

Übung Wahrscheinlichkeitsrechnung und Statistik in den Ingenieurwissenschaften

Übung 4

Aufgabe 1

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

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

Welche statistischen Kennzahlen können noch berechnet werden?

Welcher Datentyp liegt vor? -quantitativ

Also weitere mögliche Kennzahlen

- ▶ 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 $s = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2}$
- ▶ Varianz, Quartilkoeffizient, Variationskoeffizient

Aufgabe 1

```
> mean(a)
[1] 7.461538
> #Spannweite
> max(a)-min(a)
[1] 20
> #Quartilsabstand
> quantile(a,p=0.75)-quantile(a,p=0.25)
75%
  9
> #Median der absoluten Abweichung
> mad(a,constant=1)
[1] 4
> #Standardabweichung
> sqrt(1/length(a)*sum((a-mean(a))^2))
[1] 5.651622
```

Die 4 Anteile im Vergleich. Lagekennzahlen:

Komponente	Minimum	$\tilde{x}_{0.25}$	$\tilde{x}_{0.5}$	$\tilde{x}_{0.75}$	Maximum	\bar{x}
a	1	2	7	11	21	7.46
b	26	31	52	56	71	48.15
c	4	8	9	17	23	11.77
d	6	20	26	44	60	30

Die 4 Anteile im Vergleich. Streuungsparameter:

	R	Q	d_{MAD}	s	s^2	Q_{koeff}	V
a	20	9	4	5.88	34.60	1.38	0.79
b	45	25	14	15.56	242.14	0.57	0.32
c	19	9	3	6.40	41.02	0.72	0.54
d	54	24	14	16.74	280.17	0.75	0.56

Aufgabe 2

Box-Whisker-Plot:

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

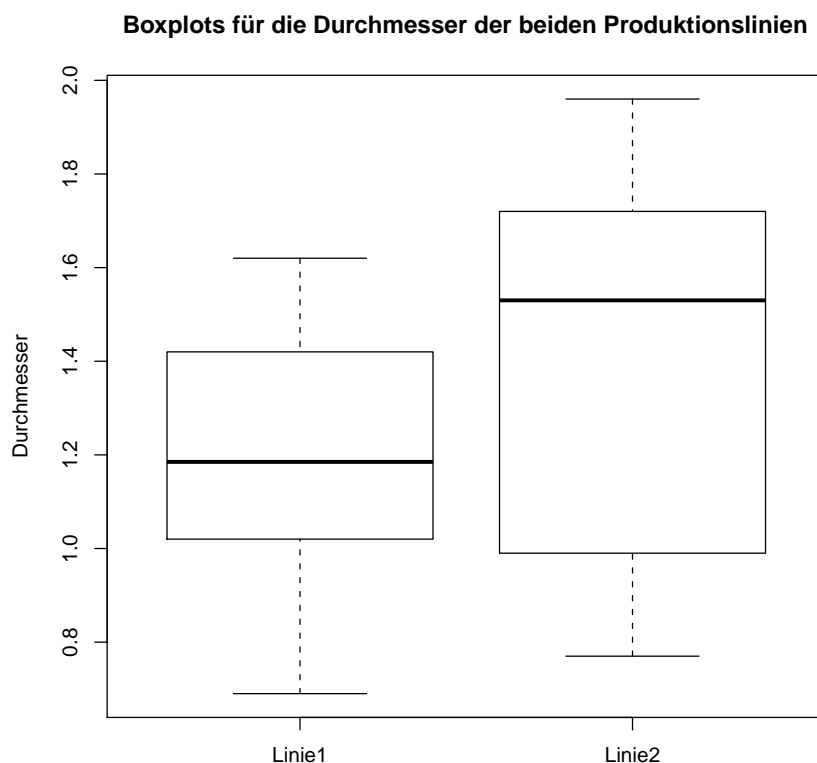
- ▶ deren linke Kante dem unteren Quartil $\tilde{x}_{0.25}$ entspricht und
- ▶ deren rechte Kante dem oberen Quartil $\tilde{x}_{0.75}$ entspricht.
- ▶ Der Median wird durch einen Strich in der Box gekennzeichnet.
- ▶ Von der Box gehen links ein Strich bis zu der kleinsten Beobachtung, die größer ist 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.

Aufgabe 2

```
> linie1<-stahl[1:10]
> linie2<-stahl[11:20]
> boxplot(linie1)
> boxplot(linie2)
> boxplot(list(linie1,linie2),ylab="Durchmesser",
+   names=c("Linie1","Linie2"),
+   main="Boxplots für die Durchmesser
+   der beiden Produktionslinien")
```

Achtung: R benutzt eine etwas 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



Aufgabe 2

```
> #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

> #Standardabweichung
> sd1<-sqrt(1/length(linie1)*sum((linie1-mean(linie1))^2))
[1] 0.2748163
> sd2<-sqrt(1/length(linie2)*sum((linie2-mean(linie2))^2))
[1] 0.4063299

> #Varianz
> 1/length(linie1)*sum((linie1-mean(linie1))^2)
[1] 0.075524
> 1/length(linie2)*sum((linie2-mean(linie2))^2)
[1] 0.165104

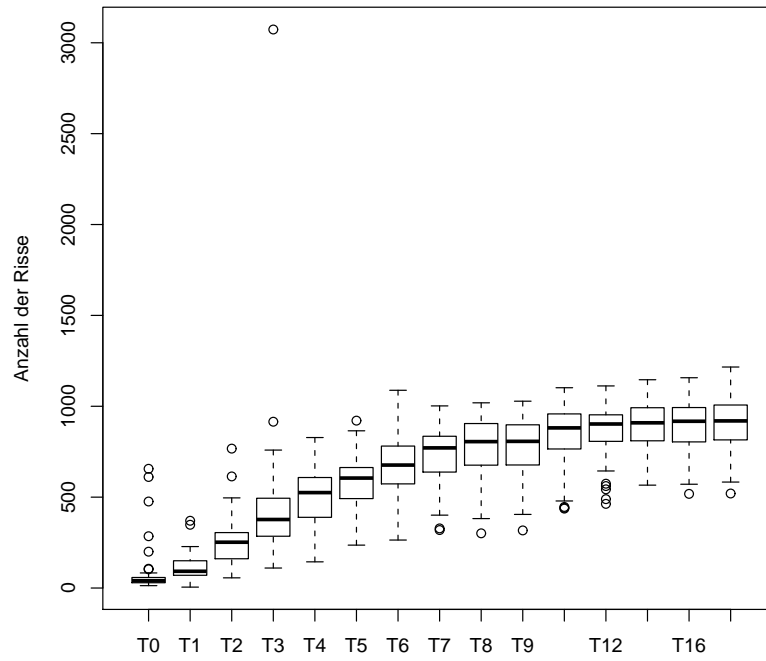
> #Quartilskoeffizient
> 2*Q1/(quantile(linie1,p=0.75)+quantile(linie1,p=0.25))
0.2939363
> 2*Q2/(quantile(linie2,p=0.75)+quantile(linie2,p=0.25))
0.4662466
```

```
> #Variationskoeffizient  
> sd1/mean(linie1)  
[1] 0.2301644  
> sd2/mean(linie2)  
[1] 0.2889971
```

Aufgabe 3

```
boxplot(cracks[,1:15],main='Boxplots der  
Rissanzahlen',ylab='Anzahl der  
Risse',names=c('T02','T1','T2','T3','T4','T5'  
, 'T6','T7','T8','T9','T10','T12','T14','T16','T18')):
```

Boxplots der Rissanzahlen



ohne "Ausreißer":

Boxplots der Rissanzahlen

