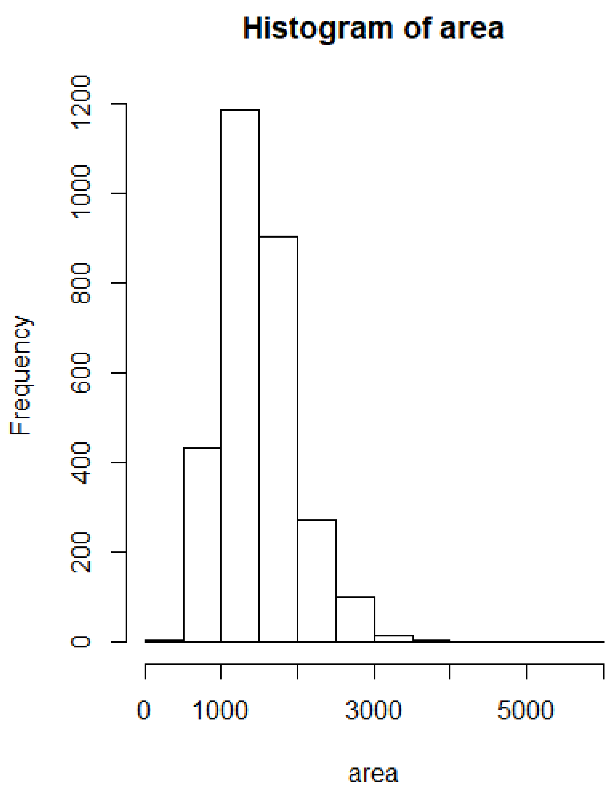
Name: Connor Lauerman Collaborators:

R Lab 5: Introduction to Inference: Sampling Distributions

Please answer all the Exercises and the questions from the “On Your Own” section. If you use any graphs or charts to justify your answer, please include them.

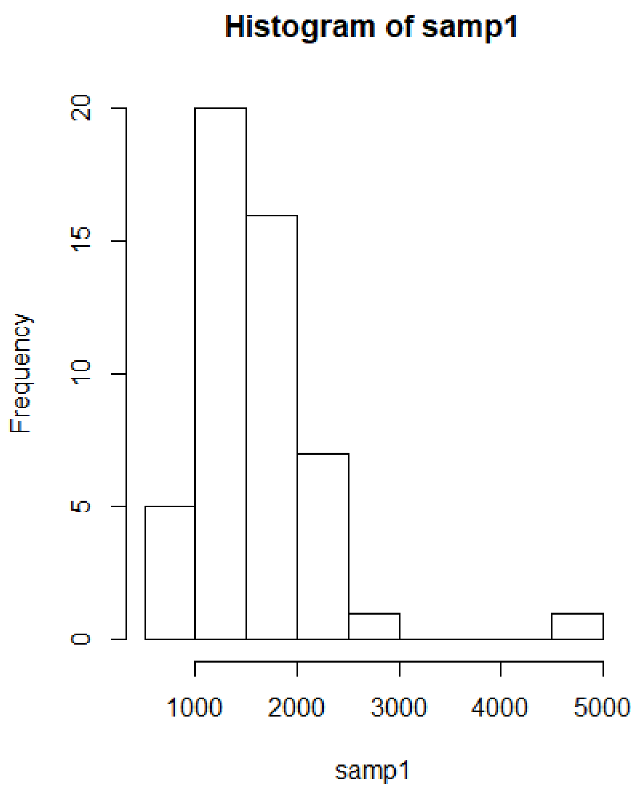
Exercise 1: Describe this population distribution. (Include plot.)

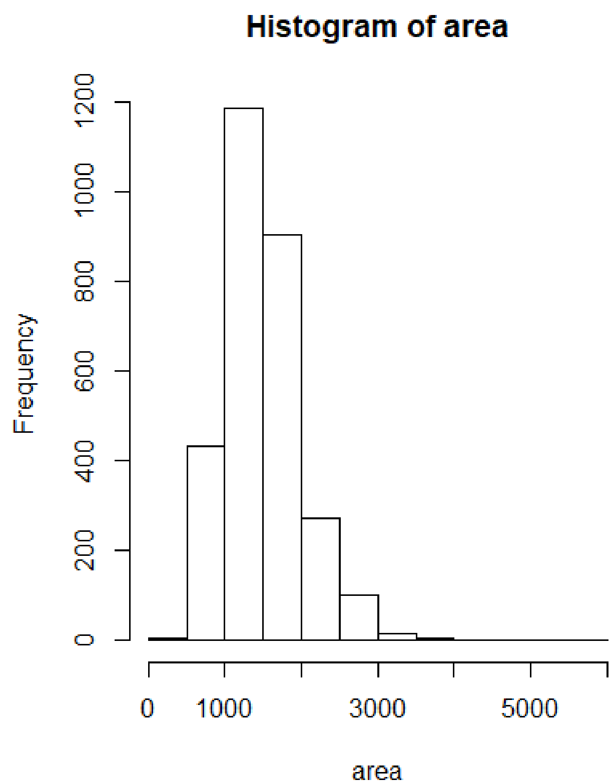
The distribution is somewhat narrow with higher frequencies from 1000-2000. There appears to me a small tail to the right so the data is minorly skewed right.



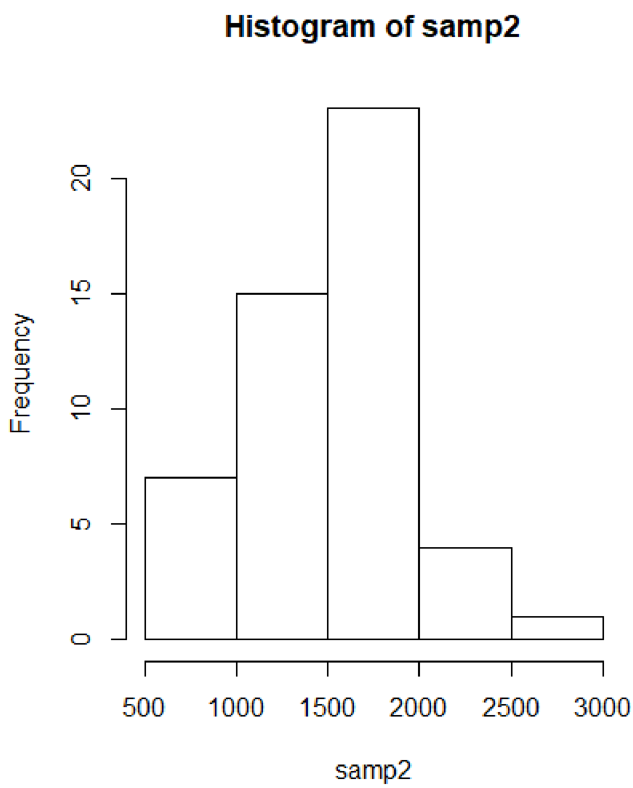
Exercise 2: Describe the distribution of this sample. How does it compare to the distribution of the population? (Include plots.)

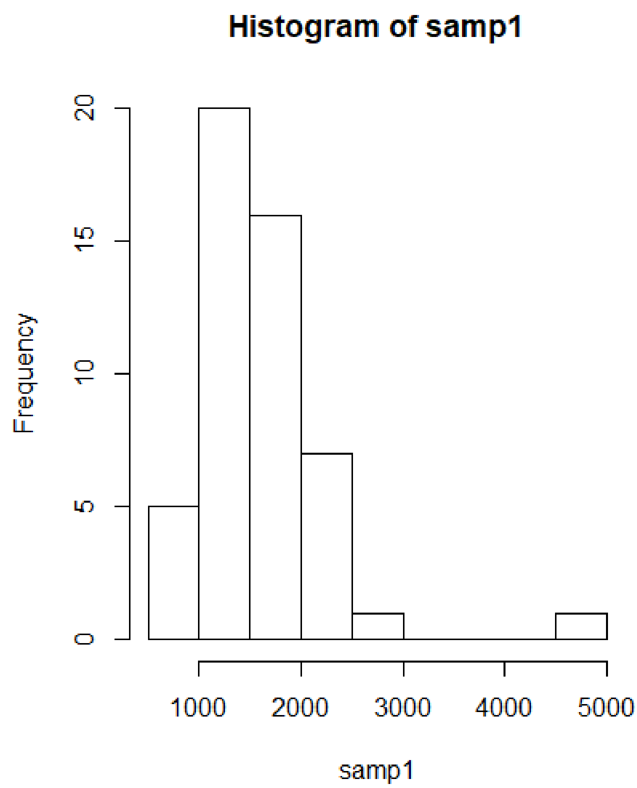
The distribution of the sample is concentrated from 1000-2000 and is very similar to the population although my sample has an outlier and a slightly higher mean than the population.





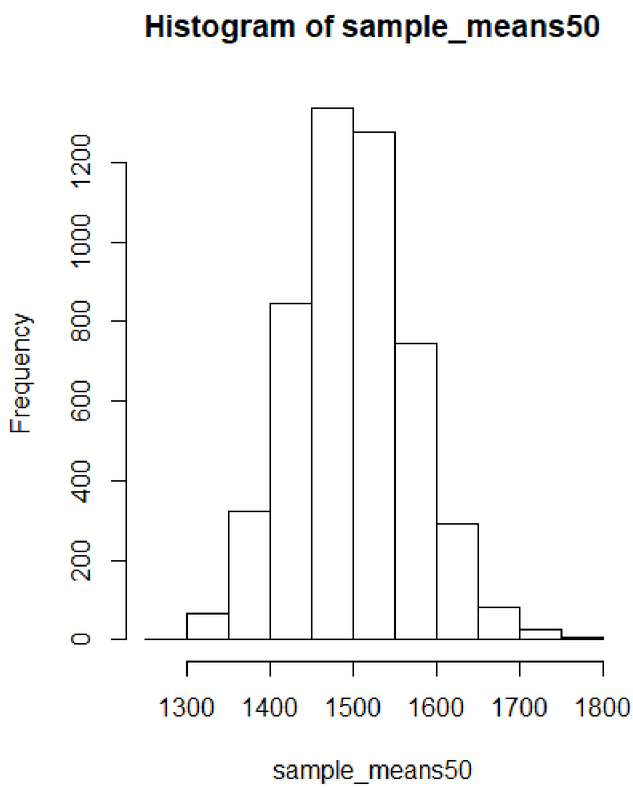
Exercise 3: Take a second sample, also of size 50, and call it `samp2`. How does the mean of `samp2` compare with the mean of `samp1`? (Include plots.) Suppose we took two more samples, one of size 100 and one of size 1000. Which would you think would provide a more accurate estimate of the population mean?
The mean of my second sample is smaller than the mean of my second sample. 1612>1490.
I think that 1000 would provide the more accurate estimate because as the sample grows it is more representative of the population similar to how the law of large numbers shows the correct probability with increased trials.





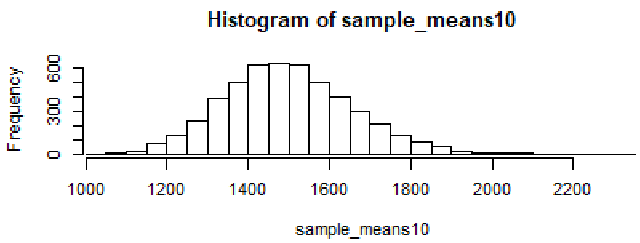
Exercise 4: How many elements are there in `sample_means50`? Describe the sampling distribution, and be sure to specifically note its center. (Include plot.) Would you expect the distribution to change if we instead collected 50,000 sample means?

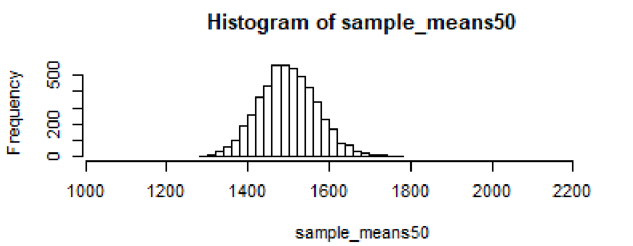
There are 5000 elements in sample_means50. The sample distribution appears to be normal with a center of about 1500. I would not expect the distribution to change since the central limit theorem states that the distribution of sample means is normal

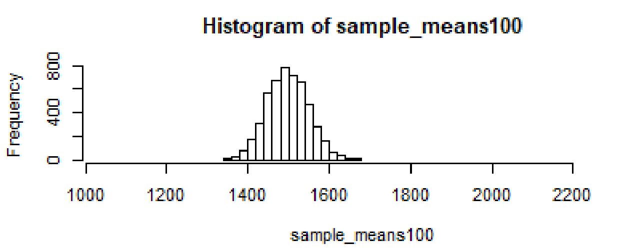


Exercise 5: To make sure you understand what you've done in this loop, try running a smaller version. Initialize a vector of 100 zeros called `sample_means_small`. Run a loop that takes a sample of size 50 from area and stores the sample mean in `sample_means_small`, but only iterate from 1 to 100. Print the output to your screen (type `sample_means_small` into the console and press enter). How many elements are there in this object called `sample_means_small`? What does each element represent?
There are 100 elements in the variable, and each element represents a mean of a sample of size 50

Exercise 6: When the sample size is larger, what happens to the center? What about the spread? (Include plots.)
When the sample size is larger the center has more frequencies but remains the same. The spread becomes narrower and taller as the sample size increases .





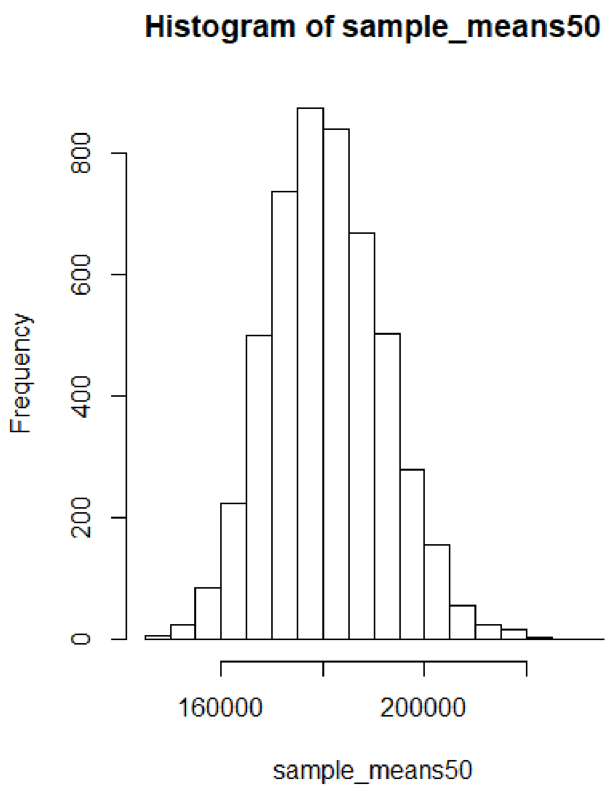


On Your Own:
1) Take a random sample of size 50 from `price`. Using this sample, what is your best point estimate of the population mean?

The best point estimate is 188083

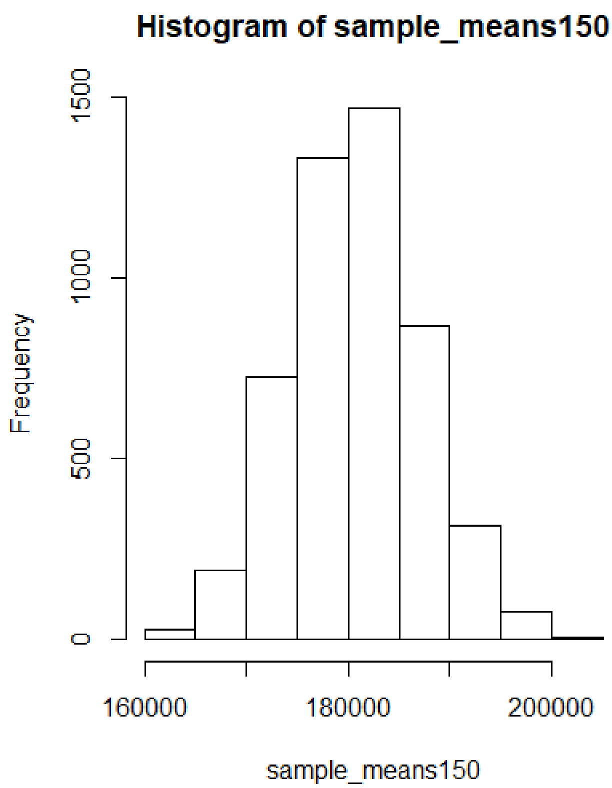
2) Since you have access to the population, simulate the sampling distribution for $\bar{x}_{price}$ by taking 5000 samples from the population of size 50 and computing 5000 sample means. Store these means in a vector called `sample_means50`. Plot the data, then describe the shape of this sampling distribution. Based on this sampling distribution, what would you guess the mean home price of the population to be? Finally, calculate and report the population mean. (Include plot.)

The sampling distribution is bell shaped and mostly symmetric and appears to be normally distributed.
My guess for the mean is 180868
The actual mean is 180796



3) Change your sample size from 50 to 150, then compute the sampling distribution using the same method as above, and store these means in a new vector called `sample_means150`. Describe the shape of this sampling distribution, and compare it to the sampling distribution for a sample size of 50. Based on this sampling distribution, what would you guess to be the mean sale price of homes in Ames? (Include plot.)

The distribution is bell shaped and relatively symmetric. Compared to sample size of 50 this plot is less spread and taller.
From this sample my guess of mean is 180752



4) Of the sampling distributions from 2 and 3, which has a smaller spread? If we're concerned with making estimates that are more often close to the true value, would we prefer a distribution with a large or small spread?

The smaller spread is the distribution from 3. We would prefer estimates with small spread because it is closer to the true value and the mean is also closer to the true mean.