

## 6. Übung(sserie) Statistik für Ingenieure WiSe 19/20

### 1. Aufgabe:

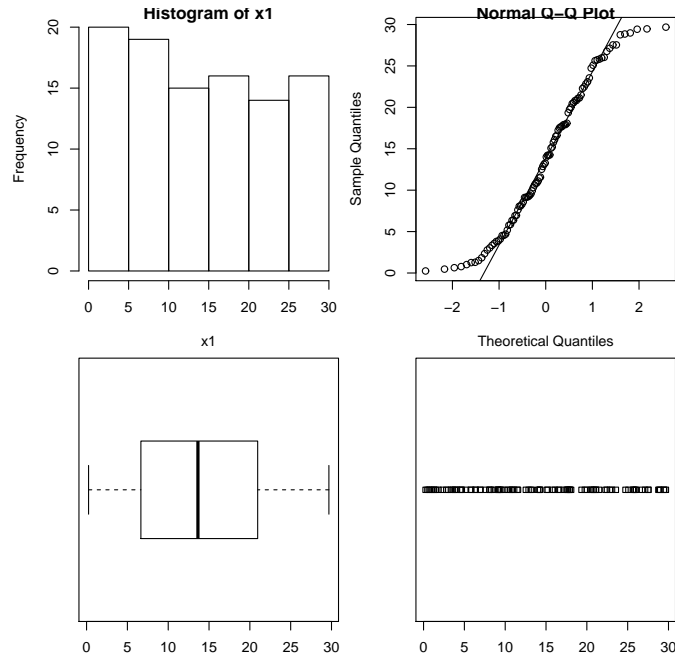
In dieser Aufgabe wurden mit R Stichproben vom Umfang  $n = 100$  erzeugt.

Die vorliegenden Daten entstammen dabei einer a) gleichverteilten, b) exponentialverteilten, c) normalverteilten bzw. d) weibullverteilten Grundgesamtheit, d.h. die vorliegenden Stichproben sind Realisierungen von unabhängig und identisch gleichverteilten, exponentialverteilten, normalverteilten und weibullverteilten Zufallsvariablen.

Was können Sie aus den einzelnen Grafiken ablesen. Bei welchen Grafiken kann man auf das Vorliegen oder Nichtvorliegen einer Normalverteilung und bei welcher auf das Vorliegen einer speziellen Verteilung schließen.

a) Stetige Gleichverteilung

```
>x1=runif(100,0,30)
>par(mfcol=c(2,2))
>hist(x1)
>boxplot(x1,horizontal =T)
>qqnorm(x1); qqline(x1)
>stripchart(x1,method="stack")
```

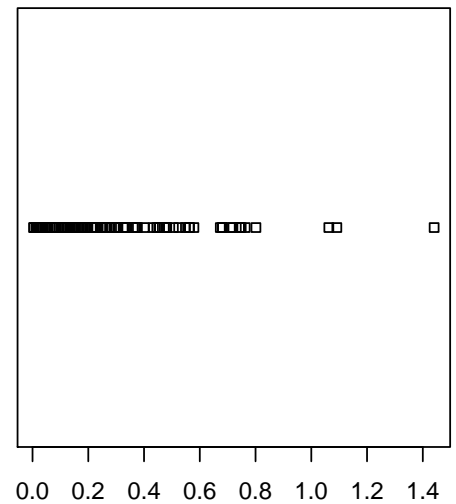
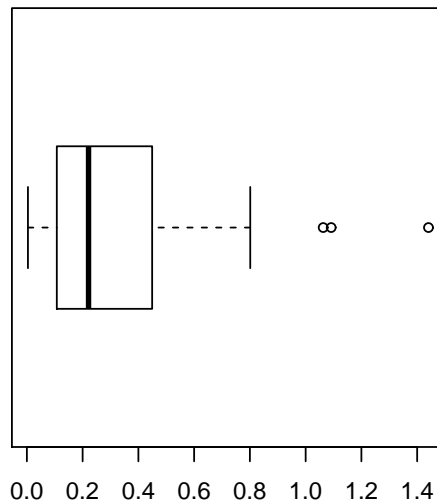
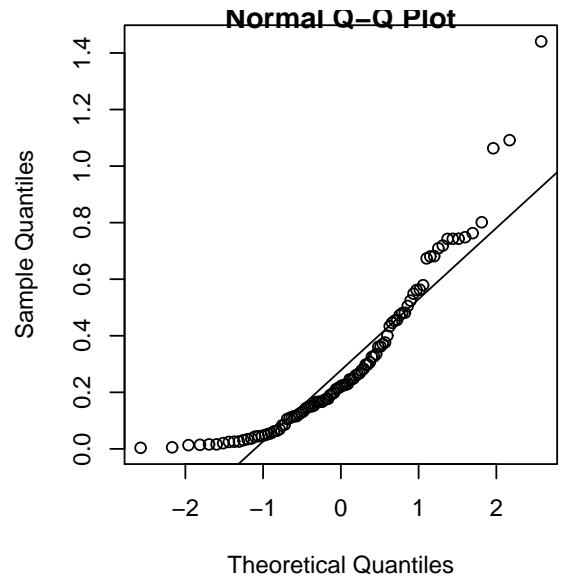
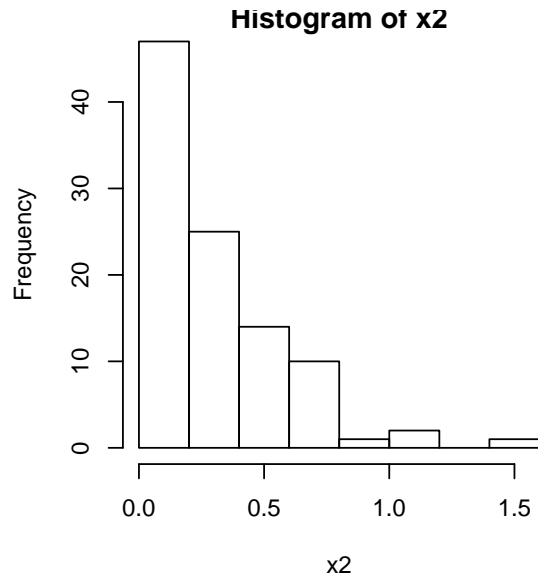


```
>shapiro.test(x1)
Shapiro-Wilk normality test
```

```
data: x1
W = 0.9529, p-value = 0.001288
```

b) Exponentialverteilung

```
>x2=rexp(100,rate=3)
>par(mfcol=c(2,2))
>hist(x2)
>boxplot(x2,horizontal =T)
>qqnorm(x2); qqline(x2)
>stripchart(x2,method="stack")
```



```
>shapiro.test(x2)
```

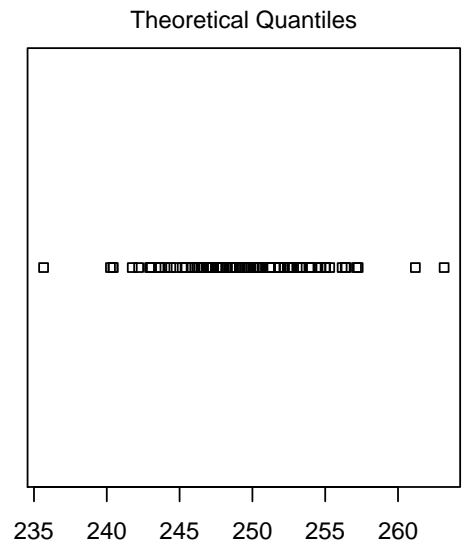
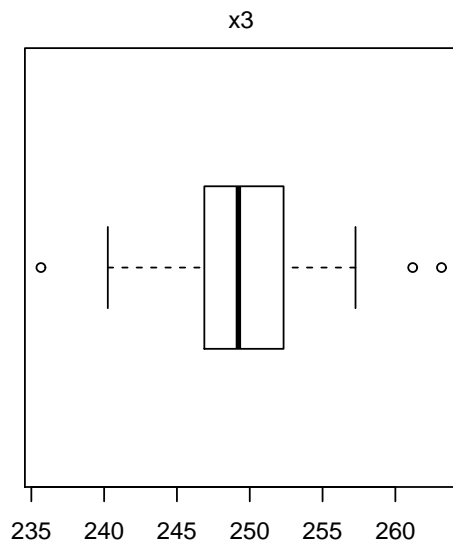
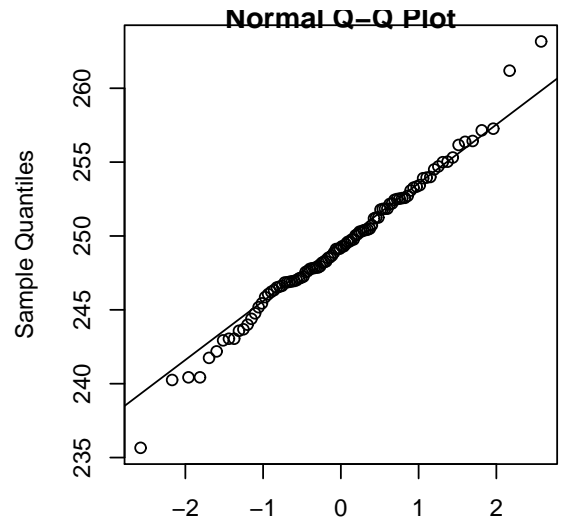
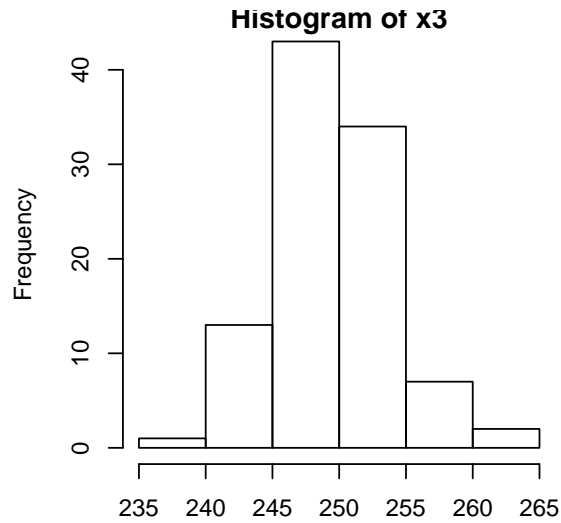
Shapiro-Wilk normality test

data: x2

W = 0.8626, p-value = 3.557e-08

c) Normalverteilung

```
x3=rnorm(100,mean=250,sd=5)
>par(mfcol=c(2,2))
>hist(x3)
>boxplot(x3,horizontal =T)
>qqnorm(x3); qqline(x3)
>stripchart(x3,method="stack")
```



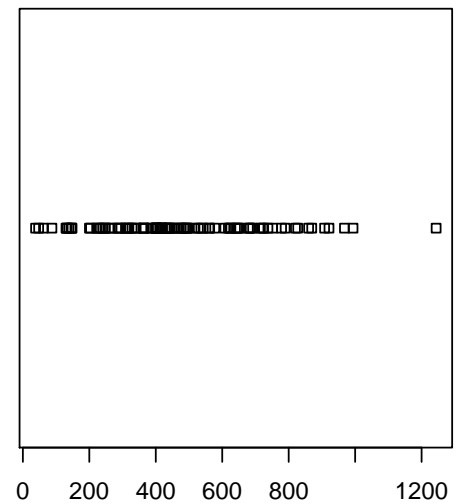
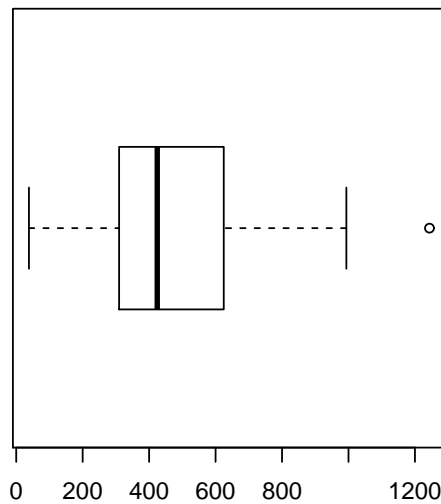
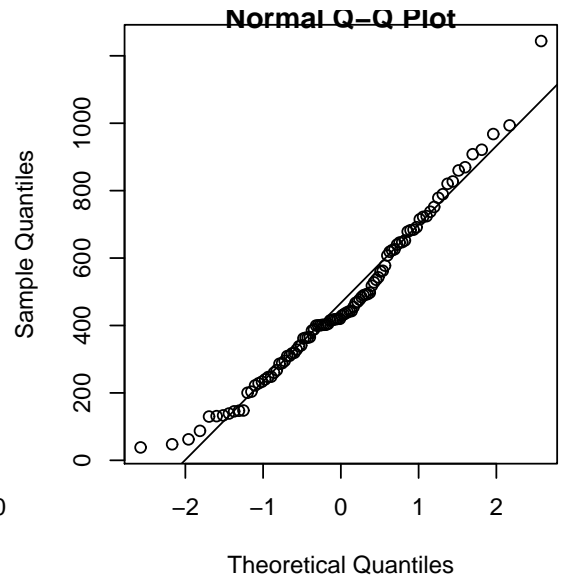
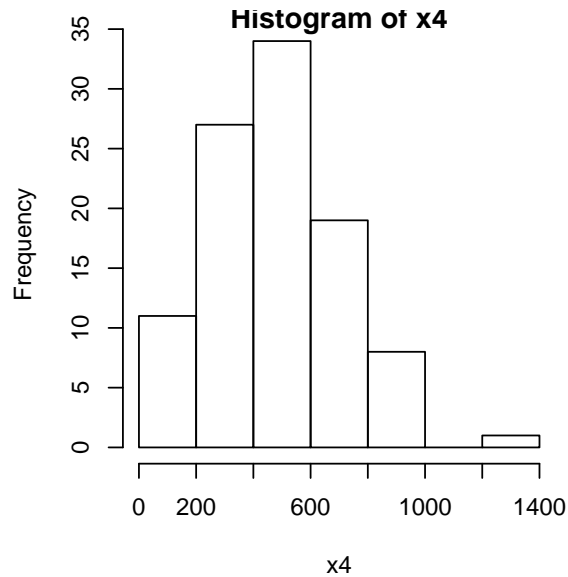
```
>shapiro.test(x3)
```

Shapiro-Wilk normality test

```
data: x3
W = 0.9888, p-value = 0.5687
```

d) Weibullverteilung

```
x4=rweibull(100,shape=2,scale=500)
>par(mfcol=c(2,2))
>hist(x4)
>boxplot(x4,horizontal =T)
>qqnorm(x4); qqline(x4)
>stripchart(x4,method="stack")
```



```
>shapiro.test(x4)
```

Shapiro-Wilk normality test

```
data: x4
W = 0.9738, p-value = 0.04339
```

## 2. Aufgabe: Ein Versuch mit einem Schlafmittel

In einem klinischen Versuch sollte die Wirksamkeit eines Schlafmittels getestet werden. Dazu wurden von den Patienten, die in einer Klinik für Schlafstörungen eingewiesen waren, zufällig 10 Patienten ausgewählt, die ein neuartiges Schlafmittel erhielten (Behandlungsgruppe), und 10 weitere Patienten, denen nur eine wie das Schlafmittel aussehende und schmeckende, wirkungslose Pille verabreicht wurde (Kontrollgruppe). Um die Wirksamkeit zu bestimmen, wurden jeweils zunächst die Schlafdauer in der Nacht vor der ersten Einnahme bestimmt. Vor der zweiten Nacht wurde den Patienten dann das Schlafmittel bzw. das Placebo gegeben. Die Schlafdauer der zweiten Nacht wurde ebenfalls ermittelt.

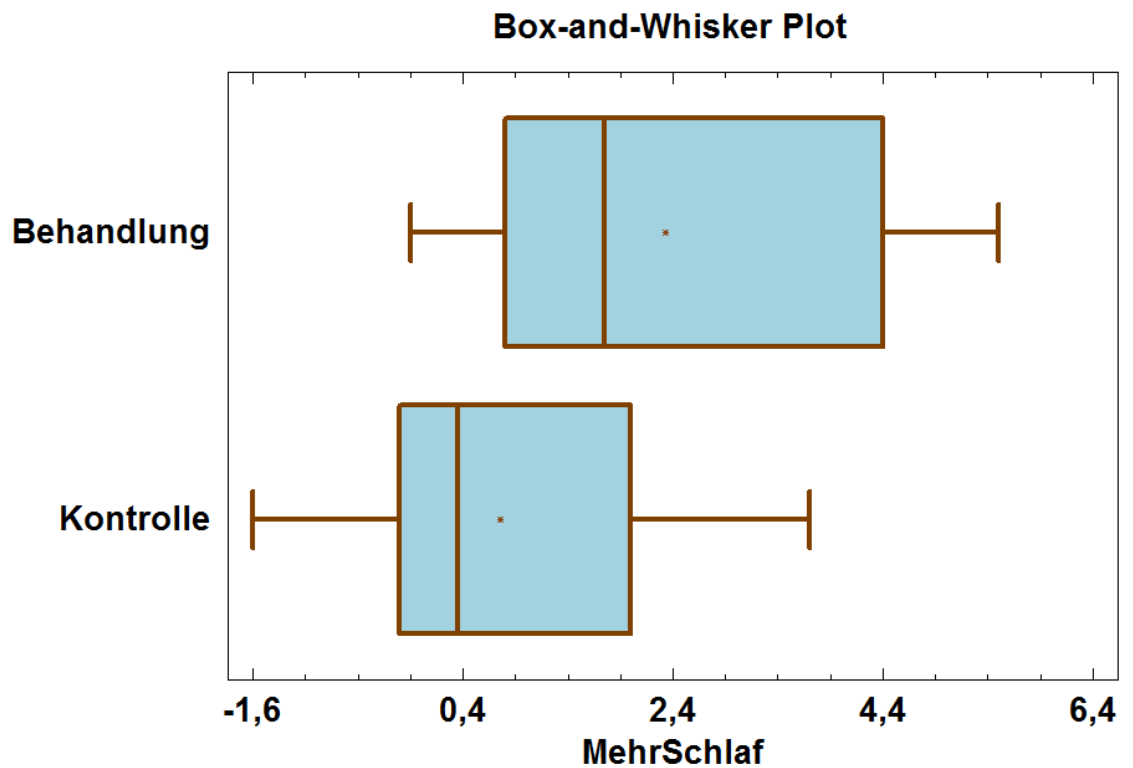
Als Daten liegen (u.a.) die Informationen vor, welches Mittel der Patient erhalten hat (Behandlung  $\hat{=}$  Schlafmittel, Kontrolle  $\hat{=}$  wirkungsloses Placebo), und wie viel länger die Patienten in der zweiten Nacht geschlafen haben (in Stunden). Ziel der Untersuchung ist es nachzuweisen, dass das Schlafmittel bei den in der Klinik behandelten Schlafstörungen wirkt. Jedoch könnte dieser Nachweis erst mit weiteren, in dieser Serie noch nicht verwendeten Methoden geführt werden.

|    | MehrSchlaf | Gruppe     |
|----|------------|------------|
| 1  | 0.7        | Kontrolle  |
| 2  | -1.6       | Kontrolle  |
| 3  | -0.2       | Kontrolle  |
| 4  | -1.2       | Kontrolle  |
| 5  | -0.1       | Kontrolle  |
| 6  | 3.4        | Kontrolle  |
| 7  | 3.7        | Kontrolle  |
| 8  | 0.8        | Kontrolle  |
| 9  | 0.0        | Kontrolle  |
| 10 | 2.0        | Kontrolle  |
| 11 | 1.9        | Behandlung |
| 12 | 0.8        | Behandlung |
| 13 | 1.1        | Behandlung |
| 14 | 0.1        | Behandlung |
| 15 | -0.1       | Behandlung |
| 16 | 4.4        | Behandlung |
| 17 | 5.5        | Behandlung |
| 18 | 1.6        | Behandlung |
| 19 | 4.6        | Behandlung |
| 20 | 3.4        | Behandlung |

- a)
- Wie liegen die Daten vor?
  - Welche Variablen gibt es und was bedeuten Sie?
  - Welche Skala haben die einzelnen Variablen?
  - Welche Annahme müssen wir über die Datenerhebung treffen, damit die Daten dafür repräsentativ sind?
  - Wofür könnten diese Daten repräsentativ sein?

b) Grafiken

- i. Welche Grafik eignet sich zur Darstellung der Häufigkeiten der Auswahl von Schlafmittel und Placebo?
- ii. Für die in der Originalpublikation verwendeten Verfahren benötigt man eine Normalverteilung der Daten in der Behandlungsgruppe. Welche statistische Grafik würde sich eignen zu erkennen, ob diese Daten einigermaßen normalverteilt sind?
- iii. Um die Behandlungserfolge mit und ohne Medikament zu vergleichen und eventuelle Ausreißer zu erkennen wurden mit Statgraphics folgende parallele Box-Plots gezeichnet.



Was können Sie aus dieser Grafik ablesen?

### 3. Aufgabe: Aquifer

Für die Untersuchung von Grundwasserleitern bezüglich Tiefe (bergmännisch Tiefe – von der Oberfläche nach unten gemessen) und Schichteigenschaften liegt der folgende *Aquifer*-Datensatz vor:

```
> aqui<-read.table("D:/Aqui.txt",header=T)
> aqui
```

|    | Teufe | Type  | Transmissivitaet |
|----|-------|-------|------------------|
| 1  | 78.64 | Poren | 0.028            |
| 2  | 49.00 | Poren | 0.073            |
| 3  | 47.00 | Poren | 0.110            |
| 4  | 43.67 | Poren | 0.144            |
| 5  | 37.00 | Poren | 0.170            |
| 6  | 23.50 | Poren | 0.440            |
| 7  | 9.00  | Poren | 1.600            |
| 8  | 80.50 | Kluft | 1.800            |
| 9  | 21.25 | Kluft | 4.300            |
| 10 | 43.50 | Kluft | 6.170            |
| 11 | 29.50 | Kluft | 13.200           |
| 12 | 16.50 | Kluft | 32.000           |
| 13 | 11.00 | Kluft | 66.000           |

Wir begutachten einige Grafiken:

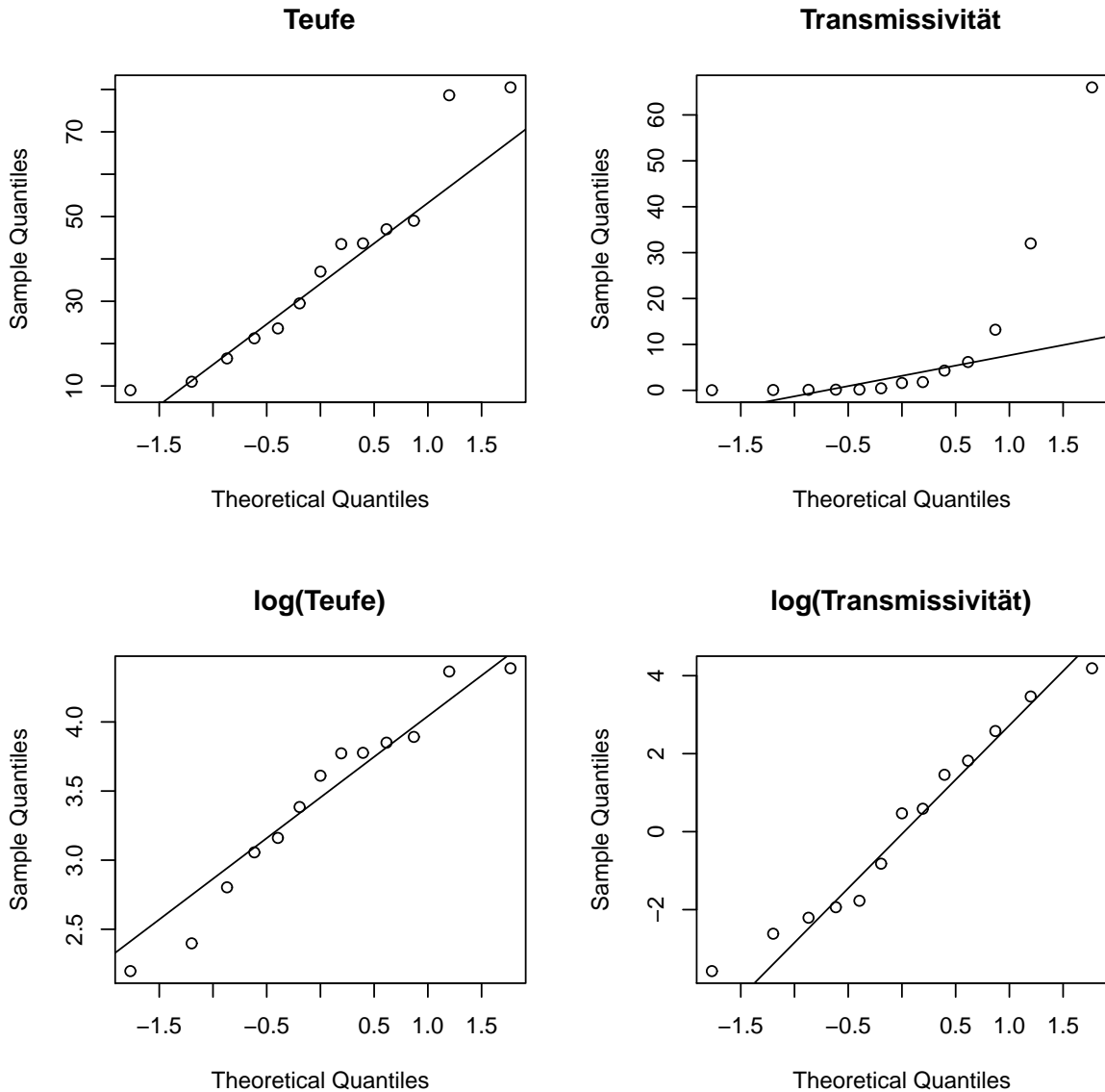
(Generierungsbefehle der Statistiksoftware **R**)

```
> attach(aqui)
```

```

> par(mfrow=c(2,2))
> qqnorm(Teufe,main="Teufe");qqline(Teufe)
> qqnorm(Transmissivitaet,main="Transmissivität");qqline(Transmissivitaet)
> qqnorm(log(Teufe),main="log(Teufe)");qqline(log(Teufe))
> qqnorm(log(Transmissivitaet),main="log(Transmissivität)")
> qqline(log(Transmissivitaet))

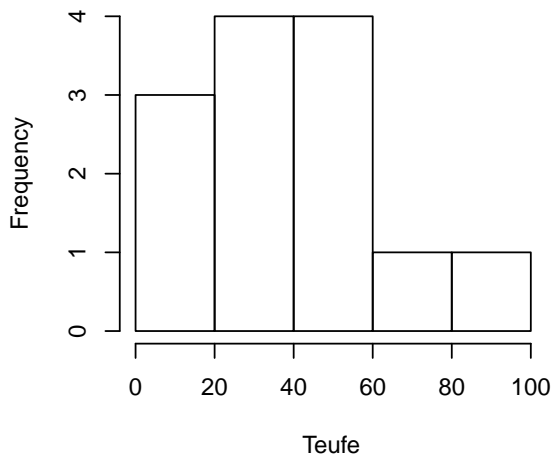
```



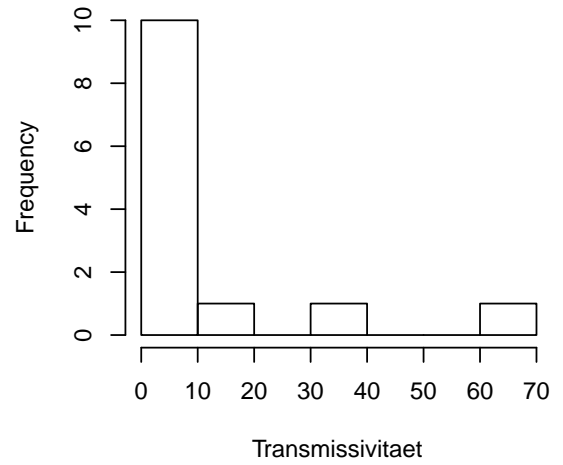


```
> par(mfrow=c(2,2))
> hist(Teufe)
> hist(Transmissivitaet)
> hist(log(Teufe))
> hist(log(Transmissivitaet))
```

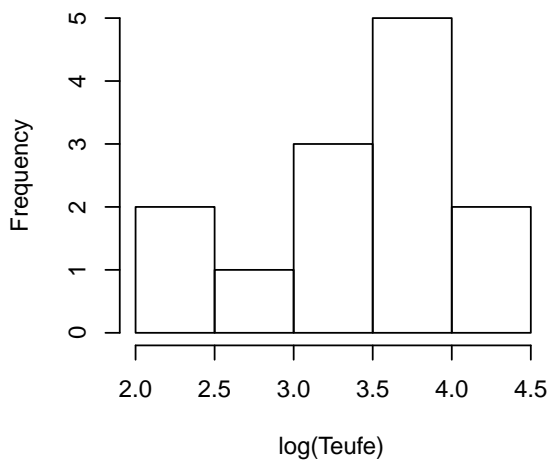
**Histogram of Teufe**



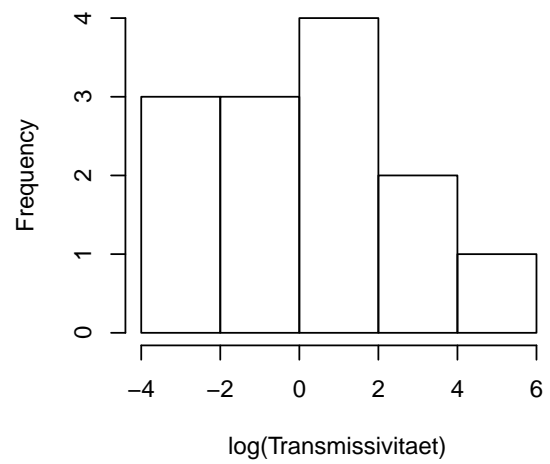
**Histogram of Transmissivitaet**



**Histogram of log(Teufe)**



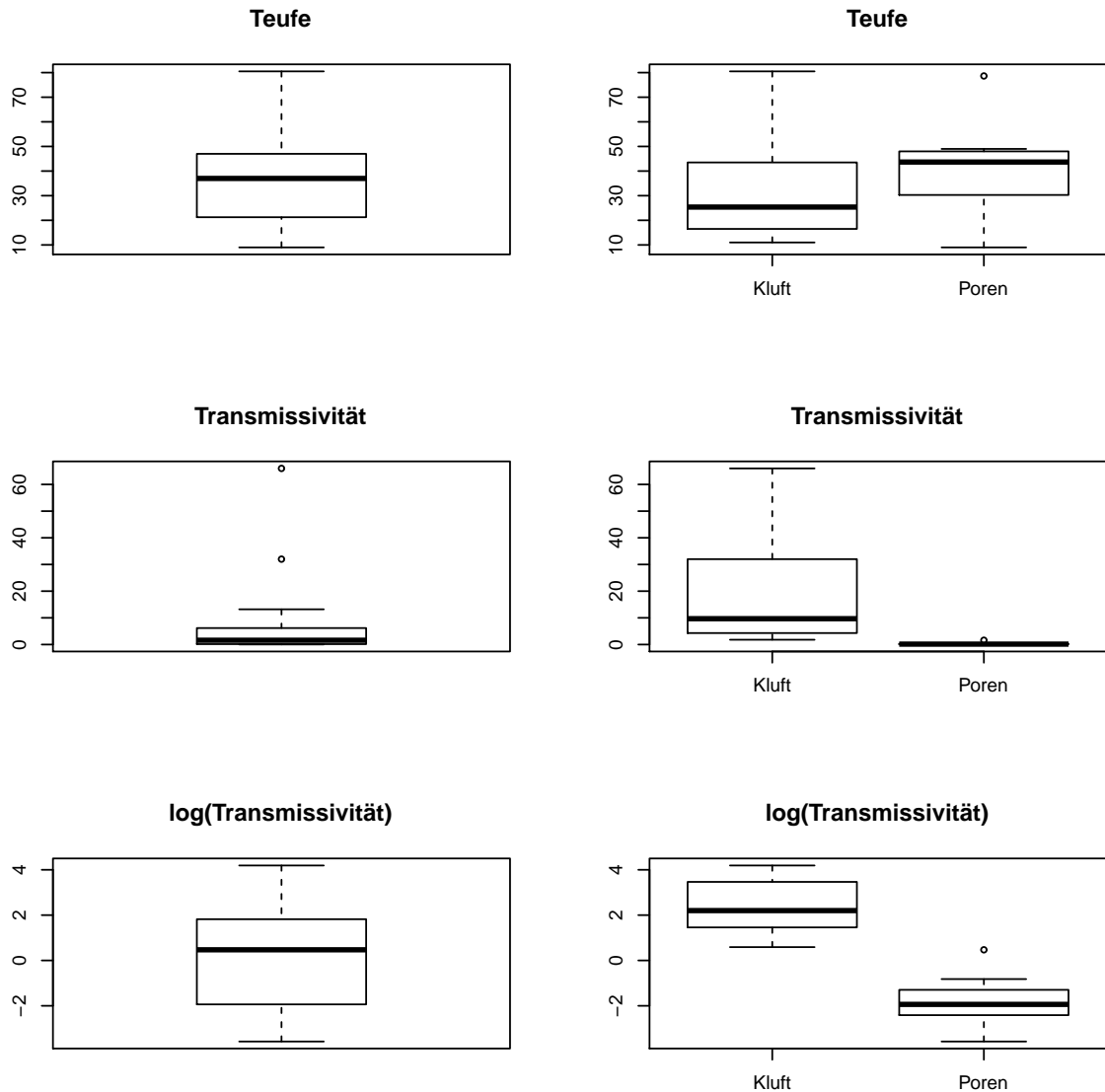
**Histogram of log(Transmissivitaet)**



```

> par(mfrow=c(3,2))
> boxplot(Teufe,main="Teufe")
> boxplot(Teufe~Type,main="Teufe")
> boxplot(Transmissivitaet,main="Transmissivität")
> boxplot(Transmissivitaet~Type, main="Transmissivität")
> boxplot(log(Transmissivitaet),main="log(Transmissivität)")
> boxplot(log(Transmissivitaet)~Type,main="log(Transmissivität)")

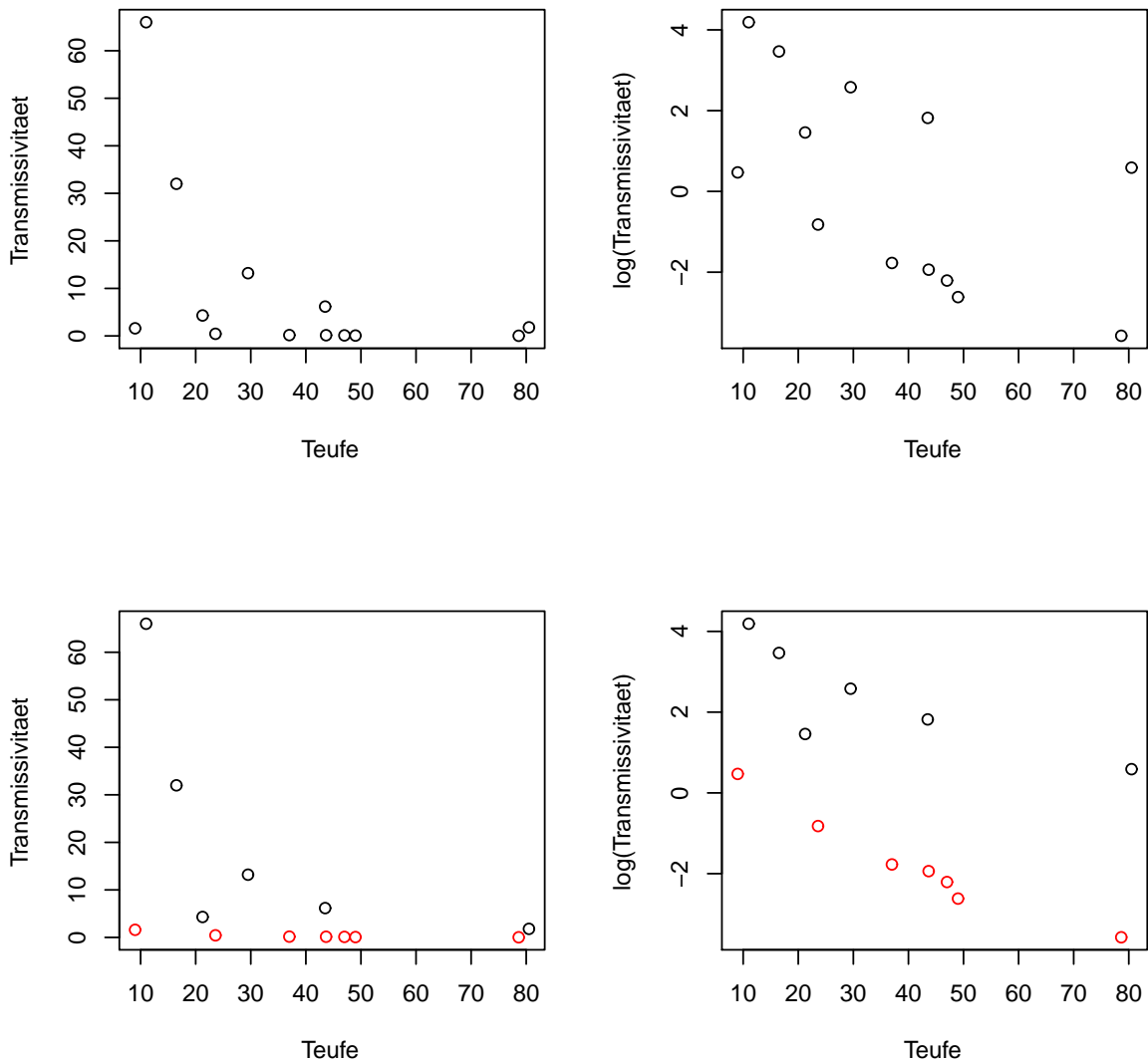
```



```

> par(mfrow=c(2,2))
> plot(Teufe, Transmissivitaet)
> plot(Teufe, log(Transmissivitaet))
> plot(Teufe, Transmissivitaet, col=Type)
> plot(Teufe, log(Transmissivitaet), col=Type)

```



- Welche Skala haben die Merkmale?
- Wie sind die Daten verteilt?
- Welche Abhängigkeiten sind erkennbar? Sind die Abhängigkeiten stark oder schwach, linear oder nichtlinear, zunehmend oder abnehmend?

4. **Aufgabe:** In der Weinabteilung eines englischen Supermarktes, der sowohl deutsche als auch französische Weine feilbot, wurde an aufeinander folgenden Tagen entweder deutsche oder französische Musik gespielt. In der ersten Woche waren die deutschen Weine auf der linken Seite des Regals, die französischen rechts, in der zweiten Woche war es umgekehrt. Nationalflaggen an den Regalen machten den Herkunftsort der Weine unmissverständlich klar. Die Musik der ersten Woche war französische (im Wesentlichen Akkordeon-Musik) und in der zweiten Woche deutsche (im Wesentlichen Blaskapellen-Musik).

| Herkunft des Weins | Musik        |          |
|--------------------|--------------|----------|
|                    | französische | deutsche |
| Frankreich         | 39           | 12       |
| Deutschland        | 8            | 22       |

(nach M. Spitzer: *Musik, Wein und Bahnungseffekte*, Geist & Gehirn, dabei ist die Studie aus: North A, Hargreaves D, McKendrick J. *The influence of in-store music on wine selections*. Journal of Applied Psychology 1999; 84: 271-6. )

```
> Tabelle <- matrix(c(39,12,8,22), 2, 2, byrow=TRUE)
> rownames(Tabelle) <- c("Frankreich", "Deutschland")
> colnames(Tabelle) <- c("französische", "deutsche")
> Tabelle # Counts
```

```

           französische deutsche
Frankreich           39         12 Deutschland           8         22

```

Was können Sie aus folgender Grafik ablesen?

```
> mosaicplot(t(Tabelle), main="Mosaikplot", color=c(2,4))
```

