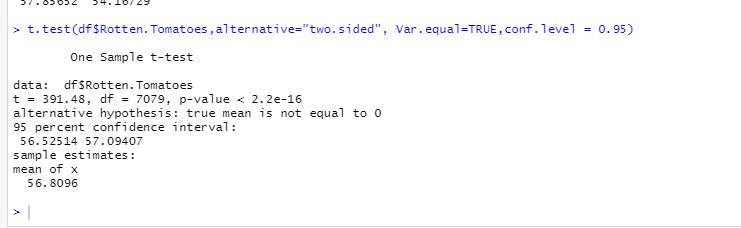
Yijun Wang

ALY 6010

12/02/2021

Question 1: The mean of Rotten.Tomatoes is greater equal to 55 or less than 55



Ho :Null Hypothesis: The mean of Rotten.Tomatoes >= 55

Ha: Alternate Hypothesis: The mean of Rotten.Tomatoes < 55

95% Confidence interval for difference: (56.52514,57.09407)

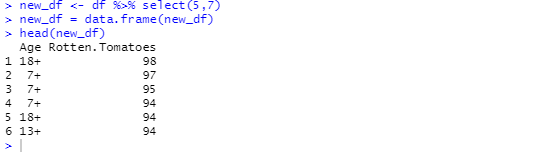
Difference: 7079

T-value = 391.48, P-value <2.2e-16

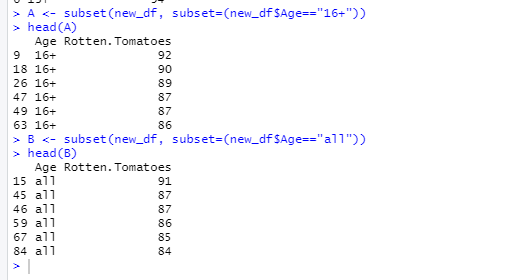
My null hypothesis is the mean of Rotten.Tomatoes >=55, and the alternative hypothesis is less than 55. From t-test, I got t-value 391.48, the p-value is <2.2e-16 which is less than the alpha value 0.05. So, we can reject the null hypothesis.

Question 2: "16+"&"all “has any difference on Rotten. Tomatoes score?

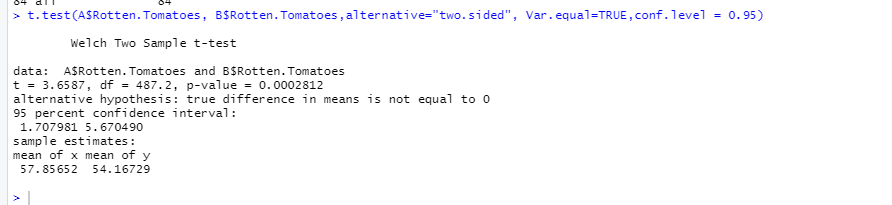
First of all, select the column of “Age” and “Rotten. Tomatoes” from the file.



And then filter the group of “16+” and “all” at “Age” column



Next, do the t-test for these two groups of data.



Difference = mu (A Sample) – mu (B Sample)

Ho: Null Hypothesis: Difference = 0

Ha: Alternate Hypothesis: Difference ≠ 0

95% Confidence interval for difference: (1.707981, 5.670490)

Difference: 487.2

T-value = 3.6587, P-value = 0.0002812

My null hypothesis is difference = 0, and the alternative hypothesis is ≠ 0. From t-test, I got t-value 3.6587, the p-value is 0.0002812which is less than the alpha value 0.05. So, we can reject the null hypothesis. In other words, the difference between “16+” and “all” is not equal to zero.

In the summary, the mean of “Rotten. Tomatoes “is 56.8096, “16+” is 57.85652, “all” is 54.16729. So, the group of “16+” is higher than the mean, and “all” is less than the mean. But in other words, "all" has a wider age group and a larger number of people, which can lead to low scores.

Code：

#Part 2

#a. The mean of Rotten.Tomatoes is greater equal to 55 or less then 55

t.test(df$Rotten.Tomatoes,alternative="two.sided", Var.equal=TRUE,conf.level = 0.95)

#b. "16+"&"all"has any difference on Rotten.Tomatoes score?

new\_df <- df %>% select(5,7)

new\_df = data.frame(new\_df)

head(new\_df)

A <- subset(new\_df, subset=(new\_df$Age=="16+"))

head(A)

B <- subset(new\_df, subset=(new\_df$Age=="all"))

head(B)

t.test(A$Rotten.Tomatoes, B$Rotten.Tomatoes,alternative="two.sided", Var.equal=TRUE,conf.level = 0.95)