

Zum Begriff der Paare:

- ab ordinalem Messniveau

Begriffsdefinition von Paaren:

- gleiche bzw. unterschiedliche Rangordnung zwischen Untersuchungsobjekten (z. B. Personen)
- Paare können **konkordant** oder **diskordant** sein

1) Konkordantes Paar \Leftrightarrow gleichsinnig bezüglich der Rangordnung (positive Beziehung)Beispiel:

Konkordantes Paar		
Schüler	Mathematiknote	Physiknote
A	1	1
B	2	4

- Es handelt sich hierbei um ein konkordantes Paar, weil Schüler A in **beiden Fächern besser** ist als Schüler B. Es handelt sich hierbei um eine positive Beziehung.

2) Diskordantes Paar \Leftrightarrow gegensinnig bezüglich der Rangordnung (negative Beziehung)Beispiel:

Diskordantes Paar		
Schüler	Mathematiknote	Physiknote
A	1	5
B	2	4

- Es handelt sich hierbei um ein diskordantes Paar, weil Schüler A zwar in Mathematik **besser**, aber in Physik **schlechter** ist als Schüler B. Es handelt sich hierbei um eine negative Beziehung.

3) Bindung bzw. Verknüpfung (Tie, T) \Rightarrow keine Rangordnung, da identische Zahlenwerte vorliegen

Beispiel:

Tie		
Schüler	Mathematiknote	Physiknote
A	1	4
B	2	4

- Hinsichtlich der Physiknote von Schüler A und B liegt ein Tie vor, weil beide die **gleiche** Note haben. Es liegt daher keine Rangordnung vor.

Allgemein:

$$\text{Anzahl der möglichen Paare} = \binom{N}{2} = \frac{N \cdot (N-1)}{2} = N_c + N_d + T_x + T_y + T_{xy}$$

N_c = Anzahl der konkordanten Paare

N_d = Anzahl der diskordanten Paare

T_x = Anzahl der Ties in Bezug auf die Variable X.
Die Untersuchungseinheiten sind in Hinblick auf X verknüpft, jedoch im Hinblick auf Y verschieden.

T_y = Anzahl der Ties in Bezug auf die Variable Y.
Die Untersuchungseinheiten sind in Hinblick auf Y verknüpft, jedoch im Hinblick auf X verschieden (siehe Beispiel).

T_{xy} = Anzahl der Ties in Bezug auf die Variablen X und Y.
Die Untersuchungseinheiten sind in Hinblick auf X und Y verknüpft.

Beispiel:

Schüler	Mathematiknote X	Physiknote Y
A	5	5
B	4	4
C	1	3
D	2	2
E	3	1

$$\text{Anzahl der möglichen Paare} = \binom{5}{2} = \frac{5 \cdot (5-1)}{2} = 10$$

Identifizierung der Paartypen		
	Paar	Paartyp
1	A - B	konkordant
2	A - C	konkordant
3	A - D	konkordant
4	A - E	konkordant
5	B - C	konkordant
6	B - D	konkordant
7	B - E	konkordant
8	C - D	diskordant
9	C - E	diskordant
10	D - E	diskordant

}

N_c = 7

}

N_d = 3

Konkordante und diskordante Paare in der bivariaten Tabelle (Ausschnitt)

		Mathematiknote X				
		1	2	3	4	5
Physiknote Y	1			1 (E)		
	2		1 (D)			
	3	1 (C)				
	4				1 (B)	
	5					1 (A)

Ermittlung der konkordanten und diskordanten Paare und der Ties in einer 2 x 2-Tabelle:

Beispiel:

	x ₁	x ₂
y ₁	a	b
y ₂	c	d

$$N_c = a(d) \qquad N_d = b(c)$$

$$T_x = a(c) + b(d) \qquad T_y = a(b) + c(d)$$

$$T_{xy} = \frac{a \cdot (a-1)}{2} + \frac{b \cdot (b-1)}{2} + \frac{c \cdot (c-1)}{2} + \frac{d \cdot (d-1)}{2}$$

$$\text{Anzahl der möglichen Paare} = \frac{N(N-1)}{2}$$

Ermittlung der konkordanten und diskordanten Paare und der Ties in einer 3 x 2-Tabelle:

Beispiel:

		Erforderliche Lernzeit in Jahren (X)		
		1 = bis 1	2 = 2 bis 4	3 = über 4
Zufriedenheit (Y)	1 = gering	12 (a)	12 (b)	5 (c)
	2 = hoch	6 (d)	14 (e)	11 (f)

$$N_c = a(e + f) + b(f) = 12(14 + 11) + 12(11) = 432$$

$$N_d = c(d + e) + b(d) = 5(6 + 14) + 12(6) = 172$$

$$T_x = a(d) + b(e) + c(f) = 12(6) + 12(14) + 5(11) = 295$$

$$T_y = a(b + c) + b(c) + d(e + f) + e(f) = 12(12 + 5) + 12(5) + 6(14 + 11) + 14(11) = 568$$

$$T_{xy} = \frac{a \cdot (a-1)}{2} + \frac{b \cdot (b-1)}{2} + \frac{c \cdot (c-1)}{2} + \frac{d \cdot (d-1)}{2} + \frac{e \cdot (e-1)}{2} + \frac{f \cdot (f-1)}{2}$$

$$= \frac{12 \cdot (12-1)}{2} + \frac{12 \cdot (12-1)}{2} + \frac{5 \cdot (5-1)}{2} + \frac{6 \cdot (6-1)}{2} + \frac{14 \cdot (14-1)}{2} + \frac{11 \cdot (11-1)}{2} = 303$$

$$\text{Anzahl der möglichen Paare} = \frac{N(N-1)}{2} = \frac{60 \cdot 59}{2} = 1.770$$

Ermittlung der konkordanten und diskordanten Paare und der Ties in einer 3 x 3-Tabelle:

	x ₁	x ₂	x ₃
y ₁	a	b	c
y ₂	d	e	f
y ₃	g	h	i

$$N_c = a(e + f + h + i) + b(f + i) + d(h + i) + e(i)$$

$$N_d = c(d + e + g + h) + b(d + g) + f(g + h) + e(g)$$

$$T_x = a(d + g) + b(e + h) + c(f + i) + d(g) + e(h) + f(i)$$

$$T_y = a(b + c) + b(c) + d(e + f) + e(f) + g(h + i) + h(i)$$

$$T_{xy} = \frac{a \cdot (a-1)}{2} + \frac{b \cdot (b-1)}{2} + \frac{c \cdot (c-1)}{2} + \frac{d \cdot (d-1)}{2} + \frac{e \cdot (e-1)}{2} + \frac{f \cdot (f-1)}{2} + \frac{g \cdot (g-1)}{2} + \frac{h \cdot (h-1)}{2} + \frac{i \cdot (i-1)}{2}$$

$$\text{Anzahl der möglichen Paare} = \frac{N(N-1)}{2}$$

Korrelationskoeffizienten für ordinale Variablen:

1) Kendall's Tau-Werte (Tau-a, Tau-b, Tau-c):

1a) Tau-a bzw. τ_a :

$$\tau_a = \frac{N_c - N_d}{\frac{N(N-1)}{2}}$$

wobei:

N_c = Anzahl der konkordanten Paare

N_d = Anzahl der diskordanten Paare

$$\frac{N(N-1)}{2} = \text{Anzahl der möglichen Paare}$$

- Tau-a setzt die Differenz zwischen der Anzahl der konkordanten Paare und der Anzahl der diskordanten Paare ins Verhältnis zur Gesamtzahl der möglichen Paare.

Interpretation:

Tau-a = -1

- Es liegen ausschließlich diskordante Paare vor.
- Es herrscht demnach eine perfekt negative Beziehung zwischen den beiden Variablen.

Tau-a = 0

- Die Anzahl der konkordanten und diskordanten Paare ist gleich groß.
- Demnach sind die Variablen statistisch unabhängig voneinander.

Tau-a = +1

- Es liegen ausschließlich konkordante Paare vor.
- Es herrscht demnach eine perfekt positive Beziehung zwischen den beiden Variablen

- Tau-a berücksichtigt keine Ties (siehe Formel).

- Wenn Ties auftreten, sollte Tau-b (τ_b) verwendet werden, da dieser Ties berücksichtigt.
- Bei Vorliegen vieler Ties wäre Tau-a kleiner als Tau-b ($\tau_a \leq \tau_b$)

Allgemein:

- Der Koeffizient kann für r x c-Tabellen beliebiger Größe berechnet werden.
- Das Assoziationsmaß Tau-a ist am ehesten für Daten geeignet, in denen keine Ties vorliegen (was in der Praxis selten vorkommt), da er die Eigenschaft besitzt die Extremwerte ± 1 nicht erreichen zu können, wenn Ties auftreten.
- Da aber ein niedriger Tau-a-Wert sowohl auf einer schwachen Korrelation als auch auf einem hohen Anteil von Verknüpfungen beruhen kann, ist der Koeffizient ausgesprochen unpopulär.

1b) Tau-b bzw. τ_b :

$$\tau_b = \frac{N_c - N_d}{\sqrt{(N_c + N_d + T_x)(N_c + N_d + T_y)}}$$

- Tau-b wird vorzugsweise für quadratische Tabellen verwendet.
- Bei nicht-quadratischen Tabellen wird Tau-c (τ_c) verwendet.

Allgemein:

- Tau-b berücksichtigt explizit T_x und T_y .
- Tau-b kann die Extremwerte ± 1 *nur* erreichen, wenn die beiden Variablen dieselbe Anzahl von Kategorien haben, d.h., wenn in einer quadratischen Tabelle die Untersuchungseinheiten entlang einer der beiden Diagonalen angeordnet sind, also keine T_x und T_y auftreten
- Tau-b kann die Extremwerte ± 1 *nicht* erreichen, wenn die Tabelle nicht quadratisch ist und infolge dessen T_x und T_y auftreten.
- Diese Eigenschaft ist allerdings kein gewichtiger Nachteil von Tau-b.
- Für 2 x 2-Tabellen berechnet, ergibt Tau-b denselben Zahlenwert wie Phi und r.

1c) Tau-c bzw. τ_c :

$$\tau_c = \frac{N_c - N_d}{\frac{1}{2} N^2 \left(\frac{m-1}{m} \right)}$$

Allgemein:

- Tau-c ist so konzipiert, dass er die Extremwerte ± 1 auch in einer nicht-quadratischen Tabelle erreichen kann.
- Trotz seiner generellen Anwendbarkeit wird Tau-b bevorzugt.

2) Pre-Maß Gamma (γ):

$$\gamma = \frac{N_c - N_d}{N_c + N_d}$$

Allgemein:

- Eine Lösung des Problems, trotz des Auftretens von Ties einen Koeffizienten zu haben, dessen Extremwerte ± 1 sind, besteht darin, die Ties als irrelevant zu betrachten und aus dem Nenner zu entfernen. Genau das geschieht beim Koeffizienten Gamma.
- Gamma ist ein symmetrisches Maß, das für Tabellen beliebiger Größe berechnet werden kann und unabhängig von der Anzahl der Ties zwischen ± 1 variiert.
- Da Gamma Ties ignoriert, die in die Berechnung von Tau-a und Tau-b eingehen, nimmt er regelmäßig höhere Zahlenwerte als diese an.
- Es besteht folgende Beziehung zwischen den Koeffizienten: $|\text{Tau-a}| \leq |\text{Tau-b}| \leq |\text{Gamma}|$.
- Als nachteilig muss gelten, dass der Zahlenwert Gammas häufig dramatisch ansteigt, wenn eine bivariate Tabelle verkleinert und dadurch die Anzahl jener Paare vermehrt wird, die der Koeffizient ignoriert.

3) Somer's d

- Für den seltenen Fall, dass keine Ties vorkommen, sind die Koeffizienten Tau-a, Tau-b, Tau-c, γ und Somer's d ein und dasselbe Maß, das denselben Zahlenwert produziert.
- Beim Auftreten von Ties können die Koeffizienten erheblich voneinander variieren. Dann gilt: $\text{Tau-a} \leq \text{Tau-b} \leq \text{Somer's d} \leq \text{Gamma}$

4) Spearman's Rangkorrelationskoeffizient (r_s):

- Der von Spearman vorgeschlagene Rangkorrelationskoeffizient gründet auf einer anderen Konzeption als die oben behandelten Maßzahlen der ordinalen Assoziation.
- Im Unterschied zu diesen Maßzahlen, bei denen Paare von Untersuchungseinheiten darauf hin betrachtet werden, ob sie konkordant oder diskordant sind, werden beim Spearmanschen r_s **Paare von Rangplätzen im Hinblick auf ihre Differenz betrachtet**.
- In den Sozialwissenschaften besteht häufig ein Interesse, die Beziehung zwischen solchen Rangreihen zu beschreiben, beispielsweise wenn man die Frage untersucht,
 - ob bestimmte Berufe in Land A und B ein gleiches oder unterschiedliches Ansehen genießen,
 - wie ein Vorgesetzter und ein positionsgleicher Mitarbeiter die Qualifikation von Stelleninhabern einschätzen,
 - wie Politiker und Wähler die Popularität oder Dringlichkeit bestimmter politischer Programme bewerten usw.
- Der Rangkorrelationskoeffizient dient also dazu, unterschiedliche Beurteilungen zu vergleichen.

$$r_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)}$$

d_i = Differenz zwischen den Rangplätzen, die die i -te Untersuchungseinheit bezüglich der Variablen X und Y aufweist.

$$d_i = \sum (x_i - y_i)$$

$\sum d_i^2$ = Summe der quadrierten Rangplatzdifferenzen

$$\sum d_i^2 = \sum (x_i - y_i)^2$$

N = Anzahl der rangplatzierten Untersuchungseinheiten

Vorgehensweise bei der Berechnung von r_s :

- Umwandlung der Originaldaten in Rangplätze (1, 2, 3, 4 usw.)
- Dem niedrigsten (höchsten) Wert wird Rangplatz 1 zugewiesen.
- Wenn Ties (gleiche Zahlenwerte) auftreten, dann werden den Werten gleiche Rangplätze zugewiesen.

Beispiel:

$r_s = +1$

- Es liegt eine perfekt positive Rangkorrelation vor.
- Es herrschen keine Differenzen zwischen den Rangplätzen.

Schüler/in	Physiknote von Lehrer 1 (x_i)		Physiknote von Lehrer 2 (y_i)	
	Originaldaten	Rangplätze	Originaldaten	Rangplätze
A	2+	1	2+	1
B	2	2	2	2
C	3	3	3	3
D	5	4	5	4

- Hierbei wurde dem niedrigsten Wert Rangplatz 1 zugewiesen.

Berechnung von r_s :

Schüler/in	$d_i = (x_i - y_i)$	$\sum d_i^2 = \sum (x_i - y_i)^2$
A	0	0
B	0	0
C	0	0
D	0	0
Σ	0	0

$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{N(N^2 - 1)} = 1 - \frac{6 \cdot 0}{4(4^2 - 1)} = 1 - \frac{0}{60} = 1 - 0 = +1$$

Beispiel:

$r_s = -1$

- Es liegt eine perfekt negative Rangkorrelation vor.
- Es herrschen maximale Differenzen zwischen den Rangplätzen.

Schüler/in	Physiknote von Lehrer 1 (x_i)		Physiknote von Lehrer 2 (y_i)	
	Originaldaten	Rangplätze	Originaldaten	Rangplätze
A	1	1	4-	4
B	3+	2	4	3
C	4	3	3+	2
D	4-	4	1	1

- Hierbei wurde dem niedrigsten Wert Rangplatz 1 zugewiesen.

Berechnung von r_s :

Schüler/in	$d_i = (x_i - y_i)$	$\sum d_i^2 = \sum (x_i - y_i)^2$
A	-3	9
B	-1	1
C	1	1
D	3	9
Σ	0	20

$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{N(N^2 - 1)} = 1 - \frac{6 \cdot 20}{4(4^2 - 1)} = 1 - \frac{120}{60} = 1 - 2 = -1$$

Beispiel:

$r_s = 0$

- Es liegt keine Rangkorrelation vor.
- In Bezug auf die Rangplatzdifferenzen ist keine Systematik erkennbar.

Schüler/in	Physiknote von Lehrer 1 (x_i)		Physiknote von Lehrer 2 (y_i)	
	Originaldaten	Rangplätze	Originaldaten	Rangplätze
A	1	1	3+	2
B	3+	2	4-	4
C	4	3	1	1
D	4-	4	4	3

- Hierbei wurde dem niedrigsten Wert Rangplatz 1 zugewiesen.

Berechnung von r_s :

Schüler/in	$d_i = (x_i - y_i)$	$\sum d_i^2 = \sum (x_i - y_i)^2$
A	-1	1
B	-2	4
C	2	4
D	1	1
Σ	0	10

$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{N(N^2 - 1)} = 1 - \frac{6 \cdot 10}{4(4^2 - 1)} = 1 - \frac{60}{60} = 1 - 1 = 0$$

Beispiel:

Umgang mit Ties

SchülerIn	Physiknote von Lehrer 1 (x _i)		Physiknote von Lehrer 2 (y _i)	
	Originaldaten	Rangplätze	Originaldaten	Rangplätze
A	1+	1	1+	1
B	<u>2</u>	Tie	2	2
C	<u>2</u>		2-	3
D	3	4	3	4

- Hierbei wurde dem niedrigsten Wert Rangplatz 1 zugewiesen.
- Man errechnet den gemeinsamen Rangplatz, indem man die Verknüpfung ignoriert und die betreffenden Fälle so behandelt, als besäßen sie unterschiedliche Rangplätze:
 - SchülerIn B = Rangplatz 2 und SchülerIn C = Rangplatz 3
oder:
 - SchülerIn B = Rangplatz 3 und SchülerIn C = Rangplatz 2
- Darauf hin errechnet man den Mittelwert der beiden Rangplätze:

$$\frac{(2+3)}{2} = 2,5 \Rightarrow \text{Dieser Rangplatz wird nun beiden Schülern zugewiesen}$$

- Beim Auftreten von Ties weist man den verbundenen Untersuchungseinheiten das arithmetische Mittel derjenigen Rangplätze zu, die man zugewiesen hätte, wenn keine Ties aufgetreten wären.

Berechnung von r_s:

Schüler/in	d _i = (x _i - y _i)	Σ d _i ² = Σ(x _i - y _i) ²
A	0	0
B	0,5	0,25
C	-0,5	0,25
D	0	0
Σ	0	0,5

$$r_s = 1 - \frac{6 \cdot \sum d_i^2}{N(N^2 - 1)} = 1 - \frac{6 \cdot 0,5}{4(4^2 - 1)} = 1 - \frac{3}{60} = 1 - 0,05 = 0,95$$