hw5

jason grahn

11/2/2018

# R portion

population <- ames$Gr.Liv.Area  
  
set.seed(40)  
samp <- sample(population, 60)  
  
a <- 1 - 0.05/2  
ci.95 <- qnorm(a)  
  
sample\_mean <- mean(samp)  
se <- sd(samp) / sqrt(60)  
lower <- sample\_mean - ci.95 \* se  
upper <- sample\_mean + ci.95 \* se  
  
mean(population)

## [1] 1499.69

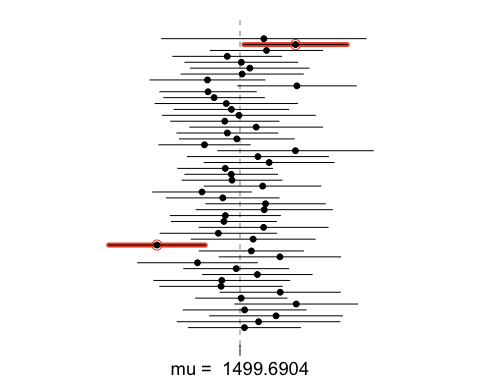
samp\_mean <- rep(NA, 50)  
samp\_sd <- rep(NA, 50)  
  
n <- 60  
  
for(i in 1:50){  
 samp <- sample(population, n)   
 samp\_mean[i] <- mean(samp)   
 samp\_sd[i] <- sd(samp)   
}  
  
lower\_vector <- samp\_mean - ci.95 \* samp\_sd / sqrt(n)   
upper\_vector <- samp\_mean + ci.95 \* samp\_sd / sqrt(n)  
  
c(lower\_vector[1], upper\_vector[1])

## [1] 1384.471 1634.829

## 1

### Using the following function (which was downloaded with the data set), plot all intervals.

plot\_ci(lower\_vector, upper\_vector, mean(population))



### What proportion of your confidence intervals include the true population mean?

p <- 1-(2/50)  
p

## [1] 0.96

The proportion of my confidence intervals that included the true population mean was 96.

### Is this proportion exactly equal to the confidence level? If not, explain why.

No, this proportion does *not* exactly equal the confidence level. This is because confidence intervals don’t capture exact values. Instead, they predict a range that is *about as likely as the confidence interval* to contain the true value of the population.

## 2.

### Pick a confidence level of your choice (NOT 95% or 99%). What is the appropriate critical value? Run it and attach the output.

I’ll use a 90% confidence interval.

ci.90 <- qnorm(.95)  
ci.90

## [1] 1.644854

### How does it differs from the 95% confidence interval?

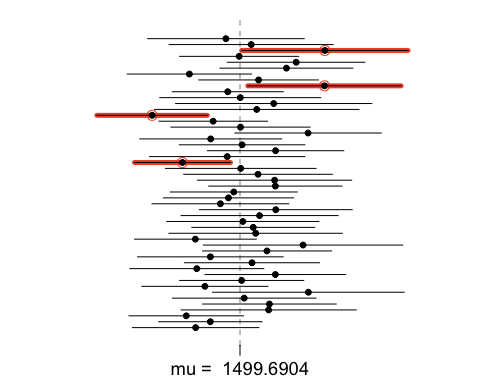
The critical value is 1.6448536. The critical value for a 95% confidence interval is approximately 1.96. This value is smaller. In application, this makes the inverval smaller, but this means a given sample is more likely to be outside the true population mean.

## 3. Calculate 50 confidence intervals at the confidence level you chose in the previous question. You do not need to obtain new samples, simply calculate new intervals based on the sample means and standard deviations you have already collected. Using the plot\_ci function, plot all intervals and calculate the proportion of intervals that include the true population mean. How does this percentage compare to the confidence level selected for the intervals?

set.seed(300)  
samp\_mean <- rep(NA, 50)  
samp\_sd <- rep(NA, 50)  
  
n <- 60  
  
for(i in 1:50){  
 samp <- sample(population, n)  
 samp\_mean[i] <- mean(samp)  
 samp\_sd[i] <- sd(samp)  
 }  
  
lower <- samp\_mean - ci.90 \* samp\_sd / sqrt(n)  
upper <- samp\_mean + ci.90 \* samp\_sd / sqrt(n)  
  
c(lower[1],upper[1])

## [1] 1330.165 1529.469

plot\_ci(lower, upper, mean(population))



q <- 1-(4/50)  
q

## [1] 0.92

Four of the 50 samples intervals did not include the true population mean of ~1499.69. This is 92%. This makes sense, given we set the CI to a lower level with 10% significance.