lab5

jason grahn

10/24/2018

# R component

population <- ames$Gr.Liv.Area  
samp <- sample(population, 60)  
  
population.tidy <- ames %>%   
 select(Gr.Liv.Area)  
  
samp.tidy <- population.tidy %>%   
 rep\_sample\_n(size = 60, reps = 1)

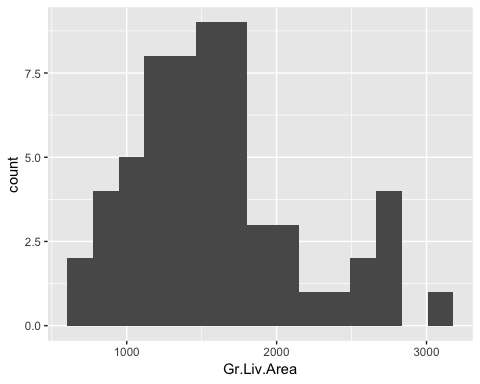
## Exercise 1

### 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.

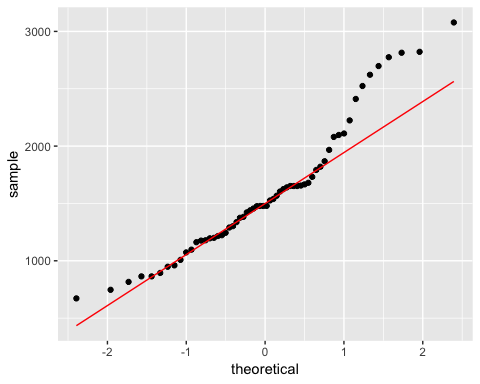
samp.describe <- describe(samp.tidy$Gr.Liv.Area)  
  
samp.histogram <- samp.tidy %>%   
 ggplot() +  
 geom\_histogram(aes(x = Gr.Liv.Area), bins = 15)  
  
samp.qqplot <- samp.tidy %>%   
 ggplot(aes(sample = Gr.Liv.Area)) +   
 stat\_qq() +  
 stat\_qq\_line(color = "red")  
  
samp.describe

## vars n mean sd median trimmed mad min max range skew  
## X1 1 60 1589.13 568.02 1478.5 1535.04 431.44 672 3078 2406 0.8  
## kurtosis se  
## X1 0.02 73.33

samp.histogram



samp.qqplot



The sample appears somewhat normal but with a very strong right skew as we can tell with a skewness of 1.02. The median is 1498 with a mean of 1574.

## Exercise 2

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

I definitely would *NOT* expect another students distribution to be identical because we pulled the descriptive statistics from a sample of 60 out of a population of 2,930. If the population is truly normal, then their sample would ideed be similar.

## Exercise 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?

For the confidence interval of a sample to be true, the sample must be random and independent. The sample size needs to be greater than 30, and there should be no skewness to the distribution.

## Exercise 4

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

The term “95% confidence” means that 95% of the time our point estimates will be within 2 standard errors of the parameter for that statistic.

## Exercise 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?

pop.mean <- mean(population.tidy$Gr.Liv.Area)   
  
#baseR  
sample\_mean <- mean(samp)  
se <- sd(samp) / sqrt(60)  
lower <- sample\_mean - 1.96 \* se  
upper <- sample\_mean + 1.96 \* se  
c(lower, upper)

## [1] 1347.676 1606.758

The population mean is 1499.6904437. The confidence interval of my sample captures the population mean, yes. I am not working on this within a classroom so I am unable to answer for my neighbors.

## Exercise 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.

We aren’t in a classroom. I would expect 95% of the intervals to capture the true population mean because that’s how confidence intervals work.