**Verwaltungs- und Wirtschafts- Akademie Potsdam e. V.**
Dipl.-Kfm. Thomas Rochow

## Notation:

### Allgemeine Notation:

nicht genormt!

$X, Y$: statistisches Merkmal, statistische Größe

$N$: Umfang der Grundgesamtheit

$n$: Stichprobenumfang

$e_i$: statistische Einheit $i$, Merkmalsträger $i$

$x_i$: Beobachtungswert Nummer $i$ bezüglich Merkmal X ohne Beachtung des Merkmalsträgers

alle Merkmalsträger, die für die Untersuchung relevant sind

### Notation für eindimensionales ungruppiertes (unklassiertes) Datenmaterial:

$a_j$ : Werte, die die betrachtete statistische Größe in unserer Stichprobe für die statistische Größe konkret annimmt

$h(a_j)$: absolute Häufigkeit der Ausprägung $a_j$

$f(a_j)$: relative Häufigkeit der Ausprägung $a_j$

$H(x)$: absolute kumulierte Häufigkeit

$F(x)$: relative kumulierte Häufigkeit

Kumulierte Häufigkeiten geben Auskunft über höchstens (also kleiner/gleich) Fragestellungen

### Notation für eindimensionales gruppiertes (klassiertes) Datenmaterial:

$j$: Nummer der Gruppe (Klasse)

$h_j$ : absolute Häufigkeit der Gruppe $j$

$f_j$ : relative Häufigkeit der Gruppe $j$

$H(x)$: absolute kumulierte Häufigkeit

$F(x)$: relative kumulierte Häufigkeit

Kumulierte Häufigkeiten geben Auskunft über höchstens (also kleiner/gleich) Fragestellungen

$b_j$ : Breite der Gruppe $j$

$$b_j = Obergrenze\ \ Gruppe\ j\ -\ Untergrenze\ der\ Gruppe\ j$$

$m_j$ : Mitte der Gruppe $j$

$$m_j = \frac{Untergrenze\ \ Gruppe\ j\ + Obergrenze\ der\ Gruppe\ j}{2}$$

## MUSTERAUFGABEN Statistik I

**Aufgabe 1 (Eindimensionales Datenmaterial – gruppiert) – Skript Aufgabe 10, S. 28**

Aus der Kriminalstatistik des Monats Juni wurden folgende Daten bezüglich der Steuerkriminalität entnommen (Schadenshöhe in 10 000 €):

| Höhe des Schadens (in 10 000 €) von ... bis unter | Anzahl |
|---|---|
| 1,5 – 2,5 | 1 |
| 2,5 – 3,5 | 1 |
| 3,5 – 4,5 | 5 |
| 4,5 – 7,5 | 3 |

1. Wie heißt das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die relativen sowie die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Bestimmen Sie aus den vorliegenden Daten das arithmetische Mittel! Welche Annahme treffen Sie bei der Berechnung des arithmetischen Mittels?
4. Stellen Sie die absoluten Häufigkeiten grafisch dar! Wie heißt die von Ihnen gewählte Darstellungsform? Welches Prinzip haben Sie bei der Darstellung hoffentlich beachtet?
5. Zeichnen Sie die empirische Verteilungsfunktion! Welche Annahme treffen Sie hier?
6. Schätzen Sie grafisch den Median der vorliegenden Häufigkeitsverteilung!
7. Berechnen Sie die absolute durchschnittliche Abweichung und die empirische Varianz, welche Annahme treffen Sie dabei für jede Gruppe?
8. Welche Skalierungsart muss gegeben sein, um die kumulierten Häufigkeiten sinnvoll tabellarisch bestimmen zu können? Begründen Sie bitte!

**Aufgabe 2 (Eindimensionales Datenmaterial–ungruppiert) – Skript Aufgabe 11, S. 28**

Aus der Kriminalstatistik des Monats April wurden folgende Daten bezüglich der Steuerkriminalität entnommen (Schadenshöhe in 10 000 €):

3 ; 7 ; 6 ; 5 ; 6 ; 4,5 ; 4,5 ; 5 ; 5 ; 4

**Hinweis: Lassen Sie sich nicht durch die beiden „Kommawerte“ irritieren!**

1. Wie lautet das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die absoluten, die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die absoluten Häufigkeiten grafisch dar!
4. Zeichnen Sie die empirische Verteilungsfunktion!
5. Berechnen Sie das arithmetische Mittel, und bestimmen Sie den Median und den Modus der vorliegenden Häufigkeitsverteilung!
6. Bestimmen Sie die Spannweite der vorliegenden Häufigkeitsverteilung!
7. Bestimmen Sie die durchschnittliche absolute Abweichung und die empirische Varianz!
8. Welche Skalierungsart muss mindestens gegeben sein, um die empirische Verteilungsfunktion sinnvoll **zeichnen** zu können! Begründen Sie bitte!

**Aufgabe 3 – Skript Spezialaufgabe S. 29**

Ein Pharmaunternehmen möchte wissen, ob sein neu entwickeltes Medikament gegen Bluthochdruck besser wirkt als ein älteres Medikament. Dazu wurde eine Gruppe von 21 Probanden untersucht, davon bekamen 10 das ältere und die restlichen 11 Probanden das neue Medikament. Der Blutdruck wurde jeweils zu Beginn des Tests und zwei Wochen später gemessen. Erfasst wurde jeweils die Höhe der Blutdrucksenkung:

Mit dem alten Medikament: 7; 15; 11; 15; 25; 4; 19; 10; 8; 31
Mit dem neuen Medikament: 11; 16; 25; 9; 36; 40; 48; 4; 26; 36; 14

Prüfen Sie mit Hilfe von Boxplots, ob das neue Medikament eine bessere Wirkung hat als das alte Medikament. Begründen Sie Ihre Aussage.

**Aufgabe 4 (Konzentrationsrechnung) – Skript Aufgabe 5, S. 39**

Gegeben seien die Jahresumsätze (in Mio. €) von den 10 Unternehmen einer Branche:

$$x_1 = 20; \quad x_2 = 30; \quad x_3 = 20; \quad x_4 = 15; \quad x_5 = 30$$
$$x_6 = 15; \quad x_7 = 40; \quad x_8 = 15; \quad x_9 = 15; \quad x_{10} = 50$$

1. Berechnen Sie die Lorenzsche Konzentrationsverteilung und zeichnen Sie die Lorenzkurve! Ist der Gini-Koeffizient gleich =0?
2. Welcher Anteil am Gesamtumsatz entfällt auf die 80% umsatzschwächsten Unternehmen?
3. Wie groß ist der Anteil am Gesamtumsatz, der auf die „mittleren“ 60% der Unternehmen entfällt?
4. Berechnen Sie die Konzentrationsraten CR1, CR3, CR5 und CR8!
5. Berechnen Sie den Herfindahl-Index!
6. Um welchen Betrag ändert sich – unter sonst gleichen Bedingungen (d. h. nicht ändernden Umsätzen) – der Herfindahl-Index, wenn die beiden Unternehmen mit Umsatz 30 Mio. € fusionieren, wobei sich deren Umsätze addieren?

**Zusätzliche Aufgaben**

# Aufgabe 1

20 Beschäftigte wurden nach der Entfernung ihres Wohnsitzes zum Arbeitsplatz (in km) befragt. Das Ergebnis der Befragung ist in nachfolgender Tabelle wiedergegeben:

| **X: „Entfernung des Wohnsitzes zum Arbeitsplatz (in km)** | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| **absolute Häufigkeit** | 5 | 4 | 1 | 3 | 3 | 4 |

1. Wie lautet das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die absoluten Häufigkeiten grafisch dar!
4. Zeichnen Sie die empirische Verteilungsfunktion!
5. Berechnen Sie das arithmetische Mittel, und bestimmen Sie den Median der Häufigkeitsverteilung! Welchen Wert hat der Modus? (Denken Sie an die Angabe der Einheit!)
6. Bestimmen Sie die Spannweite der vorliegenden Häufigkeitsverteilung! (Denken Sie an die Angabe der Einheit!)
7. Berechnen Sie die empirische Varianz und die Standardabweichung! (Denken Sie an die Angabe der Einheit!)
8. Berechnen Sie die durchschnittliche absolute Abweichung! (Denken Sie an die Angabe der Einheit!)
9. Welche Skalierungsart muss mindestens gegeben sein, um die empirische Verteilungsfunktion sinnvoll zeichnen zu können? Begründen Sie bitte!
10. Welche Skalierungsart muss mindestens gegeben sein, um die kumulierten Häufigkeiten sinnvoll tabellarisch bestimmen zu können? Begründen Sie bitte!

# Aufgabe 2

Um eine Übersicht über das Zahlungsverhalten seiner Kunden zu erhalten, lässt der Finanzvorstand eines Unternehmens die Daten von 100 Kunden erheben. Das Ergebnis ist in nachstehender Tabelle zusammengefasst und soll nun von Ihnen analysiert werden:

| **Wochen bis zur Begleichung der Rechnung von ... bis unter** | **Anzahl der Kunden (absolute Häufigkeit)** |
|---|---|
| 2 - 4 | 10 |
| 4 - 10 | 30 |
| 10 - 12 | 40 |
| 12 - 16 | 20 |

1. Wie heißt die statistische Größe und wie ist diese skaliert?
2. Stellen Sie die relativen, die absoluten kumulierten sowie die relativen kumulierten Häufigkeiten tabellarisch dar!
3. Berechnen Sie das arithmetische Mittel der vorliegenden statistischen Größe! (Denken Sie an die Angabe der Einheit!)
4. Stellen Sie die absoluten Häufigkeiten grafisch dar! Wie heißt die von Ihnen gewählte Darstellungsform? Welches Prinzip haben Sie bei der Darstellung berücksichtigt?
5. Zeichnen Sie die empirische Verteilungsfunktion und schätzen Sie anschließend grafisch den Median aus der Verteilungsfunktion! (Denken Sie an die Angabe der Einheit!)
6. Berechnen Sie den Median mit Hilfe der Formel: $x_{med} = x_e^u + \frac{0{,}5 - F\left(x_e^u\right)}{f\left(x_e\right)} \cdot \left(x_e^o - x_e^u\right)$
7. Ermitteln Sie rechnerisch den Interquartilsabstand!
   **Hinweis:** $x_{0,25}$ und $x_{0,75}$ bestimmen Sie bitte analog zu $x_{med}$ **!**

## Aufgabe 3

Die folgende Messreihe zeigt uns die Eigenkapitalrentabilität von den 20 Unternehmen einer Branche:

| | | | | |
|---|---|---|---|---|
| **5** | **16** | **11** | **19** | **14** |
| **7** | **14** | **6** | **6** | **28** |
| **13** | **8** | **9** | **11** | **7** |
| **14** | **7** | **13** | **26** | **9** |

1. Fertigen Sie ein Standard-Box-Whisker-Plot (Ausreißer werden zunächst nicht beachtet)!

   **Boxplot:** Minimum, Maximum, unteres Quartil (25%-Wert), Median (50%-Wert), oberes Quartil (75%-Wert), Spannweite, Interquartilsabstand (IQA)

2. Fertigen Sie ein Box-Whisker-Plot unter Kennzeichnung von Ausreißern, wobei die Länge eines Whisker auf 1,5*IQA beschränkt ist (das ist eine oft zitierte Definition eines Ausreißers nach Tukey) oder anders ausgedrückt: Ausreißer sind Werte, die kleiner sind als $x_{0,25} - 1{,}5 \cdot IQA$ oder größer sind als $x_{0,75} + 1{,}5 \cdot IQA$ .

## Aufgabenstellung 4

Sie wollen auf dem Markt, auf dem das Unternehmen ABC tätig ist, eine Konkurrenzanalyse vornehmen, ABC hat vier Konkurrenten, die Umsätze des letzten Jahres sind Ihnen sortiert in nachstehender Auflistung gegeben, das Unternehmen ABC ist immer noch das umsatzstärkste Unternehmen (alle Angaben in Mio. €):

$$\mathrm{x}_1 = 1; \quad \mathrm{x}_2 = 1; \quad \mathrm{x}_3 = 1; \quad \mathrm{x}_4 = 2; \quad \mathrm{x}_5 = 20$$

1. Berechnen Sie die Lorenzsche Konzentrationsverteilung und zeichnen Sie die Lorenzkurve!
2. Welcher Anteil am Gesamtumsatz entfällt auf die 80% umsatzschwächsten Unternehmen? Bestimmen Sie diesen Anteil mit Hilfe der erstellten Tabelle!
3. Berechnen Sie den Gini-Koeffizient $(Gini\ oder\ \ G)$ und den normierten Gini-Koeffizient $(Gini^*\ \ oder\ \ G^*)$!

$$Gini = G \;=\; \frac{2\sum_{i=1}^{n} i \cdot x_i}{n \cdot \sum_{i=1}^{n} x_i} - \frac{n+1}{n}$$

$$Gini^* = \; G^* \;=\; \frac{n}{n-1} \cdot Gini$$

# Skalierungsarten

Möglichkeiten der Datenauswertung

siehe Mini Skript

siehe Notizen

**Häufigkeiten** für ungruppierte Datenmaterial

$a_j$: realisierte Merkmalsausprägung

$h(a_j)$ absolute Häufigkeit (H)

ansätzlen

$f(a_j)$ relative H. prozentieren

$$\boxed{f(a_j) = \frac{h(a_j)}{n}}$$

$H(x)$ absolute kumulierte H.

absolute angehäufte H

schrittweises Aufaddieren

kleiner / gleich höchstens

$F(x)$ relative kumulierte H.

||

$\Downarrow$

$F(x)$ geschätzt

empirische Verteilungsfkt

über eine Erhebung gewonnene

## Aufgabe 1 (Musteraufgabe 2)

Aus der Kriminalstatistik des Monats April wurden folgende Daten bezüglich der Steuerkriminalität entnommen (Schadenshöhe in 10 000 €):

3 ; 7 ; 6 ; 5 ; 6 ; 4,5 ; 4,5 ; 5 ; 5 ; 4

**Hinweis: Lassen Sie sich nicht durch die beiden „Kommawerte" irritieren!**

1. Wie lautet das statistische Merkmal und wie ist es skaliert? 2 Pkt.
2. Bestimmen Sie tabellarisch die absoluten, die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!

4 Pkt

1) X: Schadenshöhe (in 10 000 €)

metrisch skaliert

3 ; 7 ; 6 ; 5 ; 6 ;
4,5 ; 4,5 ; 5 ; 5 ; 4

Urliste

unsortiert

Tabellen („sortieren")

Grafiken („visualisieren")

Maßzahlen („verdichten")

→ Lageparameter (Wo?)

→ Streuungsparameter (Wie?)

3 ; 7 ; 6 ; 5 ; 6 ; 4,5 ; 4,5 ; 5 ; 5 ; 4

| $a_j$ | Strichliste | $h(a_j)$ | $f(a_j)$ | $H(x)$ | $F(x)$ |
|---|---|---|---|---|---|
| 3 | \| | 1 | $1/10 = 0{,}1$ | 1 | 0,1 |
| 4 | \| | 1 | 0,1 | 2 | 0,2 |
| 4,5 | \|\| | 2 | 0,2 | 4 | 0,4 |
| 5 | \|\|\| | 3 | 0,3 | 7 | 0,7 |
| 6 | \|\| | 2 | 0,2 | 9 | 0,9 |
| 7 | \| | 1 | 0,1 | 10 | 1,0 |
| | | $n = 10$ | 1,0 | | |

→ addieren $h(a_j)/n$

$1{,}0 \hat{=} 100\%$

# Aufgabe 2 (Zusätzliche Aufgaben, Aufgabe 1, gekürzt)

## Aufgabe 1

20 Beschäftigte wurden nach der Entfernung ihres Wohnsitzes zum Arbeitsplatz (in km) befragt. Das Ergebnis der Befragung ist in nachfolgender Tabelle wiedergegeben:

| **X: „Entfernung des Wohnsitzes zum Arbeitsplatz (in km)** | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| **absolute Häufigkeit** | 5 | 4 | 1 | 3 | 3 | 4 |

1. Wie lautet das statistische Merkmal und wie ist es skaliert?
2. Bestimmen Sie tabellarisch die relativen, die absoluten kumulierten und die relativen kumulierten Häufigkeiten!
3. Stellen Sie die absoluten Häufigkeiten grafisch dar!
4. Zeichnen Sie die empirische Verteilungsfunktion!
5. Berechnen Sie das arithmetische Mittel, und bestimmen Sie den Median der Häufigkeitsverteilung! Welchen Wert hat der Modus? (Denken Sie an die Angabe der Einheit!)
6. Bestimmen Sie die Spannweite der vorliegenden Häufigkeitsverteilung! (Denken Sie an die Angabe der Einheit!)
7. Berechnen Sie die empirische Varianz und die Standardabweichung! (Denken Sie an die Angabe der Einheit!)

Tabellen

grafiken ...

Lageparam.

Kennzahlen

→ Streuungsparameter

1) X: Entfernung des Wohnsitzes zum Arbeitsplatz (in km)

kardinalskaliert

| | $a_j$ | $h(a_j)$ | $f(a_j)$ | $H(x)$ | $F(x)$ | $a_j \cdot h(a_j)$ | $a_j^2$ | $a_j^2 \cdot h(a_j)$ |
|---|---|---|---|---|---|---|---|---|
| MIN | 2 | 5 | 0,25 | 5 | 0,25 | 10 | 4 | 20 |
| | 3 | 4 | 0,2 | 9 | 0,45 | 12 | 9 | 36 |
| | 4 | 1 | 0,05 | 10 | 0,5 | 4 | 16 | 16 |
| | 5 | 3 | 0,15 | 13 | 0,65 | 15 | 25 | 75 |
| | 6 | 3 | 0,15 | 16 | 0,8 | 18 | 36 | 108 |
| MAX | 7 | 4 | 0,2 | 20 | 1,0 | 28 | 49 | 196 |
| | | $n=20$ | 1,0 | | | 87 | | 451 |

$h(a_j)/n$

3) Grafik für absolute Häufigkeiten

↳ Stabdiagramm

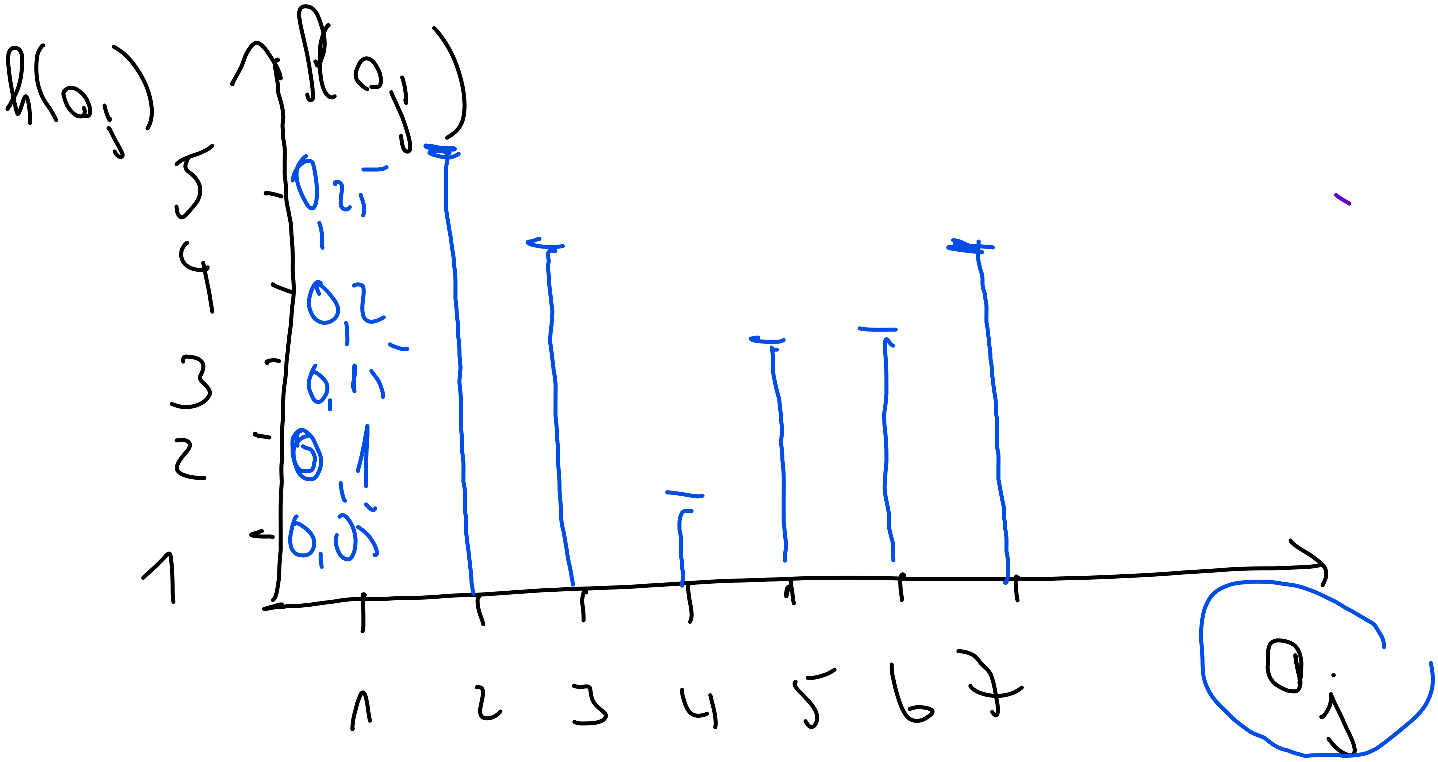
→ für kardinal / metrisch

isolierte Spitzen.

→ Maß für Häufigkeit

Fläche des Stabes

Über jeder Ausprägung $a_j$ wird ein Stab errichtet, dessen Höhe der absoluten/relativen H. entspricht.



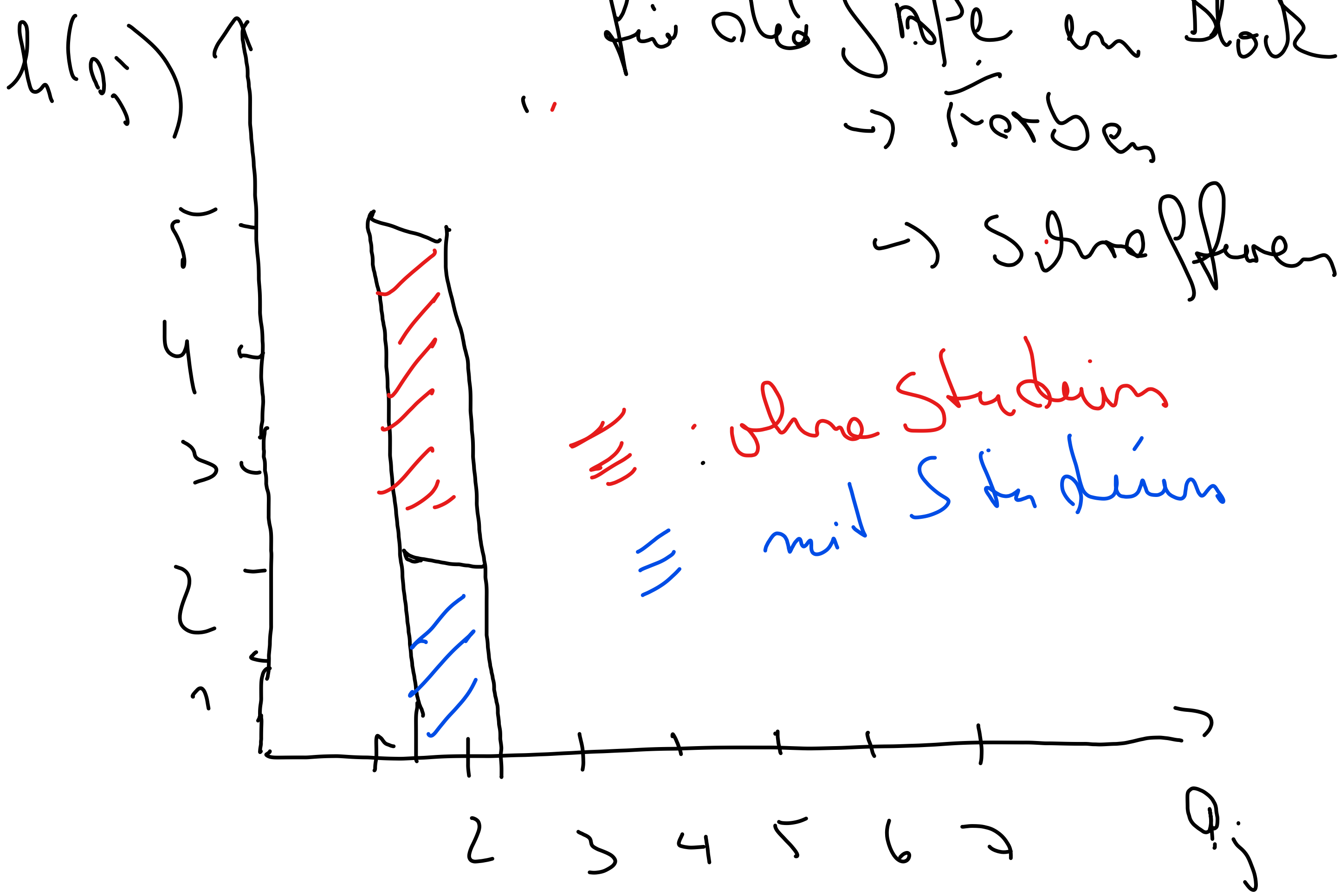

keine Flächen einbauen.

mit Flächen, Blockdiagramm

gleichzeitige Darstellung von

2 od. Größen ...

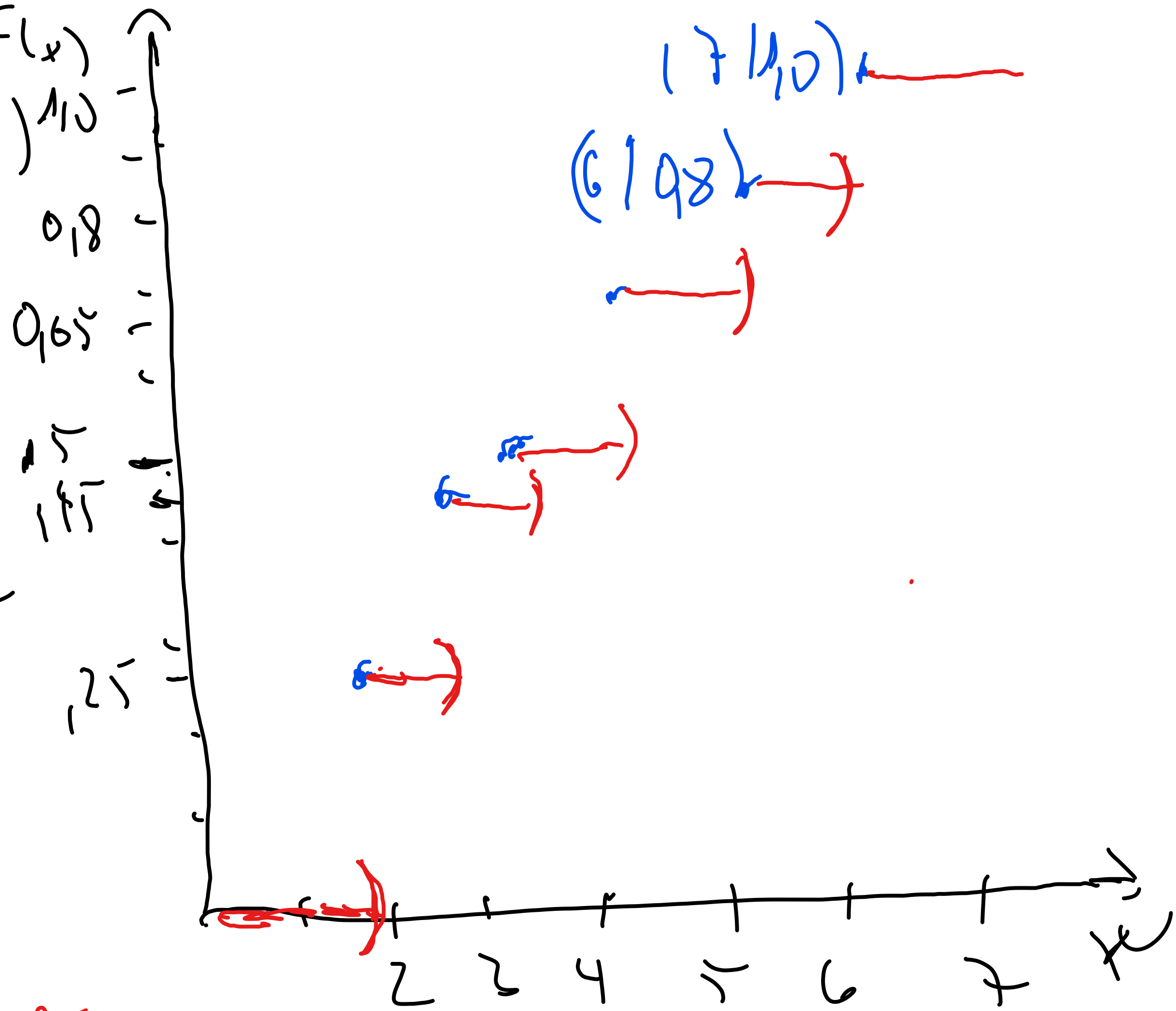
für die Größe im Block
→ Farben
→ Schraffieren




empirische Verteilungsfunktion
→ Grafik für $F(x)$
→ Treppe!

8)

| x | F(x) |
|---|---|
| 2 | 0,25 |
| 3 | 0,45 |
| 4 | 0,5 |
| 5 | 0,65 |
| 6 | 0,8 |
| 7 | 1,0 |




) runde Klammer

5) bis 7) <u>Maßzahlen</u>

<u>Lageparameter</u>

Klassi

→ ke

- arithmetisches Mittel $\bar{x}$
- Median $x_{0,5}$

Lage-
para-
meter

1. Modus, Modalwert

## Streuungsparameter

- empirische Varianz $s^2$

Standardabweichung $s$

Streuungsparameter um $\bar{x}$

- Spannweite $SP$

Berechnung in den Notizen