DarflerM - HW6

NAME: Michael Darfler

NETID: mbd25

DUE DATE: October 28, 2019 by 11:59pm

For this homework, it will be helpful to have a copy of the knitted version of this document to answer the questions as much of it is written using mathematical notation that may be difficult to read when the document is not knitted.

Instructions

For this homework:

1. All calculations must be done within your document in code chunks. Provide all intermediate steps.
2. Include any mathematical formulas you are using for a calculation. Surrounding mathematical expresses by dollar signs makes the math look nicer and lets you use a special syntax (called latex) that allows for Greek letters, fractions, etc. Note that this is not R code and therefore should not be put in a code chunk. You can put these immediately before the code chunk where you actually do the calculation.

Some Notation

Your solutions to the problems below must include the formula used for each calculation. To get you started, here is some mathematical expressions written in latex that you may find helpful when writing out the math in your answers. You can copy, paste, and edit these expressions as needed.

1. ($\bar{x}_n$ - $SE(\bar{X}_n) \times z_{\alpha/2}$, $\bar{x}_n$ + $SE(\bar{X}_n) \times z_{\alpha/2}$)
2. $n\hat{p}_n \geq 10$ and $n(1-\hat{p}_n) \geq 10$
3. ($\hat{p}_n$ - $SE(\hat{p}_n) \times z_{\alpha/2}$, $\hat{p}_n$ + $SE(\hat{p}_n) \times z_{\alpha/2}$)
4. ($\bar{x}_n$ - $SE(\bar{X}_n) \times t_{n-1,\alpha/2}$, $\bar{x}_n$ + $SE(\bar{X}_n) \times t_{n-1,\alpha/2}$)
5. $SE(\bar{X}_n) = \sigma/\sqrt{n}$
6. $s/\sqrt{n}$
7. $SE(\hat{p}_n)$ estimated by $\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$

```
icu <- read.csv("https://raw.githubusercontent.com/mdarfler/BTRY_6010/master/Home
work/HW%206/ICUAdmissions(3).csv", header = T)
flightdata <- read.csv("https://raw.githubusercontent.com/mdarfler/BTRY_6010/mast
er/Homework/HW%206/FlightData(3).csv", header = T)
homes <-  read.csv("https://raw.githubusercontent.com/mdarfler/BTRY_6010/master/H
omework/HW%206/HomesForSaleCanton(3).csv", header = T)
```

In this homework we will practice creating confidence intervals for a population mean, μ.

Problem 1

A random sample of 110 lightning flashes in a certain region resulted in a sample average radar echo duration of .81 sec and a sample standard deviation of .34 sec.

a. What distribution should be used to determine the quantiles needed to calculate a 95% confidence interval for mean radar echo duration? Why?

> The distribution that should be used to determine the questionable needed to calculate the 95% confidence interval for mean radar echo duration is the Normal Distribution $N(0, 1)$. This is because $n = 110$ which is sufficiently large to assume a Normal distribution under the Central Limit Theorem

b. Calculate a 95% confidence interval for mean radar echo duration. Interpret this confidence interval in terms of the study.

$\bar{x}_n \pm SE(\bar{X}_n) \times z_{\alpha/2}$

$SE(\bar{X}_n) = \sigma/\sqrt{n} \approx s/\sqrt{n}$

```
ci <- 0.95
alpha <- 1 - ci
qntile <- -qnorm(alpha/2)
x_bar <-  0.81
sd <-  0.34
n <-  110
SE <-  sd/sqrt(n)

lower <- x_bar - qntile * SE
upper <- x_bar + qntile * SE

paste0("ci_95 = ", lower, " ≤ mu ≤", upper)
```

```
## [1] "ci_95 = 0.746462420592059 ≤ mu ≤0.873537579407941"
```

We are 95% confident that the true mean of the random Variable X, radar echo duration, is between 0.746 seconds and 0.874 seconds.

Problem 2

Ten recently sold houses were randomly selected from Canton, NY. For each house, the sale price was recorded. The data are in the *HomesForSaleCanton.csv* file in the folder for homework 6 on blackboard. Assume the cost of homes in Canton, NY is normally distributed.

a. What distribution should be used to determine the quantiles needed to calculate a 99% confidence interval for mean cost of a home in Canton, NY? Why?

The t Distribution with 9 $(n-1)$ degrees of freedom should be used to determine the quantiles needed to calculate the 99% confidence interval for mean cost of a home in Canton, NY because we know that the population distribution from which the sample is generated is Normal, but sample size is small and the true population variance is not known.

b. Compute a 99% confidence interval for the average cost of a home in Canton, NY. Interpret this confidence interval in terms of the study. Use code chunks to compute all of the values needed for the confidence interval and to compute the lower and upper bounds of the confidence interval.

$\bar{x}_n \pm SE(\bar{X}_n) \times t_{\alpha/2,n-1}$

$SE(\bar{X}_n) = \sigma/\sqrt{n} \approx s/\sqrt{n}$

```
x_bar <-  mean(homes$Price)
sd <-  sqrt(var(homes$Price))
n <-  length(homes$Price)
SE10 <-  sd/sqrt(n)

ci <- 0.99
alpha <- 1 - ci
qntile <- -qt(alpha/2,n-1)


lower10 <- x_bar - qntile * SE10
upper10 <- x_bar + qntile * SE10

paste0("ci",(1-alpha)*100,"% = ", lower10, " ≤ mu ≤ ", upper10)
```

```
## [1] "ci99% = 49.1716612203239 ≤ mu ≤ 244.428338779676"
```

We are 99% confident that the true mean of the random Variable X, the sale price of a house in Canton, NY, is between \$49,172 and \$244,428.

c. How would this confidence interval change if the sample mean and variance are unchanged, but the sample size is 35?

```
ci <-  0.99
alpha <- 1 - ci
qntile <- -qnorm(alpha/2)
x_bar <-  mean(homes$Price)
sd <-  sqrt(var(homes$Price))
n <-  35
SE35 <-  sd/sqrt(n)

lower35 <- x_bar - qntile * SE35
upper35 <- x_bar + qntile * SE35

paste0("ci",(1-alpha)*100,"% = ", lower35, " ≤ mu ≤", upper35)
```

```
## [1] "ci99% = 105.438376894337 ≤ mu ≤188.161623105663"
```

When increasing the sample size from 10 to 35 we can use the Normal Distribution when calculating the quantiles for our confidence intraval because n is sufficiently large. In this case the 99% confidence interval decreased from 195.26 to 82.72 because the SE decreased from 30.04 to 16.06 and the distribution changed.

Problem 3

The data set *ICUAdmissions.csv* includes information on 200 randomly selected patients admitted to the Intensive Care Unit (ICU). One of the variables, *Status* indicates whether the patient lived (`Status=0`) or died (`Status = 1`).

a. Based on these data create a 99% confidence interval for the survival rate of ICU patients. Include all calculations in a code chunk and interpret this interval in the context of the study.

For this study we need to use a Bernoulli Distribution to calculate $\hat{p}_n$, the sample proportion, and use $\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$ to calculate the sample standard deviation.

$sd = \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$

```
p_hat <- 1-mean(icu$Status)
n <-  length(icu$Status)
SE <- sqrt((p_hat * (1-p_hat))/n)

ci <- 0.99
alpha <- 1 - ci
qntile <- -qnorm(alpha/2)

lower <- p_hat - qntile * SE
upper <- p_hat + qntile * SE

paste0("ci",(1-alpha)*100,"% = ", lower, " ≤ p_hat ≤", upper)
```

```
## [1] "ci99% = 0.727144545291262 ≤ p_hat ≤0.872855454708738"
```

> We are 99% confident that the true proportion of the random Variable $X = 0$, the patient survived, is between 0.73 and 0.87

b. What assumptions are necessary for the confidence interval determined in (a) to be valid? Provide evidence that each of these assumptions is reasonable.

> In order to calculate the confidence interval we have to assume that the data is iid (independent and identically distributed). The probability that patient a survives, given patient b survives, is the same as probability of patient a surviving. In other words $P(A|B) = P(A)$. We also have to know that $n\hat{p}_n \geq 10$ and $n(1 - \hat{p}_n) \geq 10$

```
n*p_hat >= 10
```

```
## [1] TRUE
```

```
n*(1-p_hat) >= 10
```

```
## [1] TRUE
```

Problem 4

A sample of 14 joint specimens of a particular type gave a sample mean proportional limit stress of 8.48

MPa. Assume proportional limit stress is approximately normally distributed. The variance of proportional limit stress is known to be $\sigma^2 = .6241$.

a. Calculate and interpret a 95% confidence interval for mean proportional limit stress.

$X \sim N(8.48, 0.6241)$ $\bar{x}_n \pm SE(\bar{X}_n) \times z_{\alpha/2}$

```
ci <-  0.95
alpha <- 1 - ci
qntile <- -qnorm(alpha/2)
x_bar <-  8.48
var <- 0.6241
n <-  14
SE <-  sqrt(var/n)

lower <- x_bar - qntile * SE
upper <- x_bar + qntile * SE

paste0("ci_95 = ", lower, " ≤ mu ≤", upper)
```

```
## [1] "ci_95 = 8.06618029719668 ≤ mu ≤8.89381970280332"
```

> We are 95% confident that the true mean of the random Variable X, the proportional limit stress, is between 8.066 MPa and 8.894 MPa.

b. Without recalculating this interval, explain how the width of this confidence interval would change if the confidence level was set to 90%.

> If the confidence interval was set to 90% its width would decrease because the $Z_{\alpha/2}$ value would decrease.

Problem 5

United Airlines Flight 179 is non-stop from Boston's Logan airport to San Francisco International Airport. An important factor in scheduling flights is the actual airborne flying time from takeoff to touchdown. In the data file, *FlightData.csv*, is the airborne time in minutes for 3 dates each month in the year 2010 for this flight and one other flight. The data of interest can be found in the column named `Flight179`. Create a 95% confidence interval for the mean airborne time for this flight based on the data and interpret it in terms of the study. Also, give support for how you chose the quantiles to calculate this interval.

```r
data <- flightdata$Flight179
ci <- 0.95
alpha <- 1 - ci
qntile <- -qnorm(alpha/2)
x_bar <-  mean(data)
var <-  var(data)
n <-  length(data)
SE <-  sqrt(var/n)

lower <- x_bar - qntile * SE
upper <- x_bar + qntile * SE

paste0("ci_95 = ", lower, " ≤ mu ≤", upper)
```

```
## [1] "ci_95 = 351.269753013826 ≤ mu ≤364.452469208396"
```

> We are 95% confident that the true mean of the random Variable X, the proportional limit stress, is between 351.27 minutes and 364.45 minutes. The Normal Distribution was used to calculate the quantiles because the sample size was 36 which is sufficiently large according to the Central Limit Theorem.