

Lab #5: Confidence intervals

Team name

Date of lab session

Lab report

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

Load data:

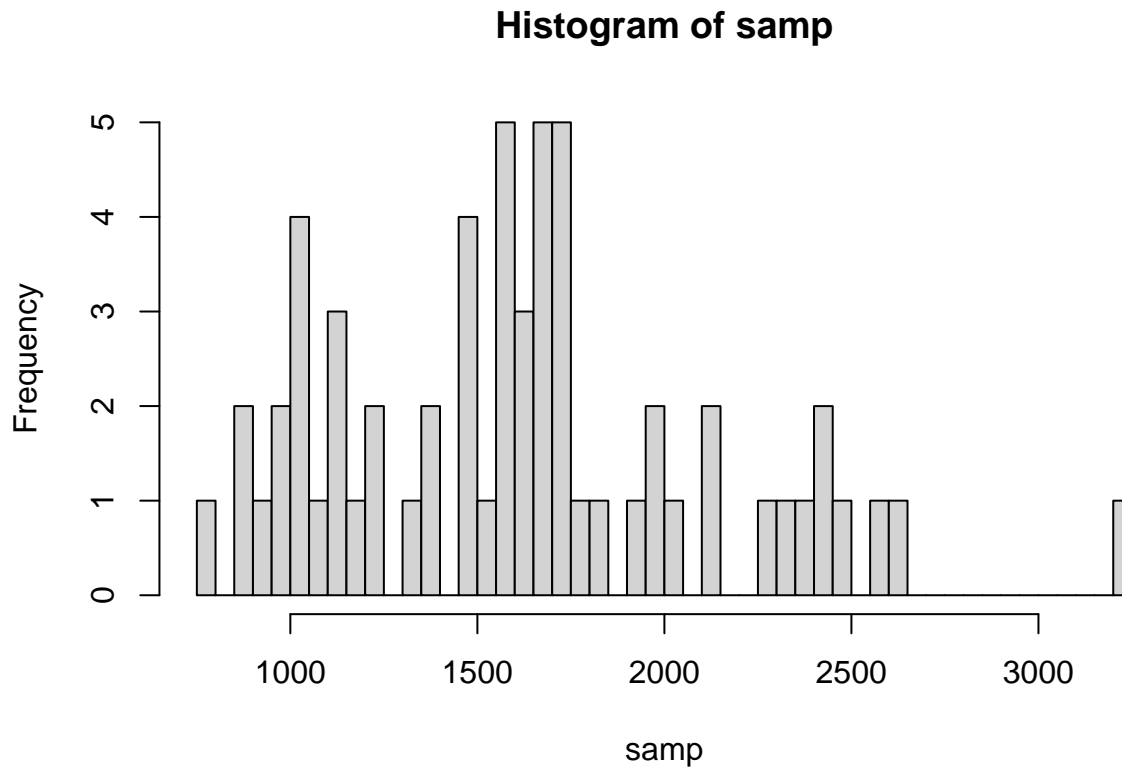
```
set.seed(432)
```

Set a seed:

Exercises:

Exercise 1: Based on this histogram, we can see that the distribution is skewed right due to outliers.

```
population <- ames$Gr.Liv.Area
samp <- sample(population, 60)
hist(samp, breaks=40)
```



Exercise 2: No, it would not be identical, however, I believe it would be similar. This is because every sample is similar, but none would be identical as the sample is random every time.

Exercise 3: Normality and Independence, these can be met if you have a sufficiently large sample size and random sampling.

Exercise 4: If we were to take infinite samples, 95% of them would contain the true population mean value.

Exercise 5: Yes, my confidence interval includes the true population mean.

```
mean(population)
```

```
## [1] 1499.69
```

```
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] 1492.704 1752.162
```

Exercise 6: I believe roughly 95% of them would include the true population mean. As I mentioned earlier, according to the definition of a confidence interval, 95% of all samples would include the true population mean, this would be relatively the same, even if the number of samples is limited to the number of students in this class. However, this does not mean it will exactly be 95% of them, but a relatively similar proportion.

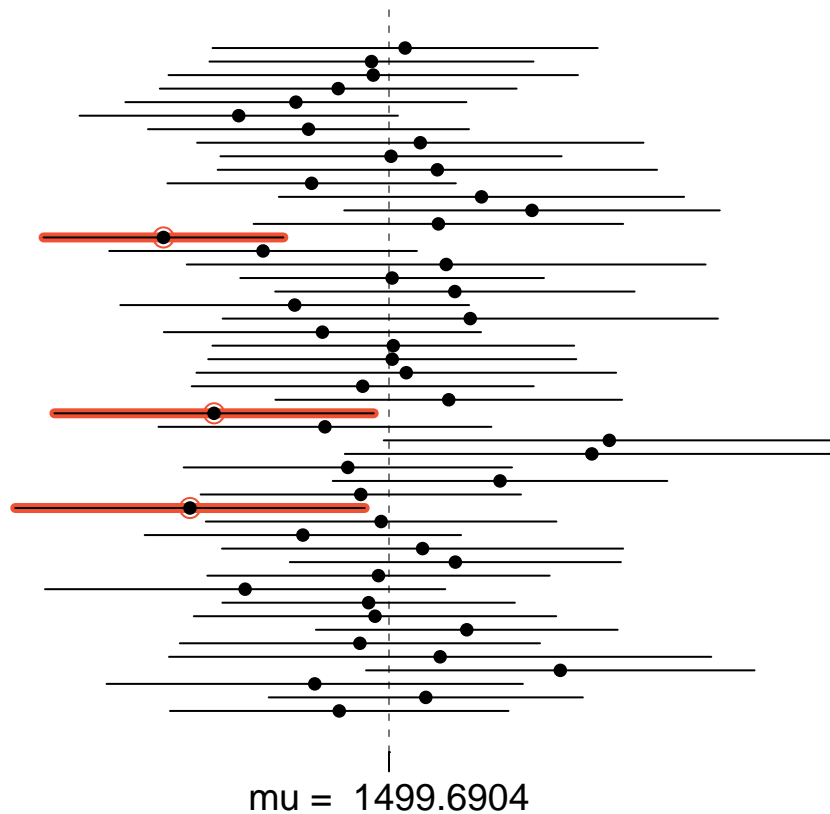
On your own:

1: 88% of the samples have the true population mean within them. The proportion does not match the confidence interval, as it is only 88% compared to the 95% the code is designed for.

```
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] 1348.079 1582.421
```

```
plot_ci(lower_vector, upper_vector, mean(population))
```



2: For 99% confidence the critical value would be 2.576.

3: My plot had 49 out of 50, which means 98% has the true population mean, compared to the 99% confidence interval, which is alot closer than the previous.

```
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 - 2.576 * samp_sd / sqrt(n)
upper_vector <- samp_mean + 2.576 * samp_sd / sqrt(n)
c(lower_vector[1], upper_vector[1])
```

```
## [1] 1410.134 1725.100
```

```
plot_ci(lower_vector, upper_vector, mean(population))
```

