Bioinformatics III

Seventh Assignment

Alexander Flohr   (2549738)
Andrea Kupitz   (2550260)

June 6, 2018

Exercise 7.1: Missing Data Imputation

Listing 1: Listing of source code

(a)
```r
# Read in data table
DFrame <- read.table(file = "ms_toy.txt", sep = "\t", header = TRUE)
DTable0 <- t(as.matrix(as.data.frame(lapply(DFrame[,1:1], as.numeric))))

# create new dataset without missing data
newdata <- na.omit(DFrame)
DTable <- t(as.matrix(as.data.frame(lapply(newdata[,1:1], as.numeric))))
colnames(DTable) <- colnames(DFrame)[1]

# Calculate the mean and standard deviation of the current data
mean <- mean(DTable)
sd <- sd(DTable)

# Derive the new mean and standard deviation for the missing data based on the current
# distribution
nansize <- dim(DTable0)[1]-dim(DTable)[1]
lowquan <- qnorm(0.25, mean, sd)
lowdata <- subset(DTable, DTable[,1] <= lowquan)
newmean <- mean(lowdata)
newsd <- sd(lowdata)

# Generate the new data based on the new mean and standard deviation from the previous step
imputdata <- rnorm(nansize, newmean, newsd)

hist1 <- hist(DTable, col="blue")
hist2 <- hist(imputdata, add=T, col="red")
#plot()
for(mean in seq(from = 20, to = lowquan, by = 2)){
  for(sd in seq(from = 1, to = sd, by = 0.2)){
    imputdata1 <- rnorm(nansize, mean, sd)
    hist3 <- hist(DTable, col="blue")
    hist4 <- hist(imputdata1, add=T, col="red")
  }
}
```

Listing 1 shows the source code of our imputation of data. Figure 1 shows the imputed data.

(b) The greater the standard deviation gets, the wider gets the histogram of the imputated data. Moreover for greater standard deviations the values tend to lay more far away from each other and thus the single bars of the histogram are lower.
The closer the new mean moves to the original mean, the more moves the distribution of the imputated data to the center of the original data.
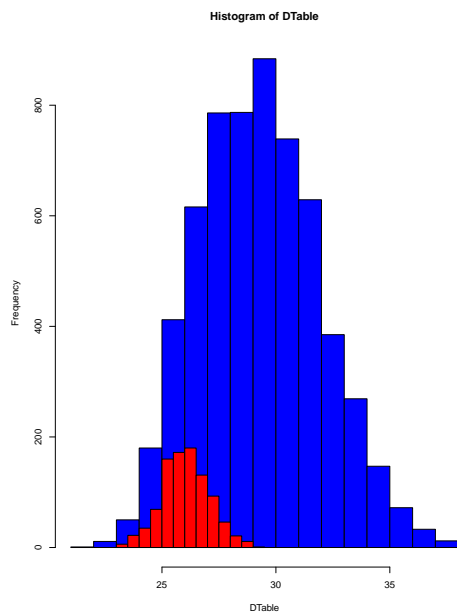Figures 2, 3, 4 show the imputated data with increasing mean.

Figure 1: Distribution of the sample with the imputed data

Figures 5, 6, 7 show the imputated data with increasing standard deviation.

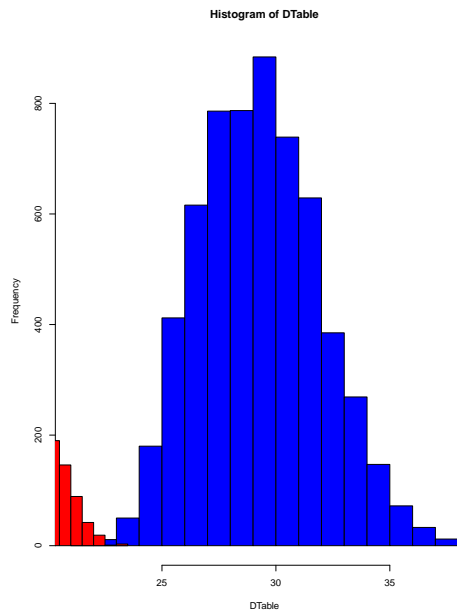## Exercise 7.2: DREAM challenge

Figure 2: Distribution of the sample with the imputed data
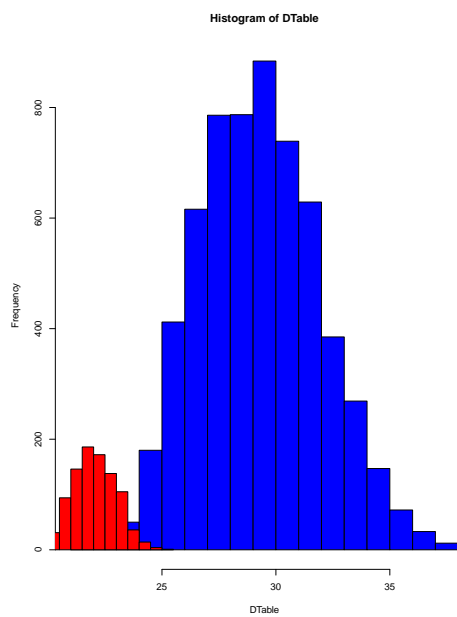

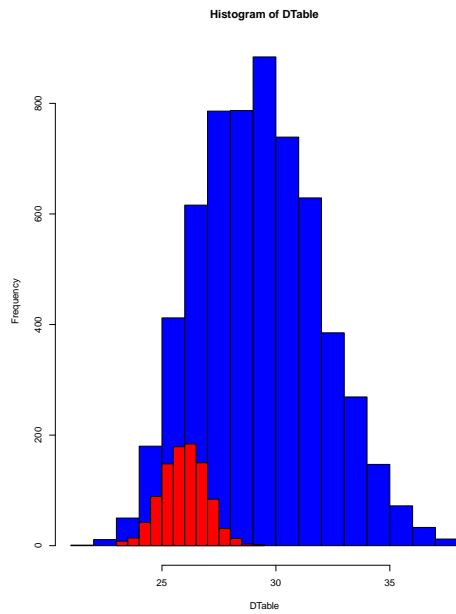Figure 3: Distribution of the sample with the imputed data

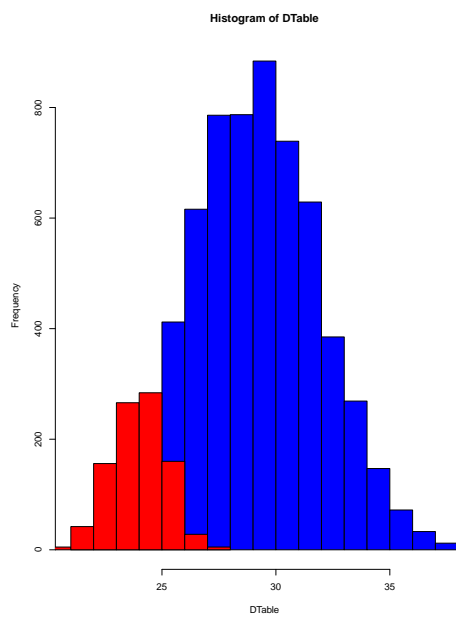Figure 4: Distribution of the sample with the imputed data

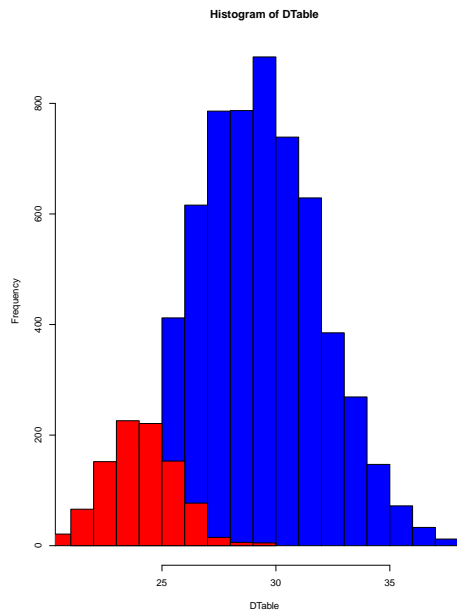Figure 5: Distribution of the sample with the imputed data
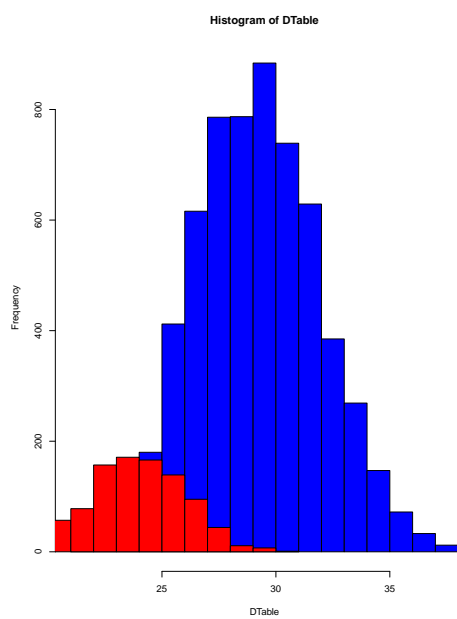
Figure 6: Distribution of the sample with the imputed data

Figure 7: Distribution of the sample with the imputed data