Week4-Lab IV

Aamol Gote

9/26/2019

download.file("http://www.openintro.org/stat/data/ames.RData", destfile = "ames.RData")  
load("ames.RData")

load("ames.RData")

population <- ames$Gr.Liv.Area  
samp <- sample(population, 60)

samp

## [1] 3194 1690 1483 1362 1558 1341 1362 848 1055 1416 1320 1144 1264 2454  
## [15] 1117 2596 907 1346 1688 1675 1103 797 1604 1992 1656 1192 1338 1280  
## [29] 972 1164 1671 2322 2787 1093 1154 2475 1142 1836 2049 1364 1869 1170  
## [43] 2473 1374 1682 1119 1585 1690 1064 767 1719 1616 1728 1217 1348 1131  
## [57] 1149 1293 1625 960

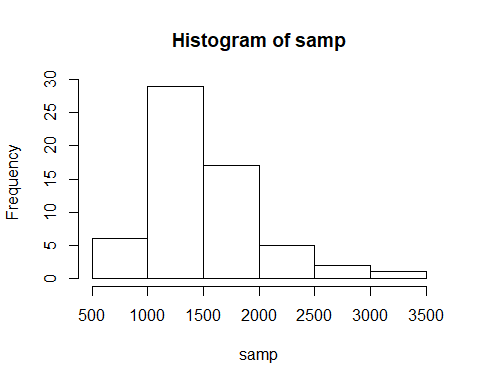
**Excercise 1**

Describe the distribution of your sample. What would you say is the “typical” size within your sample? Also state precisely what you interpreted “typical” to mean.

summary(samp)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 767 1148 1362 1506 1688 3194

hist(samp)



**Excercise 2**

Would you expect another student’s distribution to be identical to yours? Would you expect it to be similar? Why or why not?

Another student’s distribution would not be exactly identical, but it would be somewhat similar. Reason been this is randomly sampled and another student may have differet set of samples but as samples are taken from the same set of population there would be similarities between different samples.

**Excercise 3**

For the confidence interval to be valid, the sample mean must be normally distributed and have standard error s/√n What conditions must be met for this to be true?

1. Independent observations
2. Sampling size greater than 30
3. Data Not strongly skewed

All above are part of CLT (Central limit theorem) informal description

**Excercise 4**

What does “95% confidence” mean? If you’re not sure, see Section 4.2.2.

Suppose we take many samples and calculate confidence interval for each sample then 95% of the time the mean would be in between the calculated confidence interval for each sample.

mean(population)

## [1] 1499.69

**Excercise 5**

Does your confidence interval capture the true average size of houses in Ames? If you are working on this lab in a classroom, does your neighbor’s interval capture this value?

point estimate +- 1:96 \* SE

SE = SD/√n Sample Size = n = 60

Below is the 95% confidence range

sample\_mean <- mean(samp)  
se <- sd(population)/(sqrt(60))  
lowerbound <- sample\_mean - (1.96 \* se)   
upperbound <- sample\_mean + (1.96 \* se)  
c(lowerbound, upperbound)

## [1] 1378.589 1634.411

Actual mean

mean(population)

## [1] 1499.69

Actual mean falls in between the the confidence interval range and captures the true average size of houses.

Neighbours confidence interval should also capture the mean value in between them.

**Excercise 6**

Each student in your class should have gotten a slightly different confidence interval. What proportion of those intervals would you expect to capture the true population mean? Why? If you are working in this lab in a classroom, collect data on the intervals created by other students in the class and calculate the proportion of intervals that capture the true population mean.

95% confidence interval should capture the true population mean.

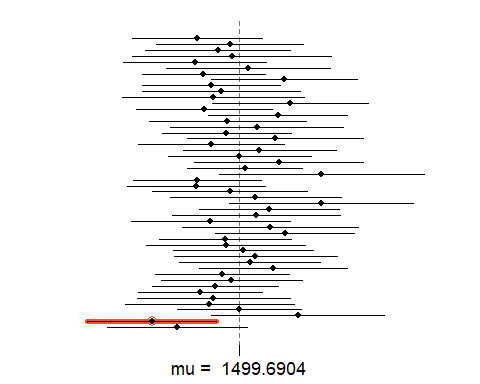
TODO……………

samp\_mean <- rep(NA, 50)  
samp\_sd <- rep(NA, 50)  
n <- 60  
for(i in 1:50){  
 samp <- sample(population, n) # obtain a sample of size n = 60 from the population  
 samp\_mean[i] <- mean(samp) # save sample mean in ith element of samp\_mean  
 samp\_sd[i] <- sd(samp) # save sample sd in ith element of samp\_sd  
}  
  
lower\_vector <- samp\_mean - 1.96 \* samp\_sd / sqrt(n)   
upper\_vector <- samp\_mean + 1.96 \* samp\_sd / sqrt(n)  
c(lower\_vector[1], upper\_vector[1])

## [1] 1278.466 1512.100

**On YOur Own** 1. Using the following function (which was downloaded with the data set), plot all intervals. What proportion of your confidence intervals include the true population mean? Is this proportion exactly equal to the confidence level? If not, explain why.

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



TODO…………….

1. Pick a confidence level of your choosing, provided it is not 95%. What is the appropriate critical value?

Confidence level = 10%

confidence\_level <- 90  
signifiance\_level <- 1-(confidence\_level/100)  
cp <- 1-(signifiance\_level/2)   
qnorm(cp)

## [1] 1.644854

Critical value is: 1.64

1. 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?

samp\_mean <- rep(NA, 50)  
samp\_sd <- rep(NA, 50)  
n <- 60  
for(i in 1:50){  
 samp <- sample(population, n) # obtain a sample of size n = 60 from the population  
 samp\_mean[i] <- mean(samp) # save sample mean in ith element of samp\_mean  
 samp\_sd[i] <- sd(samp) # save sample sd in ith element of samp\_sd  
}  
  
lower\_vector <- samp\_mean - 1.64 \* samp\_sd / sqrt(n)   
upper\_vector <- samp\_mean + 1.64 \* samp\_sd / sqrt(n)  
c(lower\_vector[1], upper\_vector[1])

## [1] 1416.482 1628.785

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

