library(UsingR)

install.packages("asbio")

library(asbio)

setwd("/Users/estherji")

#Part 1

calor <- read.csv("cs555col.csv", header=TRUE)

head(calor)

#1

summary(calor$Calorie.Intake.for.participants)

summary(calor$Calorie.intake.for.non.participants)

par(mfrow=c(1,2))

hist(calor$Calorie.Intake.for.participants, col=hcl(0),xlab = "Calorie.Intake.for.participants", main="distribution")

hist(calor$Calorie.intake.for.non.participants, col=hcl(0),xlab = "Calorie.intake.for.non.participants", main="distribution")

par(mfrow=c(1,1))

Chart, histogram

Description automatically generated

#The graph of Calorie.Intake.for.participants is not symmatric,

#The values of the two highest points are different

#The graph of Calorie.Intake.for.non.participants is also not symmatric,

#However, the values of the two highest points are same.

#2

length(calor$Calorie.Intake.for.participants)

#25

#h0 : mean = 425

#h1 : mean != 425

#df = 24

abs(qt(0.05, 24))

#critical value : 1.710882

t.test(calor$Calorie.Intake.for.participants, mu = 425, alternative = "two.sided")

'''

result :

One Sample t-test

data: calor$Calorie.Intake.for.participants

t = -0.61394, df = 24, p-value = 0.545

alternative hypothesis: true mean is not equal to 425

95 percent confidence interval:

359.9212 460.2380

sample estimates:

mean of x

410.0796

Since, -0.6 < 1.71. Thus, fail to reject null.

'''

#3

t.test(calor$Calorie.Intake.for.participants,conf.level = 0.9)

'''

One Sample t-test

data: calor$Calorie.Intake.for.participants

t = 16.874, df = 24, p-value = 8.15e-15

alternative hypothesis: true mean is not equal to 0

90 percent confidence interval:

368.5004 451.6588

sample estimates:

mean of x

410.0796

'''

#The 90% confidence interval for the mean calorie

#intake for participants in the meal preparation is (368.5-451.66)

#4

#h0 : participants do not consumed more calories than non-participants

#h1 : participants consumed more calories than non-participants

length(calor$Calorie.Intake.for.participants)

length(calor$Calorie.intake.for.non.participants)

#df = 25+22-2=45

abs(qt(0.05, 45))

#1.679427

t.test(calor$Calorie.Intake.for.participants, calor$Calorie.intake.for.non.participants )

'''

Welch Two Sample t-test

data: calor$Calorie.Intake.for.participants and calor$Calorie.intake.for.non.participants

t = 0.9636, df = 42.901, p-value = 0.3406

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-39.35666 111.37223

sample estimates:

mean of x mean of y

410.0796 374.0718

'''

#since t= 0.9636 < 1.679427, we can not reject null.

#participants do not consumed more calories than non-participants

#5

#the assumptions of the test used in (4) does not met.

#because the data is not symmetric.