

UNIVARIANTE VERTEILUNGEN III: MAßE DER ZENTRALEN TENDENZ

Martin-Luther-Universität Halle-Wittenberg

Institut für Soziologie

Übung Einführung in die deskriptive Statistik

Was machen wir heute?

- Wiederholung
- Lagemaße II: Maße der zentralen Tendenz
 - Modus
 - Median
 - arithmetisches Mittel

Aufgabe 1: Fachsemester Soziologie (Quantile)

Ein Statistikdozent interessiert sich dafür, in welchem Fachsemester die Teilnehmer der M2-Veranstaltung am Dienstagabend sind. Unter Nutzung von Stud.IP lässt er sich die entsprechenden Werte ausgeben:

i	1	2	3	4	5	6	7	8	9	10	11
x_i	7	9	3	1	3	5	5	13	7	1	3

- Bestimmen Sie das 25-%-Quantil und interpretieren Sie das Ergebnis!
- Bestimmen Sie das 75-%-Quantil und interpretieren Sie das Ergebnis!

Aufgabe 1: Lösung

i	1	2	3	4	5	6	7	8	9	10	11
x_i	7	9	3	1	3	5	5	13	7	1	3

Urliste

sortieren

(i)	1	2	3	4	5	6	7	8	9	10	11
$x_{(i)}$	1	1	3	3	3	5	5	7	7	9	13

Primärtafel

Aufgabe 1a: Lösung

- gegeben:
 - $n = 11$
 - $\alpha = 0,25$
- Position in Primärtafel bestimmen:
 - $k = n * \alpha = 11 * 0,25$
 - $k = 2,75 \rightarrow$ aufrunden!
 - $k = 3$
- Wert ablesen:

(i)	1	2	3	4	5	6	7	8	9	10	11
$x_{(i)}$	1	1	3	3	3	5	5	7	7	9	13

- $Q_\alpha = Q_{0,25} = x_{(3)} = 3$
- Mindestens 25 Prozent der Teilnehmer des Seminars befinden sich höchstens im dritten Fachsemester.

Aufgabe 1b: Lösung

- gegeben:
 - $n = 11$
 - $\alpha = 0,75$
- Position in Primärtafel bestimmen:
 - $k = n * \alpha = 11 * 0,75$
 - $k = 8,25 \rightarrow$ aufrunden
 - $k = 9$
- Wert ablesen:

(i)	1	2	3	4	5	6	7	8	9	10	11
$x_{(i)}$	1	1	3	3	3	5	5	7	7	9	13

- $Q_\alpha = Q_{0,75} = x_{(9)} = 7$
- Mindestens 75 Prozent der Teilnehmer des Seminars befinden sich höchstens im siebenten Fachsemester.

Maße der zentralen Tendenz

Welcher Wert ist charakteristisch, um die zentrale Tendenz einer Verteilung zu beschreiben?

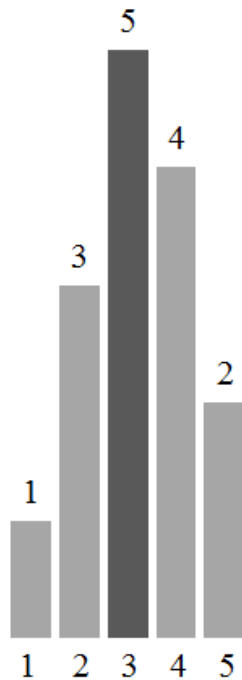
- Modus
- Median
- arithmetisches Mittel

Modus \dot{x} (Modalwert)

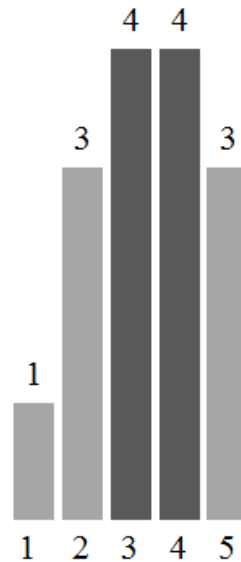
- gültiger Wert, der am häufigsten genannt wurde
- ab nominalen Messniveau
- geringster Informationsgehalt
- nicht immer eindeutig
- $\dot{x} = h = \{x_k | n_k > n_j \text{ für alle } j\}$
- bei gruppierten Daten:
 - Wert der Klassenmitte $m_k = \left(\frac{u_k + o_k}{2}\right)$

Modus \hat{x} (Modalwert) II: Visualisierung

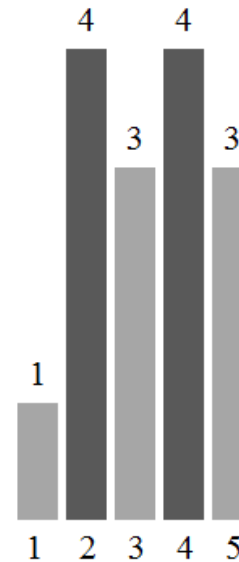
Beispiel a



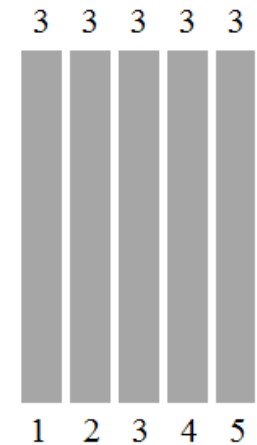
Beispiel b



Beispiel c



Beispiel d



Demokratiezufriedenheit: 1 = sehr unzufrieden, 2 = eher unzufrieden, 3 = teils/teils, 4 = eher zufrieden, 5 = sehr zufrieden

Aufgabe 2: Fachsemester Soziologie (Modus)

Ein Statistikdozent interessiert sich dafür, in welchem Fachsemester die Teilnehmer der M2-Veranstaltung am Dienstagabend sind. Unter Nutzung von Stud.IP lässt er sich die entsprechenden Werte ausgeben:

(i)	1	2	3	4	5	6	7	8	9	10	11
$x_{(i)}$	1	1	3	3	3	5	5	7	7	9	13

- Bestimmen Sie den Modus.
- Interpretieren Sie Ihr Ergebnis!

Aufgabe 2: Lösung

(i)	1	2	3	4	5	6	7	8	9	10	11
$x_{(i)}$	1	1	3	3	3	5	5	7	7	9	13

- **Modus:**
 - häufigster Wert in der Daten
 - $\hat{x} = 3$ (3x)
 - Der Modus liegt bei 3. Dies bedeutet, dass die meisten Befragten mit gültigen Werten im dritten Fachsemester sind.

Aufgabe 3: Interesse an Nachrichten

Eine Konsumforscherin interessiert sich dafür, inwieweit sich die Befragten des ALLBUS 2014 für Nachrichten interessieren. Sie erhält folgende Häufigkeitstabelle:

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UEBERHAUPT NICHT	45	1,3	1,3	1,3
	2 WENIG	95	2,7	2,8	4,2
	3 MITTEL	497	14,3	14,7	18,9
	4 STARK	1312	37,8	38,9	57,8
	5 SEHR STARK	1423	41,0	42,2	100,0
	Gesamt	3372	97,1	100,0	
Fehlend	8 KEINE ANGABE	1	,0		
	9 TRIFFT NICHT ZU	98	2,8		
	Gesamt	99	2,9		
Gesamt		3471	100,0		

- Bestimmen Sie den Modalwert und interpretieren Sie Ihr Ergebnis!
- Bestimmen Sie darüber hinaus das 1. Quartil und das 3. Quartil und interpretieren Sie Ihr Ergebnis.

Aufgabe 3a: Lösung

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UEBERHAUPT NICHT	45	1,3	1,3	1,3
	2 WENIG	95	2,7	2,8	4,2
	3 MITTEL	497	14,3	14,7	18,9
	4 STARK	1312	37,8	38,9	57,8
	5 SEHR STARK	1423	41,0	42,2	100,0
	Gesamt	3372	97,1	100,0	
Fehlend	8 KEINE ANGABE	1	,0		
	9 TRIFFT NICHT ZU	98	2,8		
	Gesamt	99	2,9		
Gesamt		3471	100,0		

Modus:

- $\hat{x} = 5$ (1423x)
- Die meisten Befragten weisen als Realisation den Wert 5 auf.
- Dies bedeutet, dass die meisten Befragten sich sehr stark für Nachrichten interessieren.

Aufgabe 3b: Lösung

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UEBERHAUPT NICHT	45	1,3	1,3	1,3
	2 WENIG	95	2,7	2,8	4,2
	3 MITTEL	497	14,3	14,7	18,9
	4 STARK	1312	37,8	38,9	57,8
	5 SEHR STARK	1423	41,0	42,2	100,0
	Gesamt	3372	97,1	100,0	
Fehlend	8 KEINE ANGABE	1	,0		
	9 TRIFFT NICHT ZU	98	2,8		
	Gesamt	99	2,9		
Gesamt		3471	100,0		

25-%-Quantil:

- $Q_{0,25} = 4$
- Mindestens 25 Prozent der Befragten mit gültigen Werten weisen höchstens einen Wert von 4 auf.
- Das bedeutet das das mindestens 25 Prozent der Befragten mit gültigen Werten höchstens ein starkes politisches Interesse besitzen.

Aufgabe 3b: Lösung II

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UEBERHAUPT NICHT	45	1,3	1,3	1,3
	2 WENIG	95	2,7	2,8	4,2
	3 MITTEL	497	14,3	14,7	18,9
	4 STARK	1312	37,8	38,9	57,8
	5 SEHR STARK	1423	41,0	42,2	100,0
	Gesamt	3372	97,1	100,0	
Fehlend	8 KEINE ANGABE	1	,0		
	9 TRIFFT NICHT ZU	98	2,8		
	Gesamt	99	2,9		
Gesamt		3471	100,0		

75-%-Quantil:

- $Q_{0,75} = 5$
- Mindestens 75 Prozent der Befragten mit gültigen Werten weisen höchstens einen Wert von 5 auf.
- Das bedeutet das das mindestens 75 Prozent der Befragten mit gültigen Werten höchstens ein sehr starkes politisches Interesse besitzen.

Aufgabe 4: Klassierte Fernsehgesamtdauer

Eine Konsumforscherin interessiert sich dafür, wie viele Minuten die ALLBUS-2014-Befragten am Tag mit Fernsehen verbringen, wenn Sie angeben, fernzuschauen. Mithilfe von SPSS erhält folgende Tabelle:

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UNTER 60 MINUTEN	126	3,6	3,7	3,7
	2 60- <180 MINUTEN	2011	57,9	59,7	63,5
	3 180- <360 MINUTEN	1103	31,8	32,8	96,3
	4 360- <540 MINUTEN	105	3,0	3,1	99,4
	5 540 MINUTEN UND MEHR	21	,6	,6	100,0
	Gesamt	3366	97,0	100,0	
Fehlend	0 SIEHT NIE FERN	98	2,8		
	9 KEINE ANGABE	7	,2		
	Gesamt	105	3,0		
Gesamt		3471	100,0		

- Bestimmen Sie die Modalkategorie und interpretieren Sie diese inhaltlich.
- Bestimmen Sie darüber hinaus die Klassenmitte der Modalkategorie als typischen Wert.

Aufgabe 4a: Lösung

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UNTER 60 MINUTEN	126	3,6	3,7	3,7
	2 60- <180 MINUTEN	2011	57,9	59,7	63,5
	3 180- <360 MINUTEN	1103	31,8	32,8	96,3
	4 360-<540 MINUTEN	105	3,0	3,1	99,4
	5 540 MINUTEN UND MEHR	21	,6	,6	100,0
	Gesamt	3366	97,0	100,0	
Fehlend	0 SIEHT NIE FERN	98	2,8		
	9 KEINE ANGABE	7	,2		
	Gesamt	105	3,0		
Gesamt		3471	100,0		

- Modalkategorie:

- $\hat{x} = 2$ (2011x)
- Die meisten Befragten mit gültigen Werten sehen zwischen 60 und 180 Minuten pro Tag fern.

Aufgabe 4b: Lösung

- Mittelwerte der Modalklasse

- $m_k = \frac{u_k + o_k}{2}$

- $m_2 = \frac{60 + 180}{2} = 120$

- Der Mittelwert der Modalkategorie liegt bei 120 Minuten.

Median \tilde{x} (Zentralwert)

- Wert, der in der Mitte der geordneten Verteilung liegt
- ab ordinalem Messniveau
- 50% der Fälle liegen sowohl oberhalb als auch unterhalb dieses Wertes
- robust gegen Ausreißer

Median \tilde{x} : Bestimmung in der Primärtafel

	ordinales Skalenniveau	metrisches Skalenniveau
gerade Fallzahl n	zwischen $x_{(\frac{n}{2})}$ und $x_{(\frac{n}{2}+1)}$ „die beiden einzelnen Werte, die sich in der Mitte befinden (Untermedian und Obermedian)“	$\frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$ „Mittelwert der beiden Werte, die sich in der Mitte befinden“
ungerade Fallzahl n	$x_{(\frac{n+1}{2})}$ „Mittlerer Wert“	$x_{(\frac{n+1}{2})}$ „Mittlerer Wert“

Quiz: Warum ist der Wert gegen Ausreißer und Extremwerte robust?

Fallbeispiel: Yogakurs

An der halleschen Volkshochschule nehmen verschiedene Leute an einem Yogakurs teil. Für das Alter der Personen in Jahren wurde hierbei folgende Werte im Anmeldebogen angegeben:

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$x_{(k)}$	22	23	28	31	35	37	41	42	46	47	51	53	55	57

- gerade Fallzahl (metrische Variable): $\frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$
- $\tilde{x} = \frac{x_{(\frac{14}{2})} + x_{(\frac{14}{2}+1)}}{2} = \frac{x_{(7)} + x_{(8)}}{2} = \frac{41+42}{2}$
- Der Median liegt bei 41,5 Jahren.

Aufgabe 5: Fachsemester Soziologie (Median)

Ein Statistikdozent interessiert sich dafür, in welchem Fachsemester die Teilnehmer der M2-Veranstaltung am Dienstagabend sind. Unter Nutzung von Stud.IP lässt er sich die entsprechenden Werte ausgeben:

i	1	2	3	4	5	6	7	8	9	10	11
x_i	7	9	3	1	3	5	5	13	7	1	3

- Welches Skalenniveau weist die Variable auf?
- Bestimmen Sie den Median.
- Interpretieren Sie Ihr Ergebnis!

Aufgabe 5: Lösung

- Skalenniveau:
 - Ratioskala
- Bestimmung Median:

i	1	2	3	4	5	6	7	8	9	10	11
x_i	7	9	3	1	3	5	5	13	7	1	3

Urliste

(i)	1	2	3	4	5	6	7	8	9	10	11
$x_{(i)}$	1	1	3	3	3	5	5	7	7	9	13

Primärtafel

Aufgabe 5: Lösung II

- Bestimmung Median II:

- $\tilde{x} = x_{\left(\frac{n+1}{2}\right)}$

- $\tilde{x} = x_{\left(\frac{11+1}{2}\right)} = x_{(6)}$

- $\tilde{x} = 5$

- Interpretation:

- Der Median liegt bei 5.

- Der mittlere Befragte befindet sich im fünften Fachsemester.

Median \tilde{x} : unklassierte Häufigkeitstabelle

- einfach bei kumulierten Prozentsätzen ablesen
- Kategorie, bei der 50 % erreicht oder übertroffen werden

Quiz: An welches Konzept, das wir bereits kennengelernt haben, erinnert uns diese Form der Bestimmung?

Aufgabe 6: Interesse an Nachrichten (Median)

Eine Konsumforscherin interessiert sich dafür, inwieweit sich die Befragten des ALLBUS 2014 für Nachrichten interessieren. Sie erhält folgende Häufigkeitstabelle:

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UEBERHAUPT NICHT	45	1,3	1,3	1,3
	2 WENIG	95	2,7	2,8	4,2
	3 MITTEL	497	14,3	14,7	18,9
	4 STARK	1312	37,8	38,9	57,8
	5 SEHR STARK	1423	41,0	42,2	100,0
	Gesamt	3372	97,1	100,0	
Fehlend	8 KEINE ANGABE	1	,0		
	9 TRIFFT NICHT ZU	98	2,8		
	Gesamt	99	2,9		
Gesamt		3471	100,0		

Bestimmen Sie den Median und interpretieren Sie Ihr Ergebnis!

Aufgabe 6: Lösung

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UEBERHAUPT NICHT	45	1,3	1,3	1,3
	2 WENIG	95	2,7	2,8	4,2
	3 MITTEL	497	14,3	14,7	18,9
	4 STARK	1312	37,8	38,9	57,8
	5 SEHR STARK	1423	41,0	42,2	100,0
	Gesamt	3372	97,1	100,0	
Fehlend	8 KEINE ANGABE	1	,0		
	9 TRIFFT NICHT ZU	98	2,8		
	Gesamt	99	2,9		
Gesamt		3471	100,0		

- Median:
 - $\tilde{x} = 4$
 - Der mittlere Befragte befindet sich in der Kategorie 4.
 - Das heißt, dass der mittlere Befragte ein starkes Interesse an Nachrichtensendungen aufweist.

Median \tilde{x} : klassierte Häufigkeitstabellen

- bester Wert zur Schätzung des Medians ist in der klassierten Häufigkeitstabelle der 50-%-Quantil
- Verwendung der Interpolationsformel:
 - $Q_{0,50} = u_k + \frac{0,50 - cp_{k-1}}{p_k} * (o_k - u_k)$

Aufgabe 7: Klassierte Fernsehgesamtdauer

Eine Konsumforscherin interessiert sich dafür, wie viele Minuten die ALLBUS-2014-Befragten am Tag mit Fernsehen verbringen, wenn Sie angeben, fernzuschauen. Mithilfe von SPSS erhält Sie folgende Tabelle:

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 UNTER 60 MINUTEN	126	3,6	3,7	3,7
	2 60- <180 MINUTEN	2011	57,9	59,7	63,5
	3 180- <360 MINUTEN	1103	31,8	32,8	96,3
	4 360- <540 MINUTEN	105	3,0	3,1	99,4
	5 540 MINUTEN UND MEHR	21	,6	,6	100,0
	Gesamt	3366	97,0	100,0	
Fehlend	0 SIEHT NIE FERN	98	2,8		
	9 KEINE ANGABE	7	,2		
	Gesamt	105	3,0		
Gesamt		3471	100,0		

- Bestimmen Sie den interpolierten Wert des Medians.
- Interpretieren Sie Ihr Ergebnis.

Aufgabe 7: Lösung

- Berechnung:

- Klasse, in der erstmals bei den kumulierten Prozenten ein Wert von 50% überschritten wird: $k = 2$

- $\tilde{x} \approx Q_{0,50} = u_k + \frac{0,50 - cp_{k-1}}{p_k} * (o_k - u_k)$

- $\tilde{x} \approx Q_{0,50} = 60 + \frac{0,50 - 0,037}{0,597} * (180 - 60)$

- $\tilde{x} \approx Q_{0,50} = 153,1$

- Interpretation:

- Der interpolierte Wert für den Median liegt bei etwa 153,1 Minuten.
- Dies bedeutet, dass der mittlere Befragte etwa 153,1 Minuten pro Tag fernsieht.

Arithmetisches Mittel \bar{x}

- bei metrischen Skalenniveau
- höchster Informationsgehalt
 - Alle Werte gehen in die Berechnung ein.
 - Auch die Abstände zwischen den Werten werden berücksichtigt.
- wichtige Eigenschaften:
 - Schwerpunkteigenschaft: Die Summe der Abweichung aller Werte vom arithmetischen Mittel beträgt 0 ($\sum_{i=1}^n (x_i - \bar{x}) = 0$).
 - Minimumeigenschaft: Die Summe der quadrierten Abweichungen vom arithmetischen Mittel ist minimal ($\sum_{i=1}^n (x_i - \bar{x})^2 = \min$)
 - anfällig für Ausreißer und Extremwerte
 -

Arithmetisches Mittel \bar{x} : Urliste

- Berechnung:

- $$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} * \sum_{i=1}^n x_i$$

- „Addieren Sie alle gültigen Wert auf und teilen Sie das Ergebnis durch die Zahl der gültigen Fälle“

Quiz: Woher kommt die Anfälligkeit für Ausreißer?

Aufgabe 8: Fachsemester Soziologie (arithmetisches Mittel)

Ein Statistikdozent interessiert sich dafür, in welchem Fachsemester die Teilnehmer der M2-Veranstaltung am Dienstagabend sind. Unter Nutzung von Stud.IP lässt er sich die entsprechenden Werte ausgeben:

i	1	2	3	4	5	6	7	8	9	10	11
x_i	7	9	3	1	3	5	5	13	7	1	3

- Bestimmen Sie das arithmetische Mittel.
- Interpretieren Sie Ihr Ergebnis!

Aufgabe 8: Lösung

i	1	2	3	4	5	6	7	8	9	10	11
x_i	7	9	3	1	3	5	5	13	7	1	3

- **Berechnung:**

- $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

- $\bar{x} = \frac{7+9+3+1+3+5+5+13+7+1+3}{11}$

- $\bar{x} = 5,18$

- **Interpretation:**

- Das arithmetische Mittel liegt bei 5,18.

- Der durchschnittliche Befragte war folglich etwa im 5. Fachsemester.

Arithmetisches Mittel \bar{x} : einfache Häufigkeitstabellen

- Berechnung über die absoluten Häufigkeiten:
 - $\bar{x} = \frac{1}{n} \sum_{j=1}^J x_j * n_j$
 - „Multiplizieren Sie zunächst den Wert der Kategorie mit Fallzahl der Kategorie, bilden Sie die Summe und teilen Sie durch die Fallzahl“
- Berechnung über die relativen Häufigkeiten:
 - $\bar{x} = \sum_{j=1}^J x_j * p_j$
 - „Multiplizieren Sie zunächst den Wert der Kategorie mit dem gültigen Anteil der Kategorie“

Aufgabe 9: Personen im Haushalt

In einer sozialwissenschaftlichen Befragung wurden die Teilnehmer gebeten anzugeben, wie viele Personen in ihrem Haushalt leben. Dabei entstand folgende Häufigkeitstabelle:

ANZAHL DER HAUSHALTSPERSONEN

	Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig 1	695	20,4	20,4	20,4
2	1384	40,7	40,7	61,2
3	623	18,3	18,3	79,5
4	532	15,7	15,7	95,1
5	165	4,9	4,9	100,0
Gesamtsumme	3399	100,0	100,0	

Berechnen Sie das arithmetische Mittel!

Aufgabe 9: Lösung

ANZAHL DER HAUSHALTPERSONEN

	Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig 1	695	20,4	20,4	20,4
2	1384	40,7	40,7	61,2
3	623	18,3	18,3	79,5
4	532	15,7	15,7	95,1
5	165	4,9	4,9	100,0
Gesamtsumme	3399	100,0	100,0	

- **Berechnung:**

- $\bar{x} = \frac{1}{n} \sum_{j=1}^J x_j * n_j$

- $\bar{x} = \frac{1*695+2*1384+3*623+4*532+5*165}{3399}$

- $\bar{x} = 2,44$

- **Antwort:**

- In einem durchschnittlichen Haushalt befanden sich in der Befragung etwa 2,44 Haushaltsmitglieder.

Arithmetisches Mittel \bar{x} : klassierte Daten

- Verwendung des Klassenmittelwertes statt Kategorie
 - $m_k = \frac{u_k + o_k}{2}$
- Berechnung über absolute Häufigkeiten:
 - $\bar{x} = \frac{1}{n} \sum_{k=1}^K n_k * m_k$
 - „Multiplizieren Sie die absoluten Häufigkeiten mit den Klassenmitten und teilen Sie das Ergebnis durch die Fallzahl“
- Berechnung über Anteilswerte:
 - $\bar{x} = \sum_{k=1}^K p_k * m_k$
 - „Multiplizieren Sie die Anteilswerte mit den Klassenmitten“

Beispiel: arithmetisches Mittel für klassierte Daten

Kategorie k	Altersgruppe	Exakte Grenzen $u_k - < o_k$	Klassenmitte m_k	Häufigkeit n_k	Klassenmitte * Häufigkeit $m_k * n_k$
1	18-29 Jahre			583	
2	30-44 Jahre			752	
3	45-59 Jahre			1083	
4	60-74 Jahre			749	
5	75-89 Jahre			296	
6	90-91 Jahre			5	
	Summe			3468	
	arithm. Mittel				

Bitte füllen Sie die Spalten aus! Hinweis $\bar{x} = \frac{1}{n} \sum_{j=1}^J n_k * m_k$ $m_k = \frac{u_k + o_k}{2}$

Beispiel: arithmetisches Mittel für klassierte Daten

Kategorie k	Altersgruppe	Exakte Grenzen $u_k - < o_k$	Klassenmitte m_k	Häufigkeit n_k	Klassenmitte * Häufigkeit $m_k * n_k$
1	18-29 Jahre	17,5-<29,5	23,5	583	13700,5
2	30-44 Jahre	29,5-<44,5	37	752	27824
3	45-59 Jahre	44,5-<59,5	52	1083	56316
4	60-74 Jahre	59,5-<74,5	67	749	50183
5	75-89 Jahre	74,5-<89,5	82	296	24272
6	90 -91 Jahre	89,5-<91,5	90,5	5	452,5
	Summe			3468	172748
	Durchschnitt				49,81

$$\bar{x} = \frac{1}{n} \sum_{k=1}^K n_k * m_k = \frac{172748}{3468} = 49,81$$

Fazit: Übersicht Lagemaße

	Modus	Median	Arithmetisches Mittel
Inhaltliche Aussage	häufigster Wert einer Verteilung	Wert, der bei nach Größe geordneter Verteilung genau in der Mitte liegt	Summe aller Realisationen mit gültigen Werten geteilt durch Anzahl
Skalenniveau	ab nominal (da ausschließlich Häufigkeit betrachtet)	ab ordinal (Ordnungsinformation fließt ein, wie Quantile erst ab diesem Niveau berechenbar)	ab metrisch (da sämtliche Abweichungen vom Mittelwert eingehen)
Robustheit	robust, wenn eindeutig	relativ unempfindlich gegen Ausreißer, da nur mittlerer Wert bzw. mittlerer Werte eingehen	reagiert empfindlich auf Ausreißer
Informationsgehalt	+	++	+++
Sonstiges	kann mehrdeutig sein	kann aus bis zu zwei Werten bestehen	eindeutig

Literaturhinweis

- Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 70-83.

Übungsaufgabe 1: Besucher Einrichtungshaus

Ein Soziologe hat in einem Einrichtungshaus die Besucher nach ihrem Alter befragt und folgende geordnete Urliste erhalten:

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$x_{(k)}$	24	26	32	33	37	37	41	45	48	52	54	58	60	61

- Welches Skalenniveau liegt vor?
- Bestimmen Sie den Median der Verteilung!

Übungsaufgabe 1: Lösung

- metrisches Skalenniveau (Ratioskala)

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$x_{(k)}$	24	26	32	33	37	37	41	45	48	52	54	58	60	61

- $$\tilde{x} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} = \frac{x_{(7)} + x_{(8)}}{2} = \frac{(41+45)}{2} = 43$$
- Der Median der Verteilung liegt bei 43 Jahren. Der mittlere Befragte war 43 Jahre alt.

Übungsaufgabe 2: Anzahl der Kinder

In einer familiensoziologischen Studie wurden zehn Befragte gebeten anzugeben, wie viele Kinder sie haben. Dabei entstand folgende Urliste:

Person	1	2	3	4	5	6	7	8	9	10
Anzahl Kinder	1	0	1	2	2	1	0	3	4	0

Bestimmen Sie das arithmetische Mittel und interpretieren Sie Ihr Ergebnis.

Übungsaufgabe 2: Lösung

k	1	2	3	4	5	6	7	8	9	10
x_k	1	0	1	2	2	1	0	3	4	0

- Berechnung:

- $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

- $\bar{x} = \frac{1+0+1+2+2+1+0+3+4+0}{10}$

- $\bar{x} = 1,4$

- Antwort:

- Das arithmetische Mittel der Verteilung liegt bei 1,4 Kindern.

- Die durchschnittliche Anzahl der Kinder pro Familie liegt bei 1,4 Kindern.

Übungsaufgabe 3: Wartezeit

An einem Bahnhof wurde die Wartezeit auf einen Anschlusszug von 15 Reisenden in Minuten erhoben.



Bild: flickr.com (User: procyonlotormotor)
Lizenz: CC-BY-2.0) Abrufbar unter
<https://flic.kr/p/7ZBnbw>.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x^{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

- Bestimmen Sie das 25%-Quantil und das 75%-Quantil. Interpretieren Sie die Werte.
- Bestimmen Sie Modus, Median und arithmetisches Mittel und interpretieren Sie die Werte.

Übungsaufgabe 3a: Lösung

- 25-%-Quantil:

- $k = n * \alpha = 15 * 0,25 = 3,75 \rightarrow \text{aufrunden} \rightarrow k = 4$
- $x_{(k)} = x_{(4)} = 15$
- Mindestens 25% der Reisenden warten höchstens 15 Minuten.

- 75-%-Quantil:

- $k = n * \alpha = 15 * 0,75 = 11,25 \rightarrow \text{aufrunden} \rightarrow k = 12$
- $x_{(k)} = x_{(12)} = 45$
- Mindestens 75% der Reisenden warten 45 Minuten oder weniger.

Übungsaufgabe 3b: Lösung

- **Modus:**

- $\dot{x} = \{x_k | n_k > n_j \text{ für alle } j\}$
- $\dot{x} = h = 12$ (2x) und 27 (2x)
- Die Verteilung hat zwei Modalwerte. Am häufigsten warteten die Reisenden 12 und 27 Minuten.

- **Median:**

- $\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_{(8)} = 25$
- Der mittlere Reisende wartete 25 Minuten.

Übungsaufgabe 3b: Lösung II

- arithmetisches Mittel:

- $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = 35,07$

- Die Reisenden warteten im Durchschnitt 35,07 Minuten.

Übungsaufgabe 4

An einem Institut für Soziologie wurde die Anzahl der Seiten der eingereichten Bachelorarbeiten erfasst. Für die 20 vorliegenden Arbeiten wurde folgende Tabelle erzielt:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	38	38	39	40	41	41	42	42	42	42	42	42	43	44	45	45	47	48	49	50

- a) Beschreiben Sie die Lage der Verteilung, in dem Sie Modus, Median und arithmetisches Mittel bestimmen.
- b) Welche der folgenden Aussagen ist richtig? Begründen Sie ihre Antwort!
- Der mittlere Umfang der Bachelor-Arbeiten beträgt 42 Seiten.
 - Der Umfang der Bachelor-Arbeiten beträgt im Durchschnitt 43 Seiten.

Übungsaufgabe 4: Lösung I

- Modus:
 - $\dot{x} = 42$ (6x)
- Median:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	38	38	39	40	41	41	42	42	42	42	42	42	43	44	45	45	47	48	49	50

- $\tilde{x} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$
- $\tilde{x} = \frac{x_{(10)} + x_{(11)}}{2}$
- $\tilde{x} = 42$

Übungsaufgabe 4: Lösung II

- arithmetisches Mittel:
 - $\bar{x} = \frac{860}{20} = 43$ Seiten
- beide Aussagen sind korrekt:
 - Der mittlere Seitenumfang liegt 42 Seiten, da der Median bei 42 liegt.
 - Der durchschnittliche Seitenumfang liegt bei 43 Seiten, da das arithmetische Mittel bei 43 liegt.