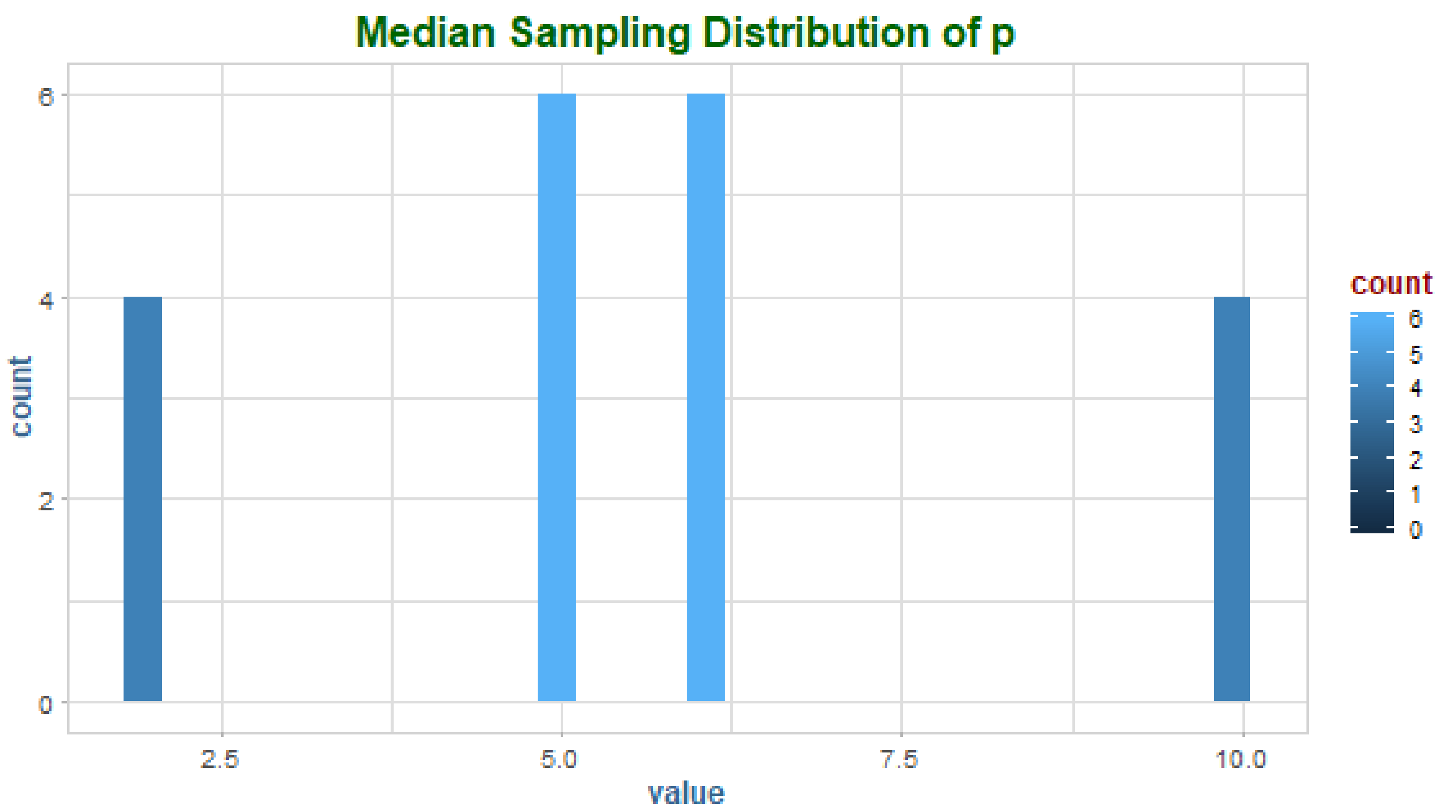
### 4.1

Consider the population {1, 2, 5, 6, 10, 12}.

Find (and plot) the sampling distribution of medians for samples of size 3 without replacement.

```
p <- c(1, 2, 5, 6, 10, 12)
c <- combinations(v = p, n = 6, r = 3)
t <- apply(c, 1, median)

ggplot(data.table(value = t), aes(value, fill = ..count..)) +
  geom_histogram(bins = 30) +
  labs(title = "Median Sampling Distribution of p")
```




Compare the median of the population to the mean of the medians.

Median of p = **5.5**. Mean of Medians of p = **5.7**

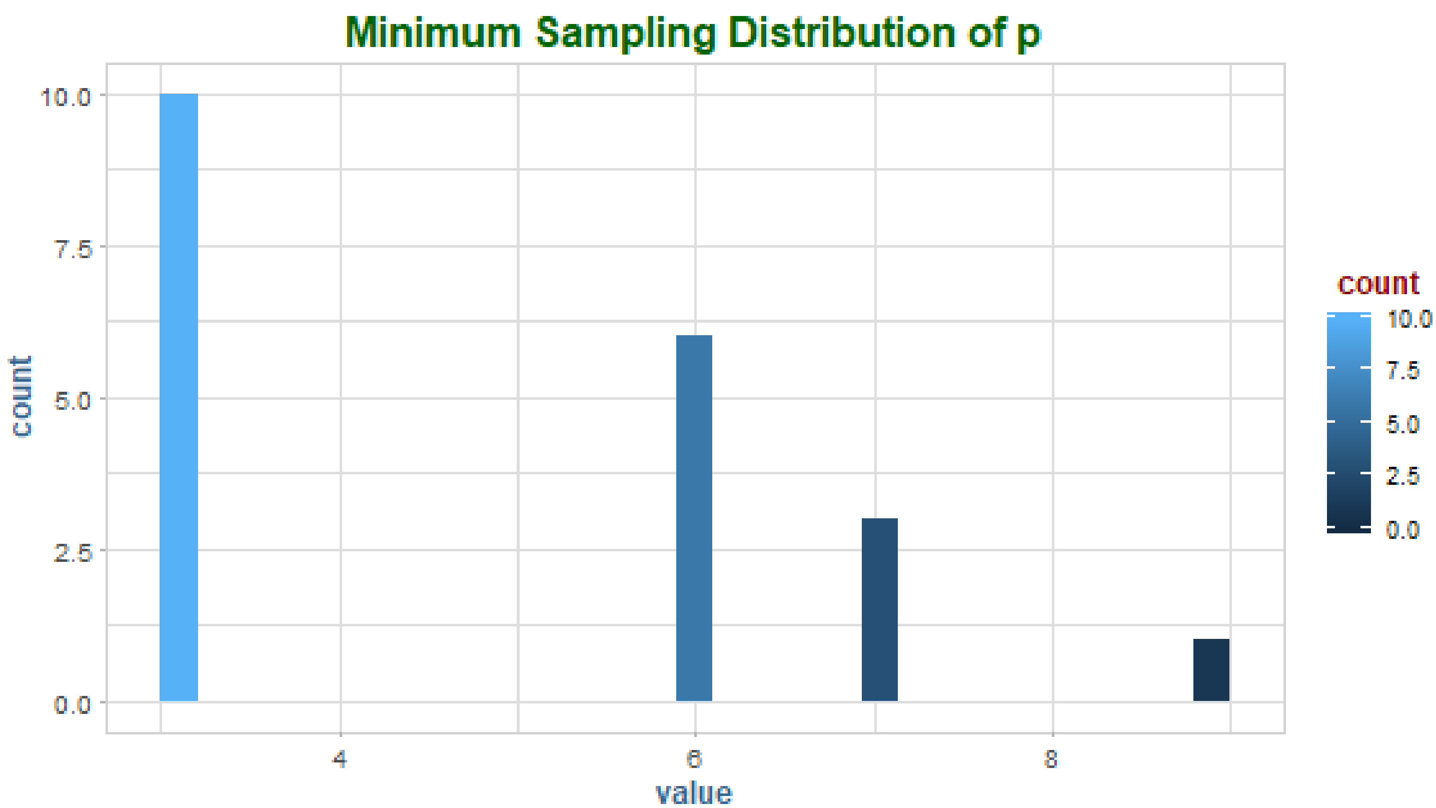
### 4.2

Consider the population {3, 6, 7, 9, 11, 14}.

For samples of size 3 without replacement, find (and plot) the sampling distribution for the minimum.

```
p <- c(3, 6, 7, 9, 11, 14)
c <- combinations(v = p, n = 6, r = 3)
t <- apply(c, 1, min)

ggplot(data.table(value = t), aes(value, fill = ..count..)) +
   geom_histogram(bins = 30) +
   labs(title = "Minimum Sampling Distribution of p")
```




What is the mean of the sampling distribution? **4.8**

The statistic is an estimate of some parameter - what is the value of that parameter?

This is an estimation of the minimum, which is: **3**

### 4.3

Let *A* denote the population {1, 3, 4, 5} and *B* the population {5, 7, 9}.

```
A <- c(1, 3, 4, 5)
B <- c(5, 7, 9)
```

Let *X* be a random value from *A*, and *Y* and random value from *B*.

a.) Find the sampling distribution of *X* + *Y*.

```
result = numeric(12)
index <- 1
for(j in 1:length(A))
{
   for(k in 1:length(B))
   {
      result[index] <- A[j] + B[k]
      index <- index + 1
   }
}

sort(result)
```

```
 [1]  6  8  8  9 10 10 10 11 12 12 13 14
```

b.) In this example, does the sampling distribution depend on whether you sample with or without replacement?

*No.*

Why or why not?

*Because 5 in is both sets.*

c.) Compute the mean of the values for each of *A* and *B* and the values in the sampling distribution of *X* + *Y*.

Mean of *A*: **3.25**. Mean of *B*: **7**.

Mean of *A* + *B*: **10.25**

How are the means related?

mean(A) + mean(B) = mean(A + B).

d.) Suppose you draw a random value from *A* and a random value from *B*.

```
prob <- sum(result >= 13) / length(result)
```

What is the probability that the sum is 13 or larger? **16.6666667%**

### 4.4

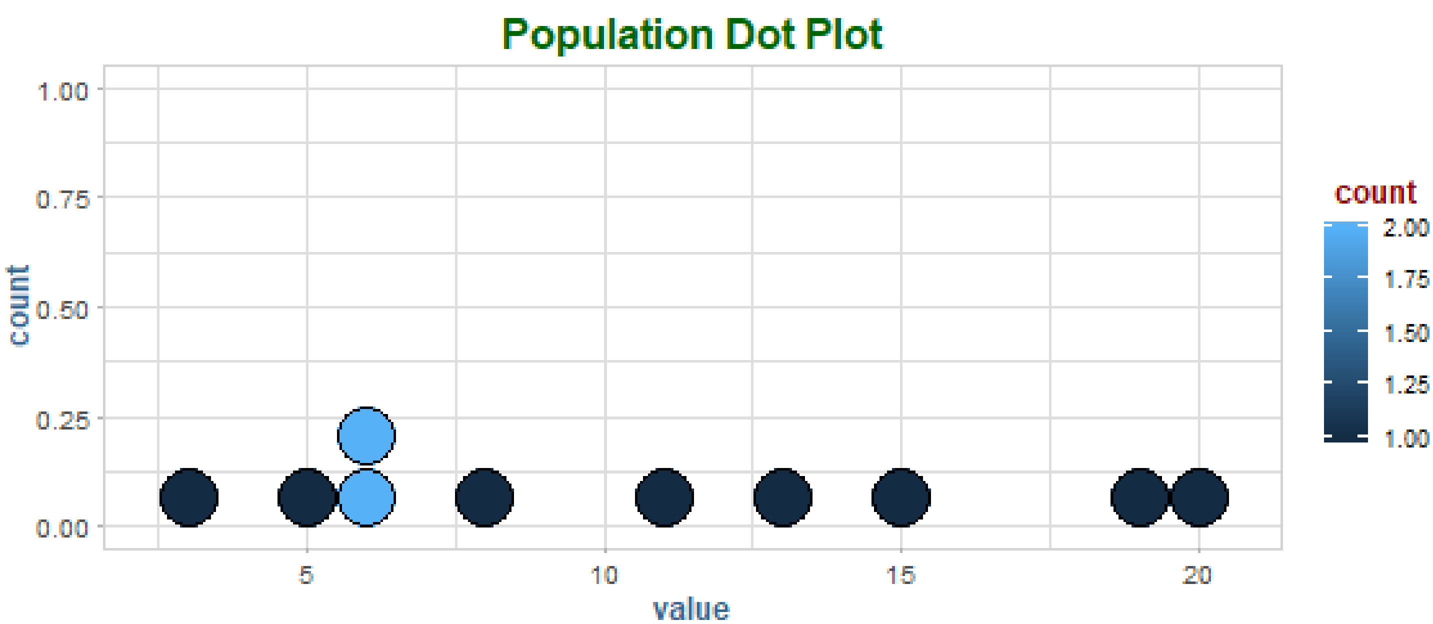
Consider the population {3, 5, 6, 6, 8, 11, 13, 15, 19, 20}.

a.) Compute the mean and standard deviation and create a dot plot of its distribution.

```
p <- c(3, 5, 6, 6, 8, 11, 13, 15, 19, 20)

mu <- mean(p)
sigma <- sd(p)

ggplot(data.table(value = p)) +
   geom_dotplot(aes(value, fill = ..count..), binwidth = 1) +
   labs(title = "Population Dot Plot")
```



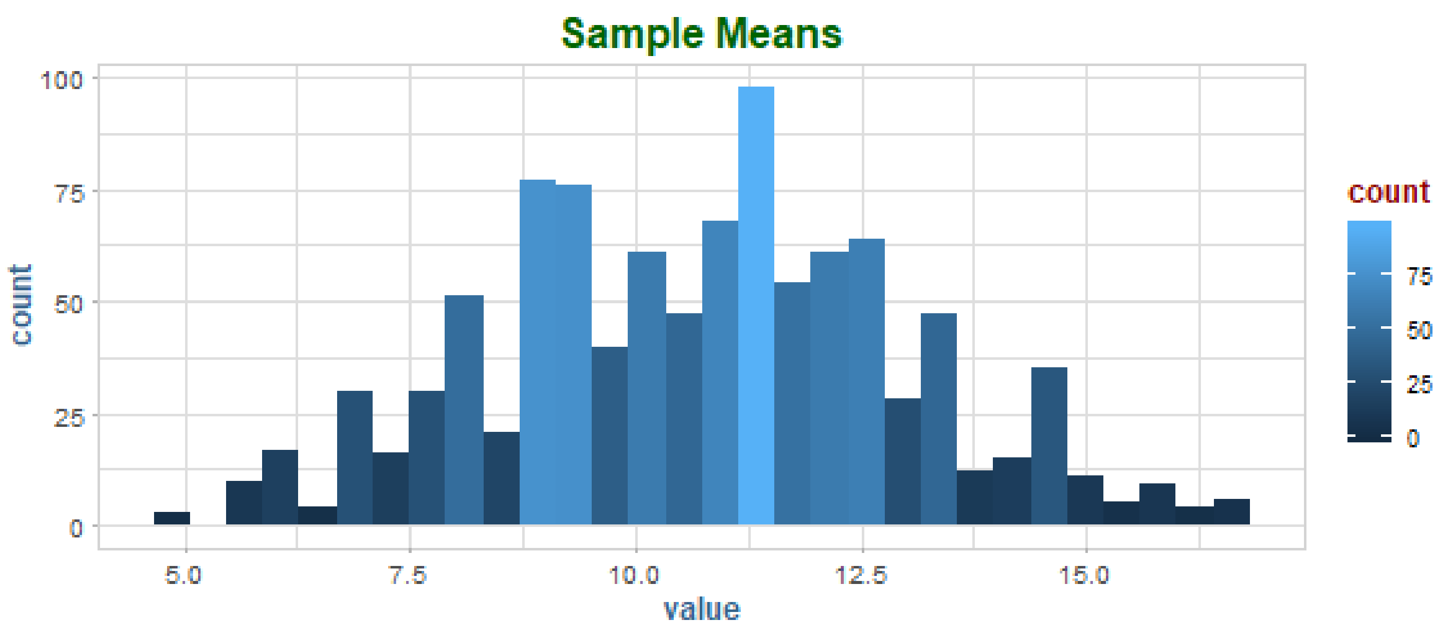

$\mu = 10.6, \sigma = 5.9851668$

b.) Simulate the sampling distribution of $\bar{X}$ by taking random samples of size 4 and plot your results.

```
N <- 10e2
results <- numeric(N)

for( i in 1:N)
{
   index <- sample(length(p), size = 4, replace = F)
   results[i] <- mean( p[index] )
}
```

```
ggplot(data.table(value = results)) +
   geom_histogram(aes(value, fill = ..count..), bins = 30) +
   labs(title = "Sample Means")
```




```
xbar <- mean(results)
se <- sd(results) / sqrt(N)
```

Compute the mean and standard error, and compare to the population mean and standard deviation.

mean: 10.704, standard error: 0.0723801

c.) Use the simulation to find $P(\bar{X} < 11)$.

```
prob <- mean(results < 11)
```

$P(\bar{X} < 11) = 50.9\%$