Assignment 1

Group 15: Ivona Bîrlad, Kaiyi Wang, Matteo Rapa

9/14/2022

Exercise 1.1 Birthweight

```
#load data
raw_data = read.table(file = "birthweight.txt", header = TRUE)
#check if the data is correctly loaded
head(raw_data)
```

```
##   birthweight
## 1        1538
## 2        2617
## 3        2691
## 4        2401
## 5        3596
## 6        3153
```

```
birthweight = raw_data[, 1]
```

a) Normality Check and μ

```
# histogram
par(mfrow= c(1,2))
hist(birthweight, main = "Histogram of Birth Weights", xlab = "Birth weight",  col="gray88", bo
# QQ Plot
qqnorm(birthweight, main = "Q-Q Plot of Birth Weights")
qqline(birthweight, col = "tomato", lwd = 2)
```

Comments: From data visualization, the data is normally distributed

```
# shapiro test for the normality of the data
shapiro.test(birthweight)
```

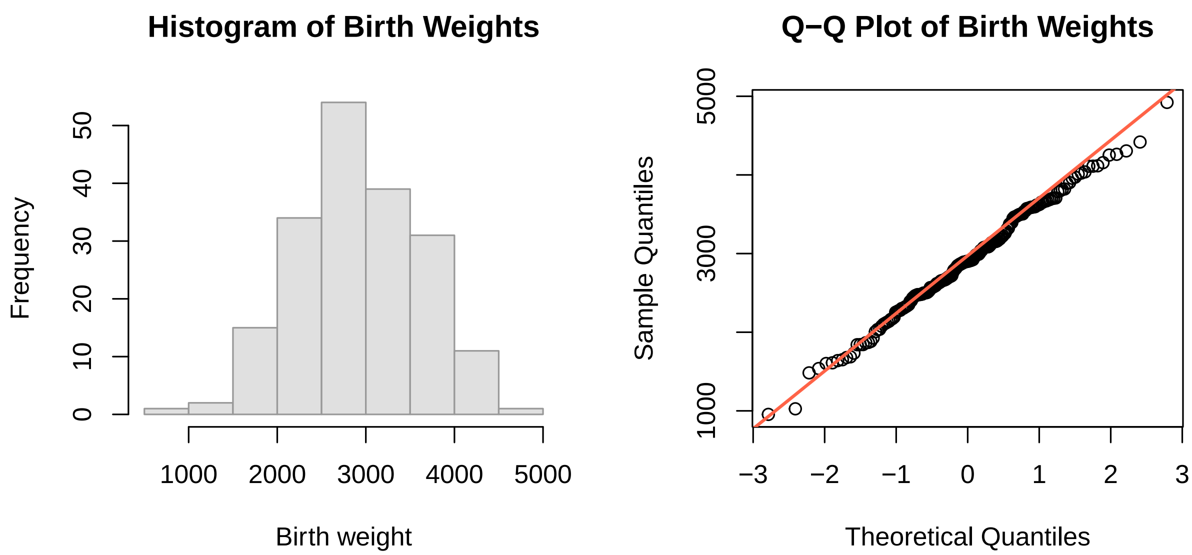


Figure 1: Normality Visualisation

```
##
##  Shapiro-Wilk normality test
##
## data:  birthweight
## W = 0.99595, p-value = 0.8995
```

Comments: Reject H0, the data is normally distributed

```
# the sample mean is used for point estimate for the data
bw_m = mean(birthweight); bw_m
```

```
## [1] 2913.293
```

Comments:

b) Confidence interval for μ

```
bw_n = length(birthweight)
bw_s = sd(birthweight)
# alpha/2 = 0.05, so 1 - alpha = 0.9
bw_t = qt(0.95, df = bw_n-1)
c(bw_m - bw_t*bw_s/sqrt(bw_n), bw_m + bw_t*bw_s/sqrt(bw_n))
```

```
## [1] 2829.202 2997.384
```

Comments: The 90% CI of μ is: [2829.20, 2997.38]

c) one sample t-test

```
# H0
# H1
# One sample t-test
t.test(birthweight, mu = 2800, alt = "g")
```

```
##
##  One Sample t-test
##
## data:  birthweight
## t = 2.2271, df = 187, p-value = 0.01357
## alternative hypothesis: true mean is greater than 2800
## 95 percent confidence interval:
##  2829.202      Inf
## sample estimates:
## mean of x
##  2913.293
```

Comments: p-value is smaller than 0.05, here we reject H0.

d) further discussion

Comments: the R-output of the test from b), indicates that we are 90% confident that the interval from 2829.20 to 2997.38 actually does contain the true value of the population mean μ. This CI is double side. In c), a single side test is conducted for the claim that the mean birthweight is bigger than 2800. To clarify this claim, we compute the alt = 'great' in the t-test. . . . .

Exercise 1.2 Kinderopvangtoeslag

a) estimate for p

```
# point estimate for p
childcare_p = 140 / 200; childcare_p
```

```
## [1] 0.7
```

Comments: the point estimate for p is 0.7

b) CI of p

```
# calculate q
childcare_q = 1 - childcare_p
# sample size
childcare_n = 200
# calculate z alpha/2.
# for 99% CI, we have alpha = 0.01
childcare_z = qnorm(1 - 0.01/2)
# calculate the 99% CI of childcare_p
c(childcare_p - childcare_z * sqrt(childcare_p * childcare_q / childcare_n),
  childcare_p + childcare_z * sqrt(childcare_p * childcare_q / childcare_n))
```

```
## [1] 0.6165336 0.7834664
```

Comments: The 99% CI of p is: [0.617, 0.783]

b) proportion test

```
bi_0.9 = binom.test(140, 200, p = 0.75,conf.level = 0.9); bi_0.9[3]
```

```
## $p.value
## [1] 0.1029764
```

```
bi_0.95 = binom.test(140, 200, p = 0.75,conf.level = 0.95); bi_0.95[3]
```

```
## $p.value
## [1] 0.1029764
```

```
bi_0.99 = binom.test(140, 200, p = 0.75,conf.level = 0.99); bi_0.99[3]
```

```
## $p.value
## [1] 0.1029764
```

```
bi_0.8 = binom.test(140, 200, p = 0.75,conf.level = 0.8); bi_0.8[3]
```

```
## $p.value
## [1] 0.1029764
```

Table 1 XXXX α | 0.8 | 0.9 | 0.95 | 0.90 —— | ———————— | —————————— | ——————————| —————————— P-value | 0.103 | 0.103 | 0.103 | 0.103

Comments: p-value doesn't change by different CI. Explanation?

Exercise 1.3 Weather

```r
#load data
raw_data_weather = read.table(file = "weather.txt", header = TRUE)
#check if the data is correctly loaded
head(raw_data_weather)
```

```
##   humidity temperature
## 1     79.7        26.2
## 2     77.4        13.8
## 3     75.8        56.7
## 4     71.7        36.9
## 5     74.0        21.8
## 6     75.5        73.9
```

```r
# assign variables
humidity = raw_data_weather$humidity
temperature = raw_data_weather$temperature
```

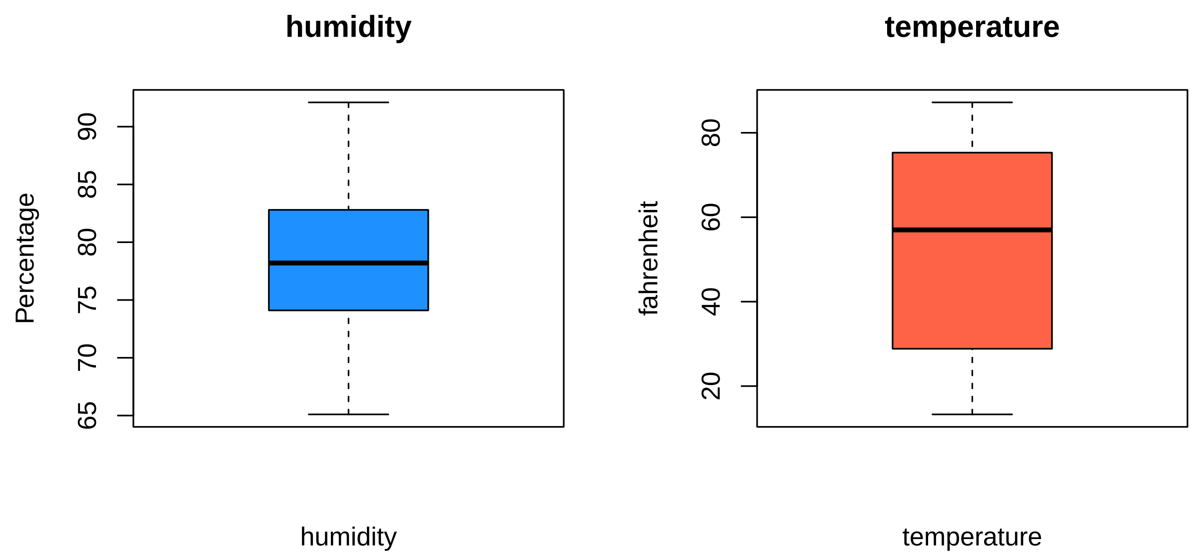
```r
summary(humidity)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   65.10   74.15   78.20   78.34   82.70   92.10
```

```r
summary(temperature)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.30   29.07   57.00   52.73   74.80   87.20
```

```r
par(mfrow= c(1,2))
boxplot(humidity, main = "humidity", xlab = "humidity", ylab = "Percentage", col="dodgerblue")
boxplot(temperature, main = "temperature", xlab = "temperature", ylab = "fahrenheit", col="toma
```




```
# histogram
par(mfrow= c(1,2))
hist(humidity, main = "Histogram of humidity", xlab = "humidity",  col="gray88", border="gray6
# QQ Plot
qqnorm(humidity, main = "Q-Q Plot of humidity")
qqline(humidity, col = "tomato", lwd = 2)
```

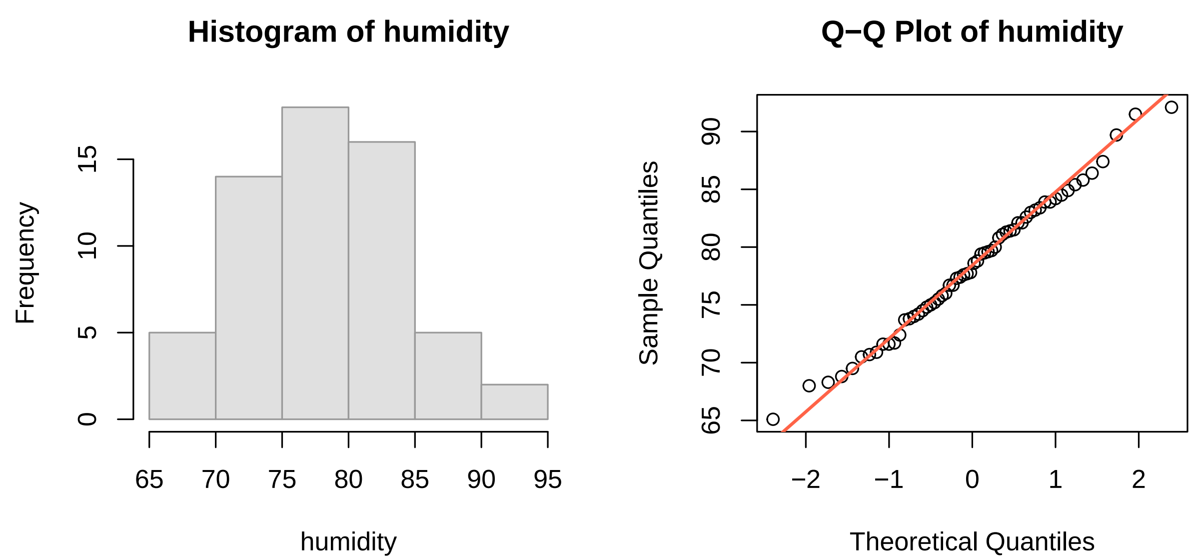


Figure 2: Normality Visualisation

```
# shapiro test for the normality of the data
shapiro.test(humidity)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  humidity
## W = 0.99129, p-value = 0.9464
```

```
# histogram
par(mfrow= c(1,2))
hist(temperature, main = "Histogram of temperature", xlab = "temperature",  col="gray88", borde
# QQ Plot
qqnorm(temperature, main = "Q-Q Plot of temperature")
qqline(temperature, col = "tomato", lwd = 2)
```

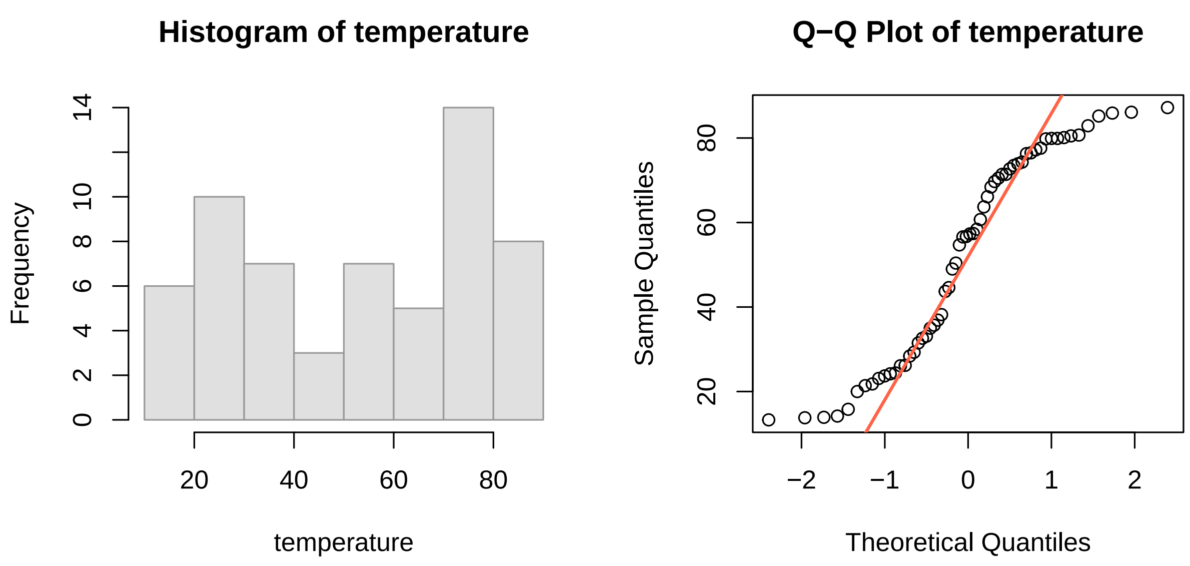


Figure 3: Normality Visualisation

```
# shapiro test for the normality of the data
shapiro.test(temperature)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  temperature
## W = 0.90696, p-value = 0.0002382
```