Exam 2

Robert Campbell

Friday, April 16, 2021

Place your answers into this markdown document, knit it, and hand in the result as a PDF. There are 10 questions, each worth 10 points.

You may use R, the internet, and any reference material, but do not work together and do not get help (except from Dr. Clair).

Honor Pledge

The work I have submitted represents my own effort. While working on this exam, I did not communicate in any form with individuals other than the instructor.

Signed: Robert V. Campbell

```
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(ggplot2))
suppressPackageStartupMessages(library(stringr))
shoes <- read.csv("http://mathstat.slu.edu/~clair/stat/data/shoes.csv")
```

Problem 1

This problem and the next use the data `Arbuthnot` from the `HistData` library.

Find all the years where death by Plague was at least 20% of all Mortality.

```
Arb<-HistData::Arbuthnot
Arb %>% group_by(Year) %>% summarize(DeathRatio = Plague/Mortality) %>% filter(DeathRatio >= .20
)
```

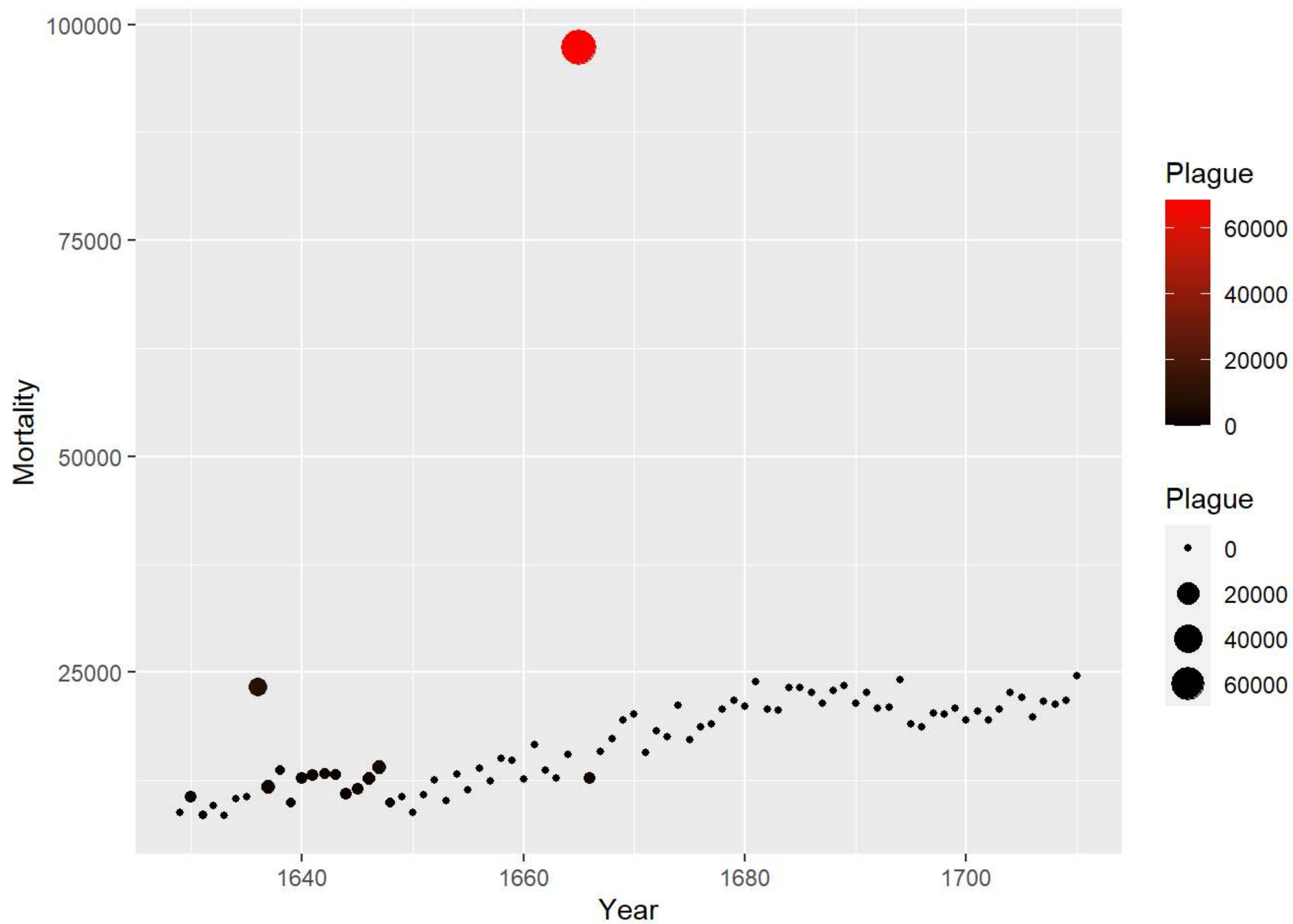
```
## # A tibble: 4 x 2
##    Year DeathRatio
##   <int>      <dbl>
## 1  1636      0.445
## 2  1637      0.262
## 3  1647      0.256
## 4  1665      0.705
```

Problem 2

Using `Arbuthnot` , make a plot showing the Mortality each Year, with Year on the x-axis. Set the size and color of your points to be the Plague variable.

Small amount of extra credit: make the color go from black to red instead of black to blue to really make the plague look bad.

```
Arb %>% ggplot(aes(x=Year,y=Mortality,size=Plague, color=Plague)) + geom_point() +
  scale_color_gradient(low="black", high="red")
```



Problem 3

This problem uses the `bechdel` data from the `fosdata` library. Each observation is a movie, and the `binary` variable tells whether that movie passed or failed the "Bechdel Test". A movie's profit (in adjusted 2013 dollars) is the `intgross_2013` international gross minus the `budget_2013` budget.

Of the 15 least profitable movies in the `bechdel` dataset, what is the only one to pass the Bechdel test?

```
fosdata::bechdel %>% mutate(profit=(intgross_2013 - budget_2013)) %>%
  group_by(title) %>% summarize(profit, binary) %>% arrange(profit) %>%
  head(15) %>% filter(binary=="PASS")
```

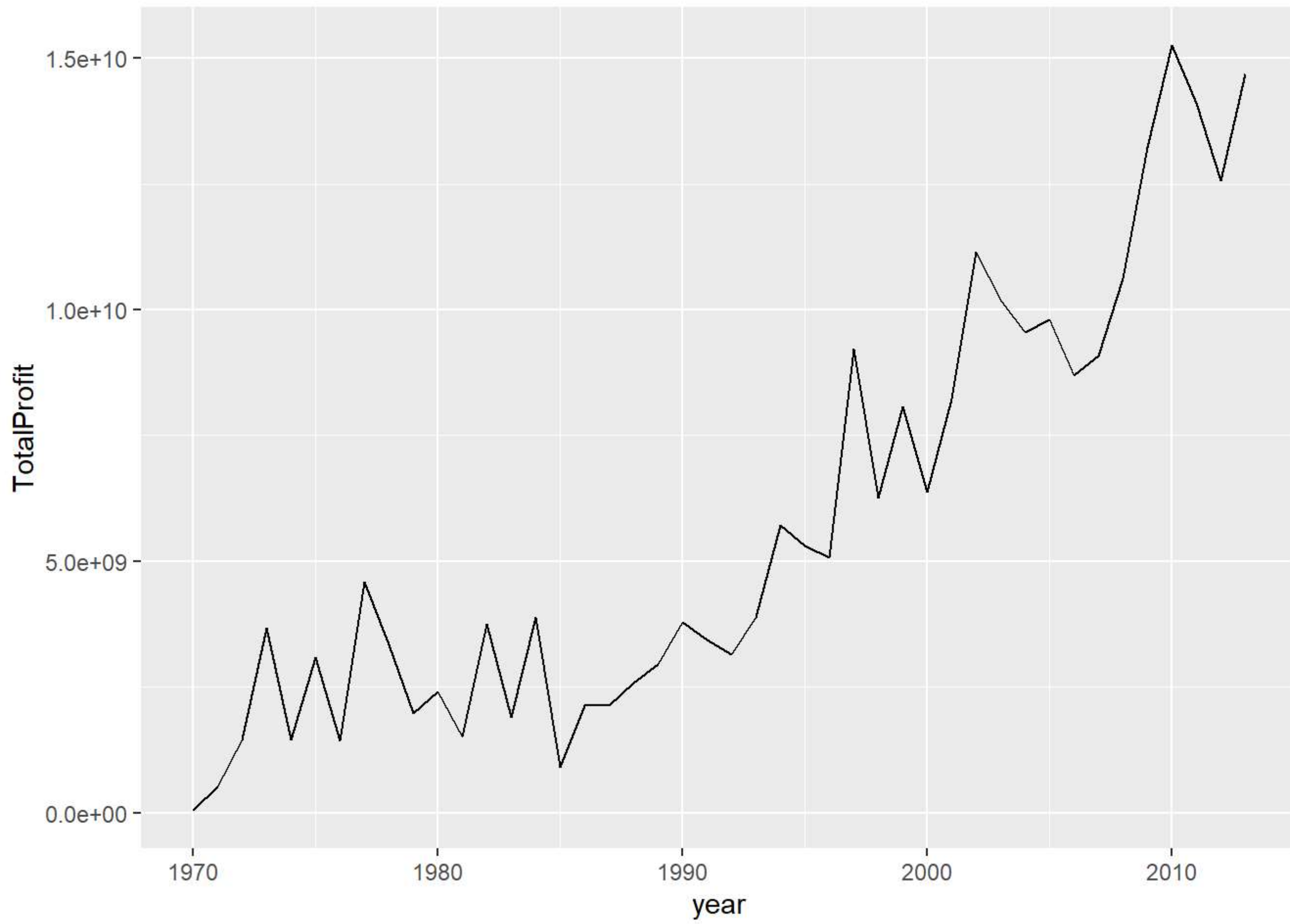
```
## # A tibble: 1 x 3
## # Groups:   title [1]
##   title         profit binary
##   <chr>          <dbl> <fct>
## 1 Hudson Hawk -81745807 PASS
```

Problem 4

Continue using `fosdata::bechdel` .

Make a line graph showing total profit of all movies added together for each year.

```
fosdata::bechdel %>% mutate(profit=(intgross_2013 - budget_2013)) %>%
  group_by(year) %>% summarize(TotalProfit=sum(profit,na.rm=TRUE)) %>%
  ggplot(aes(x=year,y=TotalProfit)) + geom_line()
```



Problem 5

Continue using `fosdata::bechdel` .

a. There are 15 movies with the word “Girl” as part of the title. What percentage of these passed the Bechdel test?

```
fosdata::bechdel %>% filter(str_detect(title, "Girl")) %>% group_by(binary) %>%
  summarize(ratio = n()/15)
```

```
## # A tibble: 2 x 2
##   binary ratio
##   <fct>  <dbl>
## 1 FAIL   0.267
## 2 PASS   0.733
```

b. There are 7 movies with the word "Boy" as part of the title (watch out for "Boynton Beach Club", which you shouldn't count). What percentage of these passed the Bechdel test?

```
fosdata::bechdel %>% filter(str_detect(title, "Boy")) %>% group_by(binary) %>%
  summarize(ratio = n()/7)
```

```
## # A tibble: 2 x 2
##   binary ratio
##   <fct>  <dbl>
## 1 FAIL   0.857
## 2 PASS   0.286
```

Problem 6

How much bigger is the tail of a t distribution than the normal distribution?

a. Compute $P(Z > 4)$ where Z is the standard normal random variable.

```
1-pnorm(4)
```

```
## [1] 3.167124e-05
```

b. Compute $P(T > 4)$ where T has a t distribution with 2 degrees of freedom.

```
1-pt(4,df=2)
```

```
## [1] 0.02859548
```

c. Compute the ratio $P(T > 4)/P(Z > 4)$.

```
(1-pnorm(4))/(1-pt(4,df=2))
```

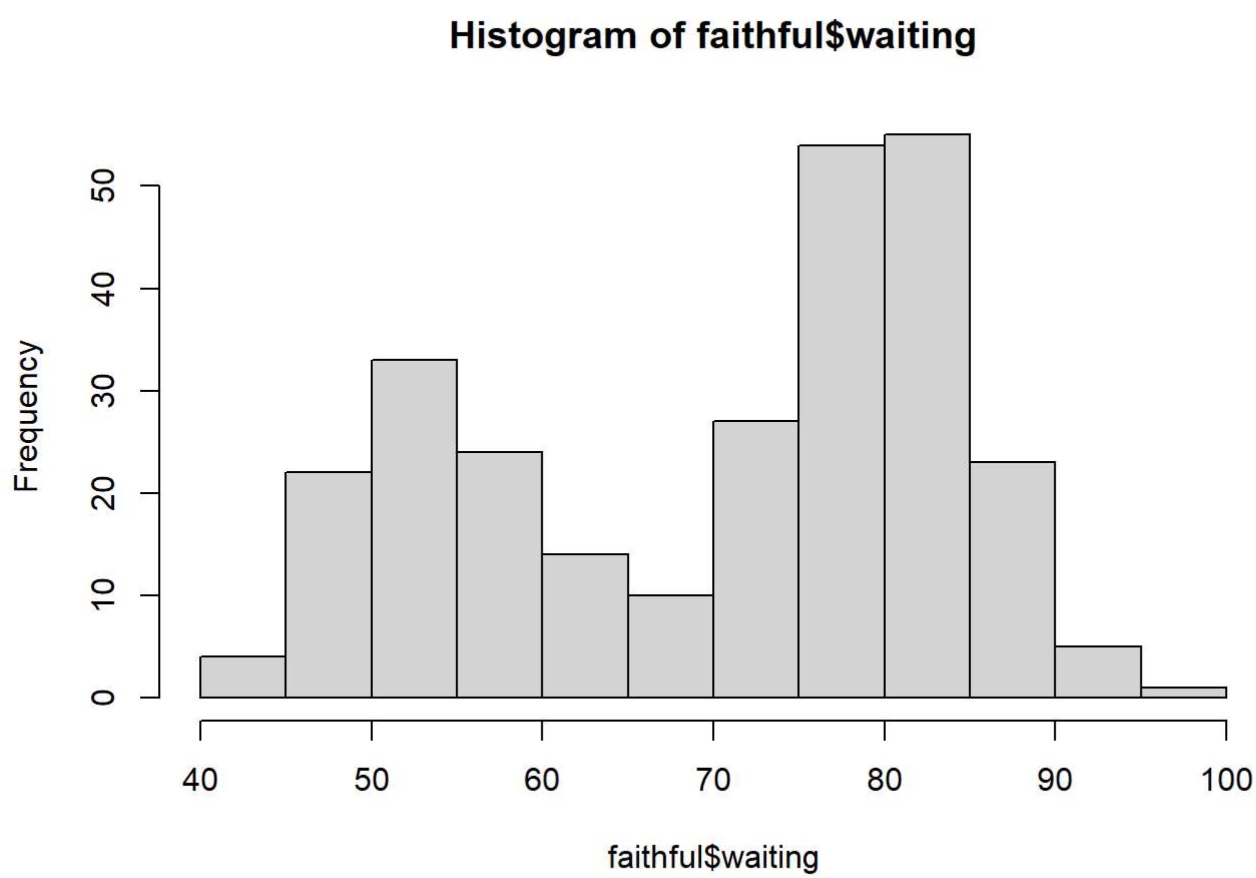
```
## [1] 0.001107561
```

Problem 7

The data set `faithful` is built in to R. It has a variable `waiting` which gives the time between eruptions of the Old Faithful geyser in Yellowstone National Park, collected in August 1985.

a. Make a histogram of the `waiting` variable. How would you describe this distribution?

```
hist(faithful$waiting)
```

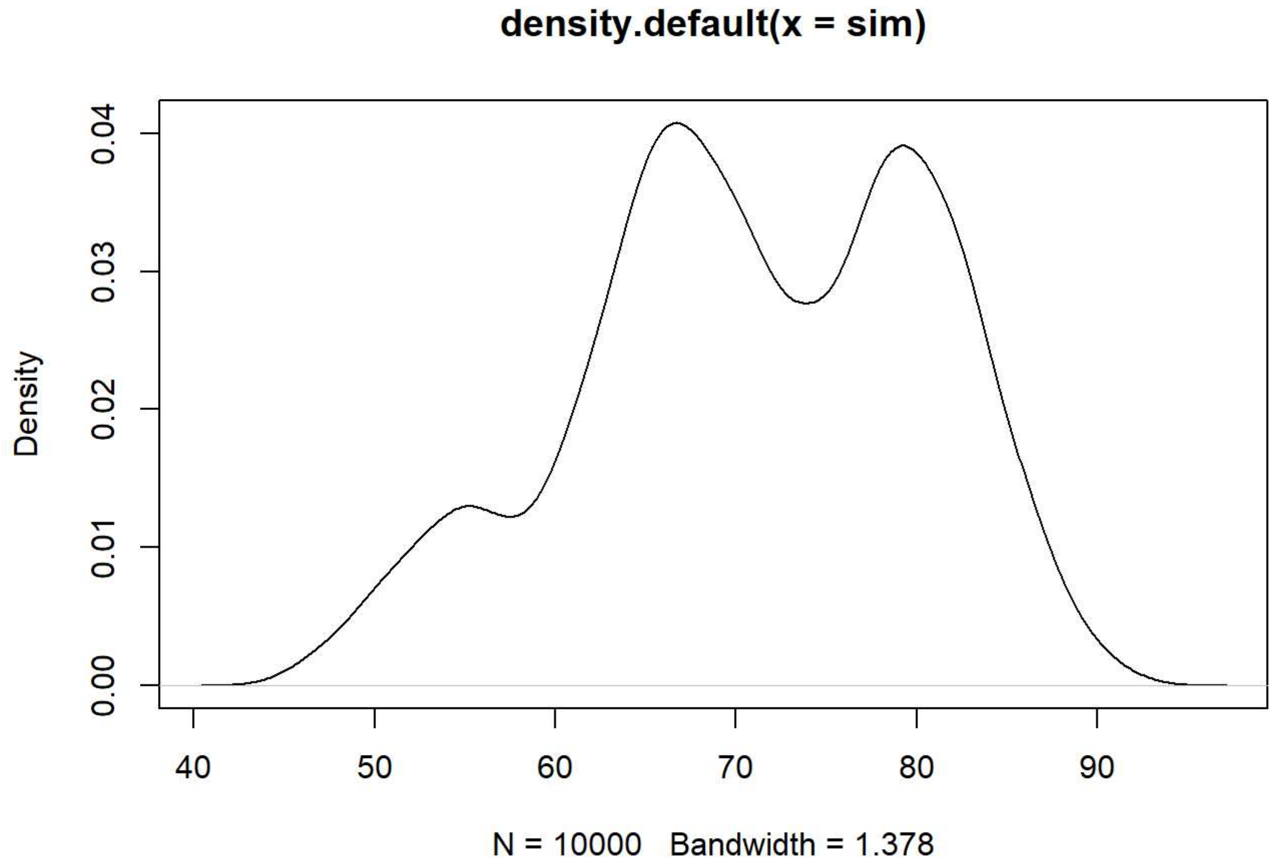


Not normal

Let $X_1, \ldots, X_n$ be a sample of size n of the `waiting` variable and $\bar{X}$ the sample mean, as usual.

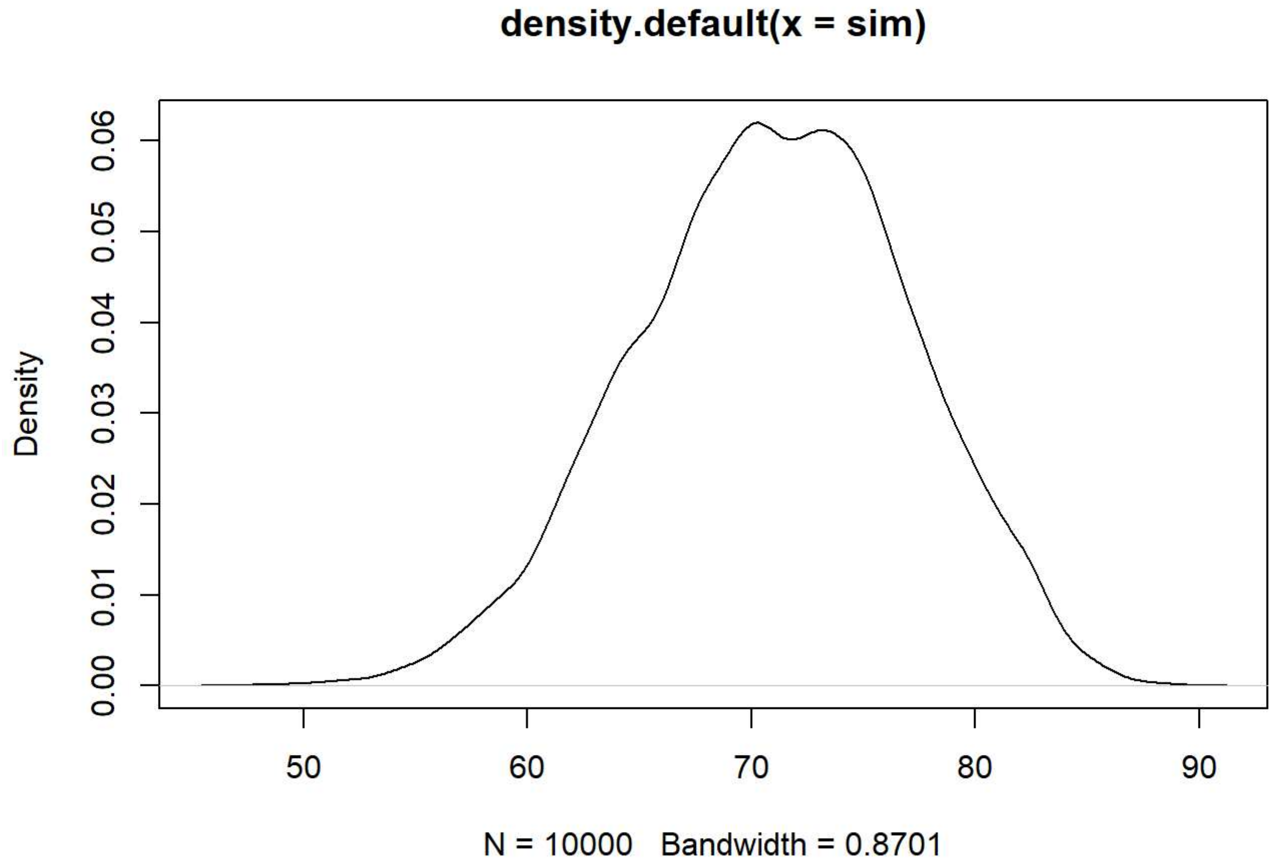
b. Simulate and plot the $\bar{X}$ distribution when $n = 2$. Does it appear to be normal? Not normal

```
pop <- faithful$waiting
sim <- replicate(10000,{
  x <- sample(pop,2,replace=TRUE); mean(x)
})
plot(density(sim))
```



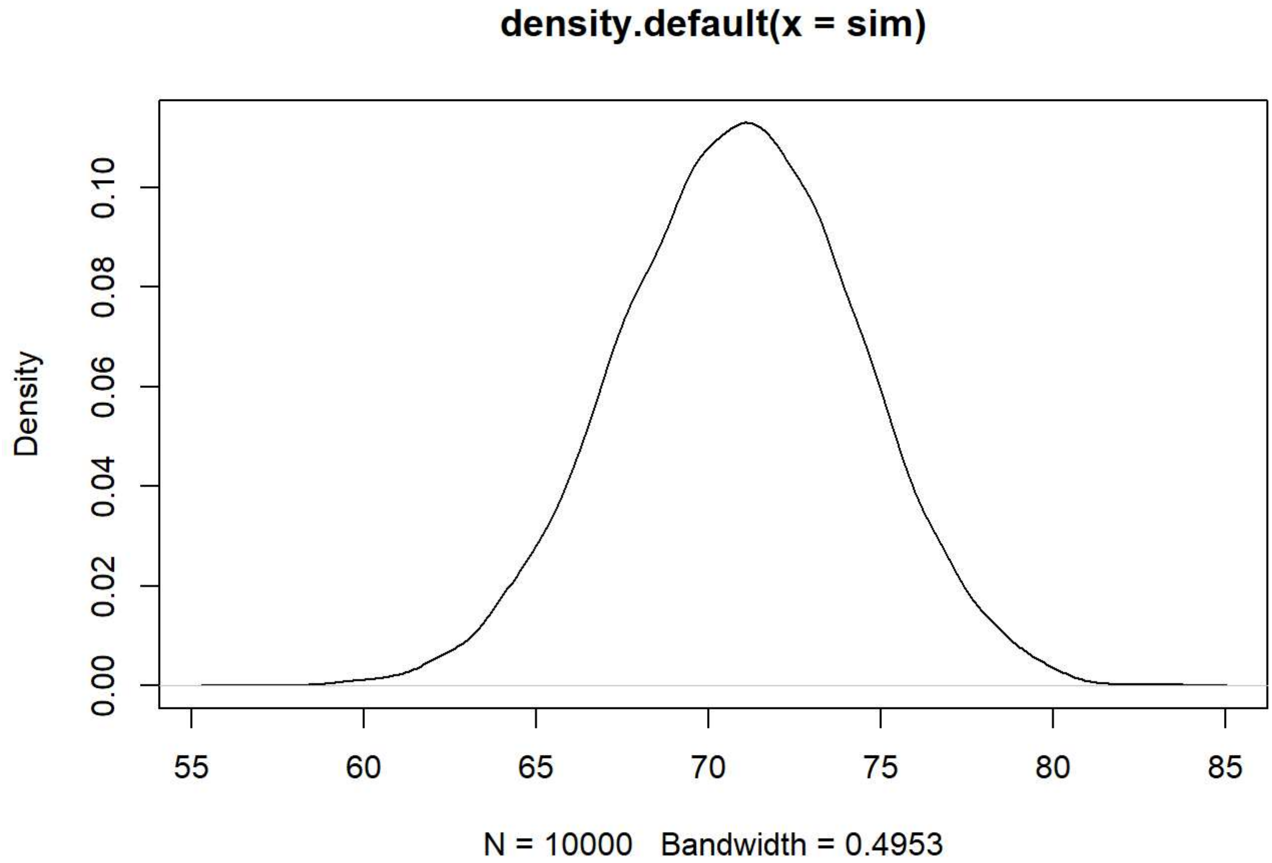
c. Simulate and plot the $\bar{X}$ distribution when $n = 5$. Does it appear to be normal? normal

```
sim <- replicate(10000,{
  x <- sample(pop,5,replace=TRUE); mean(x)
})
plot(density(sim))
```



d. Simulate and plot the $\bar{X}$ distribution when $n = 15$. Does it appear to be normal? Normal

```
sim <- replicate(10000,{
  x <- sample(pop,15,replace=TRUE); mean(x)
})
plot(density(sim))
```



The next three problems use the data set `shoes.csv` which is available on our course data page.

```
shoes <- read.csv("http://mathstat.slu.edu/~clair/stat/data/shoes.csv")
```

Studies of running mechanics often use a standardized lab shoe, ostensibly to reduce variance between subjects; however, this may induce unnatural running mechanics. Hunter et. al. (2020) tested runners in both lab shoes and in their own personal shoes.

Problem 8

a. Give the 95% confidence interval for the mean speed of male runners in their own shoes. [3.149, 3.552]

```
male<-shoes %>% filter(Sex=="M")
t.test(male$Speed.own)
```

```
##
##  One Sample t-test
##
## data:  male$Speed.own
## t = 34.701, df = 20, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.149069 3.551884
## sample estimates:
## mean of x
##  3.350476
```

b. Give the 95% confidence interval for the mean speed of female runners in their own shoes. [3.019,3.282]

```
female<-shoes %>% filter(Sex=="F")
t.test(female$Speed.own)
```

```
##
##  One Sample t-test
##
## data:  female$Speed.own
## t = 49.041, df = 28, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.019087 3.282292
## sample estimates:
## mean of x
##   3.15069
```

Problem 9

The variable `Speed.own` gives the runner's speed (while wearing their own shoes).

Test for a difference in mean speed between the male and female athletes. State your conclusion with a p-value.

With a p-value: 0.09341, there is not a significant difference based on gender of a runner's speed in their own shoes

```
t.test(Speed.own ~ Sex, data=shoes)
```

```
##
##  Welch Two Sample t-test
##
## data:  Speed.own by Sex
## t = -1.7227, df = 36.517, p-value = 0.09341
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.43487888  0.03530581
## sample estimates:
## mean in group F mean in group M
##        3.150690        3.350476
```

Problem 10

The vertical average loading rate (in bodyweight / second) tells how much load athletes experience while running. This measurement is in the `Loading.own` and `Loading.lab` variables in the data set `shoes.csv`, which is on our course data page.

Test for a difference in mean loading rate between the lab shoe and the runner's own shoe. State your conclusion with a p-value.

With a p-value of 0.00683 there is a significant difference in loading rate with a runner's own shoe vs the lab's

```
t.test(shoes$Loading.own, shoes$Loading.lab, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  shoes$Loading.own and shoes$Loading.lab
## t = 2.8245, df = 49, p-value = 0.00683
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.275272 7.564728
## sample estimates:
## mean of the differences
##                    4.42
```