Laboratory Exercise Week 8

Ryan Estes | Section 006, 10 pts

10/11/17

*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.

library(mosaic) # load the package mosaic to use its functions

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: ggformula

## Loading required package: ggplot2

##   
## New to ggformula? Try the tutorials:   
## learnr::run\_tutorial("introduction", package = "ggformula")  
## learnr::run\_tutorial("refining", package = "ggformula")

## Loading required package: mosaicData

## Loading required package: Matrix

##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.  
##   
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.

##   
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median,  
## prop.test, quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

1. The study "Digital Footprints" ([Pew Internet and American Life Project](%22Digital%20Footprints%22%20(Pew%20Internet%20and%20American%20Life%20Project)) 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.
2. Use this information to estimate the true proportion of Internet users who searched information about themselves using a default 95% confidence interval.
3. Repeat (i) but instead use a 98% Confidence Level. Use function prop.test().
4. 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 <- .47\*400 #188  
#std.error <- 1.96\*sqrt(.47\*(1-.47)/n)  
#data.frame(LL = phat - std.error, UL = phat + std.error)  
  
# i)  
prop.test(x = 188, n = 400, conf.level = 0.95)

##   
## 1-sample proportions test with continuity correction  
##   
## data: 188 out of 400  
## X-squared = 1.3225, df = 1, p-value = 0.2501  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.4203762 0.5202091  
## sample estimates:  
## p   
## 0.47

# ii)  
prop.test(x = 188, n = 400, conf.level = 0.98)

##   
## 1-sample proportions test with continuity correction  
##   
## data: 188 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)  
# Assuming the sample sizes are the same, A higher confidence level will widen the confidence interval. Meaning, a higher confidence level on the same sample sizer will yield a less precise estimate.  
  
# last R code line

1. The article ["CSI Effect Has Juries Wanting More Evidence" (USA Today, 2004)](https://usatoday30.usatoday.com/news/nation/2004-08-05-csi-effect_x.htm) 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.
   1. 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.
   2. Using the same sample proportion as part (i), construct a 95% confidence interval for same proportion but instead use a sample size of 50.
   3. 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)  
prop.test(x = 350, n = 500, conf.level = 0.95)

##   
## 1-sample proportions test with continuity correction  
##   
## data: 350 out of 500  
## X-squared = 79.202, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.6574021 0.7394725  
## sample estimates:  
## p   
## 0.7

# ii)  
prop.test(x = 35, n = 50, conf.level = 0.95)

##   
## 1-sample proportions test with continuity correction  
##   
## data: 35 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)  
# The interval of the smaller sample size are much farther apart in relation to the larger sample size. This means that a larger sample size will yield more precise results.   
  
# last R code line