Kapitel III

Lagemaße

<u>D. 3. 1.</u> (*Arithmetisches Mittel*)
Seien $x_1, x_2, ..., x_n$ die beobachteten Werte eines Merkmals X mit Ausprägungen $a_1, a_2, ..., a_k$.
Als *arithmetisches Mittel* (für nichtgruppierte Daten) bezeichnet man:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (\textit{einfaches})$$

$$= \frac{1}{n}\sum_{j=1}^{k} a_j \cdot H_n(a_j) \quad (\textit{gewogenes})$$

<u>BS. 3. 1.</u>
Betrachtet sei das Beispiel BS. 2. 1.

17.05,	17.80,	17.80,	14.70,	15.15,	18.30,
16.20,	16.20,	16.55,	17.05,	15.15,	15.60,
15.60,	15.60,	16.20,	14.00,	15.00,	15.60,
16.55,	18.00,	17.50,	17.05,	16.55,	15.60,
15.60,	15.00,	16.20,	18.30,	17.50,	15.60.

Der durchschnittliche Stundenlohn (in €) als einfaches arithmetisches Mittel lautet:

$$\bar{x} = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{489}{30} = 16.30\,€.$$

Zur Berechnung des gewogenen arithmetischen Mittels wird folgende Tabelle erstellt:

a_j	$H_n(a_j)$	$a_j \cdot H_n(a_j)$
14,00	1	14.00
14,70	1	14.70
15,00	2	30.00
15,15	2	30.30
15,60	7	109.20
16,20	4	64.80
16,55	3	49.65
17,05	3	51.15
17,50	2	35.00
17,80	2	35.60
18,00	1	18.00
18,30	2	36.60
Summe	30	489.00

Damit erhalten wir wieder:

$$\bar{x} = \frac{1}{30}\sum_{j=1}^{12} a_j \cdot H_n(a_j) = \frac{489}{30} = 16.30\,€$$

<u>B. 3. 1.</u> (*Arithmetisches Mittel bei gruppierten Daten*)
Sei p die Anzahl der Klassen. Dann gilt

$$\bar{x} \approx \frac{1}{n}\sum_{i=1}^{p} \bar{x}_i \cdot H_i$$

mit

$$\bar{x}_i := \frac{1}{H_i}\sum_{x_i \in K_i} x_i$$

Sind $\bar{x}_i,\ i = 1,\ 2, \ldots, p$, nicht bekannt, so werden sie durch die Klassenmitten

$$m_i := \frac{g_i + G_i}{2}, \quad i = 1,\ 2, \ldots, p,$$

ersetzt. Dabei sind g_i bzw. G_i die untere bzw. die Obere Grenze der Klasse K_i.
Es gilt folgende Abschätzung:

$$\frac{1}{n}\sum_{i=1}^{p} g_i \cdot H_i \le \bar{x} \le \frac{1}{n}\sum_{i=1}^{p} G_i \cdot H_i\ .$$

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

K_i	H_i	m_i	$m_i \cdot H_i$	$g_i \cdot H_i$	$G_i \cdot H_i$
$K_1 = [14.00,\ 15.00[$	2	14.50	29.00	28.00	30.00
$K_2 = [15.00,\ 16.00[$	11	15.50	170.50	165.00	176.00
$K_3 = [16.00,\ 17.00[$	7	16.50	115.50	112.00	119.00
$K_4 = [17.00,\ 18.40[$	10	17.70	177.00	170.00	184.00
	30		492.00	475.00	509.00

$$\bar{x} \approx \frac{492}{30} = 16.40\ \ €;$$

$$\frac{475}{30} = 15.83 \le \bar{x} \le 16.97 = \frac{509}{30}$$

Die Überschätzung des durchschnittlichen Stundenlohns bedeutet, dass die Werte in den einzelnen Klassen in der Regel links von dem jeweiligen arithmetischen Mittel liegen.

B. 3. 2. (***Einige wichtige Eigenschaften des arithmetischen Mittels***)
1.

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)=\sum_{i=1}^{n}x_i-n\,\bar{x}$$
$$=n\,\bar{x}-n\,\bar{x}=0\,.$$

2. (*Minimaleigenschaft des arithmetischen Mittels)*

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

3. Werden die Einzelwerte x_i einer linearen Transformation unterzogen:

$$x_i^*=\alpha+\beta x_i,\quad i=1,\ 2,...,n;\quad \alpha,\beta=const\in R^1\,,$$

dann gilt:

$$\begin{aligned}\bar{x}^*&=\frac{1}{n}\sum_{i=1}^{n}x_i^*=\frac{1}{n}\sum_{i=1}^{n}(\alpha+\beta x_i)\\&=\frac{1}{n}\left(n\cdot\alpha+\beta\cdot\sum_{i=1}^{n}x_i\right)\\&=\alpha+\frac{\beta}{n}\sum_{i=1}^{n}x_i\,,\end{aligned}$$

d. h.

$$\bar{x}^*=\alpha+\beta\,\bar{x}\,.$$

4. Zu berechnen sei das arithmetische Mittel $\bar{x}$ für eine Grundgesamtheit vom Umfang n, die in zwei oder mehr Teilgesamtheiten aufgeteilt ist, deren Umfänge und arithmetische Mittel bekannt sind.
   Beschränken wir uns auf zwei Teilgesamtheiten mit den Umfängen n_1 und n_2 $(n_1+n_2=n)$ und deren arithmetischen Mittel $\bar{x}_1$ und $\bar{x}_2$, dann ergibt sich

$$\begin{aligned}\bar{x}&=\frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1}x_{i1}+\sum_{i=1}^{n_2}x_{i2}\right)\\&=\frac{n_1\,\bar{x}_1+n_2\,\bar{x}_2}{n_1+n_2}\,.\end{aligned}$$

BS. 3. 2.
Eine Autovermietung berechnet für ihre Wagen eine feste Tagesgebühr von 40 € und einen Kilometersatz von 0.40 €/km. Ferner sei bekannt, dass die Wagen täglich im Durchschnitt 250 km zurücklegen.

Berechnen Sie die durchschnittlichen täglichen Einnahmen pro Wagen.

Lösung

$$\alpha = 40 \text{ €}, \quad \beta = 0.40 \text{ €}, \quad \bar{x} = 250 \text{ km}.$$

$$\bar{x}^* = 40 + 0.40 \cdot 250 = 140 \text{ €}.$$

<u>BS. 3. 3.</u>
Ein Unternehmen besteht aus den beiden Betrieben A und B. Die 400 Beschäftigten von A verdienen monatlich im Durchschnitt 2920.84 € und die 300 Beschäftigten von B monatlich im Durchschnitt 3012.17 €.
Berechnen Sie den durchschnittlichen Bruttomonatsverdienst sämtlicher 700 Beschäftigten von A und B. zusammen.

Lösung

$$n_1 = 400 \quad n_2 = 300, \quad \bar{x}_1 = 2920.84 \text{ €}, \quad \bar{x}_2 = 3012.98 \text{ €} .$$

$$\bar{x} = \frac{400 \cdot 2920.84 + 300 \cdot 3012.98}{400 + 300} = 2960.33 \text{ €} .$$

<u>D. 3. 2.</u> (*Median bzw. Zentralwert*)
Als *Median* bzw. *Zentralwert* bezeichnet man

$$Me := \begin{cases} x_{\left[\frac{n+1}{2}\right]} & \text{für } n: \text{ ungerade} \\ \frac{1}{2}\left(x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2}+1\right]}\right) & \text{für } n: \text{ gerade} \end{cases}$$

Dabei wird mit $[\ ]$ die Position des Merkmals in der <u>geordneten</u> Beobachtungsreihe dargestellt.

<u>BS. 3. 4.</u>
Gegeben sei die folgende Notenliste

5, 2, 2, 1, 4, 5, 3, 2, 1, 3, 4

Lösung

Position	1	2	3	4	5	6	7	8	9	10	11
Note	1	1	2	2	2	3	3	4	4	5	5

$$Me = x_{\left[\frac{11+1}{2}\right]} = x_{[6]} = 3$$

B. 3. 3.
Im Gegensatz zum arithmetischen Mittel wirken sich einzelne extrem hohe oder extrem tiefe Beobachtungswerte nicht in besonderer Weise auf den Median aus.

BS. 3. 5.
Gegeben sei folgende Liste des monatlichen Einkommens von 11 Mitarbeitern einer Firma in €.

2600 2500 2650 2720 5240
2450 2700 2750 2500 2750
2600.

Wie lautet das durchschnittliche Monatseinkommen dieser Mitarbeiter?

Lösung:
Das arithmetische Mittel

$$\bar{x} = \frac{31460}{11} = 2860.00 \text{ €}$$

stellt ein falsches Bild dar. Daher wird der Median berechnet und als durchschnittliches Monatseinkommen genommen:

Position	1	2	3	4	5	6	7	8	9	10	11
Einkommen	2450	2500	2500	2600	2600	2650	2700	2720	2750	2750	5240

$$Me = x_{\left[\frac{11+1}{2}\right]} = x_{[6]} = 2650.00 \text{ €}.$$

B. 3. 4. (***Median bei gruppierten Daten***)
Gegeben seien die Klassen $K_i,\ i = 1,\ 2,...,p$ mit den relativen Häufigkeiten $h_i,\ i = 1,\ 2,...,p$.
Gilt

$$\begin{cases} \sum_{j=1}^{i-1} h_j < 0.5 \\ \sum_{j=1}^{i} h_j \geq 0.5 \end{cases},$$

so liegt der Median in der Klasse K_i . Er wird berechnet nach der Formel:

$$Me \approx g_i + \frac{0.5 - \sum_{j=1}^{i-1} h_j}{h_i} \cdot b_i .$$

B. 3. 5.
Der Median teilt die gesamte Histogrammfläche in zwei gleichgröße Teile.

BS. 3. 1. (*Fortsetzung*)
Berechnen Sie für das gruppierte Datenmaterial den Median.

Lösung:

$$K_1 = [14.00,\ 15.00], \qquad h_1 = \frac{2}{30},$$
$$K_2 = [15.00,\ 16.00], \qquad h_2 = \frac{11}{30},$$
$$K_3 = [16.00,\ 17.00], \qquad h_3 = \frac{7}{30},$$
$$K_4 = [17.00,\ 18.40], \qquad h_4 = \frac{10}{30}.$$

$$h_1 = \frac{2}{30} < 0.5,$$
$$h_1 + h_2 = \frac{2}{30} + \frac{11}{30} = \frac{13}{30} < 0.5,$$
$$h_1 + h_2 + h_3 = \frac{2}{30} + \frac{11}{30} + \frac{7}{30} = \frac{20}{30} \geq 0.5,$$

d. h. der Median liegt in K_3. Genauer:

$$Me \approx 16 + \frac{0.5 - \frac{13}{30}}{\frac{7}{30}} \cdot 1 = 16.28571429 \approx 16.29\text{ €}.$$

D. 3. 3. ($\alpha -$ *Quantil*)
Als $\alpha -$ *Quantil*, bezeichnet mit $\tilde{x}_\alpha$ $(0 < \alpha < 1)$, einer geordneten Beobachtungsreihe bezeichnet man

$$\tilde{x}_\alpha := \begin{cases} x_{[k]} & \text{fall } n \cdot \alpha \text{ keine ganze Zahl ist.} \\ & (k \text{ ist dann die auf } n \cdot \alpha \text{ folgende ganze Zahl}) \\ \frac{1}{2}\left(x_{[k]} + x_{[k+1]}\right) & \text{falls } n \cdot \alpha \text{ eine ganze Zahl ist.} \\ & (\text{Es is dann } k = n \cdot \alpha.) \end{cases}$$

Dabei wird mit $[\ ]$ die Position des Merkmals in der <u>geordneten</u> Beobachtungsreihe dargestellt.

<u>BS. 3. 4</u> (*Fortsetzung*)
Er mitteln Sie 0.25-Quantil der Noten.

Lösung:

$$n.\alpha = 11 \cdot 0.25 = 2.75, \qquad k = 3 \bullet$$

Damit ist

$$\tilde{x}_{0.25} = x_{[3]} = 2\,.$$

<u>BS. 3.1.</u> (*Fortsetzung*)
Berechnen Sie das 0.75-Quantil für das nichtgruppierte Datenmaterial.

Lösung:

$$n.\alpha = 30 \cdot 0.75 = 22.5, \qquad k = 23\,.$$

Damit ist

$$\tilde{x}_{0.75} = x_{[23]} = 17.05 \text{ €}.$$

<u>B. 3. 6.</u> ($\alpha -$ *Quantil bei gruppierten Daten*)
Gegeben seien die Klassen $K_i,\ i = 1,\ 2,...,\ p$ mit den relativen Häufigkeiten $h_i,\ i = 1,\ 2,...,\ p$.
Gilt

$$\begin{cases} \sum_{j=1}^{i-1} h_j < \alpha \\ \sum_{j=1}^{i} h_j \geq \alpha \end{cases}, \qquad (0 < \alpha < 1)\,,$$

so liegt das $\alpha -$ Quantil in der Klasse $K_i \bullet$ Er wird berechnet nach der Formel:

$$\tilde{x}_\alpha \approx g_i + \frac{\alpha - \sum_{j=1}^{i-1} h_j}{h_i} \cdot b_i\,.$$

<u>BS. 3. 1.</u> (*Fortsetzung*)
Berechnen Sie für das gruppierte Datenmaterial das $\alpha -$ Quantil mit $\alpha = 0.25$ bzw. $\alpha = 0.75$

Lösung:
Sei $\alpha = 0.25$:

$$h_1 = \frac{2}{30} < 0.25\,,$$

$$h_1 + h_2 = \frac{2}{30} + \frac{11}{30} = \frac{13}{30} \geq 0.25\,,$$

d. h. $\tilde{x}_{0.25}$ liegt in K_2. Genauer:

$$\tilde{x}_{0.25} \approx 15 + \frac{0.25 - \frac{2}{30}}{\frac{11}{30}} \cdot 1 = 15.50\,€.$$

Sei $\alpha = 0.75$:

$$h_1 = \frac{2}{30} < 0.75\,,$$
$$h_1 + h_2 = \frac{2}{30} + \frac{11}{30} = \frac{13}{30} < 0.75\,,$$
$$h_1 + h_2 + h_3 = \frac{2}{30} + \frac{11}{30} + \frac{7}{30} = \frac{20}{30} < 0.75\,,$$
$$h_1 + h_2 + h_3 + h_4 = 1 \geq 0.75$$

d. h. $\tilde{x}_{0.75}$ liegt in K_4. Genauer:

$$\tilde{x}_{0.75} \approx 17 + \frac{0.75 - \frac{20}{30}}{\frac{10}{30}} \cdot (18.40 - 17.00) = 17.35\,€.$$

B. 3. 7.
Es ist

$$Me = \tilde{x}_{0.5}\,.$$

D. 3. 4. (*Quartile*)

$\tilde{x}_{0.25}$ wird als *unteres Quartil* bezeichnet und $\tilde{x}_{0.75}$ als *oberes Quartil*.

D. 3. 5 (*Prozentil*)
Liegt α als Prozent vor, so wird das entsprechende Quantil auch als *Prozentil* bezeichnet.

D. 3. 6 (*Modus* bzw. *Modalwert*)
Als *Modus* bezeichnet man diejenige Merkmalsausprägung, die eindeutig am häufigsten vorkommt.
Im Falle eines gruppierten Datenmaterials wird er in der Klasse K_i mit der maximalen Häufigkeit folgendermaßen berechnet:

$$M :\approx g_i + \frac{h_i - h_{i-1}}{2h_i - h_{i-1} - h_{i+1}} \cdot b_i\,.$$

<u>BS. 3. 1.</u> (*Fortsetzung*)
Berechnen Sie den Modus für das

1. nichtgruppierte
2. gruppierte

Datenmaterial.

1.

$M = 15.60$ € (mit der maximalen Häufigkeit 7).

2.

$$M \approx 15 + \frac{\frac{11}{30} - \frac{2}{30}}{2 \cdot \frac{11}{30} - \frac{2}{30} - \frac{7}{30}} \cdot 1 = 15.69 \text{ €}$$

<u>B. 3. 8.</u>
Es ist grundsätzlich möglich, die Definition des Modus auch auf den mehrdeutigen Fall zu erweitern. Für den Fall, dass zwei Modi vorkommen, spricht man von einer *bimodalen Verteilung*.

<u>B. 3. 9.</u>
Bei nominalen Merkmalen ist der Modus der entscheidende und einzig mögliche Lageparameter.

<u>B. 3. 10</u> (*Beziehung zwischen einigen Lagemaßen*)
Im Falle einer symmetrischen Verteilung gilt

$$M = Me = \bar{x}\,.$$

Bei rechtssteilen (bzw. linksschiefen) Verteilungen gilt:

$$\bar{x} < Me < M\,.$$

Analog gilt für die linkssteile (bzw. rechtsschiefe) Verteilungen

$$Me < \bar{x} < M.$$

<u>B. 3. 11.</u>
Der Modus ist besser geeignet, falls

- das Histogramm mehrere Spitzen hat.
- das Merkmal nominal skaliert ist.

Das arithmetische Mittel ist besser geeignet, falls

- der Stichprobenumfang nicht sehr klein ist.
- die Verteilung nicht allzu asymmetrisch ist.
- es nicht zu viele Ausreißer gibt.

Der Median ist besser geeignet, falls

- der Stichprobenumfang klein ist.
- die Verteilung stark asymmetrisch ist.
- es viele Ausreißer gibt.
- das Merkmal ordinal skaliert ist.

D. 3. 7 (*Ungewogenes harmonisches Mittel*)

Gegeben seien die Beobachtungswerte $x_i > 0,\ i,\ 2,...,n$, eines Merkmals X. Als *(ungewogenes) harmonisches Mittel* bezeichnet man:

$$\bar{x}_h := \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}},$$

B. 3. 12.

Die Berechnung eines harmonischen Mittels ist geboten, wenn z. B. einen Durchschnitt zu berechnen gilt aus

a) Preisen bei gegebenen Umsatzinformationen,
b) Hektarerträgen bei gegebenen Ertragsinformationen,
c) Geschwindigkeiten bei gegebenen Wegstreckeinformationen
d) Stückkosten bei gegebenen Kosteninformationen,
e) Arbeitslosenquoten bei gegebenen Arbeitsloseninformationen
f) Bevölkerungsdichten bei gegebenen Bevölkerungsinformationen.

BS. 3. 6.

Ein Pkw legt vier Teilstrecken einer Gesamtstrecke mit folgenden Geschwindigkeiten zurück:

Teilstrecke Nr.	1	2	3	4
Länge der Teilstrecke in km	30	10	40	20
Geschwindigkeit km/h	40	50	80	100

Durch welche (entlang der Gesamtstrecke konstant gehaltene) Durchschnittsgeschwindigkeit würde er die Gesamtstrecke in der gleichen Zeit bewältigen?

Lösung:

$$\tilde{x}_h = \frac{100}{\frac{30}{40} + \frac{10}{50} + \frac{40}{80} + \frac{20}{100}} = 69 \text{ km/h}.$$

BS 3. 7.

Ein Kapital von 1000 € ist in Aktien angelegt. Im 1. Jahr vermehrt sich das Kapital überhaupt nicht, im 2. Jahr erhöht es sich um 10%, im 3. Jahr um 20%.
Berechnen Sie die durchschnittliche Wachstumsrate des Kapitals.

Lösung:
Betrachten wir zunächst die tatsächliche Entwicklung des Kapitals:

Jahr n	Kapital am Ende des Jahres n K_n
1	$K_1 = 1000$ €
2	$K_2 = 1000 + 1000 \cdot 0.10 = 1100.00$ €
3	$K_3 = 1100 + 1100 \cdot 0.20 = 1320.00$ €

Es wird nun gezeigt, dass das arithmetische Mittel nicht geeignet ist, die durchschnittliche Wachstumsrate des Kapitals darzustellen:

$$\bar{x} = \frac{0.00 + 0.10 + 0.20}{3} = 0.10$$

Jahr n	Kapital am Ende des Jahres n K_n
1	$K_1 = 1000 + 1000 \cdot 0.10 = 1100.00$ €
2	$K_2 = 1100 + 1100 \cdot 0.10 = 1210.00$ €
3	$K_3 = 1210 + 1210 \cdot 0.10 = 1331.00$ €

Das Ergebnis $K_3 = 1331.00$ € entspricht nicht der Realität!

D. 3. 8. (***Geometrisches Mittel***)

1. Gegeben seien die Beobachtungswerte $x_1, x_2, ..., x_n > 0$ eines Merkmals X. Als *ungewogenes geometrisches Mittel* bezeichnet man

$$\bar{x}_g := \sqrt[n]{\prod_{i=1}^{n} x_i} \ .$$

2. Gegeben sei die Häufigkeitsverteilung des Merkmals X, dessen Ausprägungen $a_i,\ i = 1,2,...,k$, mit den absoluten Häufigkeiten $H_n(a_i),\ i = 1,2,...,k$, bzw. den relativen Häufigkeiten $h_n(a_i),\ i = 1,2,...,k$, auftreten.
   Als *gewogenes geometrisches Mittel* bezeichnet man

$$\bar{x}_g := \sqrt[n]{\prod_{i=1}^{n} a_i^{H_n(a_i)}}$$

$$= \prod_{i=1}^{k} a_i^{h_n(a_i)} .$$

<u>BS 3. 7.</u> (*Fortsetzung*)

$$\bar{x}_g := \sqrt[3]{(1+0.00)(1+0.10)(1+0.20)} = 1.09696131,$$

d, h. die tatsächliche durchschnittliche Wachstumsrate des Kapitals beträgt 9.696131 %, was die nachfolgende Tabelle bestätigt:

Jahr n	Kapital am Ende des Jahres n K_n
1	$K_1 = 1000 + 1000 \cdot 0.09696131 = 1096.96$ €
2	$K_2 = 1096.96 + 1096.96 \cdot 0.09696131 = 1203.32$ €
3	$K_3 = 1203.32 + 1203.32 \cdot 0.09696131 = 1320.00$ €

<u>B. 3. 13.</u>

1. Das geometrische Mittel wird oft auch dort als Lagemaß verwendet, wo eine logarithmische Skala sinnvoll ist, denn es gilt:

$$\ln \bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

$$= \frac{1}{n}(\ln x_1 + \ln x_2 + \ldots + \ln x_n) = \frac{1}{n}\sum_{i=1}^{n} \ln x_i \quad \text{(arithmetisches Mittel von } \ln x_i\text{)}.$$

2. Es gilt

$$\bar{x}_g \le \bar{x}$$

$$(\bar{x}_g = \bar{x} \quad \Leftrightarrow \quad x_1 = x_2 = \ldots = x_n)$$

(Letzte Aktualisierung: 23.06.08)