Lab04\_Massey

Matt Massey

6/11/2021

Part 1. Construct Confidence Intervals

# a. We can write our own functions in R  
# Credit: https://www.r-bloggers.com/calculating-confidence-intervals-for-proportions/  
  
# For the function, you must specify the variables that your function will use  
confidence <- function(n, p, z){  
   
 # Create an empty list for our confidence interval  
 out <- list()  
   
 # Formulas for calculating confidence interval for single proportion, lb is the lower bound and ub is the upper bound of the confidence interval.  
 out$lb <- p - z\*sqrt((p\*(1-p))/n)  
 out$ub <- p + z\*sqrt((p\*(1-p))/n)  
 out  
}  
  
# qnorm takes the lower bound of a standard normal distribution to find the critical z-value  
z <- qnorm(.005,lower.tail=FALSE)  
  
# We then substitute the values from the problem to get our confidence interval.  
confidence(n = 5924, p = 153/5924, z)

## $lb  
## [1] 0.02051872  
##   
## $ub  
## [1] 0.03113557

The 99% confidence interval calculated from our custom function is 0.02051872 < p < 0.03113557. In other words, we are 99% confident that the interval between 0.02051872 and 0.03113557 actually does contain the true value of the proportion of Eliquis patients that experience nausea.

#b. We can use the function prop.test()  
prop.test(x=153, n=5924, p=0.5, alternative = "two.sided", conf.level = .99 )

##   
## 1-sample proportions test with continuity correction  
##   
## data: 153 out of 5924, null probability 0.5  
## X-squared = 5325.9, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.5  
## 99 percent confidence interval:  
## 0.02094980 0.03178222  
## sample estimates:  
## p   
## 0.02582714

Using the pre-existing R function, prop.test(), we see that the 99% confidence interval is 0.02094980 to 0.03178222. This is slightly different than the CI calculated in pat 1A, and is probably due to the difference in confidence interval methods (Wald in 1A and Wilson in 1B) and the default use of a continuity correction in prop.test().

Part 2. Bootstrapping

# Original Sample - "A" for number of adverse reaction, "B" for no adverse reaction  
smp <- c(rep("A", 153), rep("B", 5924-153))  
  
# Initialize the storage of our sample statistic, single proportion   
bootstrap.stats <- c()  
  
# Create a "for loop" to run 1000 bootstrap samples   
for (i in 1:1000) {  
   
 # Sampling with replacement of the sample we created above  
 bootstrap <- sample(smp, replace = TRUE)   
   
 # Calculating the bootstrap statistics of each bootstrap sample  
 bootstrap.stats <- c(bootstrap.stats, sum(bootstrap == "A")/5924)  
}  
  
#Calculating the 99% confidence interval using our bootstrap samples   
quantile(bootstrap.stats, probs = c(0.005, 0.995))

## 0.5% 99.5%   
## 0.02093180 0.03089213

Using the bootstrap method with 1000 samples, we see that the 99% confidence interval is 0.02093180 to 0.03156735. This is close to the results obtained in part 1, but closest to the confidence interval calculated with the prop.test() function in part 1B.

Part 3. Comparing Bootstrap Samples

# The following code is the same as above, but we will be using 10000 samples  
smp <- c(rep("A", 153), rep("B", 5924-153))  
  
bootstrap.stats <- c()  
  
# Change number of bootstrap samples from 1000 to 100000  
for (i in 1:10000) {  
   
 bootstrap <- sample(smp, replace = TRUE)   
   
 bootstrap.stats <- c(bootstrap.stats, sum(bootstrap == "A")/5924)  
}  
  
#Calculating the 99% confidence interval using our bootstrap samples   
quantile(bootstrap.stats, probs = c(0.005, 0.995))

## 0.5% 99.5%   
## 0.02042539 0.03122890

Using the bootstrap method with 10,000 samples, we see that the 99% confidence interval is 0.02076215 to 0.03122890. This is close to all previous confidence intervals in part 1A, 1B, and 2, slightly lower than CI’s from 1B and 2, and slightly higher than the CI in part 1A.

Part 4. Comparing Bootstrap Samples with 95% Confidence Level

n=1000

# Original Sample - "A" for number of adverse reaction, "B" for no adverse reaction  
smp <- c(rep("A", 153), rep("B", 5924-153))  
  
# Initialize the storage of our sample statistic, single proportion   
bootstrap.stats <- c()  
  
# Create a "for loop" to run 1000 bootstrap samples   
for (i in 1:1000) {  
   
 # Sampling with replacement of the sample we created above  
 bootstrap <- sample(smp, replace = TRUE)   
   
 # Calculating the bootstrap statistics of each bootstrap sample  
 bootstrap.stats <- c(bootstrap.stats, sum(bootstrap == "A")/5924)  
}  
  
#Calculating the 95% confidence interval using our bootstrap samples   
quantile(bootstrap.stats, probs = c(0.025, 0.975))

## 2.5% 97.5%   
## 0.02211344 0.02970966

n=10,000

# Original Sample - "A" for number of adverse reaction, "B" for no adverse reaction  
smp <- c(rep("A", 153), rep("B", 5924-153))  
  
# Initialize the storage of our sample statistic, single proportion   
bootstrap.stats <- c()  
  
# Create a "for loop" to run 10000 bootstrap samples   
for (i in 1:10000) {  
   
 # Sampling with replacement of the sample we created above  
 bootstrap <- sample(smp, replace = TRUE)   
   
 # Calculating the bootstrap statistics of each bootstrap sample  
 bootstrap.stats <- c(bootstrap.stats, sum(bootstrap == "A")/5924)  
}  
  
#Calculating the 95% confidence interval using our bootstrap samples   
quantile(bootstrap.stats, probs = c(0.025, 0.975))

## 2.5% 97.5%   
## 0.02177583 0.02987846

Using the bootstrap method with 1,000 samples, we see that the 95% confidence interval is 0.02194463 to 0.02970966.

Using the bootstrap method with 10,000 samples, we see that the 95% confidence interval is 0.02177583 to 0.02987846.

Both bootstrap calculations show similar 95% confidence intervals, although the CI calculated with n=10,000 shows a slightly wider interval. Both 95% confidence intervals are more narrow than the 99% confidence intervals calculated in parts 1-3, but the 95% Cis are contained within the 99% CIs.