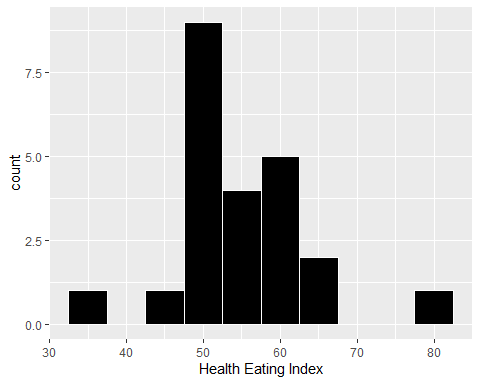
STAR511: HW 3

Megan Sears

# Q1

The population is women receiving SNAP benefits who are living in Denver, CO, and inferential statements can be made on that.

# Q2



# Q3

| Mean | Standard\_Deviation |
| --- | --- |
| 54.78 | 8.61 |

# Q4

The 95% confidence interval for the population mean is (51.06, 58.51).

##   
## One Sample t-test  
##   
## data: diet$HEI  
## t = 30.505, df = 22, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 51.05822 58.50700  
## sample estimates:  
## mean of x   
## 54.78261

# Q5

We are 95% confident that the true population mean health eating index is between 51.06 and 58.51.

# Q6

The data are normally distributed (relatively) even with a sample size smaller than 30. Therefore, the confidence intervals are valid.

# Q7

We reject the null hypothesis since the null hypothesized mean falls outside of the confidence interval. The 95% confidence interval is 51.06 to 58.51 and the null hypothesis is 60, which is not within.

# Q8

Reject the null hypothesis since the p-value is less than alpha.

## [1] 0.008206317

# Q9

As sample size increases, the CI width will decrease. This is because we are able to estimate a population proportion more accurately with an increased sample size.

# Q10

As standard deviation increases, the CI width will increase. Since standard deviation measures the deviation in the data, more variation will increase the CI width.

# Q11

The CI width will decrease as the confidence level decreases since CI width decreases as confidence level does.

# Q12

A. Reject the null hypothesis if |t| is greater than 2.26.

B. The test statistic is -2.07.

C. The conclusion is fail to reject the null hypothesis.

# Q13

A. Reject the null hypothesis if t is less than -1.83.

B. The test statistic is -2.07.

C. The conclusion is to reject the null hypothesis.

# Q14

A. Reject the null hypothesis if |t| is greater than 2.02.

B. The test statistic is -4.14.

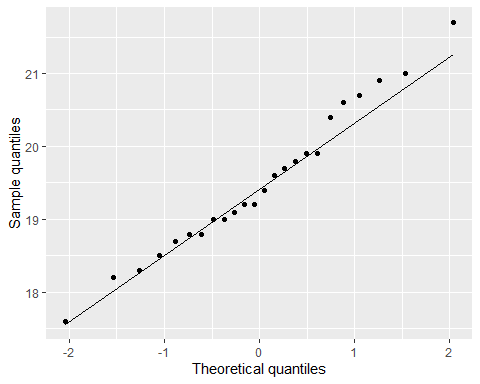
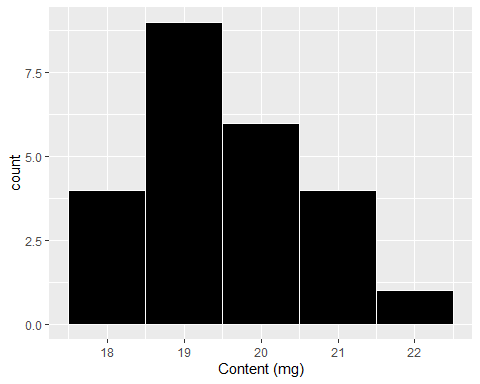
C. The conclusion is to reject the null hypothesis.

# Q15

As sample size increases, there is a stronger likelihood to reject the null hypothesis.

# Q16

Based on the histogram and qqplot, the data appear to be approximately normal. The histgram roughly has greater values towards the middle with tails on both sides. For the qqplot, since the points are near or on the line, it supports that data are approximately normal.



# Q17

The sample estimates mean is 19.5 and the 95% confidence interval is (19.08, 19.92).

##   
## One Sample t-test  
##   
## data: pills$content  
## t = 95.158, df = 23, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 19.07609 19.92391  
## sample estimates:  
## mean of x   
## 19.5

# Q18

##   
## One Sample t-test  
##   
## data: pills$content  
## t = -2.44, df = 23, p-value = 0.02281  
## alternative hypothesis: true mean is not equal to 20  
## 95 percent confidence interval:  
## 19.07609 19.92391  
## sample estimates:  
## mean of x   
## 19.5

A. The null hypothesis is the true mean equals 20 and the alternative hypothesis does not equal 20.

B. The test statistic is -2.44 and the p-value is 0.02.

C. Reject the null hypothesis (and accept the alternative hypothesis) since the p-value is less than alpha. Therefore, this batch of pill should be destroyed.

# Q19

##   
## One Sample t-test  
##   
## data: pills$content  
## t = -2.44, df = 23, p-value = 0.01141  
## alternative hypothesis: true mean is less than 20  
## 95 percent confidence interval:  
## -Inf 19.85121  
## sample estimates:  
## mean of x   
## 19.5

A. The null hypothesis is greater than or equal to 20, and the alternative hypothesis is less than 20.

B.The test satistic is -2.44 and the p-value is 0.01.

C. Reject the null hypothesis since (and accept the alternative hypothesis) since the p-value is less than alpha. Therefore, the batch of pills does not need to be destroyed.

# Q20

The estimated standard deviation is 0.09.

# Q21

The minimum sample size to achieve a 95% ME < 0.03 is 38.

## n ME  
## 1 30 0.03360655  
## 2 31 0.03301227  
## 3 32 0.03244846  
## 4 33 0.03191261  
## 5 34 0.03140248  
## 6 35 0.03091608  
## 7 36 0.03045162  
## 8 37 0.03000749  
## 9 38 0.02958226  
## 10 39 0.02917463  
## 11 40 0.02878340

# Q22

The power is 57% when there are 85 observations.

##   
## One-sample t test power calculation   
##   
## n = 85  
## delta = 0.3  
## sd = 1.5  
## sig.level = 0.05  
## power = 0.5730619  
## alternative = one.sided

# Q23

In question 23-26, the NCP equation is used to described how power changes since power is larger when NCP is.

As sample size increases, so does power since based on the NCP equation, a large sample size will increase the power. Thus, power would be higher than the power in question 22.

# Q24

Power would decrease from question 22 since the power decreases as the significance level does. We would require stronger evidence to reject the null hypothesis if alpha were smaller.

# Q25

The power would be higher than question 22. The power will increase as the delta value does based on NCP.

# Q26

As standard deviation increases, power will decrease based on NCP. Therefore, the power will decrease from question 22.

# Q27

A total of 156 samples are required to achieve a power of 80%.

##   
## One-sample t test power calculation   
##   
## n = 155.9257  
## delta = 0.3  
## sd = 1.5  
## sig.level = 0.05  
## power = 0.8  
## alternative = one.sided

# Appendix

#Retain (and do not edit) this code chunk!!!  
library(knitr)  
knitr::opts\_chunk$set(echo = FALSE)  
knitr::opts\_chunk$set(message = FALSE)  
  
library(tidyverse)  
library(here)  
  
#Q1  
diet <- read\_csv(here('./HW3/Diet.csv'))  
#str(diet)  
#diet  
  
#Q2  
ggplot(diet, aes(x=HEI)) +  
 geom\_histogram(color='white', fill='black', binwidth=5) +  
 xlab('Health Eating Index')  
  
#Q3  
diet\_summary <- diet %>%  
 summarize(Mean = round(mean(HEI), 2),  
 Standard\_Deviation = round(sd(HEI),2))  
  
kable(diet\_summary)  
  
#Q4  
t.test(diet$HEI)  
#Q8  
t.test(diet$HEI, mu=60)$p.value  
  
#Q12  
n <- 10  
sampmean <- 4.62  
sd <- 0.58  
h0 <- 5  
  
rr <- qt(0.975, df=n-1)  
  
ts <- (sampmean - h0) / (sd/sqrt(n))  
  
#Q13  
  
rr <- qt(0.05, df=n-1)  
  
#ts is the same as Q12  
#ts <- (sampmean - h0) / (sd/sqrt(n))  
  
  
#Q14  
n <- 40  
  
rr <- qt(0.975, df=n-1)  
  
ts <- (sampmean - h0) / (sd/sqrt(n))  
  
#Q16  
pills <- read\_csv(here('./HW3/Pills.csv'))  
#str(pills)  
  
ggplot(pills, aes(x=content)) +  
 geom\_histogram(color='white', fill='black', binwidth=1) +  
 xlab('Content (mg)')  
  
ggplot(pills, aes(sample = content)) +   
 stat\_qq() + stat\_qq\_line() +  
 xlab('Theoretical quantiles') +  
 ylab('Sample quantiles')  
  
#Q17  
t.test(pills$content)  
#Q18  
t.test(pills$content, mu=20)  
  
ts <- round(t.test(pills$content, mu=20)$statistic, 2)  
p <-round(t.test(pills$content, mu=20)$p.value, 2)  
  
#Q19  
t.test(pills$content, mu=20, alternative = 'less')  
  
ts <- round(t.test(pills$content, mu=20, alternative = 'less')$statistic, 2)  
p <-round(t.test(pills$content, mu=20, alternative = 'less')$p.value, 2)  
  
  
#Q20  
sd\_estimate <- (0.9 - 0.36) / 6  
  
#Q21  
#power.t.test(sd = sd\_estimate, power = 0.95, type = 'one.sample',  
# alternative = 'one.sided')  
  
  
n <- ((2\*sd\_estimate) / 0.03)^2  
  
n <- seq(from=30, to=40, by = 1)  
alpha <- 0.05  
ME <- qt(1-(alpha/2), df=n-1)\*sd\_estimate/sqrt(n)  
data.frame(n, ME)  
  
  
#Q22  
alpha <- 0.05  
n <- 85  
sd <- 1.5  
delta <- 0.3  
  
power.t.test(n, delta, sd, alpha, type = 'one.sample', alternative = 'one.sided')  
  
#Q27  
power.t.test(delta = delta, sd = sd, power = 0.8, type = 'one.sample',  
 alternative = 'one.sided')