5.1

Consider the samples 1-6. Use a six-sided die to obtain three different bootstrap samples and their corresponding means.

```
pop <- seq(from = 1, to = 6, by = 1)

n <- 6

s1 <- mean( sample(pop, n, replace = T) )
s2 <- mean( sample(pop, n, replace = T) )
s3 <- mean( sample(pop, n, replace = T) )
```

$\bar{x}_1^* = 2.5, \bar{x}_2^* = 3.3333333, \bar{x}_3^* = 4.3333333$

5.2

Consider the samples 1, 3, 4, and 6 from some distribution.

```
pop <- c(1, 3, 4, 6)

samples <- permutations(n = 4, r = 4, pop, repeats.allowed = T)
```

a.) For one random bootstrap sample, find the probability that the mean is one.

```
means <- apply(samples, 1, mean)

p <- mean( means == 1 )
```

Probability: **0.39%**

b.) For one random bootstrap sample, find the probability that the maximum is 6.

```
maxes <- apply(samples, 1, max)

p <- mean( maxes == 6 )
```

Probability: **68.36%**

c.) For one random bootstrap sample, find the probability that exactly two elements in the sample are less than 2.

```
lt2 <- apply(t(apply(samples, 1, function(x) { x < 2})), 1, sum)

p <- mean( lt2 == 2 )
```

Probability: **21.09%**

5.3

Consider the sample 1-3.

a.) List all the (ordered) bootstrap samples from this sample. How many are there?

```
samples <- permutations(n = 3, r = 3, 1:3, repeats.allowed = T)

n <- nrow(samples)
```

Samples: = 3^3 = **27**

b.) How many unordered bootstrap samples are there? For example, {1, 2, 2} and {2, 1, 2} are considered to be the same.

```
samples <- combinations(n = 3, r = 3, 1:3, repeats.allowed = T)

n <- nrow(samples)

assertthat::are_equal(n, choose(3 + 3 - 1, 3))
```

```
[1] TRUE
```

Samples: = $\binom{5}{3}$ = **10**

c.) How many ordered bootstrap samples have one occurrence of 1 and two occurences of 3?

```
samples <- permutations(n = 3, r = 3, 1:3, repeats.allowed = T)

n.ones <- apply(t(apply(samples, 1, FUN = function(x) { x == 1 })), 1, function(x) sum(x) )
n.threes <- apply(t(apply(samples, 1, FUN = function(x) { x == 3 })), 1, function(x) sum(x) )

sum((n.ones == 1 & n.threes == 2) == T)
```

```
[1] 3
```

Is this the same number of bootstrap samples that have each of 1, 2 and 3 occuring exactly once?

```
n.ones <- apply(t(apply(samples, 1, FUN = function(x) { x == 1 })), 1, function(x) sum(x) )
n.twos <- apply(t(apply(samples, 1, FUN = function(x) { x == 2 })), 1, function(x) sum(x) )
n.threes <- apply(t(apply(samples, 1, FUN = function(x) { x == 3 })), 1, function(x) sum(x) )

sum((n.ones == 1 & n.twos == 1 & n.threes == 1) == T)
```

```
[1] 6
```

No, 3 != 6.

d.) Is the probability of obtaining a bootstrap sample with one 1 and two 3's the same as the probabiliity of obtaining a bootstrap sample with each of 1, 2 and 3 occuring exactly once?

```
( sum((n.ones == 1 & n.threes == 2)) / n ) == ( sum((n.ones == 1 & n.twos == 1 &  n.threes ==
```

```
[1] FALSE
```

No, 3% and 6% chances respectfully.

5.4

Consider the samples 1, 3, 3, and 5 from some distribution.

```
samples <- c(1, 3, 3, 5)
```

a.) How many bootstrap samples are there?

```
boot <- permutations(n = 3, r = 4, v = samples, repeats.allowed = T)

n <- nrow(boot)
```

3 unique items to pick from, 4 places to put each item.

Number of permutations: 3^4 = *81*

b.) List the distinct bootstrap samples assuming order does not matter.

```
combinations(n = 3, r = 4, v = samples, repeats.allowed = T)
```

```
        [,1] [,2] [,3] [,4]
 [1,]    1    1    1    1
 [2,]    1    1    1    3
 [3,]    1    1    1    5
 [4,]    1    1    3    3
 [5,]    1    1    3    5
 [6,]    1    1    5    5
 [7,]    1    3    3    3
 [8,]    1    3    3    5
 [9,]    1    3    5    5
[10,]    1    5    5    5
[11,]    3    3    3    3
[12,]    3    3    3    5
[13,]    3    3    5    5
[14,]    3    5    5    5
[15,]    5    5    5    5
```

5.5

We determine the number of distinct bootstrap samples from a finite set.

a.) A bakery sells five types of cookies: sugar, chocolate chip, oatmeal, peanut butter, and ginger snap. Show that the number of ways to order 5 cookies is $\binom{9}{5}$

Unordered samping with replacement: $\binom{n+k-1}{k}$, $n = 5, k = 5$

```
choose(5 + 5 - 1, 5)
```

```
[1] 126
```

b.) Show that the number of sets of size n (order does not matter) drawn with replacement from the (distinct) $a_1, a_2, \ldots, a_n$ is $\binom{2n-1}{n}$

Conclude that the number of distinct bootstrap samples from the set $[a_1, a_2, \ldots, a_n]$ is $\binom{2n-1}{n}$

5.6

Let $k_1, k_2, \ldots, k_n$ denote non-negative integers satisfying $k_1 + k_2 + \ldots + k_n = n$, and suppose the elements in the set $a_1, a_2, \ldots, a_n$ are distinct.

a.) Show that the number of bootstrap samples with k_1 occurrences of a_1, k_2 occurrences of a_2, $\ldots$, k_n occurrences of a_n is $\binom{n}{k_1, k_2, \ldots, k_n}$

b.) Compute the probability that a randomly drawn bootstrap sample will have k_i occurrences of $a_i, i = 1, 2, \ldots, n$

5.7

Refer to Example 5.4 and the remark at the end of the example.

a.) What might account for the fact that there were more missing values for the men who skateboarded in front of the male experimenter? How might this bias the outcome?

It could be that the approached skateboarders refused to participate in the study of performing tricks in front of other men.

b.) Why do you suppose it was important that the two experimenters were blinded to the purpose of the study?

The female could have flurted or otherwise influenced skateboarders who were performing tricks if they knew the intent of the study.

5.8

Consider a population that has a normal distribution with mean $\mu = 36$, standard deviation $\sigma = 8$.

```
mu <- 36; sigma <- 8; n <- 200
```

```
se <- mu / sqrt(n)
```

a.) The sampling distribution of $\bar{X}$ for samples of size 200 will have what mean, standard error, and shape?

$\mu = 36$, SE = 36 / sqrt(200) = 2.5455844, shape will be approximately normal (CLT).

b.) Use R to draw a random sample of size 200 from this population. Conduct EDA on your sample.

```r
set.seed(123)

samp <- data.table(values = rnorm(200, mean = 36, sd = 8))

ggplot(samp, aes(values)) +
    geom_histogram(aes(y = ..density.., fill = ..count..), bins = 30) +
    geom_density(aes(values), color = "darkorange") +
    geom_vline(xintercept = mu, col = "black", height = .8, lty = 2) +
    geom_vline(xintercept = mu - se, col = "darkred", lty = 2) +
    geom_vline(xintercept = mu + se, col = "darkred", lty = 2)
```

```
Warning: Ignoring unknown parameters: height
```
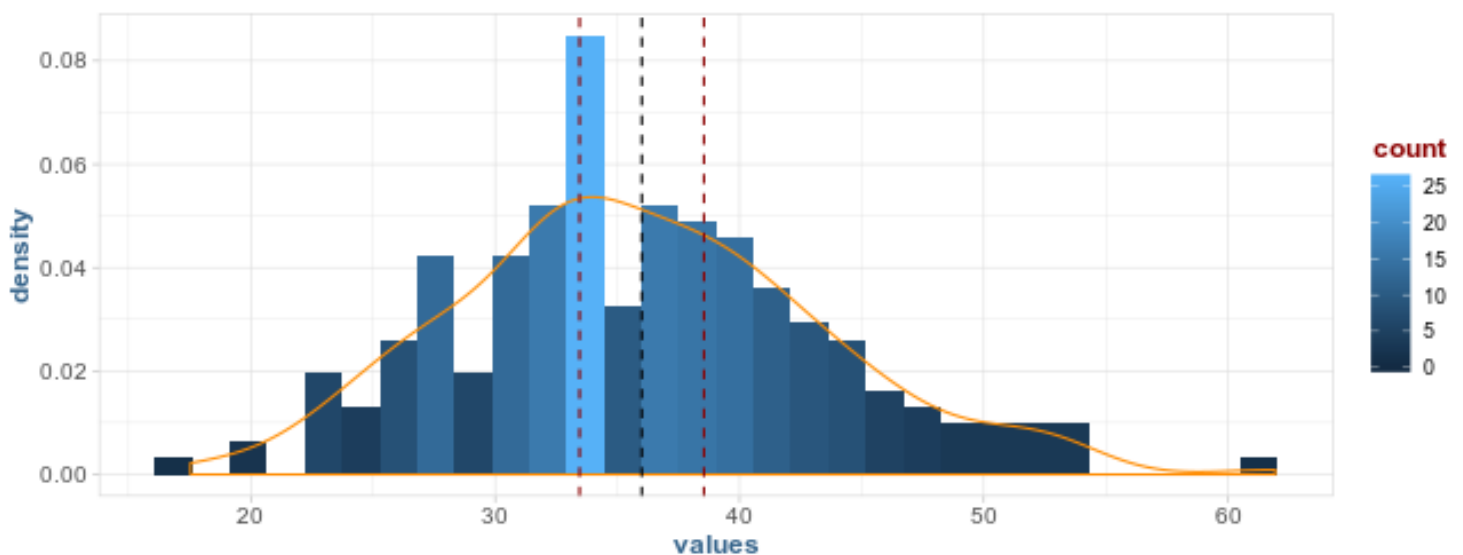

c.) Compute the bootstrap distribution for your sample, and note the bootstrap mean and standard error.

```r
boot.fn <- function(data, index){
    mean(data[index]$values)
}

I <- 10e3

boot(samp, boot.fn, R = I) # boot pkg
```

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = samp, statistic = boot.fn, R = I)
```

```
Bootstrap Statistics :
    original       bias    std. error
t1* 35.93144 0.006716095   0.5392066
```

```r
cst.boot <- function(values, n) {
   bootstrap <- numeric(I)

   for(i in 1:I)
   {
      bootstrap[i] <- mean( sample(values, n, replace = T) )
   }

   alpha <- 0.05

   observed <- mean(values)

   boot.mean <- mean(bootstrap)
   boot.bias <- observed - boot.mean
   boot.se <- sd(bootstrap)

   list(observed = observed,
        mean = boot.mean,
        bias = boot.bias,
        se = boot.se,
        conf = quantile(bootstrap, c(alpha/2, 1 - alpha/2)))
}
```

d.) Compare the bootstrap distribution to the theoretical sampling distribution by creating a table like Table 5.2:

```r
n.200 <- cst.boot(samp$values, 200)

tbl <- data.table(Data = c("Population", "Sampling Distribution", "Sample", "Bootstrap Distribu
                  Mean = c(mu, mu, n.200$observed, n.200$mean),
                  SD = c(sigma, mu/sqrt(n), sd(samp$values), n.200$se))

pretty_kable(tbl, "Sampling Statistics")
```

Table 1: Sampling Statistics

Data	Mean	SD
Population	36.00	8.00
Sampling Distribution	36.00	2.55
Sample	35.93	7.55
Bootstrap Distribution	35.94	0.53

e.) Repeat for sample sizes $n = 50$ and $n = 10$. Carefully describe yyour observations about the effects of sample size on

the bootstrap distribuiton.

```
n.50 <- cst.boot(samp$values, 50)

tbl <- data.table(Data = c("Population", "Sampling Distribution", "Sample", "Bootstrap Distribu
                  Mean = c(mu, sigma, n.50$observed, n.50$mean),
                  SD = c(sigma, mu/sqrt(n), sd(samp$values), n.50$se))

pretty_kable(tbl, "Sampling Statistics")
```

Table 2: Sampling Statistics

Data	Mean	SD
Population	36.00	8.00
Sampling Distribution	8.00	2.55
Sample	35.93	7.55
Bootstrap Distribution	35.92	1.06

```
n.10 <- cst.boot(samp$values, 10)

tbl <- data.table(Data = c("Population", "Sampling Distribution", "Sample", "Bootstrap Distribu
                  Mean = c(mu, sigma, n.10$observed, n.10$mean),
                  SD = c(sigma, mu/sqrt(n), sd(samp$values), n.10$se))

pretty_kable(tbl, "Sampling Statistics")
```

Table 3: Sampling Statistics

Data	Mean	SD
Population	36.00	8.00
Sampling Distribution	8.00	2.55
Sample	35.93	7.55
Bootstrap Distribution	35.93	2.41

The center of the bootstrap distribution doesn't vary much with smaller n, however, confidence intervals (the sd of the bootstrap dist) vary wildy.

5.9

Consider a population that has a gamma distribution with parameters r = 5, $\lambda = 4$.

a.) Use simulation (with $n = 200$) to generate an approximate sampling distribution of the mean; plot and describe the distribution.

b.) Now, draw one random sample of size 200 from this population. Create a histogram of your sample, and find the mean and standard deviation.

c.) Compute the bootstrap distribution of the mean for you sample, plot it, and note the bootstrap mean and standard error.

d.) Compare the bootstrap distribution to the approximate theoretical sampling distribution by creating a table like Table 5.2.

e.) Repeat (a-e) for sample sizes of n = 50, and n = 10. Describe carefully your observations about the effects of sample size on the bootstrap distribution.

5.10

We investigate the bootstrap distribution of the median. Create random sample of size n for various n and bootstrap the median. Describe the bootstrap distribution.