Ordinale Assoziationskoeffizienten. Mengenverknüpfungen und Paartypen

HEINZ RENN, HAMBURG

Zusammenfassung: *Gezeigt wird ein Routineverfahren, mit dessen Hilfe sich die Bestimmung der Anzahl von Paartypen anhand von Mengenverknüpfungen auch für sehr große Mengen von Untersuchungseinheiten, die gewöhnlich als Häufigkeiten in einer Kontingenztabelle zusammengestellt sind, einfach durchführen läßt. Bei dieser Gelegenheit wird auch die wirklichkeitsfremde Annahme fallen gelassen, daß für die beiden in Beziehung gesetzten Merkmale jeweils eine echte Rangreihe vorliegen muß. Neben der Ermittlung der Anzahl konkordanter Paare und der Anzahl diskordanter Paare wird das vorgestellte Verfahren somit auf die Ermittlung der Anzahl sog. ,gebundener' Paare ausgeweitet und ein entsprechender Koeffizient vorgestellt.*

1. Einleitung und Ausgangslage

In einem Beitrag dieser Zeitschrift wird mit dem KENDALLschen τ-Koeffizienten "ein alternativer Korrelationskoeffizent" vorgestellt (Schütze 1993). Die Darstellung der Korrelationsproblematik anhand dieses Koeffizienten, eines Assoziationskoeffizienten für ordinalskalierte Merkmale[1], habe – so die Verfasserin – didaktische Vorteile. Zwar fiele er „aus dem Rahmen der in der Schule üblicherweise unterrichteten Statistik" heraus, er eigne sich aber vorzüglich dazu „einen ersten anschaulichen Einblick in die Korrelationsrechnung zu geben und den Schülern den Einstieg zu erleichtern" (S. 19).

Zur Illustration wird ein einfaches Beispiel präsentiert, in dem für sechs Schüler einer Schulklasse die Leistungen in Musik und Mathematik anhand der in diesen Fächern erzielten Schulnoten verglichen werden. Die Schulnoten werden angemessen als eine Abfolge von Rängen und nicht als eine metrisch skalierte Meßreihe interpretiert. Zum Vergleich derartiger ordinalskalierter Meßreihen ist der Produkt-Moment-Korrelationskoeffizient r – wie die Verfasserin zutreffend anmerkt – nicht geeignet.

In der Tat läßt sich anhand dieses Beispiels die Logik der Korrelationsproblematik recht gut erläutern. Allerdings ist das Beispiel unrealistisch, insbesondere wenn es darum geht, für größere n den τ-Koeffizienten zu berechnen. Gleichgültig, ob sog. 'Unordnungen' (engl. „disarrays") auszuzählen sind oder beim Paarvergleich zu bestimmen ist, ob konkordante oder diskordante Paare vorliegen, im Falle von $n = 6$ ist dies noch relativ leicht zu bewältigen, da nur $\frac{1}{2}n(n-1) = \frac{1}{2}6(6-1) = 15$ mögliche Zuordnungen zu beurteilen bzw. Paare ihrem Typus nach zu bestimmen sind. Bezieht man hingegen den Vergleich schon auf eine Schulklasse normaler Größe mit $n = 25$ Schülern, so sind bereits 300 mögliche Zuordnungen zu beurteilen bzw. Paare ihrem Typus nach zu bestimmen. Die Anwendung des Kendallschen Koeffizienten kann nun nicht nur auf den Vergleich von Noten zweier Fächer in Schulklassen beschränkt werden, er wird vielfach auch in der Analyse von Erhebungen der Umfrageforschung eingesetzt. Nehmen wir an, daß – wie in unserem Beispiel, das im folgenden herangezogen wird (vgl. *Tabelle 1*, S. 36) – 715 Personen befragt werden, so sind 255.255 mögliche Zuordnungen zu beurteilen bzw. Paare ihrem Typus nach zu bestimmen. Die Berechnung des Koeffizienten in der vorgeschlagenen Art und Weise ist dann praktisch unmöglich. Durch das von der Verfasserin gewählte Beispiel wird so der falsche Eindruck vermittelt, der Koeffizient lasse sich nur bei kleinen Mengen von Untersuchungseinheiten anwenden. Darüber hinaus enthält das von der Verfasserin vorgestellte Beispiel eine weitere realitätsferne Vereinfachung, nämlich die Annahme, daß alle Schüler sich nach ihren Noten unterscheiden – und somit die Schüler in jedem der beiden miteinander verglichenen Fächer jeweils in eine echte Rangreihe gebracht werden können.

Die Realitätsferne eines Beispiels hat beträchtliche Implikationen für den didaktischen Erfolg: Sie suggeriert – zumindest implizit – auch eine Realitätsferne des Verfahrens. Nicht nur aus dem Schulunterricht, sondern auch aus der universitären Vermittlung der Statistik in den anwendenden Fächern ist bekannt, daß der Eindruck, man könne im normalen Anwendungsfall mit einem in Frage stehenden Verfahren nicht viel anfangen, da die zu machenden Annahmen in der Realität ohnehin nicht erfüllt seien, geradezu 'tödlich' für die Lernmotivation ist.

Wir zeigen im folgenden ein Routineverfahren, mit dessen Hilfe sich die Bestimmung der Anzahl von Paartypen anhand von Mengenverknüpfungen auch für sehr große Mengen von Untersuchungseinheiten, die gewöhnlich als Häufigkeiten in einer Kontingenztabelle zusammengestellt sind, einfach durchführen läßt. Dieses Verfahren ist nicht neu, es hat sogar in Lehrbücher Eingang gefunden. Dennoch scheint es allgemein nicht bekannt zu sein. Es erscheint so gerechtfertigt, das Verfahren hier in didaktisch aufbereiteter Darstellung zu präsentieren. Darüber hinaus werden dabei einige Formalisierungen eingeführt, die über die in der Lehrbuchliteratur übliche Darstellungsweise hinausgehen.

Bei dieser Gelegenheit wird auch die wirklichkeitsfremde Annahme fallengelassen, daß für die zu vergleichenden Merkmale jeweils eine echte Rangreihe vorliegen muß. Unter dieser Annahme handelt es sich bei dem dargestellten Kendallschen τ um einen speziellen τ - Koeffizienten, nämlich um den Assoziationskoeffizienten τ_a. Dieser Koeffizient wird berechnet als die Differenz zwischen der Anzahl der konkordanten Paare, C , und der Anzahl der diskordanten Paare, D , bezogen auf die Anzahl der möglichen Paare [vgl. Forme(l)] [2]:

$$\tau_a = \frac{C - D}{\frac{1}{2}n(n-1)} \tag{1}$$

Der Kendallsche τ -Koeffizient, bei dem diese Annahme *nicht* gemacht werden muß, ist der Assoziationskoeffizient τ_b (vgl. Kendall 1975, S. 34 - 47). Dieser Koeffizient wird weiter unten dargestellt. Die bei diesem Koeffizienten zu berücksichtigen 'Bindungen' (engl. 'ties'). und die Ermittlung sogenannter 'gebundener' Paare werden wir ebenfalls ausführlich behandeln.

Merkmal y: Subjektive Einschätzung der Wohnungsgröße	Merkmal x: Zufriedenheit mit der Wohnung			$\sum_{j=1}^{3} f_{ij}$
	sehr zufrieden $[x_1]$	im großen und ganzen zufrieden $[x_2]$	nicht zufrieden $[x_3]$	
viel Platz $[y_1]$	112	46	41	199
ausreichend $[y_2]$	61	143	60	264
beengt $[y_3]$	30	77	145	252
$\sum_{i=1}^{3} f_{ij}$	203	266	246	715

Tabelle 1: Zufriedenheit mit der Wohnung und subjektive Einschätzung der Wohnungsgröße

2. Paarbildung durch Mengenverknüpfungen

Nehmen wir das folgende Beispiel. In einer Untersuchung werden 715 Personen nach ihren Wohnverhältnissen befragt. Dabei wird u. a. die subjektive Einschätzung der Wohnungsgröße und die Zufriedenheit mit der Wohnung erhoben. *Tabelle 1* enthält die entsprechende bivariate Häufigkeitsverteilung. Vermutet wird ein Zusammenhang insoweit, als die ‚Zufriedenheit mit der Wohnung' (Merkmal x) und die ‚subjektive Einschätzung der Wohnungsgröße' (Merkmal y) sich gegenseitig bedingen. *Abbildung 1* (S. 37) zeigt die entsprechende allgemeine Struktur einer Kontingenztabelle mit k Zeilen und l Spalten und den Häufigkeiten der einzelnen Tabellenfelder f_{ij}, wobei $i = 1,\ldots,k$ und $j = 1,\ldots,l$. Im vorliegenden Fall handelt es sich um eine Kontingenztabelle mit k = 3 Zeilen und l = 3 Spalten, somit um eine sog. 3x3-Tabelle.

					x				
		1	2	$\dots$	j	$\dots$	$l-1$	l	$\sum_{j=1}^{l} f_{ij} = f_{i.}$
	1	f_{11}	f_{12}	$\dots$	f_{1j}	$\dots$	f_{1l-1}	f_{1l}	$f_{1.}$
	2	f_{21}	f_{22}	$\dots$	f_{2j}	$\dots$	f_{2l-1}	f_{2l}	$f_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
y	i	f_{i1}	f_{i2}	$\dots$	f_{ij}	$\dots$	f_{il-1}	f_{il}	$f_{i.}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
	$k-1$	f_{k-11}	f_{k-12}	$\dots$	f_{k-1j}	$\dots$	f_{k-1l-1}	f_{k-1l}	$f_{k-1.}$
	k	f_{k1}	f_{k2}		f_{kj}		f_{kl-1}	f_{kl}	$f_{k.}$
$\sum_{i=1}^{k} f_{ij} = f_{.j}$		$f_{.1}$	$f_{.2}$	$\dots$	$f_{.j}$	$\dots$	$f_{.l-1}$	$f_{.l}$	$\sum_{i=1}^{k} \sum_{j=1}^{l} f_{ij} = N$

Abbildung 1: Allgemeine Struktur einer Kontingenztabelle

Nicht erforderlich ist, daß – wie im vorliegenden Beispiel – die Anzahl der Zeilen und die der Spalten übereinstimmen, d.h. $k = l$ ist. Das im weiteren Dargestellte gilt auch für Tabellen mit nicht übereinstimmender Anzahl der Zeilen und der Spalten, d.h. auch für Tabellen mit $k \neq l$ [3)].

Zur Ermittlung der Art eines Paares von Untersuchungseinheiten können wir nun beispielsweise einen Befragten A aus der Menge der 112 Befragten im Tabellenfeld $[x_1 y_1]$ ('sehr zufrieden'/'viel Platz') mit einem Befragten B aus der Menge der 143 Befragten im Tabellenfeld $[x_2 y_2]$('im großen und ganzen zufrieden'/'ausreichend') hinsichtlich der jeweiligen Ordnung beurteilen, in die dieses Paar A/B bezogen auf die beiden in Frage stehenden Merkmale gebracht wird. Für das Paar A/B trifft bezogen auf die Ausprägungen der beiden Merkmale folgendes zu. Für das Merkmal x 'Zufriedenheit mit der Wohnung' gilt:

- A ist 'sehr zufrieden' $[x_1]$> B ist 'im großen und ganzen zufrieden' $[x_2]$.

Zugleich gilt für das Merkmal y 'Subjektive Einschätzung der Wohnungsgröße':

- A hat bezüglich der Wohnungsgröße die subjektive Einschätzung 'viel Platz' $[y_1]$ > B hat demgegenüber die subjektive Einschätzung 'ausreichend' $[y_2]$.

Hieraus ergibt sich:

- A besitzt mit $[x_1]$ 'sehr zufrieden' bezogen auf das Merkmal x einen höheren Rang als B mit $[x_2]$ 'im großen und ganzen zufrieden' und zugleich gilt, daß
- A mit $[y_1]$ 'viel Platz' bezogen auf das Merkmal y ebenfalls einen höheren Rang als B mit $[y_2]$ 'ausreichend' einnimmt.

Das Paar A/B ist somit ein *konkordantes* Paar:

$\left(x_1^A > x_2^B\right) \wedge \left(y_1^A > y_2^B\right)$.

Dieses Ergebnis gilt nicht allein für das Paar A/B, sondern auch für alle anderen Paare von Untersuchungseinheiten, die man zwischen Untersuchungseinheiten dieser beiden Tabellenfelder, $[x_1 y_1]$ und $[x_2 y_2]$, bilden kann. Alle diese Paare haben in der Tabelle eine äquivalente Lage und sind daher - wie das Paar A/B – konkordante Paare:

$(x_1 > x_2) \wedge (y_1 > y_2)$. Insgesamt sind dies

$$f_{11} \cdot f_{22} = 112 \cdot 143 = 16.016$$

Paare. Wir sind somit in der Lage mit einer einzigen Paarbewertung eine große Anzahl von Paaren einem bestimmten Paartyp zuzuordnen. Allgemein gilt für konkordante Paare:

$$(x_j > x_{j^*}) \wedge (y_i > y_{i^*}).$$

Bezeichnen wir die einzelnen Untersuchungseinheiten, die sich im Tabellenfeld $[x_1 y_1]$ befinden, mit X_1Y_1 und deren Menge mit F_{11} sowie die einzelnen Untersuchungseinheiten, die sich im Tabellenfeld $[x_2 y_2]$ befinden, mit X_2Y_2 und deren Menge mit F_{22}, so erhalten wir durch die Verknüpfung dieser beiden Teilmengen die Menge der konkordanten Paare, die bezogen auf die entsprechenden Untersuchungseinheiten gebildet werden können. Es handelt sich bei dieser Mengenoperation um eine zweistellige Verknüpfung der beiden Teilmengen, um deren Produktmenge bzw. ihr Kreuzprodukt (auch kartesisches Produkt genannt):

$$F_{11} \times F_{22} = \{(X_1Y_1, X_2Y_2) \mid X_1Y_1 \in F_{11} \text{ und } X_2Y_2 \in F_{22}\}$$

Die 9 Tabellenfelder können in dieser Weise jeweils paarweise zueinander in Beziehung gesetzt werden. Dabei geht es zunächst darum, die Art des Paartyps zu bestimmen, um sodann die Anzahl zu berechnen, mit der der jeweilige Paartyp vorkommt. Die entsprechenden Kreuzprodukte lauten allgemein:

$$F_{IJ} \times F_{I^*J^*} =$$
$$\{(X_JY_I, X_{J^*}Y_{I^*}) \mid X_JY_I \in F_{IJ} \text{ und } X_{J^*I^*}Y_2 \in F_{I^*J^*}\}$$

In *Abbildung 2* werden allgemeine Definitionen der Paartypen gegeben. Diese Definitionen werden in den folgenden Abschnitten im einzelnen erläutert.

3. Anzahl konkordanter Paare (C)

Gehen wir aus von der Begriffsbestimmung eines konkordanten Paares [vgl. *Abbildung 2, Zeile (1), Spalte (1)*]. Zwei äquivalente Definitionen liegen vor. Die zuerst genannte: $(x_j > x_{j^*}) \wedge (y_i > y_{i^*})$ ist diejenige, die bereits in *Abschnitt 2* verwendet wurde. Kennzeichnend ist, daß sowohl beim Merkmal x als auch beim Merkmal y von der Ausprägung mit dem höheren Rang ausgegangen wird, so daß die im Vergleich verwendete Relation 'größer als' ('>') lautet. Man hätte auch umgekehrt verfahren können: $(x_j < x_{j^*}) \wedge (y_i < y_{i^*})$.

Bezeichnung des Paartyps	Definition (1)	Anzahl (2)
(1) konkordant	$(x_j > x_{j^*}) \wedge (y_i > y_{i^*})$ bzw. $(x_j < x_{j^*}) \wedge (y_i < y_{i^*})$	C
(2) diskordant	$(x_j < x_{j^*}) \wedge (y_i > y_{i^*})$ bzw. $(x_j > x_{j^*}) \wedge (y_i < y_{i^*})$	D
(3) auf x gebunden	$(x_j = x_{j^*}) \wedge (y_i > y_{i^*})$ bzw. $(x_j = x_{j^*}) \wedge (y_i < y_{i^*})$	T_x
(4) auf y gebunden	$(x_j > x_{j^*}) \wedge (y_i = y_{i^*})$ bzw. $(x_j < x_{j^*}) \wedge (y_i = y_{i^*})$	T_y
(5) auf x und y gebunden	$(x_j = x_{j^*}) \wedge (y_i = y_{i^*})$	T_{xy}

Abbildung 2: Allgemeine Definition der Paartypen

Kennzeichnend für diese Definition ist, daß sowohl beim Merkmal x als auch beim Merkmal y nunmehr von der Ausprägung mit dem niedrigeren Rang ausgegangen wird, so daß die im Vergleich verwendete Relation 'kleiner als' ('<') lautet. Beide Definition sind gleichwertig. Im folgenden wird die zuerst angeführte Definition verwendet.

In *Abbildung 3* (S. 39) wird – unserem Beispiel gemäß – eine 3x3-Tabelle in allgemeiner Form präsentiert. Unterstellt wird, daß die Ausprägungen des Merkmals x eine absteigende Rangfolge bilden: $x_1 > x_2 > x_3$ (im Beispiel: 'sehr zufrieden' > 'im großen und ganzen zufrieden' > 'nicht zufrieden'). Gleiches gelte auch für das Merkmal y: $y_1 > y_2 > y_3$ (im Beispiel: 'viel Platz' > 'ausreichend' > 'beengt').

In *Abbildung 3 Teil (1)* ist die Bestimmung der Anzahl der konkordanten Paare dargestellt.

		x		
		x_1	x_2	x_3
y	y_1	f_{11}	f_{12}	f_{13}
	y_2	f_{21}	f_{22}	f_{23}
	y_3	f_{31}	f_{32}	f_{33}

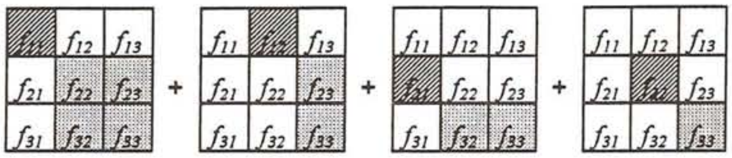
Es gilt:

$x_1 > x_2 > x_3$

$y_1 > y_2 > y_3$

(1) Bestimmung der Anzahl der *konkordanten* Paare: C

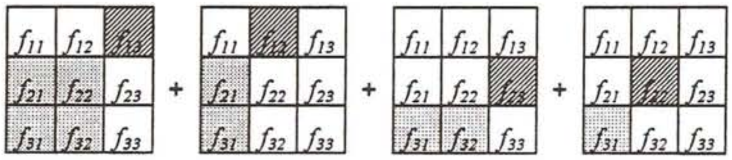
$C =$



$$C = f_{11}(f_{22}+f_{23}+f_{32}+f_{33}) + f_{12}(f_{23}+f_{33}) + f_{21}(f_{32}+f_{33}) + f_{22}(f_{33})$$

(2) Bestimmung Anzahl der *diskordanten* Paare: D

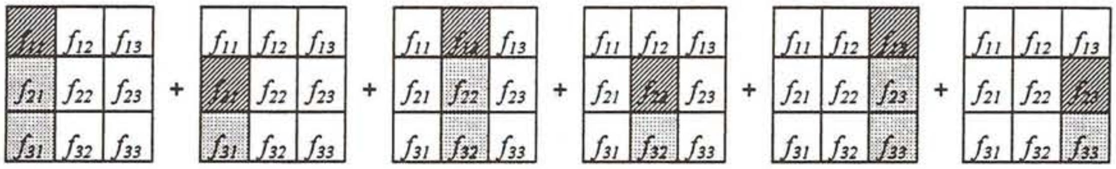
$D =$



$$D = f_{13}(f_{21}+f_{22}+f_{31}+f_{32}) + f_{12}(f_{21}+f_{31}) + f_{23}(f_{31}+f_{32}) + f_{22}(f_{31})$$

(3) Bestimmung der Anzahl der auf *x gebundenen* Paare: T_x

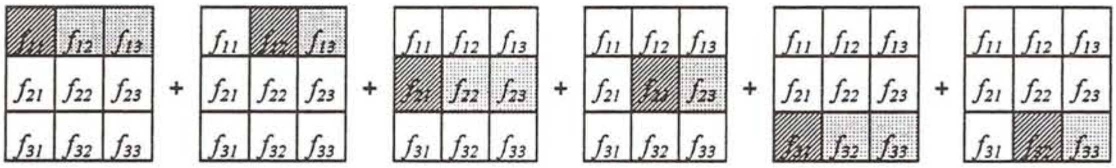
$T_x =$



$$T_x = f_{11}(f_{21}+f_{31}) + f_{21}(f_{31}) + f_{12}(f_{22}+f_{32}) + f_{22}(f_{32}) + f_{13}(f_{23}+f_{33}) + f_{23}(f_{33})$$

(4) Bestimmung Anzahl der auf *y gebundenen Paare*: T_y

$T_y =$



$$T_y = f_{11}(f_{12}+f_{13}) + f_{12}(f_{13}) + f_{21}(f_{22}+f_{23}) + f_{22}(f_{23}) + f_{31}(f_{32}+f_{33}) + f_{32}(f_{33})$$

(5) Bestimmung Anzahl der *auf x und y gebundenen Paare*: T_{xy}

$$T_{xy} = \frac{1}{2}\left[f_{11}(f_{11}-1) + f_{12}(f_{12}-1) + f_{13}(f_{13}-1) + f_{21}(f_{21}-1) + f_{22}(f_{22}-1) + f_{23}(f_{23}-1) + f_{31}(f_{31}-1) + f_{32}(f_{32}-1) + f_{33}(f_{33}-1)\right]$$

Abbildung 3: Bestimmung der Anzahl einzelner Paartypen in einer Kontingenztabelle am Beispiel einer 3x3-Tabelle

Wir gehen aus vom ersten Tabellenfeld in der *linken* oberen Ecke $[x_1 y_1]$ und 'paaren' die in diesem Tabellenfeld befindlichen Untersuchungseinheiten mit den Untersuchungseinheiten, die sich in den Tabellenfeldern befinden, die sowohl *unterhalb* als auch *rechts* von Tabellenfeld $[x_1 y_1]$ liegen. Dies sind die Tabellenfelder $[x_2 y_2]$, $[x_3 y_2]$, $[x_2 y_3]$ und $[x_3 y_3]$. Die entsprechenden Paarbildungen ergeben konkordante Paare:

$$[x_1 y_1] \text{ mit } [x_2 y_2] \Rightarrow (x_1 > x_2) \wedge (y_1 > y_2)$$

$$[x_1 y_1] \text{ mit } [x_3 y_2] \Rightarrow (x_1 > x_3) \wedge (y_1 > y_2)$$

$$[x_1 y_1] \text{ mit } [x_2 y_3] \Rightarrow (x_1 > x_2) \wedge (y_1 > y_3)$$

$$[x_1 y_1] \text{ mit } [x_3 y_3] \Rightarrow (x_1 > x_3) \wedge (y_1 > y_3)$$

Die Anzahl dieser Paare kann durch Multiplikation der Häufigkeit des Tabellenfeldes $[x_1 y_1]$ mit der Summe der Häufigkeiten der Tabellenfelder $[x_2 y_2]$, $[x_3 y_2]$, $[x_2 y_3]$ und $[x_3 y_3]$ errechnet werden. Es ergibt sich allgemein

$$f_{11}\,(f_{22}+f_{23}+f_{32}+f_{33}),$$

für unser Beispiel: 112 (143+60+77+145) = 47.600 Paare [vgl. *Abbildung 3 (1), erstes Tabellenschema*]. In diesem Tabellenschema – wie auch in den folgenden – ist das Tabellenfeld, von dem bei der Paarbildung ausgegangen wird, schraffiert, während die mit diesem Tabellenfeld ‚verpaarten' Tabellenfelder getönt sind.

Es gibt jedoch noch weitere konkordante Paare. Wir fahren deshalb fort und 'paaren' für jedes weitere Tabellenfeld die in diesem befindlichen Untersuchungseinheiten mit den Untersuchungseinheiten, die sich in Tabellenfeldern befinden, die sowohl *unterhalb* der Zeile als auch *rechts* der Spalte liegen, in der sich das jeweils betrachtete Tabellenfeld befindet [vgl. *Abbildung 3 (1), zweites, drittes und viertes Tabellenschema*]. Weitere konkordante Paare gibt es nicht, da den jeweils letzten Tabellenfeldern einer Zeile sowie allen Tabellenfeldern der letzten Zeile keine zur Bildung konkordanter Paare geeignete Tabellenfelder entsprechen. Die Anzahl der konkordanten Paare ist

$$\begin{aligned} C \;=\; & f_{11}\,(f_{22}+f_{23}+f_{32}+f_{33}) + f_{12}\,(f_{23}+f_{33}) \\ & + f_{21}\,(f_{32}+f_{33}) + f_{22}\,(f_{33}) \end{aligned}$$

Für unser Beispiel:

$$\begin{aligned} C \;=\; & 112(143+60+77+145) + 46(60+145) \\ & + 61(77+145) + 143(145) \\ =\; & 91.307 \end{aligned}$$

konkordante Paare.

Das hier am Beispiel der 3x3-Tabelle vorgestellte Berechnungsverfahren ist in analoger Weise auf Kontingenztabellen jeder anderen Größe anzuwenden. Bezogen auf die in *Abbildung 2* dargestellte allgemeine Struktur einer Kontingenztabelle mit k Zeilen und l Spalten ergibt sich die allgemeine Formel für die Anzahl der konkordanten Paare [vgl. Formel (2)]:

$$C = \sum_{i=1}^{k-1}\sum_{j=1}^{l-1}\left(f_{ij} \sum_{s=i+1}^{k}\sum_{r=j+1}^{l} f_{sr} \right) \tag{2}$$

4. Anzahl diskordanter Paare (*D*)

Wir gehen jetzt aus von der Begriffsbestimmung eines diskordanten Paares [vgl. *Abbildung 2, Zeile (2), Spalte (1)*]. Diskordanz ist gegenüber der Konkordanz dadurch definiert, daß die Rangrelationen, nach denen die beiden Untersuchungseinheiten eines Paares in Beziehung stehen, gegenläufig sind. Zwei äquivalente Definitionen sind dabei möglich:

$$(x_j < x_{j^*}) \wedge (y_i > y_{i^*}) \quad \text{und} \quad (x_j > x_{j^*}) \wedge (y_i < y_{i^*}).$$

Im weiteren wird die zuerst genannte Definition verwendet, die beim Merkmal x von der Ausprägung mit dem niedrigeren, beim Merkmal y von der Ausprägung mit dem höheren Rang ausgeht.

In *Abbildung 3, Teil (2),* ist die Bestimmung der Anzahl der diskordanten Paare dargestellt. Wegen der Gegenläufigkeit der beiden Relationen in der Definition eines diskordanten Paares ('<' gegenüber '>') gehen wir jetzt aus vom ersten Feld in der *rechten* oberen Ecke der Tabelle $[x_3 y_1]$ und 'paaren' die in diesem Tabellenfeld befindlichen Untersuchungseinheiten mit den Untersuchungseinheiten, die sich in Tabellenfeldern befinden, die sowohl *unterhalb* als auch *links* von Tabellenfeld $[x_3 y_1]$ liegen [vgl. *Abbildung 3, (2), erstes Tabellenschema*]. Die Tabelle wird in dieser Weise weiter bearbeitet [vgl. *Abbildung 3 (2), zweites, drittes und viertes Tabellenschema*]. Die Anzahl der diskordanten Paare ist

$$\begin{aligned} D \;=\; & f_{13}\,(f_{21}+f_{22}+f_{31}+f_{32}) + f_{12}\,(f_{21}+f_{31}) \\ & + f_{23}\,(f_{31}+f_{32}) + f_{22}\,(f_{31}) \end{aligned}$$

Für unser Beispiel:

$$\begin{aligned} D \;=\; & 41(61+143+30+77) + 46(61+30) \\ & + 60(30+77) + 143(30) \\ =\; & 27.647 \end{aligned}$$

diskordante Paare.

Für Kontingenztabellen jeder anderen Größe lautet die allgemeine Formel der Anzahl der diskordanten Paare [vgl. Formel (3)]:

$$D = \sum_{i=1}^{k-1} \sum_{j=2}^{l} \left(f_{ij} \sum_{s=i+1}^{k} \sum_{r=1}^{j-1} f_{sr} \right) \tag{3}$$

5. Das Problem der „Bindungen"

Die Normierung von τ_a erfolgt auf einen Bereich von $\tau_a = -1{,}0$ bei perfektem negativen Zusammenhang als Untergrenze bis zu $\tau_a = +1{,}0$ bei perfektem positiven Zusammenhang als Obergrenze:

$$-1{,}0 \le \tau_a \le +1{,}0 .$$

Ein solcher Wertebereich für τ_a ergibt sich gemäß Formel (1) allerdings nur dann, wenn ausschließlich konkordante *und* diskordante Paare vorkommen. Dies ist aber nur bei echten Rangreihen der Fall. Echte Rangreihen liegen in vielen Fällen nicht vor. Oft haben verschiedene Untersuchungseinheiten ein und denselben Rang: $n > k$ bzw. $n > l$. Dies gilt insbesondere für Kontingenztabellen.

Es gibt somit Paare von Untersuchungseinheiten, die sich bezüglich der Ausprägungen eines der Merkmale oder gar beider Merkmale nicht voneinander unterscheiden (vgl. *Tabelle 1*):

- Alle Untersuchungseinheiten, die sich in einer bestimmten *Spalte* der Tabelle befinden, unterscheiden sich bezogen auf das Merkmal *x nicht*. Beispielsweise unterscheiden sich beim Merkmal *x* ‚Zufriedenheit mit der Wohnung' 203 Untersuchungseinheiten insoweit *nicht*, als diese alle mit ihrer Wohnung ‚sehr zufrieden' sind.
- Alle Untersuchungseinheiten, die sich in einer bestimmten *Zeile* der Tabelle befinden, unterscheiden sich bezogen auf das Merkmal *y nicht*. Beispielsweise unterscheiden sich beim Merkmal *y* ‚Subjektive Einschätzung der Wohnungsgröße' 199 Untersuchungseinheiten insoweit *nicht*, als diese alle der Meinung sind, ‚viel Platz' in ihrer Wohnung zu haben.
- Alle Untersuchungseinheiten, die sich in einem bestimmten *Tabellenfeld* befinden, unterscheiden sich jeweils bezogen auf *beide* Merkmale *nicht*. Sowohl beim Merkmal *x* als auch beim Merkmal *y* unterscheiden sich z. B. 112 Untersuchungseinheiten insoweit *nicht*, als diese alle einmal mit ihrer Wohnung ‚sehr zufrieden' sind und auch alle der Meinung sind, ‚viel Platz' in ihrer Wohnung zu haben.

Es liegen sog. 'Bindungen' (engl. 'ties') liegen vor. Dies bedeutet, daß die möglichen Paare nicht mehr ausschließlich als konkordant oder diskordant eingestuft werden können. Neben den konkordanten und diskordanten Paaren gibt es noch andere Paartypen, sog. gebundene Paare [vgl. *Abbildung 2, Zeilen (3), (4) und (5)*]. Die Anzahl der auf *x* gebundenen Paare bezeichnen wir allgemein mit T_x, die der auf *y* gebundenen Paare mit T_y und die auf *x und y* gebundenen Paare mit T_{xy} [vgl. *Abbildung 2, Spalte (2)*]. Die Anzahl der möglichen Paare ist so sie Summe der Anzahl der konkordanten, der Anzahl der diskordanten, der Anzahl der auf *x* gebundenen Paare, der Anzahl der auf *y* gebundenen Paare und der Anzahl der auf *x* und *y* gebundenen Paare [vgl. Formel (4)]:

$$C + D + T_x + T_y + T_{xy} = \tfrac{1}{2} n (n-1) \tag{4}$$

Das Vorliegen gebundener Paare hat Auswirkungen auf die Normierung von τ_a:

- Im Falle eines perfekten gleichsinnigen Zusammenhangs liegen zwar keine diskordanten Paare vor, allerdings ist das Maximum der konkordanten Paare kleiner als die Anzahl der möglichen Paare:

  $D = 0$ und $C < \tfrac{1}{2} n (n-1)$.

- Im Falle eines perfekten gegenläufigen Zusammenhangs liegen zwar keine konkordanten Paare vor, allerdings ist das Maximum der diskordanten Paare kleiner als die Anzahl der möglichen Paare:

  $C = 0$ und $D < \tfrac{1}{2} n (n-1)$.

Liegen gebundene Paare vor, so erreicht der Koeffizient τ_a auch bei einem perfekten gegenläufigen Zusammenhang nicht die Untergrenze -1,0 und bei einem perfekten gleichsinnigen Zusammenhang nicht die Obergrenze +1,0:

$$-1{,}0 < \tau_a < +1{,}0 .$$

Der Koeffizient τ_a unterschätzt somit abhängig von der Anzahl der gebundenen Paare die Stärke des Zusammenhangs. Die Anzahl der gebundenen Paare muß daher bei der Normierung der Differenz der Anzahl der konkordanten und diskordanten Paare $(C - D)$ berücksichtigt werden. Eine derartige Berücksichtigung gebundener Paare führt zum Koeffizienten τ_b, der weiter unten ableitet wird.

Zunächst soll die Bestimmung der jeweiligen Anzahl der drei vorgestellten Typen gebundener Paare dargestellt werden.

6. Gebundene Paare

Wie im Falle der konkordanten und der diskordanten Paare ist die Anzahl der drei Typen gebundener Paare T_x, T_y und T_{xy}, jeweils zu bestimmen.

6.1 Anzahl auf x gebundener Paare (T_x)

Wir gehen aus von der Begriffsbestimmung eines auf x gebundenen Paares [vgl. *Abbildung 2, Zeile (3), Spalte (1)*]. Wiederum liegen zwei äquivalente Definitionen vor:

$$(x_j = x_{j^*}) \wedge (y_i > y_{i^*}) \text{ und } (x_j = x_{j^*}) \wedge (y_i < y_{i^*}).$$

Verwendet wird im folgenden die zuerst genannte Definition, die beim Merkmal y von der Ausprägung mit dem höherem Rang ausgeht.

In *Abbildung 3, Teil (3)* ist die Bestimmung der Anzahl der auf x gebundenen Paare dargestellt. Da in diesem Falle die Untersuchungseinheiten nach dem Merkmal x sich *nicht* unterscheiden, wird die Paarbildung jeweils in einer *Spalte* durchgeführt. Wir beginnen mit der ersten Spalte der Tabelle und 'paaren' die in einem Tabellenfeld befindlichen Untersuchungseinheiten mit den Untersuchungseinheiten, die sich in den übrigen Tabellenfeldern dieser Spalte befinden. [vgl. *Abbildung 3, (3), erstes und zweites Tabellenschema*]. In der gleichen Weise werden die zweite und dritte Spalte bearbeitet. [vgl. *Abbildung 3, (3), drittes, viertes, fünftes und sechstes Tabellenschema*]. Die Anzahl der auf x gebundenen Paare ist

$$\begin{aligned} T_x = {} & f_{11}(f_{21}+f_{31}) + f_{21}(f_{31}) + f_{12}(f_{22}+f_{32}) \\ & + f_{22}(f_{32}) + f_{13}(f_{23}+f_{33}) + f_{23}(f_{33}) \end{aligned}$$

Für unser Beispiel ergibt sich:

$$\begin{aligned} T_x = {} & 112(61+30) + 61(30) + 46(143+77) \\ & + 143(77) + 41(60+145) + 60(145) \\ = {} & 50.258 \end{aligned}$$

auf x gebundene Paare.

Für Kontingenztabellen jeder anderen Größe lautet die allgemeine Formel der Anzahl der auf x gebundenen Paare [vgl. Formel (5)]:

$$T_x = \sum_{i=1}^{k-1} \sum_{j=1}^{l} f_{ij} \cdot \sum_{s=i+1}^{k} f_{sj} \tag{5}$$

Wir können auch aus den *Spaltensummen* der Tabelle, $f_{.j}$, die Anzahl der auf x gebundenen Paare, T_x, errechnen [vgl. Formel (6)]:

$$T_x = \tfrac{1}{2}\left(\sum_{j=1}^{l} f_{.j}^2 - \sum_{i=1}^{k} \sum_{j=1}^{l} f_{ij}^2 \right) \tag{6}$$

Diese Formel ist im Falle großer Tabellen gegenüber Formel (6) vorteilhafter.

6.2 Anzahl auf y gebundener Paare (T_y)

Wir gehen aus von der Begriffsbestimmung eines auf y gebundenen Paares [vgl. *Abbildung 2, Zeile (4), Spalte (1)*]. Wiederum liegen zwei äquivalente Definitionen vor:

$$(x_j > x_{j^*}) \wedge (y_i = y_{i^*}) \text{ und } (x_j < x_{j^*}) \wedge (y_i = y_{i^*}).$$

Verwendet wird im weiteren die zuerst genannte Definition, die beim Merkmal x von der Ausprägung mit dem höherem Rang ausgeht.

In *Abbildung 3, Teil (4)* ist die Bestimmung der Anzahl der auf y gebundenen Paare dargestellt. Da in diesem Falle die Untersuchungseinheiten bezogen auf das Merkmal y sich nicht unterscheiden, wird die Paarbildung jeweils in einer *Zeile* durchgeführt. Wir beginnen mit ersten Zeile der Tabelle und 'paaren' die in einem Tabellenfeld befindlichen Untersuchungseinheiten mit den Untersuchungseinheiten, die sich in übrigen Tabellenfeldern dieser Zeile befinden. [vgl. *Abbildung 3, (4), erstes und zweites Tabellenschema*]. In der gleichen Weise werden die zweite und dritte Zeile bearbeitet. [vgl. *Abbildung 3, (4), drittes, viertes, fünftes und sechstes Tabellenschema*]. Die Anzahl der auf y gebundenen Paare ist

$$\begin{aligned} T_y = {} & f_{11}(f_{12}+f_{13}) + f_{12}(f_{13}) + f_{21}(f_{22}+f_{23}) \\ & + f_{22}(f_{23}) + f_{31}(f_{32}+f_{33}) + f_{32}(f_{33}) \end{aligned}$$

Für unser Beispiel ergibt sich:

$$\begin{aligned} T_y = {} & 112(46+41) + 46(41) + 61(143+60) \\ & + 143(60) + 30(77+145) + 77(145) \\ = {} & 50.418 \end{aligned}$$

auf y gebundene Paare.

Für Kontingenztabellen jeder anderen Größe lautet die allgemeine Formel der Anzahl der auf y gebundenen Paare [vgl. Formel (7)]:

$$T_y = \sum_{i=1}^{k} \sum_{j=1}^{l-1} f_{ij} \cdot \sum_{r=j+1}^{l} f_{ir} \tag{7}$$

Wir können auch aus den *Zeilensummen* der Tabelle, $f_{i.}$, die Anzahl der auf y gebundenen Paare, T_y, errechnen [vgl. Formel (8)]:

$$T_y = \tfrac{1}{2}\left(\sum_{i=1}^{k} f_{i.}^2 - \sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij}^2\right) \tag{8}$$

Diese Formel ist im Falle großer Tabellen gegenüber Formel (7) vorteilhafter.

6.3 Anzahl auf x und y gebundener Paare (T_{xy})

Wir gehen aus von der Begriffsbestimmung eines auf x und y gebundenen Paares [vgl. *Abbildung 2, Zeile (5), Spalte (1)*]. Hier liegt logischerweise nur eine Definition vor: $(x_j = x_{j^*}) \wedge (y_i = y_{i^*})$ Da die beiden Untersuchungseinheiten eines Paares bezogen auf beide Merkmale übereinstimmen, kann die Paarbildung nur zwischen den Untersuchungseinheiten innerhalb eines jeden einzelnen Tabellenfeldes erfolgen. Die Anzahl der auf x und y gebundene Paare beträgt z. B. für das Tabellenfeld $[x_1 y_1]$ $\tfrac{1}{2}[f_{11}(f_{11}-1)]$. Für jedes weitere der neun Tabellenfelder werden in dieser Weise Paarbildungen vorgenommen. [vgl. *Abbildung 3 (5)*] Die Anzahl der auf x und y gebundenen Paare ist

$$\begin{aligned} T_{xy} = {} & \tfrac{1}{2}\,[f_{11}(f_{11}-1) + f_{12}(f_{12}-1) + f_{13}(f_{13}-1) \\ & + f_{21}(f_{21}-1) + f_{22}(f_{22}-1) + f_{23}(f_{23}-1) \\ & + f_{31}(f_{31}-1) + f_{32}(f_{32}-1) + f_{33}(f_{33}-1)] \end{aligned}$$

Für unser Beispiel ergibt sich:

$$\begin{aligned} T_{xy} = {} & \tfrac{1}{2}[112(111)+46(45)+41(40) \\ & +61(60)+143(142)+60(59) \\ & +30(29)+77(76)+145(144)] \\ = {} & 35.625 \end{aligned}$$

auf x und y gebundene Paare.

Für Kontingenztabellen jeder anderen Größe lautet die allgemeine Formel für die Anzahl der auf x und y gebundenen Paare [vgl. Formel (9)]:

$$T_{xy} = \tfrac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij}\left(f_{ij}-1\right) \tag{9}$$

Anhand einer Proberechnung kann nach Formel (4) überprüft werden, ob die Summe der Häufigkeiten der einzelnen Paartypen, auch der Anzahl der möglichen Paare entspricht:

$$C + D + T_x + T_y + T_{xy} = \tfrac{1}{2}n\,(n-1).$$

Für unser Beispiel:

$$\begin{aligned} & 91.307 + 27.647 + 50.258 + 50.418 + 35.625 \\ & = \tfrac{1}{2}\cdot 715(714) = 255.255 \end{aligned}$$

Zur Bestimmung der Anzahl der auf x und y gebundenen Paare läßt sich Formel (9) in eine im Falle großer Tabellen vorteilhaftere Formel überführen [vgl. Formel (10)]:

$$T_{xy} = \tfrac{1}{2}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij}^2 - N\right) \tag{10}$$

6.4 Berücksichtigung der gebundenen Paare

Die Anzahl der gebundenen Paare ist zum Zwecke der Normierung der Differenz der Anzahl der konkordanten und diskordanten Paare, $C-D$, zu berücksichtigen. Dadurch wird erreicht, daß sich ein angemessener Assoziationskoeffizient ergibt, der bei einem perfekten gegenläufigen Zusammenhang die Untergrenze -1,0 und bei einem perfekten gleichsinnigen Zusammenhang die Obergrenze +1,0 besitzt.

Greifen wir zunächst den Kendallschen Assoziationskoeffizienten τ_a wieder auf. Bei diesem wird die Differenz $C-D$ bezogen auf die Anzahl der möglichen Paare [vgl. Formel (1)]. Die Anzahl der möglichen Paare gibt dabei das Ausmaß der *gesamten* Variation der Untersuchungseinheiten an - und zwar sowohl über die Ausprägungen des Merkmals x als auch über diejenigen des Merkmals y. Im Falle echter Rangreihen ist die Anzahl der möglichen Paare gleich der Summe der konkordanten und der diskordanten Paare:

$$\tfrac{1}{2}n(n-1) = C + D\,.$$

Eine Variation der Untersuchungseinheiten besteht insofern nur nach in Gleichsinnigkeit und Gegenläufigkeit beider Merkmale und auch $C + D$ gibt in diesem Falle das *gesamte* Ausmaß der Variation an.

Dies ist allerdings nicht so, wenn verschiedene Untersuchungseinheiten ein und denselben Rang besitzen. Dann liegen neben konkordanten und diskordanten Paaren auch gebundene Paare vor. Somit variieren die Untersuchungseinheiten nicht allein nach Gleichsinnigkeit und Gegenläufigkeit *beider* Merkmale, sondern *zusätzlich* auch noch jeweils nach *einzelnen* Merkmalen. Das Ausmaß, indem eine derartige zusätzliche Variation nach *einem* der beiden Merkmale besteht, ergibt sich aus der Anzahl der nach dem *anderen* Merkmal gebundenen Paare. Die Anzahl von gebundenen Paaren zeigt so für jedes einzelne Merkmal die ihm entsprechende zusätzliche Variation an. Wenn nun für ein einzelnes Merkmal das Ausmaß der *gesamten* Variation der Untersuchungseinheiten über die Ausprägungen dieses Merkmals bestimmt wird, muß zur Summe der konkordanten und der diskordanten Paare, $C+D$, noch die Anzahl der gebundenen Paare hinzutreten, die die zusätzliche Variation nach diesem Merkmal anzeigen.

- Bestimmen wir zuerst das Ausmaß der gesamten Variation der Untersuchungseinheiten über die Ausprägungen des Merkmals x: Hier tritt zur Summe der konkordanten und der diskordanten Paare, $C+D$, noch die Anzahl der gebundenen Paare hinzu, die die zusätzliche Variation nach dem Merkmal x anzeigen. Dies ist die Anzahl der nach dem Merkmal y gebundenen Paare T_y.

  $C+D+T_y$ entspricht so dem Ausmaß der *gesamten* Variation der Untersuchungseinheiten über die Ausprägungen des *Merkmals x*.

- Weiter ist das Ausmaß der gesamten Variation der Untersuchungseinheiten über die Ausprägungen des Merkmals y zu ermitteln: In diesem Falle tritt zur Summe der konkordanten und der diskordanten Paare, $C+D$, noch die Anzahl der gebundenen Paare hinzu, die die zusätzliche Variation nach dem Merkmal y anzeigen. Dies ist die Anzahl der nach dem Merkmal x gebundenen Paare T_x.

  $C+D+T_x$ entspricht so dem Ausmaß der *gesamten* Variation der Untersuchungseinheiten über die Ausprägungen des *Merkmals y*.

Wir definieren nun den Assoziationskoeffizienten τ_b analog dem Koeffizienten τ_a:
Wie der Koeffizient τ_a [vgl. Formel (1)] enthält der Koeffizient τ_b im *Zähler* die Differenz $C-D$, in der das in den beiden Rangreihen vorhandene Ausmaß der Gleichsinnigkeit dem in beiden Rangreihen vorhandene Ausmaß der Gegenläufigkeit gegenübergestellt wird.
Im *Nenner* der Formel hingegen steht das geometrische Mittel der Variation der Untersuchungseinheiten über die Ausprägungen des Merkmals x, $C+D+T_y$, und derjenigen über die Ausprägungen des Merkmals y, $C+D+T_x$, [vgl. Formel (11)].

$$\tau_b = \frac{C-D}{\sqrt{(C+D+T_y)(C+D+T_x)}} \tag{11}$$

Der Koeffizient τ_b hat die gewünschte Ober- und Untergrenze:

$$-1{,}0 \le \tau_b \le +1{,}0 .$$

Anzumerken ist, daß die Anzahl der auf x und y gebundenen Paare, T_{xy}, nicht berücksichtigt wird. Dies ist insoweit logisch, als ein vorhandener Zusammenhang zwischen Merkmalen etwas über das Verhältnis zwischen Zeilen und/oder Spalten einer Tabelle aussagt und nichts zu tun hat mit Gegebenheiten innerhalb einzelner Tabellenfelder.

Für unser Beispiel kann nun die Stärke des Zusammenhangs zwischen den beiden Merkmalen anhand des abgeleiteten Koeffizienten τ_b bestimmt werden. Vermutet wird, daß die ‚Zufriedenheit mit der Wohnung' (Merkmal x) und die ‚subjektive Einschätzung der Wohnungsgröße' (Merkmal y) sich gegenseitig bedingen. Eingesetzt in die Formel [vgl. Formel (11)] ergibt sich für unser Beispiel:

$$\tau_b = \frac{91.307-27.647}{\sqrt{(91.307+27.647+50.258)(91.307+27.647+50.418)}} = 0{,}376.$$

Es besteht somit ein, der obigen Vermutung entsprechender Zusammenhang im Grade von 0,376.

Berechneten wir nun statt des angemessenen Koeffizienten τ_b den Koeffizienten τ_a [vgl. Formel (1)], so ergibt sich ein geringerer Zusammenhang:

$$\tau_a = \frac{91.307-27.647}{\frac{1}{2}\cdot 715(715-1)} = 0{,}249$$

In dem gegenüber dem Wert von $\tau_b = 0{,}376$ geringeren Wert von $\tau_a = 0{,}249$ kommt die Unterschätzung des Zusammenhangs durch die Nichtberücksichtigung der gebundenen Paare zum Ausdruck.

7. Abschließende Bemerkungen

Zum Abschluß soll noch auf die Äquivalenz der logischen Struktur des Koeffizienten τ_b und des Produkt-Moment-Korrelationskoeffizienten r_{xy} hingewiesen werden. Die Korrelation der Merkmale x und y wird beim Koeffizienten r_{xy} definiert als das Verhältnis der Kovariation der Merkmale x und y zum geometrischen Mittel der Variation des Merkmals x und der des Merkmals y:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Zwischen den Koeffizienten r_{xy} und τ_b bestehen bezüglich dieser Sachverhalte folgende Äquivalenzen:

- Kovariation der Merkmale x und y:
  $$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \Leftrightarrow C - D$$
- Variation des Merkmals x:
  $$\sum_{i=1}^{n}(x_i - \bar{x})^2 \Leftrightarrow C + D + T_y$$
- Variation des Merkmals y:
  $$\sum_{i=1}^{n}(y_i - \bar{y})^2 \Leftrightarrow C + D + T_x$$

Hieraus ergibt sich hinsichtlich der Korrelation der Merkmale x und y eine Äquivalenz beider Koeffizienten.

In entsprechender Weise lassen sich analoge ordinale Koeffizienten auch zu den Regressionskoeffizienten, b_{yx} für die Regression von y auf x bzw. b_{xy} für die Regression von x auf y, ableiten. Ein Regressionskoeffizient ist allgemein definiert als das Verhältnis der Kovariation der Merkmale x und y zur Variation des Regressors, d. h. dem Merkmal, das als unabhängig gesetzt wird. Die entsprechenden äquivalenten Koeffizienten sind somit [vgl. Formel (12)]:

$$d_{yx} = \frac{C - D}{C + D + T_y} \text{ bzw. } d_{xy} = \frac{C - D}{C + D + T_x} \qquad (12)$$

Es handelt sich bei diesen Koeffizienten um die sogenannten Somers'schen Koeffizienten, die in der empirischen Sozialforschung inzwischen weit verbreitet sind.

1) Zu ordinalen Assoziationskoeffizienten – wie die Kurzbezeichnung für derartige Koeffizienten lautet – siehe die klassischen Aufsätze von Leo A. Goodman und William H. Kruskal, die im *Journal of the American Statistical Association* seit den fünfziger Jahren erschienen sind und später auch in einem Sammelband veröffentlicht wurden (Goodman; Kruskal 1979)

2) Vgl. auch Schütze (1993), Formel (3), Seite 18, in der die Anzahl der konkordanten Paare mit P und die Anzahl der diskordanten Paare mit Q bezeichnet wird.

3) Eine entsprechende Spezialisierung des Kendallschen τ-Koeffizienten, τ_c, werden wir hier nicht behandeln, da sie im vorliegenden Zusammenhang irrelevant ist.

Literatur

- Goodman, Leo A.; Kruskal, William H. (1979): Measures of Association for Cross Classifications. New York - Heidelberg - Berlin: Springer
- Kendall, Maurice (1975): Rank correlation methods. Fourth Edition. London - High Wycombe: Griffin
- Schütze, Claudia (1993): Kendall's τ - Ein alternativer Korrelationskoeffizent. In: Stochastik in der Schule 13 (H. 3) S. 10-20
- Somers, Robert H. (1962): A New Asymmetric Measure of Association for Ordinal Variables. In: American Sociological Review 27 (H. 6) S. 799-811

Adresse des Autors

Prof. Dr. rer. pol. Heinz Renn
Universität Hamburg
Institut für Soziologie
Allende-Platz 1

D-20146 Hamburg