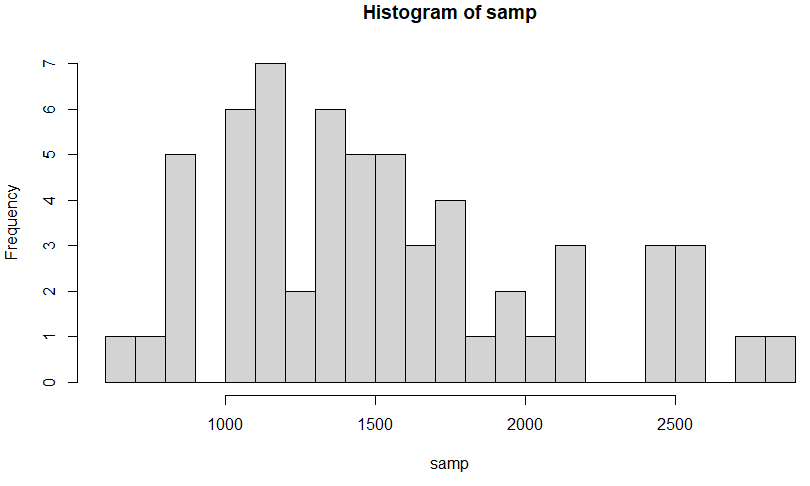
Lab 5

**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.**

hist(samp, breaks = 25)

summary(samp)

Min. 1st Qu. Median Mean 3rd Qu. Max.

 694 1139 1422 1538 1798 2828

The sample shows a slightly right skewed distribution with a lack of symmetry, and a center around 1400-1600. This adds up because the median and mean fall within this range. The shape is unimodal. The typical size in this sample would be where most data shows up like the mode, which is approximately between 1000-1200.

**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 would not expect another student’s distribution to be identical to mine because it is a random sample. There would be a very low chance that another student got the same random sample of 60 house sizes. It would be similar because the sample size is 60 because 60 is a large enough number to lower the chances of variability and standard error.

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

The data distribution for the sample mean will be normal if the sampled values are independent of each other, randomly sampled, take up less than 10% of the total population, and have a sample size that is usually 30 or above.

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

95% confidence means that we are 95% confident that the population’s unknown parameter falls within a certain range based on a representative sample that follows the CLT. 95% of the time, the values will be within two standard errors of the parameter.

**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?**

sample\_mean <- mean(samp)

se <- sd(samp) / sqrt(60)

lower <- sample\_mean - 1.96 \* se

upper <- sample\_mean + 1.96 \* se

c(lower, upper)

= 1400.426, 1675.508

mean(population)= 1499.69

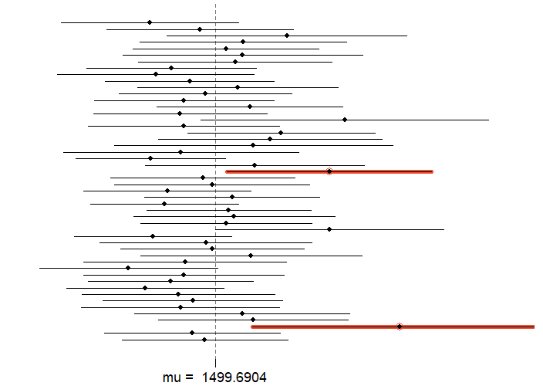
Yes, the confidence interval captures the true average size of houses in Ames, since the mean of the population fell within the boundaries of the sample’s 95% confidence interval.

**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.**

I would expect about 95% of the class to have intervals that capture the true population mean if all criteria were met. This is because we are working with a 95% confidence interval, which is slightly narrower than a 99% confidence interval. It would not collect 5% of data if, say, 100 samples were calculated. For the real class data, 17/19 (0.895 = 89.5%) of the students captured the mean in their confidence intervals.

**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))**

All the confidence intervals besides 2 contain the true population mean, which means 48/50 contained the true population mean.

48/50 = 0.96 = 96%.

This proportion is not exactly equal to the 95% confidence interval. We only created 50 random samples of size 60. If we had taken more samples, then the proportion would be closer to the confidence interval.

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

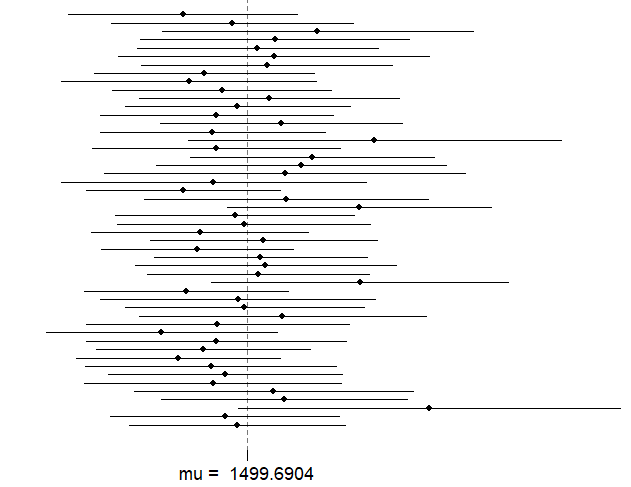
I chose 99%, and the appropriate critical value is 2.58.

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?**

lower\_vector <- samp\_mean - 2.58 \* samp\_sd / sqrt(n)

upper\_vector <- samp\_mean + 2.58 \* samp\_sd / sqrt(n)

c(lower\_vector[1], upper\_vector[1]) = 1339.423, 1631.077

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

50/50 of the confidence intervals now contain the true population mean.

50/50 = 1.0 = 100%.

This percentage is very close to the 99% confidence interval, but it is not equal. This can again be attributed to only having 50 random samples of 60 random house sizes. If the sample was maybe 100, 200, or bigger, then we would likely see a plot that accurately shows 99% of the confidence intervals containing the true population mean.