UNIVARIATE VERTEILUNGEN IV: STREUUNGSMAßE

Martin-Luther-Universität Halle-Wittenberg
Institut für Soziologie
Übung Einführung in die deskriptive Statistik

Agenda

- Wiederholung
- Streuungsmaße
  - Index qualitativer Variation
  - Spannweite und Interquartilsabstand
  - Variation / Varianz / Standardabweichung/Variationskoeffizient

Quiz: Ab welchem Skalenniveau lässt sich das arithmetische Mittel einer Verteilung bestimmen?

A: Nominalskala

B: Ratioskala

C: Ordinalskala

D: Intervallskala

Quiz: Ab welchem Skalenniveau lässt sich der Modus einer Verteilung bestimmen?

A: Intervallskala

B: Ratioskala

C: Ordinalskala

D: Nominalskala

Quiz: Ab welchem Skalenniveau lässt sich der Median bestimmen?

A: Intervallskala

B: Ratioskala

C: Ordinalskala

D: Nominalskala

Aufgabe 1: Haushaltsmitglieder 2014

Im ALLBUS 2014 wurden die Befragten gebeten anzugeben, wie viele Mitglieder in ihrem Haushalt leben. Dabei entstand folgende Tabelle.

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	1	695	20,0	20,2	20,2
	2	1384	39,9	40,2	60,3
	3	623	17,9	18,1	78,4
	4	532	15,3	15,4	93,8
	5	165	4,8	4,8	98,6
	6	31	,9	,9	99,5
	7	14	,4	,4	99,9
	8	2	,1	,1	100,0
	10	1	,0	,0	100,0
	Gesamtsumme	3447	99,3	100,0	
Fehlend	KEINE ANGABE	24	,7		
Gesamtsumme		3471	100,0		

Bestimmen Sie Modus, Median und arithmetisches Mittel und interpretieren Sie Ihr Ergebnis!

Aufgabe 1: Lösung I

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	1	695	20,0	20,2	20,2
	2	1384	39,9	40,2	60,3
	3	623	17,9	18,1	78,4
	4	532	15,3	15,4	93,8
	5	165	4,8	4,8	98,6
	6	31	,9	,9	99,5
	7	14	,4	,4	99,9
	8	2	,1	,1	100,0
	10	1	,0	,0	100,0
	Gesamtsumme	3447	99,3	100,0	
Fehlend	KEINE ANGABE	24	,7		
Gesamtsumme		3471	100,0		

Modus: $\dot{x} = 2$.
In den meisten Haushalten (1384) leben zwei Personen.

Aufgabe 1: Lösung II

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	1	695	20,0	20,2	20,2
	2	1384	39,9	40,2	60,3
	3	623	17,9	18,1	78,4
	4	532	15,3	15,4	93,8
	5	165	4,8	4,8	98,6
	6	31	,9	,9	99,5
	7	14	,4	,4	99,9
	8	2	,1	,1	100,0
	10	1	,0	,0	100,0
	Gesamtsumme	3447	99,3	100,0	
Fehlend	KEINE ANGABE	24	,7		
Gesamtsumme		3471	100,0		

Median: $\tilde{x} = 2$.
Der mittlere Haushalt besteht aus zwei Personen.

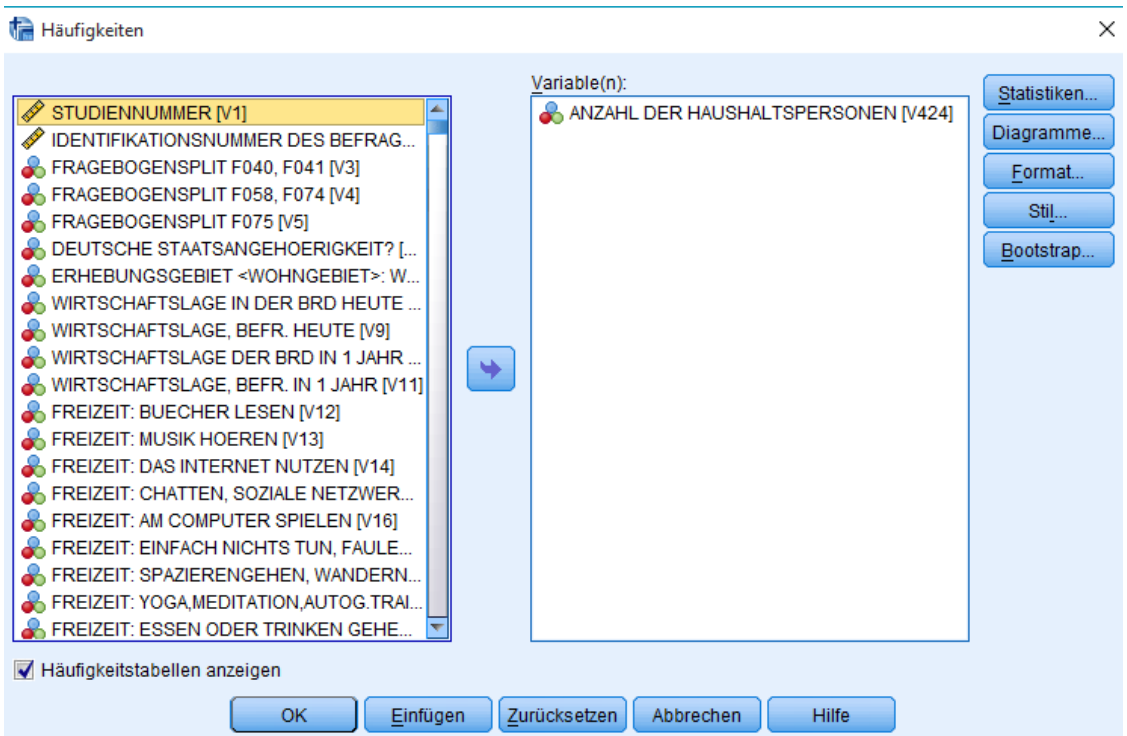
Aufgabe 1: Lösung III

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	1	695	20,0	20,2	20,2
	2	1384	39,9	40,2	60,3
	3	623	17,9	18,1	78,4
	4	532	15,3	15,4	93,8
	5	165	4,8	4,8	98,6
	6	31	,9	,9	99,5
	7	14	,4	,4	99,9
	8	2	,1	,1	100,0
	10	1	,0	,0	100,0
	Gesamtsumme	3447	99,3	100,0	
Fehlend	KEINE ANGABE	24	,7		
Gesamtsumme		3471	100,0		

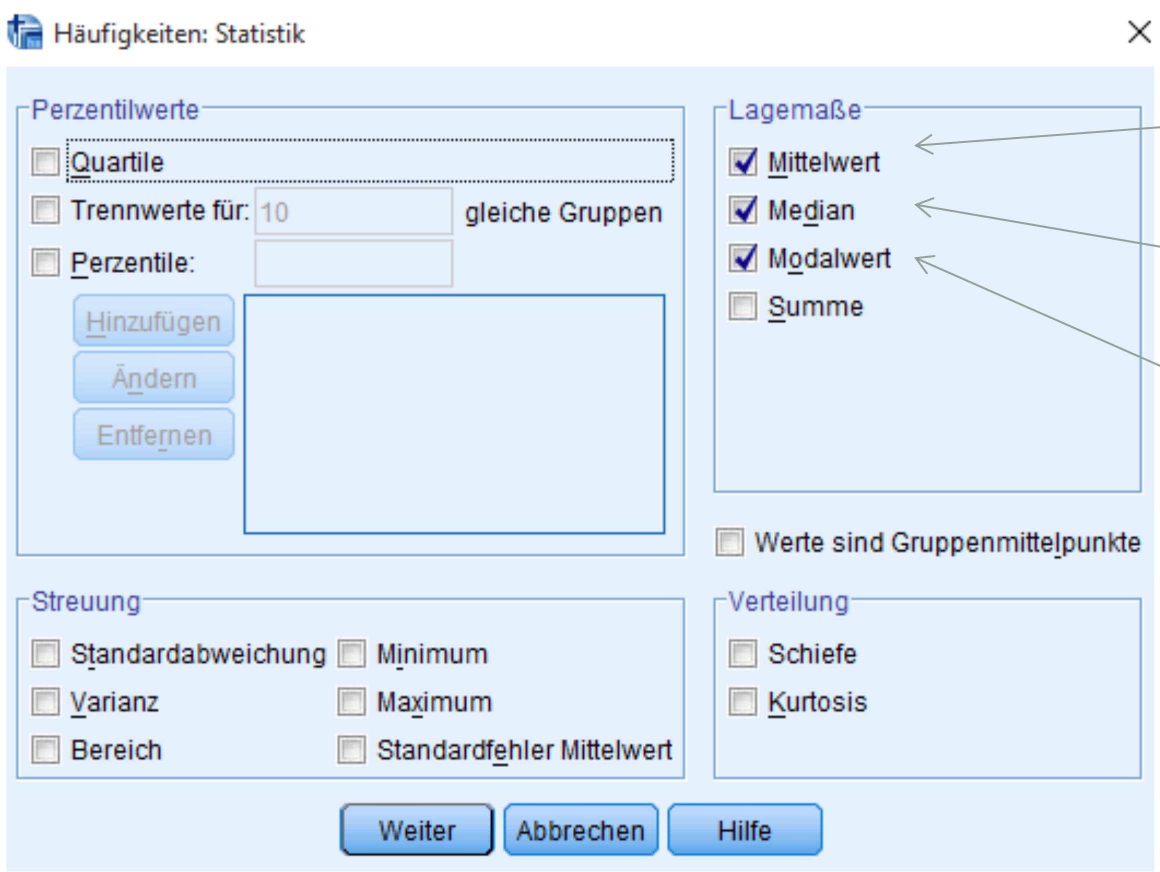
- $\bar{x} = \frac{(1*695)+(2*1384)+(3*623)+(4*532)+(5*165)+(6*31)+(7*14)+(8*2)+(10*1)}{3447}$
- $\bar{x} = 2{,}49$
- In einen durchschnittlichen Haushalt lebten 2,49 Personen.

SPSS: Maße der zentralen Tendenz

- Analysieren → Deskriptive Statistiken → Häufigkeiten…
- Variable auswählen → Statistiken…



SPSS: Maße der zentralen Tendenz II



arith. Mittel $\bar{x}$

Median $\tilde{x}$

Modus $\dot{x}$

SPSS: Maße der zentralen Tendenz III

ANZAHL DER HAUSHALTSPER:

N	Gültig	3447
	Fehlend	24
Mittelwert		2,49
Median		2,00
Modalwert		2

arith. Mittel $\bar{x}$

Median $\tilde{x}$

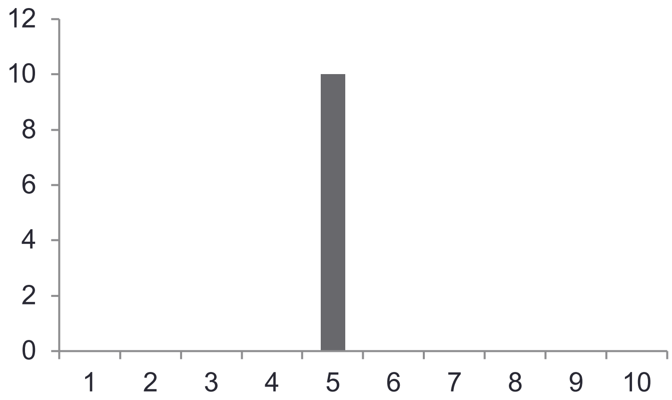
Modus $\dot{x}$

Achtung, bei mehreren Modalwerten wird nur der kleinste Wert ausgegeben.

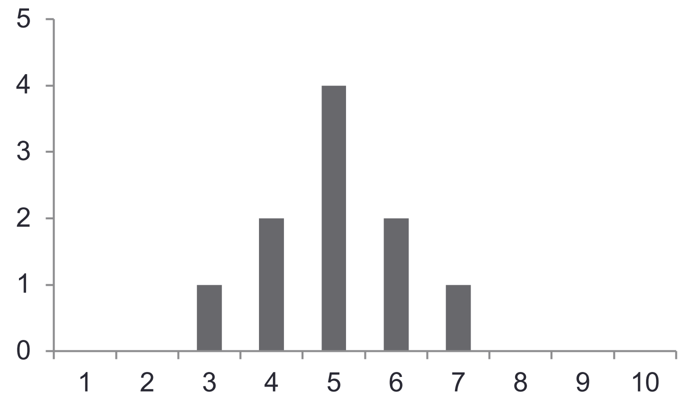
Quiz: Welches Maß der zentralen Tendenz ist im vorliegenden Fall am informativsten? Warum?

Streuungsmaße

- Aussage über Unterschiedlichkeit von Realisierungen
- Verteilungen mit gleicher zentralen Tendenz können eine sehr unterschiedliche Streuung aufweisen



$\dot{x} = 5; \tilde{x} = 5; \bar{x} = 5$



$\dot{x} = 5; \tilde{x} = 5; \bar{x} = 5$

Streuungsmaße II: Übersicht

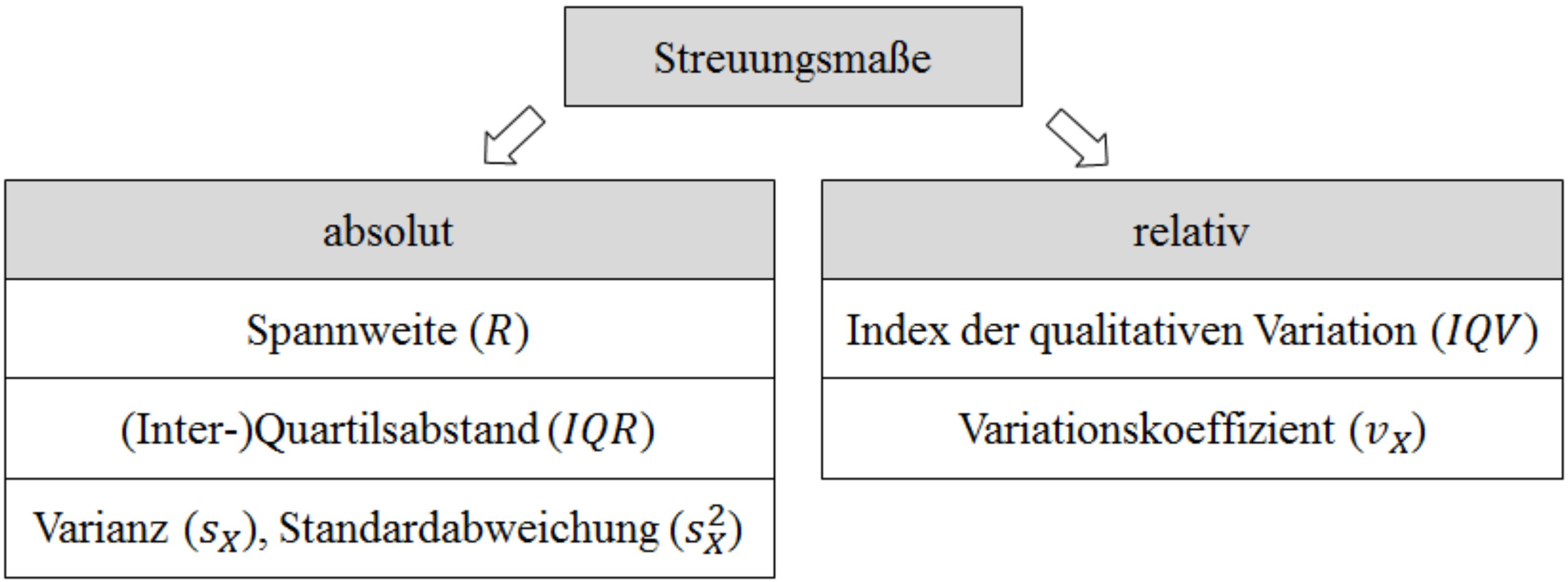


Bild: Kerstin Völkl / Christoph Korb

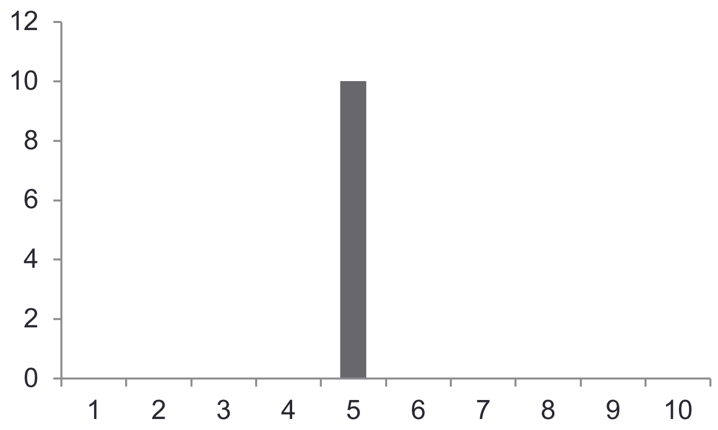
Index qualitativer Variation

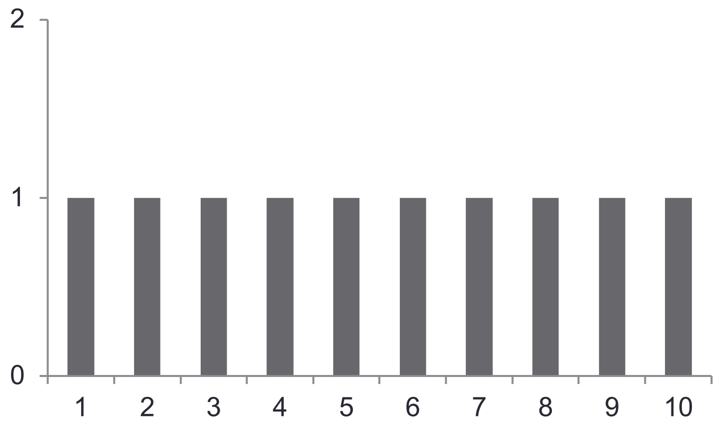
- ab nominalen Skalenniveau berechenbar
- nur die Anzahl gültiger Kategorien und die relative Häufigkeit jeder gültigen Kategorie fließt in die Berechnung des Maßes ein
- $IQV = \frac{J}{J-1} * (1 - \sum_{j=1}^{J} p_j^2)$

J: Anzahl der gültigen Kategorien

p_j^2: quadrierter relativer Anteil gültiger Werte

Index qualitativer Variation II: Bedeutung





0: es liegt keine Variation vor
(alle Fälle in einer Kategorie)

1: maximal mögliche Variation
(Kategorien sind gleich besetzt)

Interpretation: Wird der Wert mit 100 multipliziert, erhält man den prozentualen Anteil der Variation an der maximal möglichen Variation.

Index qualitativer Variation III: Beispiel

Wertelabel	Gültiger Anteil p_j	quadrierter gültiger Anteil
männlich	0,50	0,25
weiblich	0,50	0,25

- $IQV = \frac{J}{J-1} * (1 - \sum_{j=1}^{J} p_j^2)$
- $IQV = \frac{2}{2-1} * (1 - 0{,}5)$
- $IQV = 2 * 0{,}5 = 1$

- maximal mögliche Variation von 100% erreicht → Fälle sind gleich über beide Kategorien verteilt

Wertelabel	Gültiger Anteil p_j	quadrierter gültiger Anteil
männlich	1,00	1
weiblich	0,00	0

- $IQV = \frac{J}{J-1} * (1 - \sum_{j=1}^{J} p_j^2)$
- $IQV = \frac{2}{2-1} * (1 - 1)$
- $IQV = 2 * 0 = 0$

- minimal mögliche Variation von 0% erreicht → alle Fälle befinden sich in der gleichen Kategorie

Aufgabe 2: Parteimitgliedschaft (ISPP 2014)

Eine Soziologin interessiert sich für die Homogenität bzw. Heterogenität der Variable Parteimitgliedschaft. Mithilfe von SPSS erhält sie folgende Häufigkeitstabelle:

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 AKTIVES MITGLIED	47	1,4	2,8	2,8
	2 PASSIVES MITGLIED	40	1,2	2,4	5,2
	3 FRUEHER MITGLIED	122	3,5	7,3	12,4
	4 NIE MITGLIED	1472	42,4	87,6	100,0
	Gesamt	1681	48,4	100,0	
Fehlend	6 KEIN ISSP BUERGER	1753	50,5		
	8 KANN NICHT SAGEN	12	,3		
	9 KEINE ANGABE	25	,7		
	Gesamt	1790	51,6		
Gesamt		3471	100,0		

Bestimmen Sie den Index der qualitativen Variation und interpretieren Sie Ihr Ergebnis inhaltlich.

Aufgabe 2: Lösung I

		absolute Häufigkeit n_j	Gültige Prozente $p_j\%$	gültiger Anteil p_j	quadrierter gültiger Anteil p_j^2
Gültig	1 aktives Mitglied	47	2,8		
	2 passives Mitglied	40	2,4		
	3 früher Mitglied	122	7,3		
	4 nie Mitglied	1472	87,6		
	Gesamt	1681	100,0		
				Summe	

Aufgabe 2: Lösung II

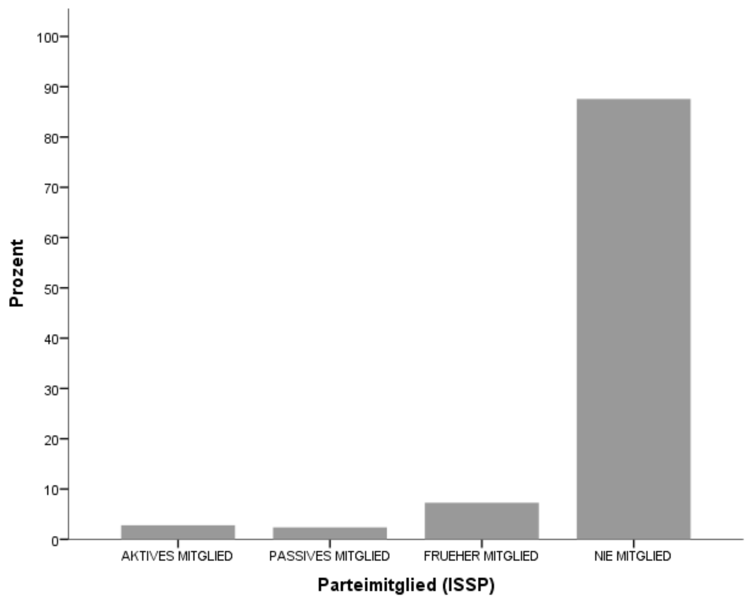
		absolute Häufigkeit n_j	Gültige Prozente $p_j\%$	gültiger Anteil p_j	quadrierter gültiger Anteil p_j^2
Gültig	1 aktives Mitglied	47	2,8	0,028	0,000784
	2 passives Mitglied	40	2,4	0,024	0,000576
	3 früher Mitglied	122	7,3	0,073	0,005329
	4 nie Mitglied	1472	87,6	0,876	0,767376
	Gesamt	1681	100,0		
				Summe	0,774065

Aufgabe 2: Lösung III

- Summe der quadrierten gültigen Anteilswerte:
  - $\sum_{j=1}^{J} p_j^2 = 0{,}774065$
- Bestimmung Anzahl der gültigen Kategorien:
  - $J = 4$
- Berechnung des IQV:
  - $IQV = \frac{J}{J-1} * (1 - \sum_{j=1}^{J} p_j^2)$
  - $IQV = \frac{4}{4-1} * (1 - 0{,}774065)$
  - $IQV = \frac{4}{3} * (1 - 0{,}774065) = 0{,}301$

Aufgabe 2: Lösung IV

- Interpretation:
  - Der Wert für den IQV beträgt im vorliegenden Fall 0,301.
  - Bei der Variable Parteimitgliedschaft liegt die Variation liegt bei 30,1 Prozent der maximal möglichen Variation.



Spannweite (Range)

- bestimmbar ab ordinalem Skalenniveau
- Berechnung:
  - $R = x_{(n)} - x_{(1)}$
  - „Nimm den größten realisierten Wert in den Daten und zieh von diesem den kleinsten realisierten Wert in den Daten ab“

Aufgabe 3: Wartezeit Sushilieferdient

Für einen Sushilieferdient in der Stadt Halle wurden für neun Auslieferungen folgende Wartezeiten in Minuten ermittelt:

k	1	2	3	4	5	6	7	8	9
$x_{(k)}$	10	12	12	15	16	17	20	70	100

Bild: „Nigiri Special“ von photoskate. Lizenz: CC-BY-2.0.

Bestimmen Sie die Spannweite!

Aufgabe 3: Lösung

k	1	2	3	4	5	6	7	8	9
$x_{(k)}$	10	12	12	15	16	17	20	70	100

- $R = x_{(n)} - x_{(1)}$
- $R = 100 - 10$
- $R = 90$ Minuten
- Die Spannweite für die Lieferzeit liegt bei 90 Minuten.

Quiz: Wie ist der Informationsgehalt des Maßes einzuschätzen?
Welche Probleme könnten sich ergeben?

Spannweite: Eigenschaften

- Eigenschaften:
  - berücksichtigt jeweils nur kleinsten und größten Wert in den Daten
  - geringer Informationsgehalt
  - ist anfällig für Ausreißer und Extremwerte

Welche Möglichkeit gibt es, um die Anfälligkeit für Ausreißer zu umgehen?

Interquartilsabstand

- Berechnung:
  - $IQR = Q_{0,75} - Q_{0,25}$
- Interpretation:
  - gibt die Länge des Bereichs der mittleren 50% der Fälle einer Verteilung an
- Eigenschaften:
  - stabiler als Spannweite, da nicht abhängig von Extremwerten, sondern nur von 25-%- und 75-%-Quantil
  - immer noch relativ geringer Informationsgehalt
  - basiert noch immer auf nur zwei Werten

Aufgabe 4: Wartezeit Sushilieferdienst

Für einen Sushilieferdienst in der Stadt Halle wurden für neun Auslieferungen folgende Wartezeiten in Minuten ermittelt:

k	1	2	3	4	5	6	7	8	9
$x_{(k)}$	10	12	12	15	16	17	20	70	100

Bild: „Nigiri Special“ von photoskate. Lizenz: CC-BY-2.0.

Bestimmen Sie den Interquartilsabstand!

Aufgabe 4: Lösung

- gegeben:

k	1	2	3	4	5	6	7	8	9
$x_{(k)}$	10	12	12	15	16	17	20	70	100

- gesucht:
  - $IQR = Q_{0,75} - Q_{0,25}$
- Berechnung:
  - $k = n * \alpha$
  - $Q_{0,75}\colon \mathrm{k} = 9 * 0{,}75 = 6{,}75 \rightarrow 7 \;\rightarrow \mathrm{x}_{(7)} = 20$
  - $Q_{0,25}\colon \mathrm{k} = 9 * 0{,}25 = 2{,}25 \rightarrow 3 \;\rightarrow \mathrm{x}_{(3)} = 12$

Aufgabe 4: Lösung II

- Berechnung:
  - $IQR = Q_{0,75} - Q_{0,25}$
  - $IQR = 20 - 12 = 8$
- Antwortsatz:
  - Der Interquartilsabstand liegt bei 8.
  - Das bedeutet das die Bandbreite der mittleren 50% der Fälle bei 8 Minuten liegt, hier also zwischen 12 und 20 Minuten.

Wie lassen sich bei metrischen Daten auch die Abstände zwischen den Werten bei der Bildung eines Streuungsmaßes berücksichtigen?

Metrische Streuungsmaße: Vorüberlegungen I

- grundlegende Idee:
  - bei metrischen Variablen lässt sich das arithmetische Mittel als Maß der zentralen Tendenz nutzen
  - Wie stark weichen die Werte von diesem?

Metrische Streuungsmaße: Vorüberlegungen II

Für 5 Befragte wurde bei einer Kundenzufriedenheits-befragung erhoben, wie lange sie an der Kasse eines Supermarktes gewartet haben. Es ergaben sich folgende Werte:

Kunde	1	2	3	4	5
Wartezeit in Minuten	0	2	4	4	10

- $\bar{x} = \frac{(0+2+4+4+10)}{5}$
- $\bar{x} = \frac{20}{5} = 4$

Metrische Streuungsmaße: Vorüberlegungen III

Warum können wir hier als Streuungsmaß nicht einfach die summierte Abweichung vom Mittelwert verwenden?

Kunde	x_i	$x_i - \bar{x}$
1	0	-4
2	2	-2
3	4	0
4	4	0
5	10	6
$\bar{x} =$	4	

$$\sum(x_i - \bar{x}) = -4 + (-2) + 0 + 0 + 6 = 0$$

→ Schwerpunkteigenschaft des arithmetischen Mittels

Was tun?

Metrische Streuungsmaße: Vorüberlegungen IV

- Lösungsmöglichkeit:
  - Betrag nutzen
    - Average (Absolute) Deviation (AD)
  - Abweichungsquadrate nutzen
    - Variation (SAQ_X)
    - Varianz (s_X^2)
    - Standardabweichung(s_X)

Exkurs: Average (Absolute) Deviation

- Berechnung:
  - $AD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$
  - „Ziehe von jedem realisierten Wert, den Mittelwert aller Werte ab, wenn der Wert negativ ist, streiche das Vorzeichen. Summiere alle Ergebnisse auf und teile Sie durch die gültigen Fälle.“
- Interpretation:
  - Durchschnitt der absoluten Abweichungen aller Messwerte einer Verteilung von ihrem arithmetischen Mittel

Exkurs: Average (Absolute) Deviation II

Kunde	x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
1	0	-4	4
2	2	-2	2
3	4	0	0
4	4	0	0
5	10	6	6
$\bar{x} =$	4		12

- Berechnung:
  - $AD = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \bar{x} \rvert$
  - $AD = \frac{12}{5} = 2{,}5$
- Interpretation:
  - Die durchschnittliche absolute Abweichung vom Mittelwert der Wartezeit der fünf Kunden liegt bei 2,5 Minuten.

Variation

- Berechnung:
  - $SS_X = SAQ_X = \sum_{i=1}^{n}(x_i - \bar{x})^2$

Sum of Squares

Wert der einzelnen Realisation

arithmetisches Mittel von x

- Interpretation:
  - Die Summe der quadrierten Abweichungen aller Realisationen vom arithmetischen Mittel liegt bei …

Variation: Beispiel

Kunde	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	4	-0	0
4	4	-0	0
5	10	6	36
$\bar{\boldsymbol{x}} =$	4		

$\sum(x_i - \bar{x})^2 = 56$

- $SAQ_X = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 56$
- Die Summe der quadrierten Ab-weichungen der Realisationen vom arithmetischen Mittel liegt bei 56 Minuten².

Was würde passieren, wenn wir die gleichen Fälle verdoppeln würden?

Variation: Beispiel II

Kunde	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	4	-0	0
4	4	-0	0
5	10	6	36
$\bar{x} =$	4		

Kunde	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	4	-0	0
4	4	-0	0
5	10	6	36
6	0	-4	16
7	2	-2	4
8	4	-0	0
9	4	-0	0
10	10	6	36
$\bar{x} =$	4		

$$SAQ_X = \sum_{i=1}^{n} (x_i - \bar{x})^2 = 56$$

$$SAQ_X = \sum_{i=1}^{n} (x_i - \bar{x})^2 = 112$$

Suboptimal☹

Variation: Eigenschaften

- Probleme:
  - leider abhängig von Fallzahl
  - Einheit der Ausgangsvariablen ist quadriert
  - kein normierter Wertebereich für Vergleiche

Wie können wir das
Problem mit der Fallzahl
beheben?

Varianz

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

Fallzahl

Variation von X

- Interpretation:
  - Die durchschnittliche quadrierte Abweichung der Realisationen vom arithmetischen Mittel liegt bei …

Varianz: Beispiel

Kunde	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	4	-0	0
4	4	-0	0
5	10	6	36
$\bar{x} =$	4		

Kunde	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	4	-0	0
4	4	-0	0
5	10	6	36
6	0	-4	16
7	2	-2	4
8	4	-0	0
9	4	-0	0
10	10	6	36
$\bar{x} =$	4		

$$SAQ_X = \sum_{i=1}^{n} (x_i - \bar{x})^2 = 56$$

$$s_X^2 = \frac{SAQ_X}{n} = \frac{56}{5} = 11{,}2$$

$$SAQ_X = \sum_{i=1}^{n} (x_i - \bar{x})^2 = 112$$

$$s_X^2 = \frac{SAQ_X}{n} = \frac{112}{10} = 11{,}2$$

Varianz: Eigenschaften

- wünschenswert:
  - unabhängig von Fallzahl
- Probleme:
  - Einheit ist immer noch quadriert
  - kein normierter Wertebereich für Vergleiche

Standardabweichung

- Berechnung:
  - $s_X = \sqrt{s_X^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{SAQ_X}{n}}$
  - „Ziehe die Wurzel aus der Varianz“
- Interpretation:
  - formal korrekt: Wurzel aus der durchschnittlichen quadrierten Abweichungen vom Mittelwert
  - Umgangssprachlich: „durchschnittliche“ Abweichung vom Mittelwert

Standardabweichung: Beispiel

Kunde	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	4	-0	0
4	4	-0	0
5	10	6	36
$\boldsymbol{\bar{x}} =$	4		

- $s_X^2 = \frac{SAQ_X}{n} = \frac{56}{5} = 11{,}2\ Minuten^2$
- $s_X = \sqrt{(s_X^2)} = \sqrt{11{,}2} = 3{,}34\ Minuten$
- formal: Die Wurzel der durchschnittlichen quadrierten Abweichungen der Werte vom arithmetischen Mittel liegt bei 3,34 Minuten.
- umgangssprachlich: Die „durchschnittliche“Abweichung vom arithmetischen Mittel liegt bei etwa 3,34 Minuten.

Standardabweichung: Eigenschaften

- wünschenswerte Eigenschaften:
  - alle Werte werden in der Berechnung berücksichtigt
  - unabhängig von Fallzahl
  - Streuungsmaß hat gleiche Einheit wie Ausgangsvariable
- Problem:
  - schwerfällige Interpretation
  - noch kein normierter Wertebereich für Vergleiche

Wie lässt sich hier der Effekt der Maßeinheit beheben?

Variationskoeffizient

- bei ratioskalierten Variablen
- Berechnung:
  - $V_X = \frac{s_X}{\bar{x}}$
  - „Quotient der Standardabweichung geteilt durch das arithmetische Mittel“
- Interpretation:
  - Grad der Variation in Einheiten des arithmetischen Mittels
- Eigenschaften:
  - einheitenlose Maßzahl
  - geeignet zum Vergleich von Streuungen von Verteilungen
  - nur sinnvoll interpretierbar bei Variablen mit nichtnegativen Wertebereich

Variationskoeffizient: Beispiel

Kunde	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	4	-0	0
4	4	-0	0
5	10	6	36
$\bar{\boldsymbol{x}} =$	4		

- Berechnung:
  - $s_X^2 = \frac{SAQ_X}{n} = \frac{56}{5} = 11{,}2\ Minuten^2$
  - $s_X = \sqrt{(s_X^2)} = \sqrt{11{,}2} = 3{,}35\ Minuten$
  - $V_X = \frac{s_X}{\bar{x}} = \frac{3{,}34\ Minuten}{4\ Minuten} = 0{,}837$
- Interpretation:
  - Die Standardabweichung der Wartezeit liegt bei 83,7% der durchschnittlichen Wartezeit liegt.

Aufgabe 5: Judith wartet…

Leider muss Judith oft auf ihren Freund Clemens warten. Da sie dabei oft gelangweilt ist, berechnet sie aus lauter Frust Streuungsmaße. Für die letzten 8 Verabredungen mit ihrem Freund hat sie folgende Wartezeiten in Minuten notiert: 5, 5, 40, 10, 30, 3, 7, 60

a) Bestimmen Sie Variation, Varianz und Standardabweichung und Interpretieren Sie die Werte.
b) Berechnen Sie zusätzlich den Variationskoeffizienten.

Aufgabe 5a: Lösung

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
5	-15	225
5	-15	225
40	20	400
10	-10	100
30	10	100
3	-17	289
7	-13	169
60	40	1600
$\bar{x} = 20$		$SAQ_X = 3108$
		$s_X^2 = 388{,}5$
		$s_X = 19{,}71$

Variation

Varianz

Standardabweichung

Aufgabe 5a: Lösung II

- Interpretation Variation:
  - $SAQ_X = 3108$
  - Die Summe der quadrierten Abweichungen vom arithmetischen Mittel liegt bei 3108 Minuten².
- Interpretation Varianz:
  - Die durchschnittliche quadrierte Abweichung aller Werte vom arithmetischen Mittel liegt bei 388,5 Minuten²
- Interpretation Standardabweichung:
  - Die Wurzel der durchschnittlichen quadrierte Abweichung aller Werte vom arithmetischen Mittel liegt bei 19,71 Minuten.

Aufgabe 5b: Lösung

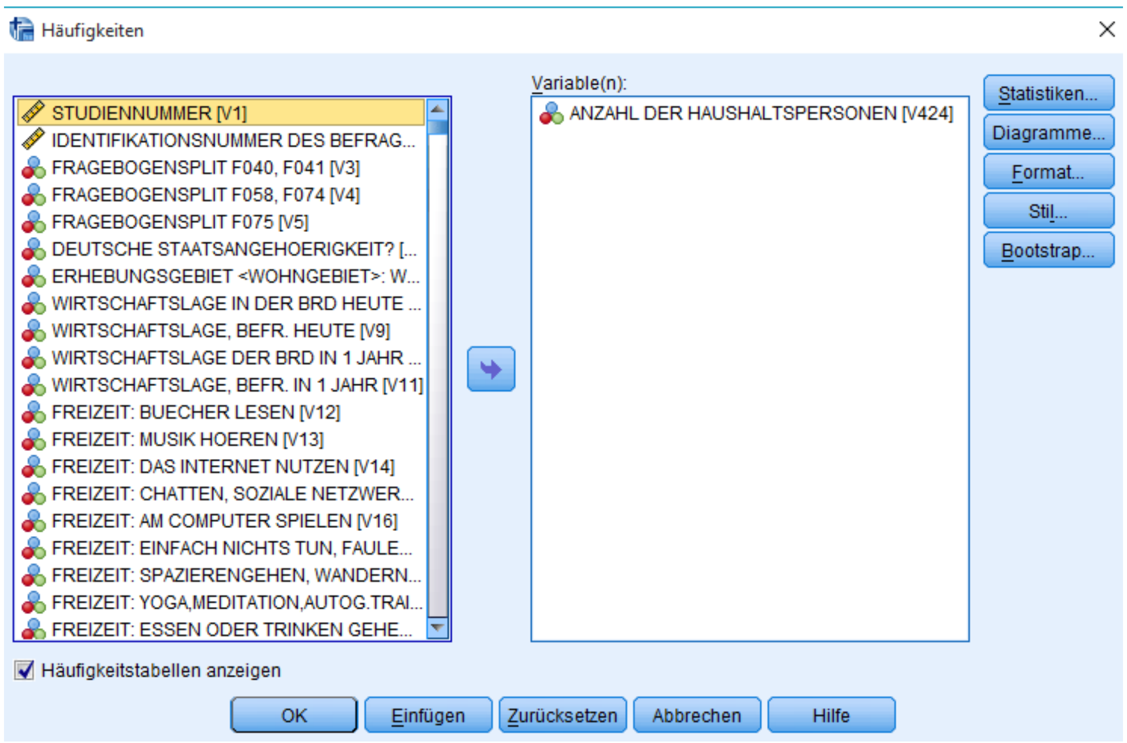
- Berechnung:
  - $V_X = \frac{s_X}{\bar{x}}$
  - $V_X = \frac{19{,}71}{20}$
  - $V_X = 0{,}986$

TIPP: Berechnung Streuungsmaße

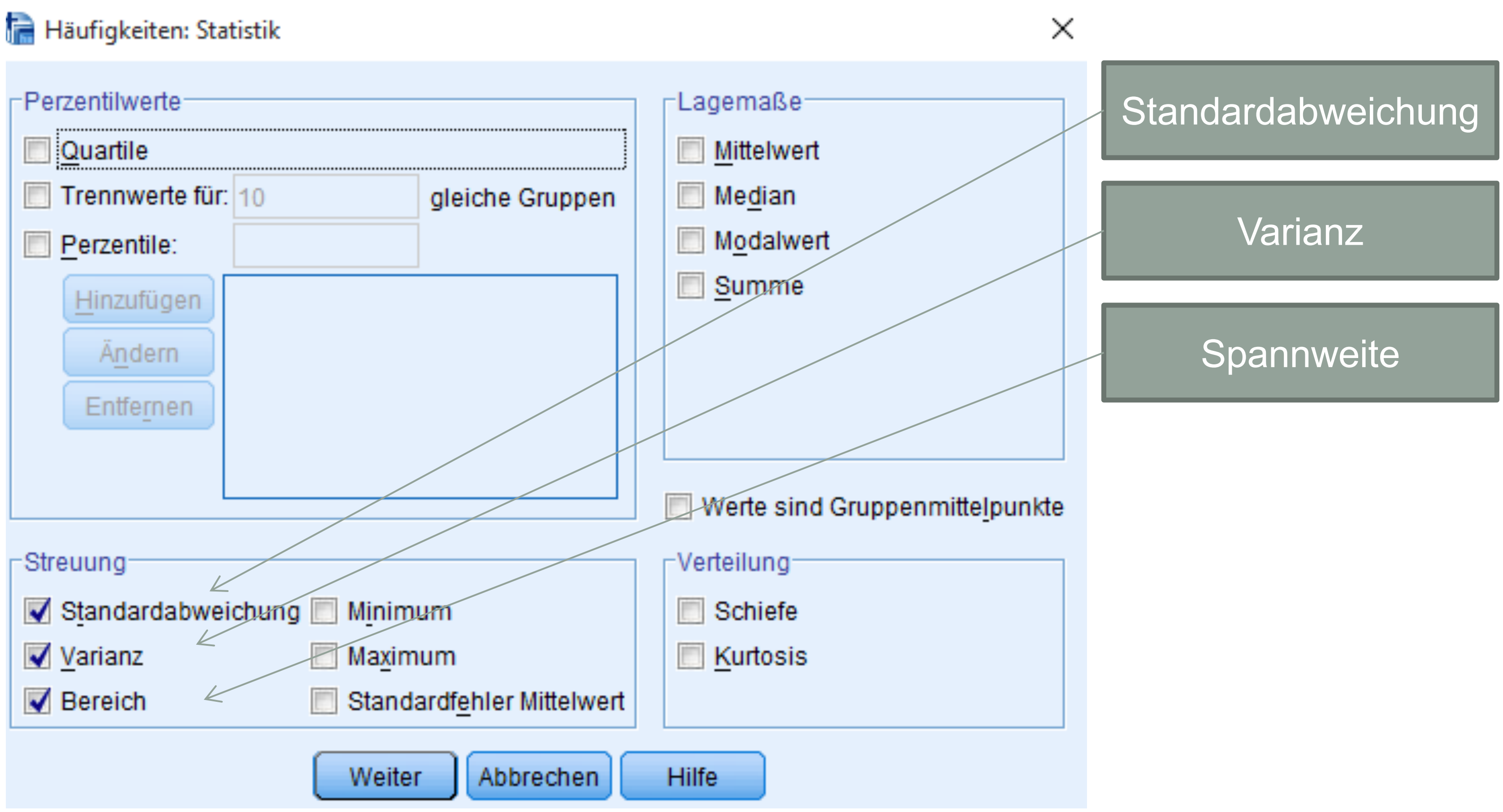
- Bei den meisten Schulstandardtaschenrechnern lässt sich die Berechnung von Mittelwerten und Standardabweichungen vereinfachen, da es einen entsprechenden Statistikmodus (SD) gibt.
- Zunächst muss oft der Statistikmodus SD ausgewählt werden. Anschließend werden die Werte in eine Datenliste eingefügt.
- Tasten wie $n, \bar{x}\ und\ \sigma_{Xn}$ stehen bei einfacheren Taschenrechnern dann für Fallzahl, arithmetisches Mittel und Standardabweichung, bei komplexeren Modellen gibt es meist einen eigenen Modus. (Stat-Taste)

SPSS: einfache Streuungsmaße

- Analysieren → Deskriptive Statistiken → Häufigkeiten…
- Variable auswählen → Statistiken…



SPSS: einfache Streuungsmaße II



SPSS: einfache Streuungsmaße III

N	Gültig	3447
	Fehlend	24
Standardabweichung		1,213
Varianz		1,472
Bereich		9

Standardabweichung

Varianz

Spannweite

Achtung: SPSS arbeitet mit einem anderen Standardabweichungs- bzw. Varianzbegriff als wir in der deskriptiv Statistik haben. Hierbei wird davon ausgegangen, dass wir es nicht mit einer Vollerhebung, sondern einer Stichprobe zu tun haben. Entsprechend wird nicht durch n, sondern n-1 geteilt. Warum das so ist, schauen wir uns in M3 näher an.

Übungsaufgabe 1: Turnschuhe

Ein Marktforschungsunternehmen interessiert sich dafür, wie viel Paar Turnschuhe Personen durchschnittlich haben und befragt hierfür 15 „repräsentativ ausgewählte" Personen.
Es ergab sich folgende Urliste:
10, 5, 2, 4, 20, 40, 3, 8, 2, 11, 12, 1, 3, 2, 10.

Bild: flickr.com (User: Jochen Handschuh Lizenz: CC-BY-2.0) Abrufbar unter https://flic.kr/p/6hxQKU.

a) Bestimmen Sie Modus, Median und arithmetisches Mittel und interpretieren Sie ihr Ergebnis!
b) Bestimmen Sie darüber hinaus Spannweite, Interquartilsabstand, Interpretieren Sie Ihr Ergebnis!
c) Berechnen Sie Variation, Varianz und Standardabweichung.

Übungsaufgabe 1a: Lösung

- Modus:
  - $\dot{x} = 2$
  - Die meisten Befragten besitzen 2 Paar Turnschuhe.
- Median:
  - $\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{15+1}{2}\right)} = x_{(8)} = 5$
  - Der mittlere Befragte besitzt 5 Paar Turnschuhe.
- arithmetisches Mittel:
  - $\bar{x} = \frac{1+2+2+2+3+3+4+5+8+10+10+11+12+20+40}{15} = 8{,}8\overline{6}$
  - Durchschnittlich besitzt ein Befragter in der Stichprobe etwas weniger als 9 Paar Turnschuhe.

Übungsaufgabe 1b: Lösung I

- Spannweite:
  - $R = x_{(n)} - x_{(1)}$
  - $R = x_{(15)} - x_{(1)}$
  - $R = 40 - 1 = 39$
  - Die Spannweite reicht von einem Paar Turnschuhen bis 40 Paar Turnschuhen. Dies ist eine Spannweite von 39 Paaren.

(i)	$x_{(i)}$	$(x_{(i)} - \bar{x})$	$(x_{(i)} - \bar{x})^2$
1	1	-7,87	61,88
2	2	-6,87	47,15
3	2	-6,87	47,15
4	2	-6,87	47,15
5	3	-5,87	34,42
6	3	-5,87	34,42
7	4	-4,87	23,68
8	5	-3,87	14,95
9	8	-0,87	0,75
10	10	1,13	1,28
11	10	1,13	1,28
12	11	2,13	4,55
13	12	3,13	9,82
14	20	11,13	123,95
15	40	31,13	969,28
	$\bar{x} = 8{,}87$		$SAQ_X = 1421{,}73$
			$s_X^2 = 94{,}78$

Übungsaufgabe 1b: Lösung II

- $Q_{0,75}$:
  - $k = 15 * 0{,}75 = 11{,}25$
  - aufrunden: $k = 12$
  - $x_{(k)} = x_{(12)} = 11$
- $Q_{0,25}$:
  - $k = 15 * 0{,}25 = 3{,}75$
  - aufrunden: $k = 4$
  - $x_{(k)} = x_{(4)} = 2$
- Interquartilsabstand:
  - $IQR = Q_{0,75} - Q_{0,25}$
  - $IQR = 11 - 2 = 9$
  - Die mittleren 50% der Befragten haben zwischen 11 und 2 Paar Turnschuhen. Dies entspricht einer Bandbreite von 9 Paar Turnschuhen.

(i)	$x_{(i)}$	$(x_{(i)} - \bar{x})$	$(x_{(i)} - \bar{x})^2$
1	1	-7,87	61,88
2	2	-6,87	47,15
3	2	-6,87	47,15
4	2	-6,87	47,15
5	3	-5,87	34,42
6	3	-5,87	34,42
7	4	-4,87	23,68
8	5	-3,87	14,95
9	8	-0,87	0,75
10	10	1,13	1,28
11	10	1,13	1,28
12	11	2,13	4,55
13	12	3,13	9,82
14	20	11,13	123,95
15	40	31,13	969,28
	$\bar{x} = 8{,}87$		$SAQ_X = 1421{,}73$
			$s_X^2 = 94{,}78$

Übungsaufgabe 1b: Lösung III

- Variation:
  - $\mathrm{SAQ}_X = 1421{,}73$
- Varianz:
  - $s_X^2 = \frac{SS_X}{\mathrm{n}} = \frac{1421{,}73}{20} = 94{,}78$
- Standardabweichung:
  - $s_X = \sqrt{s_X^2} = \sqrt{94{,}78} = 9{,}74$

(i)	$x_{(i)}$	$(x_{(i)} - \bar{x})$	$(x_{(i)} - \bar{x})^2$
1	1	-7,87	61,88
2	2	-6,87	47,15
3	2	-6,87	47,15
4	2	-6,87	47,15
5	3	-5,87	34,42
6	3	-5,87	34,42
7	4	-4,87	23,68
8	5	-3,87	14,95
9	8	-0,87	0,75
10	10	1,13	1,28
11	10	1,13	1,28
12	11	2,13	4,55
13	12	3,13	9,82
14	20	11,13	123,95
15	40	31,13	969,28
	$\bar{x} = 8{,}87$		$SAQ_X = 1421{,}73$
			$s_X^2 = 94{,}78$

Übungsaufgabe 2: Straßenbahn

In einer Zufriedenheitsbefragung der Verkehrsgesellschaft wurde gefragt, wie lange die Kunden durchschnittlich (in Minuten) auf ihre Straßenbahn warten. Dabei wurden folgende Antworten gegeben:

Bild: flickr.com (User: Harald Henkel / Lizenz: CC-BY-ND-2.0)
Abrufbar unter https://flic.kr/p/bQCpXX.

10, 5, 2, 4, 20, 40, 3, 8, 2, 11

Bestimmen Sie Variation, Varianz und Standardabweichung.

Übungsaufgabe 2: Lösung

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
10	-0,5	0,25
5	-5,5	30,25
2	-8,5	72,25
4	-6,5	42,25
20	9,5	90,25
40	29,5	870,25
3	-7,5	56,25
8	-2,5	6,25
2	-8,5	72,25
11	0,5	0,25
$\bar{x} = 10{,}5$		$SAQ_X = 1240{,}5$
		$s_X^2 = 124{,}05$
		$s_X = 11{,}14$

Variation

Varianz

Standardabweichung

Übungsaufgabe 3: Gottesglaube

Im ALLBUS 2012 wurden Befragten gebeten anzugeben, inwieweit Sie an Gott oder ein anderes höheres Wesen glauben. Dabei entstand folgende Häufigkeitstabelle:

GOTTESGLAUBEN

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	PERSOENLICHER GOTT	697	20,0	20,4	20,4
	HOEHERES WESEN	1064	30,6	31,1	51,4
	WN WAS GLAUBEN SOLL	571	16,4	16,7	68,1
	GLAUBE NICHT AN GOTT	1093	31,4	31,9	100,0
	Gesamt	3425	98,4	100,0	
Fehlend	KEINE ANGABE	55	1,6		
Gesamt		3480	100,0		

Berechnen Sie den Index qualitativer Variation und interpretieren Sie ihr Ergebnis statistisch und inhaltlich.

Übungsaufgabe 3: Lösung I

		Häufigkeit n_j	Gültige Prozente $p_j\%$	gültiger Anteil p_j	quadrierter gültiger Anteil p_j^2
Gültig	PERSOENLICHER GOTT	697	20,4	0,204	0,041616
	HOEHERES WESEN	1064	31,1	0,311	0,096721
	WN WAS GLAUBEN SOLL	571	16,7	0,167	0,027889
	GLAUBE NICHT AN GOTT	1093	31,9	0,319	0,101761
	Gesamt	3425	100,0		
Fehlend	KEINE ANGABE	55		Summe	**0,267987**
Gesamt		3480			

Übungsaufgabe 3: Lösung II

- Berechnung:
  - $IQV = \frac{J}{J-1} * (1 - \sum_{j=1}^{J} p_j^2)$
  - $IQV = \frac{4}{4-1} * (1 - 0{,}267987)$
  - $IQV = \frac{4}{3} * (1 - 0{,}267987) = 0{,}976$
- Interpretation:
  - Der Wert für den IQV beträgt im vorliegenden Fall 0,976.
  - Dies bedeutet dass bei der vorliegenden Verteilung die Variation bei 97,6% der maximal möglichen Variation beträgt. Der Gottesglaube weist folglich eine recht hohe Streuung auf.

Übungsaufgabe 4.1: Kickerturnier

An einem Kickerturnier der Soziologiestudierenden nahmen 6 Leute teil. Die Anzahl der Siege für die einzelnen Teilnehmer ist in folgender Tabelle wiedergegeben.

Teilnehmer	1	2	3	4	5	6
Siege	0	2	3	4	7	8

Berechnen Sie Variation, Varianz und Standardabweichung! Interpretieren Sie Ihr Ergebnis statistisch und inhaltlich.

Übungsaufgabe 4.1: Lösungsskizze ☺

Teilnehmer	Siege	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	0	-4	16
2	2	-2	4
3	3	-1	1
4	4	0	0
5	7	3	9
6	8	4	16
	Summe		$SAQ_X = 46$
$n = 6$	$\bar{x} = 4$		$s_X^2 = \frac{46}{6} = 7{,}67$
			$s_X = \sqrt{7{,}67} = 2{,}77$

Übungsaufgabe 4.2: Kickerturnier

An einem Kickerturnier der Soziologiestudierenden nahmen 6 Leute teil. Die Anzahl der Siege für die einzelnen Teilnehmer ist in folgender Tabelle wiedergegeben.

Teilnehmer	1	2	3	4	5	6
Siege	0	2	3	4	7	8

Berechnen Sie den Variationskoeffizienten!

Übungsaufgabe 4.2: Lösung

- Berechnung:
  - $V_X = \frac{s_X}{\bar{x}}$
  - $V_X = \frac{2{,}77}{4} = 0{,}6925$
- Interpretation:
  - Der Variationskoeffizient beträgt 0,6925. Dies bedeutet, dass die Standardabweichung der Siege bei 69,25% der durchschnittlichen Siege liegt.

Übungsaufgabe 5

In der folgenden Tabelle ist die Körpergröße von 5 Personen einmal in Zentimetern und einmal in Metern wiedergeben.

Person	1	2	3	4	5
Körpergröße in cm	160	170	180	165	175
Körpergröße in m	1,6	1,7	1,8	1,65	1,75

a) Berechnen Sie sowohl für die Körpergröße in Zentimetern als auch die Körpergröße in m jeweils getrennt die Variation, Varianz und die Standardabweichung. Vergleichen Sie die Werte! Was fällt ihnen auf?

b) Berechnen Sie für beide Variablen den Variationskoeffizienten. Was hat sich geändert? Wofür könnte dieser sinnvoll sein?

Übungsaufgabe 5a: Lösung

Körpergröße in Zentimetern

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
160	-10	100
170	0	0
180	10	100
165	-5	25
175	5	25
$\bar{x} = 170$		$SAQ_X = 250$
		$s_X^2 = 50$
		$s_X = 7{,}07$

Körpergröße in Metern

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1,60	-0,1	-0,01
1,70	0	0
1,80	0,1	0,01
1,65	-0,05	0,025
1,75	0,05	0,025
$\bar{x} = 1{,}70$		$SAQ_X = 0{,}025$
		$s_X^2 = 0{,}005$
		$s_X = 0{,}07$

Übungsaufgabe 5b: Lösung

Körpergröße in Zentimetern

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
160	-10	100
170	0	0
180	10	100
165	-5	25
175	5	25
$\bar{x} = 170$		$SAQ_X = 250$
		$s_X^2 = 50$
		$s_X = 7{,}07$

$$V_X = \frac{s_X}{\bar{x}} = \frac{7{,}07}{170} = 0{,}04$$

Körpergröße in Metern

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1,60	-0,1	-0,01
1,70	0	0
1,80	0,1	0,01
1,65	-0,05	0,025
1,75	0,05	0,025
$\bar{x} = 1{,}70$		$SAQ_X = 0{,}025$
		$s_X^2 = 0{,}005$
		$s_X = 0{,}07$

$$V_X = \frac{s_X}{\bar{x}} = \frac{0{,}07}{1{,}70} = 0{,}04$$

Übungsaufgabe 5b: Lösung II

- Interpretation:
  - sowohl bei der Varianz als auch bei der Standardabweichung unterscheiden sich beide Verteilungen um den Faktor
  - dies ist auf die unterschiedlichen Maßeinheiten der Ausgangsvariablen zurückzuführen
  - Der Variationskoeffizient bereinigt um den Effekt der Maßeinheit in dem durch das arithmetische geteilt wird
  - entsprechend weisen beide Variablen den gleichen Variationskoeffizienten auf

Literaturhinweis

- Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 87-116.
- Steffen-M. Kühnel / Dagmar Krebs (2012): Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen, Reinbek bei Hamburg: Rowohlt, S. 92-110 sowie 112-116.