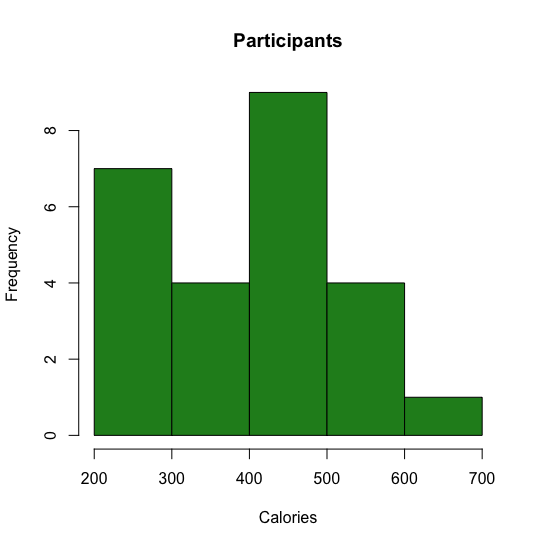
*CS555 Assignment 2*

*Zuowen Tang*

*Feb 16, 2023*

(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. **(2 points)**



The distribution of participants is left-skewed (the mean is smaller than the median.)

Summary:

Min.:211.0, Median:424.9, Mean:410.1

Chart, histogram

Description automatically generated

The distribution of participants is also left-skewed (the mean is smaller than the median.)

Summary:

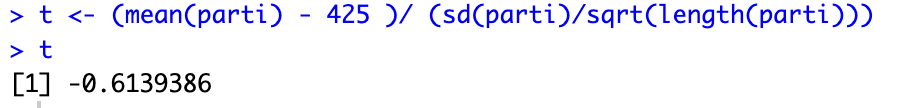
Min.:139.7, Median:374.7, Mean:374.1

The data distribution of both of them is both concentrated in the center.

(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. **(6 points)**

Text, letter

Description automatically generated



|t| < 1.711, so do not reject H0.

**Step 5: Conclusion**

We cannot say that the mean of calorie consumption for people who participated in the meal preparation is not differ from 425Text, letter

Description automatically generated.

(3) Calculate a **90% confidence interval** for the mean calorie intake for participants in the meal preparation. Interpret the confidence interval. **(4 points)**

Graphical user interface, text, application, email

Description automatically generated

90% confidence interval for participants is 325.2279 - 411.9158.

(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. **(6 points )**

Text, letter

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated with medium confidence|t| < 1.721

So do not reject H0.

**Step 5: Conclusion**

Text

Description automatically generated

The participants consumed more calories than non-participants.

p\_data <- read.csv(file = "/Users/zuowen/Documents/BU Spring 2023/CS555 Machine Learning/Homework/Assign2/p.csv", header = TRUE)

nonp\_data <- read.csv(file = "/Users/zuowen/Documents/BU Spring 2023/CS555 Machine Learning/Homework/Assign2/non-p.csv", header = TRUE)

hist\_p <- c(p\_data$X0)

hist\_nonp <- c(nonp\_data$X0)

hist(hist\_p, breaks = 5, xlab = "Calories", main = "Participants", col = "forestgreen")

hist(hist\_nonp, breaks = 5, xlab = "Calories", main = "Non Participants", col = "forestgreen")

summary(nonp\_data)

parti <- unlist(p\_data, use.names = FALSE)

parti <- subset(parti,parti!="NA")

nparti <- unlist(nonp\_data,use.names = FALSE)

nparti <- subset(nparti,nparti !="NA")

t.test(parti, mu = 425, alternative="two.sided", conf.level=0.95)

t.test(nparti, mu=mean(nparti), alternative="two.sided", conf.level = 0.90)

t.test(parti, nparti, alternative="greater", conf.level=0.95)

t <- (mean(parti) - 425 )/ (sd(parti)/sqrt(length(parti)))

t

mean(parti)

mean(nparti)

sd(parti)

sd(nparti)

length(parti)

length(nparti)