Mathematics for Engineers - II

Project Report

(MTH1711)

Heart Disease Analysis

Group Members:

Vedansh Kumawat

220489

Anusha Singh

220641

Shivanshi Goel
220457

BML MUNJAL UNIVERSITY ™

Department of Computer Science and Engineering
School of Engineering and Technology

May 2023

PREPROCESSING OF DATA

```r
library(readr)
library(dplyr)
library(car)
library(lattice)
library(ggplot2)
```

```r
heart <- read_csv("heart.csv")
```

```
Parsed with column specification:
cols(
  age = col_double(),
  sex = col_double(),
  cp = col_double(),
  trestbps = col_double(),
  chol = col_double(),
  fbs = col_double(),
  restecg = col_double(),
  thalach = col_double(),
  exang = col_double(),
  oldpeak = col_double(),
  slope = col_double(),
  ca = col_double(),
  thal = col_double(),
  target = col_double()
)
```

```r
str(heart)
```

```
tibble [303 x 14] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ age     : num [1:303] 63 37 41 56 57 57 56 44 52 57 ...
 $ sex     : num [1:303] 1 1 0 1 0 1 0 1 1 1 ...
 $ cp      : num [1:303] 3 2 1 1 0 0 1 1 2 2 ...
 $ trestbps: num [1:303] 145 130 130 120 120 140 140 120 172 150 ...
 $ chol    : num [1:303] 233 250 204 236 354 192 294 263 199 168 ...
 $ fbs     : num [1:303] 1 0 0 0 0 0 0 0 1 0 ...
 $ restecg : num [1:303] 0 1 0 1 1 1 0 1 1 1 ...
 $ thalach : num [1:303] 150 187 172 178 163 148 153 173 162 174 ...
 $ exang   : num [1:303] 0 0 0 0 1 0 0 0 0 0 ...
 $ oldpeak : num [1:303] 2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
 $ slope   : num [1:303] 0 0 2 2 2 1 1 2 2 2 ...
 $ ca      : num [1:303] 0 0 0 0 0 0 0 0 0 0 ...
 $ thal    : num [1:303] 1 2 2 2 2 1 2 3 3 2 ...
 $ target  : num [1:303] 1 1 1 1 1 1 1 1 1 1 ...
 - attr(*, "spec")=
 .. cols(
 ..   age = col_double(),
 ..   sex = col_double(),
 ..   cp = col_double(),
 ..   trestbps = col_double(),
 ..   chol = col_double(),
 ..   fbs = col_double(),
 ..   restecg = col_double(),
 ..   thalach = col_double(),
 ..   exang = col_double(),
 ..   oldpeak = col_double(),
 ..   slope = col_double(),
 ..   ca = col_double(),
 ..   thal = col_double(),
 ..   target = col_double()
 .. )
```

```
summary(heart)
```

```
      age             sex               cp            trestbps
 Min.   :29.00   Min.   :0.0000   Min.   :0.000   Min.   : 94.0
 1st Qu.:47.50   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:120.0
 Median :55.00   Median :1.0000   Median :1.000   Median :130.0
 Mean   :54.37   Mean   :0.6832   Mean   :0.967   Mean   :131.6
 3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:140.0
 Max.   :77.00   Max.   :1.0000   Max.   :3.000   Max.   :200.0
      chol           fbs             restecg          thalach
 Min.   :126.0   Min.   :0.0000   Min.   :0.0000   Min.   : 71.0
 1st Qu.:211.0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:133.5
 Median :240.0   Median :0.0000   Median :1.0000   Median :153.0
 Mean   :246.3   Mean   :0.1485   Mean   :0.5281   Mean   :149.6
 3rd Qu.:274.5   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:166.0
 Max.   :564.0   Max.   :1.0000   Max.   :2.0000   Max.   :202.0
     exang           oldpeak          slope            ca
 Min.   :0.0000   Min.   :0.00   Min.   :0.000   Min.   :0.0000
 1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:1.000   1st Qu.:0.0000
 Median :0.0000   Median :0.80   Median :1.000   Median :0.0000
 Mean   :0.3267   Mean   :1.04   Mean   :1.399   Mean   :0.7294
 3rd Qu.:1.0000   3rd Qu.:1.60   3rd Qu.:2.000   3rd Qu.:1.0000
 Max.   :1.0000   Max.   :6.20   Max.   :2.000   Max.   :4.0000
      thal           target
 Min.   :0.000   Min.   :0.0000
 1st Qu.:2.000   1st Qu.:0.0000
 Median :2.000   Median :1.0000
 Mean   :2.314   Mean   :0.5446
 3rd Qu.:3.000   3rd Qu.:1.0000
 Max.   :3.000   Max.   :1.0000
```

```
heart$sex <- heart$sex %>% factor(levels=c(0,1),
                                    labels=c("Female","Male"))
heart$target <- heart$target %>% factor(levels=c(0,1),
labels=c(0,1))


heart <- heart %>% rename(gender = sex)

str(heart)
```

```
tibble [303 x 14] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ age     : num [1:303] 63 37 41 56 57 57 56 44 52 57 ...
 $ gender  : Factor w/ 2 levels "Female","Male": 2 2 1 2 1 2 1 2 2 2 ...
 $ cp      : num [1:303] 3 2 1 1 0 0 1 1 2 2 ...
 $ trestbps: num [1:303] 145 130 130 120 120 140 140 120 172 150 ...
 $ chol    : num [1:303] 233 250 204 236 354 192 294 263 199 168 ...
 $ fbs     : num [1:303] 1 0 0 0 0 0 0 0 1 0 ...
 $ restecg : num [1:303] 0 1 0 1 1 1 0 1 1 1 ...
 $ thalach : num [1:303] 150 187 172 178 163 148 153 173 162 174 ...
 $ exang   : num [1:303] 0 0 0 0 1 0 0 0 0 0 ...
 $ oldpeak : num [1:303] 2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
 $ slope   : num [1:303] 0 0 2 2 2 1 1 2 2 2 ...
 $ ca      : num [1:303] 0 0 0 0 0 0 0 0 0 0 ...
 $ thal    : num [1:303] 1 2 2 2 2 1 2 3 3 2 ...
 $ target  : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
 - attr(*, "spec")=
  .. cols(
  ..    age = col_double(),
  ..    sex = col_double(),
  ..    cp = col_double(),
  ..    trestbps = col_double(),
  ..    chol = col_double(),
  ..    fbs = col_double(),
  ..    restecg = col_double(),
  ..    thalach = col_double(),
  ..    exang = col_double(),
  ..    oldpeak = col_double(),
  ..    slope = col_double(),
  ..    ca = col_double(),
  ..    thal = col_double(),
  ..    target = col_double()
  .. )
```

```
heart_summary1 <- heart %>% group_by(`gender`) %>% summarise(Min = min
(age,na.rm = TRUE),

                        Q1 = quantile(age,probs = .25,na.rm = TRUE),

                        Median = median(age, na.rm = TRUE),

                        Q3 = quantile(age,probs = .75,na.rm = TRUE),

                        Max = max(age,na.rm = TRUE),

                        Mean = mean(age, na.rm = TRUE),
SD = sd(age, na.rm = TRUE),                        n =
n(),

                        Missing = sum(is.na(target)))
```

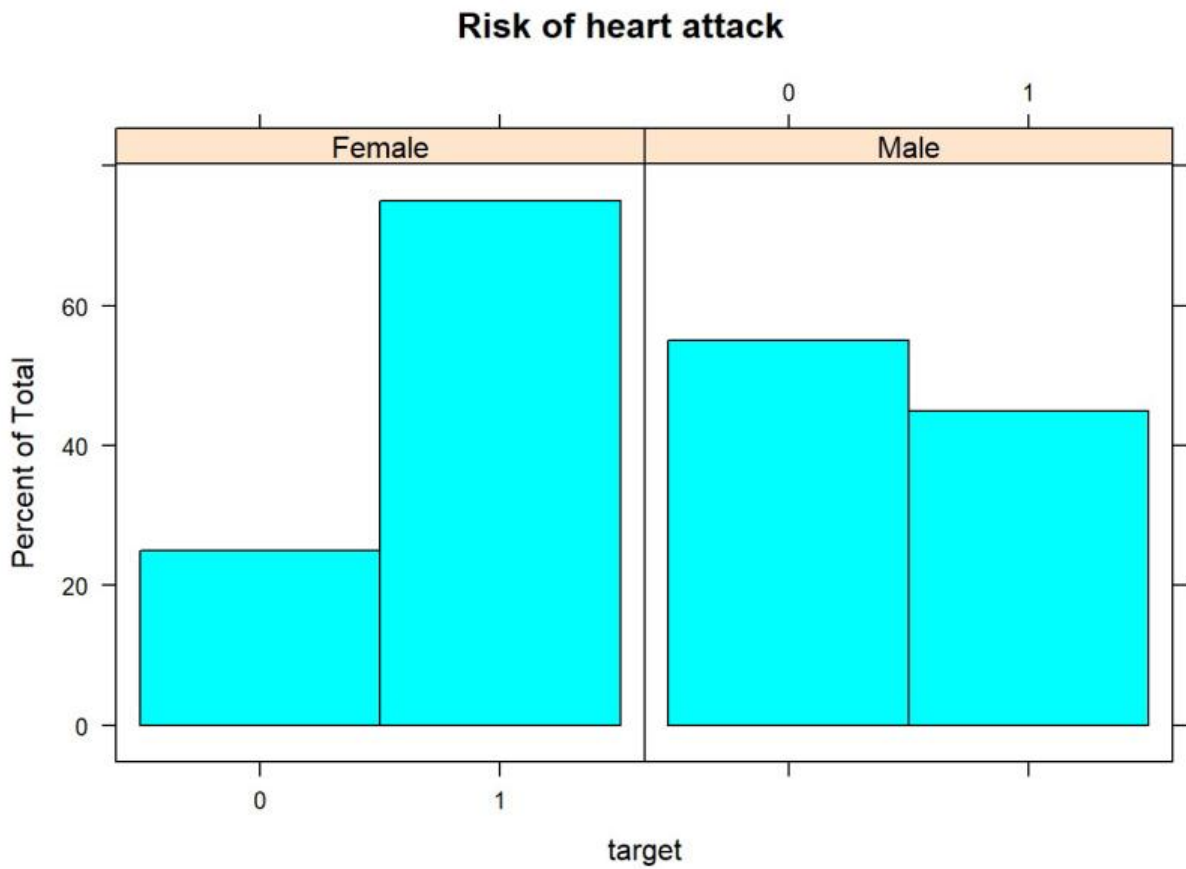`summarise()` ungrouping output (override with `.groups` argument)

```
heart_summary1
```

gender	Min	Q1	Median	Q3	Max	Mean	SD	n	Missing
<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
Female	34	49.75	57	63.0	76	55.67708	9.409396	96	0
Male	29	47.00	54	59.5	77	53.75845	8.883803	207	0

2 rows

DESCRIPTIVE STATISTICS AND VISUALISATION

```
heart %>% histogram(~target | gender, data= ., main = "Risk of heart at
tack")
```

Risk of heart attack


```
heart %>% histogram(~age | gender, data= ., main = "Age of observations", bre
aks=10)


heart %>% ggplot(aes(x=age)) + geom_histogram(aes(y=..density..), colour="bla
ck")+          geom_density(alpha=.2, fill="dodgerblue3")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
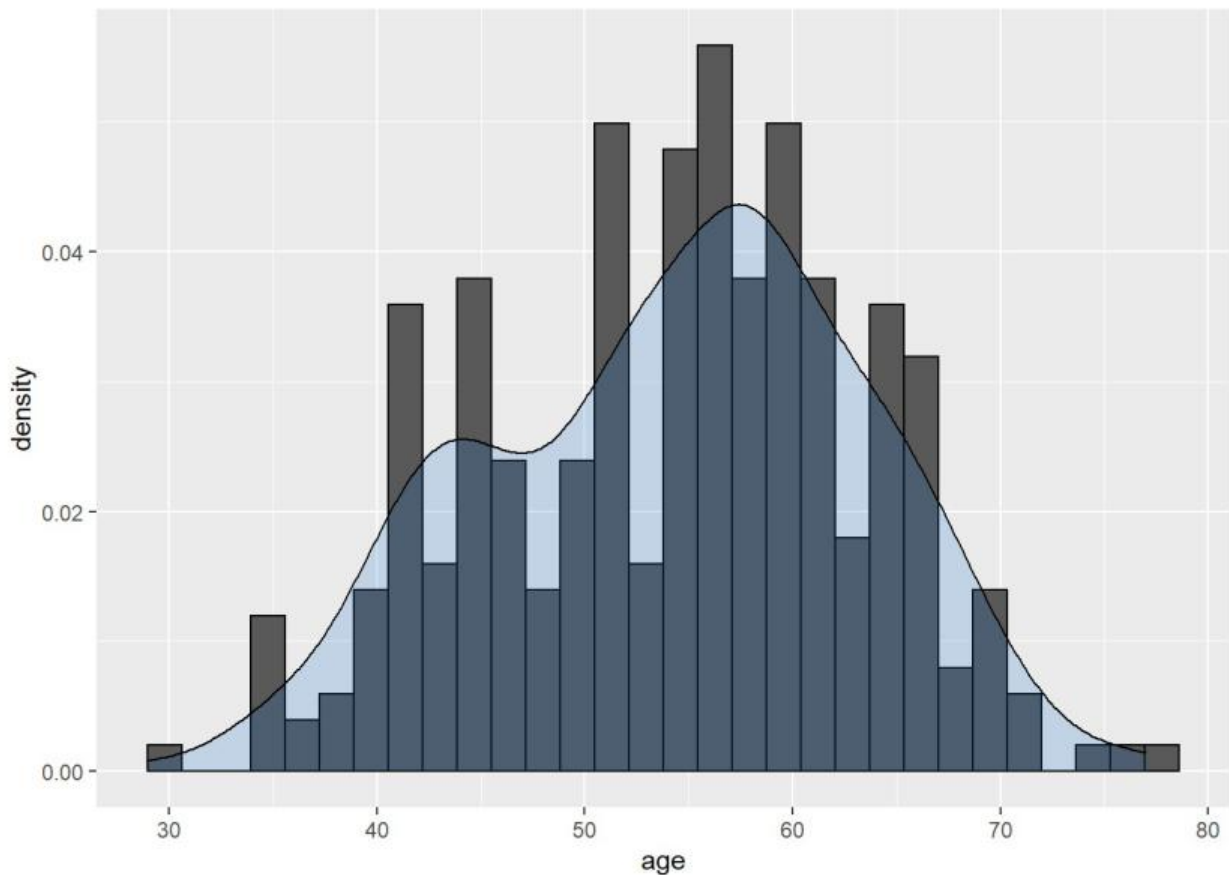

```
heart_filtered <- heart %>% filter(target == 1) table_age <-
table(heart_filtered$gender, heart_filtered$age)
 barplot(table_age, main="Bar plot of Age",
ylab="Number of observations", xlab="Age",
ylim=c(0,8),legend=row.names(table_age), beside=TRUE,
args.legend=c(x="topright",horiz=FALSE,title="Gender"),
col=c( "#FF66FF","#0099FF"))
```

Bar plot of Age
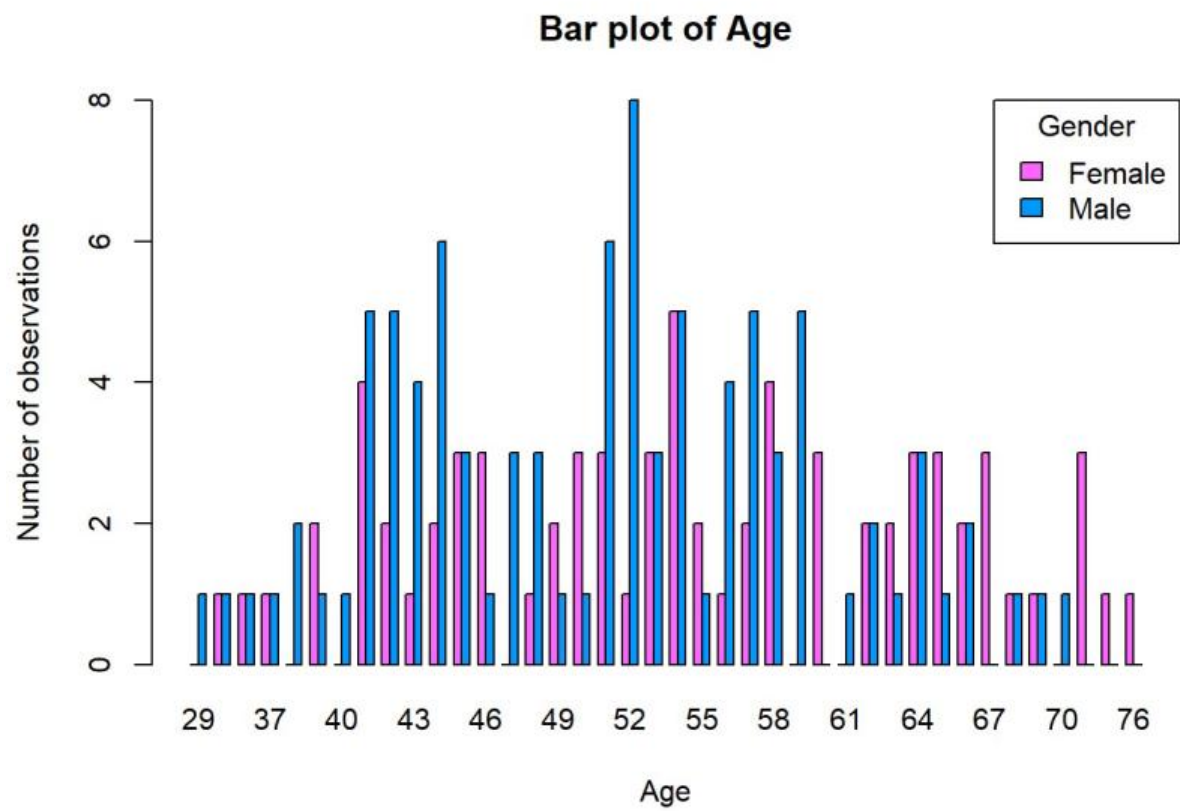

```
boxplot(age ~ gender, data=heart_filtered)
```
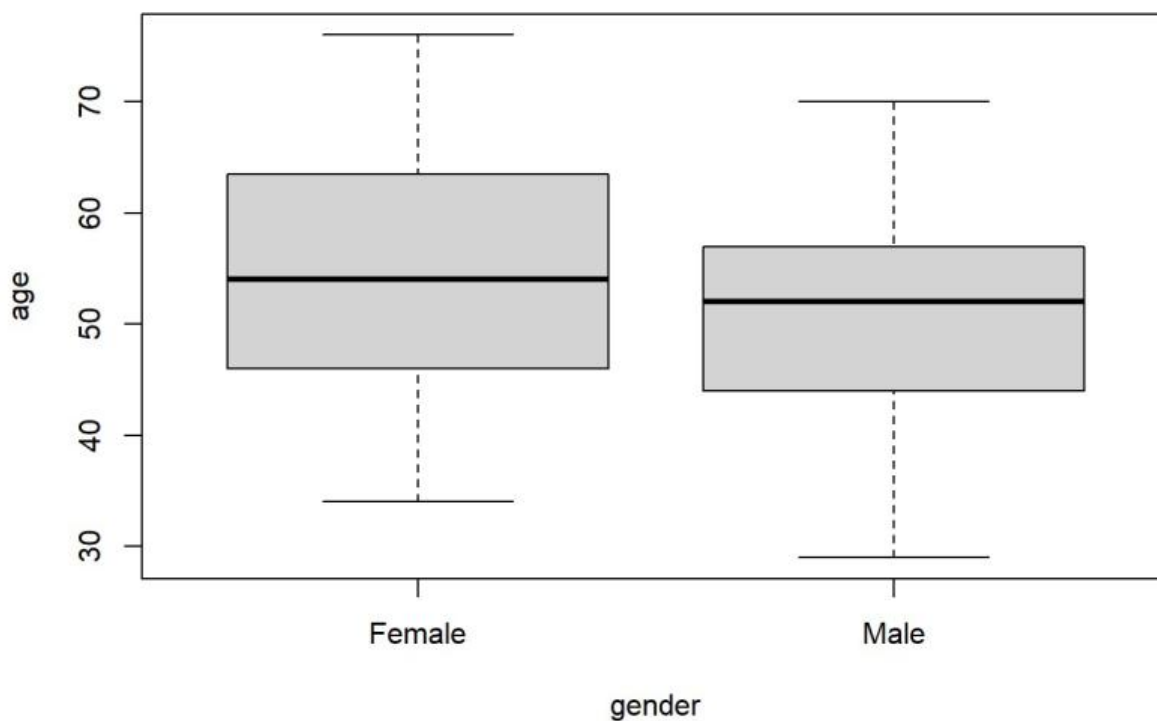
```
heart_summary2 <- heart_filtered %>% group_by(gender) %>% summarise(Mean = ro
und(mean(age, na.rm = TRUE),2),

                                                Min = min(age,na.rm = TRUE)
,

                                                Q1 = quantile(age,probs = .
25,na.rm = TRUE),

                                                Median = median(age, na.rm
= TRUE),

                                                Q3 = quantile(age,probs = .
75,na.rm = TRUE),

                                                Max= max(age,na.rm=TRUE),
n = n())
```

`summarise()` ungrouping output (override with `.groups` argument)

```
heart_summary2
```

gender	Mean	Min	Q1	Median	Q3	Max	n
<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
Female	54.56	34	46	54	63.25	76	72
Male	50.90	29	44	52	57.00	70	93

2 rows

```
heart_summary3 <- heart_filtered %>% group_by(gender) %>% summarise(Mean = ro
und(mean(age, na.rm = TRUE),2),
                                                    SD = round(sd(age, na.rm =
TRUE),3),
                                                        n = n(),
                                                    tcrit = round(qt(p = 0.975,
df = n - 1),3),
                                                    SE = round(SD/sqrt(n),3),
                                                    `95% CI Lower Bound` = roun
d(Mean - tcrit * SE,2),
                                                    `95% CI Upper Bound` = roun
d(Mean + tcrit * SE,2))
```

`summarise()` ungrouping output (override with `.groups` argument)

```
heart_summary3
```

gender	Mean	SD	n	tcrit	SE	95% CI Lower Bound	95% CI Upper Bound
<fctr>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
Female	54.56	10.265	72	1.994	1.21	52.15	56.97
Male	50.90	8.683	93	1.986	0.90	49.11	52.69

2 rows

```
table_heart <- table(heart$target, heart$gender) table_heart
```

```
    Female Male
  0     24  114
  1     72   93
```

```
table_heart %>% addmargins()
```

10

```
      Female Male Sum
0         24  114 138
1         72   93 165
Sum       96  207 303
```

```
table_heart2 <- table_heart %>% prop.table(margin=2) table_heart2
```

```
      Female      Male
0 0.2500000 0.5507246
1 0.7500000 0.4492754
```

```
barplot(table_heart2, main="Bar plot For Male and Female",
ylab="Proportion within Gender", xlab="Likelihood of heart attack",
ylim=c(0,1),legend=row.names(table_heart2), beside=TRUE,
args.legend=c(x="topleft",horiz=FALSE,title="Likelihood of heart attack"))
```

Bar plot For Male and Female
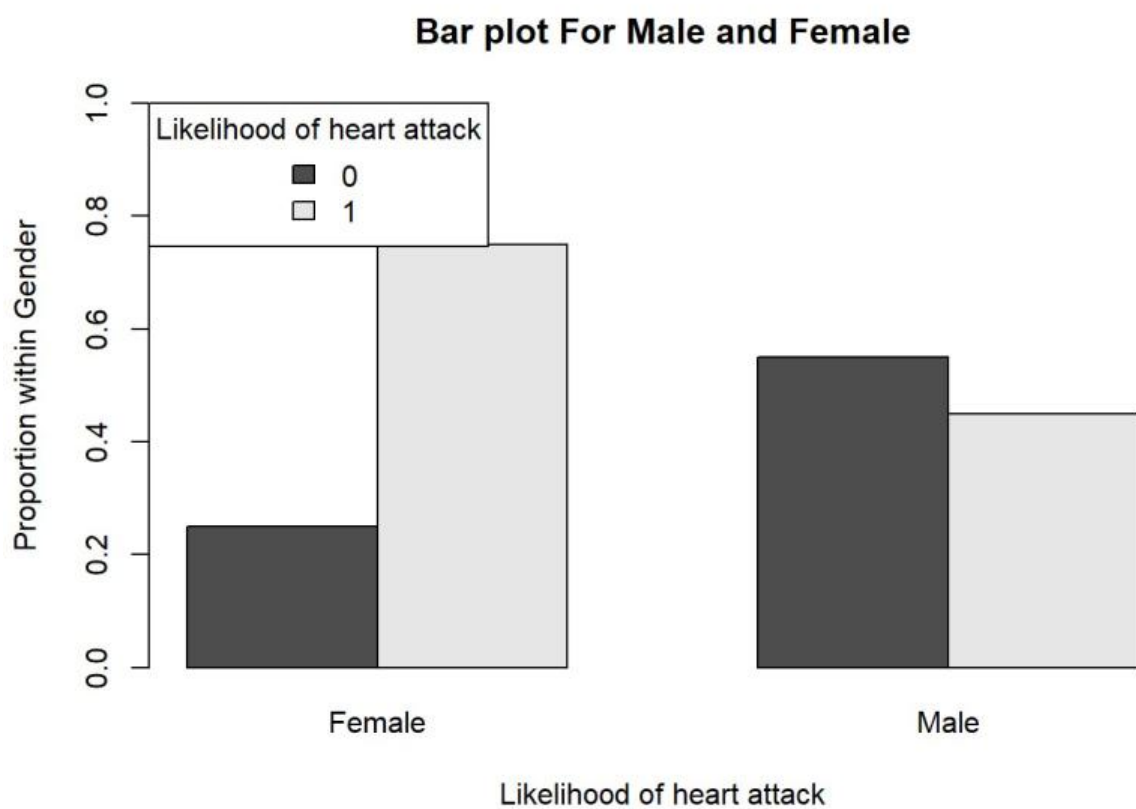
```
chi_heart <- chisq.test(table_heart, p=c(0.5,0.5)) chi_heart
```

Pearson's Chi-squared test with Yates' continuity correction

data: table_heart
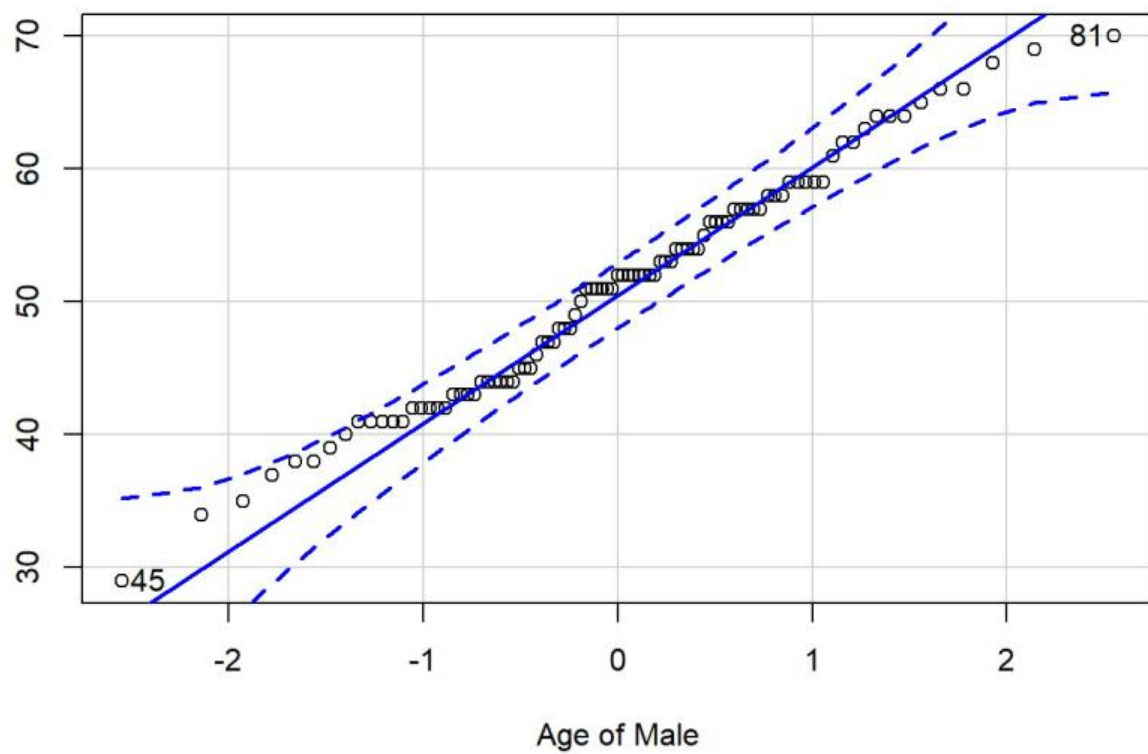X-squared = 22.717, df = 1, p-value = 1.877e-06

```
chi_heart$expected
```

```
       Female      Male
0 43.72277  94.27723
1 52.27723 112.72277
```

```
chi_heart$observed
```

```
     Female Male
0       24  114
1       72   93
```
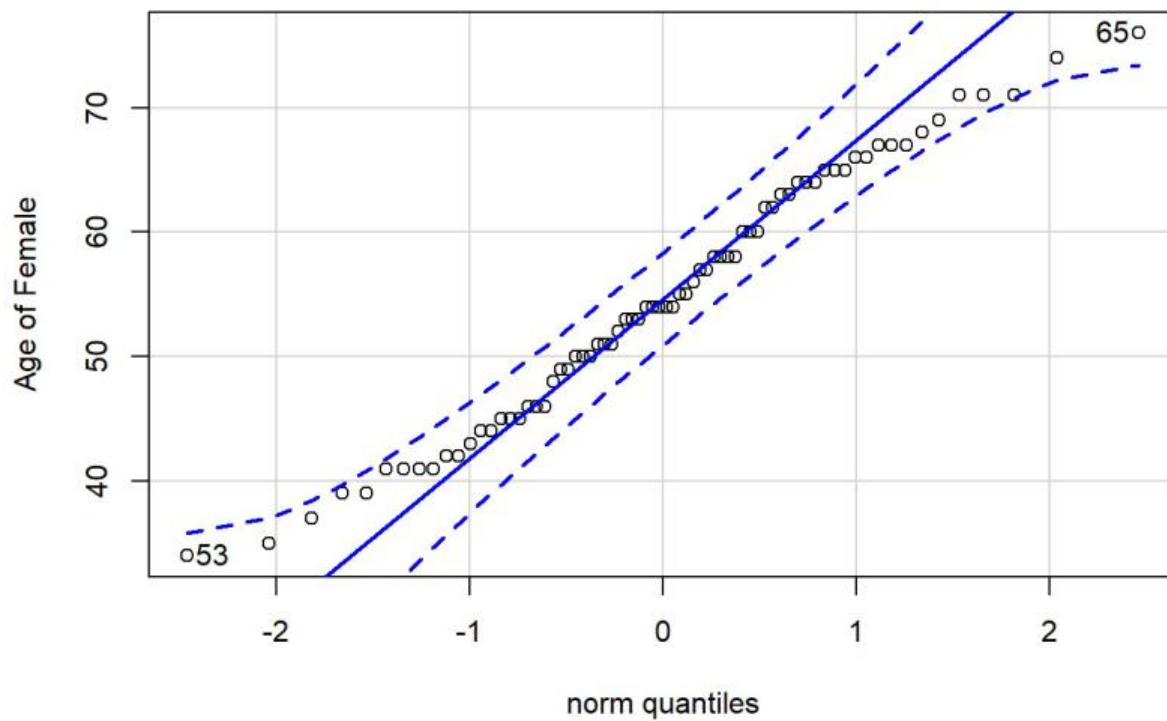
```
heart_filtered <- heart %>% filter(target == 1)
```

```
heart_male <- heart_filtered %>% filter(gender=="Male")  heart_male$age
%>% qqPlot(dist="norm", xlab = "Age of Male")
```

Age of Male

[1] 45 81

```
heart_female <- heart_filtered %>% filter(gender=="Female") heart_female$age
%>% qqPlot(dist="norm", ylab = "Age of Female")
```

```
[1] 65 53
```

```
leveneTest(age ~ gender, data = heart)
```

	Df <int>	F value <dbl>	Pr(>F) <dbl>
group	1	0.3630374	0.5472778
	301	NA	NA

2 rows

Student T Test

```
result <- t.test(age ~ gender, data=heart_filtered,
var.equal=TRUE, alternative ="two.sided")

result
```

```
   Two Sample t-test

data:  age by gender
t = 2.4739, df = 163, p-value = 0.01439
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.7370777 6.5675818
sample estimates:
mean in group Female    mean in group Male
          54.55556               50.90323
```

```
result$conf.int
```

```
[1] 0.7370777 6.5675818
attr(,"conf.level")
[1] 0.95
```

```
result$p.value
```

```
[1] 0.01439017
```