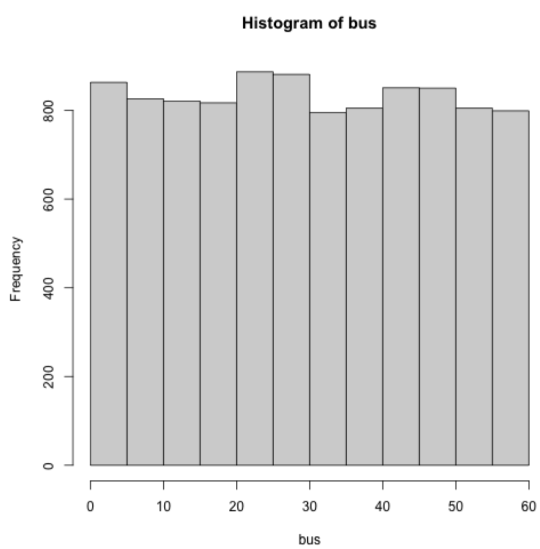
Chapter 6

Example 1: What does the distribution of the sample means of independent (continuous) uniform random variables look like? The Central Limit Theorem tells us that for a large sample size, the distribution of such means will be approximately normally distributed. Let's see this in action.

Suppose we know a bus arrives once per hour at a designated stop but we don't know when the next bus is and we don't know when the last bus came. We will model this as a continuous uniform[0,60] random variable. In other words, the next bus could come in anywhere from 0 to 60 minutes with all possibilities being equally likely. In the code below, we simulate 10,000 uniform[0,60] random variables and look at the histogram of results.

```
# Generate 10,000 uniform[0,60] random varibles
set.seed(2020)
bus = runif(10000,0,60)
hist(bus)
```



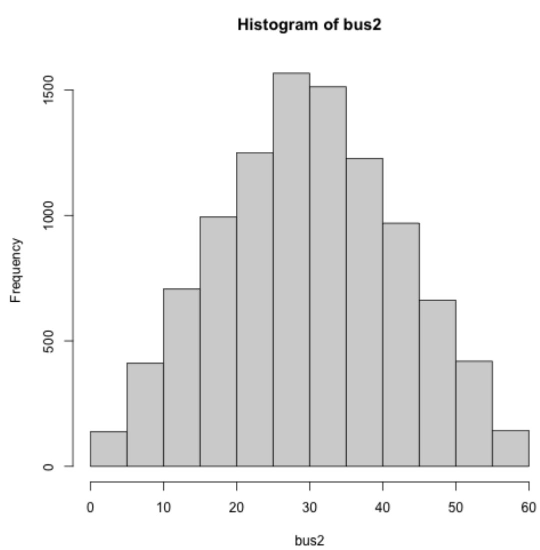
The mean and variance of this sampling distribution is calculated below.

```
mean(bus)

## [1] 29.79179

var(bus)

## [1] 298.1569
```

From chapter 4, we know that if $X \sim Uniform[a,b]$, then $E[X] = \frac{b+a}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$. For a uniform[0,60] random variable, we get the theoretical values of $E[X] = \frac{60+0}{2} = 30$ and $Var(X) = \frac{(60-0)^2}{12} = 300$. Our simulated values are very close to these theoretical values.

Now, let's suppose we have to wait for two buses, each being modeled as an independent uniform[0,60]. Let's look at the sampling distribution for the mean of the two bus wait times.

```
set.seed(2020)
bus2 = rep(0, 10000) # initialize this vector to zeros

# 10,000 simulations of 2 bus waits
for (i in 1:10000){
x1 = runif(2,0,60) # n=2
bus2[i] = mean(x1)
}

mean(bus2)

## [1] 29.90972

var(bus2)

## [1] 149.6154

hist(bus2)
```

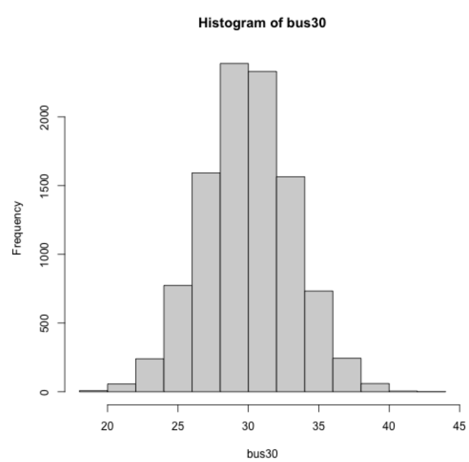


Notice that the average of two bus waits is no longer uniform and instead has more of a triangular shape. Our simulations of two bus waits had a mean of 30.05955 and variance of 151.2226. From chapter 6, we know that $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$.

For this simulation, the theoretical values for these quantities are $E[\bar{X}] = \mu = 30$ and $Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{300}{2} = 150$. Again, our simulations give values very close to the theoretical values.

Finally, let's suppose that we repeat the experiment for 30 days straight and look at the distribution of sample means of 30 days. We will simulate these 30 day periods 10,000 times.

```
set.seed(2020)
bus30 = rep(0, 10000) # initialize this vector to zeros

# 10,000 simulations of 2 bus waits
for (i in 1:10000){
x2 = runif(30,0,60) # n=30
bus30[i] = mean(x2)
}

mean(bus30)

## [1] 29.97879

var(bus30)

## [1] 10.19619

hist(bus30)
```



Our simulations of 30 bus waits had a mean of 30.00598 and variance of 10.08975. The theoretical values for this simulation are $E[\bar{X}] = \mu = 30$ and $Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{300}{30} = 10$.

Notice that the distribution of sample means is narrower for n=30 than n=2 and the variance is lower when a larger sample size is used. Also note that the distribution is somewhat bell-shaped.

In the coming chapters on Inferential Statistics, we will use the fact that the Central Limit Theorem states that the sampling distribution of sample means is approximately normally distributed.