Assignment 3 question 2

Dhanraj Patel

2021-08-08

A)

The parameter p is the probability that the Y_i'th response of ith respondent answers YES to the survey question 'Do you feel optimistic about the future of Canada?'

B)

By CLT

sampling distribution of $\hat{p}$ ~ N(μ,σ^2/n)

n = 1483

μ = 0.47

σ^2 / n = 0.47(1 - 0.47) / 1483 = 0.2491 / 1483 = 0.00016797033

$\hat{p}$ ~ N(0.47,0.00016797033)

c)

(Graph_1)

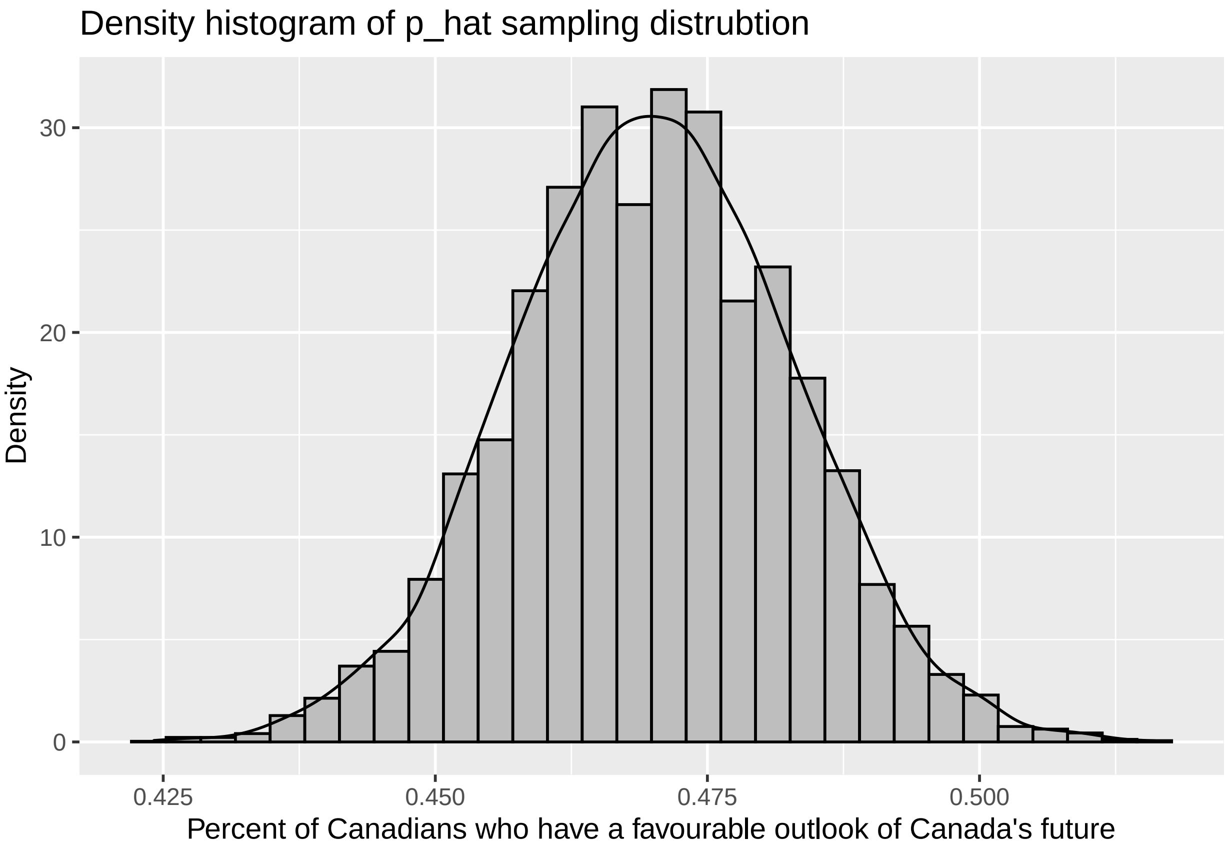
```
bar_estimates <- numeric(10000)
for (i in 1:10000)(
  bar_estimates[i] <- sum(rbernoulli(1483, 0.47)) / 1483
)




To_data_set <- data.frame(x = bar_estimates)



ggplot(To_data_set, aes(x=x)) +
  geom_histogram(aes(y=..density..), bins = 30, colour = "black",fill = "grey") +
  geom_density()+
   labs(title = "Density histogram of p_hat sampling distrubtion",
       x = "Percent of Canadians who have a favourable outlook of Canada's future",
       y = "Density")
```



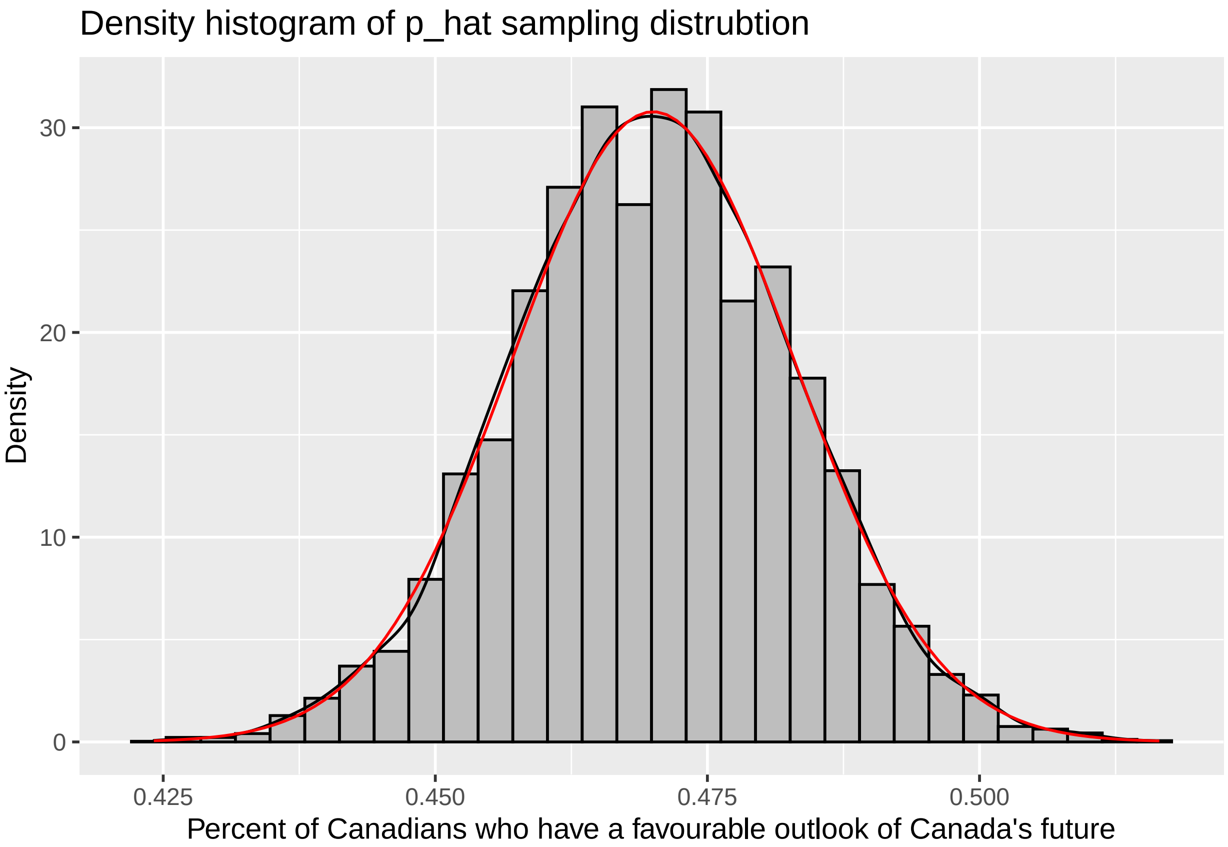
Within (Graph_1), the gray bars represents for the 10000 bootstrap samples, the density of the result of all the samples that consisted of the of the percent of Canadians who have a favorable outlook of Canada's future.

The black line represents the the density of the sampling distribution of the 10000 bootstrap samples

d)

(Graph_2)

```
ggplot(To_data_set, aes(x=x)) +
  geom_histogram(aes(y=..density..),bins = 30,colour = "black",fill = "grey") +
  geom_density()+ labs(title = "Density histogram of p_hat sampling distrubtion",
       x = "Percent of Canadians who have a favourable outlook of Canada's future",
       y = "Density")+
      stat_function(fun = dnorm,
                args = list(mean =0.47, sd = sqrt(0.00016797033)),
                col = "red")
```



Within (Graph_2), the gray bars represents for the 10000 bootstrap samples, the density of the result of all the samples that consisted of the of the percent of Canadians who have a favorable outlook of Canada's future.

The black line represents the the density of the sampling distribution of the 10000 bootstrap samples

The red line represents the overlay of the CLT approximation of the sampling distribution of $\hat{p}$.

e)

```
count2 <- sum(To_data_set$x < 0.5)
count2
```

```
## [1] 9896
```

With 10000 bootstrap samples we can see with the value of count2, the number of the samples that returned results where more than half of those surveyed answered NO to the survey question.

```
probability_to_count <- round(count2/10000,4)
probability_to_count
```

```
## [1] 0.9896
```

From this we determined that the value of probability_to_count (seen above) percent probability that a group of 1483 people surveyed would result in a majority of people voting 'NO' to the survey. From this information we can conclude that there is a very high percent that a group of 1483 people surveyed would result in a majority of people not being optimistic about the future of Canada. The news article's claim is supported by the bootstrap data due to the fact out out of the 1483 people surveyed, less than half answered 'YES', lining up with the bootstrap data where the majority of the bootstrap samples also resulted in less than half of Canadians being optimistic about the outlook in Canada's future.

f)

c repeated with n = 56

(graph_3)

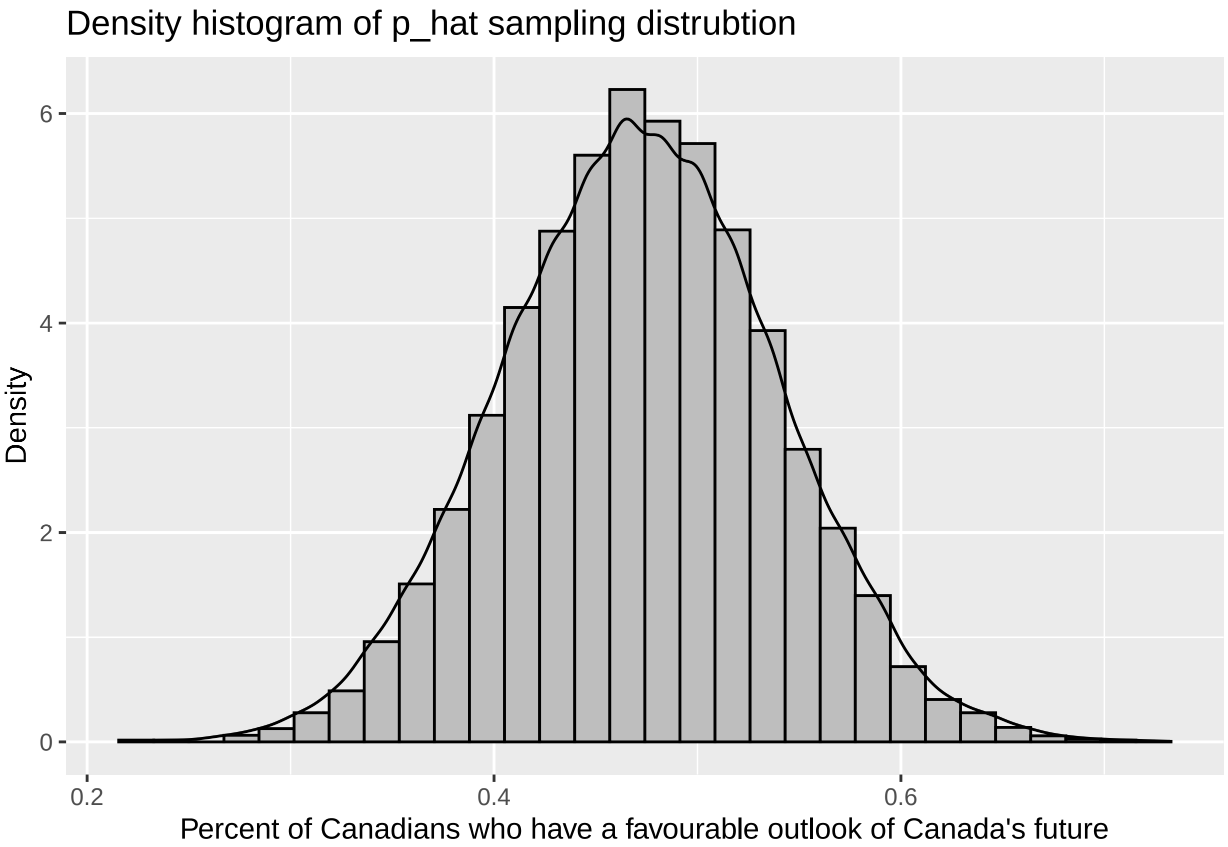
```
bar_estimates2 <- numeric(10000)
for (i in 1:10000)(
  bar_estimates2[i] <- sum(rbernoulli(56, 0.47)) / 56
)




To_data_set2 <- data.frame(x = bar_estimates2)



ggplot(To_data_set2, aes(x=x)) +
  geom_histogram(aes(y=..density..),bins = 30,colour = "black",fill = "grey") +
  geom_density()+
   labs(title = "Density histogram of p_hat sampling distrubtion",
       x ="Percent of Canadians who have a favourable outlook of Canada's future",
       y = "Density")
```



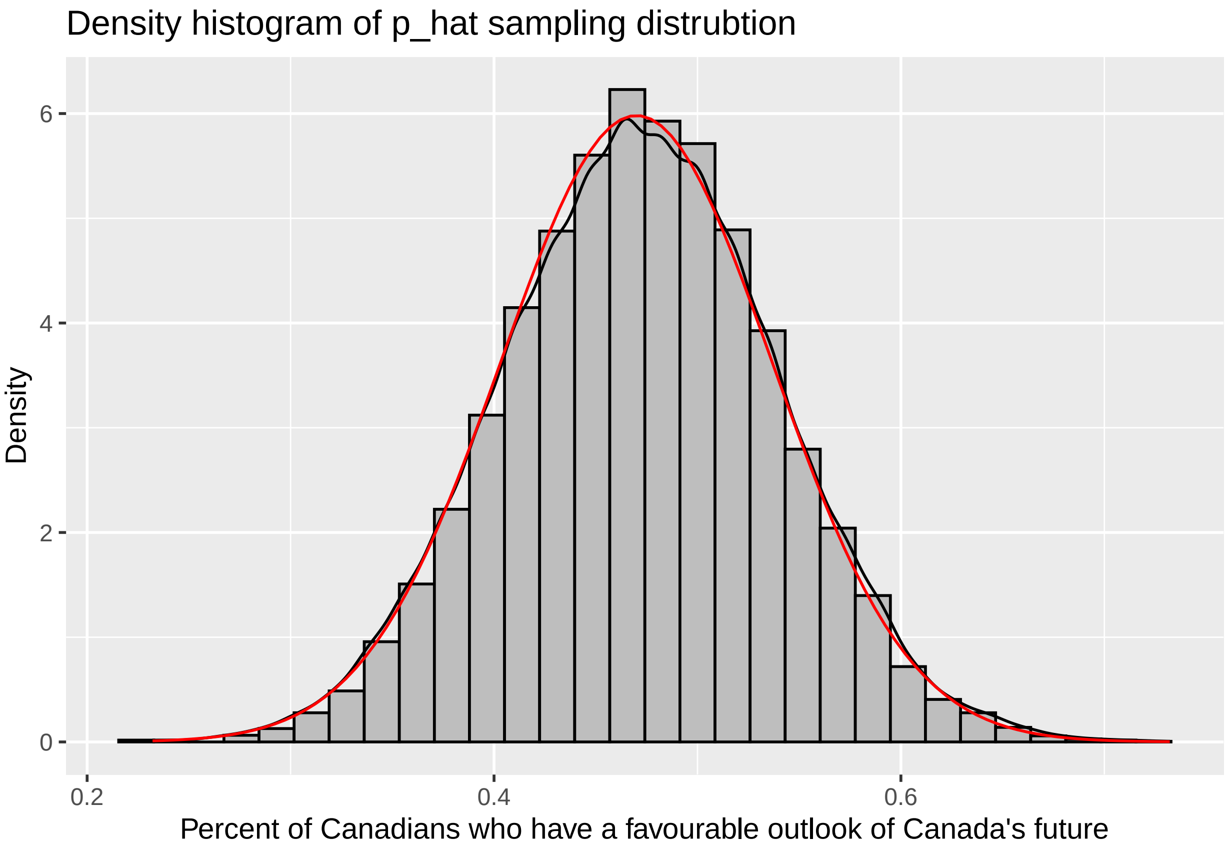
Within (Graph_3), the gray bars represents for the 10000 bootstrap samples, the density of the result of all the samples that consisted of the of the percent of Canadians who have a favorable outlook of Canada's future.

The black line represents the the density of the sampling distribution of the 10000 bootstrap samples

d repeated with n = 56

(graph_4)

```
ggplot(To_data_set2, aes(x=x)) +
  geom_histogram(aes(y=..density..), bins = 30,colour = "black",fill = "grey") +
  geom_density()+
   labs(title = "Density histogram of p_hat sampling distrubtion",
       x = "Percent of Canadians who have a favourable outlook of Canada's future",
       y = "Density") +
      stat_function(fun = dnorm,
                args = list(mean =0.47, sd = sqrt(0.004448214)),
                col = "red")
```



Within (Graph_4), the gray bars represents for the 10000 bootstrap samples, the density of the result of all the samples that consisted of the of the percent of Canadians who have a favorable outlook of Canada's future.

The black line represents the the density of the sampling distribution of the 10000 bootstrap samples

The red line represents the overlay of the CLT approximation of the sampling distribution of $\hat{p}$.

e repeated with n = 56

```
count <- sum(To_data_set2$x < 0.5)
count
```

```
## [1] 6135
```

With 10000 bootstrap samples we can see with the value of count (seen above), the number of the samples that returned results where more than half of those surveyed answered NO to the survey question.

```
probability_wanted <-count/10000
probability_wanted
```

```
## [1] 0.6135
```

From this we determined that the value of probability_wanted (seen above) percent probability that a group of 65 people surveyed would result in a majority of people voting NO to the survey. From this information we can conclude that there is a higher percent chance that a group of 65 people surveyed would result in a majority of people not being optimistic about the future of Canada, as opposed to a majority of people who are surveyed being optimistic in the future of Canada. The news article's claim is supported by the bootstrap data due to the fact that in the article's study, less than half of Canadians answered 'YES', lining up with the bootstrap data where the majority of the bootstrap samples also resulted in less than half of Canadians with an optimistic outlook in Canada's future.

g)

In both cases when n = 1483 and n = 56 of the 10000 bootstrap samples, both cases resulted in a majority of samples where more than half of Canadians who were surveyed answered that they do not having a positive outlook of Canada's future. However, in each case the proportion of samples that returned less than half of Canadians having a positive outlook of the future varied greatly. With the case of n = 1483 returning a much higher percent of the samples with a majority of those sampled answering "No" and compared to the case where n = 56. Also, in the case where n = 56 we see that the histogram is a greater variance among the values compared to when n = 1483. We know that these results are intuitively correct because the more people you survey the more accurate the results of your survey will be to the true population parameter. For example if only one person was surveyed then the results would show either 100 percent or 0 percent of Canadians have a favorable outlook of the future, but if every Canadian answered then the survey would be very accurate, in fact equal to the population parameter. This is further backed up by the Law of Large numbers.