1. New Variable:

Graphical user interface, text, application, email

Description automatically generated

1. Summary

A picture containing text

Description automatically generated

3.)

Risk Difference:

Text

Description automatically generated

Step 1:

Set up the hypotheses and select the alpha level

* H0: μFemale\_greater = μMale\_greater
* H1: μFemale\_greater  ≠ μMale\_greater
* α=0.05

Step 2: Select the appropriate test statistic

A picture containing text, clock

Description automatically generated

Step 3:

α = .05/2

Critical value = 1.96

Decision Rule: Reject H0 if Z is greater than or equal to 1.96

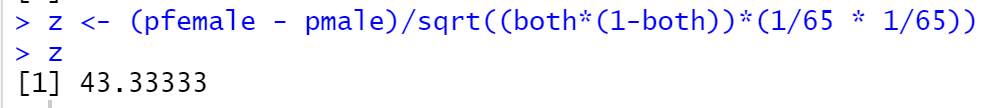
Step 4

Sample proportion:

Pmale = 14/65 = . 0.2153

Pfemale = 35/65 = 0.5385

Pboth = 49/130 = 0.3769



Step 5

Since the Z statistic is greater than the critical value, we reject the null hypothesis that the proportion of people with higher temperatures is the same across men and women. Yes, females have a bigger percentage of higher body temperature.

4.)

Logistic Regression:

Text

Description automatically generated

Text

Description automatically generated

Step 1:

Set up the hypotheses and select the alpha level

* H0: μFemale\_greater = μMale\_greater
* H1: μFemale\_greater  ≠ μMale\_greater
* α=0.05

Step 2: Select the appropriate test statistic

* z=β1SEβ^1

Step 3:

α = .05/2

Critical value = 1.96

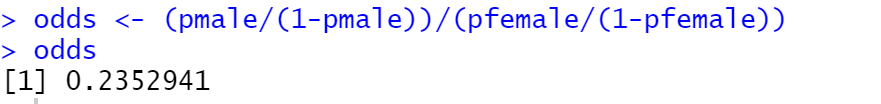
Decision Rule: Reject H0 if Z is greater than or equal to 1.96

Step 4 : Z = 1.4469/.3911 = 3.699

Step 5 :

Since the Z statistic is greater than the critical value, we reject the null hypothesis that the proportion of people with higher temperatures is the same across men and women.

Odds Ratio: Lower odds in males



Female:

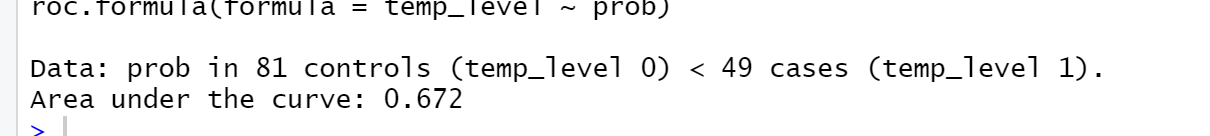


**95% Confidence Interval: (-.480, -.166)**

Text, letter

Description automatically generated

**C – Statistic: .672**



5.) **Multiple Logistic Regression**

Text

Description automatically generated with low confidence

Text

Description automatically generated

This model has a positive slope for sex & heart rate and z values higher than 1.96 rejecting the null hypothesis stated earlier.

**Odds ratio per 10 unit increase**

Text

Description automatically generated

**C-Statistic: .7289**

Text, letter

Description automatically generated

6.) Multiple Logistic regression was the better model because the c-statistic was greater.

Plot of the ROC curve: Chart, scatter chart

Description automatically generated

R Code:

rm(list=ls()); cat("\014")

#Set directory

setwd("C:/Users/HP/Documents/555")

getwd()

#1.)

#Import spreadsheet

heart = read.csv("heart.csv", header = TRUE)

as.data.frame(heart)

#Create new variable

heart$temp\_level <- ifelse(heart$temp < 98.6,

c("0"), c("1"))

heart

#2.)

#Factor Sex

is.factor(heart$sex)

heart$sex = factor(heart$sex)

levels(heart$sex) <- c("Male", "Female")

is.factor(heart$sex)

attach(heart)

aggregate(sex, by=list(temp\_level), summary)

#3

#critical value

qnorm(.975)

#sample proportions

pmale <- 14/65

pfemale <- 35/65

both <- 49/130

pmale

pfemale

both

#Risk difference

risk <- pfemale - pmale

risk

z <- (pfemale - pmale)/sqrt((both\*(1-both))\*(1/65 \* 1/65))

z

#4

#make temp\_level numeric

temp\_level<- as.numeric(temp\_level)

#logistic regression

m <- glm(temp\_level ~ sex, family=binomial)

m

summary(m)

#odds ratio

odds1 <- (pmale/(1-pmale))/(pfemale/(1-pfemale))

odds1

odds2 <- (pfemale/(1-pfemale))/(pmale/(1-pmale))

odds2

#confidence interval

prop.test(c(14, 35), c(65, 65), conf.level=0.95, correct=FALSE)

#C statistic

library(pROC)

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

g <- roc(temp\_level ~ prob)

g

#5

#Multiple Logistic Regression

#logistic regression

m2 <- glm(temp\_level ~ sex + Heart\_rate, family=binomial)

m2

summary(m2)

#Odds Ratio per 10 unit increase

exp(10 \* cbind(OR = coef(m2), confint.default(m2)))

#C statistic

prob <-predict(m2, type=c("response"))

g2 <- roc(temp\_level ~ prob)

g2

#6

plot(g)

plot(g2)

roc(temp\_level ~ prob, plot=TRUE, legacy.axes=TRUE, main = "Multiple Logistic Regression",

percent=TRUE, xlab="False Positive (%)", ylab="True Positive (%)", col="purple", lwd=4)