UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

KORRELATION

Ernste Kap. 2, Anh. A5.1-A5.2;
Ewing I: Kap. 9

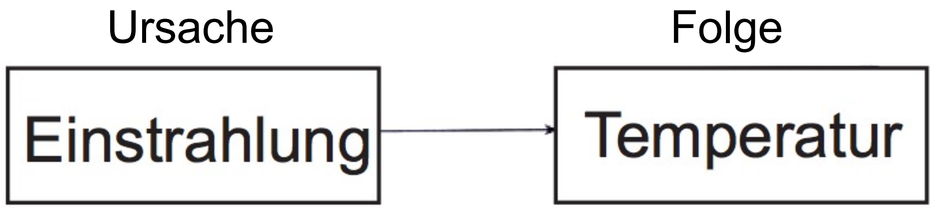
Statistische Datenanalyse

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

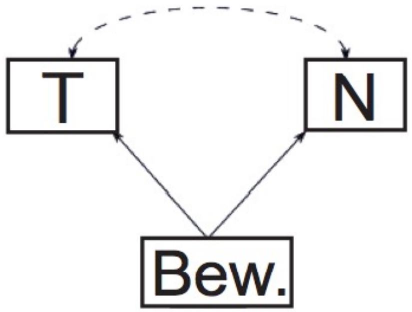
Deskriptive Statistik	Rohdaten visualisieren	Datenqualität prüfen	statistische Masszahlen
Schliessende Statistik	Unterschiede identifizieren	Zusammenhänge identifizieren	Abhängigkeiten modellieren
	Statistische Tests Konfidenzintervalle	Korrelation	Regression
	Wie wahrscheinlich sind die Daten der Stichprobe, wenn die Nullhypothese zutrifft?	Gibt es gemeinsame gleich- oder entgegengerichtete Variationen	Kausalzusammenhänge für Vorhersagen oder Interpolationen nutzen
Fallen der Statistik			
weiterführende Methoden	Daten zusammenfassen		Extremwertstatistik
	Hauptkomponentenanalyse	Clusteranalyse	Zeitreihenanal. etc.

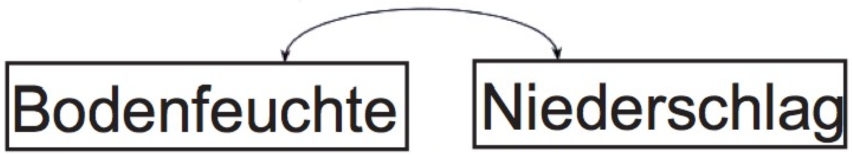
Beispiele von Beziehungen

Einfache Kausalbeziehung

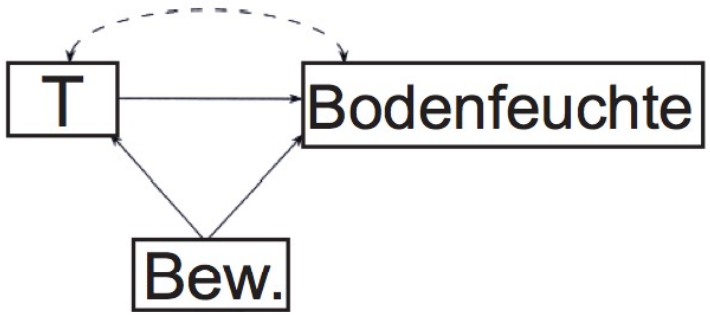


Scheinrelation





Kovariation

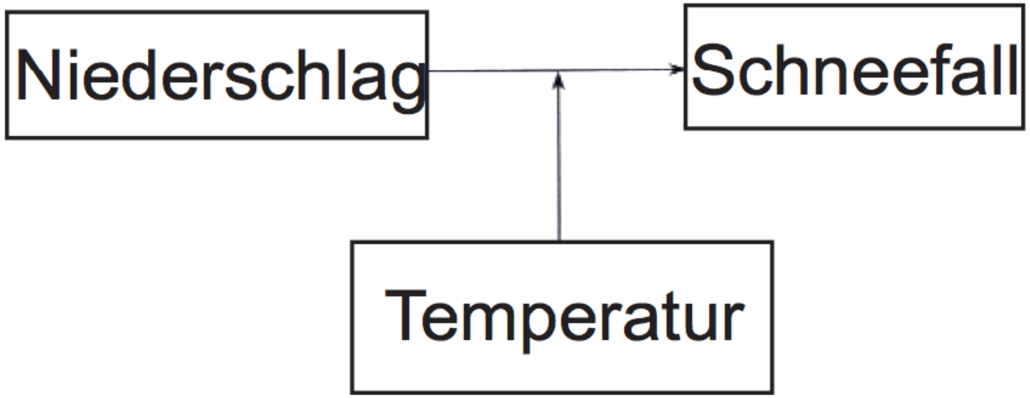


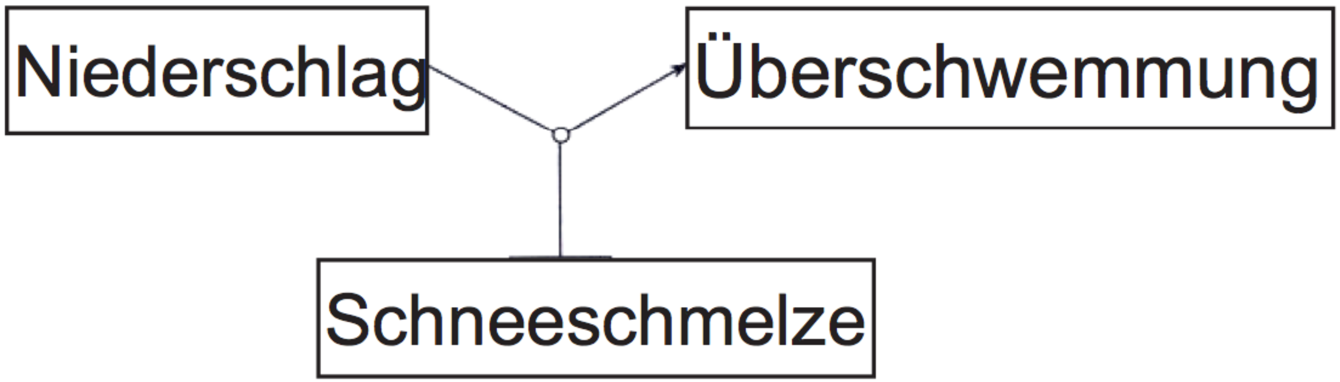
Partielle Scheinrelation

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Beispiel für konditionale Variablen





Korrelationsanalyse

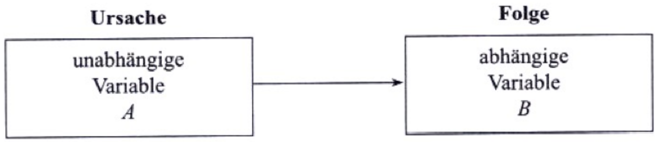


Abbildung 1.3: Pfaddiagramm einer einfachen kausalen Beziehung

Abbildung 1.4: Pfaddiagramm einer Kovariation

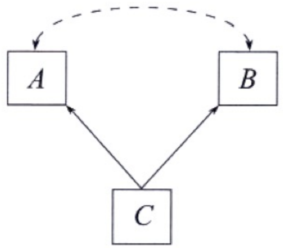


Abbildung 1.5: Pfaddiagramm einer Scheinrelation

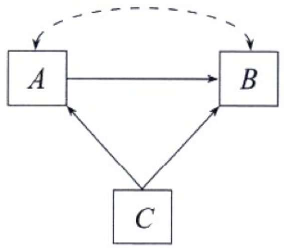


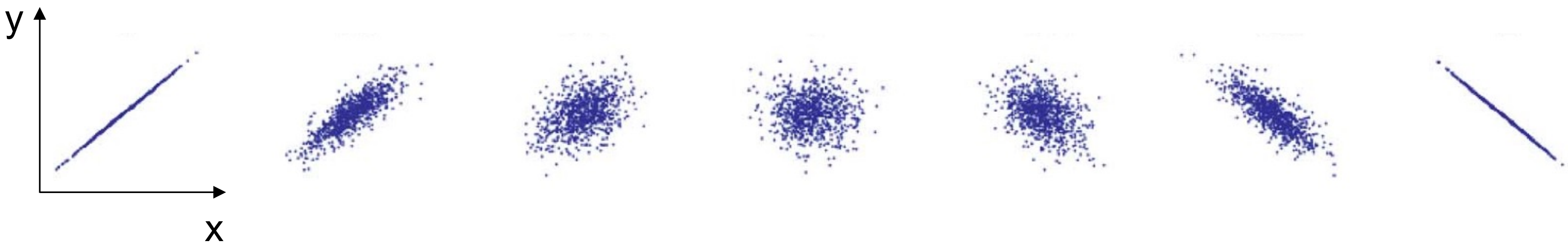
Abbildung 1.6: Pfaddiagramm einer partiellen Scheinrelation

- Masszahl für die Stärke des **linearen** Zusammenhangs **metrischer** Variablen
- sagt nichts über die Kausalität und Richtung des Zusammenhangs aus

Streudiagramme oder Scatterplots

UNIVERSITÄT
BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Zusammenhänge zwischen zwei Variablen



Korrelationskoeffizient?

Pearson-Korrelation

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Wie messen wir den Zusammenhang zweier Variablen?

1. Idee: Das Produkt der Anomalien

$$X_i^d = X_i - \mu_x \;;\; Y_i^d = Y_i - \mu_y$$

$$\sum_{i=1}^{N} (X_i - \mu_x)\left(Y_i - \mu_y\right) = \sum_{i=1}^{N} X_i^d \, Y_i^d$$

Ist von Stichprobengrösse (n) abhängig

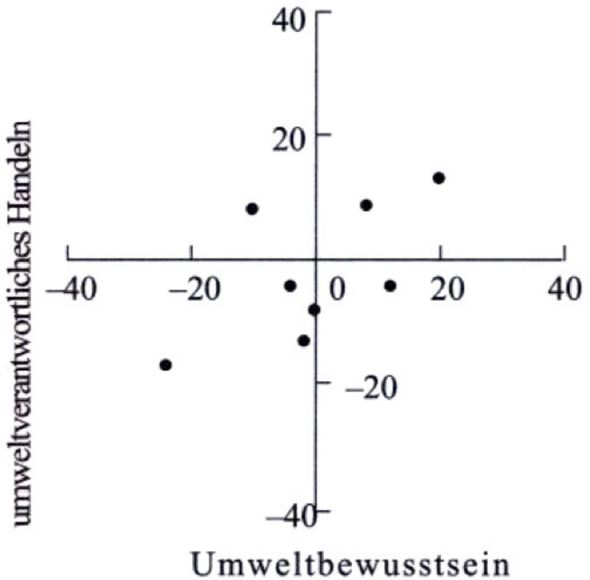


Abbildung 2.9: Transformation der ursprünglichen Werte zu Abweichungen vom Mittelwert

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Pearson-Korrelation

2. Division durch Stichprobengrösse

Kovarianz zwischen Y und X

$$\boxed{\sigma_{xy} = \frac{\sum_{i=1}^{N}(X_i - \mu_x)\left(Y_i - \mu_y\right)}{N}}$$

Ist von den Einheiten abhängig

Pearson-Korrelation

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

3. Standardisierung

$$\rho_{xy} = \frac{\sum_{i=1}^{N} \frac{(X_i - \mu_x)}{\sigma_x} \frac{(Y_i - \mu_y)}{\sigma_y}}{N}$$

mit Standardabweichungen von X und Y

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu_x)^2}{N}} \quad und$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \mu_y)^2}{N}}$$

Korrelationskoeffizient (Pearson)

$$\boxed{\rho_{xy} = \frac{\sum_{i=1}^{N} (X_i - \mu_x)(Y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N} (X_i - \mu_x)^2 \sum_{i=1}^{N} (Y_i - \mu_y)^2}}}$$

Prüfung!

Pearson-Korrelation

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

> Der Korrelationskoeffizient drückt das Verhältnis der beobachteten Kovarianz zur maximalen Kovarianz aus bzw.

> Der Korrelationskoeffizient ist die bezüglich der Standardabweichung von X und Y normierte Kovarianz.

Korrelationskoeffizient (Pearson)

$$\rho_{xy} = \frac{\sum_{i=1}^{N}(X_i - \mu_x)\left(Y_i - \mu_y\right)}{\sqrt{\sum_{i=1}^{N}(X_i - \mu_x)^2 \sum_{i=1}^{N}\left(Y_i - \mu_y\right)^2}}$$

> ρ (rho) ist der Standardbuchstabe für den Korrelationskoeffizienten der Grundgesamtheit

> r für den Korrelationskoeffizienten der Stichprobe

Ausreisser

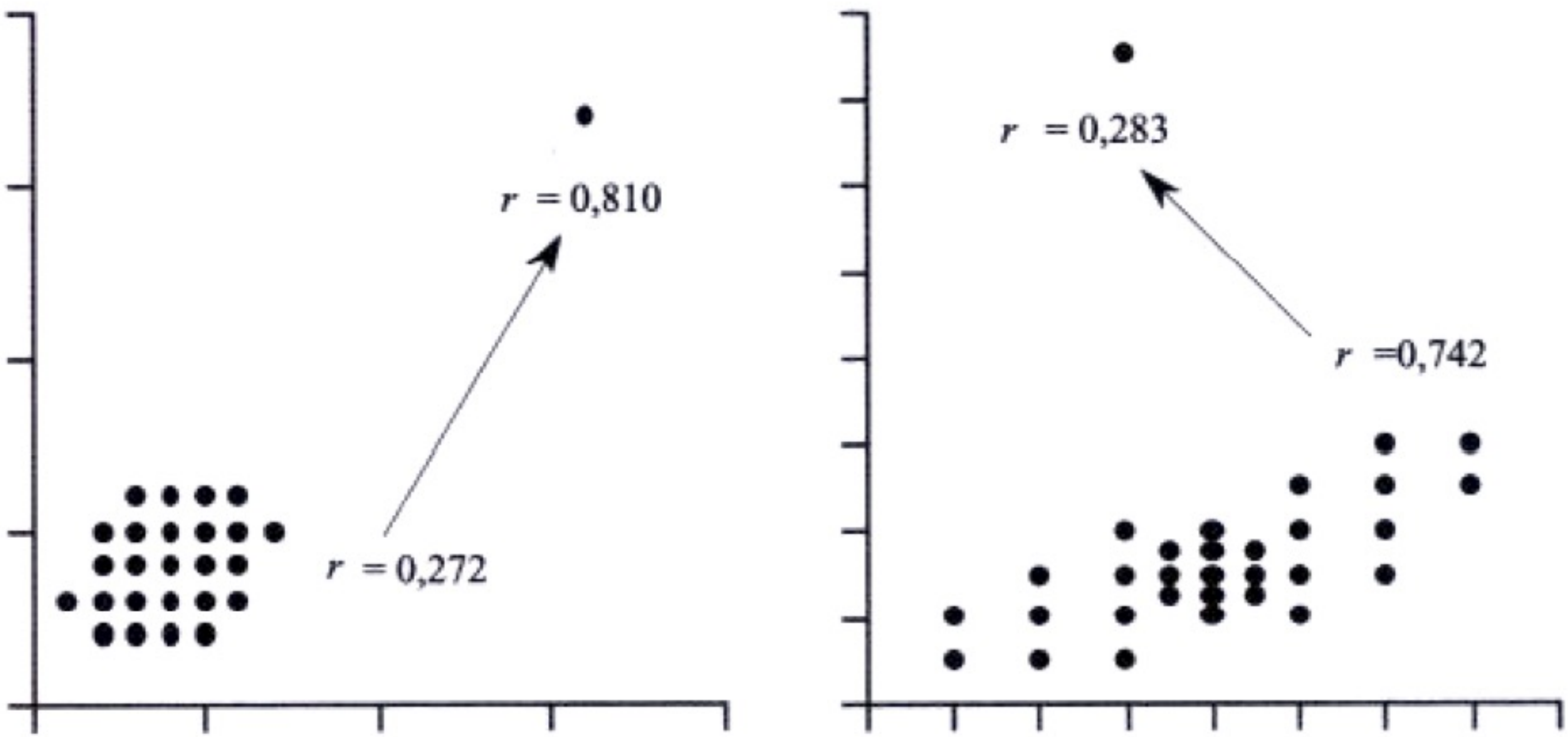


Abbildung 2.11: Einfluss eines Ausreissers auf den Korrelationskoeffizienten

Spearman Rangkorrelation

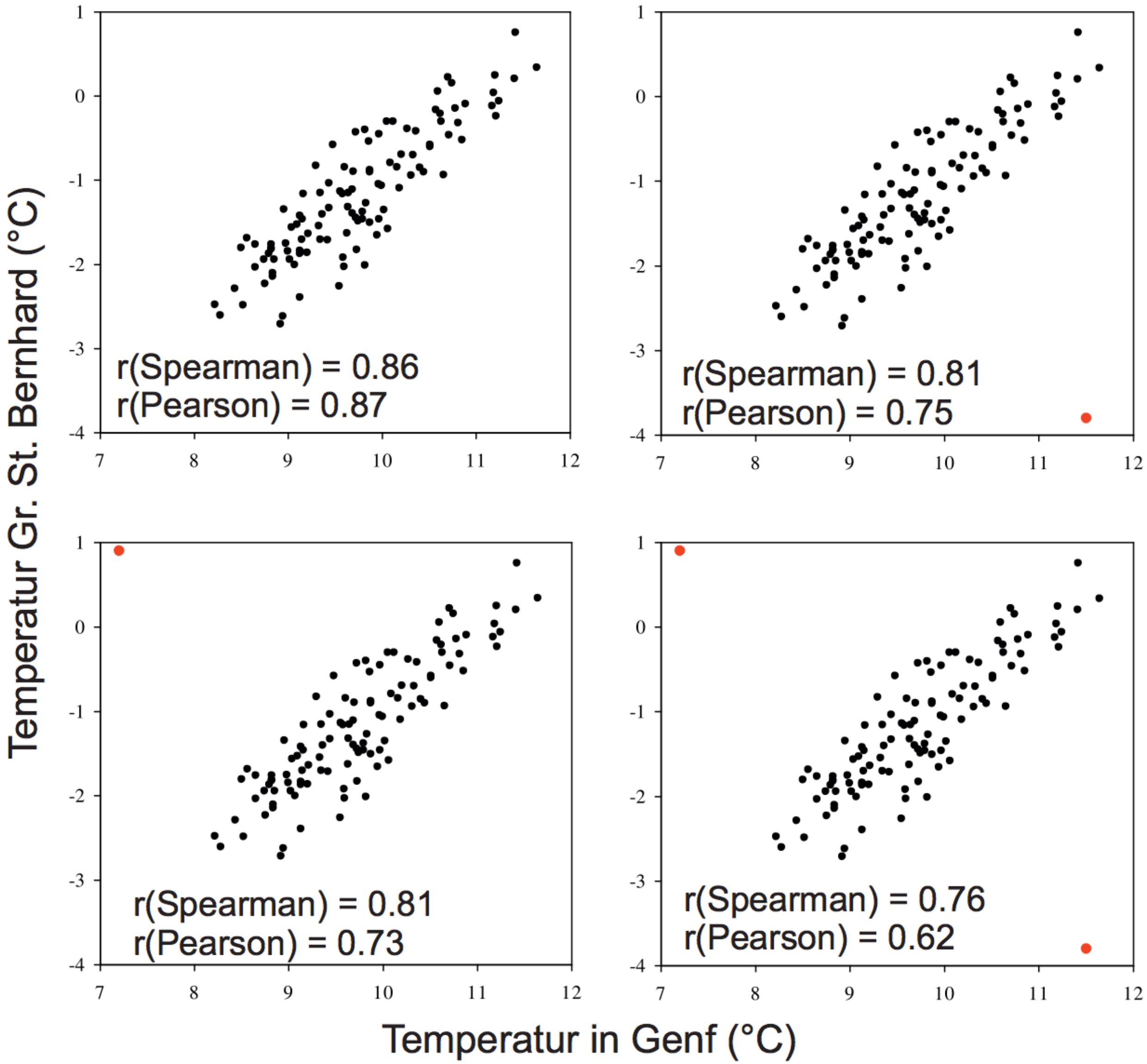
- Für ordinal skalierte Variablen kann der Rang eines Objekts (Beobachtung) in zwei Variablen verwendet werden:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n} (r_i - s_i)^2}{n^3 - n}$$

(vereinfachte Formel, wenn jeder Rank nur einmal vergeben ist)

- r_i = Rang von Objekt i in der Variablen R, s_i = Rang von Objekt i in der Variablen S.
- Der Spearman Rangkorrelationskoeffizient wird sehr oft auch für metrische Variablen verwendet, da er robust ist gegenüber Ausreissern. Im Zweifelsfall ist der Spearman-Koeffizient dem Pearson-Koeffizient vorzuziehen.

Spearman Rangkorrelation: **Ausreisser**



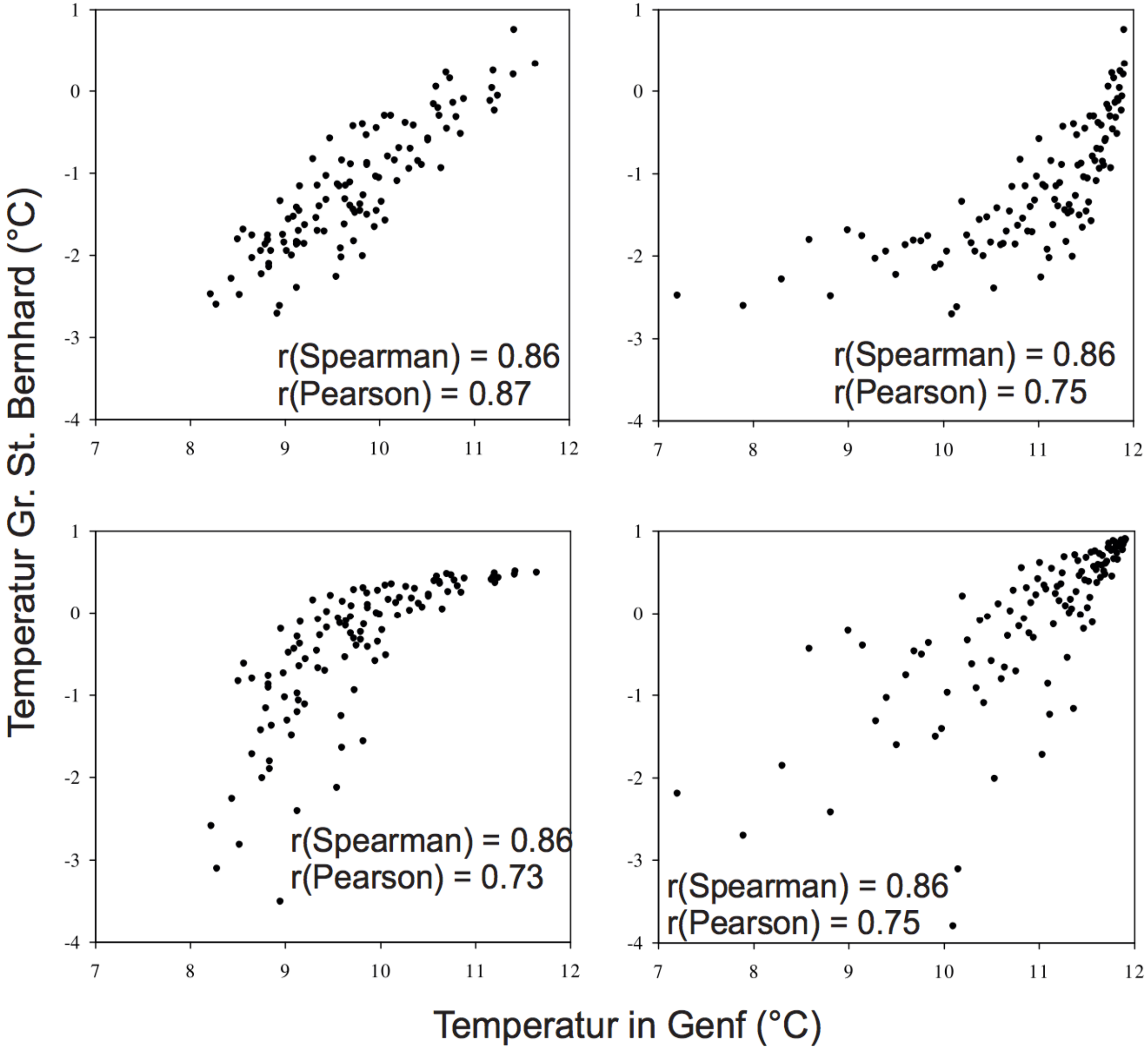
UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Spearman Rangkorrelation: **Nicht-Linearität**

b
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH



Transformation der Achsen

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Pearson-Korrelation: Misst Stärke des linearen Zusammenhangs, ist nicht sensitiv gegenüber linearen Transformationen (z.B. $y = a + bx$). Bei klarer (oder theoretisch begründeter) Nicht-linearität ist eine Transformation möglich.
- Spearman: Misst Stärke eines monotonen Zusammenhangs, ist nicht sensitiv gegenüber monotonen Transformationen (z.B. $y = \ln(x)$)

Wann machen Transformationen Sinn?

- Monotone Transformation nicht unbedingt (einfach Spearman statt Pearson verwenden)
- Nicht-monotone Transformationen (z.B. $y = x^2$) können Sinn machen.
- Alle Transformationen machen Sinn bei der Regression

Voraussetzungen für die Pearson-Korrelation

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- metrische Daten
- beide Variablen annähernd normalverteilt
- linearer Zusammenhang zwischen x und y
- statistischer Zusammenhang zwischen x und y nur, wenn der ermittelte Korrelationskoeffizient signifikant von Null abweicht (t-Test)

Rückschlüsse auf die Grundgesamtheit

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Hypothesen:

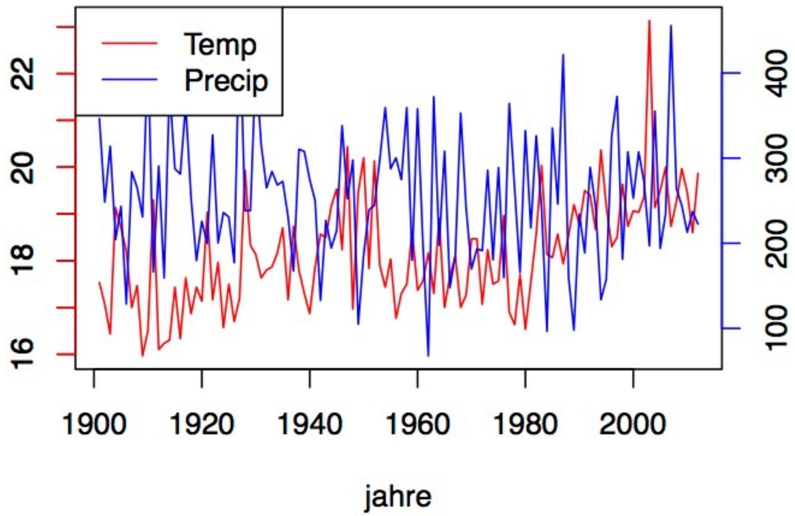
- H_0: x und y sind unkorreliert $\rho_{xy} = 0$
- H_A: x und y sind korreliert $\rho_{xy} \neq 0$
- Annahme: Verteilung von x und y in der Grundgesamtheit ist eine bivariate Normalverteilung
- Teststatistik: t-Test mit n-2 Freiheitsgraden

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

Korrelation von Zeitreihen

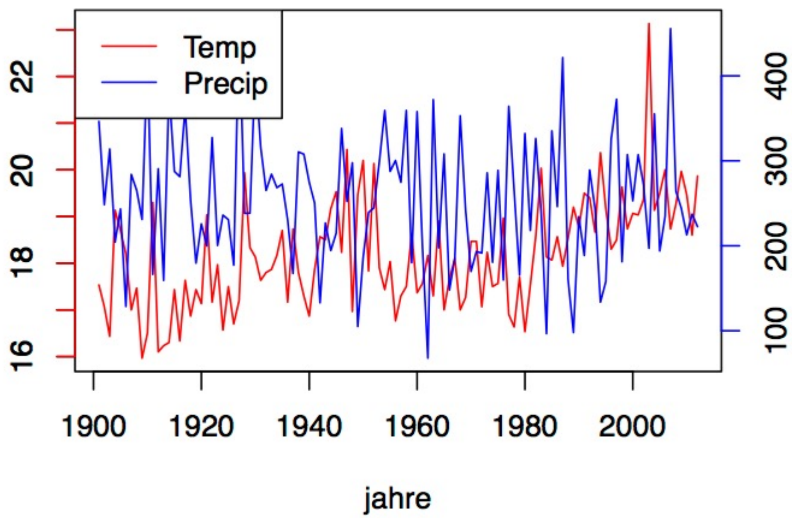
UNIVERSITÄT
BERN

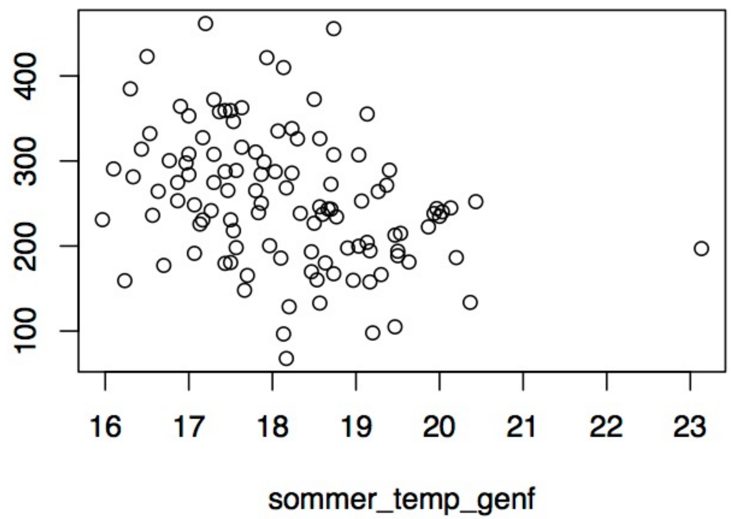
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH



Korrelationskoeffizient?

Korrelation von Zeitreihen

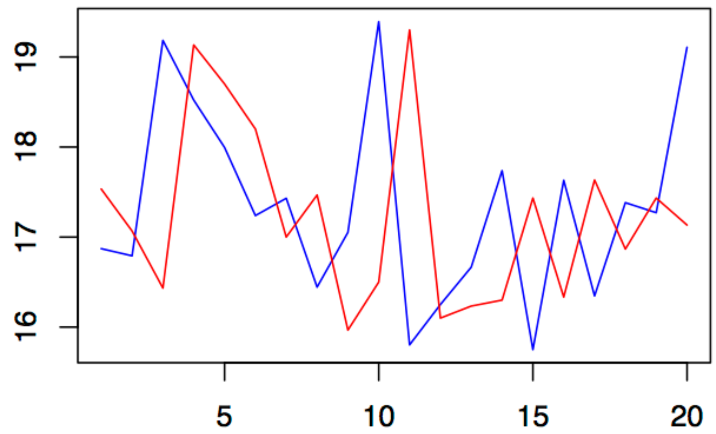


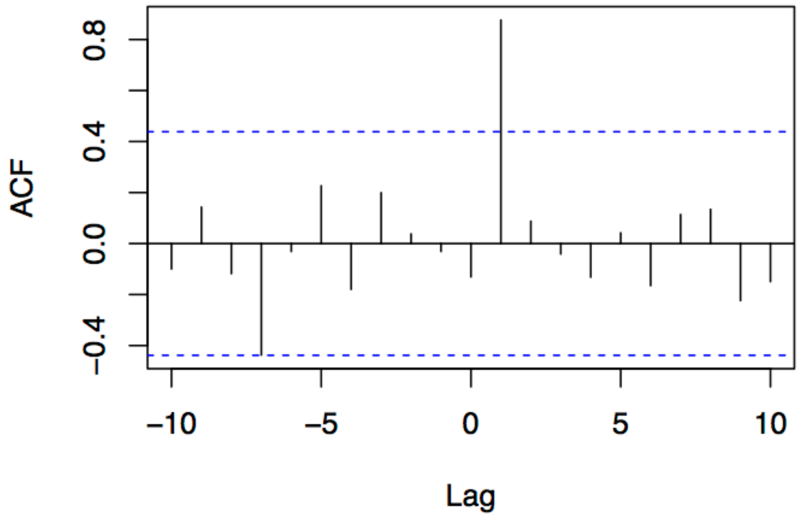


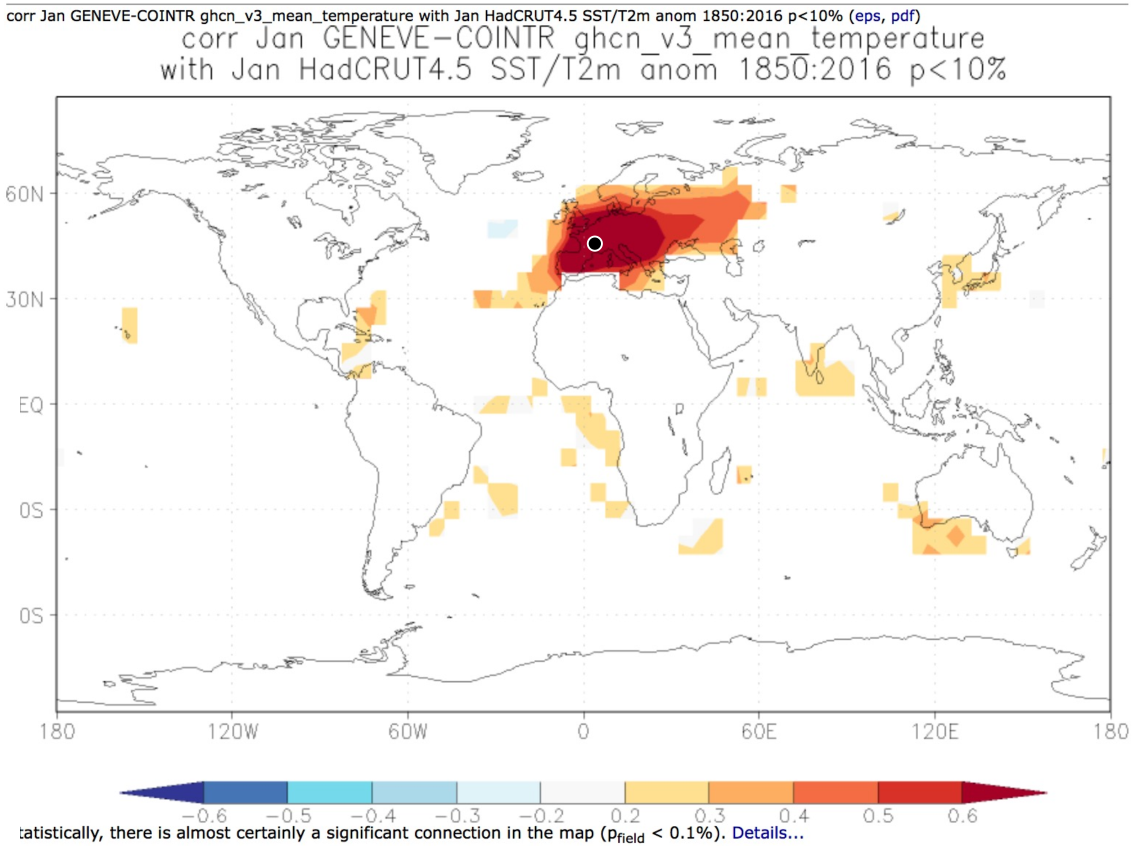

```
cor.test(sommer_temp_genf,sommer_precip_genf)

Pearson's product-moment correlation
data:  sommer_temp_genf and sommer_precip_genf
t = -3.7533, df = 110, p-value = 0.0002805
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval: -0.4917509 -0.1614801
sample estimates: cor -0.33694
```

Autokorrelation
in Zeit und Raum

UNIVERSITÄT
BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH







Varianz-Kovarianzmatrix

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

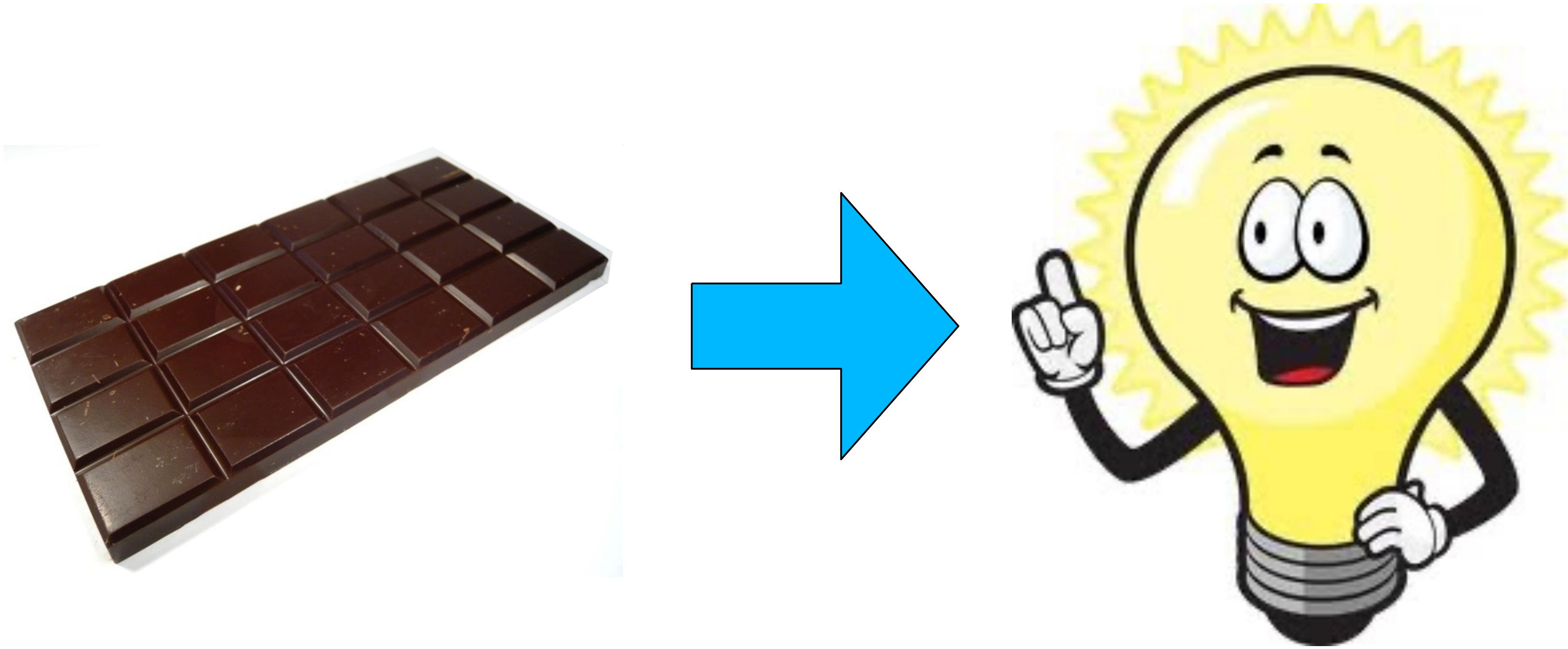
> Symmetrische Matrix mit den Varianzen in der Diagonalen und den Kovarianzen abseits der Diagonalen

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix};$$

```
R Beispiel:
A <- matrix(c(2,3,1,-1,1,1,0,4,2,-1,0,0),nrow=4,ncol=3, byrow=T)
cor(A) # oder ,method="pearson", Pearson ist Standardeinstellung
cor(A,method="spearman")
```

Korrelation

> “Dietary flavonoids ... which are widely present in cocoa ... have been shown to improve cognitive function. ... I wondered if there is a correlation between a country’s chocolate consumption and its population’s cognitive function ...” Messerli (2012)

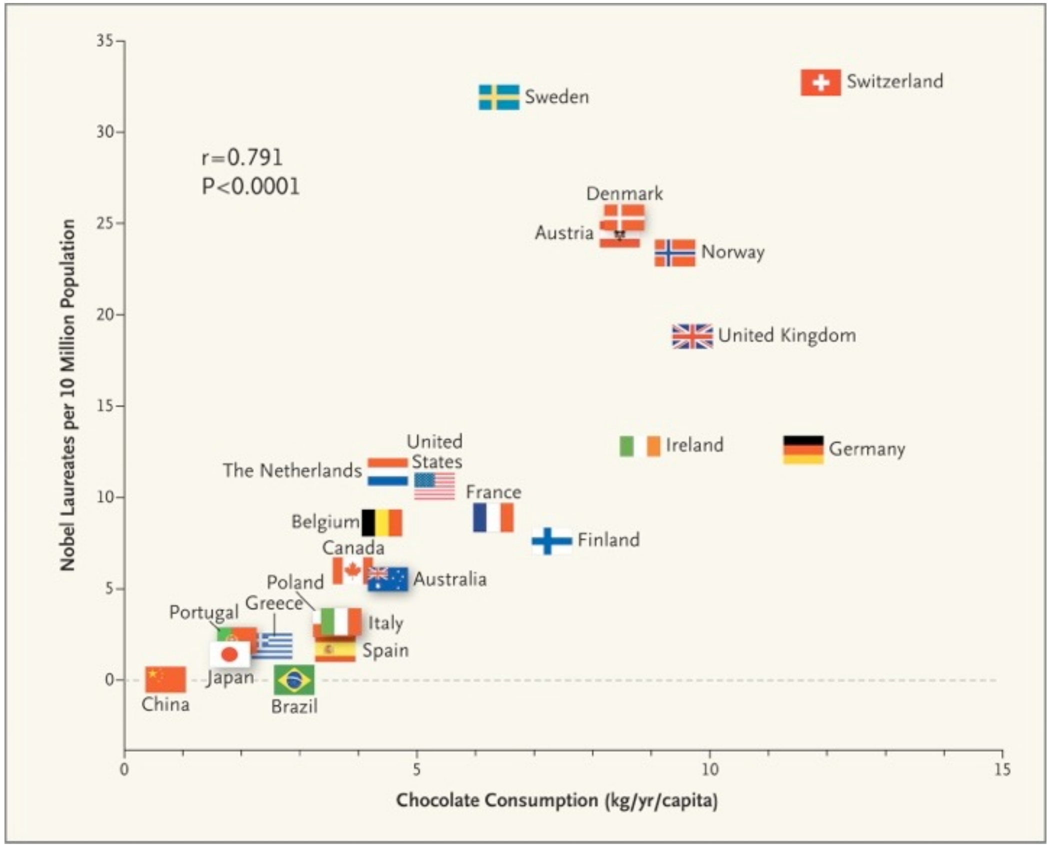


Korrelation

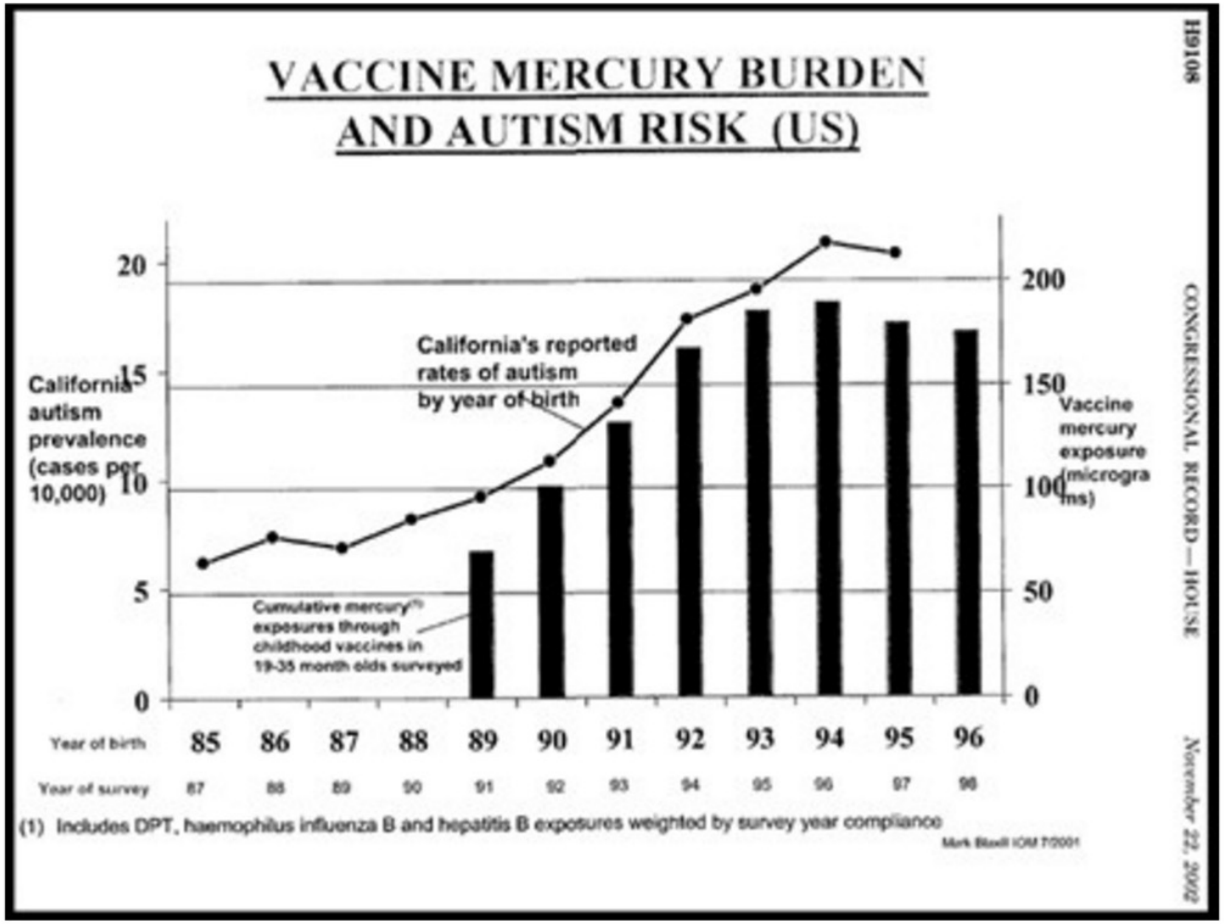
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

> "Dietary flavonoids ... which are widely present in cocoa ... have been shown to improve cognitive function. ... I wondered if there is a correlation between a country's chocolate consumption and its population's cognitive function ..." Messerli (2012)



r = 0.79

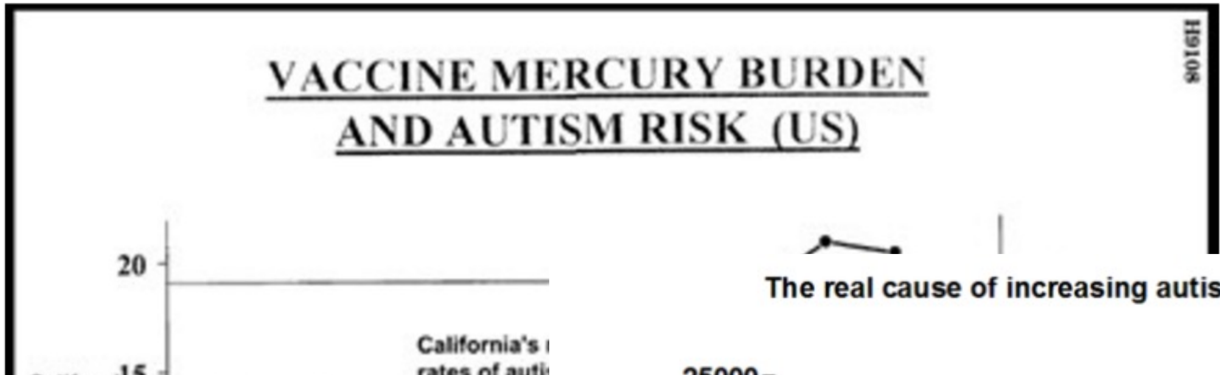

VACCINE MERCURY BURDEN
AND AUTISM RISK (US)
California autism prevalence (cases per 10,000)
Vaccine mercury exposure (micrograms)
California's reported rates of autism by year of birth
Cumulative mercury(1) exposures through childhood vaccines in 19-35 month olds surveyed
Year of birth
85 86 87 88 89 90 91 92 93 94 95 96
Year of survey
87 88 89 90 91 92 93 94 95 96 97 98
(1) Includes DPT, haemophilus influenza B and hepatitis B exposures weighted by survey year compliance
H9108
CONGRESSIONAL RECORD — HOUSE
November 22, 2002

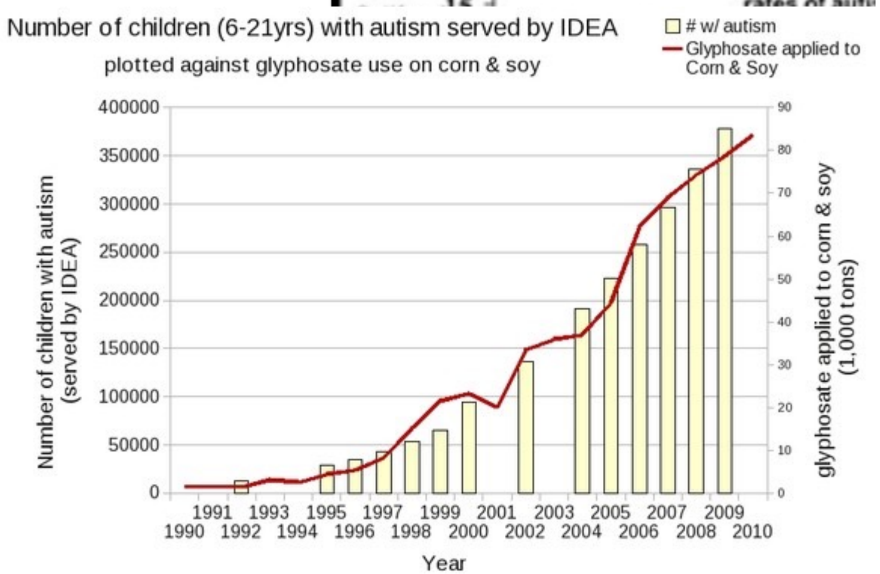
Korrelation ≠ Kausalzusammenhang

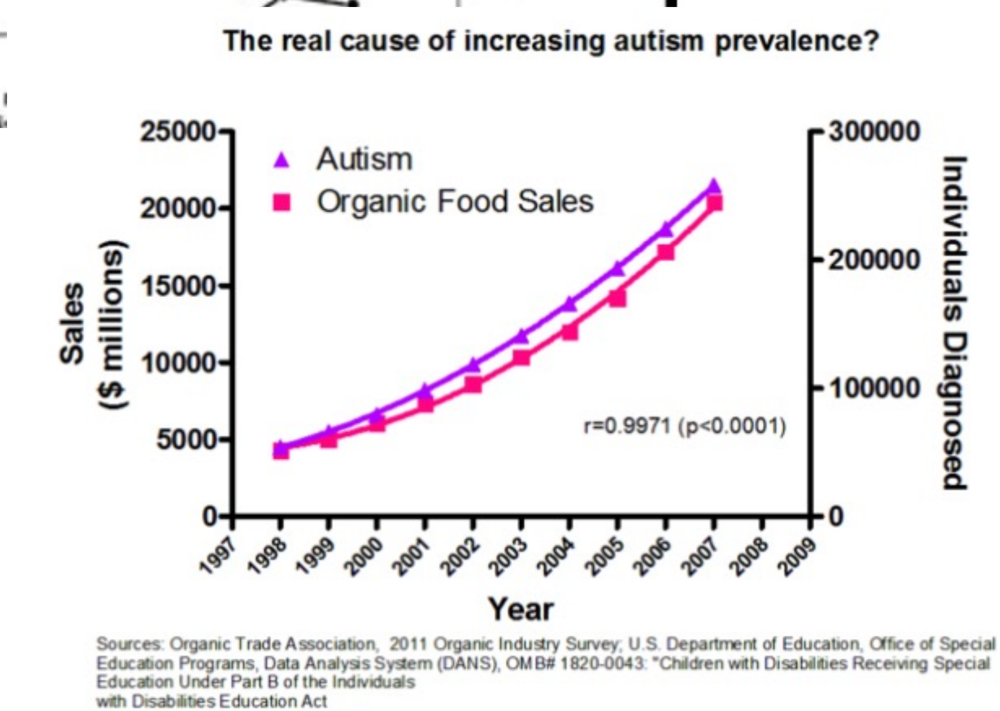
UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

> Korrelation beschreibt nur die Stärke des Zusammenhangs, KEINE Information über Kausalzusammenhänge







Take-home Message Korrelation

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Korrelation beschreibt nur die Stärke des Zusammenhangs, KEINE Information über Kausalzusammenhänge
- Pearson Korrelation nur für metrische Daten ohne Ausreisser und linearen Zusammenhang zwischen den beiden Variablen
- Spearman Rangkorrelation für metrische und kategoriale Daten, robuster gegenüber Ausreissern und Nicht-Linearität
- +1 heisst perfekter positiver Zusammenhang; 0 heisst kein Zusammenhang; -1 perfekter negativer Zusammenhang
- zyklische Schwankungen und Trends können Korrelationskoeffizienten stark beeinflussen