

# Kapitel VI - Lage- und Streuungsparameter

## Deskriptive Statistik

Prof. Dr. W.-D. Heller  
Hartwig Senska  
Carlo Siebenschuh

# Agenda

- ① Ziele
- ② Lageparameter
- ③ Streuungsparameter
- ④ Nutzen & Grenzen von Lage- und Streuungsparametern
- ⑤ Graphische Darstellung

# Agenda

- ① **Ziele**
- ② Lageparameter
- ③ Streuungsparameter
- ④ Nutzen & Grenzen von Lage- und Streuungsparametern
- ⑤ Graphische Darstellung

# Ziele

- (1) Reduzierung der Daten auf markante Vergleichsgrößen
- (2) Festlegung von Kerngrößen einer Grundgesamtheit, die durch Stichproben geschätzt werden können  
→ Solche Größen nennt man **Parameter** oder **Kennzahlen** der Häufigkeitsverteilung.

## Lageparameter:

Kennzahl über die Lage der Beobachtungswerte, Kern der Verteilung

## Streuungsparameter

Kennzahl über die Streuung der Beobachtungswerte, zusammenfassende Beurteilung der Positionierung der einzelnen Werte zum Zentrum

# Ziele

- (1) Reduzierung der Daten auf markante Vergleichsgrößen
- (2) Festlegung von Kerngrößen einer Grundgesamtheit, die durch Stichproben geschätzt werden können  
→ Solche Größen nennt man **Parameter** oder **Kennzahlen** der Häufigkeitsverteilung.

## Lageparameter:

Kennzahl über die Lage der Beobachtungswerte, Kern der Verteilung

## Streuungsparameter

Kennzahl über die Streuung der Beobachtungswerte, zusammenfassende Beurteilung der Positionierung der einzelnen Werte zum Zentrum

# Ziele

- (1) Reduzierung der Daten auf markante Vergleichsgrößen
- (2) Festlegung von Kerngrößen einer Grundgesamtheit, die durch Stichproben geschätzt werden können  
→ Solche Größen nennt man **Parameter** oder **Kennzahlen** der Häufigkeitsverteilung.

## Lageparameter:

Kennzahl über die Lage der Beobachtungswerte, Kern der Verteilung

## Streuungsparameter

Kennzahl über die Streuung der Beobachtungswerte, zusammenfassende Beurteilung der Positionierung der einzelnen Werte zum Zentrum

# Ziele

- (1) Reduzierung der Daten auf markante Vergleichsgrößen
- (2) Festlegung von Kerngrößen einer Grundgesamtheit, die durch Stichproben geschätzt werden können  
→ Solche Größen nennt man **Parameter** oder **Kennzahlen** der Häufigkeitsverteilung.

## Lageparameter:

Kennzahl über die Lage der Beobachtungswerte, Kern der Verteilung

## Streuungsparameter

Kennzahl über die Streuung der Beobachtungswerte, zusammenfassende Beurteilung der Positionierung der einzelnen Werte zum Zentrum

# Ziele

- (1) Reduzierung der Daten auf markante Vergleichsgrößen
- (2) Festlegung von Kerngrößen einer Grundgesamtheit, die durch Stichproben geschätzt werden können  
→ Solche Größen nennt man **Parameter** oder **Kennzahlen** der Häufigkeitsverteilung.

## Lageparameter:

Kennzahl über die Lage der Beobachtungswerte, Kern der Verteilung

## Streuungsparameter

Kennzahl über die Streuung der Beobachtungswerte, zusammenfassende Beurteilung der Positionierung der einzelnen Werte zum Zentrum



# Ziele

- (1) Reduzierung der Daten auf markante Vergleichsgrößen
- (2) Festlegung von Kerngrößen einer Grundgesamtheit, die durch Stichproben geschätzt werden können  
→ Solche Größen nennt man **Parameter** oder **Kennzahlen** der Häufigkeitsverteilung.

## Lageparameter:

Kennzahl über die Lage der Beobachtungswerte, Kern der Verteilung

## Streuungsparameter

Kennzahl über die Streuung der Beobachtungswerte, zusammenfassende Beurteilung der Positionierung der einzelnen Werte zum Zentrum

# Agenda

- ① Ziele
- ② **Lageparameter**
- ③ Streuungsparameter
- ④ Nutzen & Grenzen von Lage- und Streuungsparametern
- ⑤ Graphische Darstellung

# Lageparameter

## Modus oder Modalwert:

Dieser „einfachste“ Lageparameter gibt die am häufigsten auftretende Merkmalsausprägung an. Er kann sowohl bei Rang- und quantitativen Merkmalen, als auch bei qualitativen Merkmalen bestimmt werden.

Die Merkmalsausprägung(en)  $a_M$  mit

$$h(a_M) = \max_{a \in M} h(a) \qquad (p(a_M) = \max_{a \in M} p(a))$$

heißt (heißen) **Modus** oder **Modalwert**. Liegen mehrere häufigste Werte vor, so auch mehrere Modalwerte.

Geht man von einer Urliste  $x_1, \dots, x_n$  aus, so verwendet man auch die Bezeichnung  $x_M$ , um den Zusammenhang deutlich zu machen.

# Lageparameter

## Modus oder Modalwert:

Dieser „einfachste“ Lageparameter gibt die am häufigsten auftretende Merkmalsausprägung an. Er kann sowohl bei Rang- und quantitativen Merkmalen, als auch bei qualitativen Merkmalen bestimmt werden.

Die Merkmalsausprägung(en)  $a_M$  mit

$$h(a_M) = \max_{a \in M} h(a) \qquad (p(a_M) = \max_{a \in M} p(a))$$

heißt (heißen) **Modus** oder **Modalwert**. Liegen mehrere häufigste Werte vor, so auch mehrere Modalwerte.

Geht man von einer Urliste  $x_1, \dots, x_n$  aus, so verwendet man auch die Bezeichnung  $x_M$ , um den Zusammenhang deutlich zu machen.

# Lageparameter

## Modus oder Modalwert:

Dieser „einfachste“ Lageparameter gibt die am häufigsten auftretende Merkmalsausprägung an. Er kann sowohl bei Rang- und quantitativen Merkmalen, als auch bei qualitativen Merkmalen bestimmt werden.

Die Merkmalsausprägung(en)  $a_M$  mit

$$h(a_M) = \max_{a \in M} h(a) \qquad (p(a_M) = \max_{a \in M} p(a))$$

heißt (heißen) **Modus** oder **Modalwert**. Liegen mehrere häufigste Werte vor, so auch mehrere Modalwerte.

Geht man von einer Urliste  $x_1, \dots, x_n$  aus, so verwendet man auch die Bezeichnung  $x_M$ , um den Zusammenhang deutlich zu machen.

# Lageparameter

## Modus oder Modalwert:

Dieser „einfachste“ Lageparameter gibt die am häufigsten auftretende Merkmalsausprägung an. Er kann sowohl bei Rang- und quantitativen Merkmalen, als auch bei qualitativen Merkmalen bestimmt werden.

Die Merkmalsausprägung(en)  $a_M$  mit

$$h(a_M) = \max_{a \in M} h(a) \qquad (p(a_M) = \max_{a \in M} p(a))$$

heißt (heißen) **Modus** oder **Modalwert**. Liegen mehrere häufigste Werte vor, so auch mehrere Modalwerte.

Geht man von einer Urliste  $x_1, \dots, x_n$  aus, so verwendet man auch die Bezeichnung  $x_M$ , um den Zusammenhang deutlich zu machen.

# Lageparameter

## Zu Beispiel 4.2:

Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

Schüler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
Punktzahl	8	14	9	13	8	12	9	11	12	9	12	14
Schüler	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$	$s_{18}$	$s_{19}$	$s_{20}$				
Punktzahl	10	12	9	7	11	12	13	9				

Die zugehörige Urliste lautet:

8, 14, 9, 13, 8, 12, 9, 11, 12, 9, 12, 14, 10, 12, 9, 7, 11, 12, 13, 9.

Es gibt also zwei Modalwerte, da die Punktzahlen 9 und 12 übereinstimmend maximale absolute Häufigkeit aufweisen.

# Lageparameter

## Zu Beispiel 4.2:

Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

Schüler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
Punktzahl	8	14	9	13	8	12	9	11	12	9	12	14
Schüler	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$	$s_{18}$	$s_{19}$	$s_{20}$				
Punktzahl	10	12	9	7	11	12	13	9				

Die zugehörige Urliste lautet:

8, 14, 9, 13, 8, 12, 9, 11, 12, 9, 12, 14, 10, 12, 9, 7, 11, 12, 13, 9.

Es gibt also zwei Modalwerte, da die Punktzahlen 9 und 12 übereinstimmend maximale absolute Häufigkeit aufweisen.



# Lageparameter

## Zu Beispiel 4.2:

Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

Schüler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
Punktzahl	8	14	9	13	8	12	9	11	12	9	12	14
Schüler	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$	$s_{18}$	$s_{19}$	$s_{20}$				
Punktzahl	10	12	9	7	11	12	13	9				

Die zugehörige Urliste lautet:

8, 14, 9, 13, 8, 12, 9, 11, 12, 9, 12, 14, 10, 12, 9, 7, 11, 12, 13, 9.

**Es gibt also zwei Modalwerte, da die Punktzahlen 9 und 12 übereinstimmend maximale absolute Häufigkeit aufweisen.**

# Lageparameter

## Zu Beispiel 4.4:

Die Spieler eines Bundesligavereins (17 Spieler) wurden nach ihrem letzten Schul- bzw. Hochschulabschluss befragt. Dabei bedeuten: H Hauptschule, B Berufsschule, R Realschule, A Abitur, FO Fachoberschule, FH Fachhochschule, U Universität:

Spieler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$
Abschluss	A	H	H	H	U	R	B	H	B	B
Spieler	$s_{11}$	$s_{12}$	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$			
Abschluss	R	R	FO	H	A	A	R			

Die sortierte Urliste lautet:

$$A(3), H(5), U(1), R(4), B(3), FO(1), (FH(0))$$

Man erhält als Modalwert die Merkmalsausprägung Hauptschulabschluss (H), dessen absolute Häufigkeit mit 5 am größten ist.

# Lageparameter

## Zu Beispiel 4.4:

Die Spieler eines Bundesligavereins (17 Spieler) wurden nach ihrem letzten Schul- bzw. Hochschulabschluss befragt. Dabei bedeuten: H Hauptschule, B Berufsschule, R Realschule, A Abitur, FO Fachoberschule, FH Fachhochschule, U Universität:

Spieler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$
Abschluss	A	H	H	H	U	R	B	H	B	B
Spieler	$s_{11}$	$s_{12}$	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$			
Abschluss	R	R	FO	H	A	A	R			

Die sortierte Urliste lautet:

$$A(3), H(5), U(1), R(4), B(3), FO(1), (FH(0))$$

Man erhält als Modalwert die Merkmalsausprägung Hauptschulabschluss (H), dessen absolute Häufigkeit mit 5 am größten ist.

# Lageparameter

## Zu Beispiel 4.4:

Die Spieler eines Bundesligavereins (17 Spieler) wurden nach ihrem letzten Schul- bzw. Hochschulabschluss befragt. Dabei bedeuten: H Hauptschule, B Berufsschule, R Realschule, A Abitur, FO Fachoberschule, FH Fachhochschule, U Universität:

Spieler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$
Abschluss	A	H	H	H	U	R	B	H	B	B
Spieler	$s_{11}$	$s_{12}$	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$			
Abschluss	R	R	FO	H	A	A	R			

Die sortierte Urliste lautet:

$$A(3), H(5), U(1), R(4), B(3), FO(1), (FH(0))$$

Man erhält als Modalwert die Merkmalsausprägung Hauptschulabschluss (H), dessen absolute Häufigkeit mit 5 am größten ist.

# Lageparameter

Anmerkung: Bei klassierten Daten sollte bei einer Beurteilung der Klassen bezüglich ihrer Häufigkeit die Klassenbreite, also das Potenzial, einen Merkmalswert zu enthalten, berücksichtigt werden. **Modale Klasse** ist danach die Klasse mit der größten Häufigkeitsdichte.

# Lageparameter

**Zu Beispiel 5.5:** Für eine Verbrauchsstudie wurden die Nettojahreseinkommen von 100 Männern und 150 Frauen festgestellt.

Einkommen in TEuro	Männer		Frauen		Häufigkeitsdichte	
	abs. H.	rel. H.	abs. H.	rel. H.	abs. Männer	rel. Frauen
0 bis unter 10	5	0.05	9	0.06	0.5	0.006
10 bis unter 20	15	0.15	30	0.20	1.5	0.02
20 bis unter 25	20	0.20	42	0.28	4.0	0.056
25 bis unter 30	25	0.25	48	0.32	5.0	0.064
30 bis unter 40	20	0.20	15	0.10	2.0	0.01
40 bis unter 60	10	0.10	6	0.04	0.5	0.002
60 bis unter 85	5	0.05	0	0	0.2	0
$\Sigma$	100	1	150	1		

Die Modale Klasse ist bei den Frauen wie auch bei den Männern die Klasse von 25 bis unter 30.

# Lageparameter

## Zentralwert oder Median

Der Zentralwert oder Median teilt die geordnete statistische Reihe in der Mitte und damit in zwei gleichgroße Teile. Dieser Lageparameter setzt also voraus, dass die Merkmalsausprägungen zumindest ordinalskaliert sind.

Sei  $x_{(1)}, \dots, x_{(n)}$  die geordnete statistische Reihe, so ist der **Zentralwert** oder **Median**  $a_z$  (bzw.  $x_z$ , s.o.) definiert durch:

(1)  $n$  ungerade:

$$a_z = x_{\left(\frac{n+1}{2}\right)}$$

# Lageparameter

## Zentralwert oder Median

Der Zentralwert oder Median teilt die geordnete statistische Reihe in der Mitte und damit in zwei gleichgroße Teile. Dieser Lageparameter setzt also voraus, dass die Merkmalsausprägungen zumindest ordinalskaliert sind.

Sei  $x_{(1)}, \dots, x_{(n)}$  die geordnete statistische Reihe, so ist der **Zentralwert** oder **Median**  $a_z$  (bzw.  $x_z$ , s.o.) definiert durch:

(1)  $n$  ungerade:

$$a_z = x_{\left(\frac{n+1}{2}\right)}$$



# Lageparameter

## Zentralwert oder Median

Der Zentralwert oder Median teilt die geordnete statistische Reihe in der Mitte und damit in zwei gleichgroße Teile. Dieser Lageparameter setzt also voraus, dass die Merkmalsausprägungen zumindest ordinalskaliert sind.

Sei  $x_{(1)}, \dots, x_{(n)}$  die geordnete statistische Reihe, so ist der **Zentralwert** oder **Median**  $a_z$  (bzw.  $x_z$ , s.o.) definiert durch:

(1)  $n$  ungerade:

$$a_z = x_{\left(\frac{n+1}{2}\right)}$$

# Lageparameter

- (2)  $n$  gerade: Die erste Hälfte besteht aus den Beobachtungswerten

$$x_{(1)}, \dots, x_{(\frac{n}{2})},$$

die zweite Hälfte aus

$$x_{(\frac{n}{2}+1)}, \dots, x_{(n)}.$$

Fallunterscheidung:

- $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  : „Der Zentralwert  $a_z$  liegt zwischen  $x_{(\frac{n}{2})}$  und  $x_{(\frac{n}{2}+1)}$ “.

Bei Rangmerkmalen kann dies nicht weiter präzisiert werden; bei quantitativen Merkmalen setzt man

$$a_z = 0.5 \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}).$$

- $x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}$  :

$$a_z = x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}.$$

# Lageparameter

- (2)  $n$  gerade: Die erste Hälfte besteht aus den Beobachtungswerten

$$x_{(1)}, \dots, x_{(\frac{n}{2})},$$

die zweite Hälfte aus

$$x_{(\frac{n}{2}+1)}, \dots, x_{(n)}.$$

Fallunterscheidung:

- $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  : „Der Zentralwert  $a_z$  liegt zwischen  $x_{(\frac{n}{2})}$  und  $x_{(\frac{n}{2}+1)}$ “.

Bei Rangmerkmalen kann dies nicht weiter präzisiert werden; bei quantitativen Merkmalen setzt man

$$a_z = 0.5 \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}).$$

- $x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}$  :

$$a_z = x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}.$$

# Lageparameter

- (2)  $n$  gerade: Die erste Hälfte besteht aus den Beobachtungswerten

$$x_{(1)}, \dots, x_{(\frac{n}{2})},$$

die zweite Hälfte aus

$$x_{(\frac{n}{2}+1)}, \dots, x_{(n)}.$$

Fallunterscheidung:

- $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  : „Der Zentralwert  $a_z$  liegt zwischen  $x_{(\frac{n}{2})}$  und  $x_{(\frac{n}{2}+1)}$ “.

Bei Rangmerkmalen kann dies nicht weiter präzisiert werden; bei quantitativen Merkmalen setzt man

$$a_z = 0.5 \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}).$$

- $x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}$  :

$$a_z = x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}.$$

# Lageparameter

- (2)  $n$  gerade: Die erste Hälfte besteht aus den Beobachtungswerten

$$x_{(1)}, \dots, x_{(\frac{n}{2})},$$

die zweite Hälfte aus

$$x_{(\frac{n}{2}+1)}, \dots, x_{(n)}.$$

Fallunterscheidung:

- $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  : „Der Zentralwert  $a_z$  liegt zwischen  $x_{(\frac{n}{2})}$  und  $x_{(\frac{n}{2}+1)}$ “.

Bei Rangmerkmalen kann dies nicht weiter präzisiert werden; bei quantitativen Merkmalen setzt man

$$a_z = 0.5 \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}).$$

- $x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}$  :

$$a_z = x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}.$$

# Lageparameter

- (2)  $n$  gerade: Die erste Hälfte besteht aus den Beobachtungswerten

$$x_{(1)}, \dots, x_{(\frac{n}{2})},$$

die zweite Hälfte aus

$$x_{(\frac{n}{2}+1)}, \dots, x_{(n)}.$$

Fallunterscheidung:

- $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  : „Der Zentralwert  $a_z$  liegt zwischen  $x_{(\frac{n}{2})}$  und  $x_{(\frac{n}{2}+1)}$ “.

Bei Rangmerkmalen kann dies nicht weiter präzisiert werden; bei quantitativen Merkmalen setzt man

$$a_z = 0.5 \cdot (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}).$$

- $x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}$  :

$$a_z = x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}.$$

# Lageparameter

## Beispiel 6.1:

- (1) Die geordnete statistische Reihe einer Altersbefragung von 11 Personen laute: 5, 17, 23, 27, 35, 42, 43, 48, 57, 64, 70. Man erhält somit den Zentralwert  $a_z = x_{(6)} = 42$ .
- (2) Bei einer Untersuchung über die Kinderzahl von Haushalten erhielt man folgende Häufigkeitstabelle (alternativ: Häufigkeiten):

Kinderzahl	Häufigkeit
0	4
1	14
2	10
3	6
4	5
5	1
6	0
	$\Sigma = 40$

Man erhält aus der Häufigkeitstabelle die Werte  $x_{(20)} = 2$  und  $x_{(21)} = 2$ . Also ist  $a_z = 2$  der Median der Verteilung.

# Lageparameter

## Beispiel 6.1:

- (1) Die geordnete statistische Reihe einer Altersbefragung von 11 Personen laute: 5, 17, 23, 27, 35, 42, 43, 48, 57, 64, 70. Man erhält somit den Zentralwert  $a_z = x_{(6)} = 42$ .
- (2) Bei einer Untersuchung über die Kinderzahl von Haushalten erhielt man folgende Häufigkeitstabelle (alternativ: Häufigkeiten):

Kinderzahl	Häufigkeit
0	4
1	14
2	10
3	6
4	5
5	1
6	0
	$\Sigma = 40$

Man erhält aus der Häufigkeitstabelle die Werte  $x_{(20)} = 2$  und  $x_{(21)} = 2$ . Also ist  $a_z = 2$  der Median der Verteilung.



# Lageparameter

## Beispiel 6.1:

- (1) Die geordnete statistische Reihe einer Altersbefragung von 11 Personen laute: 5, 17, 23, 27, 35, 42, 43, 48, 57, 64, 70. Man erhält somit den Zentralwert  $a_z = x_{(6)} = 42$ .
- (2) Bei einer Untersuchung über die Kinderzahl von Haushalten erhielt man folgende Häufigkeitstabelle (alternativ: Häufigkeiten):

Kinderzahl	Häufigkeit
0	4
1	14
2	10
3	6
4	5
5	1
6	0
	$\Sigma = 40$

Man erhält aus der Häufigkeitstabelle die Werte  $x_{(20)} = 2$  und  $x_{(21)} = 2$ . Also ist  $a_z = 2$  der Median der Verteilung.

# Lageparameter

## Beispiel 6.1:

- (1) Die geordnete statistische Reihe einer Altersbefragung von 11 Personen laute: 5, 17, 23, 27, 35, 42, 43, 48, 57, 64, 70. Man erhält somit den Zentralwert  $a_z = x_{(6)} = 42$ .
- (2) Bei einer Untersuchung über die Kinderzahl von Haushalten erhielt man folgende Häufigkeitstabelle (alternativ: Häufigkeiten):

Kinderzahl	Häufigkeit
0	4
1	14
2	10
3	6
4	5
5	1
6	0
	$\Sigma = 40$

Man erhält aus der Häufigkeitstabelle die Werte  $x_{(20)} = 2$  und  $x_{(21)} = 2$ . Also ist  $a_z = 2$  der Median der Verteilung.

# Lageparameter

- (3) Zu Beispiel 4.5: Die Messung der Körpergröße von 20 Personen ergab folgende Urliste:

1.49, 1.87, 1.91, 1.53, 1.68, 1.75, 1.66, 1.82, 1.76, 1.80,  
1.92, 1.71, 1.77, 1.69, 1.57, 1.83, 1.84, 1.47, 1.79, 1.81.

Aus der geordneten Urliste erhält man  $x_{(10)} = 1.76$  und  $x_{(11)} = 1.77$ . Diese Werte können auch aus der empirischen Verteilungsfunktion ermittelt werden.  $x_{(10)}$  ist die Stelle, bei der erstmals  $F(x) = 0.5$ ,  $x_{(11)}$  die Stelle an der erstmals  $F(x) > 0.5$  gilt. Damit ist  $a_z = 1.765$ .

# Lageparameter

- (3) Zu Beispiel 4.5: Die Messung der Körpergröße von 20 Personen ergab folgende Urliste:

1.49, 1.87, 1.91, 1.53, 1.68, 1.75, 1.66, 1.82, 1.76, 1.80,  
1.92, 1.71, 1.77, 1.69, 1.57, 1.83, 1.84, 1.47, 1.79, 1.81.

Aus der geordneten Urliste erhält man  $x_{(10)} = 1.76$  und  $x_{(11)} = 1.77$ . Diese Werte können auch aus der empirischen Verteilungsfunktion ermittelt werden.  $x_{(10)}$  ist die Stelle, bei der erstmals  $F(x) = 0.5$ ,  $x_{(11)}$  die Stelle an der erstmals  $F(x) > 0.5$  gilt. Damit ist  $a_z = 1.765$ .

# Lageparameter

- (3) Zu Beispiel 4.5: Die Messung der Körpergröße von 20 Personen ergab folgende Urliste:

1.49, 1.87, 1.91, 1.53, 1.68, 1.75, 1.66, 1.82, 1.76, 1.80,  
1.92, 1.71, 1.77, 1.69, 1.57, 1.83, 1.84, 1.47, 1.79, 1.81.

Aus der geordneten Urliste erhält man  $x_{(10)} = 1.76$  und  $x_{(11)} = 1.77$ . Diese Werte können auch aus der empirischen Verteilungsfunktion ermittelt werden.  $x_{(10)}$  ist die Stelle, bei der erstmals  $F(x) = 0.5$ ,  $x_{(11)}$  die Stelle an der erstmals  $F(x) > 0.5$  gilt. Damit ist  $a_z = 1.765$ .

# Lageparameter

Bei der Ermittlung des Medians aus der empirischen Verteilungsfunktion müssen zwei Fälle unterschieden werden:

- (1)  $F$  nimmt den Wert 0.5 an: Dann ist  $n$  gerade mit  $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  und die kleinste Zahl  $x$  mit  $F(x) = 0.5$  ist  $x_{(\frac{n}{2})}$ :

$$x_{(\frac{n}{2})} = \min\{x \mid F(x) = 0.5\}.$$

Damit ist

$$x_{(\frac{n}{2}+1)} = \min\{x \mid F(x) > 0.5\}$$

und man erhält

$$x_z = 0.5 \cdot \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$

# Lageparameter

Bei der Ermittlung des Medians aus der empirischen Verteilungsfunktion müssen zwei Fälle unterschieden werden:

- (1)  $F$  nimmt den Wert 0.5 an: Dann ist  $n$  gerade mit  $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  und die kleinste Zahl  $x$  mit  $F(x) = 0.5$  ist  $x_{(\frac{n}{2})}$ :

$$x_{(\frac{n}{2})} = \min\{x \mid F(x) = 0.5\}.$$

Damit ist

$$x_{(\frac{n}{2}+1)} = \min\{x \mid F(x) > 0.5\}$$

und man erhält

$$x_z = 0.5 \cdot \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$

# Lageparameter

Bei der Ermittlung des Medians aus der empirischen Verteilungsfunktion müssen zwei Fälle unterschieden werden:

- (1)  $F$  nimmt den Wert 0.5 an: Dann ist  $n$  gerade mit  $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  und die kleinste Zahl  $x$  mit  $F(x) = 0.5$  ist  $x_{(\frac{n}{2})}$ :

$$x_{(\frac{n}{2})} = \min\{x \mid F(x) = 0.5\}.$$

Damit ist

$$x_{(\frac{n}{2}+1)} = \min\{x \mid F(x) > 0.5\}$$

und man erhält

$$x_z = 0.5 \cdot \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$



# Lageparameter

Bei der Ermittlung des Medians aus der empirischen Verteilungsfunktion müssen zwei Fälle unterschieden werden:

- (1)  $F$  nimmt den Wert 0.5 an: Dann ist  $n$  gerade mit  $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  und die kleinste Zahl  $x$  mit  $F(x) = 0.5$  ist  $x_{(\frac{n}{2})}$ :

$$x_{(\frac{n}{2})} = \min\{x \mid F(x) = 0.5\}.$$

Damit ist

$$x_{(\frac{n}{2}+1)} = \min\{x \mid F(x) > 0.5\}$$

und man erhält

$$x_z = 0.5 \cdot \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$

# Lageparameter

Bei der Ermittlung des Medians aus der empirischen Verteilungsfunktion müssen zwei Fälle unterschieden werden:

- (1)  $F$  nimmt den Wert 0.5 an: Dann ist  $n$  gerade mit  $x_{(\frac{n}{2})} \neq x_{(\frac{n}{2}+1)}$  und die kleinste Zahl  $x$  mit  $F(x) = 0.5$  ist  $x_{(\frac{n}{2})}$ :

$$x_{(\frac{n}{2})} = \min\{x \mid F(x) = 0.5\}.$$

Damit ist

$$x_{(\frac{n}{2}+1)} = \min\{x \mid F(x) > 0.5\}$$

und man erhält

$$x_z = 0.5 \cdot \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right).$$

# Lageparameter

- (2)  $F$  nimmt den Wert 0.5 nicht an:  $n$  ist dann gerade mit  $x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}$  oder ungerade und  $x_z$  ist die Stelle, an der erstmals der Wert 0.5 überschritten wird:

$$x_z = \min\{x \mid F(x) > 0.5\}$$

# Lageparameter

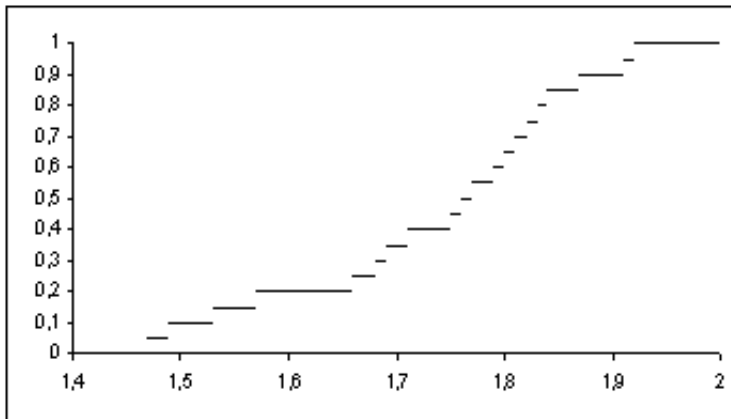
- (2)  $F$  nimmt den Wert 0.5 nicht an:  $n$  ist dann gerade mit  $x_{(\frac{n}{2})} = x_{(\frac{n}{2}+1)}$  oder ungerade und  $x_z$  ist die Stelle, an der erstmals der Wert 0.5 überschritten wird:

$$x_z = \min\{x \mid F(x) > 0.5\}$$

# Lageparameter

$F(x) =$	}	0	für	$x < 1.47$
		0.05	für	$1.47 \leq x < 1.49$
		0.1	für	$1.49 \leq x < 1.53$
		0.15	für	$1.53 \leq x < 1.57$
		0.2	für	$1.57 \leq x < 1.66$
		0.25	für	$1.66 \leq x < 1.68$
		0.3	für	$1.68 \leq x < 1.69$
		0.35	für	$1.69 \leq x < 1.71$
		0.4	für	$1.71 \leq x < 1.75$
		0.45	für	$1.75 \leq x < 1.76$
		0.5	für	$1.76 \leq x < 1.77$
		0.55	für	$1.77 \leq x < 1.79$
		0.6	für	$1.79 \leq x < 1.80$
		0.65	für	$1.80 \leq x < 1.81$
		0.7	für	$1.81 \leq x < 1.82$
		0.75	für	$1.82 \leq x < 1.83$
0.8	für	$1.83 \leq x < 1.84$		
0.85	für	$1.84 \leq x < 1.87$		
0.9	für	$1.87 \leq x < 1.91$		
0.95	für	$1.91 \leq x < 1.92$		
1	für	$1.92 \leq x$		

# Lageparameter



Zu Bsp. 4.5: Körpergröße

# Lageparameter

$\alpha$ -Quantil:  $(0 < \alpha < 1)$

Verallgemeinerung des Zentralwerts. Hierbei wird die geordnete Urliste nicht in der Mitte, sondern in den Proportionen  $\alpha$  zu  $1 - \alpha$  geteilt.

Für die genaue Festlegung muss man wieder unterscheiden, ob der „Teilungspfeil“ genau auf einen Merkmalswert (dann haben wir das  $\alpha$ -Quantil gefunden) oder auf eine Lücke trifft (dann muss das Mittel der Nachbarwerte genommen werden).

Als **oberes** und **unteres Quantil** wird das  $\alpha$ -Quantil für  $\alpha = 0.75$  bzw.  $\alpha = 0.25$  bezeichnet.

Anmerkung: Liegt ein klassiertes Merkmal vor und will man einen Zentralwert anhand der Häufigkeitsverteilung der Klassen bestimmen, so ist dies in der Regel nicht exakt möglich. Außerdem ist die bisherige Vorgehensweise nicht durchführbar.

# Lageparameter

$\alpha$ -Quantil:  $(0 < \alpha < 1)$

Verallgemeinerung des Zentralwerts. Hierbei wird die geordnete Urliste nicht in der Mitte, sondern in den Proportionen  $\alpha$  zu  $1 - \alpha$  geteilt.

Für die genaue Festlegung muss man wieder unterscheiden, ob der „Teilungspfeil“ genau auf einen Merkmalswert (dann haben wir das  $\alpha$ -Quantil gefunden) oder auf eine Lücke trifft (dann muss das Mittel der Nachbarwerte genommen werden).

Als **oberes** und **unteres Quantil** wird das  $\alpha$ -Quantil für  $\alpha = 0.75$  bzw.  $\alpha = 0.25$  bezeichnet.

Anmerkung: Liegt ein klassiertes Merkmal vor und will man einen Zentralwert anhand der Häufigkeitsverteilung der Klassen bestimmen, so ist dies in der Regel nicht exakt möglich. Außerdem ist die bisherige Vorgehensweise nicht durchführbar.



# Lageparameter

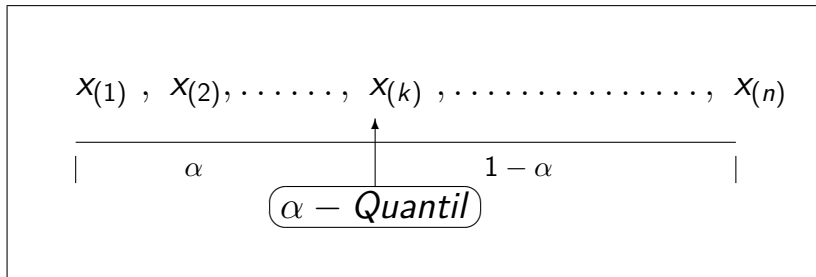


Abbildung 6.1  $\alpha$ -Quantil.

# Lageparameter

## Beispiel 6.2:

Bei den Altersangaben von 40 Beschäftigten einer Behörde erhält man als geordnete Urliste

16	17	18	19	25	26	26	27	28	28
28	29	29	35	36	36	37	37	37	37
38	39	39	42	44	45	46	46	47	48
48	55	57	58	59	59	61	62	62	63

Sei  $\alpha = 0.2$ . 20% von 40 ist 8. Der Teilungspfeil trifft also zwischen den achten und neunten Wert der geordneten Urliste. Dies sind die Werte 27 und 28. Damit ist

$$q_{0.20} = \frac{1}{2}(27 + 28) = 27.5.$$

Analog ist  $q_{0.25} = \frac{1}{2}(28 + 28) = 28$  und  $q_{0.75} = \frac{1}{2}(48 + 48) = 48$ .

# Lageparameter

## Beispiel 6.2:

Bei den Altersangaben von 40 Beschäftigten einer Behörde erhält man als geordnete Urliste

16	17	18	19	25	26	26	27	28	28
28	29	29	35	36	36	37	37	37	37
38	39	39	42	44	45	46	46	47	48
48	55	57	58	59	59	61	62	62	63

Sei  $\alpha = 0.2$ . 20% von 40 ist 8. Der Teilungspfeil trifft also zwischen den achten und neunten Wert der geordneten Urliste. Dies sind die Werte 27 und 28. Damit ist

$$q_{0.20} = \frac{1}{2}(27 + 28) = 27.5.$$

Analog ist  $q_{0.25} = \frac{1}{2}(28 + 28) = 28$  und  $q_{0.75} = \frac{1}{2}(48 + 48) = 48$ .

# Lageparameter

## Beispiel 6.2:

Bei den Altersangaben von 40 Beschäftigten einer Behörde erhält man als geordnete Urliste

16	17	18	19	25	26	26	27	28	28
28	29	29	35	36	36	37	37	37	37
38	39	39	42	44	45	46	46	47	48
48	55	57	58	59	59	61	62	62	63

Sei  $\alpha = 0.2$ . 20% von 40 ist 8. Der Teilungspfeil trifft also zwischen den achten und neunten Wert der geordneten Urliste. Dies sind die Werte 27 und 28. Damit ist

$$q_{0.20} = \frac{1}{2}(27 + 28) = 27.5.$$

Analog ist  $q_{0.25} = \frac{1}{2}(28 + 28) = 28$  und  $q_{0.75} = \frac{1}{2}(48 + 48) = 48$ .

# Lageparameter

Für die Berechnung aus der empirischen Verteilungsfunktion  $F$  erhält analog zum Zentralwert das  $\alpha$ -Quantil  $q_\alpha$ :

(1) Es gibt mindestens ein  $x$  mit  $F(x) = \alpha$ :

$$q_\alpha = \frac{1}{2} (\min\{x \mid F(x) = \alpha\} + \min\{x \mid F(x) > \alpha\})$$

(2) Es gibt kein  $x$  mit  $F(x) = \alpha$ :

$$q_\alpha = \min\{x \mid F(x) > \alpha\}$$

# Lageparameter

Für die Berechnung aus der empirischen Verteilungsfunktion  $F$  erhält analog zum Zentralwert das  $\alpha$ -Quantil  $q_\alpha$ :

(1) Es gibt mindestens ein  $x$  mit  $F(x) = \alpha$ :

$$q_\alpha = \frac{1}{2} (\min\{x \mid F(x) = \alpha\} + \min\{x \mid F(x) > \alpha\})$$

(2) Es gibt kein  $x$  mit  $F(x) = \alpha$ :

$$q_\alpha = \min\{x \mid F(x) > \alpha\}$$

# Lageparameter

Für die Berechnung aus der empirischen Verteilungsfunktion  $F$  erhält analog zum Zentralwert das  $\alpha$ -Quantil  $q_\alpha$ :

(1) Es gibt mindestens ein  $x$  mit  $F(x) = \alpha$ :

$$q_\alpha = \frac{1}{2} (\min\{x \mid F(x) = \alpha\} + \min\{x \mid F(x) > \alpha\})$$

(2) Es gibt kein  $x$  mit  $F(x) = \alpha$ :

$$q_\alpha = \min\{x \mid F(x) > \alpha\}$$

# Lageparameter

## (Feinberechner) Zentralwert:

Bei einem klassierten Merkmal mit  
Summenhäufigkeitsfunktion  $SF$  heißt  $a_z$  mit

$$SF(a_z) = \frac{1}{2}$$

(feinberechner) Zentralwert, falls  $a_z$  dadurch eindeutig  
festgelegt ist.

Die senkrechte Gerade durch diesen Wert teilt die Fläche  
unter dem Histogramm in zwei gleichgroße Teile.



# Lageparameter

## (Feinberechner) Zentralwert:

Bei einem klassierten Merkmal mit Summenhäufigkeitsfunktion  $SF$  heißt  $a_z$  mit

$$SF(a_z) = \frac{1}{2}$$

(feinberechner) Zentralwert, falls  $a_z$  dadurch eindeutig festgelegt ist.

Die senkrechte Gerade durch diesen Wert teilt die Fläche unter dem Histogramm in zwei gleichgroße Teile.

# Lageparameter

## (Feinberechner) Zentralwert:

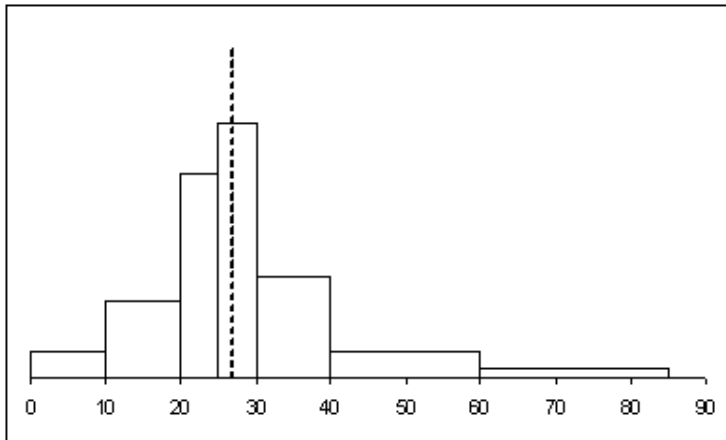
Bei einem klassierten Merkmal mit  
Summenhäufigkeitsfunktion  $SF$  heißt  $a_z$  mit

$$SF(a_z) = \frac{1}{2}$$

(feinberechner) Zentralwert, falls  $a_z$  dadurch eindeutig  
festgelegt ist.

Die senkrechte Gerade durch diesen Wert teilt die Fläche  
unter dem Histogramm in zwei gleichgroße Teile.

# Lageparameter



# Lageparameter

## Beispiel 6.3:

Bei den 30 Spielen einer Landesligamannschaft wurden folgende Zuschauerzahlen beobachtet:

14, 428, 643, 189, 206, 25, 93, 100, 612, 357,  
562, 120, 269, 309, 39, 78, 45, 142, 527, 407,  
163, 129, 44, 78, 26, 206, 345, 570, 525, 234.

Es werden folgende Klassen gebildet:

1 bis 100, 101 bis 200 usw.

# Lageparameter

## Beispiel 6.3:

Bei den 30 Spielen einer Landesligamannschaft wurden folgende Zuschauerzahlen beobachtet:

14, 428, 643, 189, 206, 25, 93, 100, 612, 357,  
562, 120, 269, 309, 39, 78, 45, 142, 527, 407,  
163, 129, 44, 78, 26, 206, 345, 570, 525, 234.

Es werden folgende Klassen gebildet:

1 bis 100, 101 bis 200 usw.

# Lageparameter

## Beispiel 6.3:

Klasse Nr.	Klasse	Strichliste	absolute Häufigkeit	relative Häufigkeit
1	1 bis 100		10	33.3%
2	101 bis 200		5	16.7%
3	201 bis 300		4	13.3%
4	301 bis 400		3	10 %
5	401 bis 500		2	6.7%
6	501 bis 600		4	13.3%
7	601 bis 700		2	6.7%

Zentralwert ist die obere Klassengrenze der Klasse 2. Wegen der allgemeinen Richtlinien für die Festlegung der Klassengrenzen (siehe Kapitel IV) also  $x_z = 200.5$ . Legt man Wert auf eine einfache und übersichtliche Darstellung, so ist auch  $x_z = 200$  genau genug.

# Lageparameter

## Beispiel 6.3:

Klasse Nr.	Klasse	Strichliste	absolute Häufigkeit	relative Häufigkeit
1	1 bis 100		10	33.3%
2	101 bis 200		5	16.7%
3	201 bis 300		4	13.3%
4	301 bis 400		3	10 %
5	401 bis 500		2	6.7%
6	501 bis 600		4	13.3%
7	601 bis 700		2	6.7%

Zentralwert ist die obere Klassengrenze der Klasse 2. Wegen der allgemeinen Richtlinien für die Festlegung der Klassengrenzen (siehe Kapitel IV) also  $x_z = 200.5$ . Legt man Wert auf eine einfache und übersichtliche Darstellung, so ist auch  $x_z = 200$  genau genug.

# Lageparameter

## (Feinberechner) Zentralwert:

Liegt der Zentralwert **nicht** auf einer Klassengrenze, so ist zunächst die Klasse zu bestimmen, in die der Zentralwert fällt. Diese Klasse heißt **Einfallsklasse**. Sei  $I_E$  mit den Klassengrenzen  $\alpha_E$  und  $\beta_E$  diese Klasse, so ist sie durch die Eigenschaften

$$SF(\alpha_E) < \frac{1}{2} \quad \text{und} \quad SF(\beta_E) > \frac{1}{2}$$

bzw.

$$\sum_{I: \beta_I \leq \alpha_E} p(I) < \frac{1}{2} \quad \text{und} \quad \sum_{I: \alpha_I \geq \beta_E} p(I) < \frac{1}{2}$$

charakterisiert.



# Lageparameter

## (Feinberechner) Zentralwert:

Liegt der Zentralwert **nicht** auf einer Klassengrenze, so ist zunächst die Klasse zu bestimmen, in die der Zentralwert fällt. Diese Klasse heißt **Einfallsklasse**. Sei  $I_E$  mit den Klassengrenzen  $\alpha_E$  und  $\beta_E$  diese Klasse, so ist sie durch die Eigenschaften

$$SF(\alpha_E) < \frac{1}{2} \quad \text{und} \quad SF(\beta_E) > \frac{1}{2}$$

bzw.

$$\sum_{I: \beta_I \leq \alpha_E} p(I) < \frac{1}{2} \quad \text{und} \quad \sum_{I: \alpha_I \geq \beta_E} p(I) < \frac{1}{2}$$

charakterisiert.

# Lageparameter

## (Feinberechner) Zentralwert:

Wegen

$$SF(z) = F(\alpha_E) + \frac{z - \alpha_E}{\beta_E - \alpha_E} p(I_E) \text{ für } \alpha_E \leq z \leq \beta_E$$

erhält man aus  $SF(a_z) = \frac{1}{2}$

$$a_z = \alpha_E + \frac{\frac{1}{2} - F(\alpha_E)}{p(I_E)} (\beta_E - \alpha_E) = \alpha_E + \frac{\frac{1}{2} - F(\alpha_E)}{F(\beta_E) - F(\alpha_E)} (\beta_E - \alpha_E).$$

# Lageparameter

## (Feinberechner) Zentralwert:

Wegen

$$SF(z) = F(\alpha_E) + \frac{z - \alpha_E}{\beta_E - \alpha_E} p(I_E) \text{ für } \alpha_E \leq z \leq \beta_E$$

erhält man aus  $SF(a_z) = \frac{1}{2}$

$$a_z = \alpha_E + \frac{\frac{1}{2} - F(\alpha_E)}{p(I_E)} (\beta_E - \alpha_E) = \alpha_E + \frac{\frac{1}{2} - F(\alpha_E)}{F(\beta_E) - F(\alpha_E)} (\beta_E - \alpha_E).$$

# Lageparameter

## Beispiel 6.4:

Bei 40 Beschäftigten einer Behörde wurden folgende Altersangaben (in vollendeten Lebensjahren) ermittelt:

16, 17, 18, 19, 25, 26, 26, 27, 28, 28,  
28, 29, 29, 35, 36, 36, 37, 37, 37, 37,  
38, 39, 39, 42, 44, 45, 46, 46, 47, 48,  
48, 55, 57, 58, 59, 59, 61, 62, 62, 63.

Nach Klassierung ergibt sich die Häufigkeitstabelle:

Alter	abs. Häufigkeit	rel. Häufigkeit
0 bis unter 20	4	0.1
20 bis unter 30	9	0.225
30 bis unter 40	10	0.25
40 bis unter 50	8	0.2
50 bis unter 60	5	0.125
60 bis unter 65	4	0.1

# Lageparameter

## Beispiel 6.4:

Bei 40 Beschäftigten einer Behörde wurden folgende Altersangaben (in vollendeten Lebensjahren) ermittelt:

16, 17, 18, 19, 25, 26, 26, 27, 28, 28,  
28, 29, 29, 35, 36, 36, 37, 37, 37, 37,  
38, 39, 39, 42, 44, 45, 46, 46, 47, 48,  
48, 55, 57, 58, 59, 59, 61, 62, 62, 63.

Nach Klassierung ergibt sich die Häufigkeitstabelle:

Alter	abs. Häufigkeit	rel. Häufigkeit
0 bis unter 20	4	0.1
20 bis unter 30	9	0.225
30 bis unter 40	10	0.25
40 bis unter 50	8	0.2
50 bis unter 60	5	0.125
60 bis unter 65	4	0.1

# Lageparameter

Alter	abs. Häufigkeit	rel. Häufigkeit
0 bis unter 20	4	0.1
20 bis unter 30	9	0.225
30 bis unter 40	10	0.25
40 bis unter 50	8	0.2
50 bis unter 60	5	0.125
60 bis unter 65	4	0.1

Man erhält als Einfallsklasse  $I_E$  die Klasse 30 Jahre b.u. 40 Jahre. Ferner gilt  $F(\alpha_E) = 0.325$  und  $F(\beta_E) = 0.575$ . Damit ist

$$a_z = 30 + \frac{0.5 - 0.325}{0.25} \cdot 10 = 30 + \frac{7}{10} \cdot 10 = 37 \text{ Jahre.}$$

(Hinweis: Analog kann ein beliebiges  $\alpha$ -Quantil  $q_\alpha$  errechnet werden,  $\alpha \in (0,1)$ )

# Lageparameter

Alter	abs. Häufigkeit	rel. Häufigkeit
0 bis unter 20	4	0.1
20 bis unter 30	9	0.225
30 bis unter 40	10	0.25
40 bis unter 50	8	0.2
50 bis unter 60	5	0.125
60 bis unter 65	4	0.1

Man erhält als Einfallsklasse  $I_E$  die Klasse 30 Jahre b.u. 40 Jahre. Ferner gilt  $F(\alpha_E) = 0.325$  und  $F(\beta_E) = 0.575$ . Damit ist

$$a_z = 30 + \frac{0.5 - 0.325}{0.25} \cdot 10 = 30 + \frac{7}{10} \cdot 10 = 37 \text{ Jahre.}$$

(Hinweis: Analog kann ein beliebiges  $\alpha$  - Quantil  $q_\alpha$  errechnet werden,  $\alpha \in (0,1)$ )

# Lageparameter

## Beispiel 6.5:

Für eine Verbrauchsstudie wurden die Nettojahreseinkommen von 100 Männern und 150 Frauen festgestellt:

Einkommen in TEuro	Männer		Frauen	
	abs. H.	rel. H.	abs. H.	rel. H.
unter 10	5	0.05	9	0.06
10 bis unter 20	15	0.15	30	0.20
20 bis unter 25	20	0.20	42	0.28
25 bis unter 30	25	0.25	48	0.32
30 bis unter 40	20	0.20	15	0.10
40 bis unter 60	10	0.10	6	0.04
über 60	5	0.05	0	0
$\Sigma$	100	1	150	1



# Lageparameter

## Beispiel 6.5:

Für eine Verbrauchsstudie wurden die Nettojahreseinkommen von 100 Männern und 150 Frauen festgestellt:

Einkommen in TEuro	Männer		Frauen	
	abs. H.	rel. H.	abs. H.	rel. H.
unter 10	5	0.05	9	0.06
10 bis unter 20	15	0.15	30	0.20
20 bis unter 25	20	0.20	42	0.28
25 bis unter 30	25	0.25	48	0.32
30 bis unter 40	20	0.20	15	0.10
40 bis unter 60	10	0.10	6	0.04
über 60	5	0.05	0	0
$\Sigma$	100	1	150	1

# Lageparameter

Bei den Einkommensverteilungen erhält man für die Männer:  
 $I_E = [25, 30)$  und damit

$$a_z = 25 + \frac{0.5 - 0.4}{0.25} \cdot 5 = 27 \text{ TEuro,}$$

und für die Frauen:  $I_E = [20, 25)$  und damit

$$a_z = 20 + \frac{0.5 - 0.26}{0.28} \cdot 5 = 24.29 \text{ TEuro.}$$

# Lageparameter

Bei den Einkommensverteilungen erhält man für die Männer:  
 $I_E = [25, 30)$  und damit

$$a_z = 25 + \frac{0.5 - 0.4}{0.25} \cdot 5 = 27 \text{ TEuro,}$$

und für die Frauen:  $I_E = [20, 25)$  und damit

$$a_z = 20 + \frac{0.5 - 0.26}{0.28} \cdot 5 = 24.29 \text{ TEuro.}$$

# Lageparameter

Bei den Einkommensverteilungen erhält man für die Männer:  
 $I_E = [25, 30)$  und damit

$$a_z = 25 + \frac{0.5 - 0.4}{0.25} \cdot 5 = 27 \text{ TEuro,}$$

und für die Frauen:  $I_E = [20, 25)$  und damit

$$a_z = 20 + \frac{0.5 - 0.26}{0.28} \cdot 5 = 24.29 \text{ TEuro.}$$

# Lageparameter

## Arithmetisches Mittel:

Im Gegensatz zum Zentralwert hat bei diesem Lageparameter die genaue Lage aller Merkmalswerte einen Einfluss. Beim Zentralwert können fast alle rechts liegenden Werte sehr weit von diesem entfernt sein, die links von ihm liegenden sehr nahe bei dem Zentralwert liegen oder umgekehrt, ohne diesen dabei zu beeinflussen. Sei  $x_1, \dots, x_n$  die statistische Reihe (Urliste) einer statistischen Untersuchung.

Das arithmetische Mittel der statistischen Reihe ist definiert durch

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$$

# Lageparameter

## Arithmetisches Mittel:

Im Gegensatz zum Zentralwert hat bei diesem Lageparameter die genaue Lage aller Merkmalswerte einen Einfluss. Beim Zentralwert können fast alle rechts liegenden Werte sehr weit von diesem entfernt sein, die links von ihm liegenden sehr nahe bei dem Zentralwert liegen oder umgekehrt, ohne diesen dabei zu beeinflussen. Sei  $x_1, \dots, x_n$  die statistische Reihe (Urliste) einer statistischen Untersuchung.

Das arithmetische Mittel der statistischen Reihe ist definiert durch

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$$

# Lageparameter

## Beispiel 6.6:

Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

Schüler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
Punktzahl	8	14	9	13	8	12	9	11	12	9	12	14
Schüler	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$	$s_{18}$	$s_{19}$	$s_{20}$				
Punktzahl	10	12	9	7	11	12	13	9				

Zugehörige Urliste lautet also:

8, 14, 9, 13, 8, 12, 9, 11, 12, 9, 12, 14, 10, 12, 9, 7, 11, 12, 13, 9.

Das arithmetische Mittel lautet

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = \frac{1}{20} \cdot 214 = 10.7$$

# Lageparameter

## Beispiel 6.6:

Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

Schüler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
Punktzahl	8	14	9	13	8	12	9	11	12	9	12	14
Schüler	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$	$s_{18}$	$s_{19}$	$s_{20}$				
Punktzahl	10	12	9	7	11	12	13	9				

Zugehörige Urliste lautet also:

8, 14, 9, 13, 8, 12, 9, 11, 12, 9, 12, 14, 10, 12, 9, 7, 11, 12, 13, 9.

Das arithmetische Mittel lautet

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = \frac{1}{20} \cdot 214 = 10.7$$



# Lageparameter

## Beispiel 6.6:

Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

Schüler	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$
Punktzahl	8	14	9	13	8	12	9	11	12	9	12	14
Schüler	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$	$s_{17}$	$s_{18}$	$s_{19}$	$s_{20}$				
Punktzahl	10	12	9	7	11	12	13	9				

Zugehörige Urliste lautet also:

8, 14, 9, 13, 8, 12, 9, 11, 12, 9, 12, 14, 10, 12, 9, 7, 11, 12, 13, 9.

Das arithmetische Mittel lautet

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = \frac{1}{20} \cdot 214 = 10.7$$

# Lageparameter

## **Gewichtetes arithmetisches Mittel der Merkmalausprägungen:**

Das arithmetische Mittel lässt auch aus der absoluten bzw. relativen Häufigkeitsverteilung berechnen. Die Gewichte sind dabei die absoluten bzw. relativen Häufigkeiten.

Absolute Häufigkeitsverteilung:

$$\bar{x} = \frac{1}{n} \sum_{a \in M} a \cdot h(a)$$

Relative Häufigkeitsverteilung:

$$\bar{x} = \sum_{a \in M} a \cdot p(a)$$

# Lageparameter

## **Gewichtetes arithmetisches Mittel der Merkmalausprägungen:**

Das arithmetische Mittel lässt auch aus der absoluten bzw. relativen Häufigkeitsverteilung berechnen. Die Gewichte sind dabei die absoluten bzw. relativen Häufigkeiten.

Absolute Häufigkeitsverteilung:

$$\bar{x} = \frac{1}{n} \sum_{a \in M} a \cdot h(a)$$

Relative Häufigkeitsverteilung:

$$\bar{x} = \sum_{a \in M} a \cdot p(a)$$

# Lageparameter

## Beispiel für gewichtete arithmetische Mittel (Bsp. 6.6: Leistungskurs Mathe):

Punktzahl	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Absolute H.	0	0	0	0	0	0	0	1	2	5	1	2	5	2	2	0

$$\bar{x} = \frac{1}{20}(7 \cdot 1 + 8 \cdot 2 + 9 \cdot 5 + 10 \cdot 1 + 11 \cdot 2 + 12 \cdot 5 + 13 \cdot 2 + 14 \cdot 2) = \frac{1}{20} \cdot 214$$

## In Beispiel 6.7 (Kinderzahl): Relative Häufigkeitsverteilung

Kinderzahl	0	1	2	3	4	5	6
Relative H.	.10	.35	.25	.15	.125	.025	0

$$\bar{x} = 0 \cdot 0.10 + 1 \cdot 0.35 + 2 \cdot 0.25 + 3 \cdot 0.15 + 4 \cdot 0.125 + 5 \cdot 0.025 + 6 \cdot 0 = 1.925.$$

# Lageparameter

## Beispiel für gewichtete arithmetische Mittel (Bsp. 6.6: Leistungskurs Mathe):

Punktzahl	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Absolute H.	0	0	0	0	0	0	0	1	2	5	1	2	5	2	2	0

$$\bar{x} = \frac{1}{20}(7 \cdot 1 + 8 \cdot 2 + 9 \cdot 5 + 10 \cdot 1 + 11 \cdot 2 + 12 \cdot 5 + 13 \cdot 2 + 14 \cdot 2) = \frac{1}{20} \cdot 214$$

## In Beispiel 6.7 (Kinderzahl): Relative Häufigkeitsverteilung

Kinderzahl	0	1	2	3	4	5	6
Relative H.	.10	.35	.25	.15	.125	.025	0

$$\bar{x} = 0 \cdot 0.10 + 1 \cdot 0.35 + 2 \cdot 0.25 + 3 \cdot 0.15 + 4 \cdot 0.125 + 5 \cdot 0.025 + 6 \cdot 0 = 1.925.$$

# Lageparameter

**Beispiel für gewichtete arithmetische Mittel (Bsp. 6.6: Leistungskurs Mathe):**

Punktzahl	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Absolute H.	0	0	0	0	0	0	0	1	2	5	1	2	5	2	2	0

$$\bar{x} = \frac{1}{20}(7 \cdot 1 + 8 \cdot 2 + 9 \cdot 5 + 10 \cdot 1 + 11 \cdot 2 + 12 \cdot 5 + 13 \cdot 2 + 14 \cdot 2) = \frac{1}{20} \cdot 214$$

**In Beispiel 6.7 (Kinderzahl): Relative Häufigkeitsverteilung**

Kinderzahl	0	1	2	3	4	5	6
Relative H.	.10	.35	.25	.15	.125	.025	0

$$\bar{x} = 0 \cdot 0.10 + 1 \cdot 0.35 + 2 \cdot 0.25 + 3 \cdot 0.15 + 4 \cdot 0.125 + 5 \cdot 0.025 + 6 \cdot 0 = 1.925.$$

# Lageparameter

## Arithmetisches Mittel bei klassierten Merkmalen:

Bei einem klassierten Merkmal ist - wie schon beim Zentralwert - der exakte Wert aus der Häufigkeitsverteilung nicht zu bestimmen. Die unbekannt Merkmalswerte in den Klassen müssen durch einen Stellvertreter (**Repräsentanten**) ersetzt werden. Dabei verwendet man die Klassenmitten, wenn keine wesentlichen Gründe dagegensprechen.

Sei also jeweils  $z_l$  die Klassenmitte der Klasse  $l$ , so ist das arithmetische Mittel eines klassierten Merkmals definiert durch

$$\bar{x} = \frac{1}{n} \sum_l z_l h(l) = \sum_l z_l p(l) \quad \text{mit} \quad n = \sum_l h(l).$$

Anmerkung: Bei offenen Randklassen ist der Repräsentant jeder Randklasse in sinnvoller Weise festzulegen.

# Lageparameter

## Arithmetisches Mittel bei klassierten Merkmalen:

Bei einem klassierten Merkmal ist - wie schon beim Zentralwert - der exakte Wert aus der Häufigkeitsverteilung nicht zu bestimmen. Die unbekanntenen Merkmalswerte in den Klassen müssen durch einen Stellvertreter (**Repräsentanten**) ersetzt werden. Dabei verwendet man die Klassenmitten, wenn keine wesentlichen Gründe dagegensprechen.

Sei also jeweils  $z_l$  die Klassenmitte der Klasse  $l$ , so ist das arithmetische Mittel eines klassierten Merkmals definiert durch

$$\bar{x} = \frac{1}{n} \sum_l z_l h(l) = \sum_l z_l p(l) \quad \text{mit} \quad n = \sum_l h(l).$$

Anmerkung: Bei offenen Randklassen ist der Repräsentant jeder Randklasse in sinnvoller Weise festzulegen.



# Lageparameter

## Beispiel 6.8:

Klassierte Daten aus Beispiel 4.5:

Klasse	1 1.395 b.u. 1.495	2 1.495 b.u. 1.595	3 1.595 b.u. 1.695	4 1.695 b.u. 1.795	5 1.795 b.u. 1.895
Absolute H.	2	2	3	5	6
Relative H.	0.1	0.1	0.15	0.25	0.3
Klasse	6 1.895 b.u. 1.995	$\Sigma$			
Absolute H.	2	20			
Relative H.	0.1	1			

Das arithmetische Mittel lautet:

$$\begin{aligned}\bar{x} &= \frac{1}{20}(144.5 \cdot 2 + 154.5 \cdot 2 + 164.5 \cdot 3 + 174.5 \cdot 5 + 184.5 \cdot 6 + 194.5 \cdot 2) \\ &= \frac{1}{20} \cdot 3460 = 173 \text{ cm.}\end{aligned}$$

Bei Verwendung der unkorrekten Klassenmitten 145, 155, ... erhält man 173.5, was als Näherung i.a. ausreichend ist; auch 173 ist nur eine Näherung.

# Lageparameter

## Beispiel 6.8:

Klassierte Daten aus Beispiel 4.5:

Klasse	1 1.395 b.u. 1.495	2 1.495 b.u. 1.595	3 1.595 b.u. 1.695	4 1.695 b.u. 1.795	5 1.795 b.u. 1.895
Absolute H.	2	2	3	5	6
Relative H.	0.1	0.1	0.15	0.25	0.3
Klasse	6 1.895 b.u. 1.995	$\Sigma$			
Absolute H.	2	20			
Relative H.	0.1	1			

Das arithmetische Mittel lautet:

$$\begin{aligned}\bar{x} &= \frac{1}{20}(144.5 \cdot 2 + 154.5 \cdot 2 + 164.5 \cdot 3 + 174.5 \cdot 5 + 184.5 \cdot 6 + 194.5 \cdot 2) \\ &= \frac{1}{20} \cdot 3460 = 173 \text{ cm.}\end{aligned}$$

Bei Verwendung der unkorrekten Klassenmitten 145, 155, ... erhält man 173.5, was als Näherung i.a. ausreichend ist; auch 173 ist nur eine Näherung.

# Lageparameter

Anmerkung: Bei der Berechnung des arithmetischen Mittels aus klassierten Daten geht man also von der Hypothese aus, dass die Merkmalswerte in den Klassen durch die Klassenmitte ersetzt werden können. Den exakten Wert allerdings, d.h. das arithmetische Mittel aus der Urliste, erhält man bei dieser Berechnung lediglich dann, wenn die Klassenmitten das arithmetische Mittel der Beobachtungswerte innerhalb der Klassen sind. Dies ist normalerweise aber nicht der Fall. In Beispiel 6.4 (Beschäftigte) ist das arithmetische Mittel aus der Urliste 39.725, aus den klassierten Daten 38.25. (innerhalb Klassen sind Daten eher in der rechten Hälfte)

# Zusammenfassung - Lageparameter

Qual. Merkmal	Rangmerkmal	Quant. Merkmal	
			klassiert
Modalwert (Modus)	Modalwert (Modus)	Modalwert	Modale Klasse
	Zentralwert (Median)	Zentralwert	Feinberechneter Zentralwert
		Arithm. Mittel	Arithm. Mittel

# Mängel - Lageparameter

Durch die Angabe eines Lageparameters ist eine Verteilung offensichtlich nicht eindeutig beschrieben. Zwar gibt ein Mittelwert einen ersten Eindruck von der Verteilung, dennoch gibt es ersichtlich mehrere recht unterschiedliche Häufigkeitsverteilungen, die denselben Mittelwert besitzen. Ein oder mehrere Lageparameter können die Verteilung der Beobachtungswerte also nicht vollständig charakterisieren.

## Abhilfe:

Charakterisierung des Streuungsverhaltens

# Mängel - Lageparameter

Durch die Angabe eines Lageparameters ist eine Verteilung offensichtlich nicht eindeutig beschrieben. Zwar gibt ein Mittelwert einen ersten Eindruck von der Verteilung, dennoch gibt es ersichtlich mehrere recht unterschiedliche Häufigkeitsverteilungen, die denselben Mittelwert besitzen. Ein oder mehrere Lageparameter können die Verteilung der Beobachtungswerte also nicht vollständig charakterisieren.

## **Abhilfe:**

Charakterisierung des Streuungsverhaltens

# Mängel - Lageparameter

## Beispiel 6.9:

Wenn das durchschnittliche Alter einer Fußballmannschaft mit 22 Jahren angegeben ist, so weiß man zwar, dass es sich um eine junge Mannschaft handelt. Es ist jedoch keine Information darüber enthalten, ob nicht einige erfahrene sprich ältere Spieler mitspielen. Man erhält dasselbe Durchschnittsalter, wenn

- a) alle Spieler 22 Jahre alt sind,
- b) 8 Spieler 20 Jahre, 2 Spieler 27 und ein Spieler 28 Jahre alt ist,
- c) 10 Spieler 21 Jahre und 1 Spieler 32 Jahre alt ist,
- d) etc.

# Mängel - Lageparameter

## Beispiel 6.9:

Wenn das durchschnittliche Alter einer Fußballmannschaft mit 22 Jahren angegeben ist, so weiß man zwar, dass es sich um eine junge Mannschaft handelt. Es ist jedoch keine Information darüber enthalten, ob nicht einige erfahrene sprich ältere Spieler mitspielen. Man erhält dasselbe Durchschnittsalter, wenn

- a) alle Spieler 22 Jahre alt sind,
- b) 8 Spieler 20 Jahre, 2 Spieler 27 und ein Spieler 28 Jahre alt ist,
- c) 10 Spieler 21 Jahre und 1 Spieler 32 Jahre alt ist,
- d) etc.



# Mängel - Lageparameter

## Beispiel 6.9:

Wenn das durchschnittliche Alter einer Fußballmannschaft mit 22 Jahren angegeben ist, so weiß man zwar, dass es sich um eine junge Mannschaft handelt. Es ist jedoch keine Information darüber enthalten, ob nicht einige erfahrene sprich ältere Spieler mitspielen. Man erhält dasselbe Durchschnittsalter, wenn

- a) alle Spieler 22 Jahre alt sind,
- b) 8 Spieler 20 Jahre, 2 Spieler 27 und ein Spieler 28 Jahre alt ist,
- c) 10 Spieler 21 Jahre und 1 Spieler 32 Jahre alt ist,
- d) etc.

# Mängel - Lageparameter

## Beispiel 6.9:

Wenn das durchschnittliche Alter einer Fußballmannschaft mit 22 Jahren angegeben ist, so weiß man zwar, dass es sich um eine junge Mannschaft handelt. Es ist jedoch keine Information darüber enthalten, ob nicht einige erfahrene sprich ältere Spieler mitspielen. Man erhält dasselbe Durchschnittsalter, wenn

- a) alle Spieler 22 Jahre alt sind,
- b) 8 Spieler 20 Jahre, 2 Spieler 27 und ein Spieler 28 Jahre alt ist,
- c) 10 Spieler 21 Jahre und 1 Spieler 32 Jahre alt ist,
- d) etc.

# Mängel - Lageparameter

## Beispiel 6.9:

Wenn das durchschnittliche Alter einer Fußballmannschaft mit 22 Jahren angegeben ist, so weiß man zwar, dass es sich um eine junge Mannschaft handelt. Es ist jedoch keine Information darüber enthalten, ob nicht einige erfahrene sprich ältere Spieler mitspielen. Man erhält dasselbe Durchschnittsalter, wenn

- a) alle Spieler 22 Jahre alt sind,
- b) 8 Spieler 20 Jahre, 2 Spieler 27 und ein Spieler 28 Jahre alt ist,
- c) 10 Spieler 21 Jahre und 1 Spieler 32 Jahre alt ist,
- d) etc.

# Mängel - Lageparameter

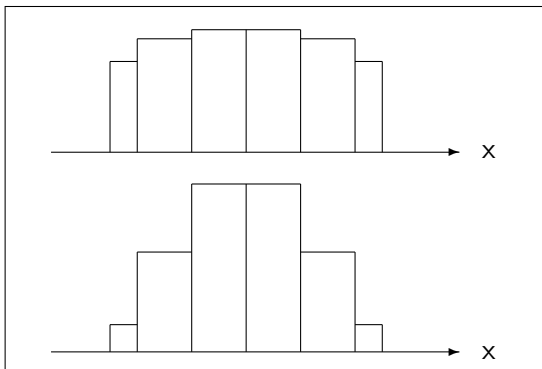


Abbildung 6.4 - Histogramme mit übereinstimmendem arithmetischem Mittel

# Agenda

- ① Ziele
- ② Lageparameter
- ③ **Streuungsparameter**
- ④ Nutzen & Grenzen von Lage- und Streuungsparametern
- ⑤ Graphische Darstellung

# Streuungsparameter

- Kenngröße für die Variation der Werte (**Volatilität**)
- Zusammenfassende Beschreibung der Position der einzelnen Werte in Bezug auf das Zentrum
- Nur für quantitative Merkmale
- Ergänzung zu Lageparametern, da unterschiedliche Verteilungen übereinstimmende Lageparameter besitzen können

# Streuungsparameter

- Kenngröße für die Variation der Werte (**Volatilität**)
- Zusammenfassende Beschreibung der Position der einzelnen Werte in Bezug auf das Zentrum
- Nur für quantitative Merkmale
- Ergänzung zu Lageparametern, da unterschiedliche Verteilungen übereinstimmende Lageparameter besitzen können

# Streuungsparameter

- Kenngröße für die Variation der Werte (**Volatilität**)
- Zusammenfassende Beschreibung der Position der einzelnen Werte in Bezug auf das Zentrum
- Nur für quantitative Merkmale
- Ergänzung zu Lageparametern, da unterschiedliche Verteilungen übereinstimmende Lageparameter besitzen können



# Streuungsparameter

- Kenngröße für die Variation der Werte (**Volatilität**)
- Zusammenfassende Beschreibung der Position der einzelnen Werte in Bezug auf das Zentrum
- Nur für quantitative Merkmale
- Ergänzung zu Lageparametern, da unterschiedliche Verteilungen übereinstimmende Lageparameter besitzen können

# Streuungsparameter

## Spannweite:

Die Spannweite ist der Abstand der am weitesten auseinanderliegenden Beobachtungswerte, also die Differenz aus dem größten und dem kleinsten Beobachtungswert.

Sei  $x_1, \dots, x_n$  die statistische Reihe eines quantitativen Merkmals, so ist die Spannweite

$$R = \max_{i=1, \dots, n} x_i - \min_{i=1, \dots, n} x_i = \max_{i, k=1, \dots, n} (x_k - x_i)$$

Bei klassierten Merkmalen ist die Spannweite definiert durch

$$R = \max_{l:p(l) \neq 0} \beta_l - \min_{l:p(l) \neq 0} \alpha_l$$

# Streuungsparameter

## Spannweite:

Die Spannweite ist der Abstand der am weitesten auseinanderliegenden Beobachtungswerte, also die Differenz aus dem größten und dem kleinsten Beobachtungswert.

Sei  $x_1, \dots, x_n$  die statistische Reihe eines quantitativen Merkmals, so ist die Spannweite

$$R = \max_{i=1, \dots, n} x_i - \min_{i=1, \dots, n} x_i = \max_{i, k=1, \dots, n} (x_k - x_i)$$

Bei klassierten Merkmalen ist die Spannweite definiert durch

$$R = \max_{l:p(l) \neq 0} \beta_l - \min_{l:p(l) \neq 0} \alpha_l$$

# Streuungsparameter

## Nachteil der Spannweite:

Einzelne u.U. weitabliegende Werte (**Ausreißer**) legen die Spannweite fest, die Zwischenwerte bleiben unberücksichtigt.

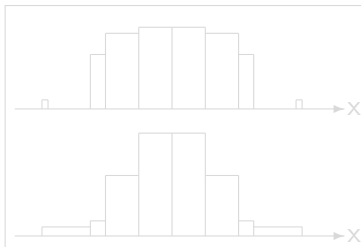


Abbildung 6.5 Histogramme mit übereinstimmender Spannweite.

# Streuungsparameter

## Nachteil der Spannweite:

Einzelne u.U. weitabliegende Werte (**Ausreißer**) legen die Spannweite fest, die Zwischenwerte bleiben unberücksichtigt.

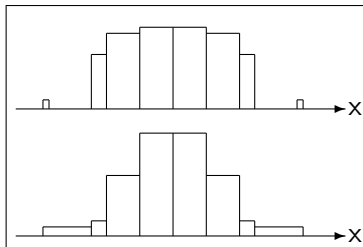


Abbildung 6.5 Histogramme mit übereinstimmender Spannweite.

# Streuungsparameter

## Quartilsabstand:

Mit Hilfe dieses Parameters umgeht man ein wenig die Ausreißerabhängigkeit der Spannweite. Er ist die Differenz aus oberem und unterem Quartil und somit ebenfalls sehr leicht zu berechnen. Jedoch werden auch hierbei die Zwischenwerte nicht berücksichtigt.

$$QA = q_{0.75} - q_{0.25}$$

Der Quartilsabstand in Beispiel 6.4 ist

$$QA = q_{0.75} - q_{0.25} = 48 - 28 = 20.$$

# Streuungsparameter

## Quartilsabstand:

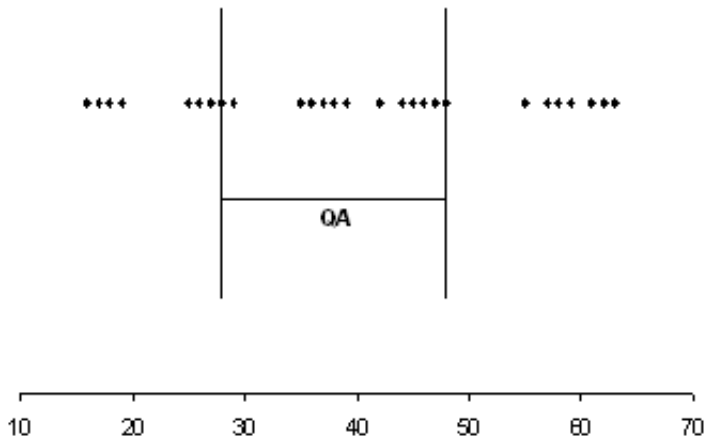
Mit Hilfe dieses Parameters umgeht man ein wenig die Ausreißerabhängigkeit der Spannweite. Er ist die Differenz aus oberem und unterem Quartil und somit ebenfalls sehr leicht zu berechnen. Jedoch werden auch hierbei die Zwischenwerte nicht berücksichtigt.

$$QA = q_{0.75} - q_{0.25}$$

Der Quartilsabstand in Beispiel 6.4 ist

$$QA = q_{0.75} - q_{0.25} = 48 - 28 = 20.$$

# Streuungsparameter





# Streuungsparameter

## **Mittlere absolute Abweichung** (vom Zentralwert):

Die mittlere absolute Abweichung ist das arithmetische Mittel der absoluten Abweichungen der Merkmalswerte vom Zentralwert.

Sei  $x_1, \dots, x_n$  eine statistische Reihe mit dem Zentralwert  $x_z$ , so heißt

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - x_z|$$

mittlere absolute Abweichung (vom Zentralwert).

# Streuungsparameter

## **Mittlere absolute Abweichung** (vom Zentralwert):

Die mittlere absolute Abweichung ist das arithmetische Mittel der absoluten Abweichungen der Merkmalswerte vom Zentralwert.

Sei  $x_1, \dots, x_n$  eine statistische Reihe mit dem Zentralwert  $x_z$ , so heißt

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - x_z|$$

mittlere absolute Abweichung (vom Zentralwert).

# Streuungsparameter

## Beispiel 6.10:

Für die Werte 7, 13, 8, 9, 3 mit  $x_z = 8$  lautet die mittlere absolute Abweichung

$$d = \frac{1 + 5 + 0 + 1 + 5}{5} = \frac{12}{5} = 2.4.$$

Anmerkung: An Stelle des Zentralwertes  $x_z$  könnte man auch einen beliebigen anderen Bezugspunkt  $m$  verwenden:

$$d(m) = \frac{1}{n} \sum_{i=1}^n |x_i - m|.$$

# Streuungsparameter

## Beispiel 6.10:

Für die Werte 7, 13, 8, 9, 3 mit  $x_z = 8$  lautet die mittlere absolute Abweichung

$$d = \frac{1 + 5 + 0 + 1 + 5}{5} = \frac{12}{5} = 2.4.$$

Anmerkung: An Stelle des Zentralwertes  $x_z$  könnte man auch einen beliebigen anderen Bezugspunkt  $m$  verwenden:

$$d(m) = \frac{1}{n} \sum_{i=1}^n |x_i - m|.$$

# Streuungsparameter

## Mittlere absolute Abweichung aus einer Häufigkeitsverteilung:

Die Berechnung der mittleren absoluten Abweichung einer Häufigkeitsverteilung erfolgt analog zur Berechnung des arithmetischen Mittels ( $a_z$  Zentralwert)

$$d = \frac{1}{n} \sum_{a \in M} |a - a_z| h(a) = \sum_{a \in M} |a - a_z| p(a).$$

Die mittlere absolute Abweichung eines **klassierten Merkmals** bestimmt man aus den Häufigkeiten der Klassen und setzt die Klassenmitte jeweils als Repräsentanten ihrer Klasse.

Sei  $z_l$  jeweils die Klassenmitte der Klasse  $l$ ,  $h(l)$ ,  $p(l)$  ihre absolute bzw. relative Häufigkeit. Sei  $x_z$  der (feinberechnete) Zentralwert, so erhält man ( $n$  Anzahl der Beobachtungswerte)

$$d = \frac{1}{n} \sum_l |z_l - x_z| h(l) = \sum_l |z_l - x_z| p(l).$$

# Streuungsparameter

## Mittlere absolute Abweichung aus einer Häufigkeitsverteilung:

Die Berechnung der mittleren absoluten Abweichung einer Häufigkeitsverteilung erfolgt analog zur Berechnung des arithmetischen Mittels ( $a_z$  Zentralwert)

$$d = \frac{1}{n} \sum_{a \in M} |a - a_z| h(a) = \sum_{a \in M} |a - a_z| p(a).$$

Die mittlere absolute Abweichung eines **klassierten Merkmals** bestimmt man aus den Häufigkeiten der Klassen und setzt die Klassenmitte jeweils als Repräsentanten ihrer Klasse.

Sei  $z_l$  jeweils die Klassenmitte der Klasse  $l$ ,  $h(l)$ ,  $p(l)$  ihre absolute bzw. relative Häufigkeit. Sei  $x_z$  der (feinberechnete) Zentralwert, so erhält man ( $n$  Anzahl der Beobachtungswerte)

$$d = \frac{1}{n} \sum_l |z_l - x_z| h(l) = \sum_l |z_l - x_z| p(l).$$

# Streuungsparameter

## Mittlere absolute Abweichung aus einer Häufigkeitsverteilung:

Die Berechnung der mittleren absoluten Abweichung einer Häufigkeitsverteilung erfolgt analog zur Berechnung des arithmetischen Mittels ( $a_z$  Zentralwert)

$$d = \frac{1}{n} \sum_{a \in M} |a - a_z| h(a) = \sum_{a \in M} |a - a_z| p(a).$$

Die mittlere absolute Abweichung eines **klassierten Merkmals** bestimmt man aus den Häufigkeiten der Klassen und setzt die Klassenmitte jeweils als Repräsentanten ihrer Klasse.

Sei  $z_l$  jeweils die Klassenmitte der Klasse  $l$ ,  $h(l)$ ,  $p(l)$  ihre absolute bzw. relative Häufigkeit. Sei  $x_z$  der (feinberechnete) Zentralwert, so erhält man ( $n$  Anzahl der Beobachtungswerte)

$$d = \frac{1}{n} \sum_l |z_l - x_z| h(l) = \sum_l |z_l - x_z| p(l).$$

# Streuungsparameter

## Mittlere absolute Abweichung aus einer Häufigkeitsverteilung:

Die Berechnung der mittleren absoluten Abweichung einer Häufigkeitsverteilung erfolgt analog zur Berechnung des arithmetischen Mittels ( $a_z$  Zentralwert)

$$d = \frac{1}{n} \sum_{a \in M} |a - a_z| h(a) = \sum_{a \in M} |a - a_z| p(a).$$

Die mittlere absolute Abweichung eines **klassierten Merkmals** bestimmt man aus den Häufigkeiten der Klassen und setzt die Klassenmitte jeweils als Repräsentanten ihrer Klasse.

Sei  $z_l$  jeweils die Klassenmitte der Klasse  $l$ ,  $h(l)$ ,  $p(l)$  ihre absolute bzw. relative Häufigkeit. Sei  $x_z$  der (feinberechnete) Zentralwert, so erhält man ( $n$  Anzahl der Beobachtungswerte)

$$d = \frac{1}{n} \sum_l |z_l - x_z| h(l) = \sum_l |z_l - x_z| p(l).$$



# Streuungsparameter

**Beispiel 6.11:** Bei den Einkommensverteilungen aus Beispiel 5.5 liegt der feinberechnete Zentralwert für die Männer bei 27 TEuro und für die Frauen bei 24.29 TEuro. Damit erhält man als mittlere absolute Abweichung:

$$\begin{aligned}\text{Männer: } d &= \frac{1}{100} (|5 - 27| \cdot 5 + |15 - 27| \cdot 15 + |22.5 - 27| \cdot 20 \\ &\quad + |27.5 - 27| \cdot 25 + |35 - 27| \cdot 20 + |50 - 27| \cdot 10 + |72.5 - 27| \cdot 5) \\ &= 10.1 \text{ TEuro,}\end{aligned}$$

$$\begin{aligned}\text{Frauen: } d &= |5 - 24.29| \cdot 0.06 + |15 - 24.29| \cdot 0.20 \\ &\quad + |22.5 - 24.29| \cdot 0.28 + |27.5 - 24.29| \cdot 0.32 + |35 - 24.29| \cdot 0.10 \\ &\quad + |50 - 24.29| \cdot 0.04 \\ &= 6.6432 \text{ TEuro.}\end{aligned}$$

Das Einkommen der Frauen variiert demnach nicht so stark wie bei den Männern.

# Streuungsparameter

**Beispiel 6.11:** Bei den Einkommensverteilungen aus Beispiel 5.5 liegt der feinberechnete Zentralwert für die Männer bei 27 TEuro und für die Frauen bei 24.29 TEuro. Damit erhält man als mittlere absolute Abweichung:

$$\begin{aligned}\text{Männer: } d &= \frac{1}{100} (|5 - 27| \cdot 5 + |15 - 27| \cdot 15 + |22.5 - 27| \cdot 20 \\ &\quad + |27.5 - 27| \cdot 25 + |35 - 27| \cdot 20 + |50 - 27| \cdot 10 + |72.5 - 27| \cdot 5) \\ &= 10.1 \text{ TEuro,}\end{aligned}$$

$$\begin{aligned}\text{Frauen: } d &= |5 - 24.29| \cdot 0.06 + |15 - 24.29| \cdot 0.20 \\ &\quad + |22.5 - 24.29| \cdot 0.28 + |27.5 - 24.29| \cdot 0.32 + |35 - 24.29| \cdot 0.10 \\ &\quad + |50 - 24.29| \cdot 0.04 \\ &= 6.6432 \text{ TEuro.}\end{aligned}$$

Das Einkommen der Frauen variiert demnach nicht so stark wie bei den Männern.

# Streuungsparameter

**Beispiel 6.11:** Bei den Einkommensverteilungen aus Beispiel 5.5 liegt der feinberechnete Zentralwert für die Männer bei 27 TEuro und für die Frauen bei 24.29 TEuro. Damit erhält man als mittlere absolute Abweichung:

$$\begin{aligned}\text{Männer: } d &= \frac{1}{100} (|5 - 27| \cdot 5 + |15 - 27| \cdot 15 + |22.5 - 27| \cdot 20 \\ &\quad + |27.5 - 27| \cdot 25 + |35 - 27| \cdot 20 + |50 - 27| \cdot 10 + |72.5 - 27| \cdot 5) \\ &= 10.1 \text{ TEuro,}\end{aligned}$$

$$\begin{aligned}\text{Frauen: } d &= |5 - 24.29| \cdot 0.06 + |15 - 24.29| \cdot 0.20 \\ &\quad + |22.5 - 24.29| \cdot 0.28 + |27.5 - 24.29| \cdot 0.32 + |35 - 24.29| \cdot 0.10 \\ &\quad + |50 - 24.29| \cdot 0.04 \\ &= 6.6432 \text{ TEuro.}\end{aligned}$$

Das Einkommen der Frauen variiert demnach nicht so stark wie bei den Männern.

# Streuungsparameter

Geschlecht	Einkommen (TEuro)	abs. H.	rel. H.	$z_j - x_z$	$ (z_j - x_z) $	$ (z_j - x_z) p(l)$	d
Männer	unter 10	5	0.05	-22	22	1.1	10.1
	10 b.u. 20	15	0.15	-12	12	1.8	
	20 b.u. 25	20	0.20	-4,5	4,5	0.9	
	25 b.u. 30	25	0.25	0,5	0,5	0.125	
	30 b.u. 40	20	0.20	8	8	1.6	
	40 b.u. 60	10	0.1	23	23	2.3	
	60 b.u. 85	5	0.05	45.5	45.5	2.275	
Frauen	unter 10	9	0.06	-19.29	19.29	1.1574	6.6432
	10 b.u. 20	30	0.20	-9.29	9.29	1.858	
	20 b.u. 25	42	0.28	-1.79	1.79	0.5012	
	25 b.u. 30	48	0.32	2.51	2.51	0.8032	
	30 b.u. 40	15	0.10	10.71	10.71	1.071	
	40 b.u. 60	6	0.04	25.71	25.71	1.0284	
	60 b.u. 85	0	0	0	0	0	

# Streuungsparameter

## Varianz:

Dieser Streuungsparameter ist das arithmetische Mittel der quadrierten Abweichungen der Beobachtungswerte vom arithmetischen Mittel.

Sei  $x_1, \dots, x_n$  eine statistische Reihe mit arithmetischem Mittel  $\bar{x}$ , so ist die Varianz der Verteilung

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

## Korrigierte Stichprobenvarianz:

Sind  $x_1, \dots, x_n$  die Werte einer Stichprobe, dann wird in der Regel

$$s^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

als Schätzwert für die Varianz der Grundgesamtheit

# Streuungsparameter

## Varianz:

Dieser Streuungsparameter ist das arithmetische Mittel der quadrierten Abweichungen der Beobachtungswerte vom arithmetischen Mittel.

Sei  $x_1, \dots, x_n$  eine statistische Reihe mit arithmetischem Mittel  $\bar{x}$ , so ist die Varianz der Verteilung

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

## Korrigierte Stichprobenvarianz:

Sind  $x_1, \dots, x_n$  die Werte einer Stichprobe, dann wird in der Regel

$$s^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

als Schätzwert für die Varianz der Grundgesamtheit

# Streuungsparameter

## Varianz:

Dieser Streuungsparameter ist das arithmetische Mittel der quadrierten Abweichungen der Beobachtungswerte vom arithmetischen Mittel.

Sei  $x_1, \dots, x_n$  eine statistische Reihe mit arithmetischem Mittel  $\bar{x}$ , so ist die Varianz der Verteilung

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

## Korrigierte Stichprobenvarianz:

Sind  $x_1, \dots, x_n$  die Werte einer Stichprobe, dann wird in der Regel

$$s^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

als Schätzwert für die Varianz der Grundgesamtheit

# Streuungsparameter

## Standardabweichung:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

## Beispiel 6.12:

Für die Werte aus dem obigen Beispiel 6.10: 7, 13, 8, 9, 3 erhält man  $\bar{x} = 8$  und

$$s^2 = \frac{1 + 25 + 0 + 1 + 25}{5} = \frac{52}{5} = 10.4,$$

$$s = \sqrt{10.4} = 3.225,$$

$$s^{*2} = \frac{52}{4} = 13.$$



# Streuungsparameter

## Standardabweichung:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

## Beispiel 6.12:

Für die Werte aus dem obigen Beispiel 6.10: 7, 13, 8, 9, 3 erhält man  $\bar{x} = 8$  und

$$s^2 = \frac{1 + 25 + 0 + 1 + 25}{5} = \frac{52}{5} = 10.4,$$

$$s = \sqrt{10.4} = 3.225,$$

$$s^{*2} = \frac{52}{4} = 13.$$

# Streuungsparameter

## Varianz einer Verteilung:

$$s^2 = \frac{1}{n} \sum_{a \in M} (a - \bar{a})^2 h(a) = \sum_{a \in M} (a - \bar{a})^2 p(a) \quad (\bar{a} \text{ arithm. Mittel})$$

## Varianz der Verteilung eines klassierten Merkmals:

$$s^2 = \frac{1}{n} \sum_l (z_l - \bar{x})^2 h(l) = \sum_l (z_l - \bar{x})^2 p(l)$$

# Streuungsparameter

## Varianz einer Verteilung:

$$s^2 = \frac{1}{n} \sum_{a \in M} (a - \bar{a})^2 h(a) = \sum_{a \in M} (a - \bar{a})^2 p(a) \quad (\bar{a} \text{ arithm. Mittel})$$

## Varianz der Verteilung eines klassierten Merkmals:

$$s^2 = \frac{1}{n} \sum_l (z_l - \bar{x})^2 h(l) = \sum_l (z_l - \bar{x})^2 p(l)$$

# Streuungsparameter

## Beispiel 6.13:

Zu Beispiel 6.7 ergibt sich folgende relative Häufigkeitsverteilung:

Kinderzahl	0	1	2	3	4	5	6
Relative H.	.10	.35	.25	.15	.125	.025	0

Das arithmetische Mittel lautet  $\bar{x} = 1.925$ . Daraus berechnet man als Varianz:

$$\begin{aligned}s^2 &= 1.925^2 \cdot 0.10 + 0.925^2 \cdot 0.35 + 0.075^2 \cdot 0.25 + 1.075^2 \cdot 0.15 \\ &\quad + 2.075^2 \cdot 0.125 + 3.075^2 \cdot 0.025 + 4.075^2 \cdot 0 \\ &= 1.619\end{aligned}$$

# Streuungsparameter

## Beispiel 6.13:

Zu Beispiel 6.7 ergibt sich folgende relative Häufigkeitsverteilung:

Kinderzahl	0	1	2	3	4	5	6
Relative H.	.10	.35	.25	.15	.125	.025	0

Das arithmetische Mittel lautet  $\bar{x} = 1.925$ . Daraus berechnet man als Varianz:

$$\begin{aligned}s^2 &= 1.925^2 \cdot 0.10 + 0.925^2 \cdot 0.35 + 0.075^2 \cdot 0.25 + 1.075^2 \cdot 0.15 \\ &\quad + 2.075^2 \cdot 0.125 + 3.075^2 \cdot 0.025 + 4.075^2 \cdot 0 \\ &= 1.619\end{aligned}$$

# Streuungsparameter

## Beispiel 6.13:

Zu Beispiel 6.7 ergibt sich folgende relative Häufigkeitsverteilung:

Kinderzahl	0	1	2	3	4	5	6
Relative H.	.10	.35	.25	.15	.125	.025	0

Das arithmetische Mittel lautet  $\bar{x} = 1.925$ . Daraus berechnet man als Varianz:

$$\begin{aligned}s^2 &= 1.925^2 \cdot 0.10 + 0.925^2 \cdot 0.35 + 0.075^2 \cdot 0.25 + 1.075^2 \cdot 0.15 \\ &\quad + 2.075^2 \cdot 0.125 + 3.075^2 \cdot 0.025 + 4.075^2 \cdot 0 \\ &= 1.619\end{aligned}$$

# Streuungsparameter

## Beispiel 6.13:

Zu Beispiel 6.7 ergibt sich folgende relative Häufigkeitsverteilung:

Kinderzahl	0	1	2	3	4	5	6
Relative H.	.10	.35	.25	.15	.125	.025	0

Das arithmetische Mittel lautet  $\bar{x} = 1.925$ . Daraus berechnet man als Varianz:

$$\begin{aligned}s^2 &= 1.925^2 \cdot 0.10 + 0.925^2 \cdot 0.35 + 0.075^2 \cdot 0.25 + 1.075^2 \cdot 0.15 \\ &\quad + 2.075^2 \cdot 0.125 + 3.075^2 \cdot 0.025 + 4.075^2 \cdot 0 \\ &= 1.619\end{aligned}$$

# Streuungsparameter

## Beispiel 6.14:

Arithmetisches Mittel der klassierten Altersangaben in Beispiel 6.4 ist  $\bar{x} = 38.25$ . Die Berechnung der Varianz wird durch folgende Tabelle erleichtert:

Alter	abs. Hfgkt.	rel. Hfgkt.	$z_l$	$z_l - \bar{x}$	$(z_l - \bar{x})^2$	$(z_l - \bar{x})^2 p(l)$
15 bis unter 20	4	0.1	17.5	20.75	430.56	43.06
20 b.u. 30	9	0.225	25	13.25	175.56	39.50
30 b.u. 40	10	0.25	35	3.25	10.56	2.64
40 b.u. 50	8	0.2	45	6.75	45.56	9.11
50 b.u. 60	5	0.125	55	16.75	280.56	35.07
60 bis unter 65	4	0.1	62.5	24.25	588.06	58.81
					$s^2 =$	188.19

**Anmerkung:** Dadurch, dass innerhalb der einzelnen Klassen die wahren Beobachtungswerte durch die Klassenmitte als Repräsentanten der Klasse ersetzt werden, bleibt bei der Berechnung der Varianz die Streuung der Werte innerhalb der Klassen unberücksichtigt.



# Streuungsparameter

## Beispiel 6.14:

Arithmetisches Mittel der klassierten Altersangaben in Beispiel 6.4 ist  $\bar{x} = 38.25$ . Die Berechnung der Varianz wird durch folgende Tabelle erleichtert:

Alter	abs. Hfgkt.	rel. Hfgkt.	$z_l$	$z_l - \bar{x}$	$(z_l - \bar{x})^2$	$(z_l - \bar{x})^2 p(l)$
15 bis unter 20	4	0.1	17.5	20.75	430.56	43.06
20 b.u. 30	9	0.225	25	13.25	175.56	39.50
30 b.u. 40	10	0.25	35	3.25	10.56	2.64
40 b.u. 50	8	0.2	45	6.75	45.56	9.11
50 b.u. 60	5	0.125	55	16.75	280.56	35.07
60 bis unter 65	4	0.1	62.5	24.25	588.06	58.81
					$s^2 =$	188.19

**Anmerkung:** Dadurch, dass innerhalb der einzelnen Klassen die wahren Beobachtungswerte durch die Klassenmitte als Repräsentanten der Klasse ersetzt werden, bleibt bei der Berechnung der Varianz die Streuung der Werte innerhalb der Klassen unberücksichtigt.

# Streuungsparameter

## Beispiel 6.14:

Arithmetisches Mittel der klassierten Altersangaben in Beispiel 6.4 ist  $\bar{x} = 38.25$ . Die Berechnung der Varianz wird durch folgende Tabelle erleichtert:

Alter	abs. Hfgkt.	rel. Hfgkt.	$z_l$	$z_l - \bar{x}$	$(z_l - \bar{x})^2$	$(z_l - \bar{x})^2 p(l)$
15 bis unter 20	4	0.1	17.5	20.75	430.56	43.06
20 b.u. 30	9	0.225	25	13.25	175.56	39.50
30 b.u. 40	10	0.25	35	3.25	10.56	2.64
40 b.u. 50	8	0.2	45	6.75	45.56	9.11
50 b.u. 60	5	0.125	55	16.75	280.56	35.07
60 bis unter 65	4	0.1	62.5	24.25	588.06	58.81
					$s^2 =$	188.19

**Anmerkung:** Dadurch, dass innerhalb der einzelnen Klassen die wahren Beobachtungswerte durch die Klassenmitte als Repräsentanten der Klasse ersetzt werden, bleibt bei der Berechnung der Varianz die Streuung der Werte innerhalb der Klassen unberücksichtigt.

# Streuungsparameter

Folgende Umformung der Berechnungsformel vereinfacht gelegentlich die Berechnung der Varianz:

$$\begin{aligned}s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) + \bar{x}^2 - 2\bar{x} \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) - \bar{x}^2\end{aligned}$$

Analog gilt für die Varianz einer Verteilung:

$$s^2 = \frac{1}{n} \sum_{a \in M} a^2 h(a) - \bar{a}^2 = \sum_{a \in M} a^2 p(a) - \bar{a}^2$$

Und für die eines klassierten Merkmals:

$$s^2 = \frac{1}{n} \sum_l z_l^2 h(l) - \bar{x}^2 = \sum_l z_l^2 p(l) - \bar{x}^2.$$

# Streuungsparameter

Folgende Umformung der Berechnungsformel vereinfacht gelegentlich die Berechnung der Varianz:

$$\begin{aligned}s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) + \bar{x}^2 - 2\bar{x} \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) - \bar{x}^2\end{aligned}$$

Analog gilt für die Varianz einer Verteilung:

$$s^2 = \frac{1}{n} \sum_{a \in M} a^2 h(a) - \bar{a}^2 = \sum_{a \in M} a^2 p(a) - \bar{a}^2$$

Und für die eines klassierten Merkmals:

$$s^2 = \frac{1}{n} \sum_l z_l^2 h(l) - \bar{x}^2 = \sum_l z_l^2 p(l) - \bar{x}^2.$$

# Streuungsparameter

Folgende Umformung der Berechnungsformel vereinfacht gelegentlich die Berechnung der Varianz:

$$\begin{aligned}s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) + \bar{x}^2 - 2\bar{x} \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 \right) - \bar{x}^2\end{aligned}$$

Analog gilt für die Varianz einer Verteilung:

$$s^2 = \frac{1}{n} \sum_{a \in M} a^2 h(a) - \bar{a}^2 = \sum_{a \in M} a^2 p(a) - \bar{a}^2$$

Und für die eines klassierten Merkmals:

$$s^2 = \frac{1}{n} \sum_l z_l^2 h(l) - \bar{x}^2 = \sum_l z_l^2 p(l) - \bar{x}^2.$$

# Streuungsparameter

## Beispiel 6.15:

In Beispiel 4.5 errechnet man auf diese Weise:

$i$	$x_i$	$x_i^2$
1	1.49	2.220
2	1.87	3.497
3	1.91	3.648
4	1.53	2.341
5	1.68	2.822
6	1.75	3.063
7	1.66	2.756
8	1.82	3.312
9	1.76	3.098
10	1.80	3.240

11	1.92	3.686
12	1.71	2.924
13	1.77	3.133
14	1.69	2.856
15	1.57	2.465
16	1.83	3.386
17	1.84	2.161
18	1.47	2.161
19	1.79	3.204
20	1.81	3.276
$\Sigma$	34.67	60.437

und damit  $\bar{x} = 1.73$ ,  $s^2 = 0.017$  und  $s = 0.130$ .

# Streuungsparameter

## Beispiel 6.15:

In Beispiel 4.5 errechnet man auf diese Weise:

$i$	$x_i$	$x_i^2$
1	1.49	2.220
2	1.87	3.497
3	1.91	3.648
4	1.53	2.341
5	1.68	2.822
6	1.75	3.063
7	1.66	2.756
8	1.82	3.312
9	1.76	3.098
10	1.80	3.240

11	1.92	3.686
12	1.71	2.924
13	1.77	3.133
14	1.69	2.856
15	1.57	2.465
16	1.83	3.386
17	1.84	2.161
18	1.47	2.161
19	1.79	3.204
20	1.81	3.276
$\Sigma$	34.67	60.437

und damit  $\bar{x} = 1.73$ ,  $s^2 = 0.017$  und  $s = 0.130$ .

# Streuungsparameter

## Lage- und Streuungsparameter bei Transformation der Merkmalswerte/-ausprägungen:

Sei

$$y = \alpha x + \beta$$

eine Transformation.

Die statistische Reihe  $x_1, \dots, x_n$  geht über in  $y_1, \dots, y_n$  mit  $y_i = \alpha x_i + \beta$ . Für Modalwert, Zentralwert und arithmetisches Mittel gilt dann dieselbe Transformationsformel:

$$y_M = \alpha x_M + \beta, \quad y_z = \alpha x_z + \beta, \quad \bar{y} = \alpha \bar{x} + \beta$$



# Streuungsparameter

## Lage- und Streuungsparameter bei Transformation der Merkmalswerte/-ausprägungen:

Sei

$$y = \alpha x + \beta$$

eine Transformation.

Die statistische Reihe  $x_1, \dots, x_n$  geht über in  $y_1, \dots, y_n$  mit  $y_i = \alpha x_i + \beta$ . Für Modalwert, Zentralwert und arithmetisches Mittel gilt dann dieselbe Transformationsformel:

$$y_M = \alpha x_M + \beta, \quad y_Z = \alpha x_Z + \beta, \quad \bar{y} = \alpha \bar{x} + \beta$$

# Streuungsparameter

## Lage- und Streuungsparameter bei Transformation der Merkmalswerte/-ausprägungen:

Sei

$$y = \alpha x + \beta$$

eine Transformation.

Die statistische Reihe  $x_1, \dots, x_n$  geht über in  $y_1, \dots, y_n$  mit  $y_i = \alpha x_i + \beta$ . Für Modalwert, Zentralwert und arithmetisches Mittel gilt dann dieselbe Transformationsformel:

$$y_M = \alpha x_M + \beta, \quad y_Z = \alpha x_Z + \beta, \quad \bar{y} = \alpha \bar{x} + \beta$$

# Streuungsparameter

## Lage- und Streuungsparameter bei Transformation der Merkmalswerte/-ausprägungen:

Sei

$$y = \alpha x + \beta$$

eine Transformation.

Die statistische Reihe  $x_1, \dots, x_n$  geht über in  $y_1, \dots, y_n$  mit  $y_i = \alpha x_i + \beta$ . Für Modalwert, Zentralwert und arithmetisches Mittel gilt dann dieselbe Transformationsformel:

$$y_M = \alpha x_M + \beta, \quad y_Z = \alpha x_Z + \beta, \quad \bar{y} = \alpha \bar{x} + \beta$$

# Streuungsparameter

## Lage- und Streuungsparameter bei Transformation der Merkmalswerte/-ausprägungen:

Damit gilt für die mittlere absolute Abweichung

$$\begin{aligned}d_y &= \frac{1}{n} \sum_{i=1}^n |y_i - y_z| = \frac{1}{n} \sum_{i=1}^n |\alpha x_i + \beta - (\alpha x_z + \beta)| \\ &= \frac{1}{n} \sum_{i=1}^n |\alpha| \cdot |x_i - x_z| = |\alpha| \cdot d_x\end{aligned}$$

und für die Varianz

$$\begin{aligned}s_y^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n (\alpha x_i + \beta - \alpha \bar{x} - \beta)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \alpha^2 \cdot (x_i - \bar{x})^2 = \alpha^2 s_x^2.\end{aligned}$$

# Streuungsparameter

## Lage- und Streuungsparameter bei Transformation der Merkmalswerte/-ausprägungen:

Damit gilt für die mittlere absolute Abweichung

$$\begin{aligned}d_y &= \frac{1}{n} \sum_{i=1}^n |y_i - y_z| = \frac{1}{n} \sum_{i=1}^n |\alpha x_i + \beta - (\alpha x_z + \beta)| \\ &= \frac{1}{n} \sum_{i=1}^n |\alpha| \cdot |x_i - x_z| = |\alpha| \cdot d_x\end{aligned}$$

und für die Varianz

$$\begin{aligned}s_y^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n (\alpha x_i + \beta - \alpha \bar{x} - \beta)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \alpha^2 \cdot (x_i - \bar{x})^2 = \alpha^2 s_x^2.\end{aligned}$$

# Streuungsparameter

## Variationskoeffizient der Verteilung:

Sei  $s$  die Standardabweichung und  $\bar{x}$  das arithmetische Mittel einer Verteilung, so ist der Variationskoeffizient

$$v = \frac{s}{\bar{x}}$$

### Beispiel 6.16:

In Beispiel 6.15 ergibt sich der Variationskoeffizient  $v_x = 0.075$ , misst man die Körpergrößen in cm (also  $x \cdot 100$ ), so erhält man  $\bar{y} = 173$  und  $s_y = 13.04$ , also ebenfalls  $v_y = 0.075$ .

Anmerkung: Führt man eine Transformation  $y = \alpha x$ ,  $\alpha > 0$  der Merkmalsausprägungen durch, so gilt

$$v_y = \frac{s_y}{\bar{y}} = \frac{\alpha \cdot s_x}{\alpha \cdot \bar{x}} = \frac{s_x}{\bar{x}} = v_x.$$

# Streuungsparameter

## Variationskoeffizient der Verteilung:

Sei  $s$  die Standardabweichung und  $\bar{x}$  das arithmetische Mittel einer Verteilung, so ist der Variationskoeffizient

$$v = \frac{s}{\bar{x}}$$

### Beispiel 6.16:

In Beispiel 6.15 ergibt sich der Variationskoeffizient  $v_x = 0.075$ , misst man die Körpergrößen in cm (also  $x \cdot 100$ ), so erhält man  $\bar{y} = 173$  und  $s_y = 13.04$ , also ebenfalls  $v_y = 0.075$ .

Anmerkung: Führt man eine Transformation  $y = \alpha x$ ,  $\alpha > 0$  der Merkmalsausprägungen durch, so gilt

$$v_y = \frac{s_y}{\bar{y}} = \frac{\alpha \cdot s_x}{\alpha \cdot \bar{x}} = \frac{s_x}{\bar{x}} = v_x.$$

# Streuungsparameter

## Variationskoeffizient der Verteilung:

Sei  $s$  die Standardabweichung und  $\bar{x}$  das arithmetische Mittel einer Verteilung, so ist der Variationskoeffizient

$$v = \frac{s}{\bar{x}}$$

### Beispiel 6.16:

In Beispiel 6.15 ergibt sich der Variationskoeffizient  $v_x = 0.075$ , misst man die Körpergrößen in cm (also  $x \cdot 100$ ), so erhält man  $\bar{y} = 173$  und  $s_y = 13.04$ , also ebenfalls  $v_y = 0.075$ .

Anmerkung: Führt man eine Transformation  $y = \alpha x, \alpha > 0$  der Merkmalsausprägungen durch, so gilt

$$v_y = \frac{s_y}{\bar{y}} = \frac{\alpha \cdot s_x}{\alpha \cdot \bar{x}} = \frac{s_x}{\bar{x}} = v_x.$$



# Streuungsparameter

## Variationskoeffizient der Verteilung:

Sei  $s$  die Standardabweichung und  $\bar{x}$  das arithmetische Mittel einer Verteilung, so ist der Variationskoeffizient

$$v = \frac{s}{\bar{x}}$$

## Beispiel 6.16:

In Beispiel 6.15 ergibt sich der Variationskoeffizient  $v_x = 0.075$ , misst man die Körpergrößen in cm (also  $x \cdot 100$ ), so erhält man  $\bar{y} = 173$  und  $s_y = 13.04$ , also ebenfalls  $v_y = 0.075$ .

Anmerkung: Führt man eine Transformation  $y = \alpha x$ ,  $\alpha > 0$  der Merkmalsausprägungen durch, so gilt

$$v_y = \frac{s_y}{\bar{y}} = \frac{\alpha \cdot s_x}{\alpha \cdot \bar{x}} = \frac{s_x}{\bar{x}} = v_x.$$

# Zusammenfassung - Streuungsparameter

## Quantitatives Merkmal

- Spannweite
- Quartilsabstand
- Mittlere absolute Abweichung
- Varianz und Standardabweichung
- Variationskoeffizient

Bei Rangmerkmalen und qualitativen Merkmalen können Streuungsparameter nicht berechnet werden.

# Zusammenfassung - Streuungsparameter

## Quantitatives Merkmal

- Spannweite
- Quartilsabstand
- Mittlere absolute Abweichung
- Varianz und Standardabweichung
- Variationskoeffizient

Bei Rangmerkmalen und qualitativen Merkmalen können Streuungsparameter nicht berechnet werden.

# Zusammenfassung - Streuungsparameter

## Quantitatives Merkmal

- Spannweite
- Quartilsabstand
- Mittlere absolute Abweichung
- Varianz und Standardabweichung
- Variationskoeffizient

Bei Rangmerkmalen und qualitativen Merkmalen können Streuungsparameter nicht berechnet werden.

# Zusammenfassung - Streuungsparameter

## Quantitatives Merkmal

- Spannweite
- Quartilsabstand
- Mittlere absolute Abweichung
- Varianz und Standardabweichung
- Variationskoeffizient

Bei Rangmerkmalen und qualitativen Merkmalen können Streuungsparameter nicht berechnet werden.

# Zusammenfassung - Streuungsparameter

## Quantitatives Merkmal

- Spannweite
- Quartilsabstand
- Mittlere absolute Abweichung
- Varianz und Standardabweichung
- Variationskoeffizient

Bei Rangmerkmalen und qualitativen Merkmalen können Streuungsparameter nicht berechnet werden.

# Zusammenfassung - Streuungsparameter

## Quantitatives Merkmal

- Spannweite
- Quartilsabstand
- Mittlere absolute Abweichung
- Varianz und Standardabweichung
- Variationskoeffizient

Bei Rangmerkmalen und qualitativen Merkmalen können Streuungsparameter nicht berechnet werden.

# Zusammenfassung - Streuungsparameter

## Quantitatives Merkmal

- Spannweite
- Quartilsabstand
- Mittlere absolute Abweichung
- Varianz und Standardabweichung
- Variationskoeffizient

Bei Rangmerkmalen und qualitativen Merkmalen können Streuungsparameter nicht berechnet werden.



# Agenda

- ① Ziele
- ② Lageparameter
- ③ Streuungsparameter
- ④ **Nutzen & Grenzen von Lage- und Streuungsparametern**
- ⑤ Graphische Darstellung

# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipflig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,

95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,

99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipflig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,

95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,

99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipflig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

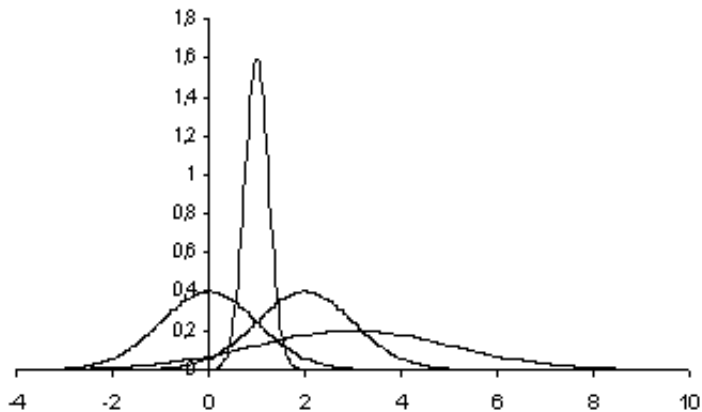
Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,

95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,

99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

# Nutzen & Grenzen von Lage- und Streuungsparametern



# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipfelig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

- 68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,
- 95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,
- 99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

## Grenzen:

Es gibt jedoch zahlreiche Verteilungen, bei denen die Angabe von zwei Parametern nicht ausreicht, um eine Verteilung hinreichend zu beschreiben.

# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipfelig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

- 68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,
- 95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,
- 99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

## Grenzen:

Es gibt jedoch zahlreiche Verteilungen, bei denen die Angabe von zwei Parametern nicht ausreicht, um eine Verteilung hinreichend zu beschreiben.

# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipfelig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

- 68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,
- 95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,
- 99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

## Grenzen:

Es gibt jedoch zahlreiche Verteilungen, bei denen die Angabe von zwei Parametern nicht ausreicht, um eine Verteilung hinreichend zu beschreiben.



# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipfelig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

- 68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,
- 95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,
- 99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

## Grenzen:

Es gibt jedoch zahlreiche Verteilungen, bei denen die Angabe von zwei Parametern nicht ausreicht, um eine Verteilung hinreichend zu beschreiben.

# Nutzen & Grenzen von Lage- und Streuungsparametern

## Nutzen:

Lage- und Streuungsparameter geben in den Fällen, in denen die Häufigkeitsverteilung eingipfelig und symmetrisch ist, ein recht genaues Bild über die Verteilung. Die Kurven geben einen idealisierten Verlauf der Histogramme, der sich bei einer Reduzierung der Klassenbreite gegen Null und wachsender Zahl von Beobachtungen ergibt.

Mittelwert und Standardabweichung charakterisieren die Verteilung sehr gut:

- 68% aller Werte im Intervall  $[\bar{x} - s, \bar{x} + s]$ ,
- 95% aller Werte im Intervall  $[\bar{x} - 2s, \bar{x} + 2s]$ ,
- 99% aller Werte im Intervall  $[\bar{x} - 3s, \bar{x} + 3s]$ .

## Grenzen:

Es gibt jedoch zahlreiche Verteilungen, bei denen die Angabe von zwei Parametern nicht ausreicht, um eine Verteilung hinreichend zu beschreiben.

# Nutzen & Grenzen von Lage- und Streuungsparametern

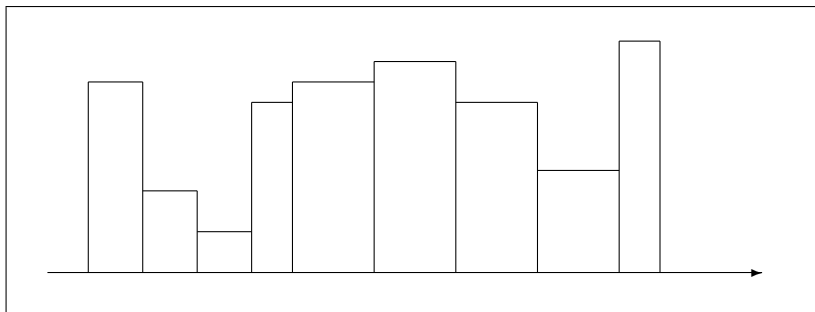


Abbildung 6.7 - Histogramm einer mehrgipfeligen Verteilung

# Nutzen & Grenzen von Lage- und Streuungsparametern

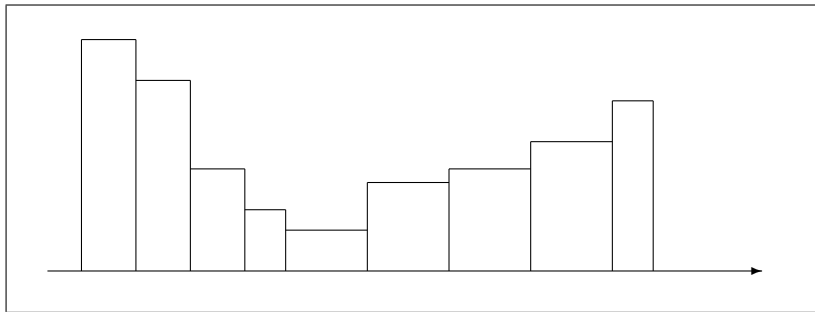


Abbildung 6.8 - Histogramm einer U-förmigen Verteilung.

# Nutzen & Grenzen von Lage- und Streuungsparametern

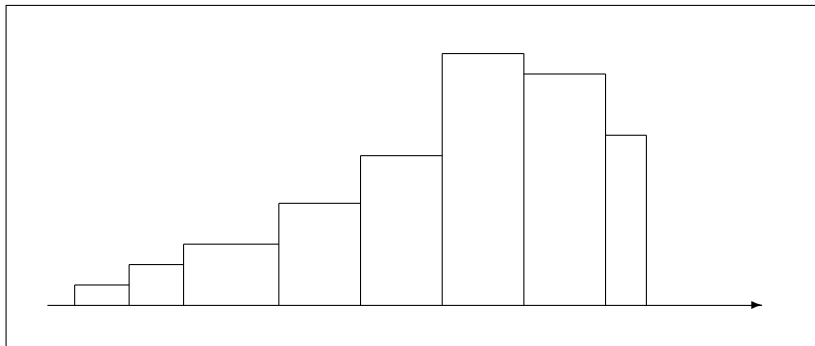


Abbildung 6.9 - Histogramm einer linksschiefen Verteilung.

# Nutzen & Grenzen von Lage- und Streuungsparametern

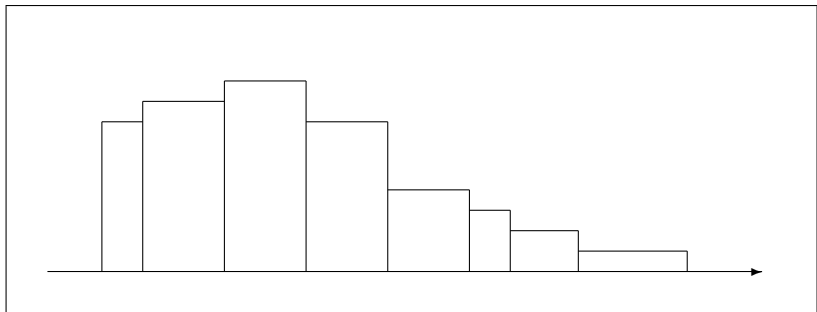


Abbildung 6.10 Histogramm einer rechtsschiefen Verteilung.

# Agenda

- ① Ziele
- ② Lageparameter
- ③ Streuungsparameter
- ④ Nutzen & Grenzen von Lage- und Streuungsparametern
- ⑤ **Graphische Darstellung**

# Graphische Darstellung

## Boxplot:

Graphische Darstellung der Daten einer Verteilung unter Benutzung von Lage- und Streuungsparametern. Dabei werden der Median, die Quartile, der Quartilsabstand und je nach Ausführung weitere Größen benutzt.

Neben einer Skala (Achse der reellen Zahlen) wird ein Rechteck vom unteren ( $q_{0.25}$ ) zum oberen ( $q_{0.75}$ ) Quartil gezeichnet, das durch den Zentralwert  $a_z$  geteilt wird.



# Graphische Darstellung

## Boxplot:

Graphische Darstellung der Daten einer Verteilung unter Benutzung von Lage- und Streuungsparametern. Dabei werden der Median, die Quartile, der Quartilsabstand und je nach Ausführung weitere Größen benutzt.

Neben einer Skala (Achse der reellen Zahlen) wird ein Rechteck vom unteren ( $q_{0.25}$ ) zum oberen ( $q_{0.75}$ ) Quartil gezeichnet, das durch den Zentralwert  $a_z$  geteilt wird.

# Graphische Darstellung

## Boxplot:

Graphische Darstellung der Daten einer Verteilung unter Benutzung von Lage- und Streuungsparametern. Dabei werden der Median, die Quartile, der Quartilsabstand und je nach Ausführung weitere Größen benutzt.

Neben einer Skala (Achse der reellen Zahlen) wird ein Rechteck vom unteren ( $q_{0.25}$ ) zum oberen ( $q_{0.75}$ ) Quartil gezeichnet, das durch den Zentralwert  $a_z$  geteilt wird.

# Graphische Darstellung

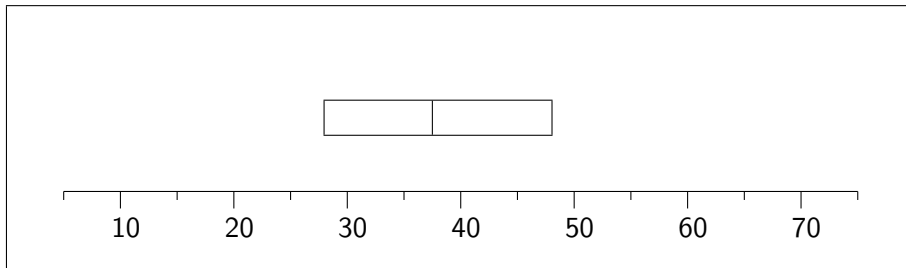


Abbildung 6.11 - Zur Konstruktion des Boxplots.

# Graphische Darstellung

## Boxplot:

Die außerhalb dieses Bereichs liegenden Merkmalswerte werden durch sogenannte **Fühler** wiedergegeben. Das sind die Werte, die am weitesten links und rechts liegen, aber nicht weiter als das 1.5-fache des Quartilsabstandes vom unteren bzw. oberen Quartil entfernt sind:

- unterer Anrainer

$$uA = \min\{x_{(i)} \mid x_{(i)} \geq q_{0.25} - 1.5QA\}$$

- oberer Anrainer

$$oA = \max\{x_{(i)} \mid x_{(i)} \leq q_{0.75} + 1.5QA\}$$

Merkmalswerte, die außerhalb der Anrainer liegen unterscheidet man noch zwischen **Außenpunkten** und **Fernpunkten**.

Außenpunkte haben einen Abstand von höchstens  $3 \cdot QA$  von den Quartilen. Fernpunkte liegen noch weiter außen.

# Graphische Darstellung

## Boxplot:

Die außerhalb dieses Bereichs liegenden Merkmalswerte werden durch sogenannte **Fühler** wiedergegeben. Das sind die Werte, die am weitesten links und rechts liegen, aber nicht weiter als das 1.5-fache des Quartilsabstandes vom unteren bzw. oberen Quartil entfernt sind:

- **unterer Anrainer**

$$uA = \min\{x_{(i)} \mid x_{(i)} \geq q_{0.25} - 1.5QA\}$$

- **oberer Anrainer**

$$oA = \max\{x_{(i)} \mid x_{(i)} \leq q_{0.75} + 1.5QA\}$$

Merkmalswerte, die außerhalb der Anrainer liegen unterscheidet man noch zwischen **Außenpunkten** und **Fernpunkten**. Außenpunkte haben einen Abstand von höchstens  $3 \cdot QA$  von den Quartilen. Fernpunkte liegen noch weiter außen.

# Graphische Darstellung

## Boxplot:

Die außerhalb dieses Bereichs liegenden Merkmalswerte werden durch sogenannte **Fühler** wiedergegeben. Das sind die Werte, die am weitesten links und rechts liegen, aber nicht weiter als das 1.5-fache des Quartilsabstandes vom unteren bzw. oberen Quartil entfernt sind:

- **unterer Anrainer**

$$uA = \min\{x_{(i)} \mid x_{(i)} \geq q_{0.25} - 1.5QA\}$$

- **oberer Anrainer**

$$oA = \max\{x_{(i)} \mid x_{(i)} \leq q_{0.75} + 1.5QA\}$$

Merkmalswerte, die außerhalb der Anrainer liegen unterscheidet man noch zwischen **Außenpunkten** und **Fernpunkten**. Außenpunkte haben einen Abstand von höchstens  $3 \cdot QA$  von den Quartilen. Fernpunkte liegen noch weiter außen.

# Graphische Darstellung

## Boxplot:

Die außerhalb dieses Bereichs liegenden Merkmalswerte werden durch sogenannte **Fühler** wiedergegeben. Das sind die Werte, die am weitesten links und rechts liegen, aber nicht weiter als das 1.5-fache des Quartilsabstandes vom unteren bzw. oberen Quartil entfernt sind:

- **unterer Anrainer**

$$uA = \min\{x_{(i)} \mid x_{(i)} \geq q_{0.25} - 1.5QA\}$$

- **oberer Anrainer**

$$oA = \max\{x_{(i)} \mid x_{(i)} \leq q_{0.75} + 1.5QA\}$$

Merkmalswerte, die außerhalb der Anrainer liegen unterscheidet man noch zwischen **Außenpunkten** und **Fernpunkten**.

Außenpunkte haben einen Abstand von höchstens  $3 \cdot QA$  von den Quartilen. Fernpunkte liegen noch weiter außen.

# Graphische Darstellung

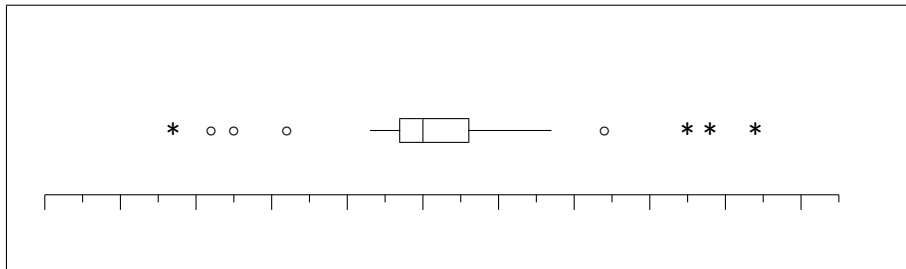


Abbildung 6.12 - Boxplot mit Außen-(o) und Fernpunkten (\*).



# Graphische Darstellung

## Beispiel 6.17:

In Beispiel 6.4 sind  $q_{0.75} = 48$ ,  $q_{0.25} = 28$  und damit der Quartilsabstand  $QA=20$ . Der Median ist  $a_z = q_{0.5} = 37.5$ . Unterer bzw. oberer Anrainer ist damit der kleinste bzw. größte beobachtete Wert.

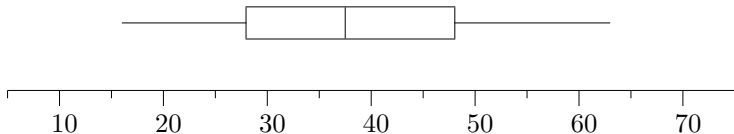


Abbildung 6.13 - Boxplot zu Beispiel 6.4.

# Graphische Darstellung

## Beispiel 6.17:

In Beispiel 6.4 sind  $q_{0.75} = 48$ ,  $q_{0.25} = 28$  und damit der Quartilsabstand  $QA=20$ . Der Median ist  $a_z = q_{0.5} = 37.5$ . Unterer bzw. oberer Anrainer ist damit der kleinste bzw. größte beobachtete Wert.

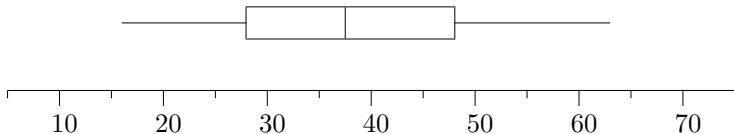


Abbildung 6.13 - Boxplot zu Beispiel 6.4.