

# Statistik I

---

## Einheit 9: Korrelation

09.01.2024 | Prof. Dr. Stephan Goerigk

# Korrelation

## Der Begriff des (bivariaten) Zusammenhangs

Was bedeutet es, wenn zwei Variablen miteinander zusammenhängen?

- Die Ausprägung, die eine Person auf der einen Variable aufweist, gibt zu gewissen Teilen auch Auskunft darüber, welche Ausprägung diese Person auf der anderen Variable erreicht.
- Beide Variablen variieren dann systematisch miteinander.

## Beispiel: Müdigkeit und Konzentrationsfähigkeit

- Müdigkeit und Konzentrationsfähigkeit stehen in Beziehung zueinander
- Überdurchschnittlich müde Menschen weisen in der Regel eine niedrigere Konzentrationsfähigkeit auf als weniger müde Menschen

Mit anderen Worten: Je höher die Müdigkeit, umso niedriger die Konzentrationsfähigkeit

→ Aussagen wie diese postulieren einen Zusammenhang

# Korrelation

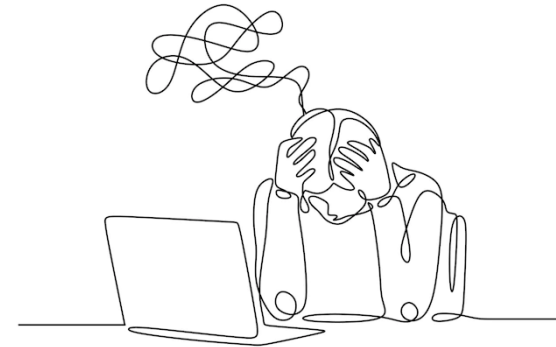
## Der Begriff des (bivariaten) Zusammenhangs

### Stärke des Zusammenhangs

- Die Stärke des Zusammenhangs ist davon abhängig, wie zwingend von der einen Variable auf die andere geschlossen werden kann.

### Beispiel: Müdigkeit und Konzentrationsfähigkeit

- wenn alle überdurchschnittlich müden Personen eine niedrige Konzentrationsfähigkeit aufweisen → starker Zusammenhang
- wenn es unter den überdurchschnittlich müden Personen auch welche mit hoher Konzentrationsfähigkeit gibt → weniger starker Zusammenhang



# Korrelation

## Der Begriff des (bivariaten) Zusammenhangs

### Richtung des Zusammenhangs

- Zusammenhänge können in zwei "Richtungen" vorliegen: positiv oder negativ.

### Positiver Zusammenhang:

- Hohe Werte auf der einen Variable entsprechen hohen Werten auf der anderen Variable

### Negativer Zusammenhang:

- Hohe Werte auf der einen Variable entsprechen niedrigen Werten auf der anderen Variable

### Kein Zusammenhang:

- Werte auf der einen Variable gehen mal mit hohen und mal mit niedrigen Werten auf der anderen Variable einher
- Beide Merkmale sind stochastisch unabhängig voneinander

# Korrelation

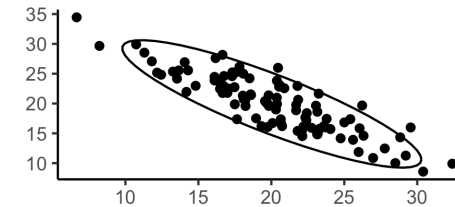
## Der Begriff des Zusammenhangs

### Richtung des Zusammenhangs

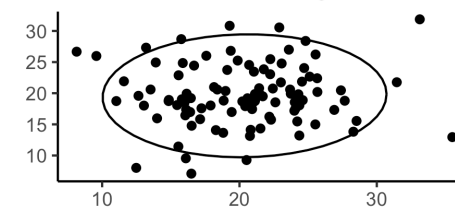
### Stärke des Zusammenhangs

- Je enger die elliptische Form eines Punkteschwarms ist, umso stärker ist der Zusammenhang
- Ist die Punktwolke kreisförmig, besteht kein Zusammenhang
- Grafische Interpretation erlaubt grobe Orientierung über den Grad des Zusammenhangs. Diesen zu quantifizieren ist Aufgabe der Kovarianz und Korrelation.

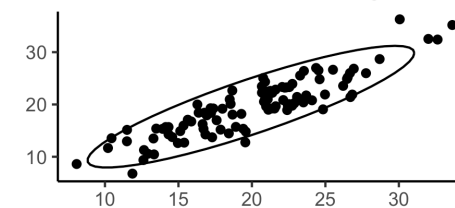
Negativer Zusammenhang



Kein Zusammenhang



Positiver Zusammenhang



# Korrelation

## Kovarianz ( $cov_{(x,y)}$ )

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

- Unstandardisiertes Maß, dass den Zusammenhang zweier Variablen erfasst (bivariat)
- Ist in ihrer Berechnung der Varianz nicht unähnlich
- Varianz: Erfasst die Abweichungen einer Variable um ihren Mittelwert
  - Liefert aufgrund der Quadrierung nur positive Ergebnisse
- Kovarianz: Gleichgerichtete Abweichungen zweier Variablen von deren Mittelwerten
  - Kann auch negative Werte annehmen

# Korrelation

## Kovarianz ( $cov_{(x,y)}$ )

### Wertespektrum:

- Betrag der Kovarianz zwischen 2 Variablen kann beliebige Werte zwischen 0 und einer maximalen Kovarianz  $|cov_{max}|$  annehmen.
- $|cov_{max}|$  ist für positiver und negative Zusammenhänge identisch

Definition von  $|cov_{max}|$ :

- Definiert als das Produkt der beiden Merkmalsstreuungen:

$$|cov_{max}| = \hat{\sigma}_x \cdot \hat{\sigma}_y$$

# Korrelation

## Kovarianz ( $cov_{(x,y)}$ )

### Wertespektrum:

$$(cov_{(x,y)}) > 0:$$

- Kovarianz ist positiv, wenn positive Abweichungen vom Mittelwert in  $X$  mit positiven in  $Y$  und negativen Abweichungen in  $X$  mit negativen in  $Y$  einhergehen.

$$(cov_{(x,y)}) < 0:$$

- Kovarianz ist negativ, wenn positive Abweichungen vom Mittelwert in  $X$  mit negativen in  $Y$  einhergehen und umgekehrt.

$$(cov_{(x,y)}) = 0:$$

- Kovarianz von 0 besagt, dass beide Variablen in keinem Zusammenhang zueinander stehen.



# Korrelation

## Kovarianz ( $cov_{(x,y)}$ )

### Wertespektrum:

- Abweichungen, die in Formel zur Berechnung der Kovarianz eingehen, sind abhängig von der Skalierung (Maßstab) der Merkmale.

### Beispiel:

Für dieselbe Länge ergibt sich eine höhere Kovarianz, wenn in cm anstatt in m gemessen wird

→ Höhe der Kovarianz ist ein unstandardisiertes Maß für den Zusammenhang

- Es lässt sich nur Richtung erkennen
- Höhen unterschiedlicher Zusammenhänge lassen sich nicht vergleichen (hierfür nutzen wir Korrelation)

# Korrelation

## Kovarianz ( $cov_{(x,y)}$ ) - Beispiel

Ein Psychologe misst die Schlafdauer (X) von ( $N = 16$ ) Proband:innen in Stunden (h). Nach dem Aufstehen müssen Sie einen Konzentrationstest (Y) durchführen (0-15 Punkte). Hohe Punktzahl entspricht guter Konzentrationsleistung.

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Schlaf (X)	6	7	13	8	8	13	9	4	6	7	12	9	9	8	6	13
Konzentration (Y)	2	2	9	3	4	8	2	4	4	2	12	7	5	7	6	12

Die Kovarianz von Schlaf und Konzentration in dieser Stichprobe berechnet sich wie folgt ( $\bar{x} = 8.62, \bar{y} = 5.56$ ):

$$cov_{x,y} = \frac{(6 - 8.62) \cdot (2 - 5.56) + (7 - 8.62) \cdot (2 - 5.56) + \dots}{15} = 7.16$$

- Der Zusammenhang zwischen Schlaf und Konzentrationsfähigkeit ist also positiv.
- Über die relative Höhe des Zusammenhangs sagt dieser Wert nichts aus. Dafür brauchen wir ein standardisiertes Maß.

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

- Standardisiertes Maß für den Zusammenhang zweier intervallskalierter Variablen
- Gebräuchlichstes Maß für die Stärke des Zusammenhangs
- Stärke des Zusammenhangs wird mit dem Korrelationskoeffizient  $r$  angegeben.
- Anderer Name für die Korrelation: Pearson-Korrelation

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Berechnung

- Standardisierung der im vorherigen Abschnitt besprochenen Kovarianz
- Die berechnete Kovarianz wird anhand der maximalen Kovarianz ( $|cov_{max}|$ ) relativiert

$$r_{xy} = \frac{cov_{emp}}{cov_{max}} = \frac{cov_{x,y}}{\hat{\sigma}_x \cdot \hat{\sigma}_y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{(n-1) \cdot \hat{\sigma}_x \cdot \hat{\sigma}_y}$$

- Dadurch wird die Kovarianz von der Streuung der Merkmale bereinigt.
- Das resultierende  $r$  (Korrelationskoeffizient) ist somit maßstabsunabhängig (kann direkt interpretiert werden)

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Wertespektrum:

Der Korrelationskoeffizient kann nur Werte zwischen -1 und +1 annehmen:

- $r = 1$  perfekter positiver Zusammenhang
- $r = -1$  perfekter negativer Zusammenhang
- $r = 0$  kein Zusammenhang (Nullkorrelation)
- $r > 0$  positiver Zusammenhang
- $r < 0$  negativer Zusammenhang

# Korrelation

## Produkt-Moment-Korrelation ( $r$ ) - Beispiel

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Schlaf (X)	6	7	13	8	8	13	9	4	6	7	12	9	9	8	6	13
Konzentration (Y)	2	2	9	3	4	8	2	4	4	2	12	7	5	7	6	12

Um den Korrelationskoeffizienten ( $r$ ) für unser Beispiel zu bestimmen, müssen wir zunächst die Standardabweichungen von X und Y als Schätzer der Merkmalsstreuung berechnen:

- $\hat{\sigma}_x = 2.8$
- $\hat{\sigma}_y = 3.37$

$$cov_{x,y} = \frac{(6 - 8.62) \cdot (2 - 5.56) + (7 - 8.62) \cdot (2 - 5.56) + \dots}{15} = 7.16$$

$$r_{xy} = \frac{cov_{x,y}}{\hat{\sigma}_x \cdot \hat{\sigma}_y} = \frac{7.16}{2.8 \cdot 3.37} = 0.76$$

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Interpretation

- Ab welcher Höhe eine Korrelation zwischen 2 Variablen als inhaltlich bedeutsam angesehen werden kann, hängt stark vom Forschungsgegenstand ab.

Beispiel: Labor- vs. Feld

- Labor: Aufgrund der Kontrollierbarkeit von Störeinflüssen werden höhere Zusammenhänge erwartet.
- Feld: Anspruch an die Höhe der Korrelation ist in der Regel niedriger.
- Unterschied des berechneten  $r$  vs. 0 muss mittels Signifikanztest geprüft werden.

Einteilung der Größe des Effekts (Cohen, 1992):

$r$	Interpretation
$r > 0.1$	kleiner Effekt
$r > 0.3$	mittlerer Effekt
$r > 0.5$	großer Effekt

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Skalierung des Korrelationskoeffizienten

- Der Korrelationskoeffizient ist nicht intervallskaliert.
- Er kann demnach nicht als Prozentmaß des Zusammenhangs interpretiert werden.
- z.B. 0.4 ist nicht doppelt so hoch wie 0.8 (Äquidistanz ist nicht gegeben)
- Relationen ist kein perfektes Maß, wenn Zusammenhänge aus mehreren Studien miteinander verglichen und zusammen gefasst werden sollen (ein  $r$  vs. ein zweites  $r$ )
- Um direkte Vergleiche zwischen Korrelationskoeffizienten zu ermöglichen, benötigt man einen Zwischenschritt (Fishers Z-Transformation)



# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Pearson-Korrelation als Signifikanztest

- Die Korrelation kann als Signifikanztest genutzt werden
- Signifikanztest der Korrelation verläuft analog zum t-Test
- Einziger Unterschied: Interessierender Wert ist eine Korrelation anstelle einer Mittelwertdifferenz
- Schreibweise:
  - Empirisch ermittelte Korrelation:  $r$
  - Zusammenhang in der Population (Grundgesamtheit):  $\rho$  (Rho)

### Ungerichtete Hypothese: ("Es besteht ein Zusammenhang zwischen X und Y.")

- $H_0: \rho = 0$
- $H_1: \rho \neq 0$

### Gerichtete Hypothese: ("Es besteht ein positiver/negativer Zusammenhang zwischen X und Y.")

- $H_0: \rho \leq 0$ , bzw.  $\rho \geq 0$
- $H_1: \rho > 0$ , bzw.  $\rho < 0$

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Pearson-Korrelation als Signifikanztest

- Die zugehörige Prüfstatistik (t-Verteilung) ist uns bereits bekannt
- Der empirische t-Wert ( $t_{emp}$ ) berechnet sich aus der empirisch ermittelte Korrelation ( $r$ ) und dem Stichprobenumfang ( $N$ )

$$t = \frac{r \cdot \sqrt{N-2}}{\sqrt{1-r^2}}; df = N - 2$$

Vorgehen:

- Empirischer t-Wert ( $t_{emp}$ ) wird gegen kritischen t-Wert ( $t_{krit}$ ) getestet (aus t-Tabelle ablesen)
- Zweiseitige  $H_0$  wird verworfen, wenn  $|t_{emp}| > t(df; 1 - \alpha/2)$  (kritischer t-Wert)
- Einseitige  $H_0$  wird verworfen, wenn Abweichung in die erwartete Richtung und  $|t_{emp}| > t(df; 1 - \alpha)$  (kritischer t-Wert)

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Pearson-Korrelation als Signifikanztest - Beispiel

Der Psychologe aus unserem Beispiel möchte wissen, ob ein signifikanter positiver Zusammenhang zwischen Schlaf und Aufmerksamkeitsleistung besteht.

Hypothesen:

- $H_0: \rho \leq 0$
- $H_1: \rho > 0$

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Pearson-Korrelation als Signifikanztest - Beispiel

Der Psychologe aus unserem Beispiel möchte wissen, ob ein positiver Zusammenhang zwischen Schlaf und Aufmerksamkeitsleistung besteht.

- $r_{xy} = 0.76$
- $N = 16$

$$t = \frac{0.76 \cdot \sqrt{16 - 2}}{\sqrt{1 - 0.76^2}}; df = 16 - 2$$

$$t = \frac{2.84}{0.65} = 4.37; df = 14$$

Bestimmung von ( $t_{krit}$ ) aus Tabelle (s.h. nächste Folie):

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Pearson-Korrelation als Signifikanztest - Beispiel

Kritischen t-Wert ( $t_{krit}$ ) für  $\alpha = .05$  nachschlagen:

Fläche*													
df	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	0,975	0,990	0,995	0,9995
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073

# Korrelation

## Produkt-Moment-Korrelation ( $r$ )

### Pearson-Korrelation als Signifikanztest - Beispiel

Der Psychologe aus unserem Beispiel möchte wissen, ob ein positiver Zusammenhang zwischen Schlaf und Aufmerksamkeitsleistung besteht.

- $r_{xy} = 0.76$
- $N = 16$

$$t = \frac{0.76 \cdot \sqrt{16 - 2}}{\sqrt{1 - 0.76^2}}; df = 16 - 2$$

$$t = \frac{2.84}{0.65} = 4.37; df = 14$$

- $t_{krit}(df=14) = 1.761 \rightarrow$  Spalte für 0.95 in der t-Tabelle, da gerichtete Hypothese
- Der empirische t-Wert ist größer als der kritische t-Wert, der Korrelationskoeffizient ist signifikant größer als 0 (positiver Zusammenhang).
- Die  $H_0$  kann verworfen werden. Es besteht ein positiver Zusammenhang zwischen Schlaf und Aufmerksamkeitsleistung.

# Korrelation

## Rangkorrelation ( $r_s$ )

- Standardisiertes Maß für den Zusammenhang zweier ordinalskalierten Variablen
- Stellt eine Analogie zur Produkt-Momentkorrelation dar.
- Unterschied: Anstelle intervallskalierter Messwerte werden Rangplätze von ordinalskalierten Rangreihen verwendet.
- Erfasst, inwieweit zwei Rangreihen systematisch miteinander variieren.
- Stärke des Zusammenhangs wird mit dem Korrelationskoeffizient  $r_s$  angegeben.
- Anderer Name für dieselbe Korrelation: Spearman-Korrelation

# Korrelation

## Rangkorrelation ( $r_s$ ) - Beispiel

- Eine Gruppe aus Frauen und Männern sollen  $N = 10$  Filme danach bewerten, wie lustig sie sind.
- Dafür sollen die Filme in eine Rangreihe (Platz 1-10) gebracht werden.
- Es sollt untersucht werden, ob ein Zusammenhang zwischen der Rangliste der Frauen und der Rangliste der Männer besteht.

Film	Rangplatz Frauen	Rangplatz Männer
1	8	6
2	10	10
3	3	1
4	1	4
5	6	7
6	7	9
7	2	3
8	5	2
9	9	8
10	4	5



# Korrelation

## Rangkorrelation ( $r_s$ ) - Beispiel

- Eine Gruppe aus Frauen und Männern sollen  $N = 10$  Filmedanach bewerten, wie lustig sie sind.
- Dafür sollen die Filme in eine Rangreihe (Platz 1-10) gebracht werden.
- Es soll untersucht werden, ob ein Zusammenhang zwischen der Rangliste der Frauen und der Rangliste der Männer besteht.

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{N \cdot (N^2 - 1)}$$

Film	Rangplatz Frauen (X)	Rangplatz Männer (Y)	Differenz (d)
1	8	6	2
2	10	10	0
3	3	1	2
4	1	4	-3
5	6	7	-1
6	7	9	-2
7	2	3	-1
8	5	2	3
9	9	8	1
10	4	5	-1

# Korrelation

## Rangkorrelation ( $r_s$ ) - Beispiel

- $d_i$  ist die Differenz der Ranglistenplätze einer Untersuchungseinheit  $i$  bezüglich der Variablen X und Y.

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{N \cdot (N^2 - 1)}$$

$$r_s = 1 - \frac{6 \cdot [2^2 + 0 + 2^2 + \dots + (-1)^2]}{10 \cdot (10^2 - 1)} = 1 - \frac{6 \cdot 204}{990} = 0.79$$

- Die Ranglisten der Frauen und Männer korrelieren also relativ stark miteinander.
- Frage: Ist dieser Zusammenhang auch statistisch signifikant?

# Korrelation

## Rangkorrelation ( $r_s$ ) - Signifikanztest

Für  $n \geq 30$  kann der Korrelationskoeffizient der Rangkorrelation mit dem t-verteilten Test näherungsweise auf Signifikanz geprüft werden ( $r_s = 0.79$ ;  $N = 10$ ).

Hypothesen:

- $H_0: \rho = 0$
- $H_1: \rho \neq 0$

$$t = \frac{r \cdot \sqrt{N-2}}{\sqrt{1-r^2}}; df = N - 2$$

$$t = \frac{0.79 \cdot \sqrt{10-2}}{\sqrt{1-0.79^2}} = \frac{2.23}{0.61} = 3.66$$

Bestimmung von ( $t_{krit}$ ) aus Tabelle (s.h. nächste Folie):

# Korrelation

## Rangkorrelation ( $r_s$ ) - Signifikanztest

Kritischen t-Wert ( $t_{krit}$ ) für  $\alpha = .05$  nachschlagen (ungerichtete Hypothese):

Fläche*														
df	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	0,975	0,990	0,995	0,9995	
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619	
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598	
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941	
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610	
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859	
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959	
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405	
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041	
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781	
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587	
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437	
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318	
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221	
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140	
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073	

# Korrelation

## Rangkorrelation ( $r_s$ ) - Signifikanztest

Für  $n \geq 30$  kann der Korrelationskoeffizient der Rangkorrelation mit dem t-verteilten Test näherungsweise auf Signifikanz geprüft werden ( $r_s = 0.79$ ;  $N = 10$ ).

Hypothesen:

- $H_0: \rho = 0$
- $H_1: \rho \neq 0$

$$t = \frac{r \cdot \sqrt{N-2}}{\sqrt{1-r^2}}; df = N - 2$$

$$t = \frac{0.79 \cdot \sqrt{10-2}}{\sqrt{1-0.79^2}} = \frac{2.23}{0.61} = 3.66$$

- $t_{krit}(df=8) = 2.306 \rightarrow$  Spalte für 0.975 in der t-Tabelle, da ungerichtete Hypothese
- Der empirische t-Wert ist größer als der kritische t-Wert, der Korrelationskoeffizient ist signifikant größer als 0 (positiver Zusammenhang).
- Die  $H_0$  kann verworfen werden. Es besteht ein positiver Zusammenhang zwischen den Ratings der Männer und Frauen.

# Korrelation

## Rangkorrelation ( $r_s$ ) - Weitere Verwendung

### Ausreißer:

- Die Rangkorrelation wird der Produkt-Moment-Korrelation auch vorgezogen, wenn Ausreißer in den Daten sind.
- Grund: Formel der Produkt-Moment-Korrelation nutzt Standardabweichungen (stark von Ausreißern beeinflusst).

### Korrelation einer intervallskalierten Variable mit einer ordinalen Variable:

- Es wird ebenfalls die Rangkorrelation verwendet.
- Hierfür wird die intervallskalierte variable auf Rangskalenniveau transformiert (z.B. kleinster Wert erhält niedrigsten Rang)
- Die Variable erfährt also eine Herabstufung des Skalenniveaus

# Korrelation

## Exkurs: Korrelation und Kausalität

- Korrelation darf nicht vorschnell als Beweis für Kausalitätsbeziehungen betrachtet werden
- Ursache-Wirkungs-Beziehungen sind durch das ledigliche gemeinsame Auftreten zweier Variablen nicht bewiesen

### Kausalmöglichkeiten:

- X verursacht Y
- Y verursacht X
- Gemeinsames Auftreten durch drittes Merkmal bedingt (Scheinkorrelation)

### Beispiel Scheinkorrelation:

Nach dem 1. Weltkrieg nahmen in Deutschland die Anzahl der Störche und die Geburtenrate gleichermaßen ab.

→ Der Storch bringt die Kinder.

## Take-aways

- Ein (bivariater) Zusammenhang zeigt sich darin, dass zwei Variablen **systematisch miteinander variieren**.
- Die **Kovarianz**, ein unstandardisiertes Zusammenhangsmaß, kann uns die Richtung des Zusammenhangs anzeigen, aber nicht direkt hinsichtlich seiner Stärke interpretiert werden.
- Der **Korrelationskoeffizient** ( $r$ ) ist ein standardisiertes Maß für den Zusammenhang zweier Variablen und kann Werte im Bereich von -1 bis +1 annehmen.
- Zusammenhänge zwischen zwei intervallskalierten Variablen werden mit der **Produkt-Moment-Korrelation** (Pearson), ordinalskalierte mit der **Rangkorrelation** (Spearman) berechnet.
- Die **Einteilung nach Cohen** erlaubt für  $|r| \geq .1$ ,  $|r| \geq .3$ ,  $|r| \geq .5$  eine Unterteilung in kleine, mittlere und starke/große Zusammenhänge.
- Ein bestehender Zusammenhang gibt **keine Auskunft über Kausalbeziehungen** zwischen den untersuchten Variablen



