Joshua Main-Smith
R 3
joshuamainsmith@ufl.edu
UFID: 7160-7655
March 15, 2020

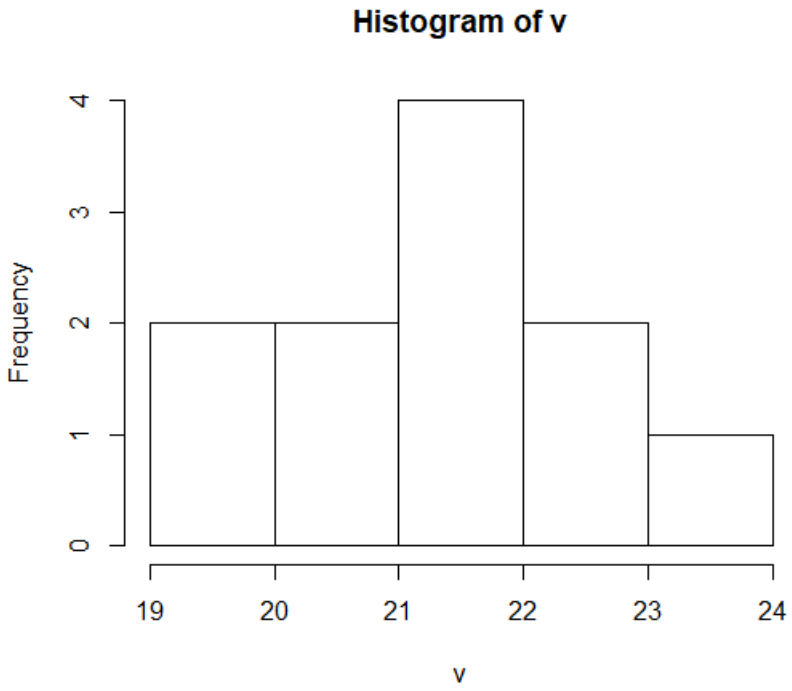
R Assignment 3

The stem length of soybeans from an experiment are:

20.2, 22.9, 23.3, 20.0, 19.4, 22.0, 22.1, 22.0, 21.9, 21.5, 20.9

1. **Create a histogram to visualize the data**

   ```
   v <- c(20.2, 22.9, 23.3, 20.0, 19.4, 22.0, 22.1, 22.0, 21.9, 21.5, 20.9)
   hist(v)
   ```



2. Test "t.test" whether the population mean is different from 23

```
> v <- c(20.2, 22.9, 23.3, 20.0, 19.4, 22.0, 22.1, 22.0, 21.9, 21.5, 20.9)
>
>
> t.test(v, mu=23)

    One Sample t-test

data:  v
t = -4.1467, df = 10, p-value = 0.00199
alternative hypothesis: true mean is not equal to 23
95 percent confidence interval:
 20.65208 22.29338
sample estimates:
mean of x
 21.47273
```

true mean is not equal to 23

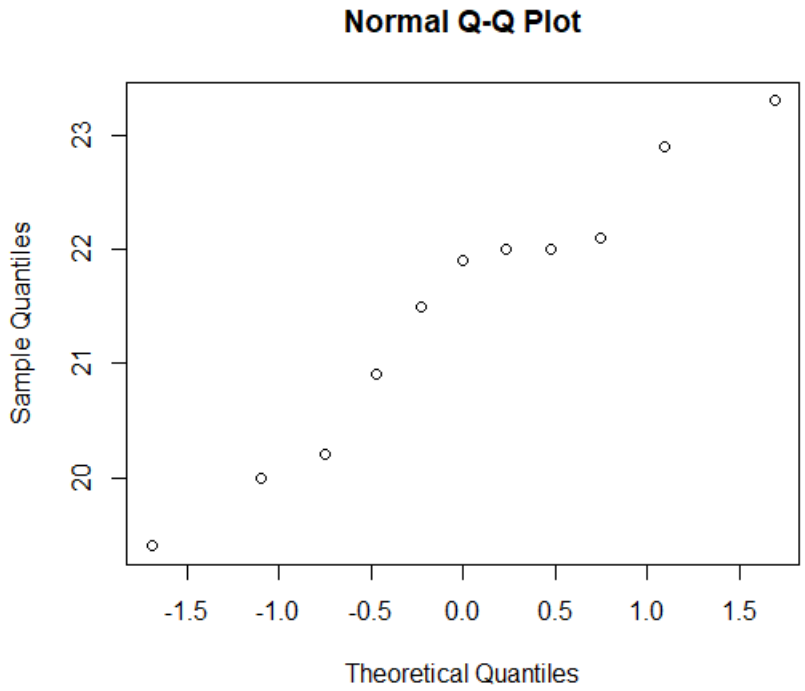
3. Obtain a 2 sided 98% confidence interval on the true mean using "t.test".

```
> v <- c(20.2, 22.9, 23.3, 20.0, 19.4, 22.0, 22.1, 22.0, 21.9, 21.5, 20.9)
>
>
> t.test(v, mu=0, alternative="two.sided",conf.level=0.98)

    One Sample t-test

data:  v
t = 58.301, df = 10, p-value = 5.352e-14
alternative hypothesis: true mean is not equal to 0
98 percent confidence interval:
 20.45480 22.49065
sample estimates:
mean of x
 21.47273
```

So, (20.4548, 22.4906)

4. **The researcher, by using "t.test" on a sample size of 11 was assuming that the data was normally distributed. Is that a valid claim? Create a QQ plot and interpret.**

It wouldn't be safe to assume a normal distribution. According to the Central Limit Theorem, a distribution is only normal with large sample sizes (generally > 30 is a safe case).



Looking at the QQ plot above, we can see that it's not normally distributed. Under the circumstance that it would be normally distributed the slope would be equal to one. In this case, it isn't.