Yiduo Feng

CS 555

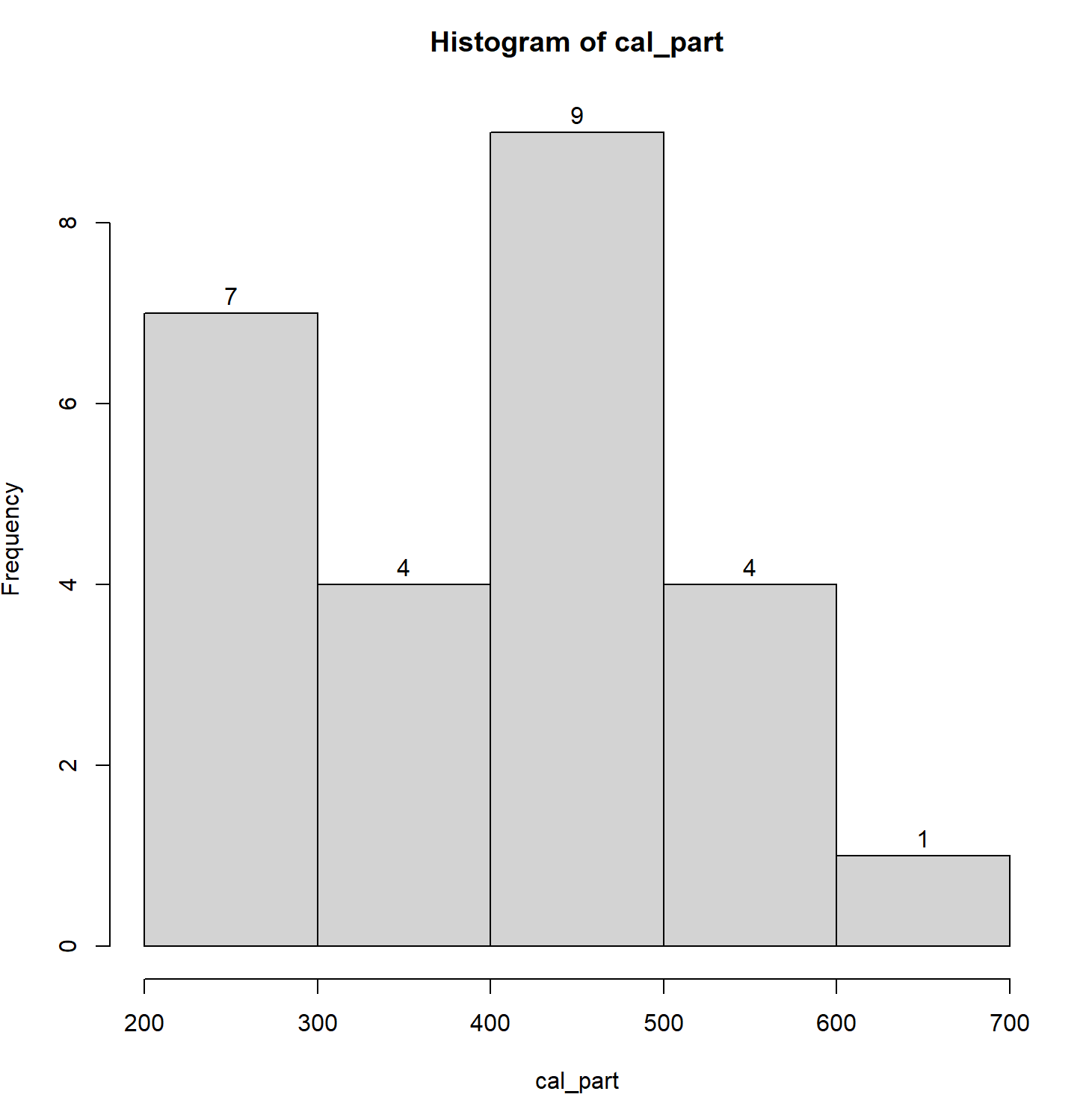
Homework 2

07/16/2022

**(1) Summarize the data by whether children participated in the meal preparation or not. Use an appropriately labelled table to show the results. Also include a graphical presentation that shows the distribution of calories for participants vs. non-participants. Describe the shape of each distribution and comment on the similarity (or lack thereof) between the distributions in each group.**

- Children participated in the meal preparation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Q | Median | Mean | 3rd Q | Max | standard deviation |
| 211 | 298.4 | 424.9 | 410.1 | 456.3 | 635.2 | 121.5138 |



The histogram shows that it is almost normal distribution but a little left skewed, and the center is median 424.9.

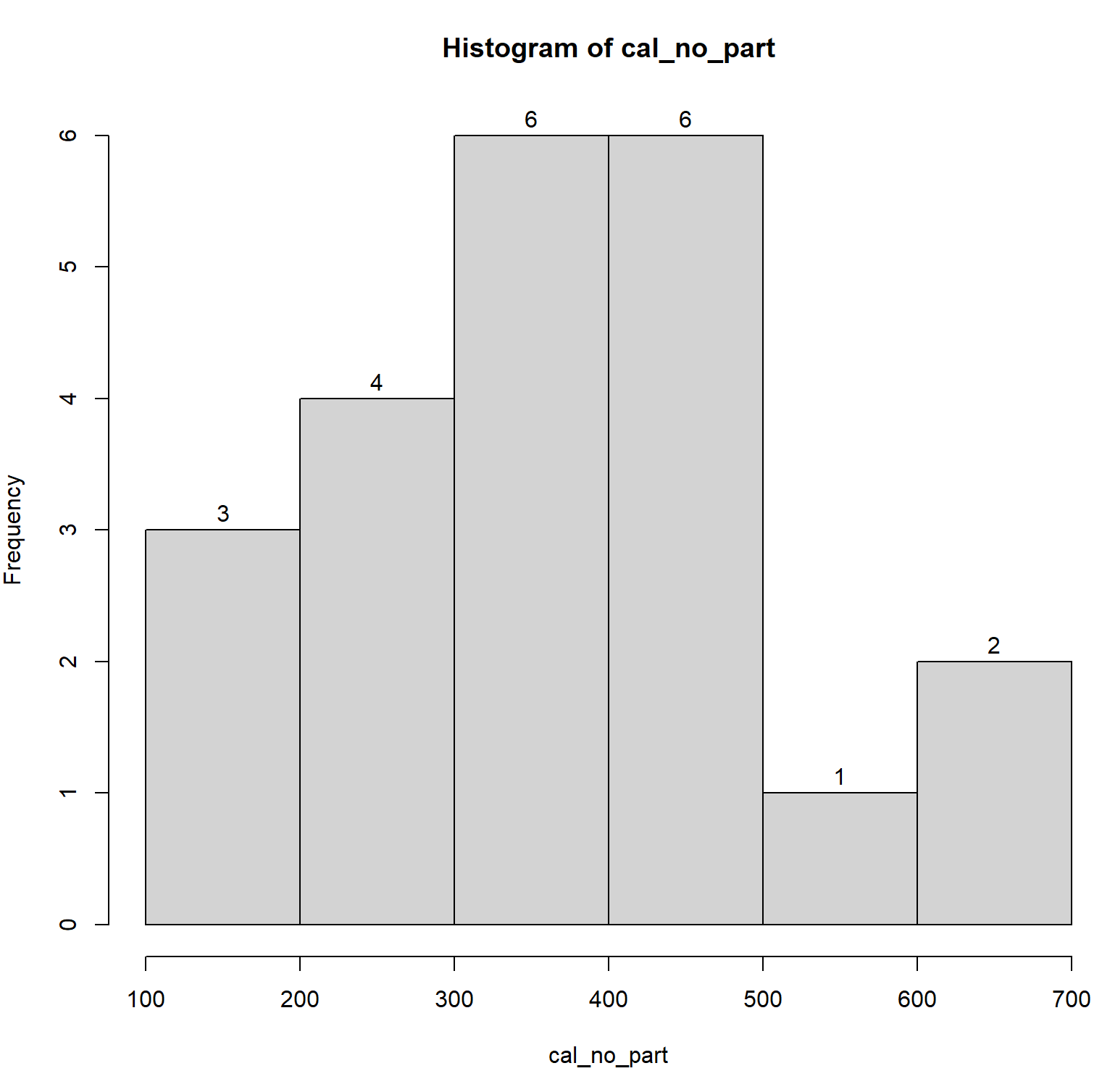
IQR is the best measure of spread which is Q3 -Q1 = 456.3-211 = 245.3.

Q1-1.5\*IQR = -156.95, Q3+1.5\*IQR = 824.25

There is no outlier.

- Children didn’t participate in the meal preparation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Q | Median | Mean | 3rd Q | Max | standard deviation |
| 139.7 | 296.4 | 374.7 | 374.1 | 445.6 | 688.8 | 133.1393 |



The histogram shows that it is almost normal distribution, and the center is median 374.7.

Standard deviation is is the best measure of spread which is 133.1393.

IQR = Q3 -Q1 = 445.6-296.4 = 149.2.

Q1-1.5\*IQR = 72.6, Q3+1.5\*IQR = 669.4

There is no outlier.

The shape of both of them are almost normal, and the numbers of Q1, Q3 and standard deviation are similar. Both of them don’t have outlier. However, the min of children participated in the meal preparation is 211 which is much higher than that of children didn’t participate in the meal preparation, but the max of participating is lower. Also, Median and mean of participating are higher.

**(2) Does the mean calorie consumption for those who participated in the meal preparation differ from 425? Formally test at the alpha = 0.05 level using the 5 steps outlined in the module.**

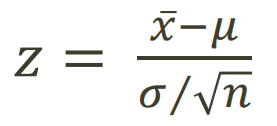
1. Set up the hypotheses and select the alpha levelH0: mean = 425 (mean calorie consumption for the participated is 425)

H1: mean != 425 (mean differs 425)

Alpha = 0.05

2. Select the appropriate test-statistic

Population sd = 119.0587

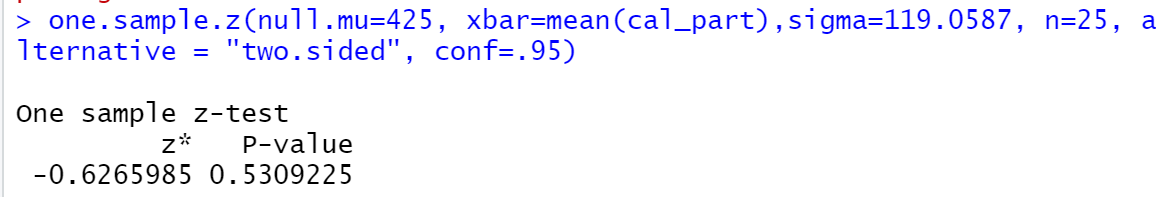
 since n < 30 and Population sd = 119.0587

3. State the decision rule

Decision Rule: Reject H0 if p <= alpha

Otherwise, do not reject H0

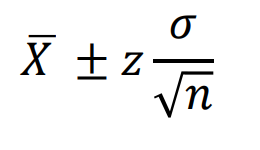
4. Compute the test statistic and the associated p-value



5. Conclusion

Fail to reject H0 since p = 0.5309225 > alpha. We do not have significant evidence at the alpha = 0.05 level that the mean calorie consumption for those who participated in the meal preparation differs from 425. We do not reject the null hypothesis that mean calorie consumption for the participated is 425.

**(3) Calculate a 90% confidence interval for the mean calorie intake for participants in the meal preparation. Interpret the confidence interval**.



At 90%, Z = 1.645

Mean = 410.1

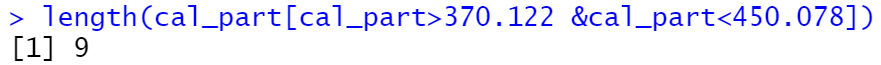
Sd = 121.5138

N = 25

410.1-1.645\*(121.5138/sqrt(25)) = 370.122

410.1+1.645\*(121.5138/sqrt(25)) = 450.078

90% confidence interval : (370.122, 450.078)



9/25 = 0.36 which is smaller than 0.5, so it would be rare for it to happen.

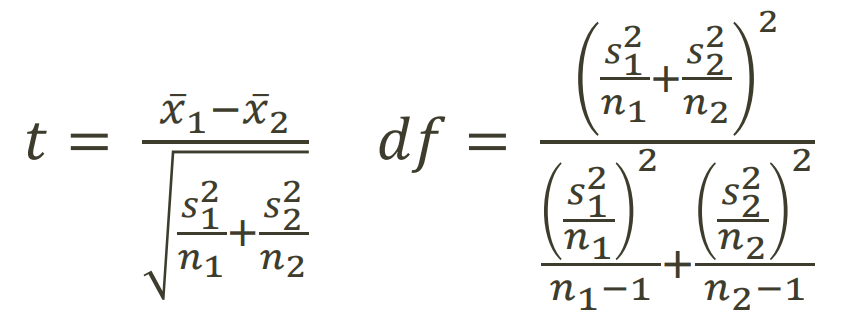
**(4) Formally test whether or not** **participants consumed more calories than non-participants at the alpha = 0.05 level using the 5 steps outlined in the module.**

1. Set up the hypotheses and select the alpha levelH0: mean1 = mean2 (participants consumed same calories as non-participants)

H1: mean1 > mean2 (participants consumed more calories than non-participants)

Alpha = 0.05

2. Select the appropriate test-statistic

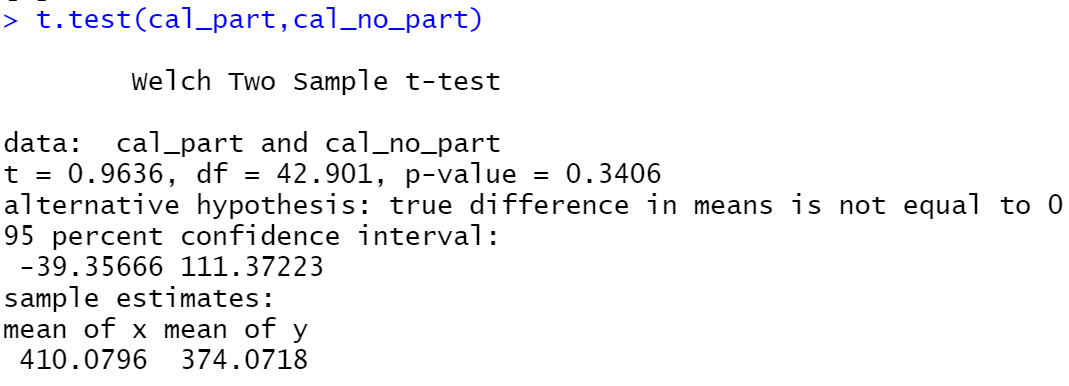


3. State the decision rule

Decision Rule: Reject H0 if p <= alpha

Otherwise, do not reject H0

4. Compute the test statistic and the associated p-value



5. Conclusion

Fail to reject H0 since p = 0.3406 > alpha. We do not have significant evidence at the alpha = 0.05 level that participants consumed more calories than non-participants. We do not reject the null hypothesis that participants consumed same calories as non-participants.

**(5) Are the assumptions of the test used in (4) met? How do you know?**

According to the data above, there is no element in confidence interval, so the assumptions in (4) can be right.

Code:

data <- read.csv(file = 'C:/Users/Yidow/Desktop/child.csv', fileEncoding="UTF-8-BOM")

data

cal\_part <- data$Calorie.Intake.for.participants

cal\_no\_part <- data$Calorie.intake.for.non.participants[!is.na(data$Calorie.intake.for.non.participants)]

summary(cal\_part)

sd(cal\_part)

#population sd

sd(cal\_part)\*sqrt((25-1)/25)

summary(cal\_no\_part)

sd(cal\_no\_part)

h\_part <- hist(cal\_part)

text(h\_part$mids,h\_part$counts,labels=h\_part$counts, adj=c(0.5, -0.5))

h\_part

h\_no\_part <- hist(cal\_no\_part)

text(h\_no\_part$mids,h\_no\_part$counts,labels=h\_no\_part$counts, adj=c(0.5, -0.5))

h\_no\_part

one.sample.z(null.mu=425, xbar=mean(cal\_part),sigma=119.0587, n=25, alternative = "two.sided", conf=.95)

410.1-1.645\*(121.5138/sqrt(25))

410.1+1.645\*(121.5138/sqrt(25))

length(cal\_part[cal\_part>370.122 &cal\_part<450.078])/25

t.test(cal\_part,cal\_no\_part)