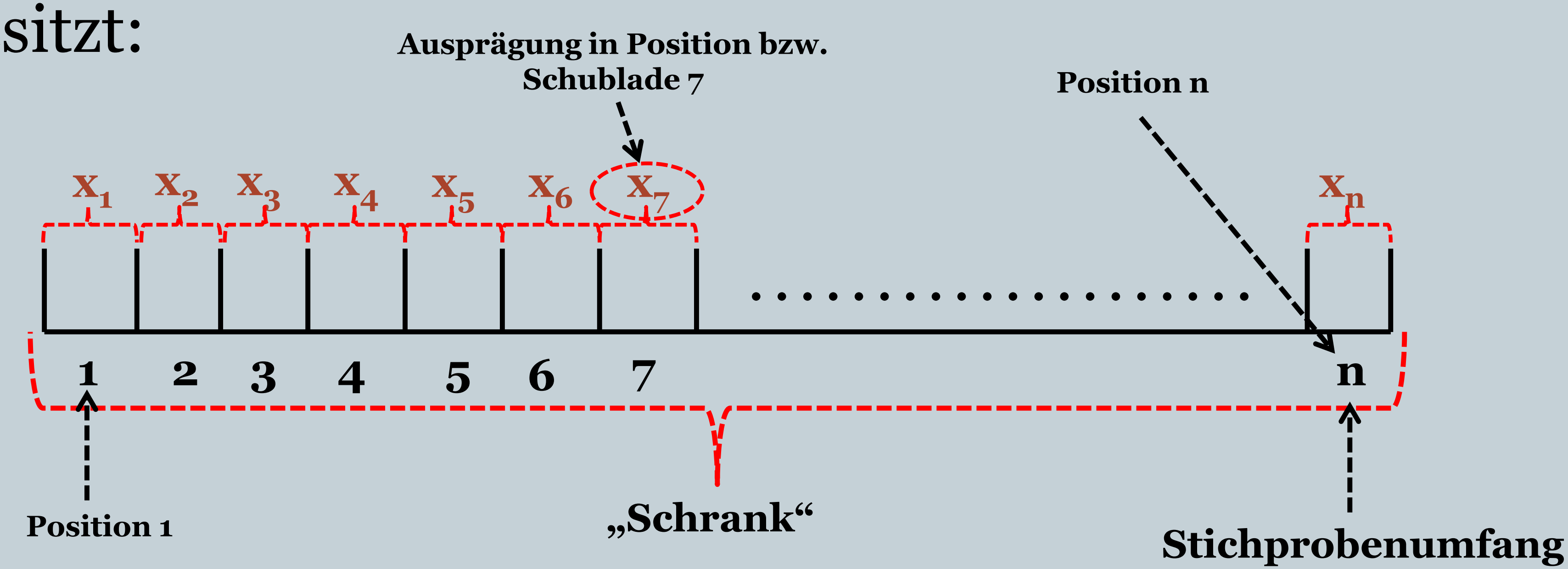
Lagemaße – Übung

MODUS, MEDIAN, MITTELWERT, MODAL KLASSE, MEDIAN, KLASSE, INTERPOLATION DER MEDIAN, KLASSE MITTE

Zentrale Methodenlehre, Europa Universität - Flensburg

Stichprobe: abstrakte Struktur

Man könnte sich eine Stichprobe als einen Schrank mit Schubladen vorstellen, dessen Inhalt nur einen Wert besitzt:



Komponenten:

- **n Positionen bzw. Schubladen bzw. Stellen: 1, 2, ….., n**
- **X_i : Ausprägung in der Position bzw. Schublade i**

Lagemaße: Modus

Modus der Stichprobe (D): Die häufigste Ausprägung eines Merkmals.

28. Denkst du, du könntest dabei mitmachen, wenn ein/e Schüler/in, den/die du nicht magst, gemobbt wird?

1 Ja
2 Ja, vielleicht
3 Ich weiß es nicht
4 Nein, ich denke eher nicht
5 Nein, auf gar keinen Fall

Quelle: Bundeszentrale für politische Bildung; Projekt: Mobbing bei uns nicht!?
http://www.bpb.de/lernen/unterrichten/grafstat/46487/projekt-mobbing-bei-uns-nicht

Diese Frage wurde 20 Schülern gestellt, es gab hierauf folgende Antworten:

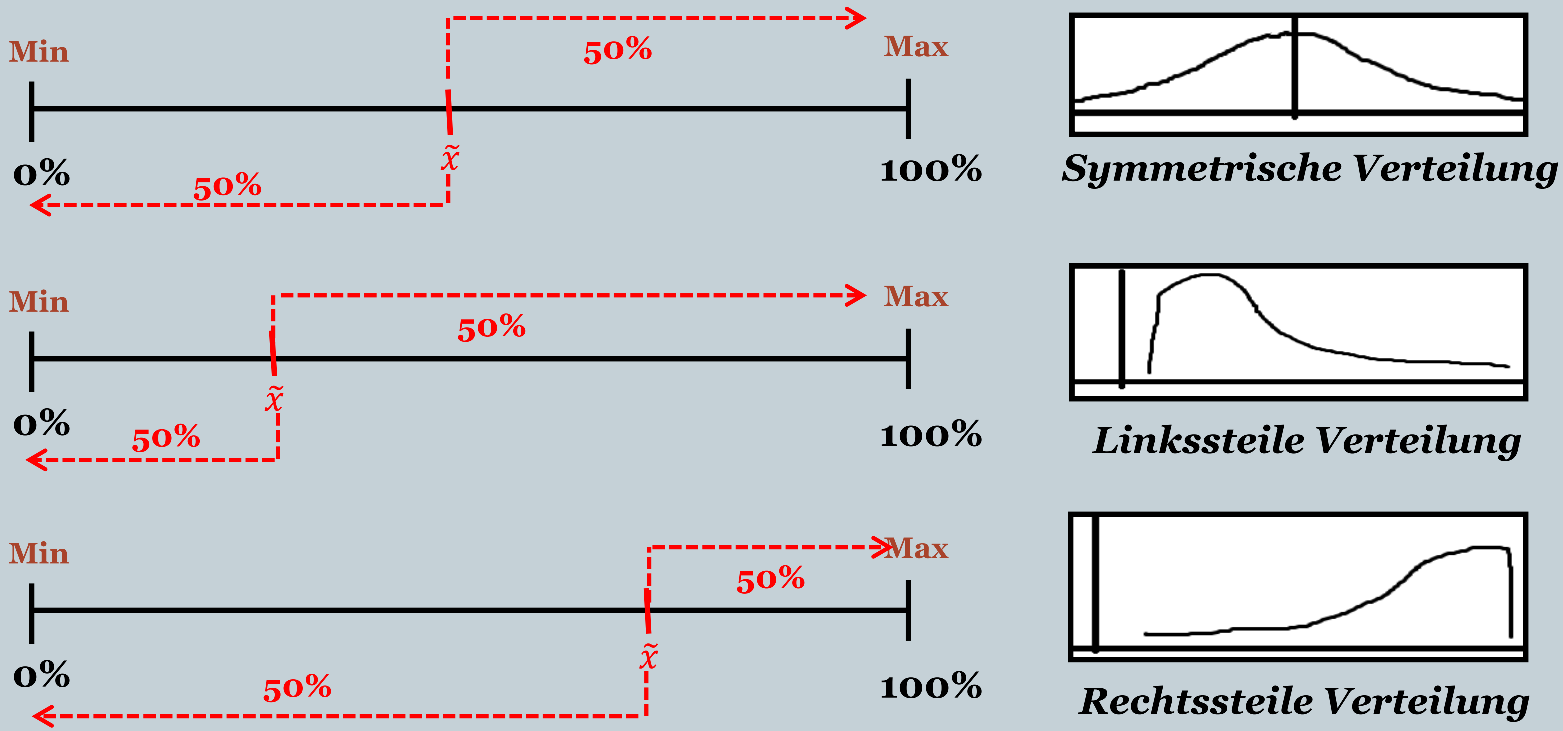
x_1 =	x_2 =																		
5	4	5	5	5	5	1	5	5	3	4	2	3	4	3	1	4	5	5	5
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

$f_5 = 10 \rightarrow D = 5$, das heißt: die häufigste Antwort der befragten Schüler/-innen war „Nein, auf gar keinen Fall“.

Lagemaße: Median

Median der Stichprobe ($\tilde{x}$): Die Ausprägung eines Merkmals, die genau in der Mitte liegt.

Wichtig: Um den Median zu bestimmen, muss die Stichprobe nach Größe geordnet werden.



Lagemaße: Median

- Median mit Häufigkeitstabellen:

USA: GSS 2010

THINK OF SELF AS LIBERAL OR CONSERVATIVE

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte abs. Häufigkeit	Kumulierte Prozente
Gültig	EXTREMELY LIBERAL	76	3.7	3.9	76	3.9
	LIBERAL	259	12.7	13.1	335	17.0
	SLIGHTLY LIBERAL	232	11.4	11.8	567	28.7
	MODERATE	746	36.5	37.8	1313	66.5
	SLGHTLY CONSERVATIVE	265	13.0	13.4	1578	80.0
	CONSERVATIVE	315	15.4	16.0	1893	95.9
	EXTRMLY CONSERVATIVE	80	3.9	4.1	1973	100.0
	Gesamt	1973	96.5	100.0		
Fehlend	DK	61	3.0			
	NA	10	.5			
	Gesamt	71	3.5			
Gesamt		2044	100.0			

n $\qquad h_i \qquad f_i \qquad H_i \qquad F_i$

Mit P_i:

50% der Daten überschritten

$$\tilde{x} = Moderate$$

Mit F_i:

n ist ungerade:

$$\tilde{x} = X_{(n+1)/2}$$

$$\tilde{x} = X_{(1973+1)/2} = X_{987} = ??$$

Stelle 987 wird überschritten

$$\tilde{x} = X_{987} = Moderate$$

Lagemaße: Mittelwert

- Mittelwert der Stichprobe mit Häufigkeitstabellen: Man hat eine Stichprobe mit Umfang „n“ und „k“ verschiedenen Ausprägungen: i=1,2,….,k.

Mittelwert mit absoluter Häufigkeit (f_i): $\bar{x} = \frac{\sum_{i=1}^{k} f_i * i}{n}$

Mittelwert mit relativer Häufigkeit (p_i): $\bar{x} = \sum_{i=1}^{k} p_i * i$

2	4	3	2	6	2	5	1	1	3	5	5	2	1	5	2	4	3	2	2
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Note für die Schule				
i	f_i	p_i	F_i	P_i
1	3	0,15	3	0,15
2	7	0,35	10	0,5
3	3	0,15	13	0,65
4	2	0,1	15	0,75
5	4	0,2	19	0,95
6	1	0,05	20	1
Gesamt	20	1		

Mit f_i:

$$\bar{x}=\frac{\sum_{i=1}^{6} f_i * i}{20}=\frac{3*1+7*2+3*3+2*4+4*5+1*6}{20} = \frac{60}{20} = 3$$

Mit p_i:

$$\bar{x}=\sum_{i=1}^{6} p_i * i=0.15*1+0.35*2+0.15*3+0.1*4+0.2*5+0.05*6$$

$$\bar{x}= 3$$

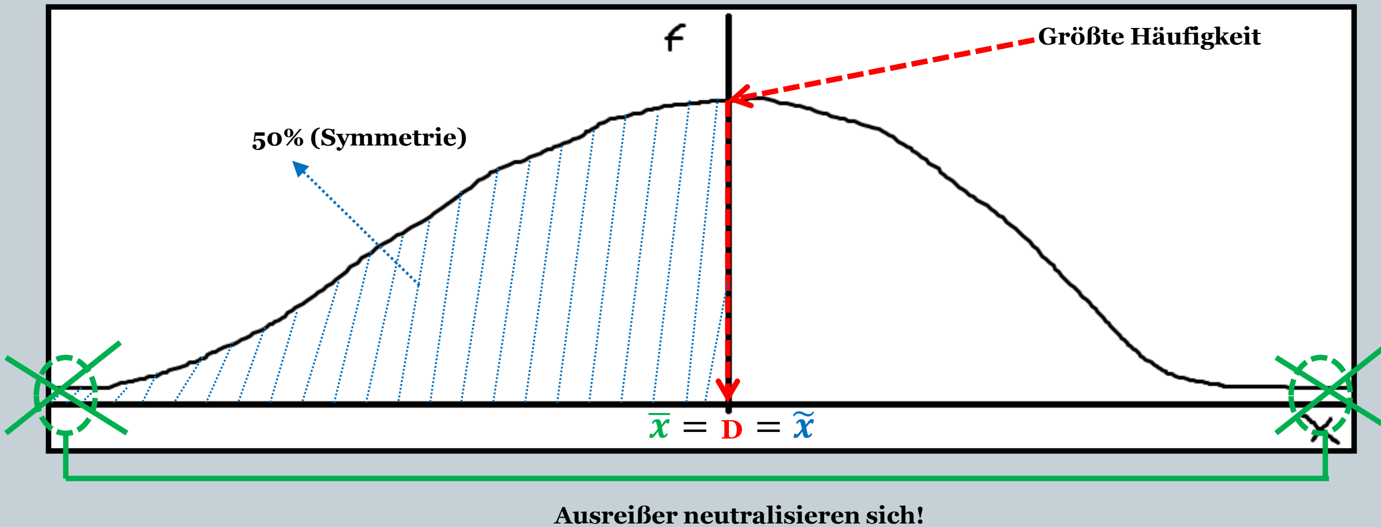
Lagemaße

- Zusammenfassung: Lagemaße für die verschiedenen Skalenniveaus

	D	$\tilde{x}$	$\bar{x}$
Nominal	👍	👎	👎
Ordinal	👍	👍	👎
Quantitativ	👍	👍	👍

Lagemaße: Verhältnis zueinander

- Symmetrische Verteilung

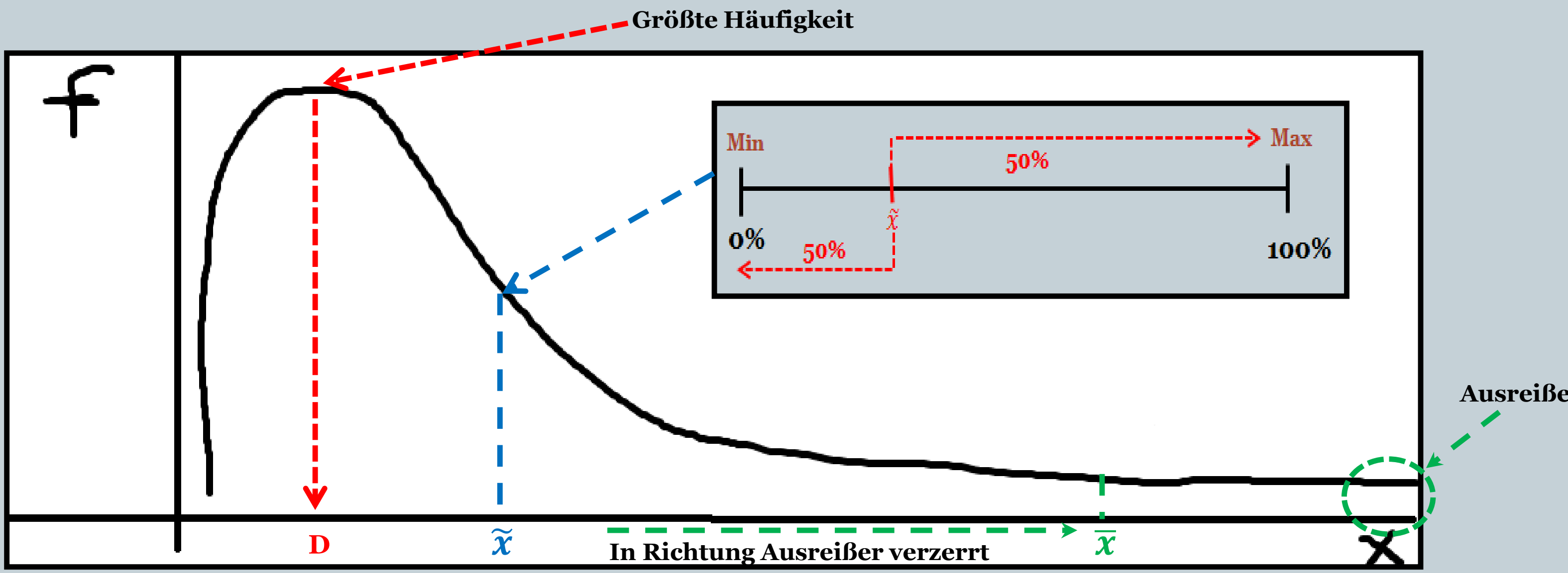


Eine Häufigkeitsverteilung ist symmetrisch, wenn:

$$\mathbf{D} = \tilde{x} = \bar{x}$$

Lagemaße: Verhältnis zueinander

- Linkssteile Verteilung



Eine Häufigkeitsverteilung „tendiert" dazu, linkssteil zu sein, wenn

$$D < \tilde{x} < \bar{x}$$

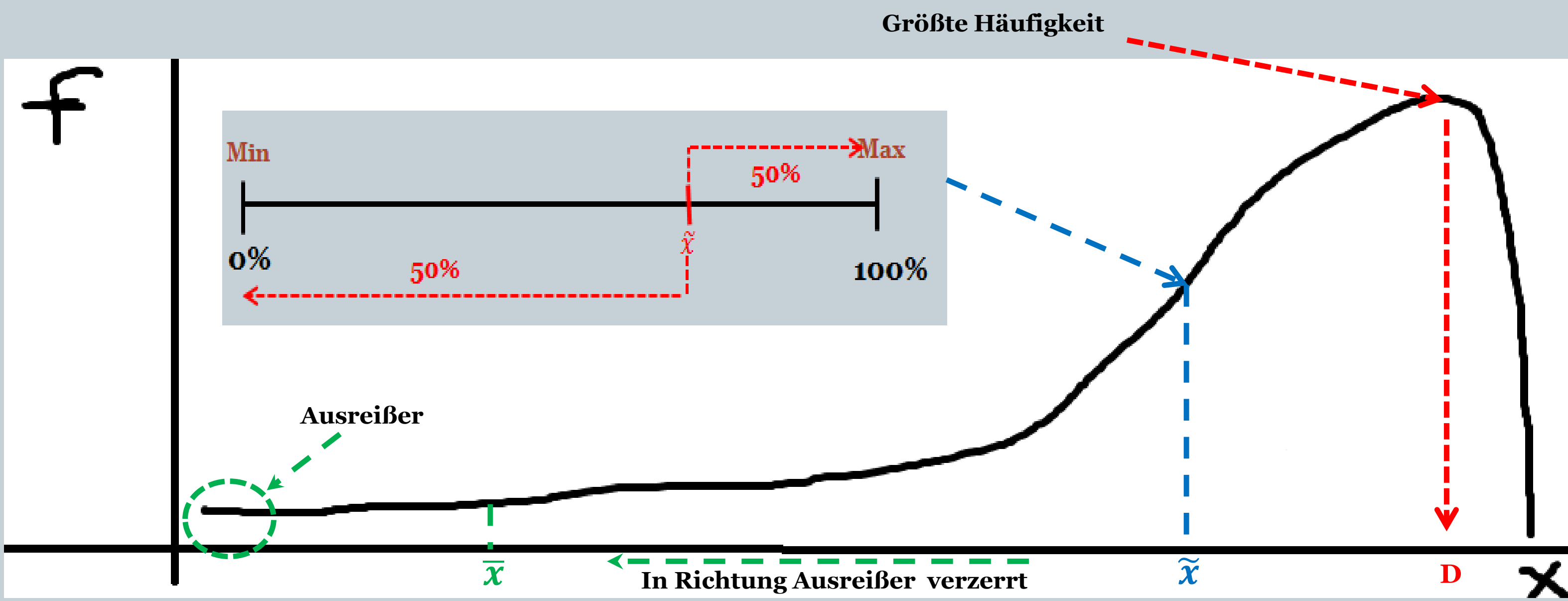
Aber auch, wenn

$$D = \tilde{x} < \bar{x}$$

$$D < \tilde{x} = \bar{x}$$

Lagemaße: Verhältnis zueinander

- Rechtssteile Verteilung



Eine Häufigkeitsverteilung „tendiert“ dazu, rechtssteil zu sein, wenn

$$\bar{x} < \tilde{x} < D$$

Aber auch, wenn

$$\bar{x} = \tilde{x} < D$$

$$\bar{x} < \tilde{x} = D$$

Lagemaße – Aufgabe 1

Aufgabe 1: Die Deutsch-Noten von 20 Schülern der Klasse 4.1 und Klasse 4.2:

Klasse 4.1	1	1	3	5	1	3	5	6	1	3	4	1	2	4	6	2	4	2	6	4
Klasse 4.2	1	5	4	1	6	4	5	5	2	3	5	6	5	6	4	2	3	3	4	3

- Berechnen Sie jeweils Mittelwert und Median
- Formulieren Sie jeweils einen Ergebnissatz ohne statistische Begriffe
- Vergleichen Sie die Klassen

Lösung: Als erstes werden die Häufigkeitstabellen für beide Klassen erstellt. So ordnen wir die Stichproben effizient ein:

Klasse 4.1	fi	Fi	Pi
1	5	5	25%
2	3	8	40%
3	3	11	55%
4	4	15	75%
5	2	17	85%
6	3	20	100%
Total	20		

Klasse 4.2	fi	Fi	Pi
1	2	2	10%
2	2	4	20%
3	4	8	40%
4	4	12	60%
5	5	17	85%
6	3	20	100%
Total	20		

Lagemaße – Aufgabe 1

Aufgabe 1: Lagemaße der Klasse 4.1:

$D_{Klasse4.1}$ = 1

Größte Häufigkeit

Häufigste Note

Klasse 4.1	fi	Fi	Pi
1	5	5	25%
2	3	8	40%
3	3	11	55%
4	4	15	75%
5	2	17	85%
6	3	20	100%
Total	20		

Für den Median - (n=20) - Mit F_i:

$$\tilde{x} = \left(x_{n/2} + x_{(n/2)+1}\right)/_2 = \left(x_{10} + x_{11}\right)/_2$$

Von Position 9 bis Position 11 gibt es eine 3. Also x_{10}=3 und x_{11}=3

$$\tilde{x}_{Klasse4.1} = \left(x_{10} + x_{11}\right)/_2 = \frac{(3+3)}{2} = 3$$

Für den Mittelwert - Mit f_i:

$$\bar{x}_{Klasse4.1} = \frac{\sum_{i=1}^{6} f_i * i}{20} = \frac{5*1+3*2+3*3+4*4+2*5+3*6}{20} = \frac{64}{20} = 3{,}2$$

Also:

- **Modus:** Die häufigste Note in Klasse 4.1 ist 1.
- **Median:** 50% der Schüler in Klasse 4.1 haben die Note 3 oder besser/schlechter.
- **Mittelwert:** Die durchschnittliche Note der Klasse 4.1 ist 3,2.

Lagemaße – Aufgabe 1

Aufgabe 1: Lagemaße der Klasse 4.2:

Klasse 4.2	fi	Fi	Pi
1	2	2	10%
2	2	4	20%
3	4	8	40%
4	4	12	60%
5	5	17	85%
6	3	20	100%
Total	20		

Für den Median - (n=20) - Mit P_i:

50% wird hier überschritten

$$\tilde{x}_{Klasse4.2} = 4$$

Für den Mittelwert - Mit f_i:

$$\bar{x}_{Klasse4.2} = \frac{\sum_{i=1}^{6} f_i * i}{20} = \frac{2*1+2*2+4*3+4*4+5*5+3*6}{20} = \frac{77}{20} = 3{,}85$$

Häufigste Note

Größte Häufigkeit

$D_{Klasse4.2} = 5$

Also:

- **Modus:** Die häufigste Note in Klasse 4.2 ist 5.
- **Median:** 50% der Schüler in Klasse 4.2 haben die Note 4 oder besser/schlechter.
- **Mittelwert:** Die durchschnittliche Note der Klasse 4.2 ist 3,85.

Lagemaße – Aufgabe 1

Aufgabe 1: Vergleich der Klassen 4.1 und 4.2:

Lagemaße	Klasse 4.1	Klasse 4.2
D	**1**	**5**
$\tilde{x}$	**3**	**4**
$\bar{x}$	**3,2**	**3,85**

Klasse 4.1 ist entsprechend aller drei Lagemaße besser als Klasse 4.2. Klasse 4.1 hat z.B. eine bessere durchschnittliche Note, und Klasse 4.2 hat am häufigsten Schüler/-innen, die nicht bestanden haben. Der Median ist auch in Klasse 4.1 besser.

Welche Verteilungsformen haben die zwei Klassen?

$D_{Klasse\ 4.1} < \tilde{x}_{Klasse4.1} < \bar{x}_{Klasse4.1}$ → Die Verteilungsform der Noten der Klasse 4.1 ist laut Folie 9 **Linkssteil**.

$\bar{x}_{Klasse\ 4.2} < \tilde{x}_{Klasse4.2} < D_{Klasse4.2}$ → Die Verteilungsform der Noten der Klasse 4.2 ist laut Folie 10 **Rechtssteil**.

Lagemaße – Aufgabe 2

Lesekompetenz wird als ordinal behandelt. Lagemaße: Modal Klasse und Median Klasse:

Größte Häufigkeit

Überschreitet die Position 16208

Median Klasse

Modal Klasse

Lesekompetenz							
Untergrenze	Obergrenze	fi	Fi	Pi	Mi	Mi*fi
250	300	1013	1013	3.13%	275	278575
300	350	1875	2888	8.91%	325	609375
350	400	3150	6038	18.63%	375	1181250
400	450	4659	10697	33.00%	425	1980075
450	500	5976	16673	51.43%	475	2838600
500	550	6189	22862	70.53%	525	3249225
550	600	5159	28021	86.44%	575	2966425
600	650	3008	31029	95.72%	625	1880000
650	700	1387	32416	100.00%	675	936225
Summe		32416			Summe	15919750

Für den Median:

n=32416 (gerade) → Wir suchen in Position n/2= 16208

Lagemaße – Aufgabe 2: Mittelwert

Lesekompetenz wird als Intervall behandelt. Lagemaße: Modus, Median und Mittelwert.

Addiert, um alle Untergrenzen und Obergrenzen zu erhalten

$$M_1 = \frac{(250+300)}{2}$$

Lesekompetenz						
Untergrenze	Obergrenze	fi	Fi	Pi	Mi	Mi*fi
250	300	1013	1013	3.13%	275	278575
300	350	1875	2888	8.91%	325	609375
350	400	3150	6038	18.63%	375	1181250
400	450	4659	10697	33.00%	425	1980075
450	500	5976	16673	51.43%	475	2838600
500	550	6189	22862	70.53%	525	3249225
550	600	5159	28021	86.44%	575	2966425
600	650	3008	31029	95.72%	625	1880000
650	700	1387	32416	100.00%	675	936225
Summe		32416			**Summe**	15919750

$325 * 1875$

$575 * 5159$

n

$$M_5 = \frac{(450+500)}{2}$$

$$\sum f_i * M_i$$

Damit haben wir folgendes:

$$D = 525 \text{ Punkte} \qquad \tilde{x} = 475 \text{ Punkte} \qquad \bar{x} = \frac{\sum f_i * M_i}{n} = \frac{15919750}{32416} = 491.11 \text{ Punkte}$$

Lagemaße – Aufgabe 2: Mittelwert

Die Verhältnis der Lagemaße zueinander ist nur eine empirische Betrachtung – es passt ziemlich oft, aber es könnte sein, dass manchmal keiner der Fälle vorgekommen ist; in diesen Fällen würde es helfen, ein Histogramm der Daten zu zeichnen, um die Verteilungsform zu erkennen. Zum Beispiel in Aufgabe 2:

Welche Verteilungsform hat das Merkmal Lesekompetenzpunkte?

$D = 525$ Punkte

$\tilde{x} = 475$ Punkte

$\bar{x} = 491.11$ Punkte

$$\tilde{x} < \bar{x} < D$$

In diesem Beispiel haben die Lagemaße des Merkmals keine der Verhältnisse zueinander, die die verschiedenen Verteilungsformen erkennen lassen. Das heißt, man kann mit Hilfe der Lagemaße keine eindeutige Verteilungsform erkennen. Man könnte aber die Vermutung haben, dass die Verteilung wahrscheinlich rechtssteil ist, weil D das größte Lagemaße ist. Um sicher zu sein, könnte man das Histogramm der Daten zeichnen oder später (Streuungsmaße Vorlesung) ein Boxplot erstellen.