_tbl <- results %>%
  summarise(mean = mean(count),
            sd = sd(count))

sum_1_b <- sum_tbl %>%
  gt()

plot_1_b
```




```
sum_1_b
```

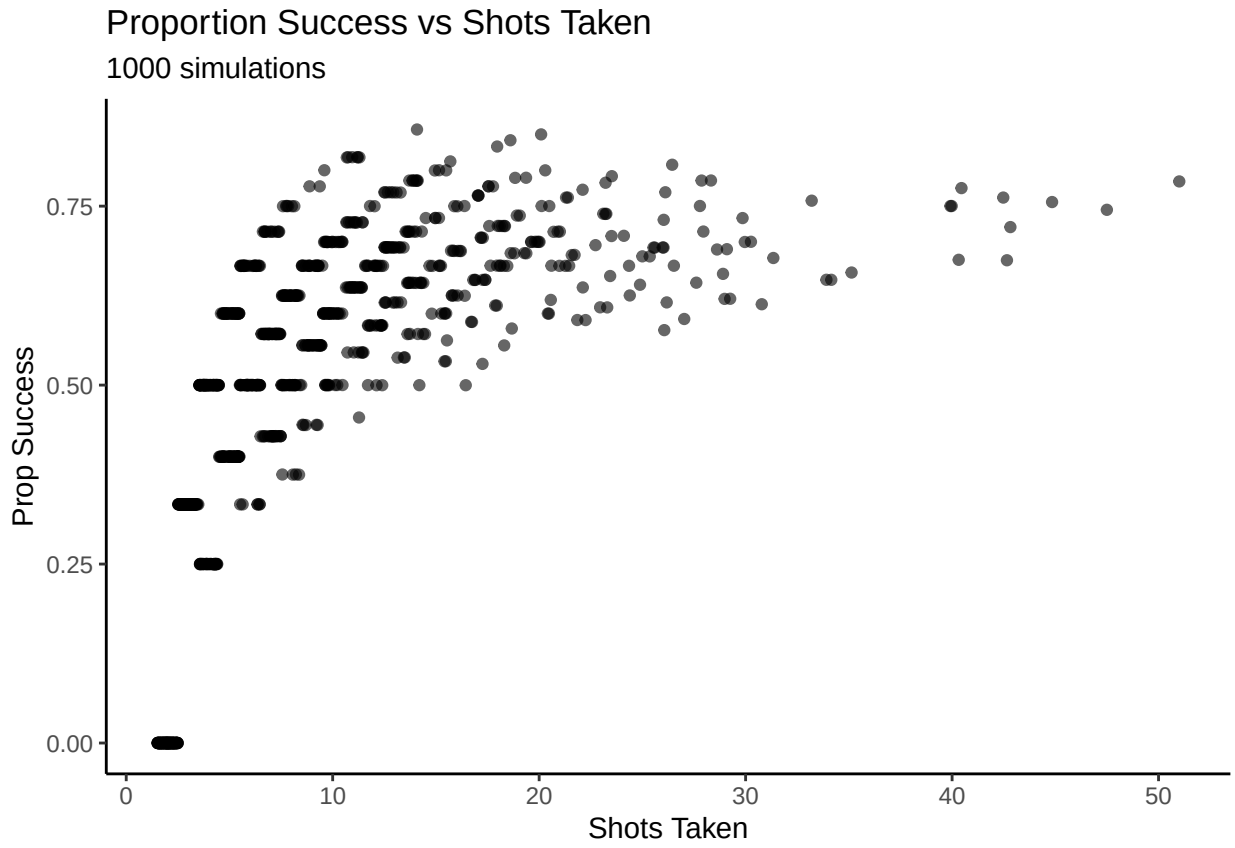
mean	sd
8.843	7.292018

Part c

In the code below, I calculate the proportion of shots made for each shoot function. I then plot the proportion of successes against the shots taken. I used geom_jitter to more clearly see all 1000 data points.

```
# creating prop success col

results <- results %>%
  mutate(prop_success = success/count)

# plotting with gitter and alpha making semi transparent

ggplot(results, aes(x = count,  y = prop_success)) +
  geom_jitter(width = .5,
              alpha = .6) +
  labs(title = "Proportion Success vs Shots Taken",
       subtitle = "1000 simulations",
       x = "Shots Taken",
```

```
       y = "Prop Success") +
  theme_classic()
```



Question 7.4

Part a

In the code below, I use the sim function to simulate 1000 predictions for teacher A and teacher B.

```
# reading in the data

beauty_data <- read.csv("http://www.stat.columbia.edu/~gelman/arm/examples/beauty/ProfEvaltnsBeautyPubli

# cleaning the data

beauty_data <- beauty_data %>%
  select(profevaluation,
               female, courseevaluation,
               btystdave, nonenglish, age,
               profnumber)

# creating model
```

```r
mod <- lm(courseevaluation ~ age + female + nonenglish + btystdave, beauty_data)

# creating data for teacher A and teacher B with teacher A first

age <- c(50,60)
female <- c(1,0)
nonenglish <- c(0,0)
btstydave <- c(-1,-.5)

# pretty much copying from Le's example

n.tilde <- length(age) # n.tilde is the number of predictions

# joining the data together by column to create a matrix

X.tilde <- cbind(rep(1, n.tilde), age, female, nonenglish, btstydave)

# conducting 1000 sims

n.sims <- 1000
sim <- sim(mod, n.sims) # Make simulations


# creating sims

n.sims <- 1000
sim <- sim(mod, n.sims)


# Create an array to save the predicted results.

y.tilde <- array(NA, c(n.sims, n.tilde))

# Predict loop

for (s in 1:n.sims){
  y.tilde[s,] <- rnorm (n.tilde,
                        X.tilde %*% sim@coef[s,],
                        sim@sigma[s]
                        )
}
```

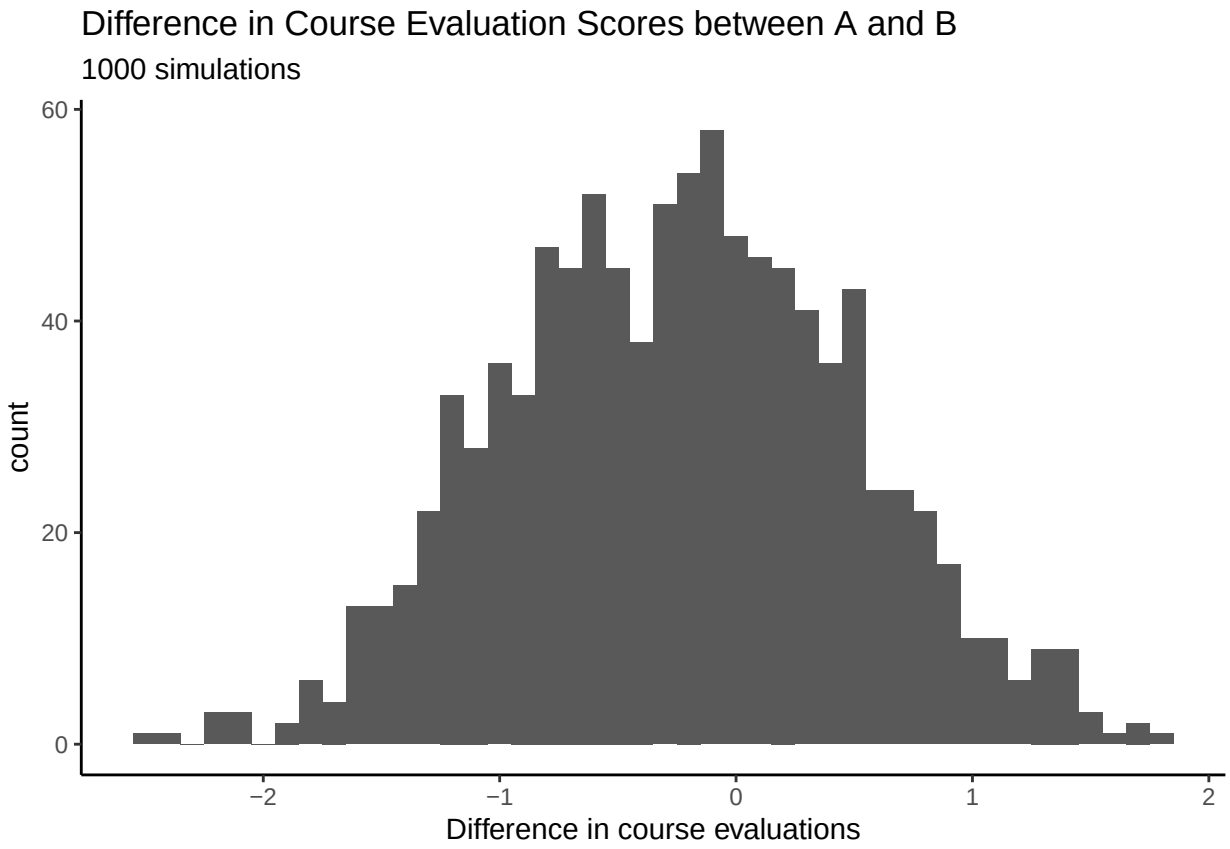
Part b

In the code below, I created a histogram of the distribution of the difference in course evaluation scores of A and B from the 1000 simulations created in part a. I then calculated the probability teacher A's course evaluation score being greater than teacher B's course evaluation score by calculating the proportion of the difference values above 0.

** The probability that A has a higher score than B is 37.1% **

```r
# creating data frame from sim results above. creating new column that is the
# difference between a and b where v1 is a and v2 is b. If diff is positive,
```

```r
# then a had a better course evaluation than b.

new_df <- y.tilde %>%
  as_tibble() %>%
  mutate(diff = V1 - V2)

hist <- new_df %>%
  ggplot(aes(x = diff)) +
  geom_histogram(binwidth = .1) +
  labs(title = "Difference in Course Evaluation Scores between A and B",
       subtitle = "1000 simulations",
       x = "Difference in course evaluations") +
  theme_classic()

# calculating the probability that diff is greater than 0 by calculating the
# proportion of values above 0 in the distribution

prop <- new_df %>%
  mutate(greater = if_else(diff > 0, 1, 0)) %>%
  count(greater) %>%
  pivot_wider(id_cols = everything(),
              names_from = "greater",
              values_from = "n",
              names_prefix = "count") %>%
  summarise(prop = count1/(count0 + count1))

hist
```


Difference in Course Evaluation Scores between A and B
1000 simulations
count
60
40
20
0
−2
−1
0
1
2
Difference in course evaluations