

ORDINALE ZUSAMMENHÄNGE II UND METRISCHE ZUSAMMENHÄNGE I

Martin-Luther-Universität Halle-Wittenberg

Institut für Soziologie

Übung Einführung in die deskriptive Statistik

Was machen wir heute?

- Wiederholung
- asymmetrische ordinale Zusammenhangsmaße:
 - Somers d_{YX}
 - Somers d_{XY}
- symmetrische metrische Zusammenhangsmaße:
 - Kovariation und Kovarianz
 - Pearsons Produktmomentkorrelation $r_{X,Y}$

Warm-up: Beutebuch

Gamma

Cramers V

Prozentsatz-
differenz

Tau-a

Tau-b

Tau-c

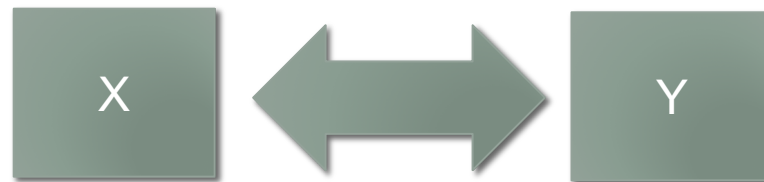
Somers Dyx

Yules Q

Lambda

Sortieren Sie die genannten Maße nach ihrem Skalenniveau und Eignung für symmetrische / asymmetrische Zusammenhänge!

Quiz: Welche der Darstellungen visualisiert einen symmetrischen Zusammenhang?



Quiz: Welches Maß eignet sich, wenn der Einfluss einer nominalen Variablen auf eine ordinale Variable untersucht werden soll?

Gamma

Cramér's V

Lambda

Somers D

Quiz: Welches Maß eignet sich, um einen Einfluss einer ordinalen Variablen (X) auf eine andere ordinale Variable (Y) zu untersuchen?

Lambda xy

Lambda yx

Somers Dyx

Somers Dxy

Quiz: Welche der folgenden Variablen weisen ein metrisches Skalenniveau auf?

Alter in Jahren

Geschlecht

Zustimmung zu einer These
(1=stimme voll zu, 5=stimme
überhaupt nicht zu)

Nettoeinkommen in Euro

Aufgabe 1: Allgemeiner Schulabschluss und Teilnahme an politischer Diskussion 2014

Eine Bildungsforscherin interessiert sich dafür, ob es einen Zusammenhang zwischen dem Schulabschluss und der Teilnahme an politischen Diskussionen gibt. Mithilfe des ALLBUS 2014 erhält sie folgende Kreuztabelle:

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

- Welches Zusammenhangsmaß ist hier geeignet, wenn Sie an einer Aussage zur proportionalen Fehlerreduktion nicht interessiert sind und Sie weder besonders konservativ noch euphorisch sein wollen?
- Berechnen Sie dieses Zusammenhangsmaß.
- Interpretieren Sie ihr Ergebnis statistisch und inhaltlich.

Aufgabe 1a: Lösung

symmetrischen
Fragestellung

Eine Bildungsforscherin interessiert sich dafür, ob es einen Zusammenhang zwischen dem Schulabschluss und der Teilnahme an politischen Diskussionen gibt.

quadratische
Mehrfeldertafel mit zwei
ordinalen Variablen →
größer/kleiner Relation
vorhanden

geeignetes Maß:
 τ_b (überschätzt nicht wie
 γ , unterschätzt nicht wie
 $\tau_a \dots$)

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

Aufgabe 1b: Lösung

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

- Konkordante Paare:

- $C = 56 * (264 + 325 + 218 + 140) + 92 * (325 + 140) + 189 * (218 + 140) + 264 * 140$
- $C = 53032 + 42780 + 67662 + 36960$
- $C = 200.434$

Aufgabe 1b: Lösung II

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

- Diskordante Paare:

- $D = 149 * (189 + 264 + 258 + 218) + 92 * (189 + 258) + 325 * (258 + 218) + 264 * 258$
- $D = 138421 + 41124 + 154700 + 68112$
- $D = 402.357$

Aufgabe 1b: Lösung III

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

- x-verbundene Paare:

- $T_X = 56 * (189 + 258) + 92 * (264 + 218) + 149 * (325 + 140) + 189 * 258 + 264 * 218 + 325 * 140$
- $T_X = 25032 + 44344 + 69285 + 48762 + 57552 + 45500$
- $T_X = 290.475$

Aufgabe 1b: Lösung IV

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

- **y-verbundene Paare:**

- $T_Y = 56 * (92 + 149) + 92 * 149 + 189 * (264 + 325) + 264 * 325 + 258 * (218 + 140) + 218 * 140$
- $T_Y = 13496 + 13708 + 111321 + 85800 + 92364 + 30520$
- $T_Y = 347.209$

Aufgabe 1b: Lösung IV

- Berechnung τ_b :

- $$\tau_b = \frac{C-D}{\sqrt{(C+D+T_X)*(C+D+T_Y)}}$$

- $$\tau_b = \frac{200.434-402.357}{\sqrt{(200.434+402.357 +290.475)*(200.434+402.357+347.209)}}$$

- $$\tau_b = -0,219$$

Aufgabe 1b: Lösung V

- Interpretation:
 - $\tau_b = -0,219$
 - Es besteht ein niedriger negativer Zusammenhang zwischen dem Schulabschluss und der Teilnahme an politischen Diskussionen. Hohe Werte in X gehen tendenziell eher mit niedrigen Werten in Y einher und umgekehrt.
 - Da hohe Werte in der X-Variablen für einen hohen Schulabschluss stehen und hohe Werte in der Y-Variablen für eine seltene Teilnahme bedeutet dies, dass Menschen mit hoher formaler Bildung tendenziell häufiger an politischen Diskussionen teilnehmen und umgekehrt.

Schulabschluss und politische Diskussion II

Eine Bildungsforscherin interessiert sich dafür, ob es einen Einfluss des Schulabschlusses auf die Teilnahme an politischen Diskussionen gibt. Mithilfe des ALLBUS 2014 erhält sie folgende Kreuztabelle:

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

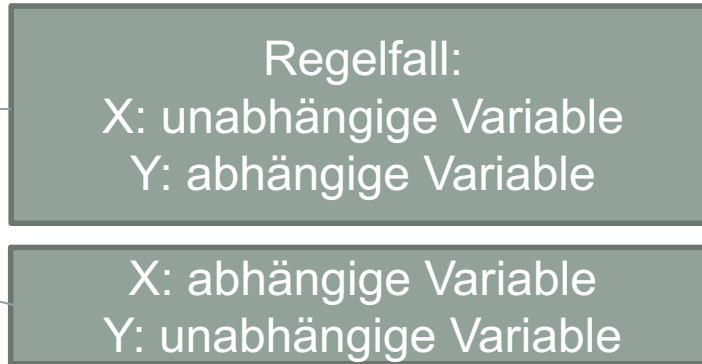
Quiz: Um welche Art von Zusammenhang handelt es sich bei der aufgestellten Aufgabenformulierung?

Asymmetrische Zusammenhänge zwischen ordinalen Variablen: Somers d_{YX} und d_{XY}

- Berechnung:

- $d_{YX} = \frac{C-D}{C+D+T_Y}$

- $d_{XY} = \frac{C-D}{C+D+T_X}$



- Interpretation:

- analog zu Cramérs V mit zusätzlicher Richtungsaussage
- Es besteht ein starker/mittlerer/schwacher positiver/negativer **Einfluss** der unabhängigen Variablen auf die abhängige Variable.

Aufgabe 2:

Schulabschluss und politische Diskussion

Eine Bildungsforscherin interessiert sich dafür, ob es einen Einfluss des Schulabschlusses auf die Teilnahme an politischen Diskussionen gibt. Mithilfe des ALLBUS 2014 erhält sie folgende Kreuztabelle:

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

$$C = 200.434, D = 402.357, T_X = 290.475, T_Y = 347.209$$

- Welches Maß ist hier zur Überprüfung des postulierten Zusammenhangs geeignet?
- Berechnen Sie dieses Maß.
- Interpretieren Sie Ihr Ergebnis inhaltlich und statistisch.

Aufgabe 2a: Analyse

asymmetrische
Fragestellung:
Bildungsabschluss(X) →
politische Diskussion(Y)

Eine Bildungsforscherin interessiert sich dafür, ob es einen Einfluss des Schulabschlusses auf die Teilnahme an politischen Diskussionen gibt.

Mehrfeldertafel mit zwei
ordinalen Variablen →
größer/kleiner Relation
vorhanden

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

geignetes Maß:
asymmetrisches
Paarvergleichsmaß
Somers d_{YX}

Aufgabe 2b: Lösung

- gegeben:

- $C = 200.434$
- $D = 402.357$
- $T_X = 290.475$
- $T_Y = 347.209$

- Berechnung:

- $d_{YX} = \frac{C-D}{C+D+T_Y}$
- $d_{YX} = \frac{200.434-402.357}{200.434+402.357+347.209}$
- $d_{YX} = -0,213$

Aufgabe 2c: Lösung

		Schulabschluss (rek)			Gesamtsumme
		1 - max. Hauptschule	2 - mittlere Reife	3 - Abitur	
Wie oft politische Diskussionen (rek)	1 - oft	56	92	149	297
	2 - manchmal	189	264	325	778
	3 - selten / nie	258	218	140	616
Gesamtsumme		503	574	614	1691

- Interpretation:

- $d_{YX} = -0,213$
- Es besteht ein niedriger negativer Einfluss des Schulabschlusses auf die Häufigkeit der politischen Diskussionen.
- Dies bedeutet, dass Leute mit höherem Bildungsabschluss tendenziell häufiger an politischen Diskussionen teilnehmen.

Übungssitzungen und Klausurpunkte

Ein M2-Dozent interessiert sich dafür, ob es einen Zusammenhang zwischen der Anzahl der verpassten Übungstermine und Punktzahl in der Klausur gibt. Für einen kleinen Durchgang mit 10 Personen erhält er folgende Daten:

Person	A	B	C	D	E	F	G	H	I	J
Verpasste Sitzungen (X)	0	5	4	3	2	0	1	0	1	0
Punkte (Y)	15	25	60	60	80	90	90	95	100	100

Quiz: Welches Skalenniveau liegt bei den Variablen „verpasste Sitzungen“ und „Punkte“ vor?

Kreuztabelle noch adäquate Darstellungsform?

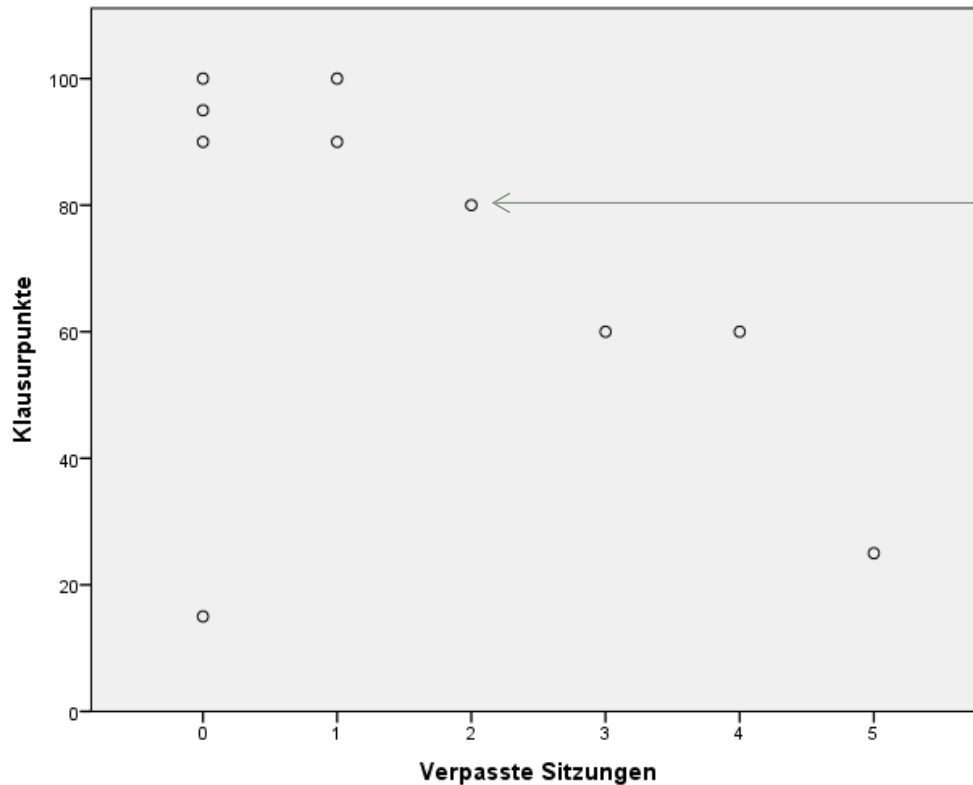
Kreuztabelle Klausurpunkte * Verpasste Sitzungen

Anzahl

		Verpasste Sitzungen						Gesamtsumme
		0	1	2	3	4	5	
Klausurpunkte	15	1	0	0	0	0	0	1
	25	0	0	0	0	0	1	1
	60	0	0	0	1	1	0	2
	80	0	0	1	0	0	0	1
	90	1	1	0	0	0	0	2
	95	1	0	0	0	0	0	1
	100	1	1	0	0	0	0	2
Gesamtsumme		4	2	1	1	1	1	10

Quiz: Ist hier die Darstellung des Zusammenhangs in einer Kreuztabelle adäquat? Warum oder warum nicht?

Streudiagramm I

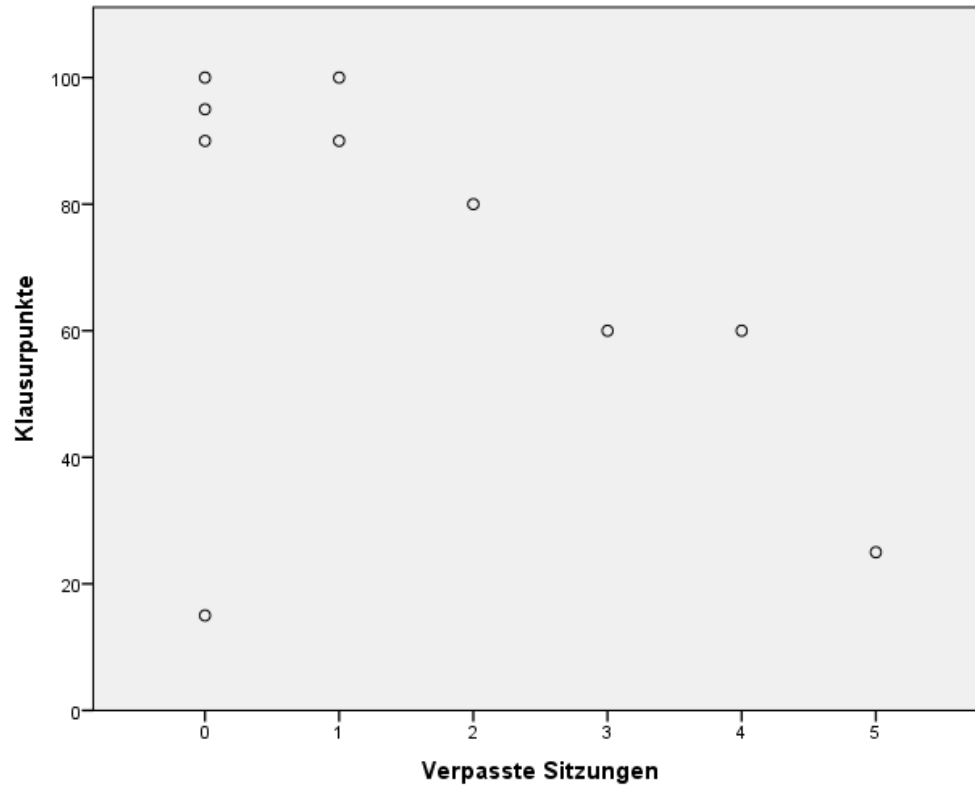


einzelner Fall

Lesehilfe: eine Person hat zwei Stunden verpasst und trotzdem 80 Punkte erzielt.

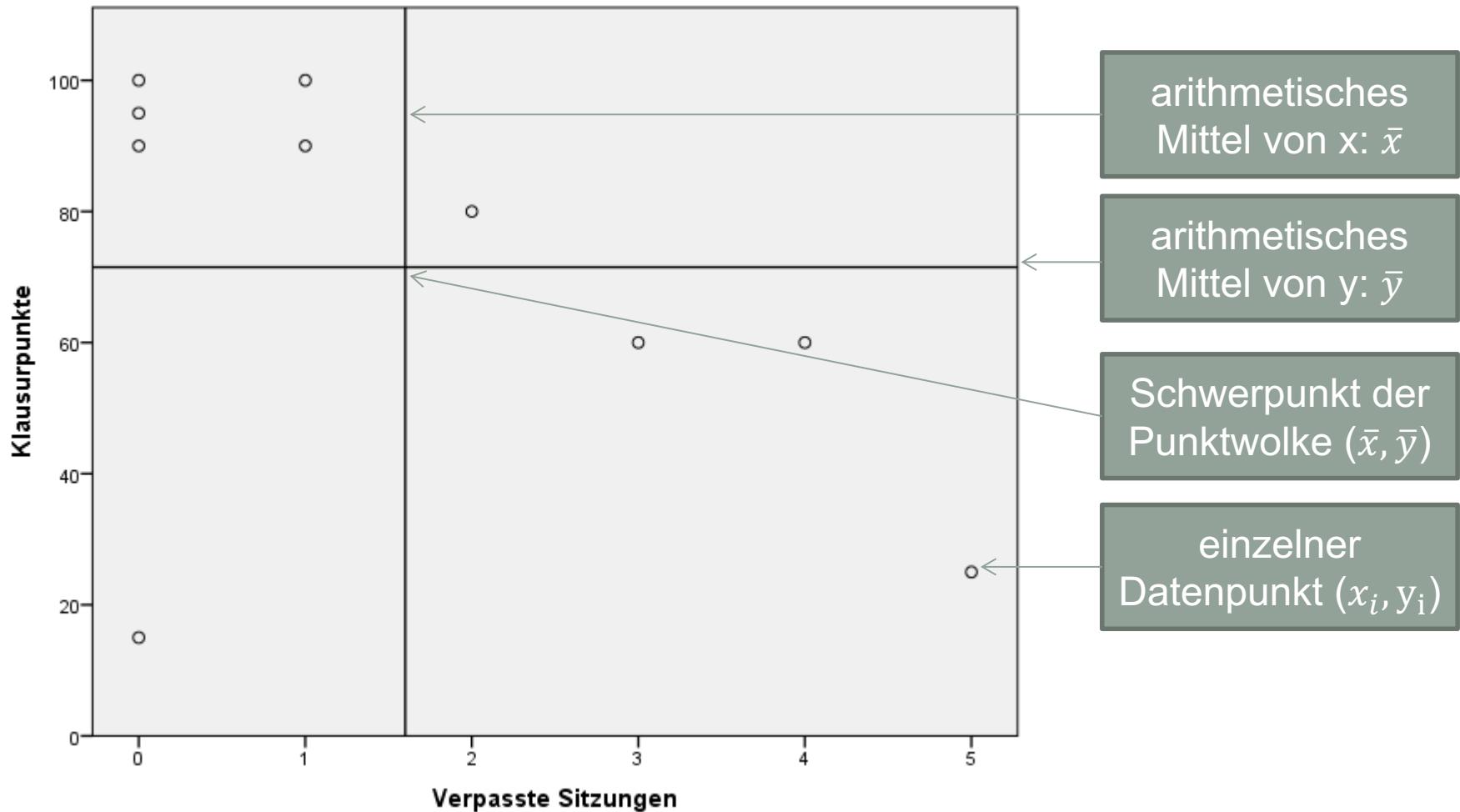
Im Scatterplot sind alle Fälle mit ihren Ausprägungen auf der X- und Y-Achse in einem Koordinatensystem abgetragen.

Streudiagramm II

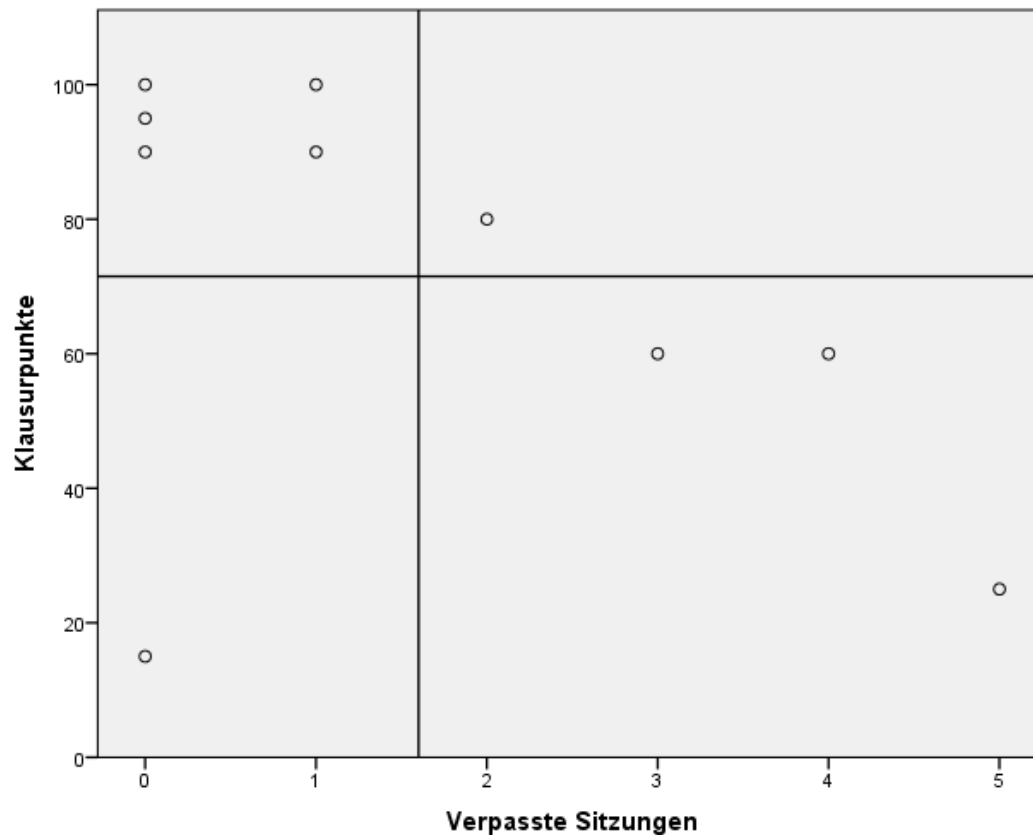


Quiz: Was für einen Zusammenhang zwischen den verpassten Übungen und den Klausurpunkten würden wir hier vermuten?

Kovariation I (Kreuzproduktsumme)



Kovariation II

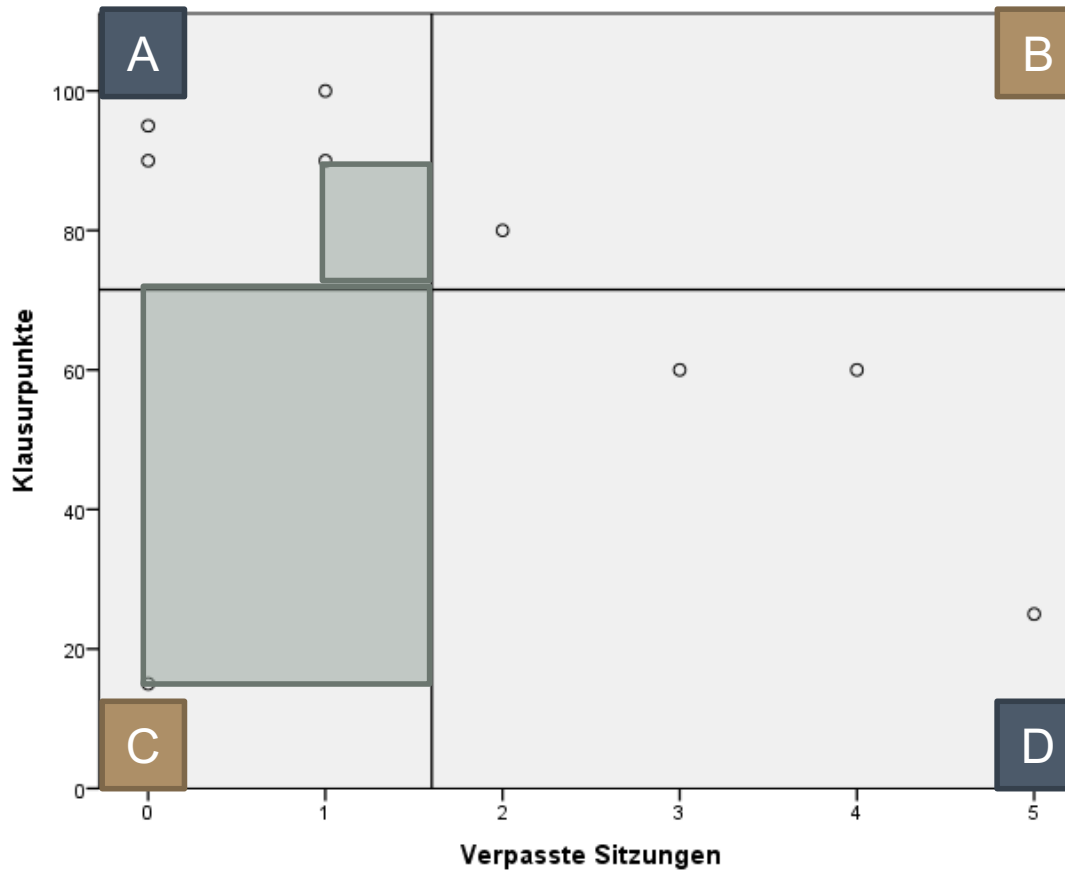


Kovariation betrachtet
Abweichungen vom
Schwerpunkt der
Punktwolke
 $(x_i - \bar{x}; y_i - \bar{y})$

Kovariation:

$$SP_{X,Y} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Kovariation III



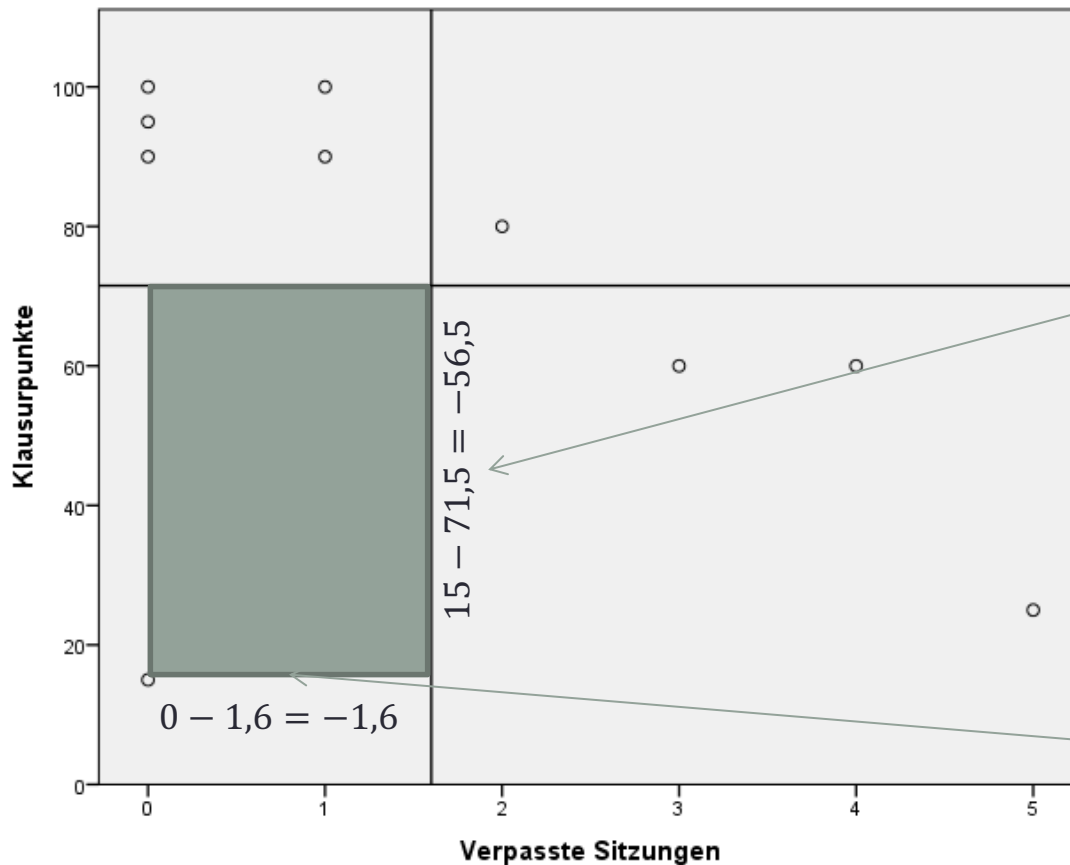
Kovariation:

$$SP_{X,Y} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Werte im Quadranten B und C deuten auf positive Beziehung hin (positives Vorzeichen)

Werte im Quadranten A und D deuten auf eine negative Beziehung hin (negatives Vorzeichen)

Kovariation: Beispiel I

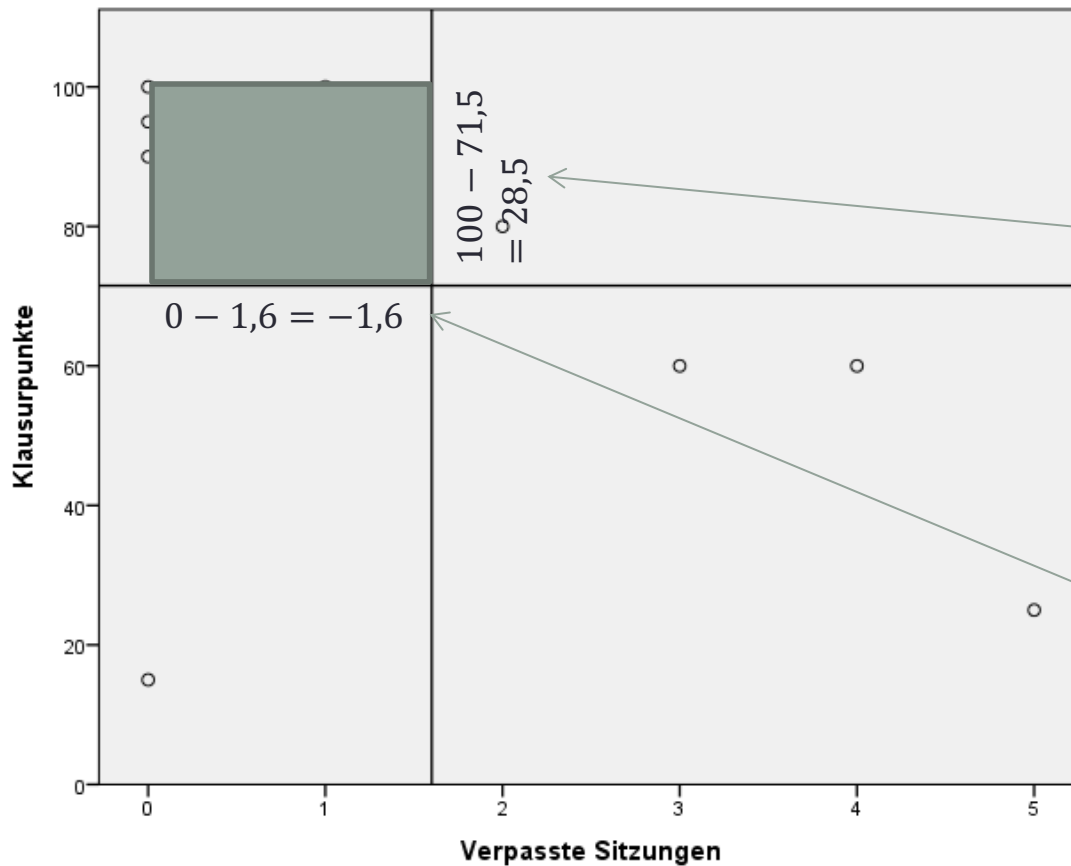


Person A hat mit 15 Punkten 56,5 Punkte weniger als der Durchschnitt
 $(y_i - \bar{y}) = -56,5$

Person A hat mit 0 Fehlstunden 1,6 Stunden weniger verpasst als der Durchschnitt
 $(x_i - \bar{x}) = -1,6$

$$(x_i - \bar{x}) * (y_i - \bar{y}) = -1,6 * (-56,5) = 90,4$$

Kovariation: Beispiel II



Person J hat mit 100 Punkten 28,5 Punkte mehr als der Durchschnitt
 $(y_i - \bar{y}) = 28,5$

Person J hat mit 0 Fehlstunden 1,6 Stunden weniger verpasst als der Durchschnitt
 $(x_i - \bar{x}) = -1,6$

$$(x_i - \bar{x}) * (y_i - \bar{y}) = -1,6 * 28,5 = -45,6$$

Kovariation IV

- Formel:
 - $SP_{X,Y} = \sum_{i=1}^n ((x_i - \bar{x}) * (y_i - \bar{y}))$
- Interpretation:
 - $SP_{X,Y} > 0 \rightarrow$ positiver linearer Zusammenhang
 - $SP_{X,Y} = 0 \rightarrow$ kein linearer Zusammenhang
 - $SP_{X,Y} < 0 \rightarrow$ negativer linearer Zusammenhang

Aufgabe 3.1: Kovariation

ID	Verpasste Übungen (x_i)	Klausur- punkte (y_i)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15			
B	5	25			
C	4	60			
D	3	60			
E	2	80			
F	0	90			
G	1	90			
H	0	95			
I	1	100			
J	0	100			
	$\bar{x} =$	$\bar{y} =$			$SP_{X,Y} =$

Berechnen Sie die Kovariation
 $SP_{XY} = \sum_{i=1}^n ((x_i - \bar{x}) * (y_i - \bar{y}))$
 und interpretieren Sie Ihr Ergebnis!

Aufgabe 3.1: Lösung

ID	Verpasste Übungen (x_i)	Klausur- punkte (y_i)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15			
B	5	25			
C	4	60			
D	3	60			
E	2	80			
F	0	90			
G	1	90			
H	0	95			
I	1	100			
J	0	100			
	$\bar{x} = 1,6$	$\bar{y} = 71,5$			$SP_{X,Y} =$

Aufgabe 3.1: Lösung II

ID	Verpasste Übungen (x_i)	Klausur- punkte (y_i)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15	-1,6	-56,5	
B	5	25	3,4	-46,5	
C	4	60	2,4	-11,5	
D	3	60	1,4	-11,5	
E	2	80	0,4	8,5	
F	0	90	-1,6	18,5	
G	1	90	-0,6	18,5	
H	0	95	-1,6	23,5	
I	1	100	-0,6	28,5	
J	0	100	-1,6	28,5	
	$\bar{x} = 1,6$	$\bar{y} = 71,5$			$SP_{X,Y} =$

Aufgabe 3.1: Lösung III

ID	Verpasste Übungen (x_i)	Klausur- punkte (y_i)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15	-1,6	-56,5	90,4
B	5	25	3,4	-46,5	-158,1
C	4	60	2,4	-11,5	-27,6
D	3	60	1,4	-11,5	-16,1
E	2	80	0,4	8,5	3,4
F	0	90	-1,6	18,5	-29,6
G	1	90	-0,6	18,5	-11,1
H	0	95	-1,6	23,5	-37,6
I	1	100	-0,6	28,5	-17,1
J	0	100	-1,6	28,5	-45,6
	$\bar{x} = 1,6$	$\bar{y} = 71,5$			$SP_{X,Y} = -249$

Aufgabe 3.1: Lösung IV

- Interpretation Kovariation:
 - $SP_{X,Y} = -249$
 - Es besteht ein negativer Zusammenhang zwischen der Anzahl der verpassten Stunden und der Klausurpunktzahl.
 - Menschen, die mehr Übungen verpasst haben, schneiden tendenziell in der Klausur eher schlechter ab.

Problem: Kovariation ist abhängig von Fallzahl und Maßeinheit
→ keine Aussage über die Stärke des Zusammenhangs möglich

Kovarianz

- Verwendung:
 - „unnormiertes“ Zusammenhangsmaß für symmetrische Zusammenhänge zwischen metrischen Variablen
- Berechnung:
 - $s_{X,Y} = \frac{1}{n} * \sum_{i=1}^n ((x_i - \bar{x}) * (y_i - \bar{y}))$
 - $s_{X,Y} = \frac{SP_{X,Y}}{n}$

Hinweis: Dies ist die Formel für die deskriptive Statistik. In der Inferenzstatistik bzw. bei Statistikprogrammen wird die Kovarianz wie auch die Varianz ermittelt, in dem durch n-1 geteilt wird.

Aufgabe 3.2: Kovarianz

ID	Verpasste Übungen (x_i)	Klausur- punkte (y_i)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15	-1,6	-56,5	90,4
B	5	25	3,4	-46,5	-158,1
C	4	60	2,4	-11,5	-27,6
D	3	60	1,4	-11,5	-16,1
E	2	80	0,4	8,5	3,4
F	0	90	-1,6	18,5	-29,6
G	1	90	-0,6	18,5	-11,1
H	0	95	-1,6	23,5	-37,6
I	1	100	-0,6	28,5	-17,1
J	0	100	-1,6	28,5	-45,6
	$\bar{x} = 1,6$	$\bar{y} = 71,5$			$SP_{X,Y} = -249$

Berechnen Sie die Kovarianz $s_{X,Y} = \frac{1}{n} * \sum_{i=1}^n ((x_i - \bar{x}) * (y_i - \bar{y})) = \frac{SP_{X,Y}}{n}$

Aufgabe 3.2: Lösung

ID	(x_i)	(y_i)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15	-1,6	-56,5	90,4
B	5	25	3,4	-46,5	-158
C	4	60	2,4	-11,5	-27,6
D	3	60	1,4	-11,5	-16,1
E	2	80	0,4	8,5	3,4
F	0	90	-1,6	18,5	-29,6
G	1	90	-0,6	18,5	-11,1
H	0	95	-1,6	23,5	-37,6
I	1	100	-0,6	28,5	-17,1
J	0	100	-1,6	28,5	-45,6
	$\bar{x} = 1,6$	$\bar{y} = 71,5$			$SP_{X,Y} = -249$
					$s_{X,Y} = \frac{SP_{X,Y}}{10} = -24,9$

Kovarianz II

- Wertebereich:
 - $[-\infty; \infty]$
- Eigenschaften:
 - Kovarianz normiert Kovariation an Fallzahl
 - wird dadurch Fallzahl unabhängig
 - Höhe des Wertes dennoch abhängig von Maßeinheit
- Lösung:
 - Normierung an Standardabweichung der Variablen X und Y
 - Berechnung von Pearsons Produktmomentkorrelation $r_{X,Y}$

Pearsons Produktmomentkorrelation $r_{X,Y}$

- Verwendung:
 - standardisiertes symmetrisches Maß für Zusammenhänge zwischen metrischen Variablen
- Eigenschaften:
 - unabhängig von Fallzahl und Maßeinheiten

- Berechnung:

$$r_{X,Y} = \frac{s_{X,Y}}{s_X * s_Y}$$

Kovarianz von X,Y

Standardabweichung von X *
Standardabweichung von Y

Pearsons Produktmomentkorrelation $r_{X,Y}$ II

- Alternative Berechnung:

$$r_{X,Y} = \frac{SP_{XY}}{\sqrt{SAQ_X * SAQ_Y}}$$

Kovariation von X,Y

Wurzel der (Variation von X * Variation von Y)

- Interpretation:

- $r_{X,Y} = -1$: perfekter negativer **linearer** Zusammenhang
- $r_{X,Y} = 0$: kein **linearer** Zusammenhang
- $r_{X,Y} = +1$: perfekter positiver **linearer** Zusammenhang

Aufgabe 3.3: Pearsons $r_{X,Y}$

ID	(x_i)	(y_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15	-1,6		-56,5		90,4
B	5	25	3,4		-46,5		-158,1
C	4	60	2,4		-11,5		-27,6
D	3	60	1,4		-11,5		-16,1
E	2	80	0,4		8,5		3,4
F	0	90	-1,6		18,5		-29,6
G	1	90	-0,6		18,5		-11,1
H	0	95	-1,6		23,5		-37,6
I	1	100	-0,6		28,5		-17,1
J	0	100	-1,6		28,5		-45,6
	$\bar{x} = 1,6$	$\bar{y} = 71,5$		$SAQ_X =$		$SAQ_Y =$	$SP_{X,Y} = -249$

Berechnen Sie bitte Pearson $r_{X,Y} = \frac{SP_{XY}}{\sqrt{SAQ_X * SAQ_Y}}$
und interpretieren Sie Ihr Ergebnis.

Aufgabe 3.3: Lösung

ID	(x_i)	(y_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
A	0	15	-1,6	2,56	-56,5	3192,3	90,4
B	5	25	3,4	11,56	-46,5	2162,3	-158,1
C	4	60	2,4	5,76	-11,5	132,25	-27,6
D	3	60	1,4	1,96	-11,5	132,25	-16,1
E	2	80	0,4	0,16	8,5	72,25	3,4
F	0	90	-1,6	2,56	18,5	342,25	-29,6
G	1	90	-0,6	0,36	18,5	342,25	-11,1
H	0	95	-1,6	2,56	23,5	552,25	-37,6
I	1	100	-0,6	0,36	28,5	812,25	-17,1
J	0	100	-1,6	2,56	28,5	812,25	-45,6
	$\bar{x} = 1,6$	$\bar{y} = 71,5$		$SAQ_X = 30,4$		$SAQ_Y = 8552,5$	$SP_{X,Y} = -249$

Aufgabe 3.3: Lösung II

- Berechnung $r_{X,Y}$:

- $r_{X,Y} = \frac{SP_{XY}}{\sqrt{SAQ_X * SAQ_Y}}$

- $r_{X,Y} = \frac{-249}{\sqrt{30,4 * 8552,5}}$

- $r_{X,Y} = -0,488$

- Interpretation:

- $r_{X,Y} = -0,488$. Es besteht ein mittlerer negativer linearer Zusammenhang zwischen der Anzahl der verpassten Stunden und den Punkten in der Klausur. Eine höhere Anzahl verpasster Stunden geht einher mit einer geringeren Klausurpunktzahl.

Literaturhinweise

- Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 204-234.
- Steffen-M. Kühnel / Dagmar Krebs (2012): Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. S. 433-444.
- Hans Benninghaus (2007): Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler. S. 185-227.

Übungsaufgabe 1: Parteimitgliedschaft und Wahlbeteiligung junger Europäer

Ein Sozialwissenschaftler möchte wissen, ob es einen Zusammenhang zwischen der Mitgliedschaft junger Europäer und der Wahlteilnahme auf Landesebene gibt. Mithilfe des Flash Eurobarometers erhält er folgende Daten.

- Welches Maß ist hier zur Beschreibung des Zusammenhangs geeignet?
- Berechnen Sie dieses Maß!
- Interpretieren Sie dieses Maß inhaltlich und statistisch!

Land	Mitgliedschaft politische Organisation in % (X)	Wahlteilnahme in % (Y)
BE	8,2	92,4
BG	6,8	69,0
DK	8,3	87,5
EE	2,6	69,5
GB	4,4	52,3
HR	6,2	83,2
HU	1,2	53,4
MT	14,0	95,4
PL	2,6	71,8
SE	7,4	83,8

Übungsaufgabe 1a: Analyse

symmetrische
Fragestellung:
keine Kausalrichtung
angegeben

Ein Sozialwissenschaftler möchte wissen, ob es einen Zusammenhang zwischen der Mitgliedschaft junger Europäer und der Wahlteilnahme auf Landesebene gibt.

zwei metrische
Variablen, Abstände
interpretierbar

geeignetes Maß:
Pearson
Produktmoment-
korrelation $r_{X,Y}$

Land	Mitgliedschaft politische Organisation in % (X)	Wahlteilnahme in % (Y)
BE	8,2	92,4
BG	6,8	69,0
DK	8,3	87,5
EE	2,6	69,5
GB	4,4	52,3
HR	6,2	83,2
HU	1,2	53,4
MT	14,0	95,4
PL	2,6	71,8
SE	7,4	83,8

Übungsaufgabe 1b: Lösung I

ID	(x_i)	(y_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
BE	8,2	92,4					
BG	6,8	69,0					
DK	8,3	87,5					
EE	2,6	69,5					
GB	4,4	52,3					
HR	6,2	83,2					
HU	1,2	53,4					
MT	14,0	95,4					
PL	2,6	71,8					
SE	7,4	83,8					
	$\bar{x} =$	$\bar{y} =$		$SAQ_X =$		$SAQ_Y =$	$SP_{X,Y} =$

Übungsaufgabe 1b: Lösung II

ID	(x_i)	(y_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
BE	8,2	92,4	2,03	4,1209	16,57	274,5649	33,6371
BG	6,8	69,0	0,63	0,3969	-6,83	46,6489	-4,3029
DK	8,3	87,5	2,13	4,5369	11,67	136,1889	24,8571
EE	2,6	69,5	-3,57	12,7449	-6,33	40,0689	22,5981
GB	4,4	52,3	-1,77	3,1329	-23,53	553,6609	41,6481
HR	6,2	83,2	0,03	0,0009	7,37	54,3169	0,2211
HU	1,2	53,4	-4,97	24,7009	-22,43	503,1049	111,4771
MT	14,0	95,4	7,83	61,3089	19,57	382,9849	153,2331
PL	2,6	71,8	-3,57	12,7449	-4,03	16,2409	14,3871
SE	7,4	83,8	1,23	1,5129	7,97	63,5209	9,8031
	$\bar{x} =$ 6,17	$\bar{y} =$ 75,83		$SAQ_x =$ 125,201		$SAQ_y =$ 2071,301	$SP_{x,y} = 407,559$

Übungsaufgabe 1b: Lösung III

- gegeben:

- $SAQ_X = 125,201$
- $SAQ_Y = 2071,301$
- $SP_{X,Y} = 407,559$

- Berechnung:

- $r_{X,Y} = \frac{SP_{XY}}{\sqrt{SAQ_X * SAQ_Y}}$
- $r_{X,Y} = \frac{407,559}{\sqrt{125,201 * 2071,301}}$
- $r_{X,Y} = 0,800$

Übungsaufgabe 1c: Lösung

- Interpretation:
 - $r_{X,Y} = 0,800$
 - Es besteht ein hoher positiver linearer Zusammenhang zwischen dem prozentualen Anteil der jungen Mitglieder einer politischen Organisation eines Landes und der entsprechenden Wahlteilnahme.
 - Je höher der Anteil der Anteil der jungen Menschen in politischen Organisationen war, umso höher war auch der Anteil derer, die an einer Wahl in den letzten drei Jahren teilgenommen haben.

Übungsaufgabe 2: Verpflichtung gegenüber Natur und Biodiversität

Im Rahmen des Eurobarometers wurden Befragte aus 10 Ländern gebeten anzugeben, ob sie glauben, dass es eine Verpflichtung gibt auf die Natur aufzupassen und das das Wohlergehen der Menschen von der Biodiversität abhängt. Bei beiden Fragen wurde die Zustimmung der Befragten innerhalb des jeweiligen Landes in Prozent erhoben. Gibt es einen linearen Zusammenhang zwischen beiden Variablen?

- a) Berechnen Sie mithilfe der dargestellten aggregierten Eurobarometerdaten Kovariation, Kovarianz und Pearsons r .
- b) Interpretieren Sie Ihr Ergebnis!

Land	Zustimmung Verpflichtung Natur aufpassen in % (X)	Zustimmung Wohlergehen abhängig von Biodiversität in % (Y)
AT	98	95,7
CY	99,2	98
ES	98,6	95,3
FI	97,2	93,7
HR	98,9	98
IE	97,6	96,7
IT	99,2	96,7
PL	96,7	90,7
PT	98,4	96,7
RO	98,2	97,7

Übungsaufgabe 2a: Lösung

ID	$\%(x_i)$	$\%(y_i)$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
AT	98	95,7					
CY	99,2	98					
ES	98,6	95,3					
FI	97,2	93,7					
HR	98,9	98					
IE	97,6	96,7					
IT	99,2	96,7					
PL	96,7	90,7					
PT	98,4	96,7					
RO	98,2	97,7					
	$\bar{x} = 98,2$	$\bar{y} = 95,92$		$SAQ_X =$		$SAQ_Y =$	$SP_{XY} =$

Übungsaufgabe 2a: Lösung II

ID	$\%(x_i)$	$\%(y_i)$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
AT	98	95,7	-0,2	0,04	-0,22	0,0484	0,044
CY	99,2	98	1	1	2,08	4,3264	2,08
ES	98,6	95,3	0,4	0,16	-0,62	0,3844	-0,248
FI	97,2	93,7	-1	1	-2,22	4,9284	2,22
HR	98,9	98	0,7	0,49	2,08	4,3264	1,456
IE	97,6	96,7	-0,6	0,36	0,78	0,6084	-0,468
IT	99,2	96,7	1	1	0,78	0,6084	0,78
PL	96,7	90,7	-1,5	2,25	-5,22	27,2484	7,83
PT	98,4	96,7	0,2	0,04	0,78	0,6084	0,156
RO	98,2	97,7	0	0	1,78	3,1684	0
	$\bar{x} = 98,2$	$\bar{y} = 95,92$		$SAQ_X = 6,34$		$SAQ_Y = 46,256$	$SP_{XY} = 13,85$

Übungsaufgabe 2a: Lösung III

- Kovariation:

- $SP_{XY} = 13,85$

- Kovarianz:

- $s_{XY} = \frac{SP_{XY}}{N} = \frac{13,85}{10} = 1,385$

- Pearsons r_{XY} :

- $r_{XY} = \frac{SP_{XY}}{\sqrt{SAQ_X * SAQ_Y}}$

- $r_{XY} = \frac{13,85}{\sqrt{6,34 * 46,256}} = 0,81$

Übungsaufgabe 2b: Lösung

- Interpretation:
 - $r_{XY} = 0,81$
 - Es besteht ein hoher positiver linearer Zusammenhang zwischen beiden Variablen. Hohe Werte auf der einen Variablen gehen tendenziell mit hohen Werten auf der anderen Variablen einher und umgekehrt.
 - In Ländern, in denen die Menschen zustimmen, dass es eine moralische Verpflichtung zum Erhalt der Natur gibt, stimmen tendenziell auch der These zu, dass das Wohlergehen des Menschen von der Biodiversität abhängig ist und umgekehrt.

Übungsaufgabe 3: Guter Ort für Minderheiten?

Eine Sozialwissenschaftlerin vermutet, dass es einen Zusammenhang zwischen der Einschätzung eines Landes als guter Ort für ethnische Minderheiten und für Migranten gibt. Bitte berechnen Sie alle Maße, die Sie für die Beschreibung einer symmetrischen Beziehung kennengelernt haben. Erstellen Sie eine Arbeitstabelle und dokumentieren Sie jeden Schritt. Interpretieren Sie alle Maßzahlen.

Land	Guter Ort für ethnische Minderheiten in Prozent	Guter Ort für Schwule/Lesben in Prozent	Guter Ort für Migranten in Prozent
Australien	86,2	79,0	87,2
Belgien	63,5	66,8	67,6
Irland	82,6	69,8	82,2
Israel	41,1	32,3	38,3
Mexiko	50,9	44,9	46,4
Niederlande	79,4	82,9	81,8
Norwegen	86,2	61,1	71,9
Spanien	74,4	76,7	81,6
Schweden	76,8	69,7	77,1
UK	76,1	65,1	70,7

Übungsaufgabe 3: Lösung

ethnische Minderheiten (x_i)	Migranten (y_i)	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
86,2	87,2	14,48	16,72	209,67	279,558	242,106
63,5	67,6	-8,22	-2,88	67,5684	8,2944	23,6736
82,6	82,2	10,88	11,72	118,374	137,358	127,514
41,1	38,3	-30,6	-32,2	937,584	1035,55	985,352
50,9	46,4	-20,8	-24,1	433,472	579,846	501,346
79,4	81,8	7,68	11,32	58,9824	128,142	86,9376
86,2	71,9	14,48	1,42	209,67	2,0164	20,5616
74,4	81,6	2,68	11,12	7,1824	123,654	29,8016
76,8	77,1	5,08	6,62	25,8064	43,8244	33,6296
76,1	70,7	4,38	0,22	19,1844	0,0484	0,9636
$\bar{x} = 71,72$	$\bar{y} = 70,48$			$SAQ_X = 2087,5$	$SAQ_Y = 2338,3$	$SP_{XY} = 2051,88$

Übungsaufgabe 3: Lösung II

- Kovariation:
 - $SP_{XY} = 2051,88$
 - Es besteht ein positiver Zusammenhang zwischen beiden Variablen, wobei anhand der Kovariation keine Aussage über die Stärke getroffen werden kann, da dieses Maß abhängig ist von Fallzahl und Maßeinheiten.

Übungsaufgabe 3: Lösung IV

- Kovarianz:

- $s_{XY} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) * (y_i - \bar{y}) = \frac{SP_{XY}}{N}$

- $s_{XY} = \frac{2051,88}{10} = 205,188$

- Auch die Kovarianz zeigt den positiven Zusammenhang an, lässt sich aber ebenfalls nicht hinsichtlich der Stärke deuten, da hier noch die Maßeinheit der Variablen einen Einfluss ausübt.

Übungsaufgabe 3: Lösung IV

- Pearsons r

- $$r_{XY} = \frac{SP_{XY}}{\sqrt{SAQ_X * SAQ_Y}} = \frac{2051,88}{\sqrt{2087,5 * 2338,3}} = 0,929$$

- Es besteht ein starker positiver Zusammenhang zwischen beiden Variablen. Hohe Werte auf der einen Variablen hängen tendenziell mit hohen Werten auf der anderen Variablen zusammen.
- Für Länder, wo die Lage der ethnischen Minderheiten als gut eingeschätzt wird, wird diese auch für die Migranten als gut eingeschätzt.