

W-Rechnung und Statistik für Ingenieure

Übung 4

Lösung in R:

Lade den Datensatz SETTING.DAT:

```
>setting<-read.table("SETTING.DAT")
```

Die erste Spalte von setting gibt die prozentualen Anteile von a wieder, die zweite die von b, die dritte für c, die vierte die Anteile von d:

```
>a<-setting[,1]
```

```
>b<-setting[,2]
```

```
>c<-setting[,3]
```

```
>d<-setting[,4]
```

Aufgabe 1 : Quantile

Bestimme zunächst den Median und die 25%- und 75%-Quantile:

```
> quantile(a,p=c(0.25,0.5,0.75))
```

```
25% 50% 75%
```

```
2    7   11
```

```
> quantile(b,p=c(0.25,0.5,0.75))
```

```
25% 50% 75%
```

```
31   52   56
```

```
> quantile(c,p=c(0.25,0.5,0.75))
```

```
25% 50% 75%
```

```
8    9   17
```

```
> quantile(d,p=c(0.25,0.5,0.75))
```

```
25% 50% 75%
```

```
20   26   44
```

Aufgabe 1 : Statistische Kennzahlen

Welche statistischen Kennzahlen können noch berechnet werden?

Welcher Datentyp liegt vor? -quantitativ

Also weitere mögliche Kennzahlen (Vgl. VL Kap.5 und Kap.6)

- arithmetisches Mittel $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$
- Spannweite $R = x_{(N)} - x_{(1)}$
- Quartilsabstand $Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}$
- Median der absoluten Abweichung
 $\text{Median}(|x_1 - \tilde{x}_{0.5}|, \dots, |x_N - \tilde{x}_{0.5}|)$
- Standardabweichung $s = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2}$

Aufgabe 1 : Abgeleitete Kennzahlen

Aus den bisherigen Kennzahlen abgeleitet werden können noch:

- Varianz $s^2 = \frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2$
- Quartilskoeffizient $Q_{koeff} = \frac{2Q_3 - Q_1}{\tilde{x}_{0.25} + \tilde{x}_{0.75}}$
- Variationskoeffizient $V = \frac{s}{\bar{x}}$

Aufgabe 1 : Berechnung in R

```
k<-summary(data) #Max,Min,Quartile und Mittelwert  
R<-k[6]-k[1] #Spannweite Max-Min  
Q<-k[5]-k[2] #Quartilsabstand  $q(0.75)-q(0.25)$   
dMad<-mad(data,constant=1)#Median der abs. Abweichungen  
stdev<-sqrt(var(data)) #Standardabweichung  
var<-var(data) #Varianz  
Qkoeff<-2*Q/(k[2]+k[5]) #Quartilskoeffizient  
Vkoeff<-stdev/k[4] #Variationskoeffizient
```

Natürlich können auch die durch
`quantile()`, `mean()`, `min()`, `max()` gewonnenen Daten statt der
Funktion `summary()` verwendet werden.

Aufgabe 1 : Vergleich

	a	b	c	d
Lageparameter				
Min	1.00	26.00	4.00	6.00
$\tilde{x}_{0.25}$	2.00	31.00	8.00	20.00
Median	7.00	52.00	9.00	26.00
Mittelwert	7.46	48.15	11.77	30.00
$\tilde{x}_{0.75}$	11.00	56.00	17.00	44.00
Max	21.00	71.00	23.00	60.00
Streuungsparameter				
R	20.00	45.00	19.00	54.00
Q	9.00	25.00	9.00	24.00
d_{MAD}	4.00	14.00	3.00	14.00
s	5.88	15.56	6.41	16.74
s^2	34.60	242.14	41.03	280.17
Q_{koeff}	1.38	0.57	0.72	0.75
V	0.79	0.32	0.54	0.56

Aufgabe 2 : Box-Whisker-Plot

Box-Whisker-Plot:

Grafische Darstellung von Lage- und Streuungsparametern. Besteht aus einer Box,

- deren linke Kante dem unteren Quartil $\tilde{x}_{0.25}$ entspricht und
- deren rechte Kante dem oberen Quartil $\tilde{x}_{0.75}$ entspricht.
- Der Median wird durch einen Strich in der Box gekennzeichnet.
- Von der Box gehen links ein Strich bis zu der kleinsten Beobachtung, die größer ist als der unter Whisker $\tilde{x}_{0.25} - 1.5Q$.
- Von der Box gehen rechts ein Strich bis zu der größten Beobachtung, die kleiner ist als der obere Whisker $\tilde{x}_{0.75} + 1.5Q$.
- Alle Beobachtungen, die über bzw. unter den Whiskers liegen werden mit Kreisen eingezeichnet und gelten als Ausreißer.

Aufgabe 2 : Erstellen des Box-Whisker-Plot in R

Verwende Datenvektoren aus Blatt 2 Aufgabe 2.

Für einzelne Box-Whisker-Plots nutze

```
> boxplot(linie1)
> boxplot(linie2)
```

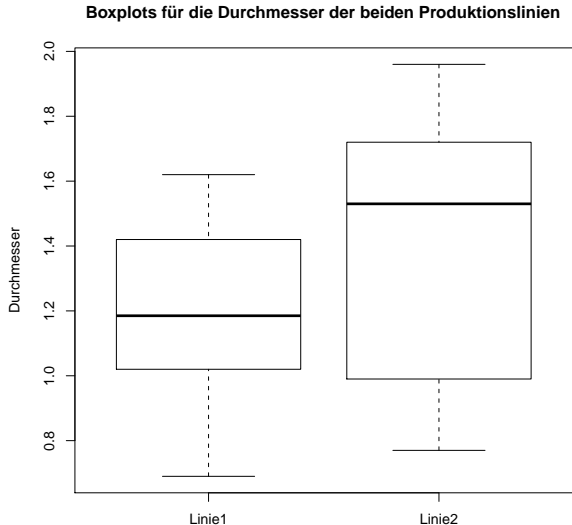
Für mehrere Box-Whisker-Plots in einer Abbildung nutze

```
> boxplot(list(linie1,linie2),ylab="Durchmesser",
+  names=c("Linie1","Linie2"),
+  main="Boxplots für die Durchmesser
+  der beiden Produktionslinien")
```

Achtung: R benutzt eine andere Definition des Boxplots:

Untere Grenze der Box ist der Median der unteren Hälfte (Minimum bis zum Median) des sortierten Datensatzes und die obere Grenze wird bestimmt als der Median der oberen Hälfte des Datensatzes.

Aufgabe 2 : Box-Whisker-Plot der Produktionsdaten



Aufgabe 2 : Streuungsparameter

```
> #Spannweite
> max(linie1)-min(linie1)
[1] 0.93
> max(linie2)-min(linie2)
[1] 1.19

> #Quartilsabstand
> Q1<-quantile(linie1,p=0.75)-quantile(linie1,p=0.25)
0.3575
> Q2<-quantile(linie2,p=0.75)-quantile(linie2,p=0.25)
0.6475

> #Median der absoluten Abweichung
> mad(linie1,constant=1)
[1] 0.2
> mad(linie2,constant=1)
[1] 0.25
```

Aufgabe 2 : Streuungsparameter

```
> #Standardabweichung
> sd1<-sqrt(1/(length(linie1)-1)*sum((linie1-mean(linie1))^2))
[1] 0.2896818
> sd2<-sqrt(1/(length(linie2)-1)*sum((linie2-mean(linie2))^2))
[1] 0.4283093

> #Varianz
> 1/(length(linie1)-1)*sum((linie1-mean(linie1))^2)
[1] 0.083915
> 1/(length(linie2)-1)*sum((linie2-mean(linie2))^2)
[1] 0.18344

> #Quartilskoeffizient
> 2*Q1/(quantile(linie1,p=0.75)+quantile(linie1,p=0.25))
0.2939363
> 2*Q2/(quantile(linie2,p=0.75)+quantile(linie2,p=0.25))
0.4662466
```

Aufgabe 2 : Streuungsparameter

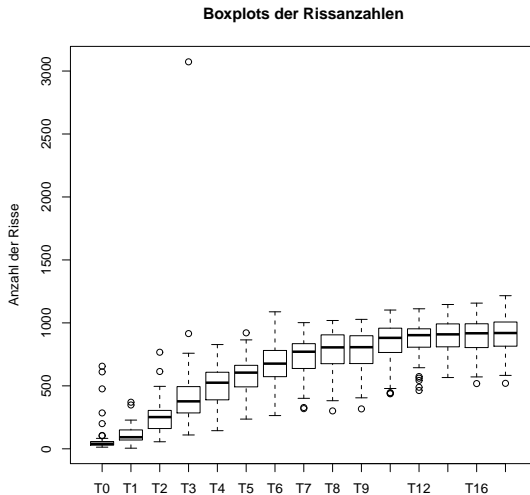
```
> #Variationskoeffizient  
> sd1/mean(linie1)  
[1] 0.2426146  
> sd2/mean(linie2)  
[1] 0.3046297
```

Aufgabe 3 : Boxplots für Rissdaten

Erstelle Boxplots für die Rissdaten aus dem Crackcounts Datensatz.

```
boxplot(cracks[,1:15],main='Boxplots der  
Rissanzahlen',ylab='Anzahl der  
Risse',names=c("T0","T1","T2","T3","T4","T5"  
,"T6","T7","T8","T9","T10","T12","T14","T16","T18")):
```

Aufgabe 3 : Boxplots mit Extremwerten



Aufgabe 3 : Boxplots ohne Extremwerte

Mit der Option `outline=F` der Funktion `boxplot()` werden Ausreißer nicht abgebildet.

