hw08

Victor Huang

3/4/2022

Q1. Exercise 8.2

We are able to get the folloing hypothesis: the null states that the mean choleserol level for vegeterians is equal to 161 while the alternative states that it isn't equal to 161. Given the information from the questionm we get the following t-score and p-value. Since our p-value is less than 0.05, we can reject the null hypothesis and conclude that there exists a statistically discernable difference between vegeterian average cholesterol levels.

```r
t <- (164 - 161)/(5/sqrt(24))
p <- 2* pt(q = t, df = 23, lower.tail=FALSE)
```

Q2. Exercise 8.5. Part (b) asks you to conduct a bootstrap t test. This is very similar to the bootstrap t interval, but instead of finding quantile, you compare your observed T statistic to the bootstrap distribution and see what proportion of the bootstrap statistics exceed that value. See Section 8.2 for an overview and code to adapt.
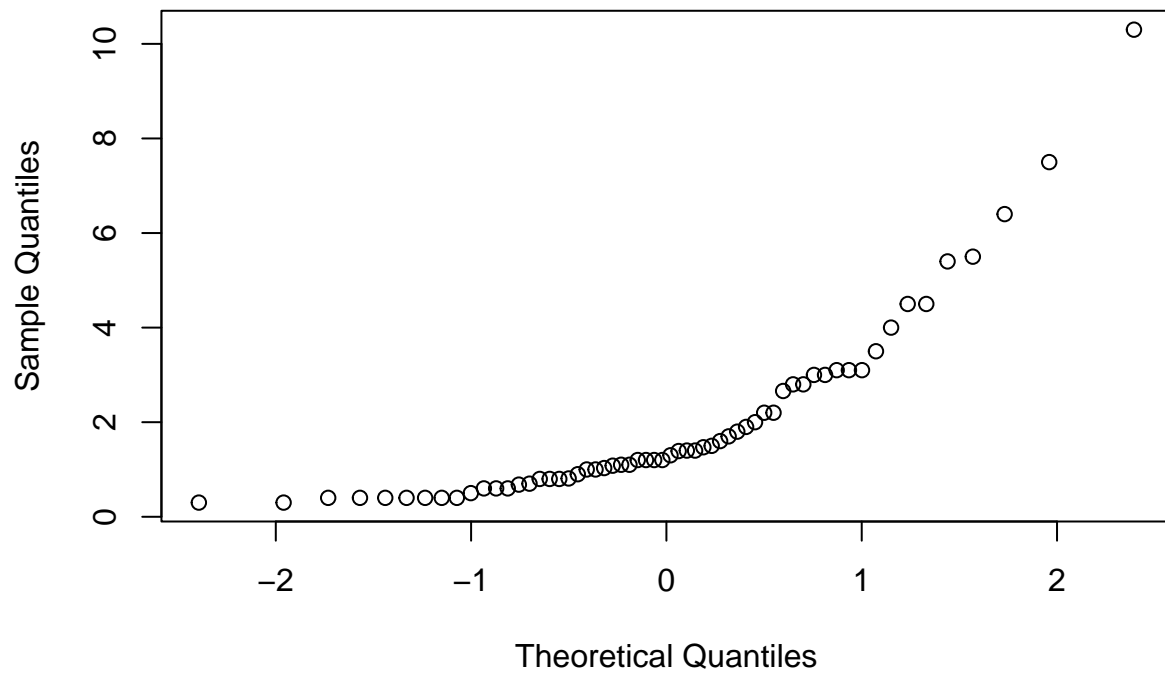
a).

```r
Walleye <- read.csv("https://sites.google.com/site/chiharahesterberg/data2/Walleye.csv")
hist(Walleye$Weight)
```

Histogram of Walleye$Weight
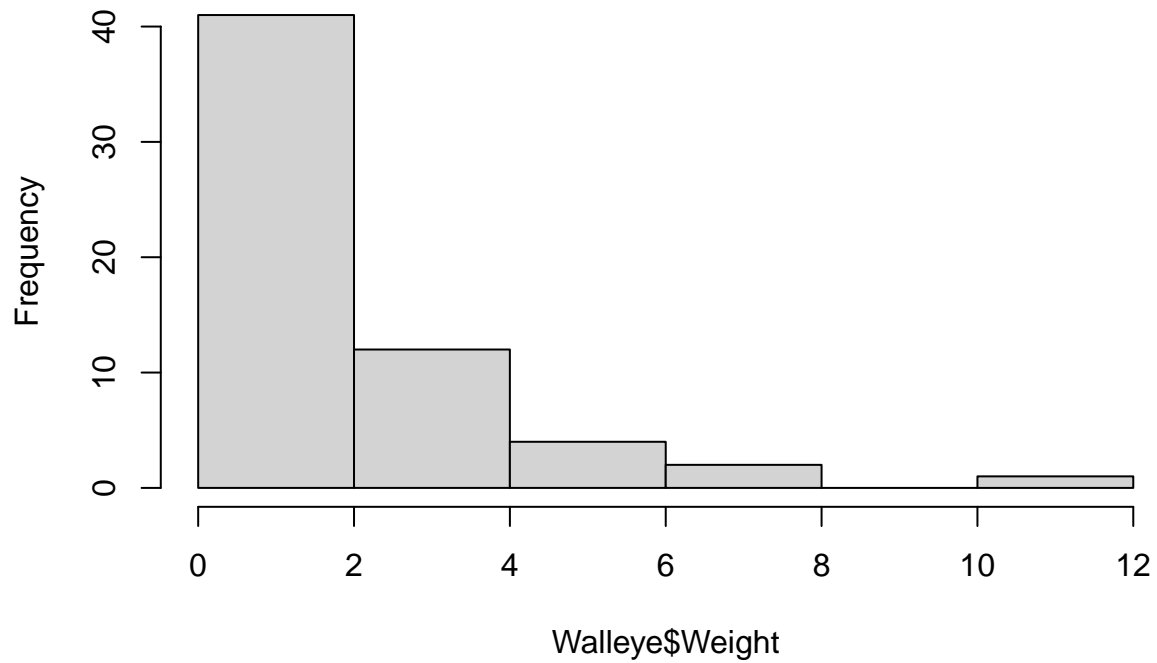
qqnorm(Walleye$Weight)
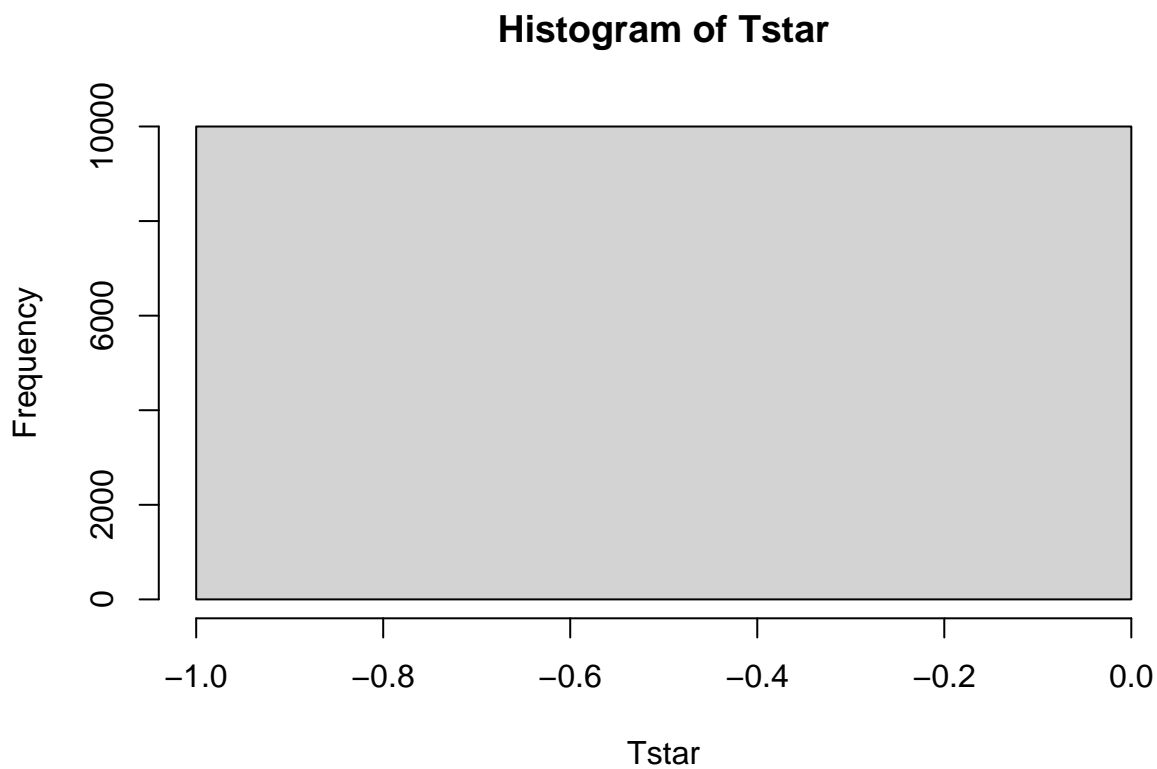
Normal Q–Q Plot

b).

```
Weight <- Walleye$Weight
observedT <- t.test(Walleye$Weight, mu = 2.5, alt = "less")$statistic
xbar <- length(Weight)
N <- 10^4
Tstar <- numeric(N)
for(i in 1:N){
  bootx <- sample(Weight, n, replace = TRUE)
  Tstar[i] <- (mean(bootx) - xbar)/(sd(bootx)/sqrt(n))
}
```

```
## Error in sample.int(length(x), size, replace, prob): invalid 'size' argument
```

```
hist(Tstar)
```

Histogram of Tstar


```
p2 <- (sum(Tstar <= observedT) + 1)/(N+1)
p2
```

```
## [1] 9.999e-05
```

Given that our p-value is significantly less than 5% , we reject the null hypothesis and conclude that there is a difference in average weight for current walleye compared to the average weight of walleye from the 1990.

c).

```
t.test(Walleye$Weight, mu = 2.5, alt = "less")
```

```
##
```

```
##  One Sample t-test
##
## data:  Walleye$Weight
## t = -2.0462, df = 59, p-value = 0.0226
## alternative hypothesis: true mean is less than 2.5
## 95 percent confidence interval:
##      -Inf 2.406267
## sample estimates:
## mean of x
##  1.988667
```

Since our p-value for the t.test is less than 5%, we are able to reject the null hypothesis and conclude that there is indeed a difference/decrease in average wieght of the walleye of now compared to the average weight of walleye in 1990

Q3. Exercise 8.8

a).

```
Spruce <- read.csv("https://sites.google.com/site/chiharahesterberg/data2/Spruce.csv")
t.test(Di.change ~ Competition, data = Spruce)
```

```
##
##  Welch Two Sample t-test
##
## data:  Di.change by Competition
## t = -4.1477, df = 67.717, p-value = 9.573e-05
## alternative hypothesis: true difference in means between group C and group NC is not equal to 0
## 95 percent confidence interval:
##  -2.3357470 -0.8182461
## sample estimates:
##   mean in group C mean in group NC
##          3.207292         4.784288
```

Since we get a p-value that is less than 5%, we can reject the null hypothesis and conclude that there is indeed a difference in diameter change when compaing the two groups of trees (competing and non-competing)

b). This means that the data is randomly sampled, as such, we can confidantly accept the conclusion we arrived at in part a). as all confounding variables have been negated.

Q4. Exercise 8.15

```
prop.test(x = 69, n = 310, p = 0.26, correct = FALSE, alternative = "less")
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  69 out of 310
## X-squared = 2.2561, df = 1, p-value = 0.06655
## alternative hypothesis: true p is less than 0.26
## 95 percent confidence interval:
##  0.0000000 0.2637481
## sample estimates:
##         p
## 0.2225806
```

We get a p-value greater than 5%, as such we fail to reject the null hypothesis and do not have sufficient evidence to support the professor's hypothesis.

Q5. Exercise 8.22

Type I Error: The drug does indeed lower chlosterol levels but is found and concluded that it does not (false positive).

Type II Error: The drug does not lower chlosterol levels but is found and concluded that it does (false negative).

Q6. Exercise 8.24

```
zval <- 1.645
sm <- 25 + zval*(4/sqrt(30))
z <- (sm - 27)/(4/sqrt(30))
power <- 1 - pnorm(z)
power
```

[1] 0.8629376

We get the power of the test to be 0.8629376

Q7. Exercise 8.28 Given $y = \sum_{i=1}^{12} x_i \ Bin(12, p)$, we can get using PMF $f_y(y) = P(Y = y) = \binom{12}{y} p^y (1 - p)^{12-y}; y = 0, 1, ..., 12$. As such, we can get type ! error probablity to be:

$$P(y \in 0, 1|p = 0.3) = P(y = 0|p = 0.3) + P(y = 1|p = 0.3)$$
$$(1 - 0.3)^1 2 + 12(0.3)(1 - 0.3)^{11}$$
$$(0.7)^{12} + 12(0.3)(0.7)^{11}$$
$$0.085$$

As such, we see that we have proabblity 0.085 of getting a type I error.

b).

$$1 - \beta = P(reject)$$
$$p(y \in 0, 1|p) = p(y = 0|p) + P(y = 1|p)$$
$$(1 - p)^{12} + 12p + (1 - p)^n$$
$$(1 - p)^{11} + (1 + 11p)$$

As such, we get that $1 - \beta = (1 - p)^{11} + (1 + 11p)$

c). (reference picture)

Q8. Exercise 8.30 (reference picture)

Q9. Exercise 8.25

```
N <-  ((-2.3263 - 1.2816)^2 * (0.3)^2)/(1.2 - 1)^2
N
```

[1] 29.28812

We get N = 29.29, due to ethics issues we will round it up to 30. As such, is she wanted to have a 90% chance of detecting a mean of 1.2 or higher, she would need a sample larger than 30.

Q10. Exercise 8.38

```
alphastar <- 1 - (1 - 0.1)^{1/12}
alphastar
```

[1] 0.008741611

The significance level for each test should be 0.0087 accoridng to the Sidak correction.