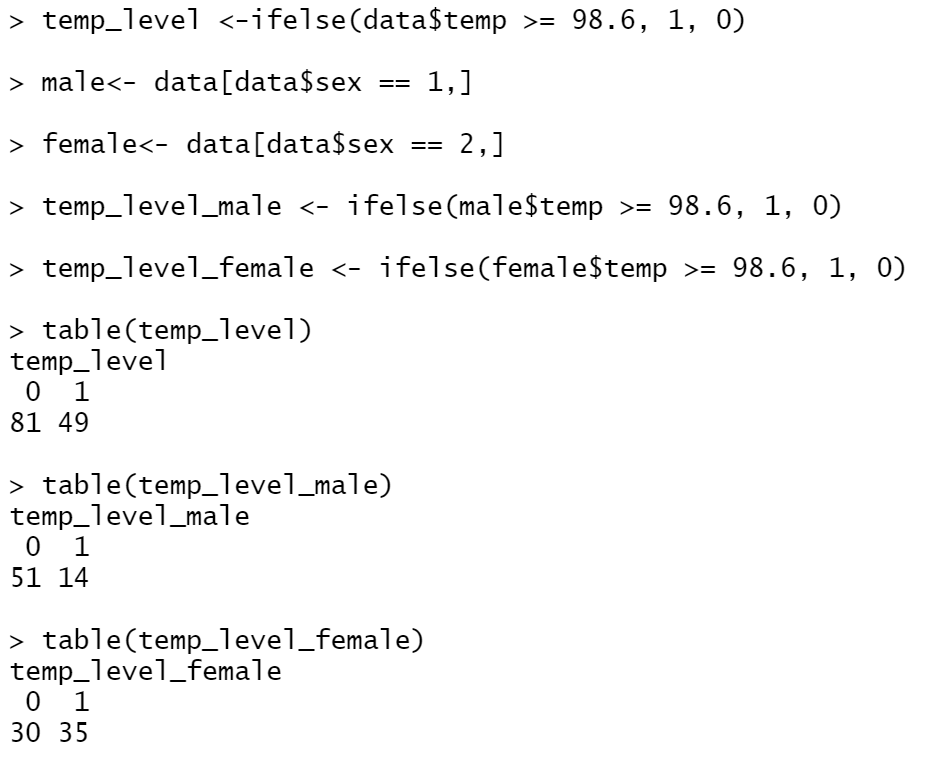
Yiduo Feng

CS 555

Homework 6

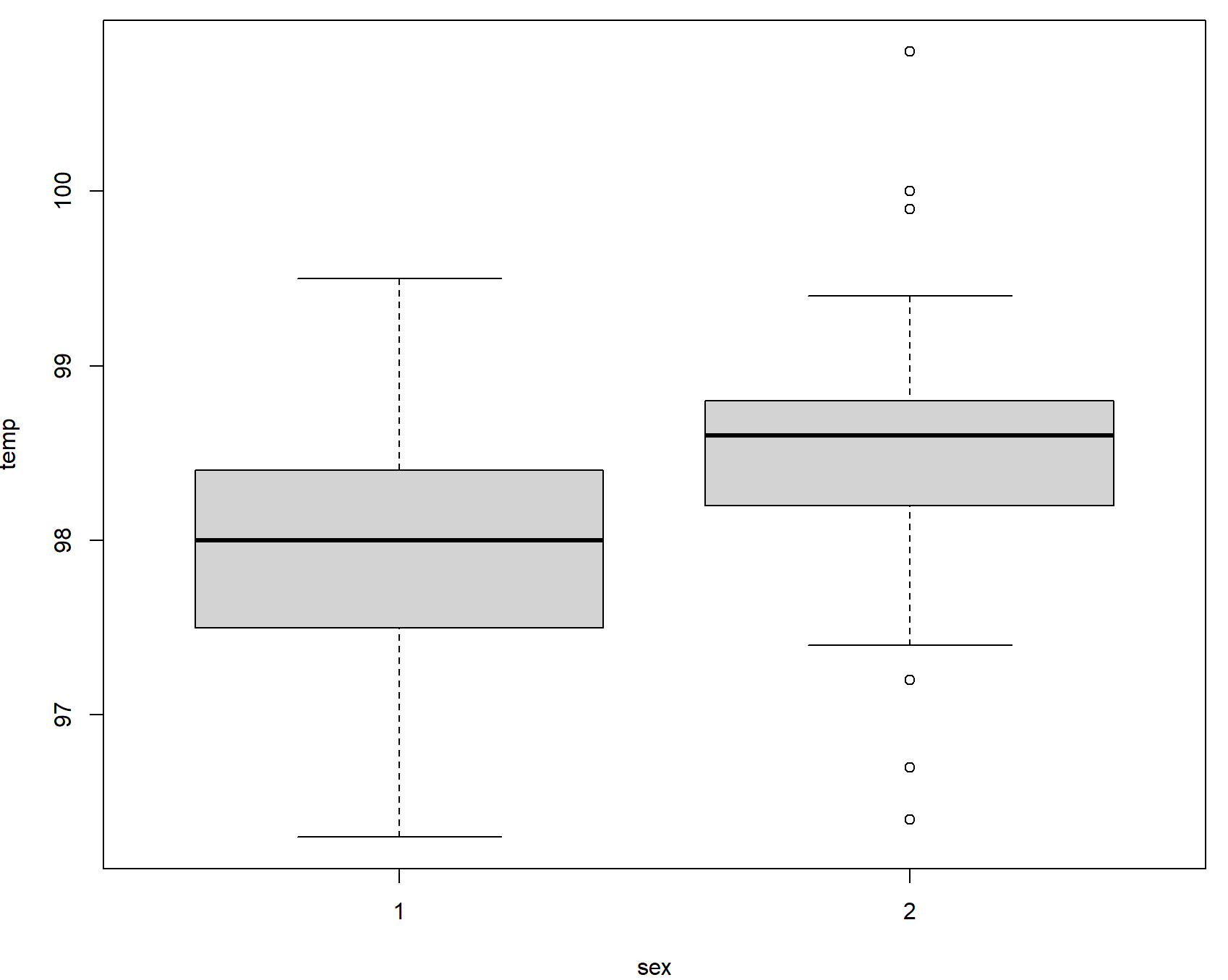
08/15/2022

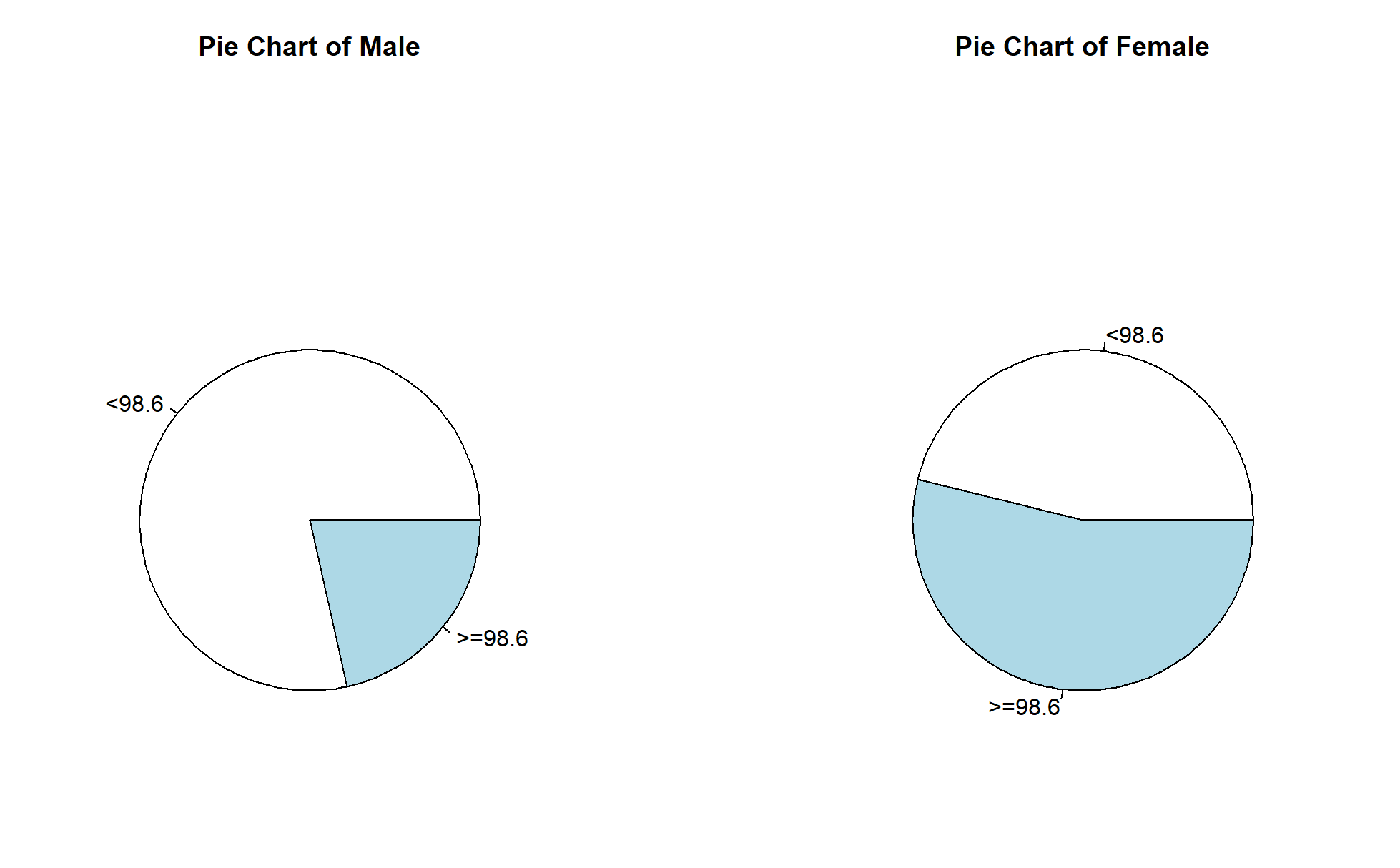
**(1) We are interested in whether the proportion of men and women with body temperatures greater than or equal to 98.6 degrees Fahrenheit are equal. Therefore, we need to dichotomize the body temperature variable. Create a new variable, called “temp\_level” in which temp\_level = 1 if body temperature >= 98.6 and temp\_level=0 if body temperature < 98.6.**

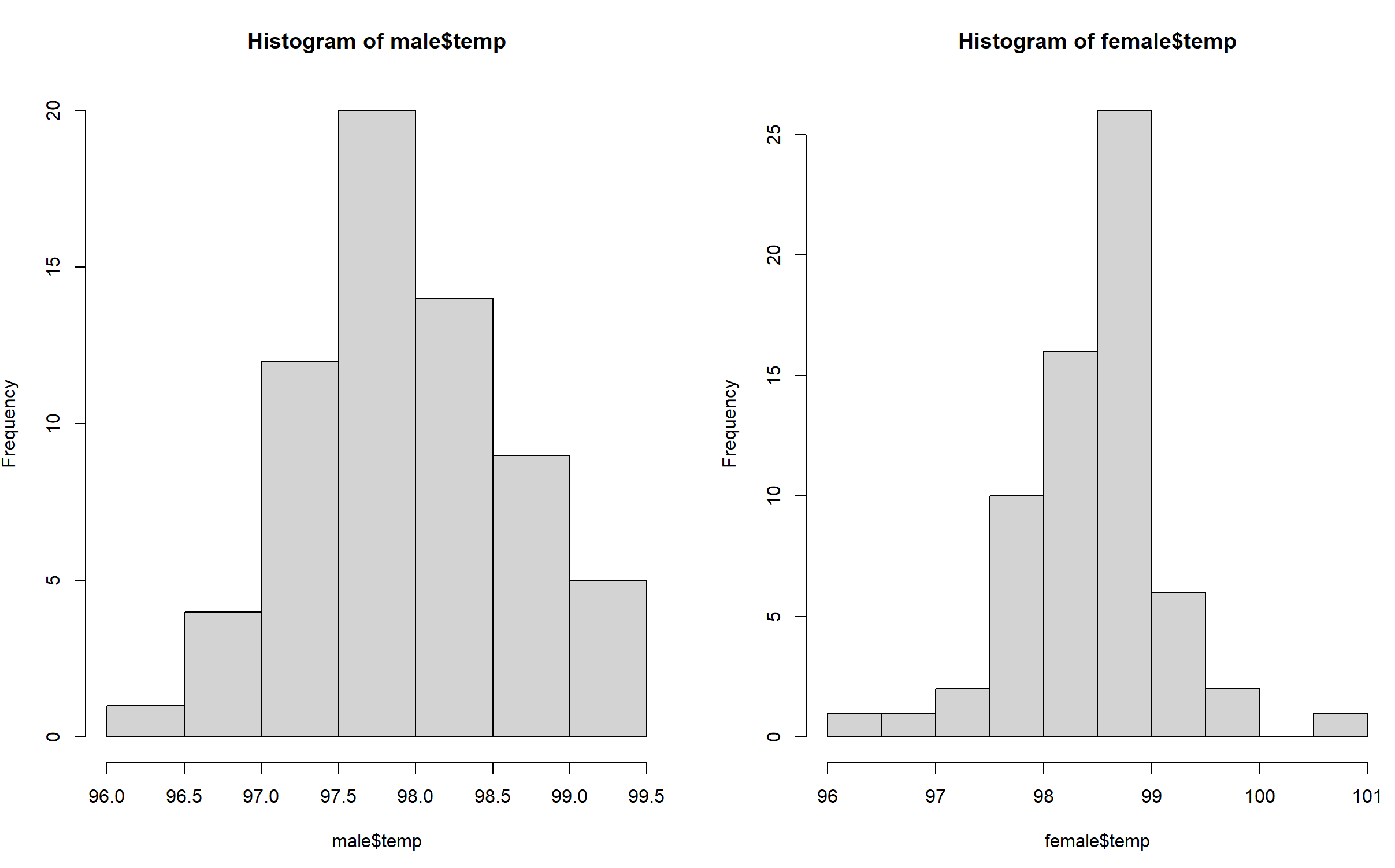


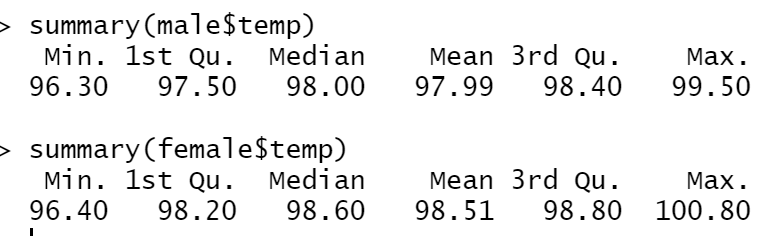
**(2) Summarize the data relating to body temperature level (i.e., the variable you created above) by sex.**

|  |  |  |  |
| --- | --- | --- | --- |
| Gender\Temperature | <98.6 | >=98.6 | Sum |
| Male | 14 | 51 | 65 |
| Female | 35 | 30 | 65 |
| Sum | 49 | 81 | 130 |

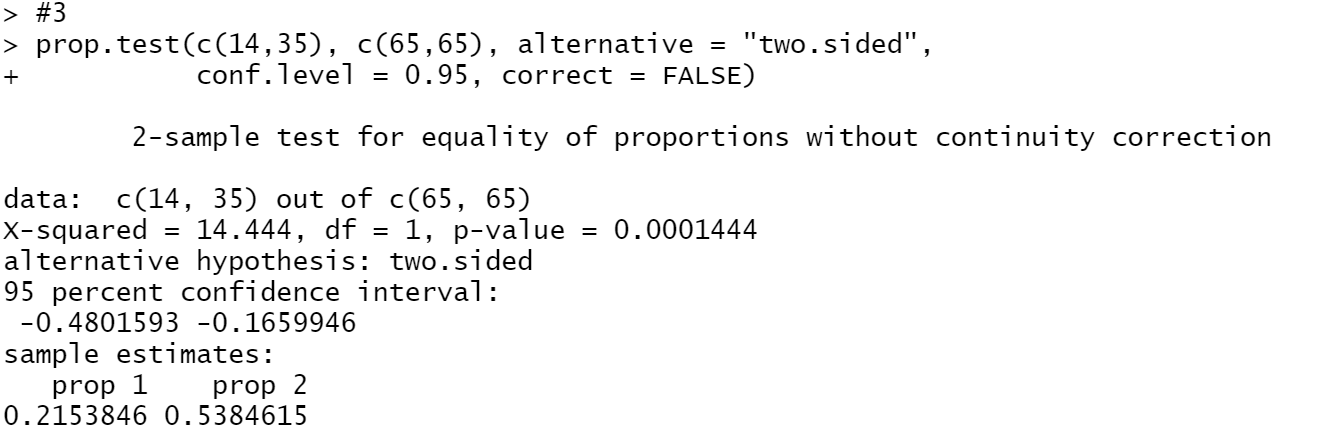








(3) Calculate the risk difference for high body temperature level between men and women. Formally test (at the **alpha=0.05** level) whether the proportion of people with higher body temperatures (greater than or equal to **98.6**) is the same across men and women based on this effect measure. You should be showing all 5 steps in the 5-step recipe for testing.



So risk difference = 0.5384615-0.2153846 = 0.3230769

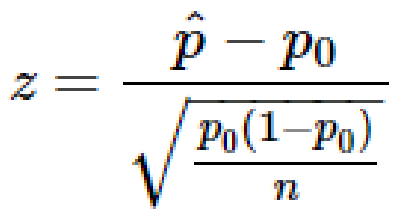
1. Set up the hypotheses and select the alpha level

H0 : 𝛽1= 0 (the proportion of people with higher body temperatures is the same across men and women based on this effect measure.)

H1 : 𝛽1≠ 0 (the proportion of people with higher body temperatures is not the same across men and women based on this effect measure.)

𝛼 = 0.05

1. Select the appropriate test-statistic



1. State the decision rule

Decision Rule: Reject H0 if 𝑝 ≤ 𝛼. Otherwise, do not reject H

1. Compute the test statistic

According to the code above,



P < 0.05

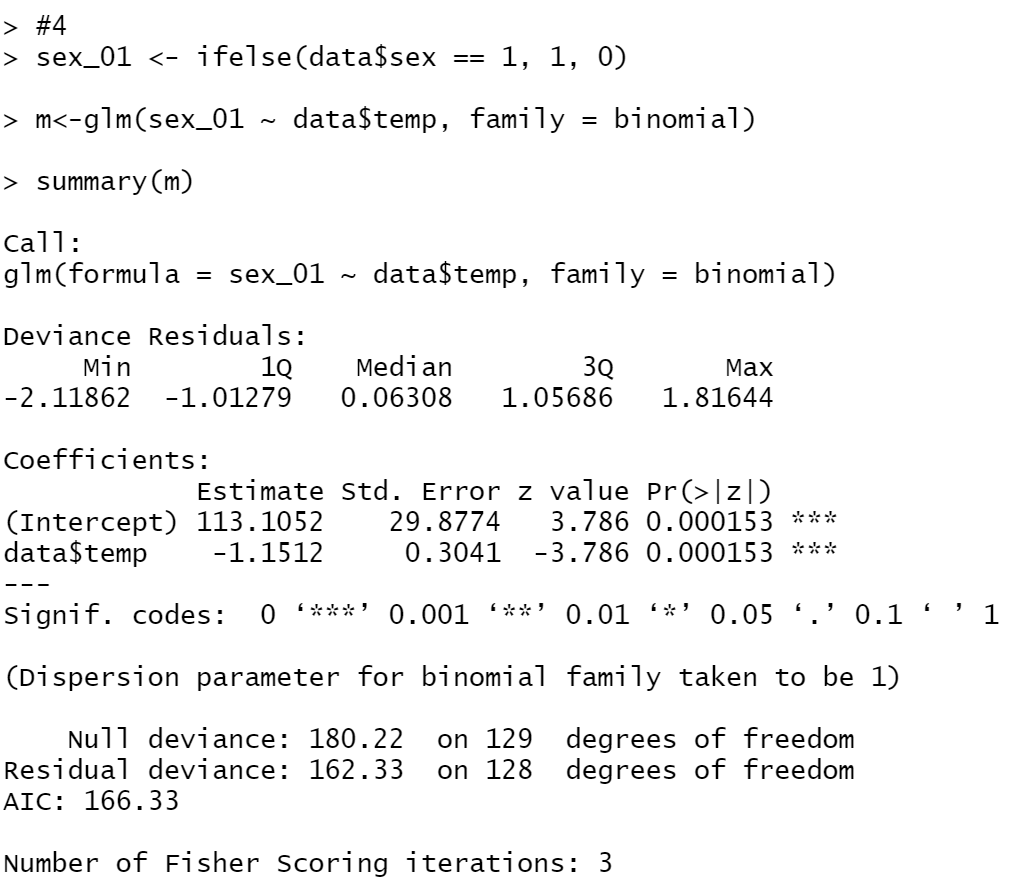
1. Conclusion

Reject H0

since 𝑝 ≤ 𝛼. We have significant evidence at the 𝛼 = 0.05 level

that the proportion of people with higher body temperatures is not the same across men and women based on this effect measure. That is, there is evidence that the proportion of people with higher body temperatures is not the same across men and women based on this effect measure.

(4) Perform a logistic regression with sex as the only explanatory variable. Formally test (at the **alpha=0.05** level) if the odds of having a temperature **greater than or equal to 98.6** is the same between males and females. Again, please show all 5 steps. Additionally, include the odds ratio for sex and the associated **95%** confidence interval in your summary, and interpret the value of the odds ratio. Lastly, what is the c-statistic for this model?



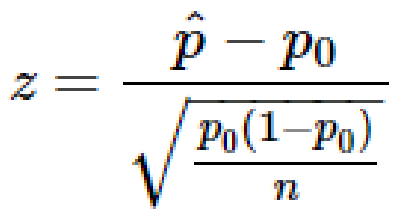
1. Set up the hypotheses and select the alpha level

H0 : 𝛽1= 0 (the odds of having a temperature greater than or equal to 98.6 is the same between males and females..)

H1 : 𝛽1≠ 0 (the odds of having a temperature greater than or equal to 98.6 is not the same between males and females..)

𝛼 = 0.05

1. Select the appropriate test-statistic

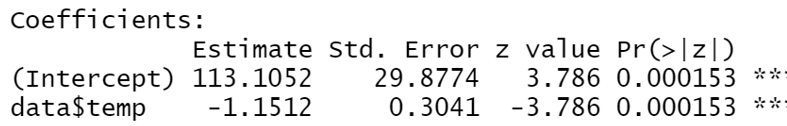


1. State the decision rule

Decision Rule: Reject H0 if 𝑝 ≤ 𝛼. Otherwise, do not reject H

1. Compute the test statistic

According to the code above,



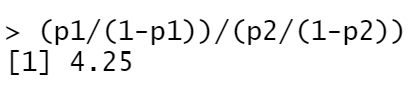
P < 0.05

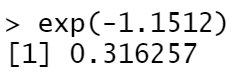
1. Conclusion

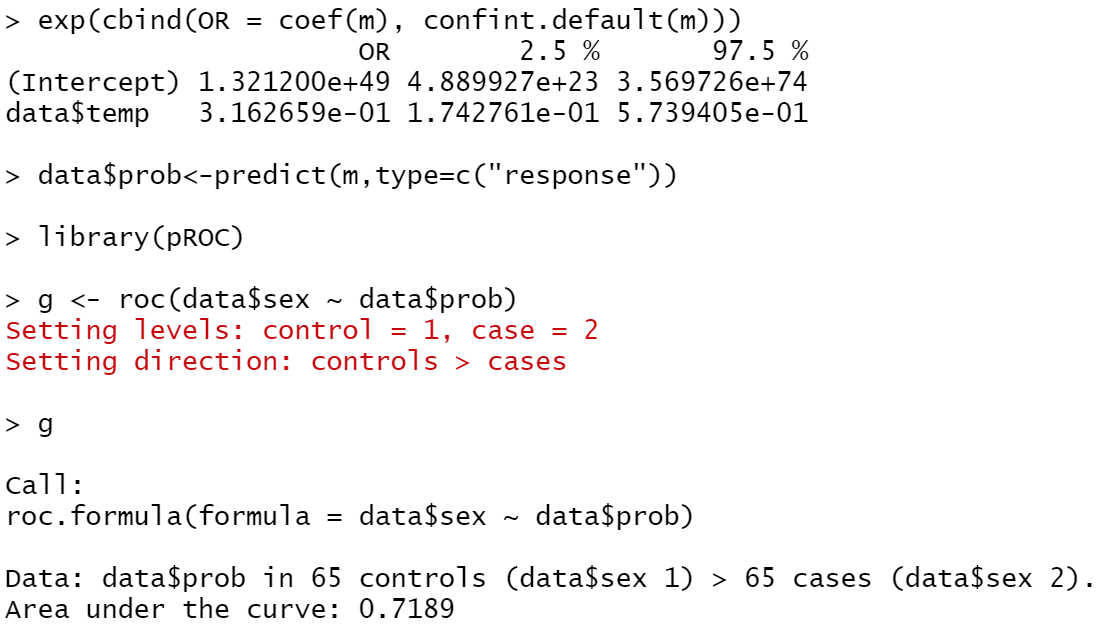
Reject H0

since 𝑝 ≤ 𝛼. We have significant evidence at the 𝛼 = 0.05 level

that the odds of having a temperature greater than or equal to 98.6 is not the same between males and females.. That is, there is evidence that the odds of having a temperature greater than or equal to 98.6 is not the same between males and females..

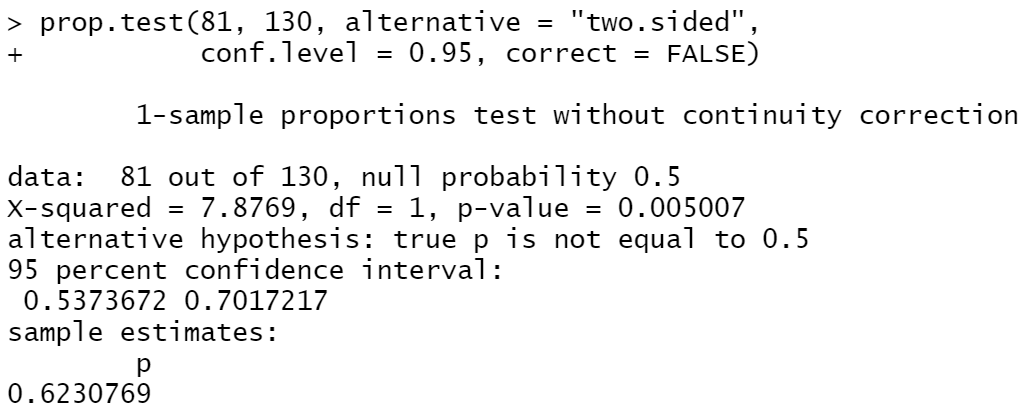
Odd ratio: 





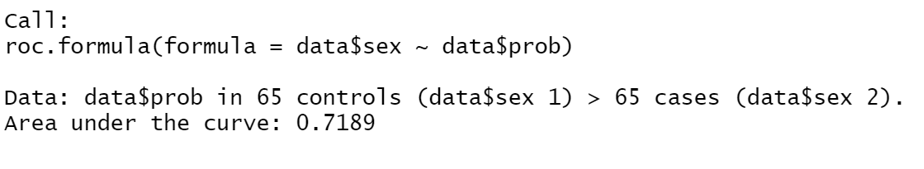
Odd ratio is larger than 1, and the odds of high temperature is 2.6 times higher among female with male.

Confidence interval:

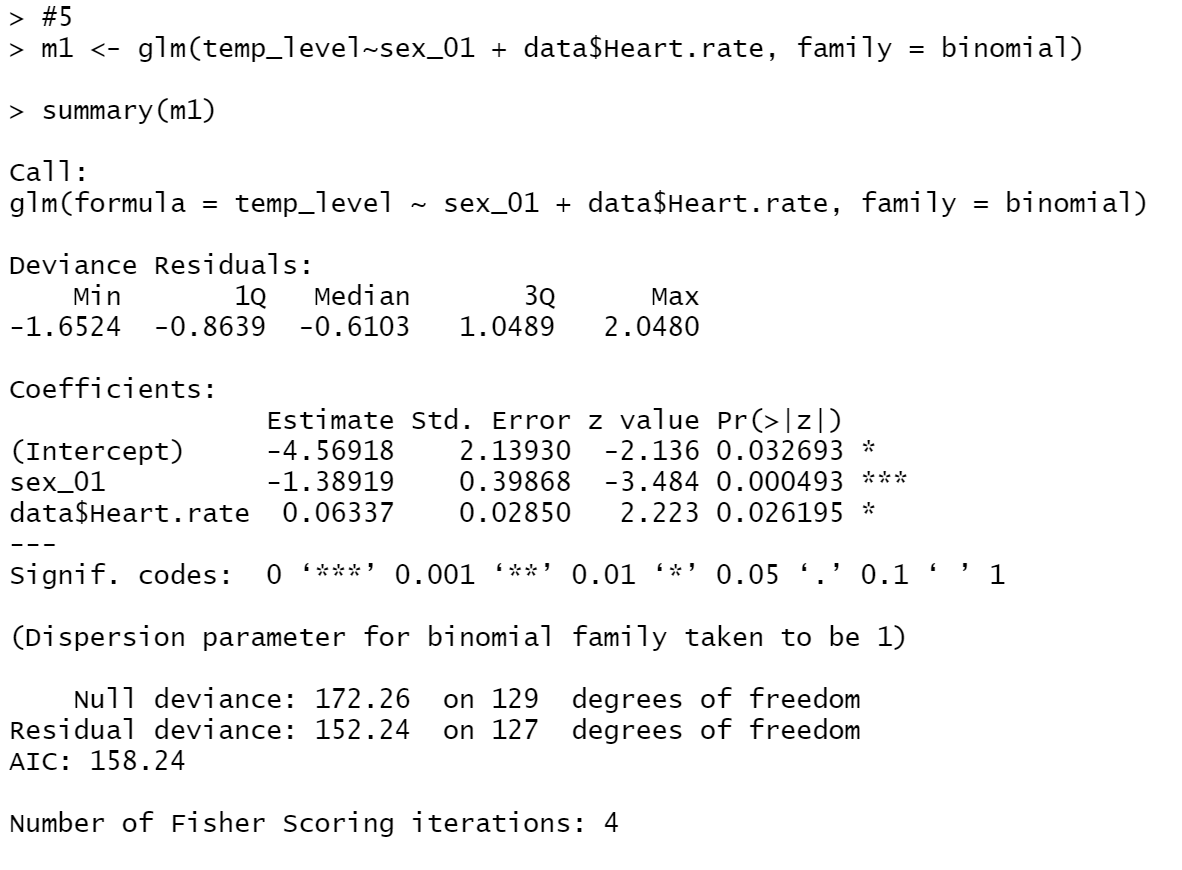


(0.5373672, 0.7017217)

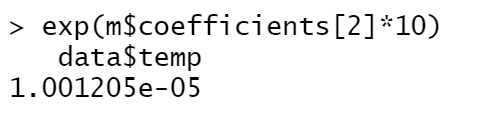
c-statistic：

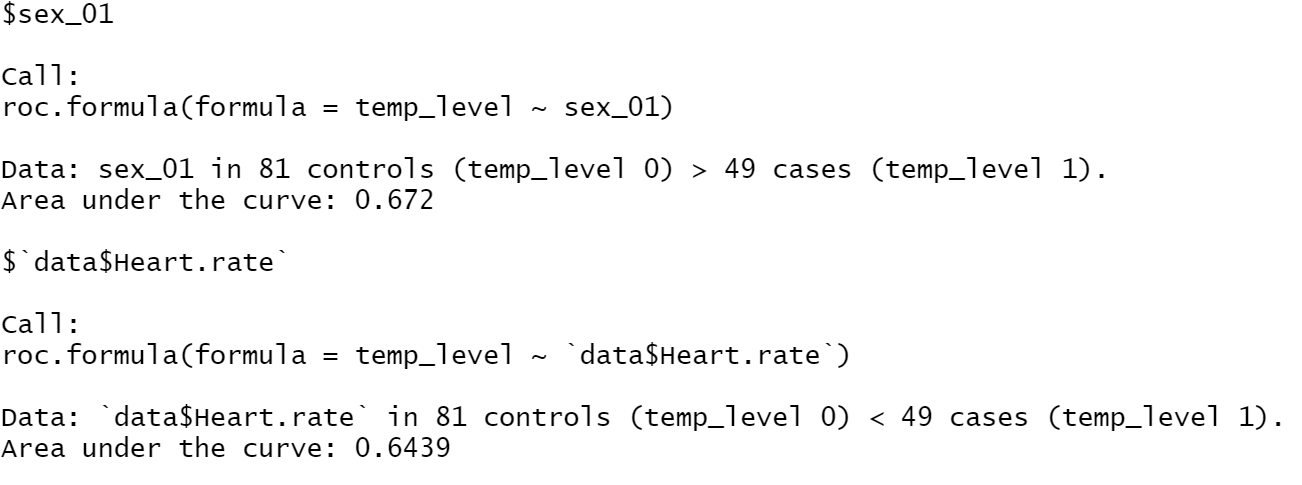


(5) Perform multiple logistic regression predicting body temperature level from sex and heart rate. Briefly summarize the output from this mode (no need to go through all 5 steps). Give the odds ratio for sex. Also, report the odds ratio for heart rate **(for a 10-beat increase)**. What is the c-statistic of this model? **(5 points)**



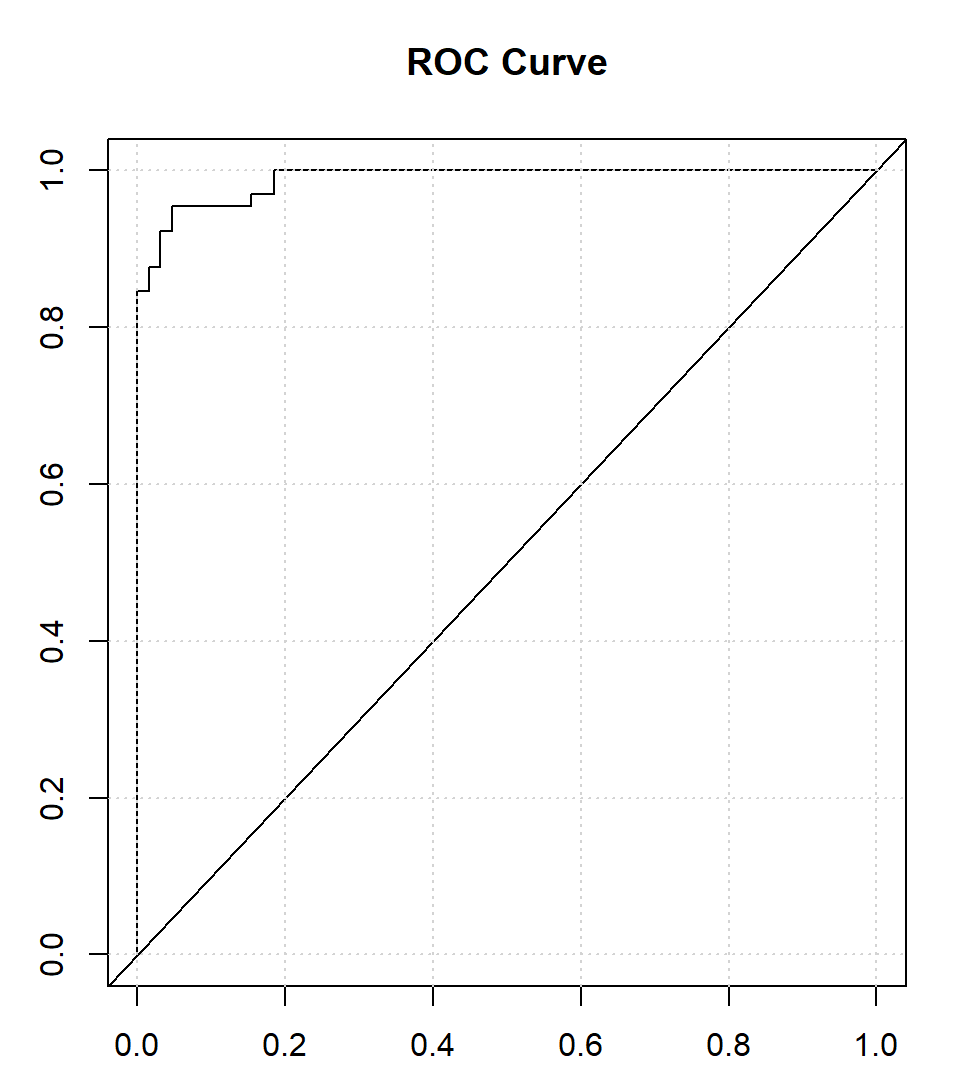
It shows that the coefficient estimate between sex is -1.38919 which is bigger than 0.06337 for heart beat, so gender is a risker variable for temperature..

Odd ratio:



(6) Which model fit the data better? Support your response with evidence from your output. Present the ROC curve for the model you choose.

According to the table in question 2, we can find that the sex variable should impact the temperate a lot and it should be more effect more than heart rate. However, only logistic regression gives us reasonable results, and especially multiple logistic regression, it shows us the more details information among three elements which are temperature, gender and heart rate.



Code:

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

data

temp\_level <-ifelse(data$temp >= 98.6, 1, 0)

male<- data[data$sex == 1,]

female<- data[data$sex == 2,]

temp\_level\_male <- ifelse(male$temp >= 98.6, 1, 0)

temp\_level\_female <- ifelse(female$temp >= 98.6, 1, 0)

table(temp\_level)

table(temp\_level\_male)

table(temp\_level\_female)

#boxplot(temp~sex, data=data)

# par(mfrow=c(1,2))

# pie(table(temp\_level\_male),labels = c("<98.6",">=98.6"),

# main="Pie Chart of Male")

# pie(table(temp\_level\_female),labels = c("<98.6",">=98.6"),

# main="Pie Chart of Female")

# hist(male$temp)

# hist(female$temp)

summary(male$temp)

summary(female$temp)

#3

prop.test(c(14,35), c(65,65), alternative = "two.sided",

conf.level = 0.95, correct = FALSE)

#4

sex\_01 <- ifelse(data$sex == 1, 1, 0)

m<-glm(sex\_01 ~ data$temp, family = binomial)

summary(m)

exp(-1.1512)

exp(cbind(OR = coef(m), confint.default(m)))

data$prob<-predict(m,type=c("response"))

library(pROC)

g <- roc(data$sex ~ data$prob)

g

prop.test(81, 130, alternative = "two.sided",

conf.level = 0.95, correct = FALSE)

p1<-0.5384615

p2<-0.2153846

(p1/(1-p1))/(p2/(1-p2))

#5

m1 <- glm(temp\_level~sex\_01 + data$Heart.rate, family = binomial)

summary(m1)

exp(m$coefficients[2]\*10)

roc(temp\_level~sex\_01+data$Heart.rate)

#6

data$prob<-predict(m1,type=c("response"))

g <- roc(data$sex ~ data$prob)

par(mfrow=c(1,1))

plot(1-g$specificities, g$sensitivities, type = "l",

xlab = "1 - Specificity", ylab = "Sensitivity", main = "ROC Curve")

abline(a=0,b=1)

grid()