Homework 6: Confidence Intervals

NAME: Your Name

NETID: Your NetID

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. Incude 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 $\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$

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? *A N(0,1) distribution should be used because the sample size is large.*

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

$(\bar{x}_n - SE(\bar{X}_n) \times z_{\alpha/2}, \bar{x}_n + SE(\bar{X}_n) \times z_{\alpha/2}) = (.81 - .34/\sqrt{110} \times z_{.025}, .81 + .34/\sqrt{110} \times z_{.025}) = (.75,.87)$

```
smean=.81
se=.34/sqrt(110)
quant = -qnorm(.025)

Lower = smean - se*quant
Upper = smean + se*quant
Lower
```

```
## [1] 0.7464624
```

```
Upper
```

```
## [1] 0.8735376
```

We are 95% confident that the mean radar echo duration of lightning strikes is between 0.75 sec. and 0.87 sec.

Problem 2

Ten recently sold houses were randomly selected from Canton, NY. For each house, the sale price in thousands 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? *A t_9 distribution since the data are sampled from a normal population and the sample size is 10.*

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.

```
HomesForSaleCanton <- read.table("HomesForSaleCanton.csv", header = TRUE, quote = "\"")
smean = mean(HomesForSaleCanton$Price)
sd = sd(HomesForSaleCanton$Price)
se = sd/sqrt(10)
quant = -qt(0.005, 9)
```

$(\bar{x}_n - SE(\bar{X}_n) \times t_{n-1,\alpha/2}, \bar{x}_n + SE(\bar{X}_n) \times t_{n-1,\alpha/2}) = (146.8 - 95/\sqrt{10} \times t_{9,.005}, 146.8 + 95/\sqrt{10} \times t_{9,.005}) = (49.17,244.43)$

```r
Lower = smean - se*quant
Upper = smean + se*quant
smean
```

```
## [1] 146.8
```

```r
sd
```

```
## [1] 94.99801
```

```r
Lower
```

```
## [1] 49.17166
```

```r
Upper
```

```
## [1] 244.4283
```

We are 99% confident that the mean sale price of houses in Canton, NY is between $49.17 and $244.43.

c) How would this confidence interval change if the sample mean and variance are unchanged, but the sample size is 35? *The standard error and the quantile used to determine the confidence interval would both be smaller. Thus, the width of the confidence interval would be smaller than the one calculated in (b).*

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.

```r
ICUAdmissions <- read.csv("ICUAdmissions.csv")
length(ICUAdmissions$Status)
```

```
## [1] 200
```

```r
smean=1-mean(ICUAdmissions$Status)
svar = (smean*(1-smean))
smean
```

```
## [1] 0.8
```

```r
svar
```

```
## [1] 0.16
```

```r
se = sqrt(svar/length(ICUAdmissions$Status))
quant = -qnorm(.005)
```

```r
Lower = smean - se*quant
Upper = smean + se*quant
Lower
```

```
## [1] 0.7271445
```

```r
Upper
```

```
## [1] 0.8728555
```

($\hat{p}_n$ - $SE(\hat{p}_n)$ $\times$ $z_{\alpha/2}$, $\hat{p}_n$ + $SE(\hat{p}_n)$ $\times$ $z_{\alpha/2}$)=(.8 - $\sqrt{\frac{.8(1-.8)}{200}}$ $\times$ $z_{.005}$, .8 + $\sqrt{\frac{.8(1-.8)}{200}}$ $\times$ $z_{.005}$) = (.73,.87).
We are 99% confident that the survival rate of ICU patients is in the interval (.73,.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.

The first assumption made is that the observations are independent and sampled from the population of interest. Here we assume this assumption is met since a random sample of ICU patients is selected. The second assumption we are making is that the CLT applies to determine the distribution of $\hat{p}_n$. To check this we should check that $n\hat{p}_n \geq 10$ and $n(1-\hat{p}_n) \geq 10$. $200 \times .8 = 160$ and $200 \times (1-.8) = 40$. So, this assumption seems reasonable.

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 distrubuted. 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.

($\bar{x}_n$ - $SE(\bar{X}_n)$ $\times$ $z_{\alpha/2}$, $\bar{x}_n$ + $SE(\bar{X}_n)$ $\times$ $z_{\alpha/2}$) = (8.48 - $\sqrt{\frac{.6241}{14}}$ $\times$ $z_{.025}$, 8.48 + $\sqrt{\frac{.6241}{14}}$ $\times$ $z_{.025}$) =

```r
smean=8.48
se=sqrt(.6241/14)
quant = -qnorm(.025)

Lower = smean - se*quant
Upper = smean + se*quant
Lower
```

```
## [1] 8.06618
```

```r
Upper
```

```
## [1] 8.89382
```

We are 95% confident that mean proportional limit stress for this type of joint specimens 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%.

The interval would have a smaller width because we are reducing how confident we are about the range of values for μ. This results in a smaller quantile being used to compute the confidence interval.

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.

Since the sample size of 36 is likely to be large enough for the CLT to apply to these data, the standard normal distribution can be used to determine the quantiles. Without evidence that the population distribution of airborne flight time is approximately normally distributed, we should refrain from using the t-distribution. However, even if this was a valid assumption, a t-distribution with 35 degrees of freedom and the N(0,1) distribution will have similar quantiles.

```
FlightData <- read.csv("FlightData.csv")
smean <- mean(FlightData$Flight179)
length(FlightData$Flight179)
```

```
## [1] 36
```

```
ssd <- sd(FlightData$Flight179)
ssd
```

```
## [1] 20.178
```

```
se <- sd(FlightData$Flight179) / sqrt(length(FlightData$Flight179))
smean
```

```
## [1] 357.8611
```

```
se
```

```
## [1] 3.363
```

```
quant = -qnorm(.025)

Lower = smean - se*quant
Upper = smean + se*quant
Lower
```

```
## [1] 351.2698
```

```
Upper
```

```
## [1] 364.4525
```

($\bar{x}_n$ - $SE(\bar{X}_n) \times z_{\alpha/2}$, $\bar{x}_n$ + $SE(\bar{X}_n) \times z_{\alpha/2}$) = (357.9 - 20.18/6 $\times$ $z_{.025}$, 357.9 + 20.18/6 $\times$ $z_{.025}$) = (351.3, 364.5)

We are 95% confident that mean airborne flight time for Flight 179 is between 351.3 minutes and 364.5 minutes.