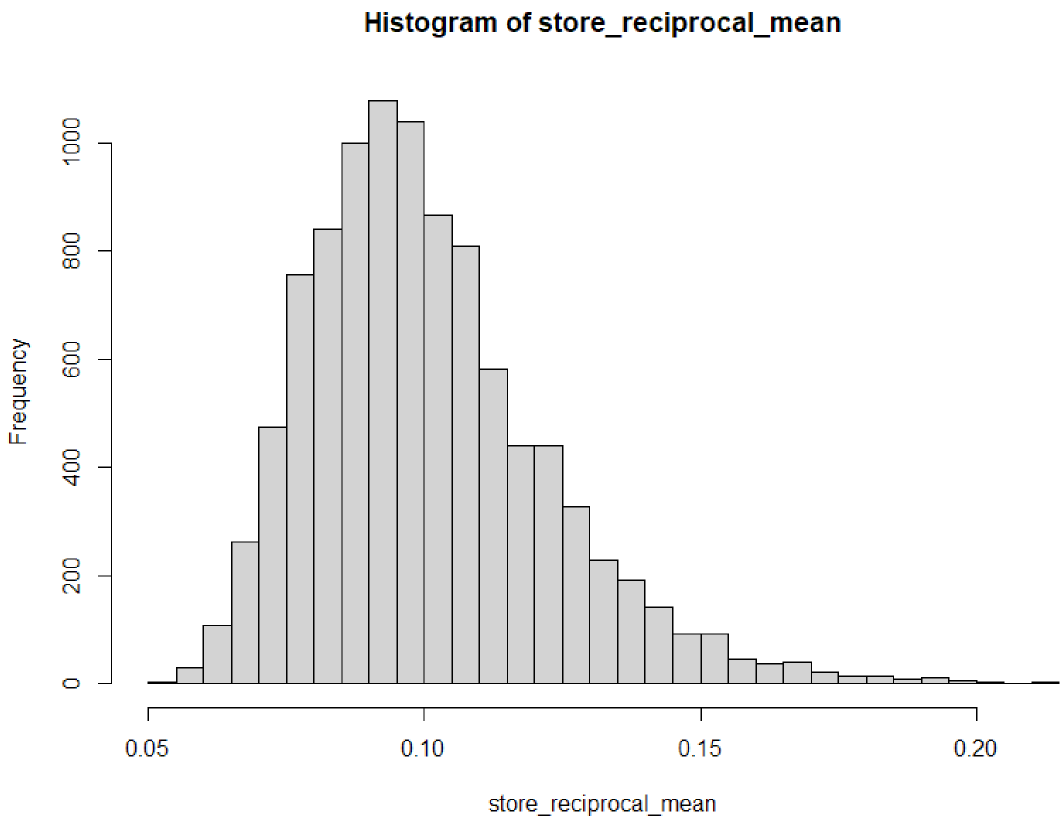
Exercise Bootstrap: Bootstrapping is often needed when we have an unusual statistic and don't know its standard error (Ch. 6) or else when we can't trust the central limit theorem to help us (by allowing us, in Ch. 7, to use plus or minus 2 standard errors to get a confidence interval) and need to find the 95% confidence interval directly. Suppose we have collected the following data from an exponential distribution: (4.1, 15.2, 14.3, 3.1, 25.8, 9.6, 4.1, 14.7, 4.9, 8.7) and we want to estimate the rate ($1/\mu$) with the estimator $1/\overline{x}$. I have know idea what the standard error is of that estimator. Calculate the estimate for this data and then use bootstrapping to get an estimate of the standard error. (Hint: follow the example from class

Solution:
Using r we can create a large Bootstrap sample to estimate ($1/\mu$), calculate the 95% confidence interval and standard error. We end up with a sample estimator of $1/\overline{x} = .1002935$, confidence interval of $(0.06756757, 0.1517451)$ and a standard error of .02161504.

Console:

```
> dat = c(4.1, 15.2, 14.3, 3.1, 25.8, 9.6, 4.1, 14.7, 4.9, 8.7)
> n = length(dat)
> store_reciprocal_mean = rep(NA,10000)

> for (i in 1:10000){
    + resample = sample(dat, size = n, replace = TRUE)
    + store_reciprocal_mean[i] = 1/mean(resample)
    + }

> hist(store_reciprocal_mean, n = 40)
> lower = quantile(store_reciprocal_mean, 0.025)
> upper = quantile(store_reciprocal_mean, 0.975)

> cat("(", lower, ",", upper, ")\n")
( 0.06756757 , 0.1517451 )

> sd(store_reciprocal_mean)
[1] 0.02161504

> mean(store_reciprocal_mean)
[1] 0.1002935
```

Figure 1: Histogram of Bootstrap sample



Exercise 7.4: A CI is desired for the true average stray-load loss μ (watts) for a certain type of induction motor when the line current is held at 10 amps for a speed of 1500 rpm. Assume that stray-load loss is normally distributed with $\sigma = 3.0$

1. Compute a 95% CI for μ when $n = 25$ and $\overline{x} = 58.3$.

   Solution:
   Computing the 95% CI from the definition,

$$95\%CI = (\overline{x} - z_{.025}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{.025}\frac{\sigma}{\sqrt{n}})$$

$$95\%CI = (58.3 - 1.96\frac{3.0}{\sqrt{25}}, 58.3 + 1.96\frac{3.0}{\sqrt{25}}),$$
$$95\%CI = (57.1, 59.5).$$

2. Compute a 95% CI for μ when $n = 100$ and $\overline{x} = 58.3$.

Solution:
Computing the 95% CI from the definition,

$$95\%CI = (58.3 - 1.96\frac{3.0}{\sqrt{100}}, 58.3 + 1.96\frac{3.0}{\sqrt{100}}),$$
$$95\%CI = (57.7, 58.9).$$

3. Compute a 99% CI for μ when $n = 100$ and $\overline{x} = 58.3$.

Solution:
Computing the 99% CI from the definition and $z_{.005} = 2.58$,

$$99\%CI = (58.3 - 2.58\frac{3.0}{\sqrt{100}}, 58.3 + 2.58\frac{3.0}{\sqrt{100}}),$$
$$99\%CI = (57.5, 59.1).$$

4. Compute a 82% CI for μ when $n = 100$ and $\overline{x} = 58.3$.

Solution:
Computing the 82% CI from the definition and $z_{.09} = 1.34$,

$$82\%CI = (58.3 - 1.34\frac{3.0}{\sqrt{100}}, 58.3 + 1.34\frac{3.0}{\sqrt{100}}),$$
$$82\%CI = (57.9, 58.7).$$

5. How large must n be if the width of the 99% CI for μ is to be 1.0?

Solution:
Solving for n by the definition of a 99% CI with width $w = 1$,

$$w = 2z_{.005}\frac{\sigma}{\sqrt{n}},$$
$$n = (2z_{.005}\frac{\sigma}{w})^2.$$

Substituting and solving for n,

$$n = (2(2.58)\frac{3.0}{1})^2 = 239.62.$$

Exercise 7.10: A random sample of $n = 15$ heat pumps of a certain type yielded the following observations on lifetime (in years) ... Assume tha the lifetime distribution is exponential and use an argument parallel to that of example 7.5 to obtain a 95% CI for the expected average lifetime.

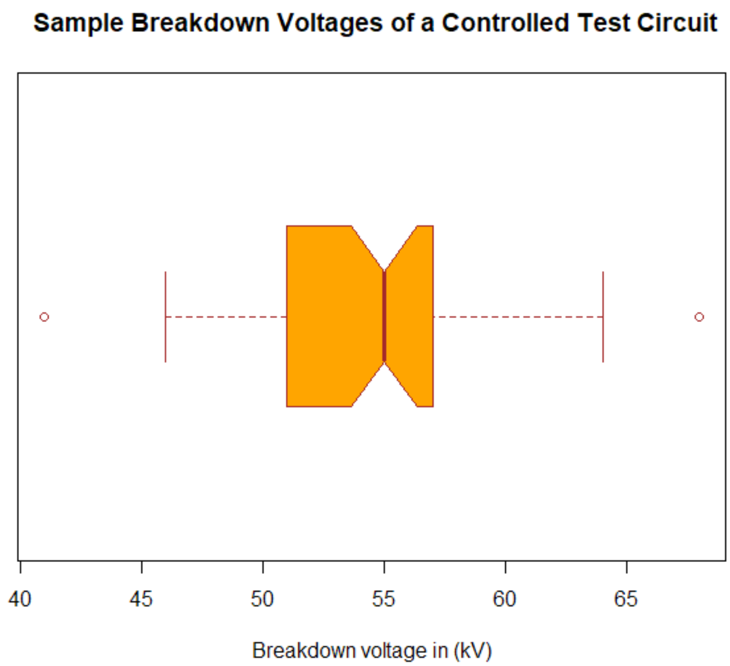
Solution:
Consider the random variable $h(X_1, X_2, \ldots, X_{15}) = 2\lambda \sum X_i$. Following the example we know that this variable is a chi-square distribution with $2n = 2(15) = 30$ degrees of freedom. From Table A.7 we know the critical values $\chi^2_{.975,30} = 16.791$ and $\chi^2_{.025,30} = 46.976$. From the example we get that the confidence interval for the parameter $\mu = 1/\lambda$ and where $\sum x_i = 63.2$ from our given data,

$$95\% = (\frac{2\sum x_i}{16.791}, \frac{2\sum x_i}{46.976}) = (2.69, 7.53).$$

Exercise 7.16: The alternating current (AC) breakdown voltage of an insulating liquid indicates its dielectric strength...

1. Construct a boxplot of data and comment on any interesting features.

   Solution:
   The following is a boxplot generated in r



Looking at the box plot we can see that the data is fairly, normally distributed. There is a little bit of a left skew but we also have outliers on either side of the median. We can also see this looking at the five number summary,

$$\begin{bmatrix} Min. & 1stQu. & Median & 3rdQu. & Max. \\ 41 & 51 & 55 & 57 & 68 \end{bmatrix}$$

2. Calculate and interpret a 95% CI for true average breakdown voltage μ. Does it appear that μ has been precisely estimated? Explain.

Solution:
Computing the sample mean and the variance/standard deviation of the sample mean in order to calculate the large-sample confidence interval. **Console:**

```
> x
 [1] 62 50 53 57 41 53 55 61 59 64
[11] 50 53 64 62 50 68 54 55 57 50
[21] 55 50 56 55 46 55 53 54 52 47
[31] 47 55 57 48 63 57 57 55 53 59
[41] 53 52 50 55 60 50 56 58

> mean(x)
```

```
[1] 54.70833

> var(x)
[1] 27.35993

> sd(x)
[1] 5.230672
```

Since we are calculating the 95% CI we use $x_{0.025} = 1.96$ we get the following,

$$95\% = (54.7083 - 1.96\frac{5.2306}{\sqrt{48}}, 54.7083 + 1.96\frac{5.2306}{\sqrt{48}})$$

$$95\% = (53.2285, 56.1881)$$

With a margin of error around 1.5m, $\overline{x}$ seems to be a relatively precise estimate for μ.

3. Suppose the investigator believes that virtually all values of breakdown voltage are between 40 and 70. What sample size would be appropriate for the 95% CI to have a width of 2 kV (so that μ is estimated to within 1 kV with 95% confidence)?

Solution:
Solving for n with a new width for the confidence interval,

$$2 = 2(1.96)\frac{\sigma}{\sqrt{n}}.$$

Given what the investigator believes we can estimate the standard deviation σ by taking the difference between the largest and smallest values and dividing it by 4,

$$\sigma \approx \frac{70 - 40}{4} = 7.5.$$

Using that we can solve for n,

$$n = (2(1.96)\frac{7.5}{2})^2 = 216.09.$$

A sample size above 220 would likely be the safest bet.

Exercise 7.20: TV advertising agencies face increasing challenges in reaching audience members because viewing TV programs via digital streaming is gaining in popularity. . .

1. Calculate and interpret a confidence interval at the 99% confidence level for the proportion of all adult Americans who watched streamed programming up to that point in time.

   Solution:
   Given our point estimation for p, $\hat{p} = .53$ and our sample size $n = 2343$. Using Proposition 7.11 we can calculate the 99% confidence level with critical point $z_{.005} = 2.575$ for the population proportion,

$$99\%CI = (.53 - 2.575\sqrt{\frac{.53(1-.53)}{2343}}), (.53 + 2.575\sqrt{\frac{.53(1-.53)}{2343}}),$$
$$99\%CI = (.5034, .5566).$$

2. What sample size would be required for the width of a 99% CI to be at most .05 irrespective of the value of $\hat{p}$?

   Solution:
   When $\hat{p}$ is unknown we approximate with $\hat{p} = .50$. solving for n using Proposition 7.11 with a given width of .5,

$$.05 = 2(2.575)\sqrt{\frac{.50(1-.50)}{n}},$$
$$.05 = 2(2.575)\sqrt{\frac{.25}{n}},$$
$$n = 2(2.575)^2\frac{.25}{.05^2},$$
$$n = 1326.125.$$

   Therefore we would need a sample size of approximately $n = 1327$ to get a width of .05 in an 99% CI.

Exercise 7.32: According to the article Fatigue... Calculate and interpret a confidence interval at the 99% confidence level for the true average number of cycles to break.

Solution:
Given that $n = 20$, we have a sample mean of $\overline{x} = 1584$ with a standard deviation of $s = 607$ all we have to do is find the critical value at $t_{.005}$ for a t distribution with 19 degrees

of freedom. Through table $A.5$ we get that $t_{.005,19} = 2.861$ and by Proposition 7.15 we get that the confidence interval is,

$$99\%CI = (1584 - 2.861\frac{607}{\sqrt{20}}), 1584 + 2.861\frac{607}{\sqrt{20}}),$$

$$99\%CI = (1195.68, 1972.32).$$

Exercise 7.36: A normal probability plot of the $n = 26$ observations on escape time given in Exercise 36 of Chapter 1 shows a substantial linear pattern; the sample mean and sample standard deviation are 370.69 and 24.36, respectively.

1. Calculate an upper confidence bound for population mean escape time using a confidence level of 95%

   Solution:
   To calculate the upper bound for the 95% confidence interval we must determine the critical value at $t_{.025}$ for a t distribution with 25 degrees of freedom. Through table $A.5$ we get that $t_{.025,25} = 1.708$ and by Proposition 7.15 we get that the upper bound is,

   $$370.69 + 2.861\frac{24.36}{\sqrt{26}} = 378.84.$$

2. Calculate an upper prediction bound for the escape time of a single additional worker using a prediction level of 95%. How does this bound compare with the confidence bound of part (a)?

   Solution:
   Using the same critical value at $t_{.025}$ we can calculate the upper bound of the prediction interval by Proposition 7.16

   $$370.69 + 2.861(24.36)\sqrt{1 + \frac{1}{26}} = 42.39.$$