Ökologischer Fehlschluss

Inwieweit können wir von statistischen Beziehungen zwischen **Aggregat**merkmalen auf die Beziehungen zwischen den korrespondierenden **Individual**merkmalen schließen?

Kann z.B. aus einer Korrelation, die **„im Aggregat“** zwischen zwei Merkmalen besteht, auf die Korrelation zwischen den korrespondierenden **Individual**merkmalen geschlossen werden?

W.S.Robinson (1950) klassisches Beispiel

„foreign-born“ $\xrightarrow{?}$ „literate in English“

Wenn es sich um verschiedene Gruppen handelt,
die sich z.B. in großen Städten konzentrieren,
dann befinden sich in Aggregateinheiten
mit vielen „foreign-born“ Bewohnern auch viele Bewohner,
die des Lesens und Schreibens mächtig sind

	„foreign-born“ residents	„literate in English“
Agg.Einheit 1	n_{11}	n_{12}
Agg.Einheit 2	n_{21}	n_{22}
Agg.Einheit 3	n_{31}	n_{32}
.	.	.
.	.	.

Analyseebenen

Aggregatebene(n)	**Kollektive** Eigenschaften	Aggregatdaten-analyse
	[Inklusion]	
Individualebene	**Individuelle** Eigenschaften	Individualdaten-analyse

Hypothesen

- Wenn/dann und je/desto Hypothesen
- deterministisch vs. probabilistisch
- Informationsgehalt einer Hypothese
- Im Allgemeinen: Erwartung, wie Untersuchungseinheiten in einem analytischen Raum von Variablen verteilt sind
- Über Drittvariablen bestimmbar: Allgemeinheits- und Komplexitätsgrad einer Hypothese
- Falsifizierbarkeit und Spezifizität einer Hypothese

Wenn – dann – Hypothesen

dann: \ wenn:	A	~A
B	K (I)	K (II)
~B	F (III)	K (IV)

dann: \ wenn:	A	~A
B	K (I)	F (II)
~B	F (III)	K (IV)

K = Konfirmatoren

F = Falsifikatoren

(1) Wenn – dann – Hypothese

(deterministische **Implikation**)

(2) Wenn – und – nur – wenn – dann – Hypothese

(deterministische **Äquivalenz**)

A = hinreichende Bedingung

A = hinreichende und notwendige Bedingung

Quelle: A. Diekmann (1995) Empirische Sozialforschung. Reinbek bei Hamburg: Rowohlt, S. 109

Deterministische und probabilistische Hypothesen

„wenn A, dann B":

kann in die Erwartung transformiert werden:

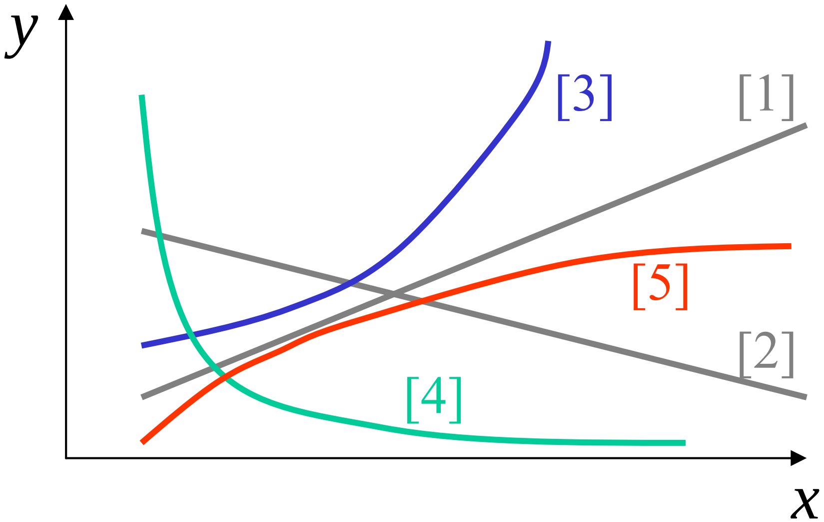
Prob (B) bei Auftreten von A
ist größer als
Prob (B) bei Auftreten von ~A

$$P(B \mid A) > P(B \mid \sim A)$$

Je – desto – Hypothesen

Monoton steigende bzw. monoton fallende Zusammenhänge

Ggf. Spezifikation des Zusammenhangs als spezielle mathematische Funktion möglich (Beispiele):



[1] linear steigend

[2] linear fallend

[3] exponentiell steigend

[4] exponentiell fallend

[5] logarithmisch

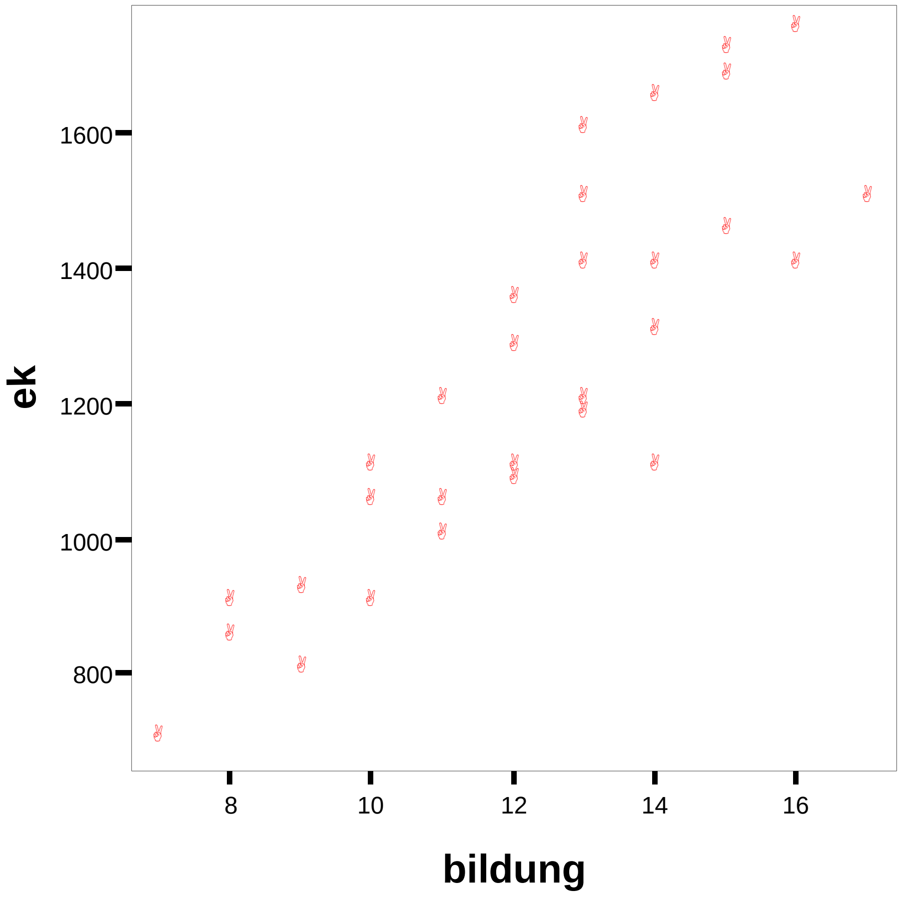
Nach Abb. *IV.3* (S. 115) in Diekmann, Andreas (1995) Empirische Sozialforschung. Reinbek bei Hamburg: Rowohlt

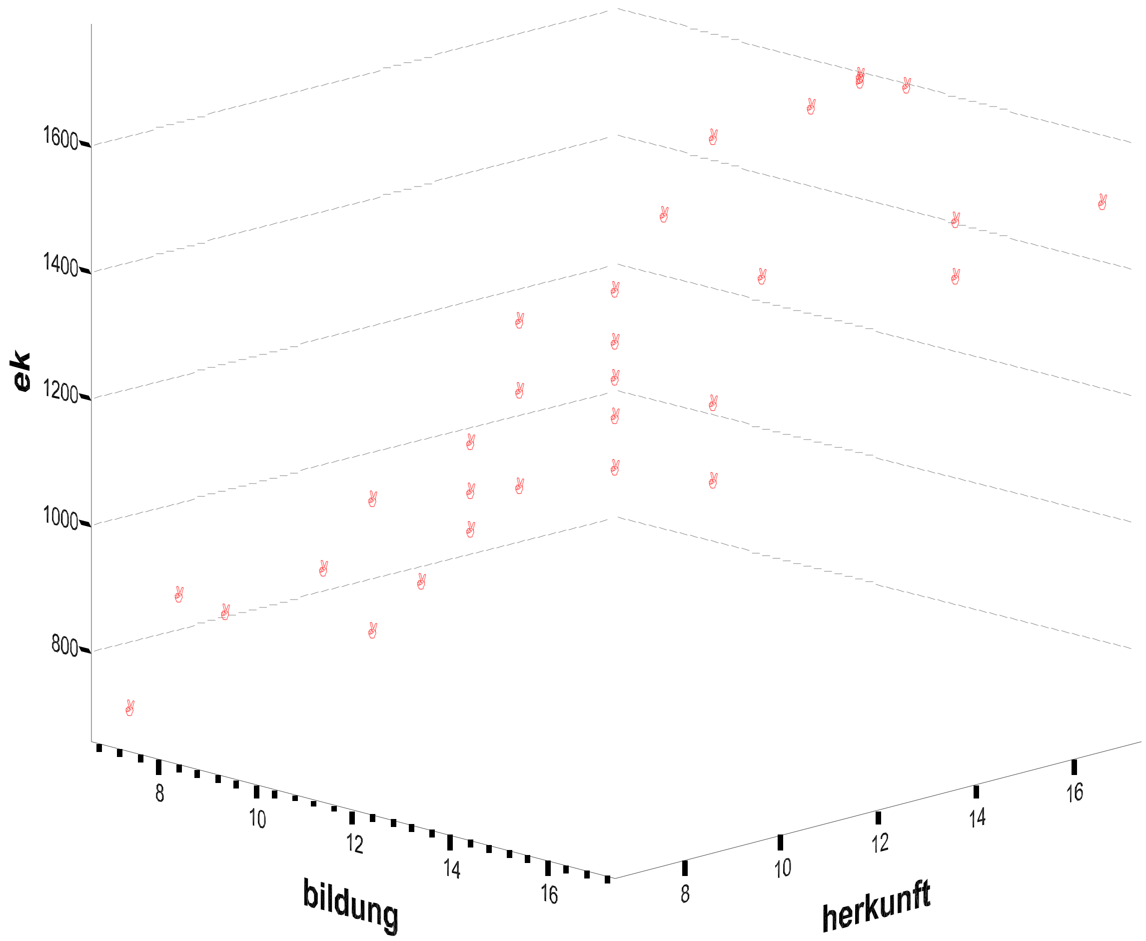
» **Informationsgehalt** (empirischer Gehalt) eines Satzes ist die Menge der von diesem Satz ausgeschlossenen Sätze.

» Der Gehalt einer Wenn-dann-Hypothese wächst **gleich**sinnig mit dem Gehalt der Dann-Komponente und **gegen**sinnig zum Gehalt der Wenn-Komponente

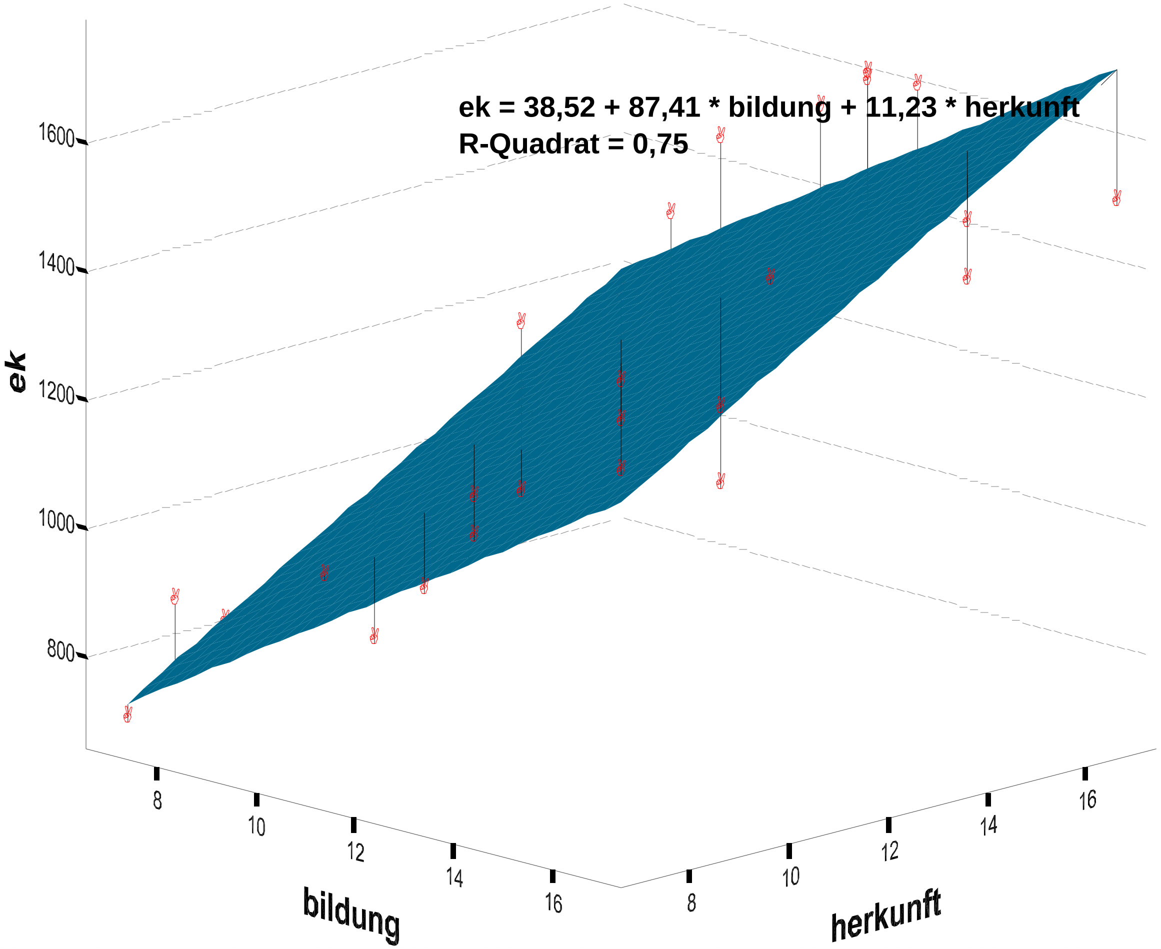
Nach Diekmann (1995) Empirische Sozialforschung. Reinbek bei Hamburg: Rowohlt, S. 132f.

Eine **Hypothese** ist eine Behauptung darüber, wie ein Set **S** von Untersuchungs-einheiten in einem Raum von Variablen $X_1, X_2, \ldots, X_n$ verteilt ist.




1600
1400
1200
1000
800
ek
8
10
12
14
16
bildung
8
10
12
14
16
herkunft

Lineare Regression



Allgemeinheitsgrad …

.. ist bestimmt durch das Set von Bedingungen **B**, unter denen die Hypothese haltbar ist

„[intended] field of tenability“

Forschungsleitende **Frage**

„Gegeben die Hypothese H, zeige mir das Set von Bedingungen **B**, unter denen sie haltbar ist“

Bedingungen B = Ausprägungen von Drittvariablen

Nach: Galtung, Johan (1969) Theory and Methods of Social Research. Oslo: Universitetsforlaget, pp. 309 -339

Komplexitätsgrad …

.. über die **Anzahl** der Variablen, die den untersuchtenVariablenraum aufspannen

H (X,Y) getestet für T=$\mathbf{t_1}$

? ***H***(X,Y) für T=$\mathbf{t_2}$

Mögliche **Ergebnisse**:

H hängt von T ab (gilt z.B. nur für T=$\mathbf{t_1}$) und ist entsprechend zu **spezifizieren**

H hängt **nicht** von T ab und ist entsprechend auf T=$\mathbf{t_2}$ **generalisiert**

Falsifizierbarkeit …

$$F = \frac{n_F}{E}$$

Anteil falsifizierender Ergebnisse an allen möglichen Ergebnissen
- bei a priori Annahme, dass alle E Verteilungen gleich wahrscheinlich sind

Spezifizität

Anzahl möglicher Unterscheidungen, die in den empirischen Ergebnissen getroffen werden können

Nach: Galtung, Johan (1969) Theory and Methods of Social Research. Oslo: Universitetsforlaget, p. 188

Im Fall probabilistischer Hypothesen:

Anzahl der Partitionen von N in R Teile:

$$E = \binom{N+R-1}{R-1}$$

Anzahl der möglichen Verteilungen eines Samples **S** mit **N** Fällen über die **R** [= $r_1 \times r_2 \times \ldots \times r_k$ -] Ausprägungs-kombinationen der *k* Variablen, die den analytischen Raum (Ereignisraum) aufspannen

Nach: Galtung, Johan (1969) Theory and Methods of Social Research. Oslo: Universitetsforlaget, p. 188

Beispiel

Mögliche Verteilungen oder Partitionen von 3 Fällen in 2 Teile (**R** = 2; **N** = 3):

3 + 0 = 3
2 + 1 = 3
1 + 2 = 3
0 + 3 = 3

$$E = \binom{3+2-1}{2-1} = \binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4 \times 3 \times 2 \times 1}{1 \times 3 \times 2 \times 1} = \frac{24}{6} = 4$$

$$E = \binom{N+R-1}{R-1}$$

Größe des Ereignisraumes, **R**

Stichproben-Umfang, **N**

	1	2	3	4	5	…10
0	1	1	1	1	1	.
1	1	2	3	4	5	.
2	1	3	6	10	15	.
3	1	4	10	20	35	.
4	1	5	15	35	70	.
5	1	6	21	56	126	.
.						.
.						.
10	.	.	.	.	.	92.378

1. Vereinfachung von **E** durch ..

.. **deterministische** Repräsentation probabilistischer Hypothesen über eine **Korrelations**sprache

2. Abbildung des „Ereignissets“ auf ein Set mit den Ausprägungen

wahr	– unentschieden –	falsch
Korrespondenz	– unentschieden –	Nicht Korresp.
bestätigt	– unentschieden –	entkräftet

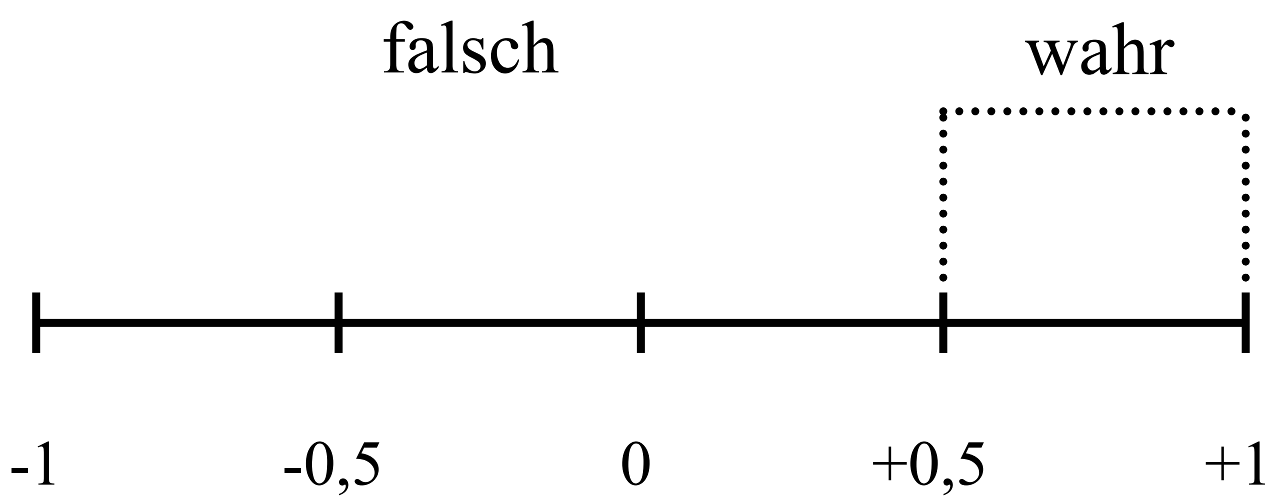
Nach: Galtung, Johan (1969) Theory and Methods of Social Research. Oslo: Universitetsforlaget, pp. 309 - 315

3. Vergleich der **erwarteten** Korrelationen mit den **beobachteten** Korrelationen

$C\ (X,Y)$	$C^{+}\ (X,Y)$
	$C^{-}\ (X,Y)$
	$C^{0}\ (X,Y)$
	$C^{\geq 0,5}(X,Y)$

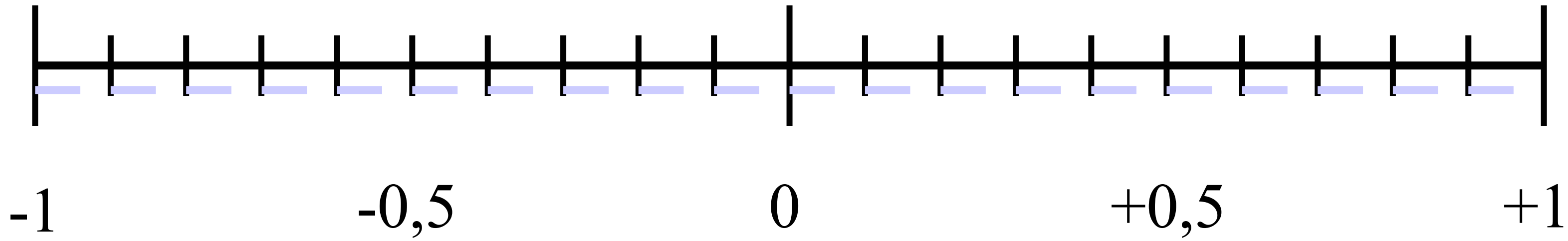
Nach: Galtung, Johan (1969) Theory and Methods of Social Research. Oslo: Universitetsforlaget, pp. 309 - 315

Beispiel:
Hypothese, derzufolge r ≥ 0,5 erwartet wird



Nach: Galtung, Johan (1969) Theory and Methods of Social Research.
Oslo: Universitetsforlaget, p. 314

Korrelation



$$E=20$$

$\boxed{-1{,}0 \le r < -0{,}9}$ $\boxed{-0{,}9 \le r < -0{,}8}$ $\boxed{-0{,}8 \le r < -0{,}7}$ $[\ \dots$

$\dots\ ]$ $\boxed{0{,}7 \le r < 0{,}8}$ $\boxed{0{,}8 \le r < 0{,}9}$ $\boxed{0{,}9 \le r < 1{,}0}$

Falsifizierbarkeit …

$$F = \frac{n_F}{E}$$

Anteil falsifizierender Ergebnisse an allen möglichen Ergebnissen

- bei a priori Annahme, dass alle E Verteilungen gleich wahrscheinlich sind

Empirische **Haltbarkeit**

$$DC' = \frac{DC^* - (1-F)}{1-(1-F)} = 1 - \frac{1-DC^*}{F}$$

DC* „naives“ Maß zur Bestimmung des Bestätigungsgrades, da es nicht die a priori Wahrscheinlichkeit (1 - F) berücksichtigt, die Hypothese zu bestätigen

DC‘ Überschuss von DC* über die *a priori* Bestätigungs-wahrscheinlichkeit (1 - F), relativ zum maximalen Überschuss, der erreicht wird, wenn DC*=1

Je höher F, desto höher ist DC‘ für ein gegebenes DC*

DC* - Naiver Bestätigungsgrad

- Anteil korrekt vorhergesagter Ereignisse

  z.B.:

  % der Fälle, die eine Hyp. bestätigen univariat oder über ein bivariates Korrelations-/Assoziationsmaß, als Grad der Annäherung an eine deterministische Beziehung

- Relativer Anstieg der Vorhersagegenauigkeit

$$DC' = \frac{DC^* - (1-F)}{1-(1-F))} = 1 - \frac{1-DC^*}{F}$$

$$\boxed{r \geq 0{,}1} \qquad E = 20 \qquad F = \frac{11}{20} = 0{,}55$$

Hypothese

$$DC' = 1 - \frac{1-0{,}6}{0{,}55} = 0{,}27$$

$$\boxed{r \geq 0{,}5} \qquad E = 20 \qquad F = \frac{15}{20} = 0{,}75$$

$$DC' = 1 - \frac{1-0{,}6}{0{,}75} = 0{,}47$$

Empirischer Befund: jeweils r = 0,6 (Beispiel)

Systematische Replikation

Veränderungen im DC‘ Grad

1. Bestätigungsgrad für jede Replikation
2. Replikationsdistanz
3. Anzahl der Replikationen

$$DC = \frac{DC' + \sum_{1}^{p} d_i \cdot dc_i}{1 + \sum_{1}^{p} d_i}$$

d_i Replikationsdistanz
(z.B.: $1 - r_i$)

p Anzahl Replikationen
(i= i-te Replikation)

DC „degree of confirmation“

dc_i „degree of confirmation“
der i-ten Replikation

Nach: Galtung, Johan (1969) Theory and Methods of Social Research.
Oslo: Universitetsforlaget, p. 443