02441 Applied Statistics and Statistical Software

Exercise 4D - KFM

The dataset kfm contains measurements of the newborn babies, their mother and milk consumption

Variable name	Description
dl.milk	amount of breast milk (dl)
sex	gender of body
weight	baby weight (kg)
ml.suppl	amount of milk supplement (ml)
mat.weight	mothers weight (kg)
mat.height	mothers height (cm)

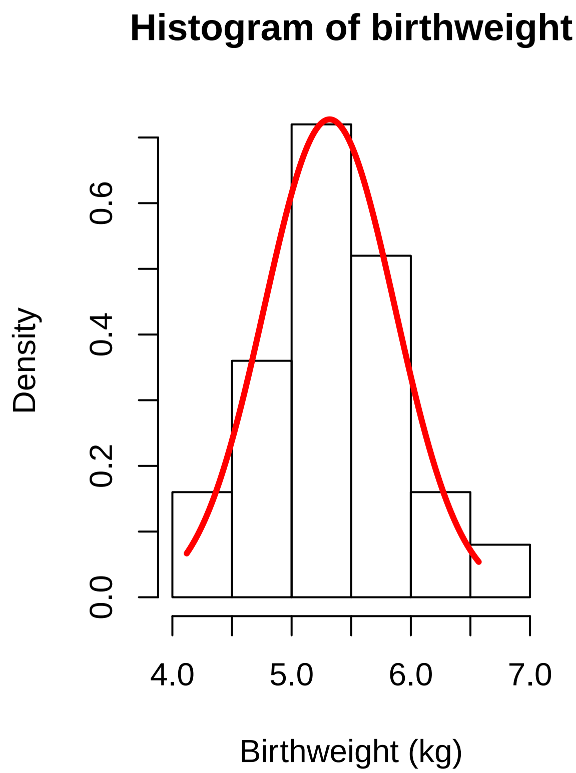
1. Make apropriate plots of the birthweight in order to check wether the weight is normally distributed

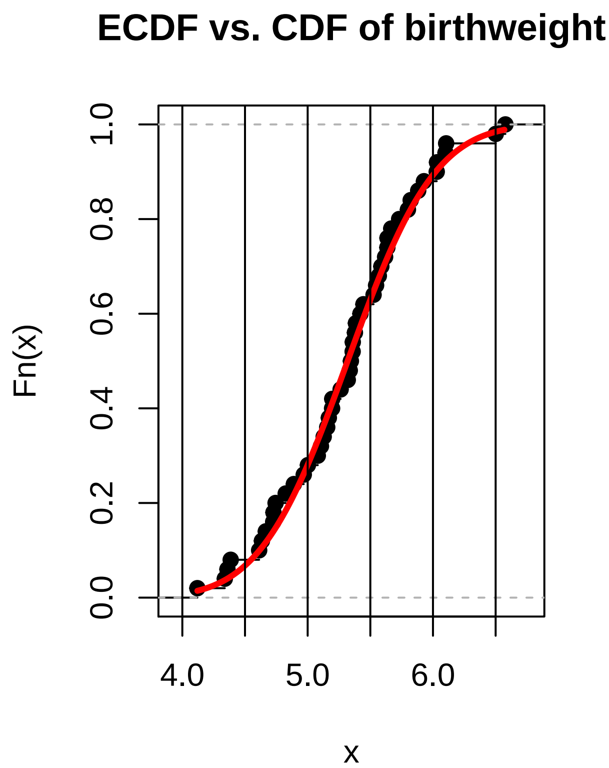
```
# Read data
df <- read.table("kfm.txt", header=TRUE, row.names=1)
df <- df[,-1]
head(df)
```

```
##   dl.milk sex weight ml.suppl mat.weight mat.height
## 1    8.42 boy  5.002      250         65        173
## 2    8.44 boy  5.128        0         48        158
## 3    8.41 boy  5.445       40         62        160
## 4    9.65 boy  5.106       60         55        162
## 5    6.44 boy  5.196      240         58        170
## 6    6.29 boy  5.526        0         56        153
```

```
# Visual comparison of histogram with normal pdf
par(mfrow=c(1,2))
hist1 <- hist(df$weight, freq = FALSE, main="Histogram of birthweight", xlab="Birthweight (kg)")
x_range <- seq(min(df$weight), max(df$weight), by = 0.01)
lines(x_range, dnorm(x_range, mean(df$weight), sd(df$weight)), lw = 3, col = "red")
color1 <- rgb(1,0,0, alpha = 0.5)
# polygon(c(seq(3.5,4,0.05),4,3.5), c(dnorm(seq(3.5,4,0.05), mean(x), sd(x)),0,0), col = color1, border

# Visual comparison of ecdf with normal cdf
plot(ecdf(df$weight), main="ECDF vs. CDF of birthweight")
lines(x_range, pnorm(x_range, mean(df$weight), sd(df$weight)), lw = 3, col = "red")
abline(v=hist1$breaks)
```





From the visual inspection it seems that birthweight is normally distributed.

2. Make a χ^2-test to test if the birthweights for the babies can be assumed normally distributed

```r
shapiro.test(df$weight)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$weight
## W = 0.98976, p-value = 0.9405
```

```r
# Get expected counts for all histogram bins
breaks <- hist1$breaks
prob <- pnorm(c(-Inf, breaks[-c(1, length(breaks))], Inf), mean(df$weight), sd(df$weight))
prob <- diff(prob)
expected <- length(df$weight)*prob

# Chi Square test for goodness-of-fit
counts <- hist1$counts
chi_obs <- sum((counts - expected)^2 / expected)
p_value <- 1 - pchisq(chi_obs, df = length(hist1$mids) - 3)
```

Birthweight is normally distributed since p-valye is greather than α, i.e. we fail to reject H_0 since 0.5005768 $>$ 0.05.