

M24 Statistik 1: Wintersemester 2024 / 2025

Seminar 04: Korrelation

MSc Albert Anoschin & Prof. Matthias Guggenmos
Health and Medical University Potsdam



Beispiel: Berechnung eines Scores für Sinnerleben

Meaning in Life Measure (Experience Subscale)

MILM_E *

Bitte beantworten Sie die folgenden Aussagen so ehrlich und genau wie möglich. Beachten Sie, dass dies sehr subjektive Fragen sind. Es gibt keine falschen oder richtigen Antworten.

	1 = stimme gar nicht zu	2	3	4 = neutral	5	6	7 = stimme sehr zu
Ich erlebe mein Leben als sinnvoll.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich werde anderen in Erinnerung bleiben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe etwas, das ich in meinem Leben erreichen will.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann Ereignisse in meiner Vergangenheit und Gegenwart sinnvoll verknüpfen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Option 1: Drag & Drop

Name: MILM_Mean Langer Name: MILM_Mean

Spaltentyp: Metrisch Beschreibung: ...

Berechneter Typ: Computed with drag-and-drop

Definition berechneter Spalte Fehlende Werte

+ - * ÷ / ^ √ % = ≠ < ≤ > ≥ ∧ ∨ | ¬

BFI_10
MILM_01
MILM_02
MILM_03
MILM_04

$$(((\text{MILM_01} + \text{MILM_02}) + \text{MILM_03}) + \text{MILM_04})/4$$

Option 2: R-code Befehl „rowMean“

Name: MILM_Mean Langer Name:

Spaltentyp: Metrisch Beschreibung: ...

Berechneter Typ: Computed with R code

Definition berechneter Spalte Fehlende Werte

$$\text{rowMean}(\text{MILM_01}, \text{MILM_02}, \text{MILM_03}, \text{MILM_04})$$

Übungen

1. Wir haben aktuell empfundene Sinnlosigkeit mit 6 Items gemessen (MLD_01 bis MLD_06). Berechnen Sie einen „Mean Score“ der Sinnlosigkeit per Drag & Drop oder per „rowMean“ Befehl. Benennen Sie die Variable mit „MLD_Mean“.
2. Berechnen Sie einen Summenscore für Mobiltelefonabhängigkeit, basierend auf den 11 Items des TMD-Brief. Nutzen Sie entweder Drag & Drop oder den Befehl „rowSum“ in JASP. Benennen Sie die Variable mit „TMD_Sum“.
3. Lassen Sie sich deskriptive Statistiken (Standardabweichung, Mittelwert, Median) von TMD_Sum ausgeben.
4. Wir definieren alle Summenwerte, die eine Standardabweichung über dem Mittelwert liegen als „problematische“ Handynutzung. Dies entspricht Summenwerten über 28. Berechnen Sie eine neue Variable namens „TMD_above_cutoff“, die anzeigt, ob der Summenwert einer Person über oder unter diesem Cutoff von 28 liegt.
5. Prüfen Sie, ob sich Personen mit einer „problematischen“ Handynutzung in a) Bildschirmzeit, b) Konsumhäufigkeit von „Reels“ c) Stillsitz-Fähigkeit, d) Sozialkontakte, e) Einsamkeit, f) Lebenszufriedenheit, g) Lebenssinn und h) empfundener Sinnlosigkeit von Personen mit einer unproblematischen Handynutzung unterscheiden.

Übungen

6. Wie würden Sie einen „Score“ für Extraversion berechnen?

Items				
Nr.	Deutsche Items	Englische Items	Polung	Subskalen
1	Ich bin eher zurückhaltend, reserviert.	I see myself as someone who is reserved.	-	Extraversion
2	Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	I see myself as someone who is generally trusting.	+	Verträglichkeit
3	Ich bin bequem, neige zur Faulheit.	I see myself as someone who tends to be lazy.	-	Gewissenhaftigkeit
4	Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	I see myself as someone who is relaxed, handles stress well.	-	Neurotizismus
5	Ich habe nur wenig künstlerisches Interesse.	I see myself as someone who has few artistic interests.	-	Offenheit
6	Ich gehe aus mir heraus, bin gesellig.	I see myself as someone who is outgoing, sociable.	+	Extraversion
7	Ich neige dazu, andere zu kritisieren.	I see myself as someone who tends to find fault with others.	-	Verträglichkeit
8	Ich erledige Aufgaben gründlich.	I see myself as someone who does a thorough job.	+	Gewissenhaftigkeit
9	Ich werde leicht nervös und unsicher.	I see myself as someone who gets nervous easily.	+	Neurotizismus
10	Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	I see myself as someone who has an active imagination.	+	Offenheit

BFI-10: Kurzskala der „Big Five“ Persönlichkeitseigenschaften

„Umpolung“ von Items am Beispiel Extraversion

		Trifft überhaupt nicht zu	Trifft eher nicht zu	Weder noch	Eher zutreffend	Trifft voll und ganz zu
BFI_01	Ich bin eher zurückhaltend, reserviert.	1	2	3	4	5
BFI_06	Ich gehe aus mir heraus, bin gesellig.	1	2	3	4	5

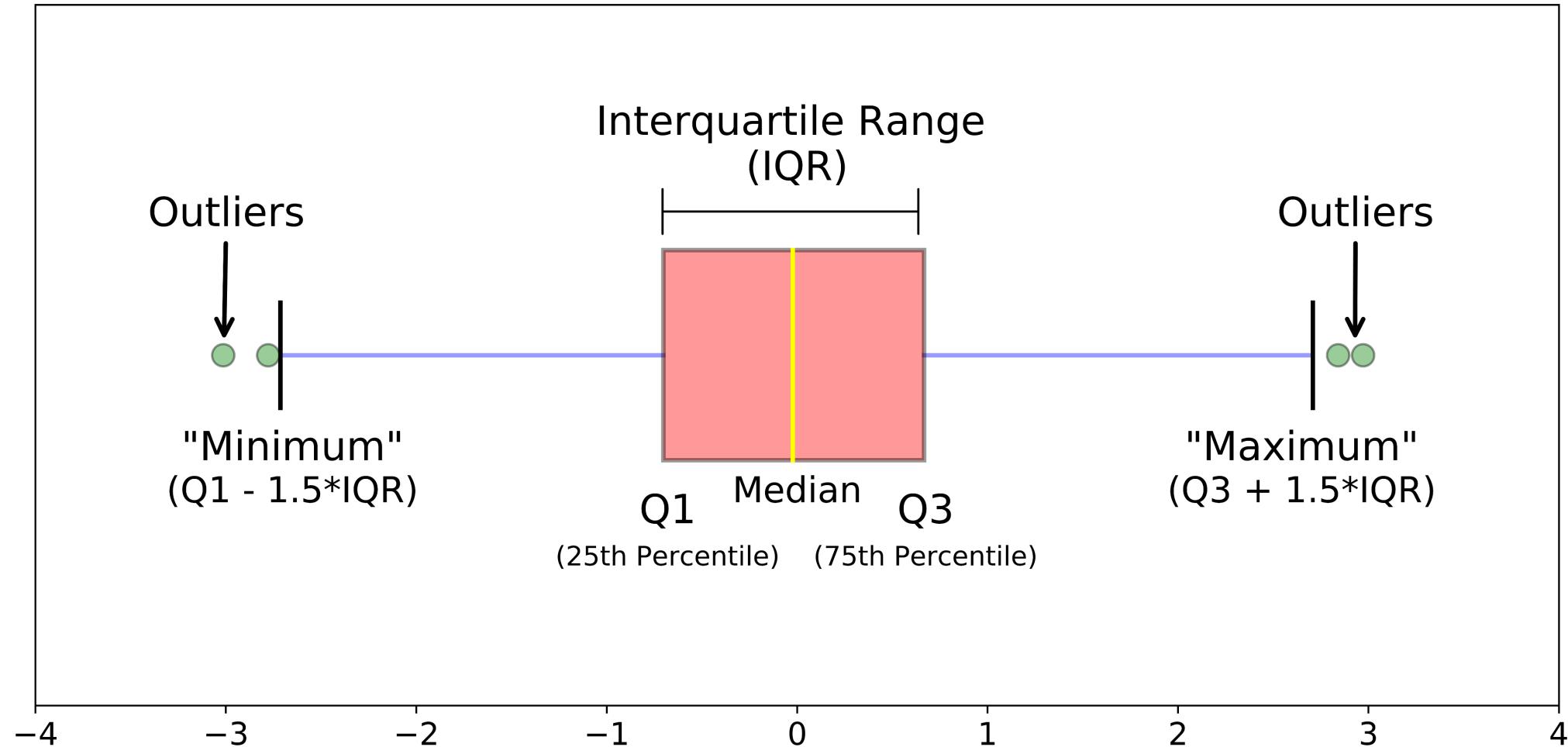
$$\text{Extraversion} = \frac{(6 - BFI_{01}) + BFI_{06}}{2}$$

Welche Aussagen treffen zu?

Der Interquartilabstand...

1. basiert auf der Einteilung der Daten in 4 Abschnitte gleicher Größe
2. kann mit einem Boxplot dargestellt werden
3. wird durch Ausreißer stark verfälscht
4. umfasst 50% der Werte in der Mitte der Verteilung

Visualisierung der Streuung: Boxplots



Bestimmen Sie den IQR für die Variable „Lebenszufriedenheit“

Statistik Vorlesung 03: Lage- und Streuungsmaße

- Berechnung des IQR:

1. Sortiere alle Werte von klein nach groß

2. Bestimme die Tiefe des Medians: $Tiefe_{Median} = \frac{n+1}{2}$

3. Bestimme die Tiefe des Quartils: $Tiefe_{Quartil} = \frac{Tiefe_{Median} + 1}{2} = \frac{\frac{n+1}{2} + 1}{2} = \frac{n+3}{4}$

4. Für das 25%-Quantil (Q_1) geht man **von vorne** in die Datenreihe

5. Für das 75%-Quantil (Q_3) geht man **von hinten** in die Datenreihe

$$IQR = Q_3 - Q_1$$

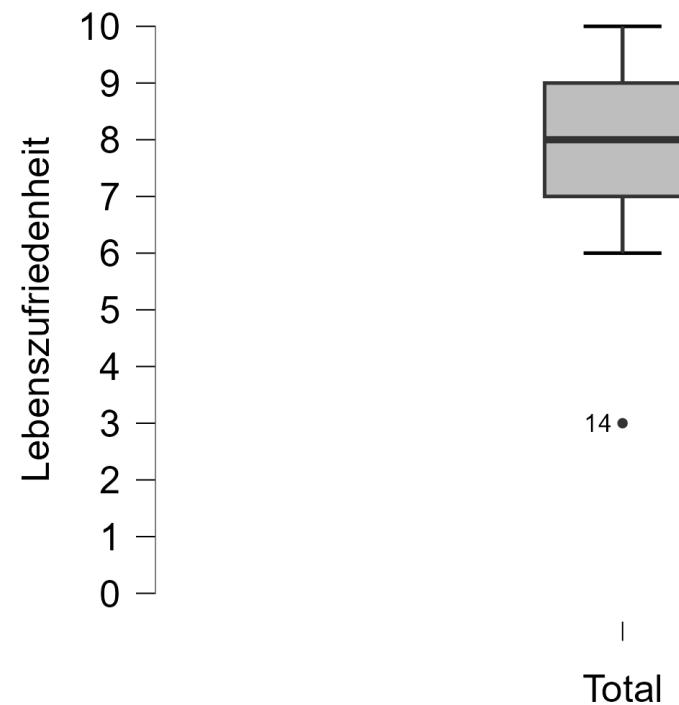
Bestimmung des IQR für die Lebenszufriedenheit

- Sortierung der Daten: {3, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10}
- Tiefe des Medians: $(n + 1) / 2 = 9$
- Median = 8
- Tiefe des Quartils: $(\text{TiefeMedian} + 1) / 2 = 5$
- $Q1 = 7$
- $Q3 = 9$
- $IQR = Q3 - Q1 = 9 - 7 = 2$

Zeichnen Sie den
Boxplot!

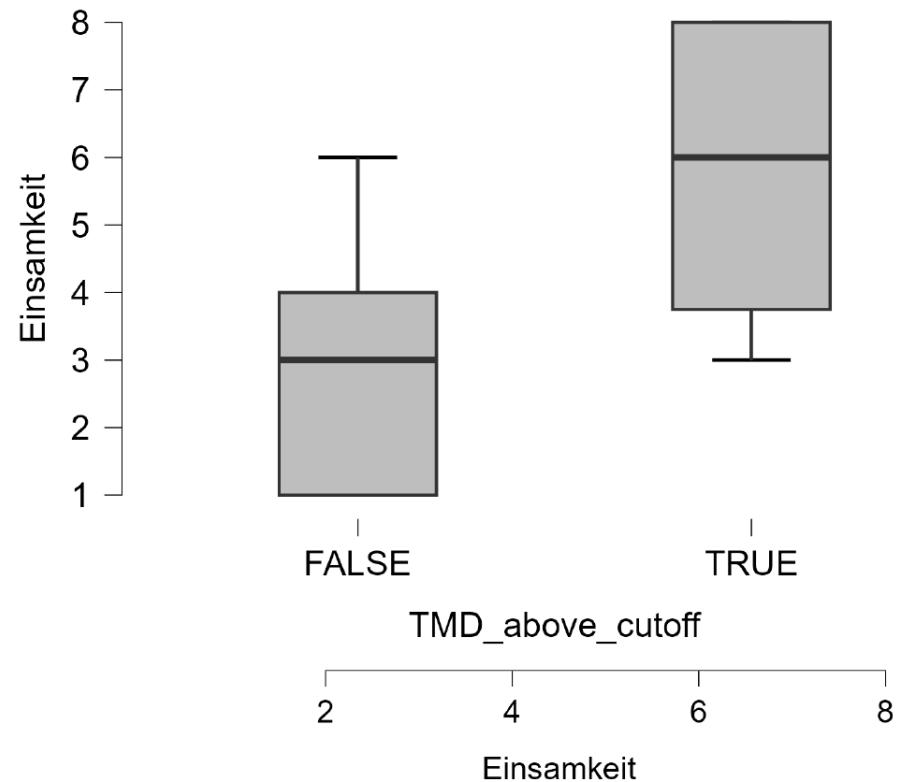
Lösung

{3, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 10}



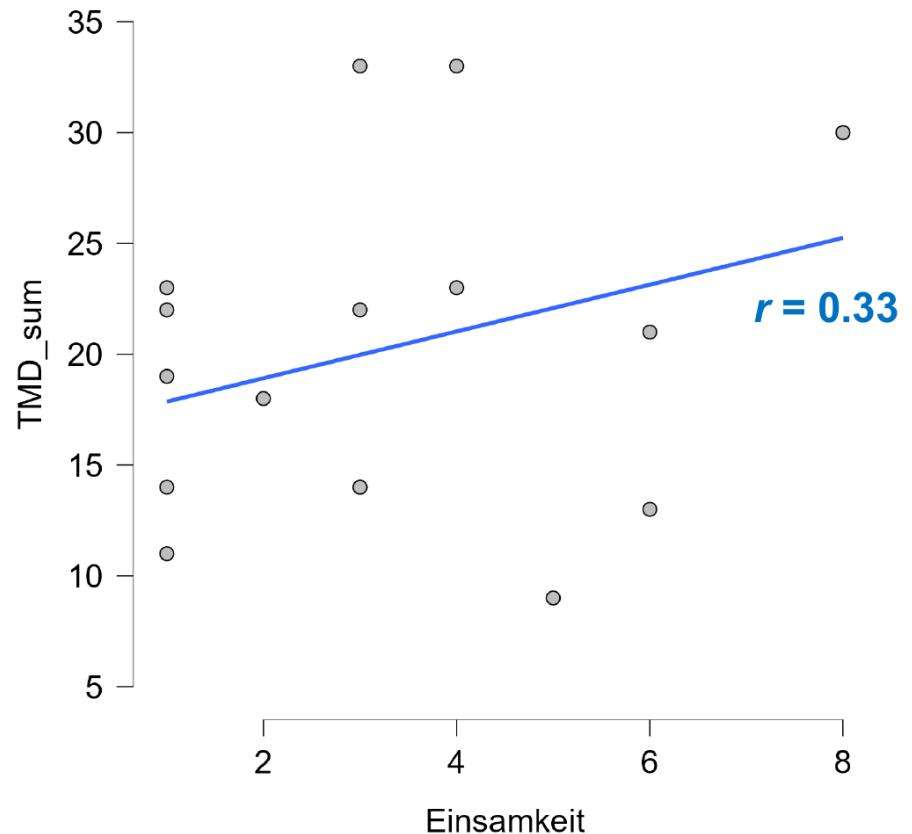
Ausblick: Korrelation

- Bislang haben wir Unterschiede in einer metrischen Variable *in Abhangigkeit von einer kategorialen Variable* betrachtet (vgl. Ubung 5).
- Korrelationen zeigen Zusammenhange zwischen *zwei metrischen Variablen*.
- Der Zusammenhang zwischen zwei metrischen Variablen lsst sich mit einem *Streudiagramm* abbilden.



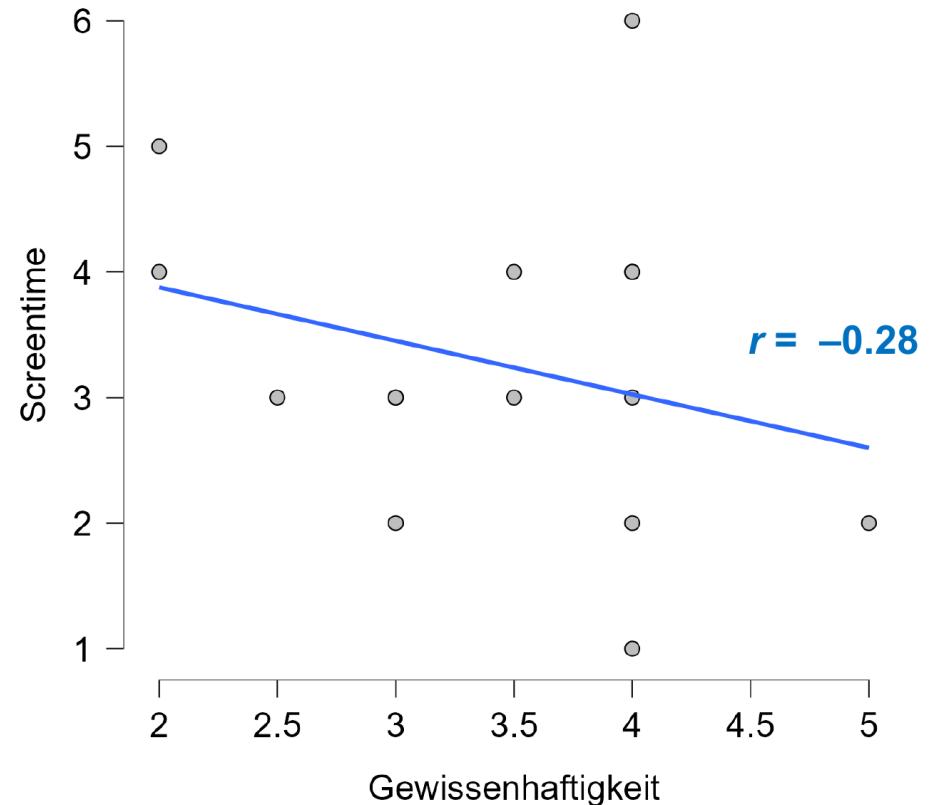
Ausblick: Korrelation

- Bislang haben wir Unterschiede in einer metrischen Variable *in Abhangigkeit von einer kategorialen Variable* betrachtet (vgl. Ubung 5).
- Korrelationen zeigen Zusammenhange zwischen zwei *metrischen Variablen*.
- Der Zusammenhang zwischen zwei metrischen Variablen lsst sich mit einem *Streudiagramm* abbilden.



Ausblick: Korrelation

- Bislang haben wir Unterschiede in einer metrischen Variable *in Abhangigkeit von einer kategorialen Variable* betrachtet (vgl. Ubung 5).
- Korrelationen zeigen Zusammenhange zwischen *zwei metrischen Variablen*.
- Der Zusammenhang zwischen zwei metrischen Variablen lsst sich mit einem *Streudiagramm* abbilden.

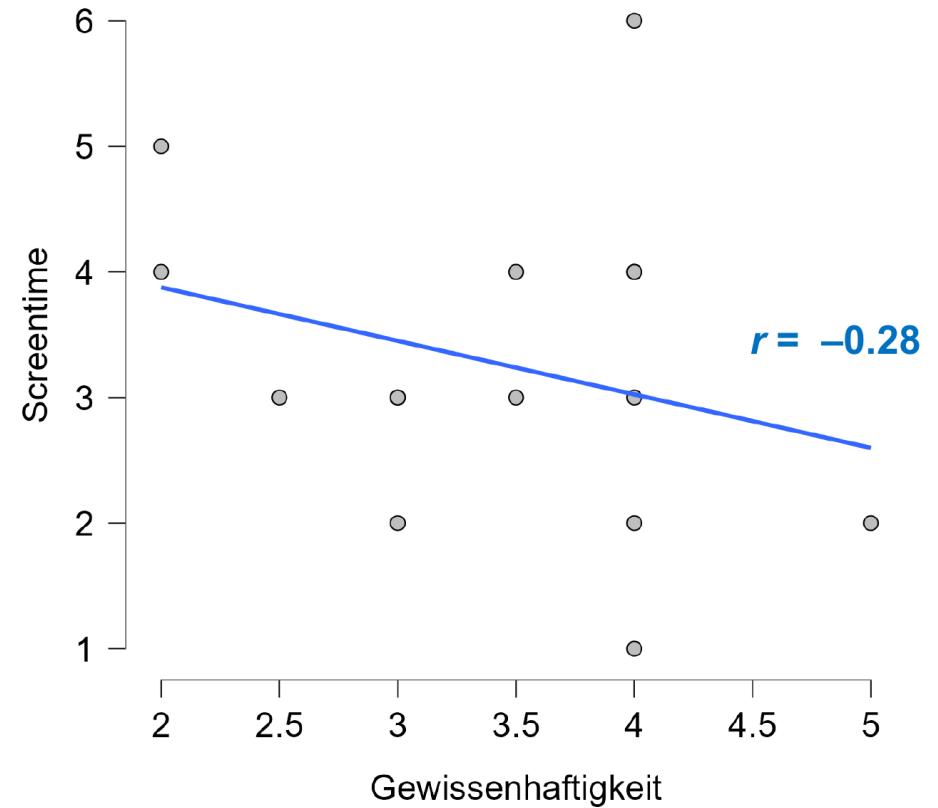


Korrelation

Der Zusammenhang zwischen zwei metrischen Variablen kann quantifiziert werden. Das gebräuchlichste Maß für den linearen Zusammenhang zweier Variablen ist der sog. **Pearson-Korrelationskoeffizient r** (auch Produkt-Moment-Korrelation genannt).

Wertebereich von r : -1 bis +1

Daumenregel (nach Cohen)	
.10	schwache Korrelation
.30	moderate Korrelation
über .50	starke Korrelation



Mathematische Grundlagen der Kovarianz und Korrelation

Varianz einer Variable

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Kovarianz zweier Variablen

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

Stichprobengröße

Abweichung vom Mittelwert
(Variable X)

Abweichung vom Mittelwert
(Variable Y)

Übungen

Berechnen Sie:

1. Den durchschnittlichen Eiskonsum und die durchschnittliche Temperatur über alle Länder (Mittelwerte und Median).
2. Die Varianzen beider Variablen.
3. Die Standardabweichungen beider Variablen.
4. Die Kovarianz beider Variablen.

Verwenden Sie die Formeln mit Bessel-Korrektur aus Ihrer Formelsammlung!

Country	X Ice Cream (kg)	Y Temp (°C)
Andorra	16.00	16.00
Belgium	12.00	8.00
Croatia	9.00	14.00
Denmark	6.00	10.00

Median

Mittelwert $\hat{\mu}$ 10.75

Varianz $\hat{\sigma}^2$

SD $\hat{\sigma}$

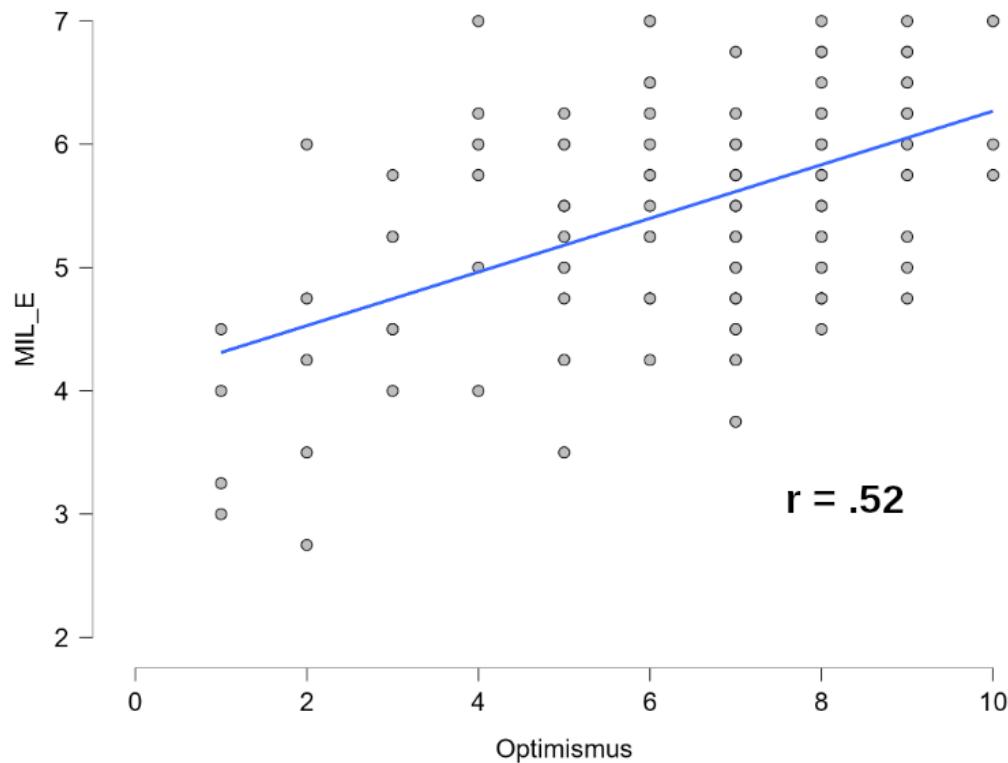
$\text{Cov}(X,Y) \hat{\sigma}_{x,y}$

Korrelation

1. Was gibt die Korrelation an?
2. Worauf weist das Vorzeichen der Korrelation hin?
3. Welcher Zusammenhang ist stärker: $r = -.48$ oder $r = .37$
4. Wie würden Sie die folgende Aussage interpretieren / einem Laien erklären: „*Wir fanden einen signifikanten Zusammenhang zwischen Prokrastination und Gewissenhaftigkeit, $r = -.48$* “
5. Wodurch unterscheidet sich die Kovarianz von der Korrelation?

Korrelation vs. Kausalität

Schlagzeile: “*Studierende in Potsdam fanden heraus, dass Menschen optimistischer denken sollten, um ihr Sinnerleben zu erhöhen.*“ Ist diese Aussage korrekt?



Zusammenhänge

Psychologen interessieren sich nicht nur für die Verteilung von Merkmalen (Variablen) in der Bevölkerung, sondern für Zusammenhänge zwischen diesen Merkmalen.

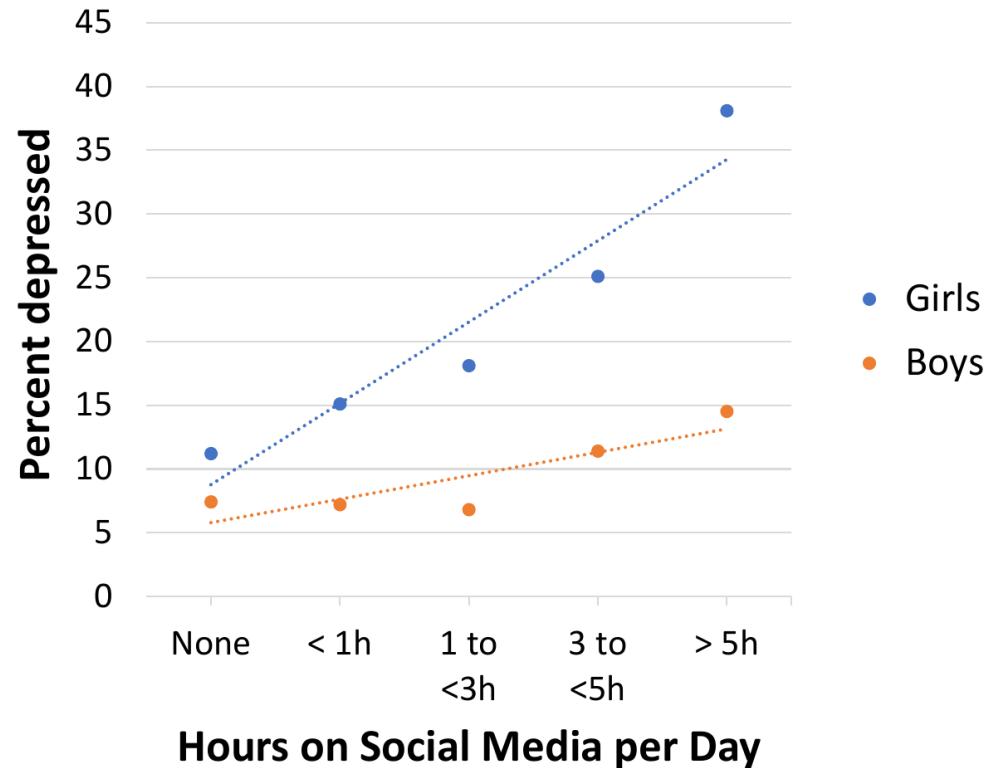
Das Vorliegen einer Korrelation beweist nicht, dass eine Variable die andere ursächlich beeinflusst.

Research Paper

Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study

Yvonne Kelly ^{a,*}, Afshin Zilanawala ^a, Cara Booker ^b, Amanda Sacker ^a

Daten aus Kelly et al. (2018)



Mathematische Grundlagen der Kovarianz und Korrelation

Varianz einer Variable

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Kovarianz zweier Variablen

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

↑
Stichprobengröße

Abweichung vom Mittelwert
(Variable X)

Abweichung vom Mittelwert
(Variable Y)

Übungen

Berechnen Sie:

1. Den durchschnittlichen Eiskonsum und die durchschnittliche Temperatur über alle Länder (Mittelwerte und Median).
2. Die Varianzen beider Variablen.
3. Die Standardabweichungen beider Variablen.
4. Die Kovarianz beider Variablen.

Verwenden Sie die Formeln mit Bessel-Korrektur aus Ihrer Formelsammlung!

Country	X Ice Cream (kg)	Y Temp (°C)
Andorra	16.00	16.00
Belgium	12.00	8.00
Croatia	9.00	14.00
Denmark	6.00	10.00

Median

Mittelwert $\hat{\mu}$ 10.75

Varianz $\hat{\sigma}^2$

SD $\hat{\sigma}$

$\text{Cov}(X,Y) \hat{\sigma}_{x,y}$

Übungen

Berechnen Sie:

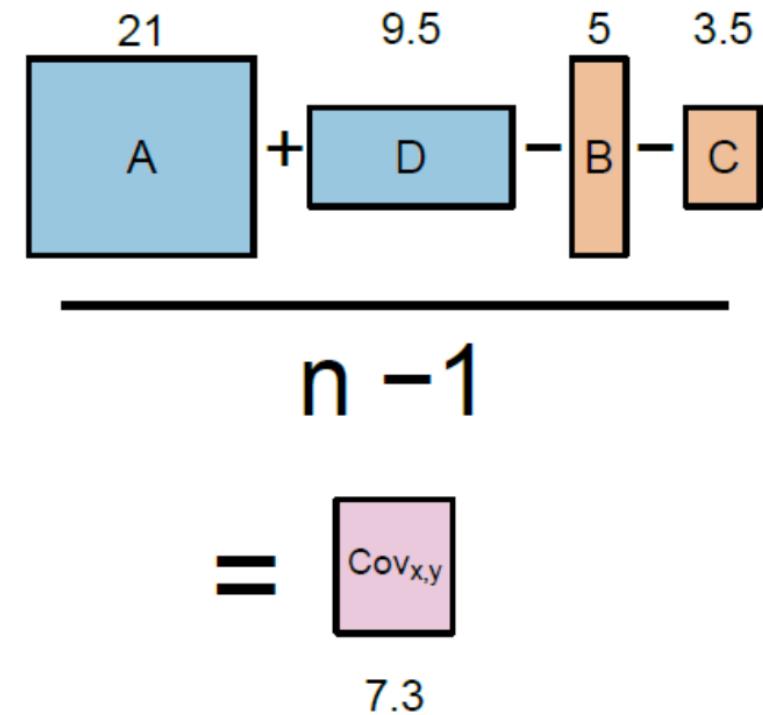
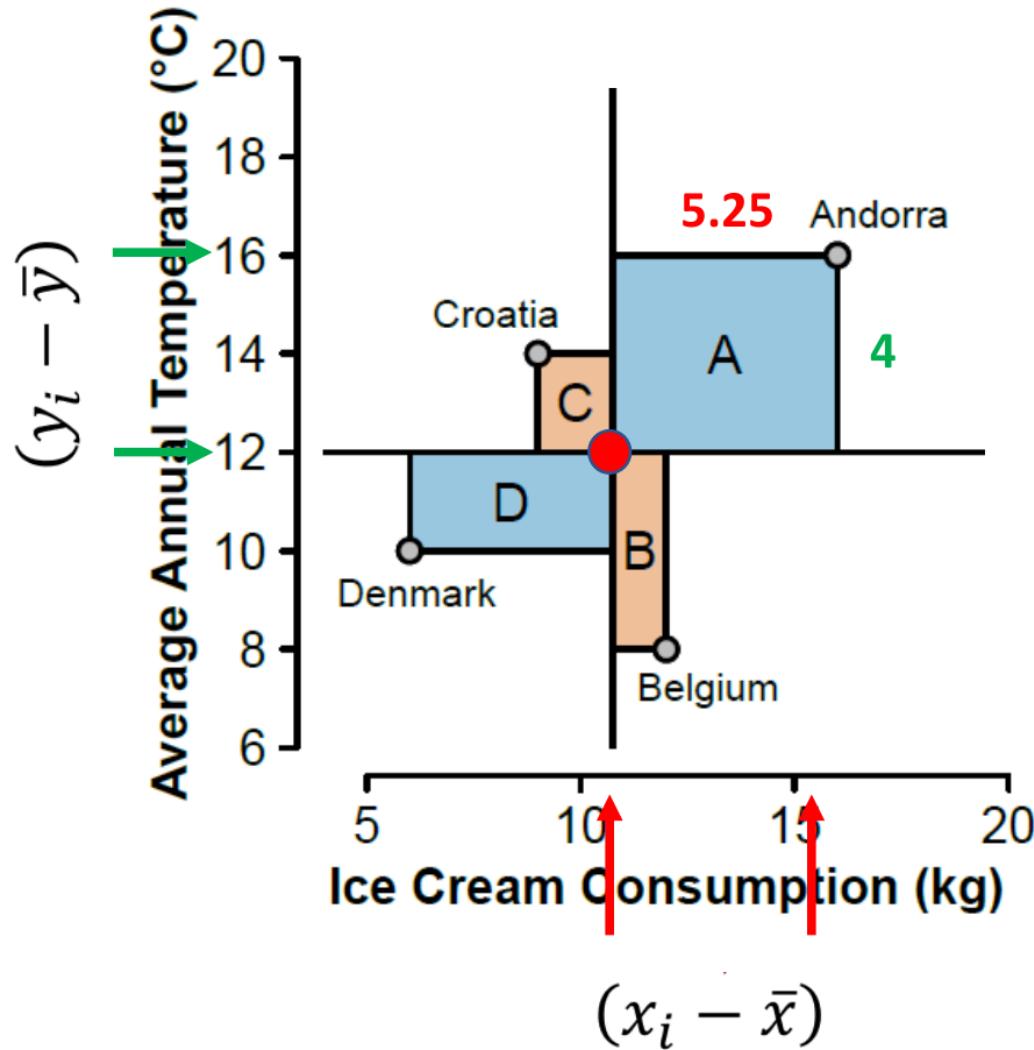
1. Den durchschnittlichen Eiskonsum und die durchschnittliche Temperatur über alle Länder (Mittelwerte und Median).
2. Die Varianzen beider Variablen.
3. Die Standardabweichungen beider Variablen.
4. Die Kovarianz beider Variablen.

Verwenden Sie die Formeln mit Bessel-Korrektur aus Ihrer Formelsammlung!

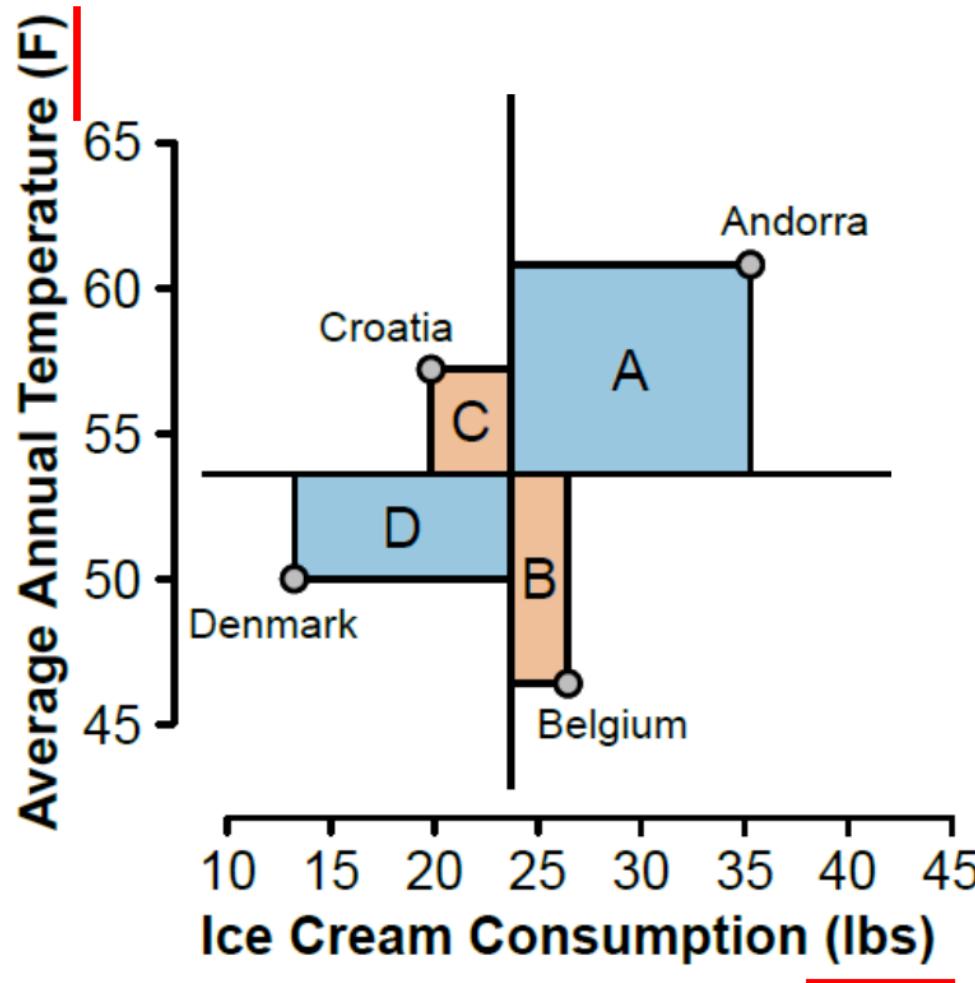
Country	X Ice Cream (kg)	Y Temp (°C)
Andorra	16.00	16.00
Belgium	12.00	8.00
Croatia	9.00	14.00
Denmark	6.00	10.00
Median	10.50	12.00
Mittelwert $\hat{\mu}$	10.75	12.00
Varianz $\hat{\sigma}^2$	18.25	13.33
SD $\hat{\sigma}$	4.27	3.65
$\text{Cov}(X,Y) \hat{\sigma}_{x,y}$		7.33

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

Grafische Darstellung der Kovarianz



Abhangigkeit der Kovarianz von den Einheiten der Skala



$$\begin{array}{ccccccc}
 & 83.3 & + & 37.7 & - & 19.8 & - 13.9 \\
 \boxed{A} & & & \boxed{D} & & \boxed{B} & - \boxed{C} \\
 & & & & & & \\
 & & & & & n - 1 & \\
 & & & & = & & \\
 & & & & \boxed{\text{Cov}_{x,y}} & & \\
 & & & & & & \\
 & & & & & 29.1 &
 \end{array}$$

Die Lösung: Korrelationskoeffizient $\hat{\rho}$

- Die Kovarianz muss in eine einheitslose Größe überführt werden, um sie vergleichbar zu machen.
- Dazu wird die Kovarianz mit dem Produkt der Standardabweichung beider Variablen „normalisiert“.
- Der Wertebereich von $\hat{\rho}$ ist immer: -1 bis $+1$

$$\hat{\rho} = \frac{\hat{Cov}(X, Y)}{\hat{\sigma}_X \hat{\sigma}_Y}$$

- Je näher der empirische Wert an $|1|$ liegt, desto stärker ist der lineare Zusammenhang beider Variablen.

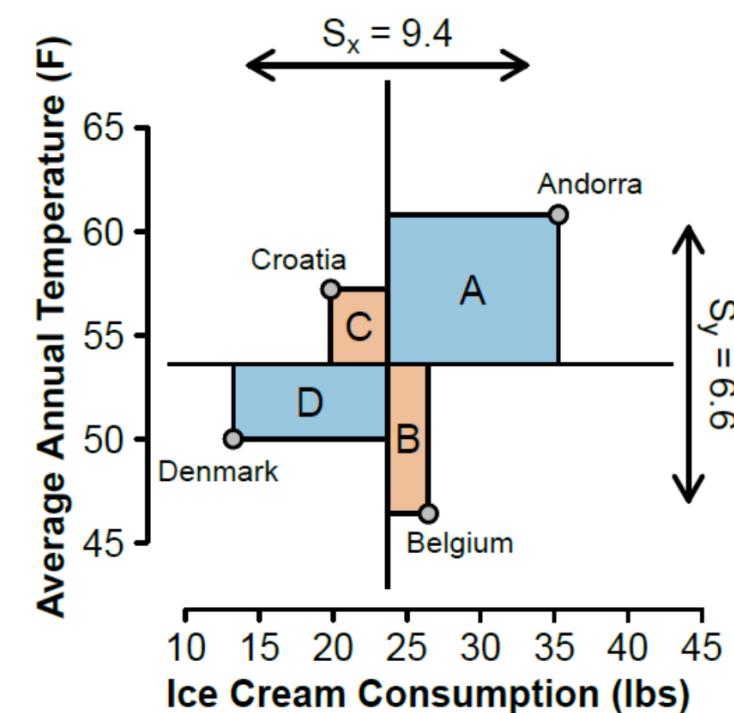
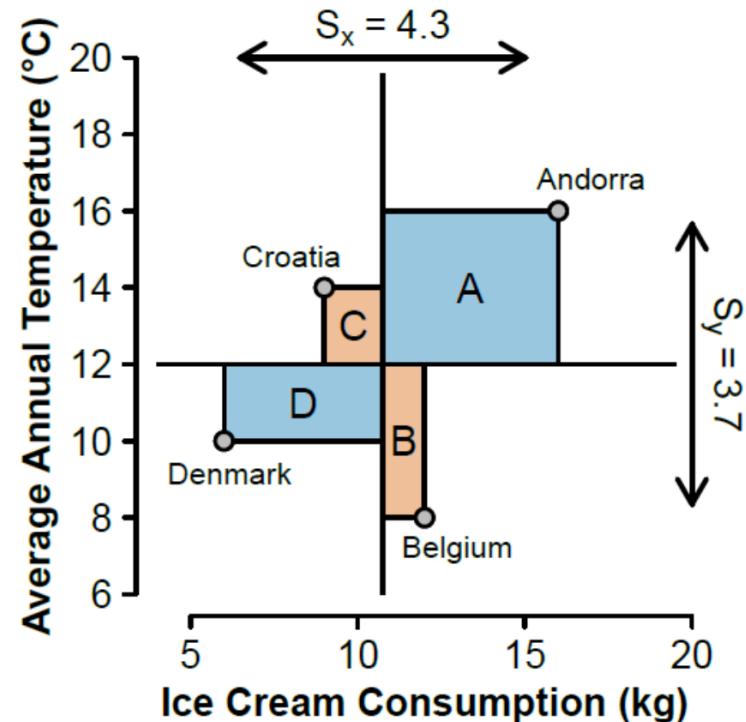
Übungen

Berechnen Sie:

- Die Korrelation zwischen Eis-Konsum und Temperatur (Pearson-Korrelationskoeffizient)

Country	X Ice Cream (kg)	Y Temp (°C)
Andorra	16.00	16.00
Belgium	12.00	8.00
Croatia	9.00	14.00
Denmark	6.00	10.00
Median	10.50	12.00
Mittelwert $\hat{\mu}$	10.75	12.00
Varianz $\hat{\sigma}^2$	18.25	13.33
SD $\hat{\sigma}$	4.27	3.65
Cov(X,Y) $\hat{\sigma}_{x,y}$		7.33

Berechnung des Korrelationskoeffizienten $\hat{\rho}$



$$\hat{\rho} = \frac{\hat{Cov}(X, Y)}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{7.3}{4.3 \cdot 3.7} = 0.47$$

$$\hat{\rho} = \frac{\hat{Cov}(X, Y)}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{29.1}{9.4 \cdot 6.6} = 0.47$$

JASP-Übungen Korrelation

1. Prüfen Sie die Zusammenhänge zwischen den 5 Persönlichkeitsmerkmalen des BFI
 - a. Lassen Sie sich dazu Streudiagramme ausgeben (Menü: Deskriptive Statistiken)
 - b. Welche Persönlichkeitsmerkmale hängen deutlich positiv miteinander zusammen, welche negativ?
 - c. Lassen Sie sich die Pearson-Korrelationen zwischen allen 5 Persönlichkeitsmerkmalen ausgeben (Menü: Regression -> Korrelation. Stimmen Ihre Annahmen aus Punkt b)?
2. Prüfen Sie, ob es einen Zusammenhang zwischen dem Persönlichkeitsmerkmal „Gewissenhaftigkeit“ und der Variable „Schwierigkeiten beim Stillsitzen“ gibt.
3. a) Formulieren Sie eine Zusammenhangshypothese bezogen auf zwei gemessene Variablen. b) Prüfen Sie Ihre Hypothese, indem Sie den Korrelationskoeffizienten berechnen. c) Prüfen Sie mittels Boxplots auf Ausreißer, filtern Sie die Ausreißer heraus und berechnen Sie erneut die Korrelation. (Siehe Codebook für gemessenen Variablen)

BFI-10

Kurzskala der „Big Five“ Persönlichkeitseigenschaften

Items				
Nr.	Deutsche Items	Englische Items	Polung	Subskalen
1	Ich bin eher zurückhaltend, reserviert.	I see myself as someone who is reserved.	-	Extraversion
2	Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	I see myself as someone who is generally trusting.	+	Verträglichkeit
3	Ich bin bequem, neige zur Faulheit.	I see myself as someone who tends to be lazy.	-	Gewissenhaftigkeit
4	Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	I see myself as someone who is relaxed, handles stress well.	-	Neurotizismus
5	Ich habe nur wenig künstlerisches Interesse.	I see myself as someone who has few artistic interests.	-	Offenheit
6	Ich gehe aus mir heraus, bin gesellig.	I see myself as someone who is outgoing, sociable.	+	Extraversion
7	Ich neige dazu, andere zu kritisieren.	I see myself as someone who tends to find fault with others.	-	Verträglichkeit
8	Ich erledige Aufgaben gründlich.	I see myself as someone who does a thorough job.	+	Gewissenhaftigkeit
9	Ich werde leicht nervös und unsicher.	I see myself as someone who gets nervous easily.	+	Neurotizismus
10	Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	I see myself as someone who has an active imagination.	+	Offenheit

Big Five Persönlichkeitsmerkmale

- **Offenheit: Kreativität, Neugierde und Interesse an neuen Erfahrungen und Ideen.**

- Berechnung:
$$\frac{(6-BFI_{05})+BFI_{10}}{2}$$

- **Gewissenhaftigkeit: Zuverlässigkeit und eine organisierte, sorgfältige Arbeitsweise.**

- Berechnung:
$$\frac{(6-BFI_{03})+BFI_{08}}{2}$$

- **Extraversion: Geselligkeit und das Bedürfnis nach sozialer Interaktion.**

- Berechnung:
$$\frac{(6-BFI_{01})+BFI_{06}}{2}$$

- **Verträglichkeit: Freundlichkeit und die Tendenz, kooperativ und hilfsbereit zu sein.**

- Berechnung:
$$\frac{(6-BFI_{07})+BFI_{02}}{2}$$

- **Neurotizismus: Anfälligkeit für Stress und emotionale Instabilität.**

- Berechnung:
$$\frac{(6-BFI_{04})+BFI_{09}}{2}$$

Guess the Correlation!

<https://istics.net/Correlations/>

Spearman Rangkorrelation

- Die Rangkorrelation erfasst, inwieweit zwei Rangreihen systematisch miteinander variieren.
- Für ordinalskalierte Variablen, die als Ränge vorliegen oder für Daten, die in Ränge überführt worden sind.
- Wertebereich ebenfalls -1 bis +1

$$\hat{\rho}_s = \frac{\hat{Cov}(R(X), R(Y))}{\hat{\sigma}_{R(X)} \hat{\sigma}_{R(Y)}}$$



Students	Maths	Science
A	35	24
B	20	35
C	49	39
D	44	48
E	35	45

Demonstration: Spearman Rangkorrelation

Rangkorrelation_Data (C:\Users\albert.anoschin\OneDrive - HMU Health and Medical University\Lehrvorbereitung)

	Student	Maths	Science	Maths_Rank	Science_Rank
1	A	35	24	J.	J.
2	B	20	35	J.	J.
3	C	49	39	J.	J.
4	D	44	48	J.	J.
5	E	35	45	J.	J.
6					

Kendall's Tau

- Wird u.a. zur Bestimmung der Interrater-Reliabilität zwischen zwei Beurteilern verwendet: „Wie groß ist die Übereinstimmung zwischen zwei Beurteilern?“

Wasserspringer	Beurteiler 1	Beurteiler 2
Springer 1	8	7
Springer 2	6	5
Springer 3	9	8
Springer 4	5	6

Die möglichen Paare von Wasserspringern sind:

(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)

Bestimmen der konkordanten und diskordanten Paare:

Springer 1 und Springer 2:

- Rater 1: 8 > 6
- Rater 2: 7 > 5
- Konkordant

Springer 1 und Springer 3:

- Rater 1: 8 < 9
- Rater 2: 7 < 8
- Konkordant

Anzahl der konkordanten Paare: 5

Anzahl der diskordanten Paare: 1

$$\Rightarrow \tau \rightarrow (5 - 1)/6 = 0.67$$