

Central Limit Theorem

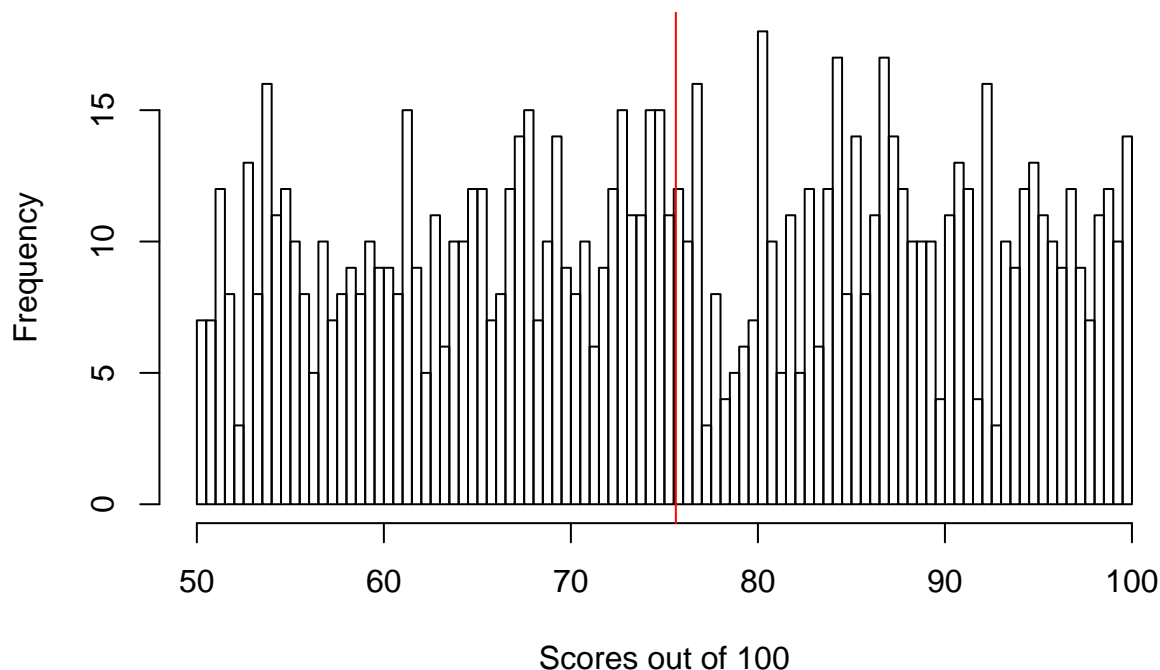
Exploration of the Central Limit Theorem

The Central Limit Theorem is one of the most fundamental concepts in statistics. In this document, I will be visualizing the central limit theorem to better drive my understanding of this concept.

Assume we have a 'population' of 1000 students in a university who have completed a placement exam, and received scores following the distribution below (assume no one received a grade below 50/100, and the highest grade is 100/100). The mean of the distribution is denoted by the red vertical line:

```
test_dist <- runif(1000, 50, 100)
mean_score <- mean(test_dist)
hist(test_dist,
      main = 'Exam Distribution for 5000 Students',
      xlab = 'Scores out of 100',
      breaks = 100)
abline(v=mean_score,col="red")
```

Exam Distribution for 5000 Students



Given this population, we hope to run an experiment to try and approximate the scores of this class of 100 students. Since we “can’t” sample all 100 students immediately, we decide our best course of action is to complete simple random sampling (SRS). Lets start with a sample of size of 3, and draw 5 samples using the `sample()` function of base R. We can write a function to accomplish this task.

```
# Function takes the sample size N, the number of sampling repetitions to complete and distribution to  
samplel <- function(n, reps, dist) {  
  sample_dist <- c(rep(c(sample(dist, n)), reps))  
  sample_means <- lapply(sample_dist, mean)  
  return(c(sample_dist, sample_means))  
}
```

Plotting the mean of these samples gives us the following sample distribution:

```
means <- rep(mean(sample(test_dist, 3)), 10)
```