

## exercise2

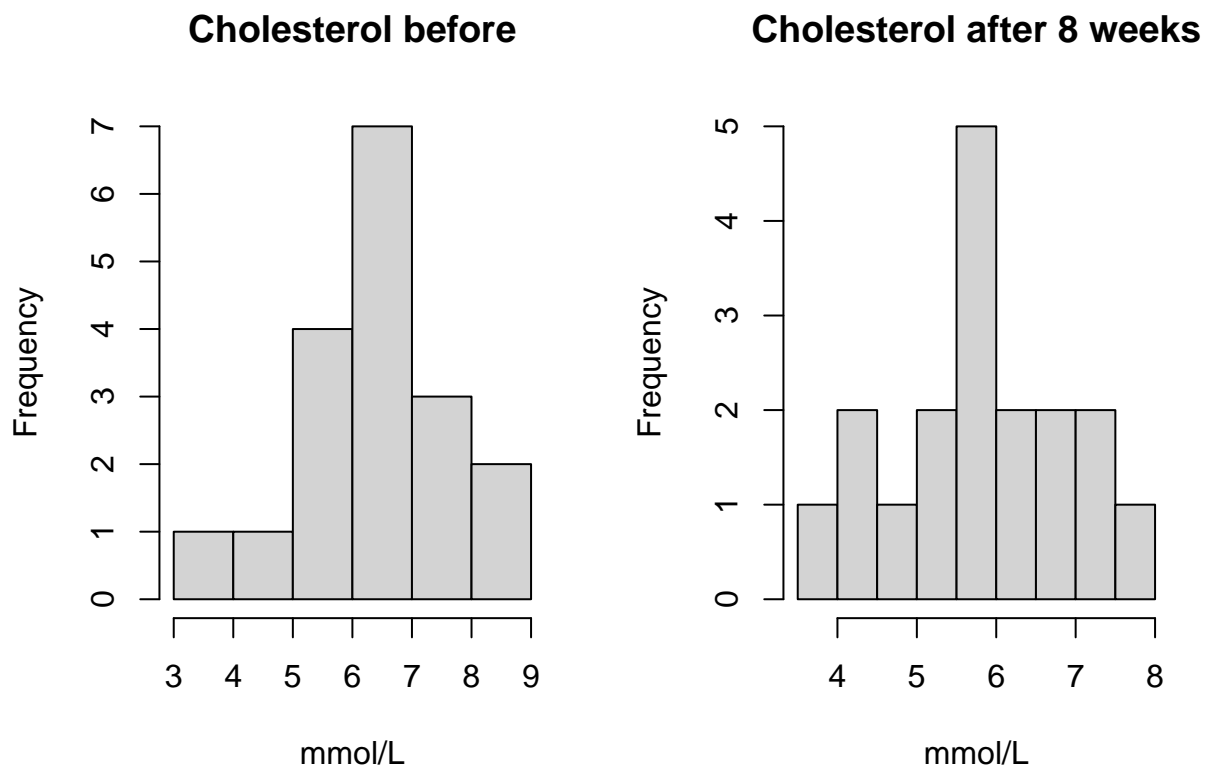
2023-02-20

### R Markdown

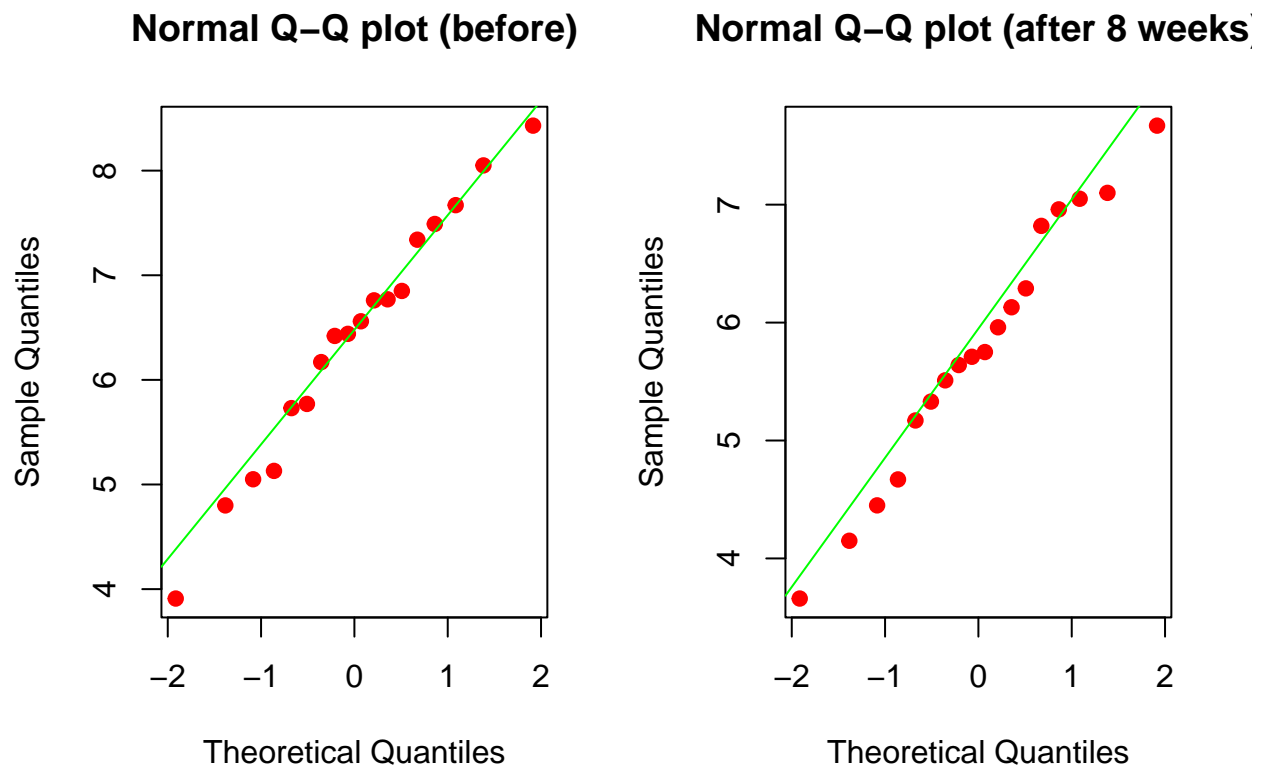
Exercise 2:

- a) Histogram plots for cholesterol samples before low fat diet and after 8 weeks of low fat diet imply that data is normally distributed, since both histograms show symmetrical “bell” shape distribution. Normality is also implied by QQ plots in which a straight diagonal line show that theoretical quantiles of normal distribution match with sample quantiles (true for both samples). However, it should be taken in to the account, when assuming normality, that in both samples we only have 18 observations. Correlation between samples of cholesterol levels before, and after 8 week diet was computed to be 0.991. Since cholesterol levels were measured for the same sample of people before and after low fat diet, it can be expected that data will be highly correlated.

```
df = read.csv('Data/cholesterol.txt', header = TRUE, sep = "")
par(mfrow=c(1,2))
hist(df$Before, main = 'Cholesterol before', xlab = 'mmol/L')
hist(df$After8weeks, main = 'Cholesterol after 8 weeks', xlab = 'mmol/L')
```

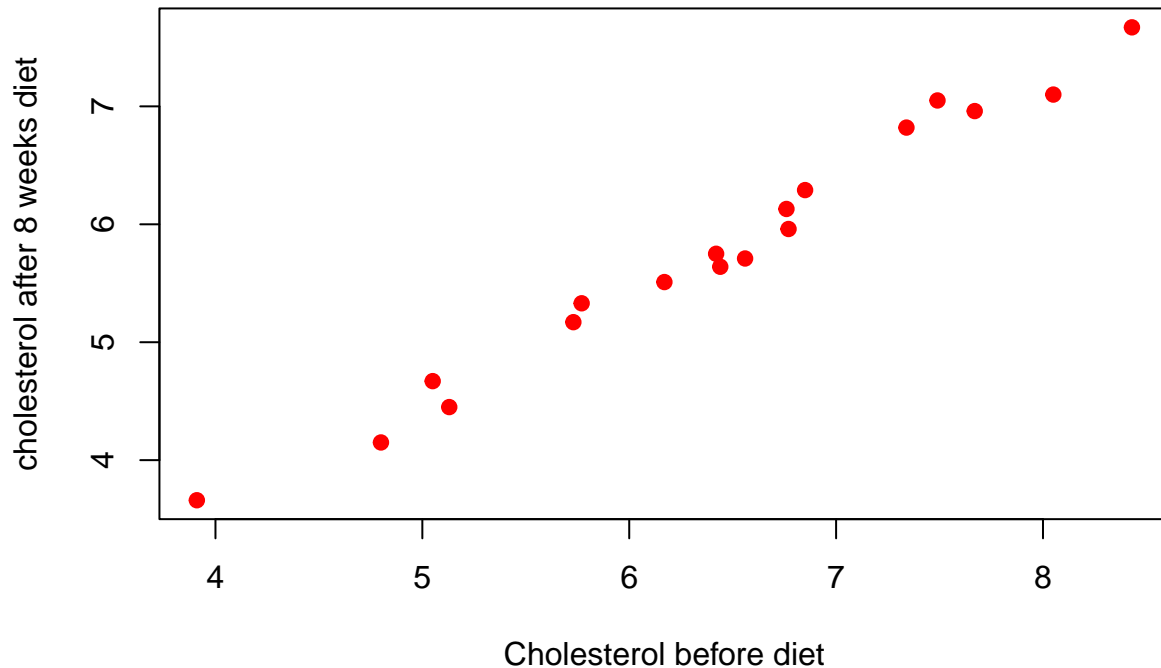


```
par(mfrow=c(1,2))
qqnorm(df$Before, col='red', pch=19, main="Normal Q-Q plot (before)")
qqline(df$Before, col='green')
qqnorm(df$After8weeks, col='red', pch=19, main="Normal Q-Q plot (after 8 weeks)")
qqline(df$After8weeks, col='green')
```



```
plot(df$Before, df$After8weeks, xlab = "Cholesterol before diet",
     ylab = "cholesterol after 8 weeks diet",
     main = paste("Scatter Plot (correlation = ", round(cor(df$Before,df$After8weeks), 3),
                  "collapses="), sep=""), col='red', pch = 19)
```

### Scatter Plot (correlation = 0.991)



- b) To verify that low fat diet is effective in lowering cholesterol levels, paired t-test and paired Wilcoxon signed rank test were constructed, where  $H_0 : \mu_{before} \leq \mu_{after8weeks}$  and  $H_1 : \mu_{before} > \mu_{after8weeks}$ . T-test provided us with p-value equal to 0.000 which allowed us to reject  $H_0$ , therefore we can conclude that low fat diet is indeed effective in lowering cholesterol levels. The Wilcoxon signed rank test having the same hypothesis resulted in p-value also equal to 0.000, which also allows us to confirm alternative hypothesis that  $\mu_{before} > \mu_{after8weeks}$  is true. Our motives for choosing t-test and Wilcoxon signed rank test come from our data properties. The data set cholesterol features two-paired samples, in which experimental units (18 people) have two numerical outcomes (cholesterol levels (mmol/L)) - before treatment (diet) and after it. Also, it must be mentioned that both samples imply to be normally distributed (see Q-Q plots above). Therefore, two-paired nature of data and normality allows us to conduct paired t-test, and symmetry of data allows us computing Wilcoxon signed rank test.

Permutation test can also be applied in this case since we have a setting of two normally distributed paired samples.

```
ttest = t.test(df$Before, df$After8weeks, alt='g', paired=TRUE)
print(paste("p-value of two-paired t-test: ",round(ttest$p.value,3)))
```

```
## [1] "p-value of two-paired t-test:  0"
```

```
wilcox_test = wilcox.test(df$Before, df$After8weeks, alt='g',paired = TRUE)
print(paste("p-value of two-paired Wilcoxon signed rank test: ",round(wilcox_test$p.value,3)))
```

```
## [1] "p-value of two-paired Wilcoxon signed rank test:  0"
```

- c) Assuming that  $X_1, \dots, X_{18} \sim Unif[3, \theta]$ , where  $X_1, \dots, X_{18}$  is random variable from column *after8weeks*, we applied central limit theorem by drawing 18 samples with replacement from column *after8weeks* and calculating max cholesterol level in drawn sample, this step is repeated 1000 times to collect a set of maximum values. By computing mean for aforementioned maximum values set we estimate that  $\hat{\theta} = 7.43$ . Our computed 95% confidence interval - [6.96, 7.67].

```
sample_maxs <- c()
n = 1000
for (i in 1:n){
  sample_maxs[i] = max(sample(df$After8weeks, nrow(df), replace=TRUE))
}
estimated_upper_limit = mean(sample_maxs)
cat("Estimated Theta:", estimated_upper_limit, "\n")
```

```
## Estimated Theta: 7.47
```

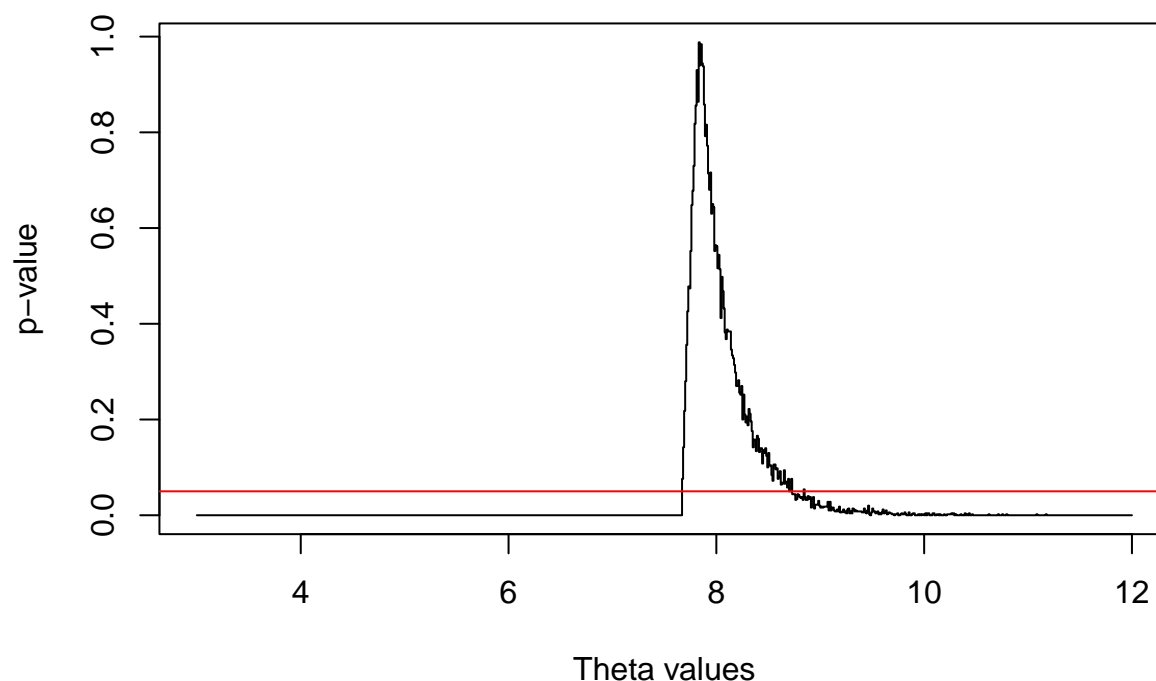
```
cat("Confidence interval:", quantile(sample_maxs, probs=c(0.025,0.975 )))
```

```
## Confidence interval: 6.96 7.67
```

d)

```
theta = 3.00; t=max(df$After8weeks); counter = 1; B=1000;
tstar=numeric(B); p_values = c(); thetas = c();
while (theta <= 12) {
  for (i in 1:B){
    xstar = runif(n=nrow(df), min=3, max=theta)
    tstar[i]=max(xstar)
  }
  p_left=sum(tstar<t)/B; p_right=sum(tstar>t)/B;
  p_values[counter]= 2*min(p_left,p_right)
  thetas[counter] = theta
  counter = counter + 1
  theta = theta + 0.01 #increment theta by 0.01
}
plot(x=thetas, y = p_values, type = "S", xlab = "Theta values", ylab="p-value", main = "Theta distribut")
abline(a=0.05,b=0, col='red')
```

## Theta distribution according to p-values



e)

```
binom.test(sum(df$After8weeks < 6), nrow(df), alt='l')
```

```
##
## Exact binomial test
##
## data: sum(df$After8weeks < 6) and nrow(df)
## number of successes = 11, number of trials = 18, p-value = 0.9
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.000 0.801
## sample estimates:
## probability of success
##           0.611
```

```
binom.test(sum(df$After8weeks < 4.5), nrow(df), alt='l', p=0.25)
```

```
##
## Exact binomial test
##
## data: sum(df$After8weeks < 4.5) and nrow(df)
## number of successes = 3, number of trials = 18, p-value = 0.3
## alternative hypothesis: true probability of success is less than 0.25
## 95 percent confidence interval:
##  0.000 0.377
```

```
## sample estimates:  
## probability of success  
##           0.167
```