Kapitel IV

Streuungs-, Schiefe und Wölbungsmaße[1]

B. 4. 1.
Lagemaße von Häufigkeitsverteilungen geben allein wenig Auskunft über eine Häufigkeitsverteilung. Sie beschreiben zwar ein Zentrum dieser Verteilung, geben aber keinen Anhaltspunkt dafür, wie weit ein konkreter Merkmalswert von einem solchen Zentrum abweichen kann.
Maße, die die Abweichung von einem Zentrum einer Häufigkeitsverteilung beschreiben, nennt man *Streuungsmaße* oder *Dispersionsmaße*.

D. 4. 1. (***Spannweite***)
Als *Spannweite* (bei einem nichtgruppierten Datenmaterial) bezeichnet man

$$R := x_{\max} - x_{\min}$$

B. 4. 2.
Bei einem gruppierten Datenmaterial ist

$$R \approx G_p - g_1$$

[1] "Ein Mensch, der von Statistik hört,
denkt dabei nur an Mittelwert,
Er glaubt nicht dran und ist dagegen,
ein Beispiel soll es gleich belegen:
Ein Jäger auf der Entenjagd
hat einen ersten Schuss gewagt.
Der Schuss, zu hastig aus dem Rohr,
lag eine gute Handbreit vor.
Der zweite Schuss mit lautem Krach
lag eine gute Handbreit nach.
Der Jäger spricht ganz unbeschwert
voll Glauben an den Mittelwert:
Statistisch ist die Ente tot.
Doch wär´ er klug und nähme Schrot
- dies sei gesagt ihn zu bekehren -
er würde seine Chancen mehren:
Der Schuss geht ab, die Ente stürzt,
weil Streuung ihr das Leben kürzt."
P. H. List
(Professor für pharmazeutische Technologie Marburg)

<u>BS. 4. 1.</u> (*Siehe BS. 3. 1.*)

1. <u>Nichtgruppiert</u>

$$R = 18.30 - 14.00 = 4.30 \text{ €}.$$

2. <u>Gruppiert</u>

$$R \approx 18.40 - 14.00 = 4.40 \text{ €}.$$

<u>B. 4. 3.</u>
Die Spannweite ist geeignet, falls

- man sich für den gesamten Streubereich interessiert.
- die beiden Randwerte eine bedeutende Rolle spielen.

Die Spannweite ist nicht geeignet

- bei großen Stichprobenumfängen.
- beim Auftreten von Ausreißern.
- um die Streuung der Grundgesamtheit zu schätzen.

<u>D. 4. 2.</u> (*Quartilsabstand*)
Als *Quartilsabstand* bezeichnet man

$$QA := \tilde{x}_{0.75} - \tilde{x}_{0.25}\,.$$

<u>BS. 3. 1.</u> (*Fortsetzung*)
1. *Nichtgruppiert*

$$QA = 17.05 - 15.60 = 1.45 \text{ €}.$$

2. *Gruppiert*

$$QA = 17.35 - 15.50 = 1.85 \text{ €}.$$

<u>D. 4. 3.</u> (*Mittlerer Quartilsabstand*)

$$\bar{Q} := \frac{1}{2}\left(\tilde{x}_{0.75} - \tilde{x}_{0.25}\right)$$

<u>BS. 3. 1.</u> (*Fortsetzung*)
1. *Nichtgruppiert*

$$\bar{Q} = \frac{1}{2}\left(17.05 - 15.60\right) = 0.725 \text{ €}.$$

2. *Gruppiert*

$$\bar{X} = \frac{1}{2}(17.35 - 15.50) = 0.925\,€.$$

<u>B. 4. 4.</u>
Im Vergleich zur Spannweite haben der Quartilsabstand und der mittlere Quartilsabstand den Vorteil, von den Extremwerten der Verteilung nicht beeinflusst zu werden.

<u>D. 4. 4.</u> (*Mittlere absolute Abweichungen*)
1. Als *mittlere absolute* (bzw. *lineare*) *Abweichung vom Median* bezeichnet man

$$d_{\tilde{x}_{0.5}} := \frac{1}{n}\sum_{i=1}^{n} |\, x_i - \tilde{x}_{0.5}\,|.$$

2. Als *mittlere absolute* (bzw. *lineare*) *Abweichung vom arithmetischen Mittel* bezeichnet man

$$d_{\bar{x}} := \frac{1}{n}\sum_{i=1}^{n} |\, x_i - \bar{x}\,|.$$

<u>BS. 3. 1.</u> (*Fortsetzung*)
1.
$$d_{\tilde{x}_{0.5}} = \frac{1}{30}\sum_{i=1}^{30} |\, x_i - 16.20\,| = \frac{1}{30}\cdot 27.8 \approx 0.93.$$
2.
$$d_{\bar{x}} = \frac{1}{30}\sum_{i=1}^{30} |\, x_i - 16.30\,| = \frac{1}{30}\cdot 28.2 \approx 0.94\,.$$

<u>B. 4. 5.</u> (*Minimaleigenschaft des Medianes*)
Es gilt:

$$\sum_{i=1}^{n} |\, x_i - \tilde{x}_{0.5}\,| \le \sum_{i=1}^{n} |\, x_i - Z\,|, \quad Z \in R^1: \text{ beliebig}.$$

<u>B. 4. 6.</u> (*Mittlere absolute Abweichungen bei einem gruppierten Datenmaterial*)
Es gilt:

$$d_{\tilde{x}_{0.5}} :\approx \frac{1}{n}\sum_{i=1}^{p} |\, \bar{x}_i - \tilde{x}_{0.5}\,| \cdot H_i\,,$$

$$d_{\bar{x}} :\approx \frac{1}{n}\sum_{i=1}^{p} |\, \bar{x}_i - \bar{x}\,| \cdot H_i\,.$$

Falls die Klassenmittel $\bar{x}_i$ nicht vorhanden sind, werden sie durch die Klassenmitten m_i ersetzt.

<u>BS. 3. 1.</u> (*Fortsetzung*)

Es ist

$$\tilde{x}_{0.5} = 16.29, \qquad \bar{x} = 16.40.$$

K_i	H_i	m_i	$\lvert m_i - 16.29 \rvert \cdot H_i$	$\lvert m_i - 16.40 \rvert \cdot H_i$
$K_1 = [14.00,\ 15.00[$	2	14.50	3.58	3.80
$K_2 = [15.00,\ 16.00[$	11	15.50	8.69	9.90
$K_3 = [16.00,\ 17.00[$	7	16.50	1.47	0.70
$K_4 = [17.00,\ 18.40[$	10	17.70	14.1	13.00
	30		27.84	27.40

$$d_{\tilde{x}_{0.5}} \approx \frac{27.84}{30} \approx 0.93, \qquad d_{\bar{x}} \approx \frac{27.4}{30} \approx 0.91.$$

<u>D. 4. 5.</u> (*Mittlere quadratische Abweichungen*)

1. Als *mittlere quadratische Abweichung vom Median* bezeichnet man

$$d^2_{\tilde{x}_{0.5}} := \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x}_{0.5})^2 .$$

2. Als *mittlere Abweichung vom arithmetischen Mittel* (bzw. *Varianz*) bezeichnet man

a) bei einer *Grundgesamtheit*

$$\sigma^2 := \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\left(= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2 \right).$$

b) bei einer *Stichprobe*

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

$$= \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}.$$

3. Als *Standardabweichung* bezeichnet man die <u>positive</u> Wurzel der Varianz.

<u>BS. 3. 1.</u> (*Fortsetzung*)

1.

$$d^2_{\tilde{x}_{0.5}} = \frac{1}{30}\sum_{i=1}^{30}(x_i - 16.20)^2 = \frac{37.79}{30} \approx 1.26\,.$$

2.

$$s^2 = \frac{1}{29}\sum_{i=1}^{30}(x_i - 16.30)^2 = \frac{37.49}{29} \approx 1.29\,.$$

3.. $s \approx 1.14$.

B. 4. 7.

1.

Es gilt:

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \le \sum_{i=1}^{n}\left(x_i - Z\right)^2, \quad Z \in R^1, \text{ beliebig.}$$

(*Minimaleigenschaft des arithmetischen Mittels*)

2.

Sei

$$d^2_{\bar{x}}(Z) := s^2 + (\bar{x} - Z)^2\,, \quad Z \in R^1 : \text{ beliebig.}$$

Hieraus folgt:

$$d^2_{\bar{x}}(Z) \ge s^2 \qquad \qquad (\textit{Verschiebungssatz})$$

B. 4. 8. (***Mittlere quadratische Abweichungen bei einem gruppierten Datenmaterial***)

Es gilt:

$$d^2_{\bar{x}} :\approx \frac{1}{n}\sum_{i=1}^{p}(\bar{x}_i - \tilde{x}_{0.5})^2 \cdot H_i\,,$$

$$\sigma^2 :\approx \frac{1}{N}\sum_{i=1}^{p}(\bar{x}_i - \mu)^2 \cdot H_i\,, \text{ (bei einer } \textit{Grundgesamtheit}\text{)}$$

$$s^2 :\approx \frac{1}{n-1}\sum_{i=1}^{p}(\bar{x}_i - \bar{x})^2 \cdot H_i\,, \text{ (bei einer } \textit{Stichprobe}\text{)}$$

Falls die Klassenmittel $\bar{x}_i$ nicht vorhanden sind, werden sie durch die Klassenmitten m_i ersetzt.

BS. 3. 1. (*Fortsetzung*)

Es ist

$$\tilde{x}_{0.5} = 16.29, \qquad \bar{x} = 16.40\,.$$

K_i	H_i	m_i	$(m_i - 16.29)^2 \cdot H_i$	$(m_i - 16.40)^2 \cdot H_i$
$K_1 = [14.00,\ 15.00[$	2	14.50	6.4082	7.22
$K_2 = [15.00,\ 16.00[$	11	15.50	6.8651	8.91
$K_3 = [16.00,\ 17.00[$	7	16.50	0.3087	0.07
$K_4 = [17.00,\ 18.40[$	10	17.70	19.8810	16.90
	30		33.4630	33.10

$$d_{\bar{x}}^2 \approx \frac{33.4630}{30} \approx 1.12, \qquad s^2 \approx \frac{33.10}{29} \approx 1.14, \qquad s \approx 1.07\ .$$

<u>B. 4. 7.</u> (*Sigma-Regeln)*

Im Intervall $\left[\bar{x}-\sigma \;\; \bar{x}+\sigma\right]$ liegt stets die Mehrheit, also mindestens 50 % aller Merkmalswerte

Für den Fall, dass die Merkmalswerte hinreichend genau normalverteilt sind, gilt die Folgende Sigma-Regel.

Im Intervall $\left[\bar{x}-k\cdot\sigma,\ \bar{x}+k\cdot\sigma\right]$ liegen für $k=1$ rund 68%, für $k=2$ rund 95% und für $k=3$ rund 99% aller Merkmalswerte.

<u>D. 4. 6.</u> (*Variationskoeffizient*)

Das Merkmal *X* möge nur positive Werte annehmen.

Als *Variationskoeffizient bezeichnet* man

$$v := \frac{\sigma}{\mu}, \quad \mu > 0 \quad \text{(bei einer } \textit{Grundgesamtheit}\text{)},$$

$$v := \frac{s}{\bar{x}}, \quad \bar{x} > 0 \quad \text{(bei einer } \textit{Stichprobe}\text{)}.$$

<u>B. 4. 8.</u>

Der Variationskoeffizient ist ein relatives Streuungsmaß, das keine Maßeinheit besitzt und in der Praxis meist in Prozent angegeben wird.

Er ist vor allem in zweierlei Hinsicht von praktischer Bedeutung:

1. Der Variationskoeffizient wird als eine Maßzahl benutzt, um einschätzen zu können, wie gut das arithmetische Mittel alle Einzelwerte repräsentiert. Dabei verwendet man die folgende <u>Faustregel</u>:

   Ein Variationskoeffizient größer als 0.5 bzw. 50% ist ein Indiz dafür, dass der Durchschnitt wegen einer zu großen Streuung kein geeigneter statistischer Repräsentant der Einzelwerte ist.

2. Der Variationskoeffizient ist eine geeignete Maßzahl für den Streuungsvergleich von gleich und/oder unterschiedlich dimensionierten Merkmals.

<u>BS. 3. 1.</u> (*Fortsetzung*)

1. Nichtgruppiert

$$v = \frac{1.14}{16.30} \approx 0.07\,.$$

2. Gruppiert

$$v = \frac{1.07}{16.40} \approx 0.07$$

<u>BS. 4. 1.</u>
Ein Fachgeschäft für Schrauben weist an einem bestimmten Wochentag folgende Verkaufszahlen in den Abteilungen A_1 und A_2:

Abteilung A_1		Abteilung A_2	
Verkaufsbetrag [€]	Anzahl der Verkäufe	Verkaufsbetrag [€]	Anzahl der Verkäufe
16.00	8	$[1.50,\ 2.50[$	4
30.00	20	$[2.50,\ 3.50[$	8
40.00	16	$[3.50,\ 4.50[$	20
50.00	12	$[4.50,\ 5.50[$	4
60.00	4	$[5.50,\ 6.50[$	2
		$[6.50,\ 7.50[$	2

1. Nennen und charakterisieren Sie das statistische Merkmal.

2. Berechnen Sie für jede Abteilung den durchschnittlichen Verkaufsbetrag.

3. Überprüfen Sie die Richtigkeit folgender Aussage mithilfe der entsprechenden *Variationskoeffizienten*: „Die Verkaufsbeträge in der Abteilung A_1 streuen stärker als in der Abteilung A_2.“

Lösung:

1.

Das Merkmal heißt: verkaufte Beträge. Es handelt sich (praktisch) um ein diskretes Merkmal.

2. und 3.

Abteilung A_1 :

Arbeitstabelle

a_j	$H(a_j)$	$a_j \cdot H(a_j)$	$(a_j - 36.80)^2 \cdot H(a_j)$
16	8	128	3461.12
30	20	600	924.80
40	16	640	163.84
50	12	600	2090.88
60	4	240	2152.96
	60	2208	8793.60

$$\bar{x} = \frac{2208}{60} = 36.80\ €, \quad s^2 = \frac{8793.60}{59} = 149.04, \quad s \approx 12.21, \quad v \approx 0.33. \quad .$$

Abteilung A_2 :

Arbeitstabelle

K_i	H_i	m_i	$m_i \cdot H_i$	$(m_i - 3.95)^2 \cdot H_i$
$[1.50,\ 2.50[$	4	2	8	15.210
$[2.50,\ 3.50[$	8	3	24	7.220
$[3.50,\ 4.50[$	20	4	80	0.050
$[4.50,\ 5.50[$	4	5	20	4.410
$[5.50,\ 6.50[$	2	6	12	8.405
$[6.50,\ 7.50[$	2	7	14	18.605
	40		158	53.900

$$\bar{x} = \frac{158}{40} = 3.95\ €, \quad s^2 = \frac{53.9}{39} = 1.38, \quad s \approx 1.18, \quad v \approx 0.30. \quad .$$

Wegen

$$v_{A_1} = 0.33 > 0.30 = v_{A_2}$$

ist die Aussage „Die Verkaufsbeträge in der Abteilung A_1 streuen stärker als in der Abteilung A_2 “ wahr.

D. 4. 7. (***k-tes Zentralmoment***)
Das $k-te$ *Zentralmoment* von n kardinalskalierten Merkmalswerten ist gegeben durch:

$$M_k := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{(für ein nichtgruppiertes Datenmaterial)}$$

$$M_k := \frac{1}{n}\sum_{i=1}^{p}(m_i - \bar{x})^2 \cdot H_i \qquad \text{(für ein gruppiertes Datenmaterial)}$$

B. 4. 9.
Die Varianz ist gleich M_2.

D. 4. 8. (***Schiefe***)
Die *Schiefe* der Verteilung eines kardinalskalierten Merkmals X sei gegeben durch:

$$S := \frac{M_3}{s^3}$$

B. 4. 10.
Die Schiefe gibt an, ob die Werte der Verteilung vom Modus aus links ($S > 0$) oder rechts ($S < 0$) schneller abfallen; das „lange Ende“ der Verteilung ist jeweils auf der anderen Seite. Im ersten Falle ist die Verteilung *linkssteil* bzw. *rechtsschief*; im zweiten Falle *rechtssteil* bzw. *linksschief*.

BS. 3. 1. (*Fortsetzung*)
1. Nichtgruppiert

$$S = \frac{\frac{6.32775}{30}}{1.14^3} \approx 0.14.$$

2. Gruppiert

H_i	m_i	$(m_i - 16.40)^3 \cdot H_i$
2	14.5	-13.718
11	15.5	-8.019
7	16.5	0.007
10	17.7	21.970
30		0.240

$$S = \frac{\frac{0.240}{30}}{1.07^3} \approx 0.007 \bullet$$

B. 4. 11. (***Weitere Schiefemaße***)
Es gibt weitere Schiefemaße, u. a.

1. Das Schiefemaß aus den Quartilen:

$$S_{\tilde{x}_\alpha} := \frac{\tilde{x}_{0.75} + \tilde{x}_{0.25} - 2\tilde{x}_{0.50}}{\tilde{x}_{0.75} - \tilde{x}_{0.25}} \qquad (-1 \le S_{\tilde{x}_\alpha} \le +1).$$

2. Das Schiefemaß nach Pearson:

$$S_P := \frac{\bar{x} - M}{s} \qquad (-1 \le S_{P1} \le +1).$$

3. Das Schiefemaß nach Yule-Pearson:

$$S_{Y-P} := \frac{3\left(\bar{x} - Me\right)}{s} \qquad (-3 \le S_{P2} \le +3).$$

BS. 3. 1. (*Fortsetzung*)

1. Nichtgruppiert:

$$S_{\tilde{x}_\alpha} = \frac{17.05 + 15.60 - 2 \cdot 16.20}{17.05 - 15.60} \approx 0.17,$$

$$S_P = \frac{16.30 - 15.60}{1.14} = 0.61,$$

$$S_{Y-P} = \frac{3(16.30 - 16.20)}{1.14} \approx 0.26.$$

2. Gruppiert

$$S_{\tilde{x}_\alpha} \approx \frac{17.35 + 15.50 - 2 \cdot 16.29}{17.35 - 15.50} \approx 0.15,$$

$$S_P \approx \frac{16.40 - 15.69}{1.05} \approx 0.66,$$

$$S_{Y-P} \approx \frac{3(16.40 - 16.29)}{1.14} \approx 0.29.$$

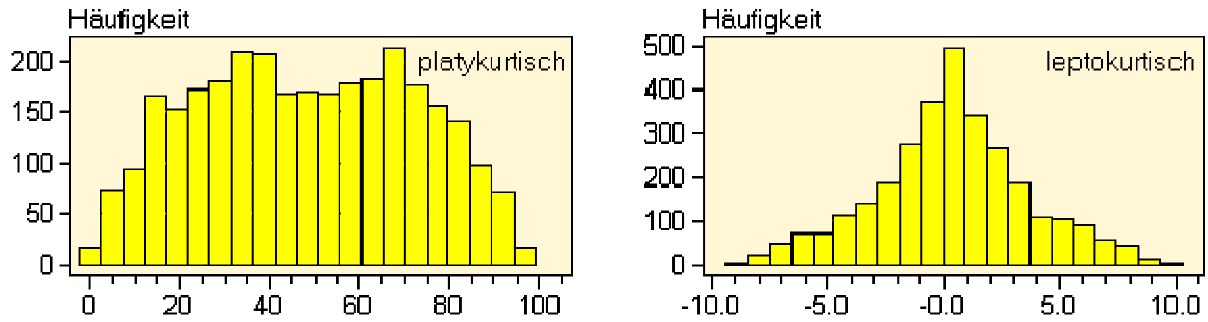
D. 4. 8. (***Exzess, Wölbung, „Kurtosis"***)
Der *Exzess* der Verteilung eines kardinalskalierten Merkmals X sei gegeben durch:

$$K := \frac{M_4}{s^4} - 3 \qquad \text{(für eine Stichprobe),}$$

$$K := \frac{n(n+1)}{(n-1)(n-2)(n-3)} \cdot \left(\frac{M_4}{d^4} \right) - \frac{3(n-1)^2}{(n-2)(n-3)} \qquad \text{(für eine Grundgesamtheit).}$$

B. 4. 12.
Der *Exzess* ist ein Maß für die relative „Flachheit" einer Verteilung (im Vergleich zur Normalverteilung, die einen Exzess von null aufweist.).
Ein positiver Exzess zeigt eine spitz zulaufende Verteilung (eine sog. *Leptokurtische* Verteilung), wohingegen ein negativer Exzess eine flache Verteilung (*platykurtische* Verteilung) anzeigt.

Hier zwei Beispiele von Verteilungen mit unterschiedlichem Exzess:



BS. 3. 1. (*Fortsetzung*)

1. Nichtgruppiert:

$$K = \frac{\frac{101.01425}{30}}{1.14^4} - 3 \approx -1.01$$

2. Gruppiert:

H_i	m_i	$(m_i - 16.40)^4 \cdot H_i$
2	14.5	26.0642
11	15.5	7.2171
7	16.5	0.0007
10	17.7	28.561
30		61.843

$$K \approx \frac{\frac{61.843}{30}}{1.07^4} - 3 \approx -1.43$$

(Letzte Aktualisierung: 01.03.2018)