

## 8. Lageparameter

- Leitfragen

1) Warum ist die Lage einer Verteilung für das Ergebnis einer statistischen Analyse von Bedeutung?

2) Was ist der Unterschied zwischen Parametern der Lage und der Streuung?

3) Was sind die wichtigsten Lageparameter?

4) Wie berechnet und interpretiert man Lageparameter?

5) Gibt es Besonderheiten bei Lagemaßen klassierter Verteilungen?

6) Wie erkennt man Lageparameter in grafischen Darstellungen?

### 8.1 Lage einer Verteilung

Neben der Beurteilung der Verteilungsform, die in den letzten beiden Kapiteln ausführlich diskutiert wurde, ist die Lage einer Verteilung von Bedeutung.

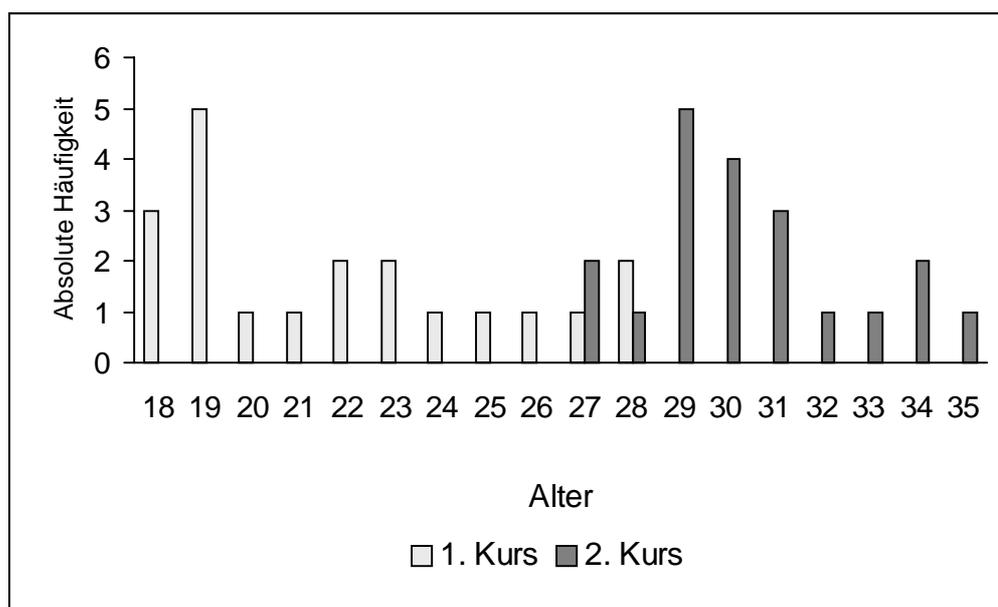


Abb. 8.1: Lage zweier Verteilungen

Die obige Abb. 8.1 stellt zwei Altersverteilungen mit jeweils unterschiedlichem Zentrum dar. Der **erste Kurs** entspricht den Seminarteilnehmern unseres **Master-Beispiels** und wird im weiteren Verlauf dieses Kapitels mittels verschiedener Lageparameter genauer analysiert.

Wir sehen, daß die Teilnehmer des ersten Kurses insgesamt jünger sind als die des zweiten Kurses. Der grafische Vergleich der Verteilungen deutet auch eine etwas größere „Streubreite“ der linken Verteilung an. Die exakte Streuung wird mit speziellen Parametern gemessen (siehe Kapitel 9).

Lageparameter werden benötigt, um typische, zentrale Werte einer Verteilung zu erhalten. Sie sagen grundsätzlich nichts über den Streubereich der Verteilung aus. Die Lageparameter charakterisieren als Kennzahlen das Zentrum (die Mitte, den Durchschnitt) und damit die Lage der Verteilung. Die Berechnung von Lageparametern kann anhand der ungeordneten Urliste, der geordneten Urliste oder der Häufigkeitstabelle erfolgen. Hierzu die Angaben des **Beispiels aus unserem Master-Projekt**:

- die ungeordnete Urliste der Altersangaben:  
19, 18, 22, 23, 19, 20, 25, 27, 18, 23, 19, 19, 26, 28, 24, 28, 19, 18, 21, 22
- die geordnete Urliste:  
18, 18, 18, 19, 19, 19, 19, 19, 20, 21, 22, 22, 23, 23, 24, 25, 26, 27, 28, 28
- die Häufigkeitsverteilung; siehe Abb. 8.1 (Kurs 1) oder Abb. 8.3.

Zunächst werden die **elementaren Lageparameter** vorgestellt: der Modus (häufigster Wert), der Median (Zentralwert) und das arithmetische Mittel (Mittelwert):

- **Modus ( Modalwert, häufigster Wert)**

= Wert mit maximaler Häufigkeit

⇒ **19** (da Häufigkeit für den Wert 19: 5-mal = Maximum aller Häufigkeiten)

- **Median (Zentralwert)**

= Wert, der die Verteilung in zwei gleich große Teile aufteilt

= Wert, der die geordnete Urliste halbiert

= Mitte der Verteilung mit jeweils 50% der Werte „darunter und darüber“

⇒ Mittelwert des 10. und 11. Werts der geordneten Urliste:

$$21,5 = \frac{1}{2} (21 + 22)$$

- **Arithmetisches Mittel (Mittelwert)**

= Summe aller Einzelwerte / Anzahl der Elemente

= Durchschnittsalter

⇒ **21,90** =  $(19 + 18 + 22 + \dots + 22) / 20 = 438 / 20$ .

Für die exakte Erläuterung der Berechnungsformeln der einzelnen Lageparameter werden im weiteren Verlauf Arbeitstabellen verwendet.

- **Berechnung von Lageparametern am PC**

Die Berechnung der Lageparameter erfolgt mit dem PC-Programm EXCEL grundsätzlich anders als mit dem Statistikprogramm SPSS.

➤ **Berechnung mit EXCEL**

Bei der Berechnung mit **EXCEL** wird die ungeordnete Urliste als EXCEL-Tabelle angelegt. Nach dem Markieren der Werte können dann über den FUNKTIONSSASSISTENTEN diverse Lageparameter (**MITTELWERT, MEDIAN, MODALWERT, QUANTIL, GEOMITTEL** etc.) abgerufen und in einem festgelegten Tabellenfeld ausgegeben werden.

Diese Berechnung lässt sich leicht anhand einer EXCEL-Tabelle nachvollziehen (vgl. Kapitel 2, Abb. 2.4). Eine weitere Möglichkeit besteht darin, die Arbeitstabellen der Lageparameter (siehe Abb. 8.3) in EXCEL abzubilden und die Berechnung selbst durchzuführen. Hier bleibt es dem Leser überlassen, die Tabellen in EXCEL selbst einzugeben und durchzurechnen (siehe auch Kapitel 2, Abb. 2.2).

➤ **Berechnung mit SPSS**

Mit dem Statistikprogramm SPSS können Lageparameter über mehrere Programmfunktionen abgerufen werden. Dies sind insbesondere die Funktion **DESKRIPTIVEN STATISTIKEN** (siehe im weiteren Verlauf dieses Kapitels) oder die **HÄUFIGKEITEN**, die wie folgt zu bedienen sind:

**SPSS-Dialog: ANALYSIEREN – DESKRIPTIVE STATISTIKEN – HÄUFIGKEITEN – STATISTIK (Häufigkeitstabelle nicht anzeigen).**

Für unser obiges Beispiel ergibt folgende Ergebnisausgabe:

ALTER		
N	Gültig	20
	Fehlend	0
Mittelwert		21,9000
Median		21,5000
Modus		19,00

Abb. 8.2: Lageparameter mit SPSS

Auf die PC-Aspekte der Berechnung von Lageparametern wird im weiteren Verlauf immer dann eingegangen, wenn sich Besonderheiten ergeben, die von der eben besprochenen, sehr einfachen Anwendung abweichen. Nun zu den Lageparametern und ihren Berechnungsformeln im einzelnen.

## 8.2 Modus

Der Modus ist der häufigste Wert einer Verteilung. Wir können den Modus, wie in Abschnitt 8.1 gezeigt, sehr einfach ermitteln, wenn eine der Häufigkeiten größer ist als alle anderen. Bedeutung und Interpretation des Modus sind danach zu unterscheiden, ob eine einfache oder eine klassierte Verteilung – möglicherweise mit ungleichen Klassenbreiten – vorliegt. Auch macht es einen Unterschied, ob eine „maximale“ Häufigkeit nur unwesentlich über den anderen Häufigkeiten liegt, oder ob dieses Maximum deutlich aus der Verteilung herausragt.

- **Modus der Häufigkeitsverteilung**

Zur Ermittlung des Modus bei nicht-klassierten Häufigkeitsverteilungen genügt eine Analyse der absoluten oder der relativen Häufigkeiten. Im Beispiel unserer Altersverteilung (siehe Abb. 8.3) weist der Wert 19 mit einer absoluten Häufigkeit von 5 und einer relativen Häufigkeit von 0,25 bzw. 25% ein deutliches Maximum auf und ist daher der Modus: Teilnehmer mit einem Alter von 19 Jahren sind in der untersuchten Gesamtheit am häufigsten vertreten.

Fragen nach dem häufigsten Wert lassen sich in der Praxis sehr oft finden: Welche Klausurnote ist am häufigsten? Welche Zeitung wird am meisten gelesen? Welche Partei hat die meisten Stimmen? Welches Land wird am häufigsten als Reiseziel gewählt?

j	$x_j$	$f_j$	$h_j$	$h_j'$
1	18	3	0,15	15,0
2	19	5	0,25	25,0
3	20	1	0,05	5,0
4	21	1	0,05	5,0
5	22	2	0,10	10,0
6	23	2	0,10	10,0
7	24	1	0,05	5,0
8	25	1	0,05	5,0
9	26	1	0,05	5,0
10	27	1	0,05	5,0
11	28	2	0,10	10,0
Summe	-	20	1,00	100,0

Abb. 8.3: Arbeitstabelle Modus

Die Berechnung des Modus ist nicht mehr ganz so einfach wie oben gezeigt, wenn die Verteilung bestimmte Besonderheiten aufweist.

- **Modus der klassierten Verteilung mit gleichen Klassenbreiten**

Bei einer klassierten Verteilung mit gleichen Klassenbreiten kann der Modus ähnlich wie bei der nicht-klassierten Verteilung ermittelt werden. Die Klassenhäufigkei-

ten können wie die gewöhnlichen Häufigkeiten behandelt werden, nur bei der Bestimmung des konkreten Werts für den Modus müssen wir speziell vorgehen:

- Einfallsklasse des Modus ermitteln (= Klasse mit der maximalen Häufigkeit)
- Klassenmitte der Einfallsklasse ist der Näherungswert für den Modus.

Greifen wir auf das Klassierungsbeispiel aus Kapitel 7 zurück (siehe Histogramm in Abb. 7.5), dann ist die erste Klasse die Einfallsklasse des häufigsten Werts. Mit der Klassenmitte von 17,5 erhalten wir den Näherungswert für den Modus.

#### • **Modus der klassierten Verteilung mit ungleichen Klassenbreiten**

Bei klassierten Häufigkeitsverteilungen mit unterschiedliche Klassenbreiten, dürfen wir nicht die Klassenhäufigkeit zur Ermittlung der Einfallsklasse des Modus heranziehen. Jetzt wird die Besetzungsdichte als „normierte“ Klassenhäufigkeit verwendet. Damit gilt für die Ermittlung des Modalwerts:

- Einfallsklasse des Modus ermitteln (= Klasse mit der maximalen Besetzungsdichte)
- Klassenmitte der Einfallsklasse ist der Näherungswert für den Modus.

Auch hier wollen wir auf ein Beispiel aus Kapitel 7 und das entsprechende Histogramm in Abb. 7.7 zurückkommen. Während die Häufigkeit der 1. Klasse größer ist als alle anderen Klassenhäufigkeiten (siehe Tabelle in Abb. 7.6), stellt bei den Besetzungsdichten (und damit auch im Histogramm) die 2. Klasse das Maximum dar. Damit können wir die Klassenmitte von 250 DM als häufigste BAFÖG-Zahlung angeben.

#### • **Verteilungen mit mehreren Modi**

Werden in einer Verteilung mehrere Werte oder Klassen mit exakt gleich großer „maximaler“ Häufigkeit bzw. Besetzungsdichte vorgefunden, spricht man von einer multimodalen Verteilung. Die Häufigkeitsverteilung hat mehrere Modi. Die rechnerische Ermittlung des einzelnen Modus erfolgt wie oben gezeigt.

Hinweis zur PC-Anwendung: **SPSS** und **EXCEL** gehen bei multimodalen Verteilungen unterschiedlich vor. Bei mehreren gleich großen Häufigkeiten gibt SPSS an, daß mehrere Modalwerte vorhanden sind und zeigt den kleinsten davon an. EXCEL gibt „ohne Kommentar“ den kleinsten der gleich häufigen Werte als Modalwert an.

#### • **Verteilungen ohne Modus**

Sind alle Häufigkeiten einer Verteilung gleich groß, so ergibt sich kein Modus. Es liegt eine Gleichverteilung (ohne häufigsten Wert) vor.

Ergänzend sei noch darauf hingewiesen, daß zur Interpretation des Modus auch die Unterschiede der Häufigkeiten betrachtet werden sollten. So kann es sein, daß in einer Verteilung mit mehreren Tausend Elementen jeder Wert etwa 100-mal vorkommt. Hier würde ein Modus mit einer Häufigkeit von z.B. 103 nur unwesentlich häufiger auftreten als andere Werte mit einer Häufigkeit von knapp unter 100.

Und schließlich ist auch der Fall multimodaler Verteilungen so vorstellbar, daß die (lokalen) Häufigkeitsmaxima nicht exakt gleich groß sind. Die einzelnen Modalwerte können in ihrer unmittelbaren Umgebung ein lokales Maximum darstellen und trotzdem Häufigkeiten aufweisen, die sich untereinander unterscheiden. Hier sind dann spezielle Verteilungsanalysen gefragt, z.B. die Clusteranalyse, die Gruppierungen und Ballungen von Elementen entdecken will.

Zusammengefaßt gelten für den Modus folgende Definitionen und Formeln.

**Modus** = Häufigster Wert einer Verteilung

= Wert mit maximaler absoluter oder maximaler relativer Häufigkeit

= Klassenmitte der Klasse mit maximaler Klassenhäufigkeit  
(bei gleichen Klassenbreiten)

= Klassenmitte der Klasse mit maximaler Besetzungsdichte  
(bei ungleichen Klassenbreiten)

### Modus

#### - bei nicht-klassierter Verteilung

(8-1)  $M = x_D$                       wobei  $f_D = \max(f_j)$  bzw.  $h_D = \max(h_j)$   
mit  
 $f_j$  = absolute Häufigkeit  
 $h_j$  = relative Häufigkeit

#### - bei klassierter Verteilung mit gleichen Klassenbreiten

(8-2)  $M = (a_E + b_E) / 2$             wobei  $f_E = \max(f_j)$  bzw.  $h_E = \max(h_j)$   
mit  
 $a_E, b_E$  = Klassenunter-/obergrenze der  
Einfallsklasse E des Modus  
 $f_j, h_j$  = Klassenhäufigkeiten

#### - bei klassierter Verteilung mit ungleichen Klassenbreiten

(8-3)  $M = (a_E + b_E) / 2$             wobei  $f_E^* = \max(f_j^*)$  bzw.  $h_E^* = \max(h_j^*)$   
mit  
 $a_E, b_E$  = Klassenunter-/obergrenze der  
Einfallsklasse E des Modus  
 $f_j^*, h_j^*$  = Besetzungsdichten der Klassen