Laboratory Exercise Week 8

Brian Tipton - STAT-380-001

10/11/23

Directions:

- Write your R code inside the code chunks after each question.
- Write your answer comments after the `#` sign.
- To generate the word document output, click the button `Knit` and wait for the word document to appear.
- RStudio will prompt you (only once) to install the `knitr` package.
- Submit your completed laboratory exercise using Blackboard's Turnitin feature. Your Turnitin upload link is found on your Blackboard Course shell under the Laboratory folder.

For this exercise, you will need to use the package `mosaic` to find numerical and graphical summaries.

```
# install package if necessary
if (!require(mosaic)) install.packages(`mosaic`)
# load the package in R
library(mosaic) # load the package mosaic to use its functions
```

1. The study "Digital Footprints" (Pew Internet and American Life Project reported that 47% of Internet users have searched for information about themselves online. The 47% figure was based on a random sample of Internet users. Suppose that the sample size was `n = 400`.

i) Use this information to estimate the true proportion of Internet users who searched information about themselves using a default 95% confidence interval.

ii) Repeat (i) but instead use a 98% Confidence Level. Use function `prop.test()`.

iii) What is the main difference between the intervals computed in (i) and (ii)? What do you think is the role of the confidence level in determining the precision of the confidence interval estimate?

Code chunk

```
# start your code
n = 400
phat = 0.47
std.error <- 1.96 * sqrt(phat * (1 - phat) / n)

# i) Use this information to estimate the true proportion of Internet users who searched information ab
data.frame(LL = phat - std.error, UL = phat + std.error)
```

```
##          LL        UL
## 1 0.4210883 0.5189117
```

```
#  ii) Repeat (i) but instead use a 98\% Confidence Level. Use function `prop.test()`.
prop.test(x = phat * n, n = 400, conf.level = 0.98)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  * out of 400phat out of 400n out of 400
## X-squared = 1.3225, df = 1, p-value = 0.2501
## alternative hypothesis: true p is not equal to 0.5
## 98 percent confidence interval:
##  0.4115094 0.5293087
## sample estimates:
##    p
## 0.47
```

```
# iii) What is the main difference between the intervals computed in (i) and (ii)?
# What do you think is the role of the confidence level in determining the precision of the confidence

# The main difference between part i) and part ii) is that, in part ii) has a wider range.
# This is because with a standard distribution the higher the confidence level should mean the numbers

# last R code line
```

2. The article "CSI Effect Has Juries Wanting More Evidence" (USA Today, 2004) examines how the popularity of crime-scene investigation television shows in influencing jurors' expectation of what evidence should be produced at a trial. In a survey of 500 potential jurors, one study found that 350 were regular watchers of at least one crime-scene forensics television series.

   i) Assuming that it is reasonable to regard this sample of 500 potential jurors as representative of potential jurors in the US, use the given information to construct and interpret a 95% confidence interval for the proportion of all potential jurors who regularly watch at least one crime-scene investigation series.

   ii) Using the same sample proportion as part (i), construct a 95% confidence interval for same proportion but instead use a sample size of 50.

   iii) What is the main difference between the confidence intervals computed in part (i) and (ii)? What do you think is the role of the sample size in determining the precision the confidence interval estimate?

Code chunk

```
# start your code
# i) Assuming that it is reasonable to regard this sample of 500 potential jurors as representative of
n = 500
phat = 350 / n
std.error = 1.96 * sqrt(phat * (1 - phat) / n)
data.frame(LL = phat - std.error, UL = phat + std.error)
```

```
##         LL       UL
## 1 0.659832 0.740168
```

```
# ii) Using the same sample proportion as part (i), construct a 95\% confidence interval for same propo
prop.test(x = (phat * n) / 10, n = 50, conf.level = 0.95)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  / out of 50(phat * n) out of 5010 out of 50
## X-squared = 7.22, df = 1, p-value = 0.00721
```

```
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.5521660 0.8171438
## sample estimates:
##   p
## 0.7
```

```
# iii) What is the main difference between the confidence intervals computed in part (i) and (ii)? What

# The main and of course the biggest difference here is that the smaller the sample size the lesser to
# When we use n=500, it had about one third of the range of the part we used n=50.
# From this we can conclude that the size of the sample proportional to the confidence level we can gai

# iv) looking at the x-squares and p values in both the first and second part, this shows that the bigg

# last R code line
```