Statistik &
Methodenlehre

Prof. Dr. G. Meinhardt
6. Stock, Wallstr. 3
(Raum 06-206)

Sprechstunde jederzeit nach Vereinbarung und nach der Vorlesung.

Mathematische und statistische Methoden II

Dr. Malte Persike
persike@uni-mainz.de
http://psymet03.sowi.uni-mainz.de/methods/

SS 2010
Fachbereich Sozialwissenschaften
Psychologisches Institut
Johannes Gutenberg Universität Mainz

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Wahrscheinlichkeitsverteilungen

Uni- und multivariate Zufallsvariablen

- **Univariate Zufallsvariable X**: Zuweisung nur einer Zahl zu jedem Ergebnis eines Zufallsexperiments (d.h. zu jedem zufällig ausgewählten Merkmalsträger)
- **Multivariate Zufallsvariable X , Y, …**: Zuweisung mehrerer Zahlen zu jedem Ergebnis
- Einer der häufigsten Spezialfälle multivariater Zufalls-variablen ist der bivariate Fall mit zwei ZVn X und Y
- Beispiele für mehrdimensionale Zufallsvariablen:
  - Erfassung von mathematischem und verbalem IQ
  - Messung von Ölverbrauch und Ausfallrate
  - Erfassung von Einkommen und Parteipräferenz

Wahrscheinlichkeitsverteilungen

Uni- und multivariate Wahrscheinlichkeitsverteilungen

- Die Wahrscheinlichkeitsverteilung einer univariaten Zufallsvariablen wird zumeist mit $f(x)$ bezeichnet.
- Sie beschreibt die Punktwahrscheinlichkeit für das Auftreten **einer Ausprägung** $X=x$.
- Die Wahrscheinlichkeitsverteilung einer multivariaten Zufallsvariablen wird zumeist mit $f(x,y,\ldots)$ bezeichnet.
- Sie beschreibt die Punktwahrscheinlichkeit für das Auftreten **einer Wertekombination** $X=x, Y=y, \ldots$.

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Wahrscheinlichkeitsverteilungen

Der bivariate Fall – Multivariate Verteilungsfunktion

- Die **Verteilungsfunktion** zweier **diskreter** Zufallsvariablen X und Y ist definiert als

$$\boxed{f(X \leq x_m, Y \leq y_n) = \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i, y_j)}$$

- Die **Verteilungsfunktion** zweier **stetiger** Zufallsvariablen X und Y ist definiert als

$$\boxed{F(x_u \leq X \leq x_o, y_u \leq Y \leq y_o) = \int_{x_u}^{x_o} \int_{y_u}^{y_o} f_{XY}(x, y)\, dxdy}$$

- Wie bei univariaten stetigen Verteilungen ist die **Punktwahrscheinlichkeit stets Null**, nur die **Intervallwahrscheinlichkeit** ist ein sinnvoller Wert.

Wahrscheinlichkeitsverteilungen

Der bivariate Fall – Beispiel Normalverteilung

- Die univariate Wahrscheinlichkeitsfunktion der Normalverteilung war

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{x-\mu_x}{\sigma_x}\right)^2\right)$$

- Die bivariate Wahrscheinlichkeitsfunktion der Normalverteilung lautet

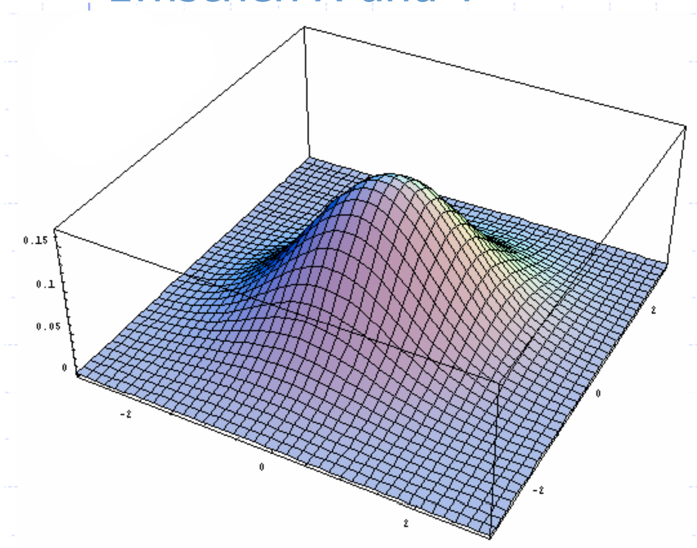
$$f_X(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho \cdot \frac{(x-\mu_x)}{\sigma_x} \cdot \frac{(y-\mu_y)}{\sigma_y}\right]\right)$$

- **Frage:** Welche Rolle spielt der Parameter ρ?

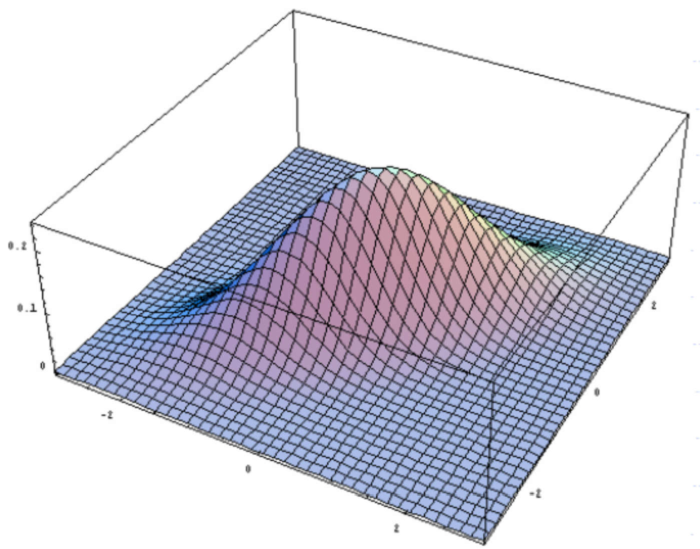
Wahrscheinlichkeitsverteilungen

Der bivariate Fall – Beispiel Normalverteilung

Niedriger Zusammenhang zwischen X und Y



Hoher Zusammenhang zwischen X und Y



Die **Elongation** oder „Länglichkeit" der multivariaten Wahrscheinlichkeitsverteilung ist ein Kennzeichen für den Zusammenhang der Zufallsvariablen

(Parameter ρ im Fall der bivariaten Normalverteilung)

Wahrscheinlichkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

- Die Größe des Zusammenhangs zweier Zufallsvariablen wird als **Kovarianz** bezeichnet

$$\boxed{\operatorname{cov}_{XY} = E\left[\left(X - E(X)\right)\cdot\left(Y - E(Y)\right)\right]}$$

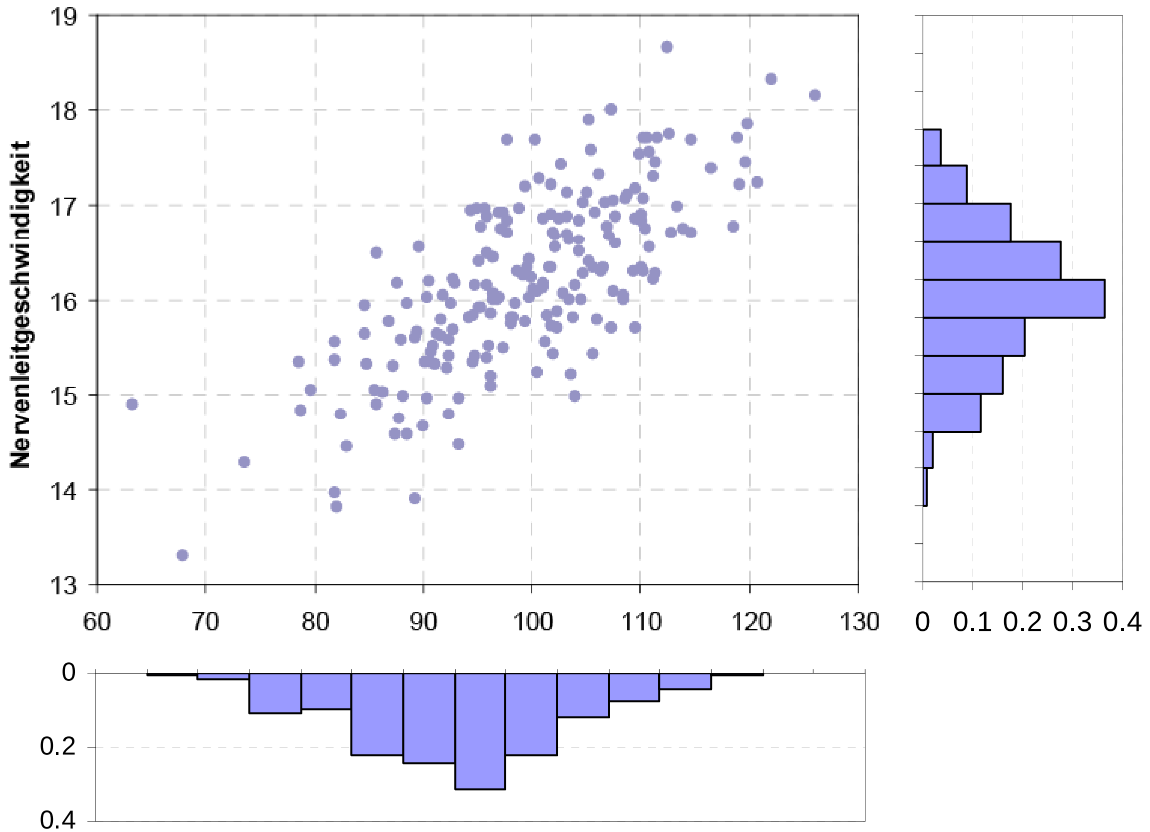
- Die Kovarianz ist Null, wenn **kein Zusammenhang** zwischen den Zufallsvariablen besteht.
  Wahrscheinlichkeitsfunktionen für X und Y hängen nicht zusammen

- Die Kovarianz ist positiv, wenn ein **gleichsinniger Zusammenhang** besteht
  Hohe (niedrige) Werte von X treten mit hohen (niedrigen) Werten von Y auf

- Die Kovarianz ist negativ, wenn ein **gegensinniger Zusammenhang** besteht
  Niedrige (hohe) Werte von X treten mit hohen (niedrigen) Werten von Y auf

Häufigkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung



Häufigkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

- Für n Beobachtungen aus einem Zufallsexperiment $x_1 \ldots x_n$ und $y_1 \ldots y_n$ ist die **Kovarianz** definiert als

$$\boxed{s_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}$$

- Es gelten bei der empirischen Kovarianz dieselben Prinzipien wie für die Kovarianz zwischen theoretischen Wahrscheinlichkeitsverteilungen.

- Die so berechnete Kovarianz ist nur bei **mindestens intervallskalierten Zufallsvariablen** ein sinnvolles Maß.

Definition

Kovarianz

Korrelation

Wahrscheinlichkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

- Problem: Die Kovarianz **erfüllt nicht die Forderung der Invarianz** gegenüber erlaubten Transformationen
- Für die Formel der empirischen Kovarianz ist das leicht zu zeigen und gilt ebenso für die theoretische Verteilung
- Zwar gilt: Addition einer Konstanten zu X und Y:

$$\boxed{\operatorname{cov}_{X+a,Y+B} = \operatorname{cov}_{X,Y}} \qquad \boxed{s_{xy}(x+a, y+b) = s_{xy}}$$

- **Aber:** Multiplikation von X und X mit einer Konstanten

$$\boxed{\operatorname{cov}_{a \cdot X, b \cdot Y} = a \cdot b \cdot \operatorname{cov}_{X,Y}} \qquad \boxed{s_{xy}(a \cdot X, b \cdot Y) = a \cdot b \cdot s_{xy}}$$

- Die Kovarianz ist also numerisch schwer zu interpretieren

Wahrscheinlichkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

- Die Korrelation zweier Zufallsvariablen ist definiert als

$$\boxed{\rho = \frac{E\left[\left(X - E(X)\right)\cdot\left(Y - E(Y)\right)\right]}{E\left(X - E(X)\right)^2 \cdot E\left(Y - E(Y)\right)^2} = \frac{\operatorname{cov}_{XY}}{\sigma_X \cdot \sigma_Y}}$$

- Für die **Richtungsinformation** gelten dieselben Regeln wie bei der Kovarianz

- Der Wert der Korrelation schwankt zwischen -1 und 1

- Bei der Korrelation ist neben der **Richtung** (Vorzeichen) also auch die **Stärke** (Betrag) des Zusammenhangs zwischen X und Y interpretier- und vergleichbar.

Definition

Kovarianz

Korrelation

Häufigkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

- Für empirische Daten gibt es je nach Skalenniveau verschiedene Berechnungsformel für die Korrelation.
- Für n **intervallskalierte** Beobachtungen aus einem Zufallsexperiment $x_1 \ldots x_n$ und $y_1 \ldots y_n$ ist der **Korrelationskoeffizient** definiert als

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{s_{xy}}{s_x \cdot s_y}$$

Häufigkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

- Der so definierte Korrelationskoeffizient r_{xy} wird auch als **Produkt-Moment-Korrelation** oder **Korrelationskoeffizient nach Pearson** bezeichnet.
- Die Korrelation liegt immer zwischen -1 und 1.
- Die Korrelation ist Null, wenn kein Zusammenhang zwischen den Ausprägungen der Zufallsvariablen besteht
- Negative Werte zeigen einen gegensinnigen, positive Werte einen gleichsinnigen Zusammenhang an
- Die Korrelation ist **anfällig gegenüber Ausreißern**

Definition

Kovarianz

Korrelation

Häufigkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

Kovarianz	Korrelation
$s_{x,y} = s_{y,x}$	$r_{x,y} = r_{x,y}$
$s_{a,x} = 0$	$r_{a,y}$ = nicht def.
$s_{a,b} = 0$	$r_{a,b}$ = nicht def.
$s_{x,x} = s^2_x$	$r_{x,x} = 1$
$s_{a\cdot x+b,c\cdot x+d} = a\cdot c\cdot s_{x,y}$	$r_{a\cdot x+b,\, c\cdot y+d} = r_{x,\, y}$
Mit a, b, c, d = konstante Werte	Achtung: Ist a oder b negativ, verändert sich das Vorzeichen von r, sind beide negativ, bleibt r gleich.

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Definition

Kovarianz

Korrelation

Häufigkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung

- Für die Bewertung der absoluten Höhe der Produkt-Moment-Korrelation existieren Faustregeln nach Cohen (1988)

$r = \pm\, 0.10 \quad \rightarrow$ kleine Korrelation

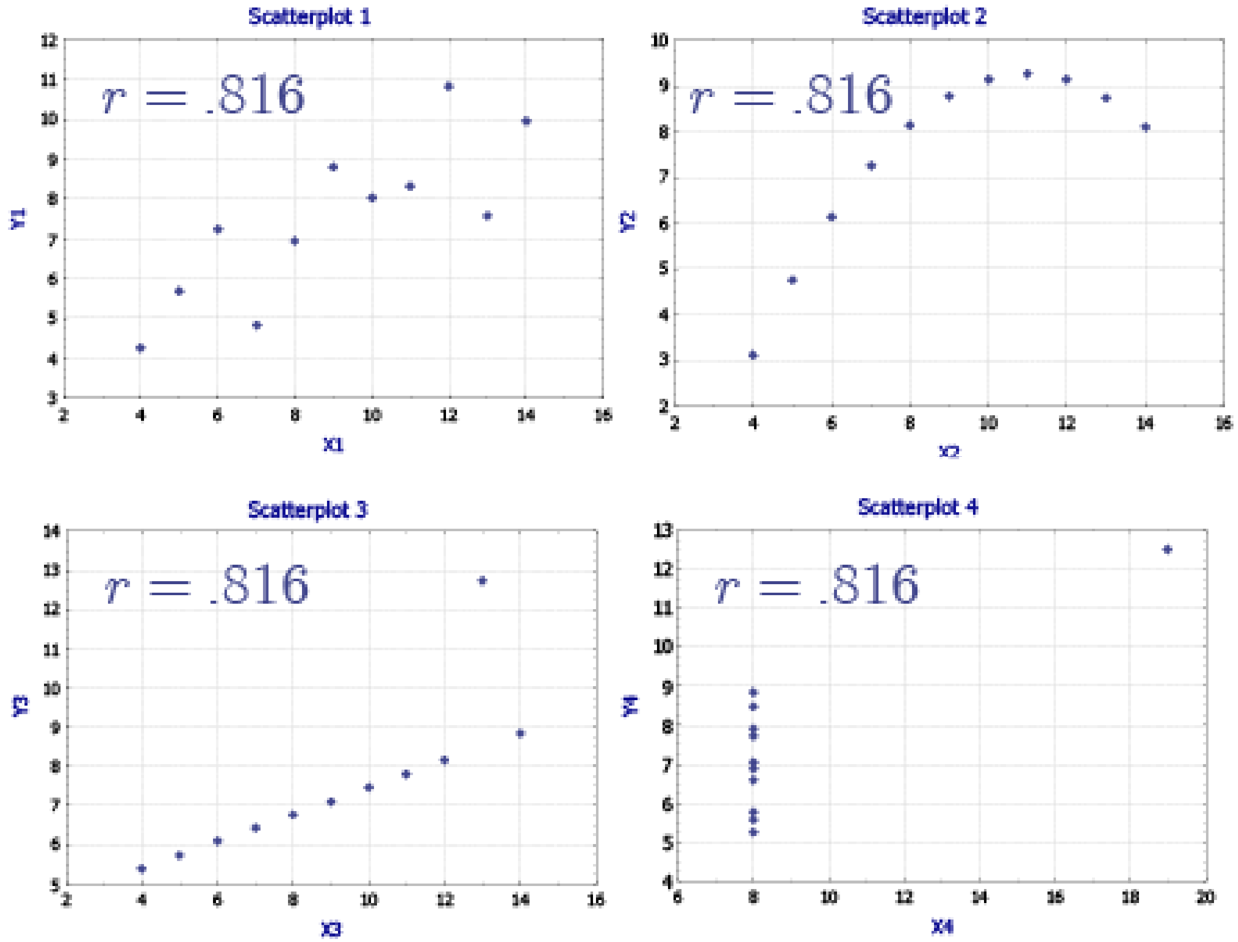
$r = \pm\, 0.30 \quad \rightarrow$ mittlere Korrelation

$r = \pm\, 0.50 \quad \rightarrow$ hohe Korrelation

- In der nicht-experimentellen Psychologie liegen Korrelationen selten über $r=0.75$.

Häufigkeitsverteilungen

Der bivariate Fall – Numerische Beschreibung



JOHANNES GUTENBERG UNIVERSITÄT MAINZ

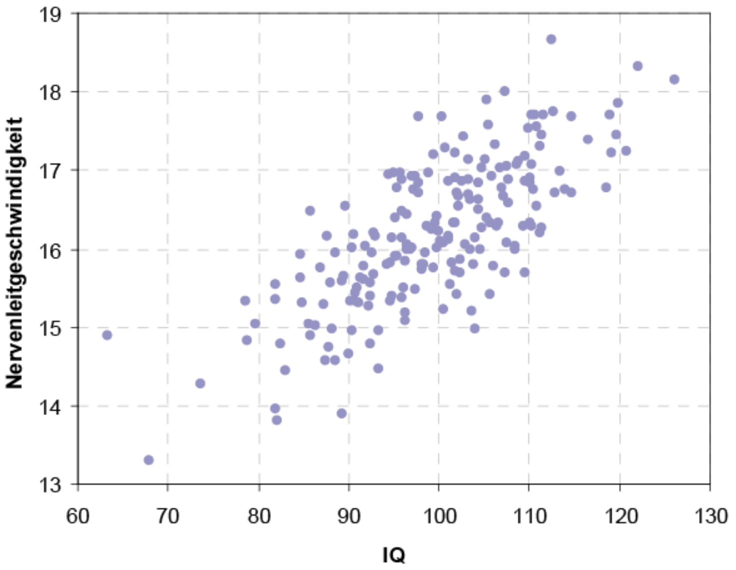
Häufigkeitsverteilungen

Der bivariate Fall – Grafische Beschreibung

- Liegen für eine Stichprobe Messungen der Ausprägung zweier Zufallsvariablen X und Y vor, so ist jeder Merkmalsträger durch ein **Wertepaar** gekennzeichnet.

- Die Merkmalsträger können nun als Punkte in einem Koordinaten-system dargestellt werden, wobei das Wertepaar die Koor-dinaten festlegt.

- Dies ist ein **Scatterplot**



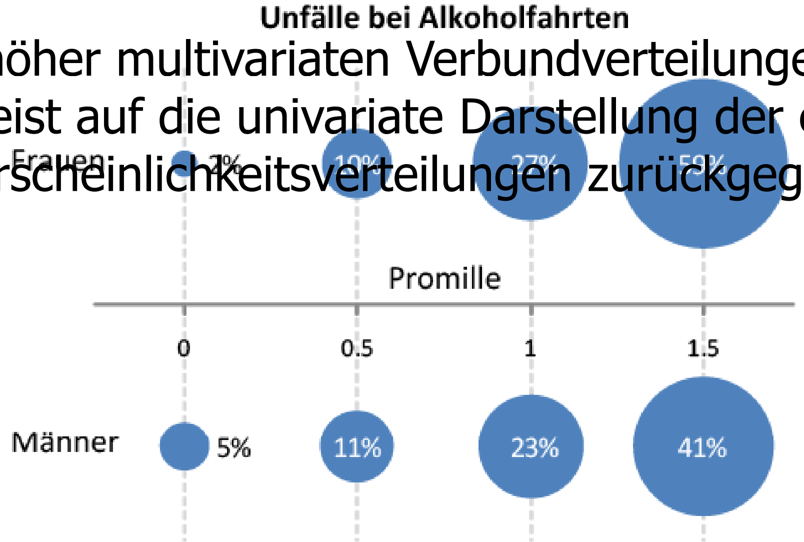
Häufigkeitsverteilungen

Der multivariate Fall – Grafische Beschreibung

- Die grafische Beschreibung multivariater Wahrscheinlichkeitsverteilungen ist aufwändig und vollständig nur in niedrig multivariaten Fällen zu leisten.

- Die Darstellungen sind dann oft dreidimensional.

- Bei höher multivariaten Verbundverteilungen wird zumeist auf die univariate Darstellung der einzelnen Wahrscheinlichkeitsverteilungen zurückgegriffen.



Korrelationstest

Varianten

- Für den statistischen Test von Korrelationen anhand von Stichprobendaten existieren verschiedene Verfahren, abhängig von der jeweiligen Forschungsfrage
    - Test einer Korrelation gegen Null
    - Test einer Korrelation gegen einen gegebenen Populationswert
    - Vergleich von Korrelationskoeffizienten aus zwei unabhängigen Stichproben
    - Vergleich von Korrelationskoeffizienten aus abhängigen Stichproben

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Korrelationstest

Schätzung des Populationsparameters ρ

- Aus Stichprobendaten erhält man eine **empirisch beobachtete** Produkt-Moment-Korrelation r.
- **Frage:** Wie erhält man daraus die Schätzung des Populationsparameters ρ für die Korrelation?
- Es lässt sich zeigen, dass die Produkt-Moment-Korrelation eine erwartungstreue Schätzung des Populationsparameters ist.
- Also gilt für die Zufallsvariablen X und Y:

$$\boxed{\hat{\rho}_{XY} = r_{XY}}$$

Korrelationstest

Test einer Korrelation gegen Null

- Bei empirisch beobachteten Korrelationen zweier Zufallsvariablen ist die erste Forschungsfrage oft, ob es in der Population überhaupt einen Zusammenhang gibt.
- Dies entspricht dem Test der aus einer Stichprobe geschätzten Korrelation gegen den Erwartungswert $\rho = 0$.
- Die zu testenden Hypothesen sind

$$\boxed{\begin{array}{lll} \text{a)} & H_0: \hat{\rho} = 0; & H_1: \hat{\rho} \neq 0 \\ \text{b)} & H_0: \hat{\rho} \leq 0; & H_1: \hat{\rho} > 0 \\ \text{c)} & H_0: \hat{\rho} \geq 0; & H_1: \hat{\rho} < 0 \end{array}}$$

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

1 Stichprobe
2 Stichproben (abhängig)
2 Stichproben (unabhängig)
Konfidenzint.

Korrelationstest

Test einer Korrelation gegen Null

- Die Verteilung von Korrelationskoeffizienten bzw. ihre Parameter sind nicht einfach bestimmbar.
- Es lässt sich aber eine geeignete **Transformation** finden, so dass der Kennwert approximativ **t-verteilt** ist um einen Erwartungswert von $\rho = 0$.

$$\boxed{t = \frac{\hat{\rho}_t - 0}{\sigma_{\hat{\rho}_t}}} \quad \text{mit} \quad \boxed{\hat{\rho}_t = \frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}}} \quad \text{und} \quad \boxed{\sigma_{\hat{\rho}_t} = \sqrt{\frac{1}{n-2}}}$$

- Die Prüfgröße ist t-verteilt mit $df = n - 2$ Freiheitsgraden
- Wie beim t-Test kann ein- oder zweiseitig geprüft werden.

Korrelationstest

Test einer Korrelation gegen Null

- Die Verteilung von Korrelationskoeffizienten bzw. ihre Parameter sind nicht einfach bestimmbar.
- Es lässt sich aber eine geeignete **Transformation** finden, so dass der Kennwert approximativ **t-verteilt** ist um einen Erwartungswert von ρ = 0.

$$\boxed{t = \frac{\hat{\rho}_t - 0}{\sigma_{\hat{\rho}_t}}} \quad \text{mit} \quad \boxed{\hat{\rho}_t = \frac{r}{\sqrt{1-r^2}}} \quad \text{und} \quad \boxed{\sigma_{\hat{\rho}_t} = \sqrt{\frac{1}{n-2}}}$$

- Die Prüfgröße ist t-verteilt mit $df = n - 2$ Freiheitsgraden
- Wie beim t-Test kann ein- oder zweiseitig geprüft werden.

Korrelationstest

Test einer Korrelation gegen einen Populationswert

- Bei bestimmten Fragestellung ist der Populationsparameter für die Korrelation zweier Zufallsvariablen bereits bekannt.
- **Beispiel:** Korrelation zwischen IST-2000R und Berufserfolg ist ρ = .47.
- Dies entspricht dem Test einer geschätzten Korrelation gegen den Erwartungswert ρ = c.
- Die zu testenden Hypothesen sind

$$
\begin{aligned}
&\text{a)}\quad H_0: \hat{\rho} = c; \quad H_1: \hat{\rho} \neq c \\
&\text{b)}\quad H_0: \hat{\rho} \leq c; \quad H_1: \hat{\rho} > c \\
&\text{c)}\quad H_0: \hat{\rho} \geq c; \quad H_1: \hat{\rho} < c
\end{aligned}
$$

Korrelationstest

Test einer Korrelation gegen einen Populationswert

- Soll geprüft werden, ob ein beobachtetes r einer Population mit dem wahren Parameter $\rho = c \neq 0$ entstammt, ist die t-Prüfgröße nicht anwendbar.
- Die Fisher-Z Transformation überführt eine Korrelation in einen approximativ **normalverteilten** Kennwert.
- Es gilt: $$\boxed{Z = \frac{1}{2} \cdot \ln\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) = \frac{1}{2} \cdot \ln\left(\frac{1+r}{1-r}\right)}$$

  ist unter der H0 approximativ normalverteilt mit $\boxed{\mu_Z = Z(\rho)}$ und $\boxed{\sigma_Z = \sqrt{\frac{1}{n-3}}}$

- Die Prüfgröße $\boxed{z = \frac{Z - \mu_Z}{\sigma_Z}}$ ist **standardnormalverteilt** mit μ=0 und σ=1.

Korrelationstest

Vergleich von zwei Korrelationskoeffizienten I

- Liegen aus zwei **unabhängigen Stichproben** Messungen zweier Zufallsvariablen vor, so kann für jede Stichprobe eine Korrelation zwischen den ZVn berechnet werden.
- Es kann nun geprüft werden, ob beide Stichproben zu einer Population **mit demselben Erwartungswert** der Korrelation zwischen den Zufallsvariablen gehören.
- Die zu testenden Hypothesen sind

$$\begin{aligned}
&\text{a)}\quad H_0 : \hat{\rho}_1 = \hat{\rho}_2; \quad H_1 : \hat{\rho}_1 \neq \hat{\rho}_2 \\
&\text{b)}\quad H_0 : \hat{\rho}_1 \leq \hat{\rho}_2; \quad H_1 : \hat{\rho}_1 > \hat{\rho}_2 \\
&\text{c)}\quad H_0 : \hat{\rho}_1 \geq \hat{\rho}_2; \quad H_1 : \hat{\rho}_1 < \hat{\rho}_2
\end{aligned}$$

Korrelationstest

Vergleich von zwei Korrelationskoeffizienten I

- Mithilfe der Fisher Z-Transformation können zwei Korrelationskoeffizienten r_1 und r_2 **aus zwei unabhängigen Stichproben** der Größen n_1 und n_2 auf Unterschiedlichkeit geprüft werden

- Die Prüfgröße z ist nach Fisher-Z Transformation der Korrelationen **standardnormalverteilt** mit μ=0 und σ=1 und berechnet sich als

$$\boxed{z = \frac{Z_1 - Z_2}{\sigma_{\Delta_{Z_1 Z_2}}}}$$

- Der Standardfehler $\sigma_{\Delta z}$ ist $\boxed{\sigma_{\Delta_{Z_1 Z_2}} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$

Korrelationstest

Vergleich von zwei Korrelationskoeffizienten II

- Beim Test von Korrelationen **aus einer Stichprobe** (also „abhängige Korrelationen") unterscheidet man zwei Fälle:

  1. 2 Korrelationen von Zufallsvariablen $X1$ und $X2$ mit einer Drittvariable Y

     $\rightarrow r_{X1Y}$ vs. r_{X2Y}

  2. 2 Korrelationen von Variablen $X1$ mit $X2$ und $Y1$ mit $Y2$

     $\rightarrow r_{X1X2}$ vs. r_{Y1Y2}

- Die Hypothesen sind dieselben wie beim Korrelationstest für zwei unabhängige Stichproben.

1 Stichprobe
2 Stichproben (unabhängig)
2 Stichproben (abhängig)
Konfidenzint.

Korrelationstest

Vergleich von zwei Korrelationskoeffizienten IIa

- Die Prüfgröße für den Vergleich von zwei **abhängigen** Korrelationen aus Stichproben der Größe n ist immer:

$$z = \frac{Z_{X_1 Y} - Z_{X_2 Y}}{\sigma_{\Delta_{Z_{X_1 Y} Z_{X_2 Y}}}} \quad \text{mit} \quad \sigma_{\Delta_{Z_{x_1 y} Z_{x_2 y}}} = \sqrt{\frac{2 - 2 \cdot C}{n - 3}}$$

- Der Term C variiert abhängig davon, ob es sich um den Vergleich von r_{X1Y} mit r_{Y2Y} oder r_{X1X2} mit r_{Y1Y2} handelt

- Die Prüfgröße z ist **standardnormalverteilt** mit μ=0 und σ=1.

Korrelationstest

Vergleich von zwei Korrelationskoeffizienten IIa

- Für den Vergleich von zwei **abhängigen** Korrelationen r_{X1Y} und r_{X2Y} ist der Term C:

$$C = \frac{1}{\left(1 - r_{pooled}^2\right)^2} \cdot \left(r_{x_1 x_2} \cdot \left(1 - 2 \cdot r_{pooled}^2\right) - \frac{1}{2} \cdot r_{pooled}^2 \cdot \left(1 - 2 \cdot r_{pooled}^2 - r_{x_1 x_2}^2\right) \right)$$

mit $r_{pooled} = \left(r_{x_1 y} + r_{x_2 y}\right)/2$

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Korrelationstest

Vergleich von zwei Korrelationskoeffizienten IIa

- Für den Vergleich von zwei **abhängigen** Korrelationen r_{X1X2} und r_{Y1Y2} ist der Term C:

$$\boxed{\begin{aligned}C = \frac{1}{2\cdot\left(1-r_{pooled}^2\right)^2}\cdot\; & [(r_{x_1y_1} - r_{x_1x_2}\cdot r_{x_2y_1})\cdot(r_{x_2y_2} - r_{x_2y_1}.r_{y_1y_2}) \\ & +(r_{x_1y_2} - r_{x_1y_1}\cdot r_{y_1y_2})\cdot(r_{x_2y_1} - r_{x_1x_2}\cdot r_{x_1y_1}) \\ & +(r_{x_1y_1} - r_{x_1y_2}\cdot r_{y_1y_2})\cdot(r_{x_2y_2} - r_{x_1x_2}\cdot r_{x_1y_2}) \\ & +(r_{x_1y_2} - r_{x_1x_2}\cdot r_{x_2y_2})\cdot(r_{x_2y_1} - r_{x_2y_2}\cdot r_{y_1y_2})]\end{aligned}}$$

mit $\boxed{r_{pooled} = \left(r_{x_1x_2} + r_{y_1y_2}\right)/2}$

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Korrelationstest

Konstruktion von Konfidenzintervallen für *r*

- Mithilfe der Fisher Z-Transformation lassen sich Konfidenzintervalle für einen beobachteten Korrelationskoeffizienten r konstruieren

Kritischer Wert aus der **Standard-normalverteilung**

$$\boxed{\left[Z_r - z_{1-\alpha/2} \cdot \sigma_{Z_r} ; Z_r + z_{1-\alpha/2} \cdot \sigma_{Z_r}\right]}$$

- Die **Konfidenzintervallgrenzen** konstruiert man zunächst in Z-Metrik über

$$\boxed{Z_r \pm z_{1-\alpha/2} \cdot \sigma_Z} \quad \text{bzw.} \quad \boxed{Z_r \pm z_{1-\alpha/2} \cdot \sqrt{\frac{1}{N-3}}}$$

- Dann können die Grenzen über die inverse Fisher Z-Transformation wieder in r-Metrik übersetzt werden:

$$\boxed{r = \frac{e^{2Z} - 1}{e^{2Z} + 1}}$$

Korrelationstest

Interpretation signifikanter Korrelationen

- Die Verteilungsannahme der Prüfgröße bei Korrelationen ist besonders bei kleinen Stichprobengrößen eher heikel.
- Eine signifikante Korrelation zwischen zwei Zufallsvariablen X und Y darf **nicht ohne weiteres als Kausalität** zwischen den Variablen interpretiert werden.
- Eine signifikante Korrelation zeigt zunächst nur eine **Assoziation** an. Diese kann viele Ursachen haben, z.B.

$$X \longrightarrow Y$$

$$X \longleftarrow Y$$

$$X \longleftarrow Z \longrightarrow Y$$

Korrelationstest

Interpretation signifikanter Korrelationen

- **Frage:** Wann darf in einer psychologischen Untersuchung auf Kausalität geschlossen werden?

1. Das Treatment und der Effekt müssen **kovariieren**
   $\rightarrow$ der Korrelationstest muss eine Signifkanz anzeigen

Probleme:

– Standards (z.B. Signifkanzniveau) sind normativ

– Je höher n, desto eher werden kleinste Effekte signifkant

Assoziation

Kausalität

Korrelationstest

Interpretation signifikanter Korrelationen

- **Frage:** Wann darf in einer psychologischen Untersuchung auf Kausalität geschlossen werden?

1. Das Treatment und der Effekt müssen **kovariieren**
   → der Korrelationstest muss eine Signifkanz anzeigen
2. Die Ursache muss der Wirkung **zeitlich vorausgehen** (z.B. Pretest – Treatment – Posttest)
3. Andere plausible Erklärungen für die Kovariation müssen ausgeschlossen werden können
4. Die Kovariation muss raum-zeitlich indifferent sein
   → Generalisierung auf eine Population zu jeder Zeit

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Relevante Excel Funktionen

- Zusammenhangsmaße
  - KORREL()
  - LN()
  - EXP()