





Sandra Hansen-Morath Sascha Wolfer



STATISTIK MIT R

Korrelation





JE... DESTO...

Korrelationsanalyse:

Feststellen von Richtung und Stärke eines Zusammenhanges zwischen zwei oder mehr Variablen. Man kann jedoch nicht eine kausale Beziehung beweisen oder im Sinne von Ursache-Wirkung erklären.

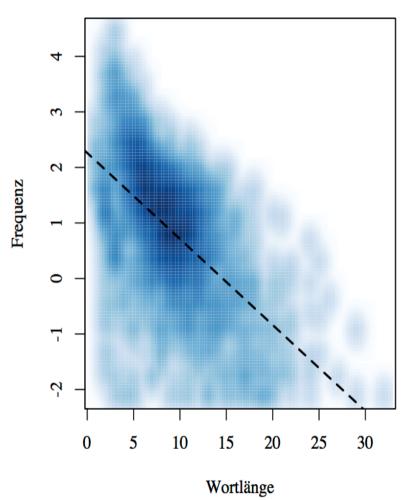
Korrelationskoeffizient:

Die Stärke der linearen Assoziation der Variablen wird durch einen Korrelationskoeffizienten r festgestellt.



JE... DESTO... KORRELATIONEN SIND ÜBERALI

- Je schneller man etwas tut, desto ungenauer tut man es (speed-accuracy trade-off).
- Je mehr Hubraum ein Auto hat, desto mehr PS hat es.
- Je mehr Feuerwehrleute vor Ort, desto höher die Schadenssumme.
- Je häufiger ein Wort ist, desto kürzer ist es (Zipf's law).





- Die wichtigste Maßzahl der Korrelation ist der Korrelationskoeffizient r nach Pearson.
- Korrelationen zwischen zwei intervallskalierten (kontinuierlichen) Variablen werden zwischen -1 und +1 angegeben. Dabei bedeutet:
 - +1: positiver Zusammenhang (Je mehr X, desto mehr Y)
 - -1: negativer Zusammenhang (Je mehr X, desto weniger Y)
 - 0: kein Zusammenhang



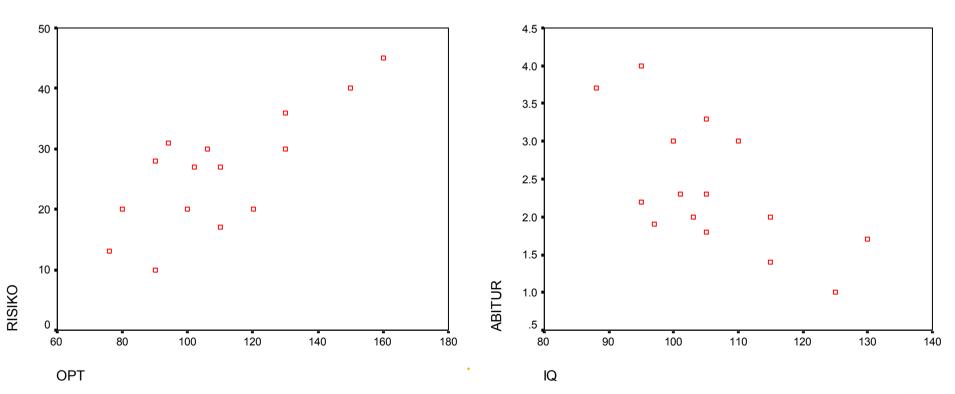
- Die Korrelation ist ein Maß für den (linearen) Zusammenhang zwischen zwei Variablen.
- Eine *positive Korrelation* ist dann gegeben, wenn ein hoher Wert auf einer Variable häufig mit einen hohen Wert auf der anderen Variable einhergeht (z.B. Optimismus und Risikobereitschaft).
- Eine *negative Korrelation* ist dann gegeben, wenn ein hoher Wert auf einer Variable häufig mit einen niedrigen Wert auf der anderen Variable einhergeht (z.B. Optimismus und Ängstlichkeit oder Intelligenz und Abi-Note).



Grafisch kann man Zusammenhänge zwischen zwei Variablen gut in einem Scatterplot darstellen:

positiver Zusammenhang

negativer Zusammenhang





- Folgende Schwellen gelten als anerkannt:
 - ab r = 0,1: schwache Korrelation
 - ab r = 0,3: mittlere Korrelation
 - ab r = 0,5: starke Korrelation

Achtung:

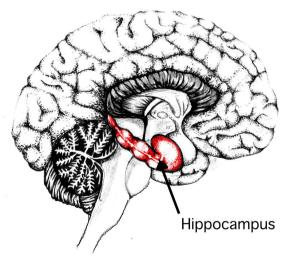
Auch starke Korrelationen müssen auf Signifikanz (Verallgemeinerbarkeit auf die Grundgesamtheit) geprüft werden! Auch recht kleine Korrelationen können hochsignifikant (p < 0,001) sein.



VON TAXIFAHRERN UND HIPPOCAMPI

- Wissenschaftlicher haben herausgefunden: Taxifahren verbessert den Orientierungssinn!
- Datenlage: Anzahl der Dienstjahre als Taxifahrer in London korreliert positiv mit der Größe eines Teils des Hippocampus (zuständig u.a. für Orientierung).
- Folgerung: Taxifahren führt zur Vergrößerung des Hippocampus!
- Oder etwa nicht????







Nehmen wir an, zwei Variablen X und Y korrelieren positiv

→ Je mehr X, desto mehr Y

Interpretation 1: Einseitige Steuerung

X bewirkt Y





Nehmen wir an, zwei Variablen X und Y korrelieren positiv

→ Je mehr X, desto mehr Y

Interpretation 2: Gegenseitige Steuerung

X wirkt auf Y, Y wirkt auf X zurück





Nehmen wir an, zwei Variablen X und Y korrelieren positiv

→ Je mehr X, desto mehr Y

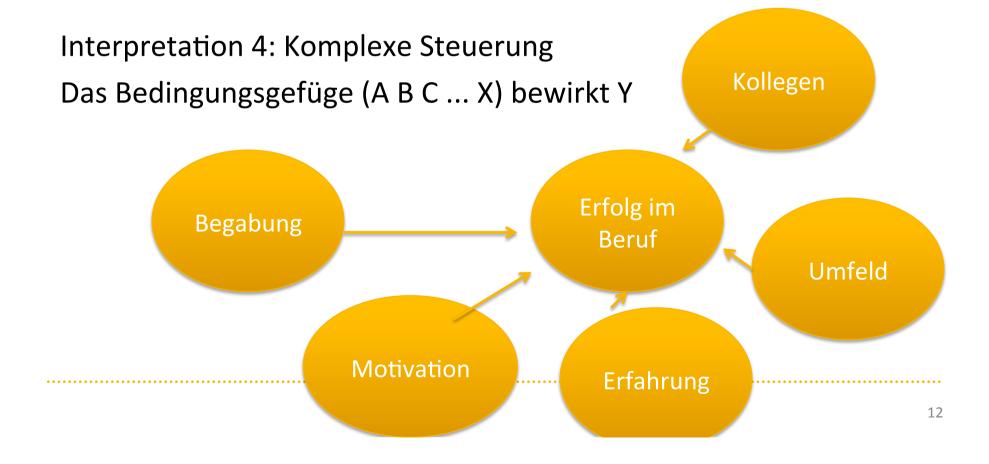
Interpretation 3: Drittseitige Steuerung X und Y hängen von einer dritten Variable Z ab.





Nehmen wir an, zwei Variablen X und Y korrelieren positiv

→ Je mehr X, desto mehr Y





- Wir haben es (nahezu) immer mit komplexen Steuerungen zu tun...
- ... die Frage ist nur, wie wir mit den Kovariaten umgehen:
 - Konstant halten (Laboruntersuchungen)
 - Zufällig verteilen ("Masse statt Klasse")
 - Vernachlässigen
- Korrelation = gemeinsames Variieren von Variablen
- Bei der kausalen Interpretation von Korrelationen ist stets Vorsicht geboten!



- Korrelationen können zunächst nur als Koinzidenzen interpretiert werden.
- Ob ein kausaler Zusammenhang besteht, kann wenn überhaupt - nur dann gesagt werden, wenn eine Variable im Vorhinein systematisch variiert wurde:
 - Trainingszeit → Erfolg
 - Dosis → Wirkung
- Experimentelle Designs!
- Richtung immer bidirektional!





FORMEL

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot s_x \cdot s_y}$$



FORMEL

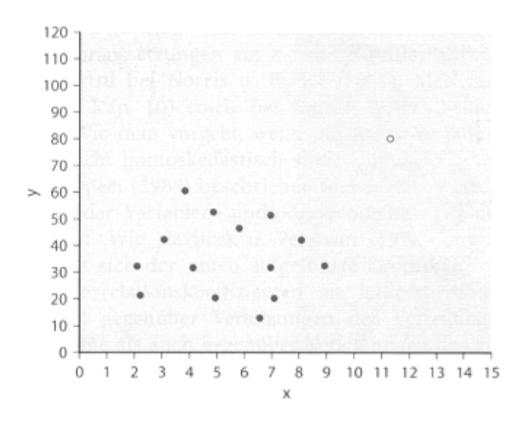
- Wenn x und y gemeinsam in eine Richtung von ihrem Mittelwert abweichen, wird der Zähler größer, damit r auch.
- Wenn x und y in entgegengesetzte Richtungen, aber gemeinsam von ihren Mittelwerten abweichen, wird der Zähler negativ groß.
- Wenn x und y beide unabhängig voneinander von ihrem jeweiligen Mittelwert abweichen, geht der Zähler gegen 0.
- Ist das Rauschen (Standardabweichungen) in den Stichproben groß, wird der Nenner größer und damit r kleiner.

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot s_x \cdot s_y}$$



UMGANG MIT AUSREIßERWERTEN

- Korrelationen können von Ausreißerwerten künstlich vergrößert werden!
- Lösungsansatz: Immer ein Streudiagramm anzeigen lassen und evtl. Ausreißerwerte von der Analyse ausschließen.





ANDERE KORRELATIONSKOEFFIZIENTEN

Merkmal y	Merkmal x		
	Intervallskala	Ordinalskala	dichotom
Intervallskala	Produkt- Moment- Korrelation (Pearsons r)	Rangkorrelation (Spearmans p)	Punkt-biseriale Korrelation
Ordinalskala	Rangkorrelation (Spearmans p)	Rangkorrelation (Spearmans p)	Biseriale Rangkorrelation
dichotom	Punkt-biseriale Korrelation	Biseriale Rangkorrelation	Φ -Koeffizient



BEGRIFFE

Korrelationsanalyse
Korrelationskoeffizient
positive und negative Korrelation
Scatterplot

Signifikanz

Ausreißerwerte

Vorsicht bei kausaler Interpretation!!!