

DER KAPPA-KOEFFIZIENT:
DISKUSSION EINES MISSVERSTÄNDNISSES
UND EIN
MODIFIZIERUNGSVORSCHLAG

Marcus Kutschmann
Institut für Rehabilitationsforschung
Norderney

Gerd Rippin
Omnicare Clinical Research GmbH
Köln

Vortragsinhalt:

- *Was wird mit Kappa berechnet?*
- *Wie wird Kappa berechnet?*
- *Welche Probleme gibt es bei der Anwendung?*
- *Warum werden diese Probleme als solche empfunden? („Missverständnis“)*
- *Wie lassen sich diese Probleme lösen? („Modifizierungsvorschlag“)*
- *Welche Vorteile hat der Modifizierungsvorschlag“?*
- *Welche Schlussfolgerungen ergeben sich?*

Der Kappa-Koeffizient von Cohen (1960):

„It is now fairly well disseminated as one of the standard summary statistics used in the medical literature.“ (Guggenmoos-Holzmann, 1993)

Berechnung der zufallskorrigierten Übereinstimmung zweier Gutachter, die eine bestimmte Anzahl von Objekten/Subjekten mit Hilfe eines vorgegebenen Kategoriensystems klassifizieren.

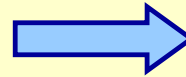
Beispiel zur Berechnung von Kappa:

Zuordnung von **neun Patienten** zu einer der **drei Kategorien** *Schizophren (S)*, *Neurotisch (N)* und *Hirngeschädigt (H)* durch **zwei Psychiater**:

	<i>Patienten</i>								
	1	2	3	4	5	6	7	8	9
<i>Rater A</i>	<i>H</i>	<i>S</i>	<i>N</i>	<i>N</i>	<i>S</i>	<i>H</i>	<i>H</i>	<i>N</i>	<i>H</i>
<i>Rater B</i>	<i>S</i>	<i>S</i>	<i>N</i>	<i>H</i>	<i>S</i>	<i>N</i>	<i>H</i>	<i>N</i>	<i>S</i>

Beispiel (Fortsetzung):

	<i>S</i>	<i>N</i>	<i>H</i>	
<i>S</i>	p_{11}	p_{12}	p_{13}	$p_{1.}$
<i>N</i>	p_{21}	p_{22}	p_{23}	$p_{2.}$
<i>H</i>	p_{31}	p_{32}	p_{33}	$p_{3.}$
	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{..}$



	<i>S</i>	<i>N</i>	<i>H</i>	
<i>S</i>	$\frac{2}{9}$	0	0	$\frac{2}{9}$
<i>N</i>	0	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{3}{9}$
<i>H</i>	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{4}{9}$
	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	1

$$\left. \begin{aligned}
 B &= p_{11} + p_{22} + p_{33} = 0,56 \\
 E &= p_{1.}p_{.1} + p_{2.}p_{.2} + p_{3.}p_{.3} = 0,31
 \end{aligned} \right\} \Rightarrow \kappa = \frac{B - E}{1 - E} = 0,36$$

„Chance-corrected measures of agreement are prone to exhibit paradoxical and counter-intuitive results when used as measures of reliability.“
(Guggenmoos-Holzmann, 1993)

Gleiches B, unterschiedliches κ :

$$\begin{array}{ccc|c} \frac{3}{9} & 0 & 0 & \frac{3}{9} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{6}{9} & 0 & \frac{6}{9} \\ \hline \frac{3}{9} & \frac{6}{9} & 0 & 1 \end{array}$$

$$\kappa = 0,25$$

$$\begin{array}{ccc|c} \frac{3}{9} & \frac{3}{9} & 0 & \frac{6}{9} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{3}{9} & 0 & \frac{3}{9} \\ \hline \frac{3}{9} & \frac{6}{9} & 0 & 1 \end{array}$$

$$\kappa = 0,14$$

$$\begin{array}{ccc|c} \frac{3}{9} & 0 & 0 & \frac{3}{9} \\ 0 & 0 & \frac{4}{9} & \frac{4}{9} \\ 0 & \frac{2}{9} & 0 & \frac{2}{9} \\ \hline \frac{3}{9} & \frac{2}{9} & \frac{4}{9} & 1 \end{array}$$

$$\kappa = 0,04$$

$$\begin{array}{ccc|c} \frac{3}{9} & 0 & \frac{6}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{3}{9} & 0 & \frac{6}{9} & 1 \end{array}$$

$$\kappa = 0,00$$

$$\begin{array}{ccc|c} \frac{3}{9} & 0 & \frac{5}{9} & \frac{8}{9} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{9} & 0 & \frac{1}{9} \\ \hline \frac{3}{9} & \frac{1}{9} & \frac{5}{9} & 1 \end{array}$$

$$\kappa = -0,04$$

$$\begin{array}{ccc|c} \frac{3}{9} & 0 & \frac{5}{9} & \frac{8}{9} \\ 0 & 0 & 0 & 0 \\ \frac{1}{9} & 0 & 0 & \frac{1}{9} \\ \hline \frac{4}{9} & 0 & \frac{5}{9} & 1 \end{array}$$

$$\kappa = -0,23$$

$$\begin{array}{ccc|c} \frac{3}{9} & \frac{2}{9} & \frac{1}{9} & \frac{6}{9} \\ \frac{2}{9} & 0 & 0 & \frac{2}{9} \\ \frac{1}{9} & 0 & 0 & \frac{1}{9} \\ \hline \frac{6}{9} & \frac{2}{9} & \frac{1}{9} & 1 \end{array}$$

$$\kappa = -0,35$$

$$\begin{array}{ccc|c} \frac{3}{9} & 0 & \frac{3}{9} & \frac{6}{9} \\ 0 & 0 & 0 & 0 \\ \frac{3}{9} & 0 & 0 & \frac{3}{9} \\ \hline \frac{6}{9} & 0 & \frac{3}{9} & 1 \end{array}$$

$$\kappa = -0,50$$

Unterschiedliches B , gleiches κ :

$$\begin{array}{ccc|c} \frac{8}{9} & 0 & \frac{1}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{8}{9} & 0 & \frac{1}{9} & 1 \end{array}$$

$\kappa = 0$

$$\begin{array}{ccc|c} \frac{7}{9} & 0 & \frac{2}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{7}{9} & 0 & \frac{2}{9} & 1 \end{array}$$

$\kappa = 0$

$$\begin{array}{ccc|c} \frac{6}{9} & 0 & \frac{3}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{6}{9} & 0 & \frac{3}{9} & 1 \end{array}$$

$\kappa = 0$

$$\begin{array}{ccc|c} \frac{5}{9} & 0 & \frac{4}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{5}{9} & 0 & \frac{4}{9} & 1 \end{array}$$

$\kappa = 0$

$$\begin{array}{ccc|c} \frac{4}{9} & 0 & \frac{5}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{4}{9} & 0 & \frac{5}{9} & 1 \end{array}$$

$\kappa = 0$

$$\begin{array}{ccc|c} \frac{3}{9} & 0 & \frac{6}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{3}{9} & 0 & \frac{6}{9} & 1 \end{array}$$

$\kappa = 0$

$$\begin{array}{ccc|c} \frac{2}{9} & 0 & \frac{7}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{2}{9} & 0 & \frac{7}{9} & 1 \end{array}$$

$\kappa = 0$

$$\begin{array}{ccc|c} \frac{1}{9} & 0 & \frac{8}{9} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline \frac{1}{9} & 0 & \frac{8}{9} & 1 \end{array}$$

$\kappa = 0$

1. Kappa-Paradoxon: grosses B, kleines Kappa

„If E is large, the correction process can convert a relatively high value of K[appa] ... Thus, with different values of E the K[appa] for identical values of B can be more than twofold higher in one instance than the other.“ (Feinstein & Cicchetti, 1990)

kleines E:

$\frac{2}{9}$	0	0	$\frac{2}{9}$	$B = 0,67$ $E = 0,33$
$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{4}{9}$	
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	
$\frac{3}{9}$	$\frac{3}{9}$	$\frac{3}{9}$	1	

$\Rightarrow \kappa = 0,51$

grosses E:

$\frac{6}{9}$	0	0	$\frac{6}{9}$	$B = 0,67$ $E = 0,55$
$\frac{1}{9}$	0	$\frac{1}{9}$	$\frac{2}{9}$	
0	$\frac{1}{9}$	0	$\frac{1}{9}$	
$\frac{7}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	1	

$\Rightarrow \kappa = 0,27$

2. Kappa-Paradoxon: Randverteilungen

„Unbalanced marginal totals produce higher values of Kappa than more balanced totals.“ (Feinstein & Cicchetti, 1990)

unbalancierte Rangverteilungen:

$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{5}{9}$	↑	$B = 0,33$	
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$			$E = 0,23$
0	0	$\frac{1}{9}$	$\frac{1}{9}$			
$\frac{1}{9}$	$\frac{3}{9}$	$\frac{5}{9}$	1			$\Rightarrow \kappa = 0,13$

→

balancierte Rangverteilungen:

$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{5}{9}$	↑	$B = 0,33$	
$\frac{2}{9}$	$\frac{1}{9}$	0	$\frac{3}{9}$			$E = 0,43$
$\frac{1}{9}$	0	0	$\frac{1}{9}$			
$\frac{5}{9}$	$\frac{3}{9}$	$\frac{1}{9}$	1			$\Rightarrow \kappa = -0,18$

←

3. Kappa-Paradoxon: $B=0$, trotzdem Zufallskorrektur

0	$\frac{8}{9}$	0	$\frac{8}{9}$
0	0	0	0
0	$\frac{1}{9}$	0	$\frac{1}{9}$
0	1	0	1

$$\kappa = 0$$

0	0	0	0
$\frac{2}{9}$	0	0	$\frac{2}{9}$
$\frac{5}{9}$	$\frac{2}{9}$	0	$\frac{7}{9}$
$\frac{7}{9}$	$\frac{2}{9}$	0	1

$$\kappa = -0,04$$

0	0	0	0
$\frac{2}{9}$	0	0	$\frac{2}{9}$
$\frac{2}{9}$	$\frac{5}{9}$	0	$\frac{7}{9}$
$\frac{4}{9}$	$\frac{5}{9}$	0	1

$$\kappa = -0,14$$

0	0	0	0
$\frac{2}{9}$	0	0	$\frac{2}{9}$
$\frac{1}{9}$	$\frac{6}{9}$	0	$\frac{7}{9}$
$\frac{3}{9}$	$\frac{6}{9}$	0	1

$$\kappa = -0,17$$

0	0	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{2}{9}$	0	0	$\frac{2}{9}$
$\frac{4}{9}$	$\frac{2}{9}$	0	$\frac{6}{9}$
$\frac{6}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	1

$$\kappa = -0,25$$

0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$
$\frac{2}{9}$	0	$\frac{1}{9}$	$\frac{3}{9}$
$\frac{1}{9}$	$\frac{2}{9}$	0	$\frac{3}{9}$
$\frac{3}{9}$	$\frac{3}{9}$	$\frac{3}{9}$	1

$$\kappa = -0,50$$

0	0	$\frac{4}{9}$	$\frac{4}{9}$
$\frac{2}{9}$	0	0	$\frac{2}{9}$
$\frac{1}{9}$	$\frac{2}{9}$	0	$\frac{3}{9}$
$\frac{3}{9}$	$\frac{2}{9}$	$\frac{4}{9}$	1

$$\kappa = -0,53$$

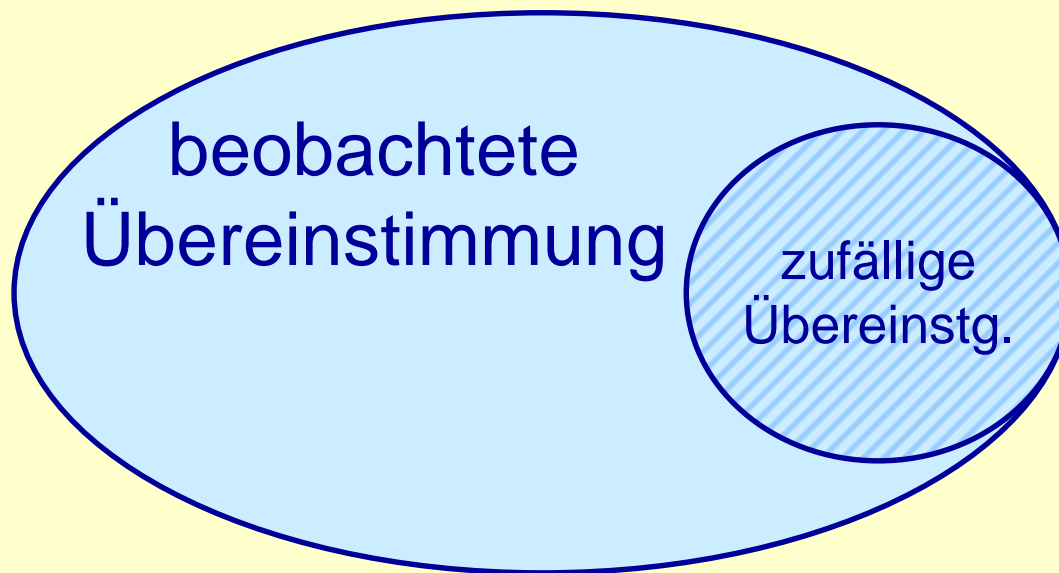
0	0	$\frac{4}{9}$	$\frac{4}{9}$
0	0	0	0
$\frac{5}{9}$	0	0	$\frac{5}{9}$
$\frac{5}{9}$	0	$\frac{4}{9}$	1

$$\kappa = -0,98$$

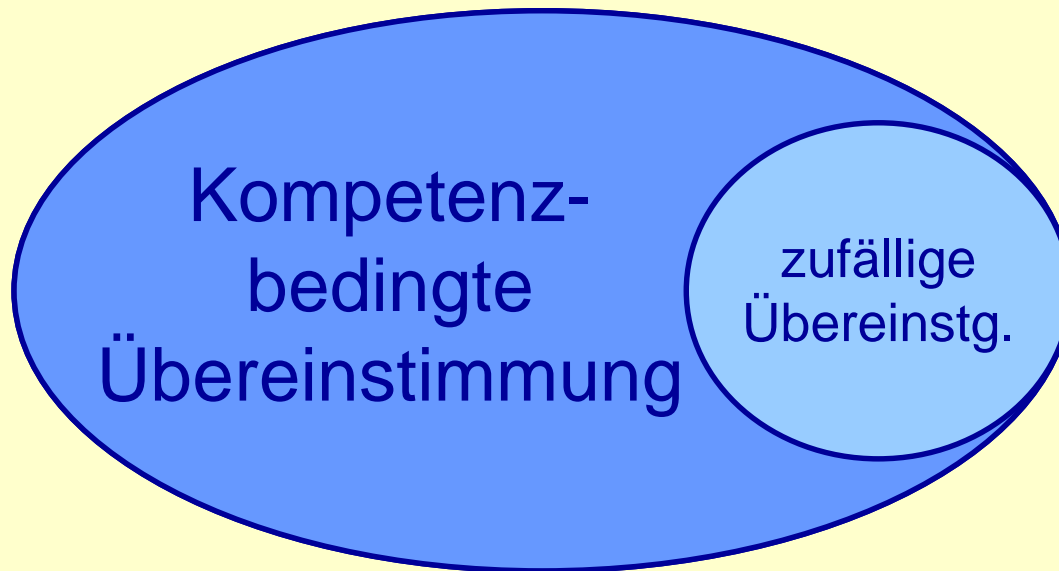
Warum bezeichnet man diese Ergebnisse als paradox bzw. warum erscheinen sie unplausibel?

„The coefficient kappa ... is the proportion of agreement *after* chance agreement is removed from consideration.“ (Cohen, 1960)

Suggestierter Zusammenhang zwischen beobachteter und zufälliger Übereinstimmung:

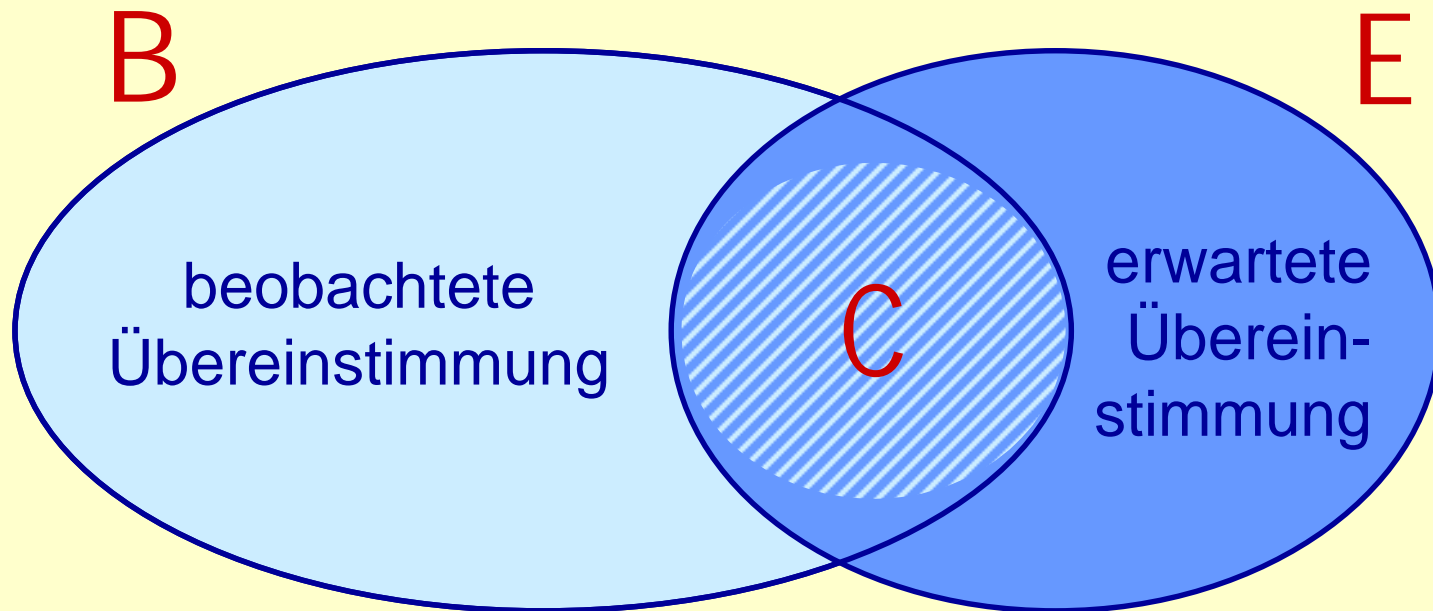


Suggestierter Zusammenhang zwischen beobachteter und zufälliger Übereinstimmung (Fortsetzung):



„Zwei Befunder werden bei einem gewissen Anteil der Individuen bereits rein zufällig den gleichen Befund erstellen, auch wenn gar kein Bezug des Befundes zum Individuum besteht (mithin zufällige Übereinstimmungen auftreten). Nun interessiert in erster Linie, **wie weit die durch echte Kompetenz bedingte Übereinstimmung den Anteil dieser 'Zufalls-Treffer' überschreitet**. Gemessen wird also das **Kompetenz-bedingte Agreement** zweier oder mehrerer Befunde(r) ...“ (Krummenauer, 1999)

Tatsächlicher Zusammenhang zwischen beobachteter und erwarteter Übereinstimmung bei Cohen:



„Kappa is not really a chance-corrected measure of agreement.“ (Übersax, 2002)

Ursache der Paradoxa:

In Cohens Korrekturterm fließen Matrixelemente ein, die nichts zur beobachteten Übereinstimmung beitragen!

	<i>S</i>	<i>N</i>	<i>H</i>	
<i>S</i>	p_{11}	p_{12}	p_{13}	$p_{1.}$
<i>N</i>	p_{21}	p_{22}	p_{23}	$p_{2.}$
<i>H</i>	p_{31}	p_{32}	p_{33}	$p_{3.}$
	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{..} = 1$

$$B = p_{11} + p_{22} + p_{33}$$

$$E = p_{1.}p_{.1} + p_{2.}p_{.2} + p_{3.}p_{.3}$$

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

Modifizierungsvorschlag:

Verwende für den Korrekturterm nur die Anteile der Randverteilungen, die auch Anteil an der beobachteten Übereinstimmung haben!

	<i>S</i>	<i>N</i>	<i>H</i>	
<i>S</i>	p_{11}	p_{12}	p_{13}	$p_{1.}$
<i>N</i>	p_{21}	p_{22}	p_{23}	$p_{2.}$
<i>H</i>	p_{31}	p_{32}	p_{33}	$p_{3.}$
	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{..} = 1$

$$B = p_{11} + p_{22} + p_{33}$$

$$\begin{aligned}
 E &= p_{1.} p_{.1} + p_{2.} p_{.2} + p_{3.} p_{.3} \\
 &= (p_{11} + p_{12} + p_{13}) \cdot (p_{11} + p_{21} + p_{31}) + \\
 &\quad (p_{21} + p_{22} + p_{23}) \cdot (p_{12} + p_{22} + p_{32}) + \\
 &\quad (p_{31} + p_{32} + p_{33}) \cdot (p_{13} + p_{23} + p_{33})
 \end{aligned}$$

$$C = p_{11}^2 + p_{22}^2 + p_{33}^2$$

$$\Rightarrow \kappa_{ney} = \frac{B - C}{1 - C}$$

Mengentheoretische Betrachtungsweise:

$B :=$ Menge der Elemente, deren Summe B ergibt

$E :=$ Menge der Elemente deren Summe E ergibt

$C :=$ Menge der Elemente, deren Summe C ergibt

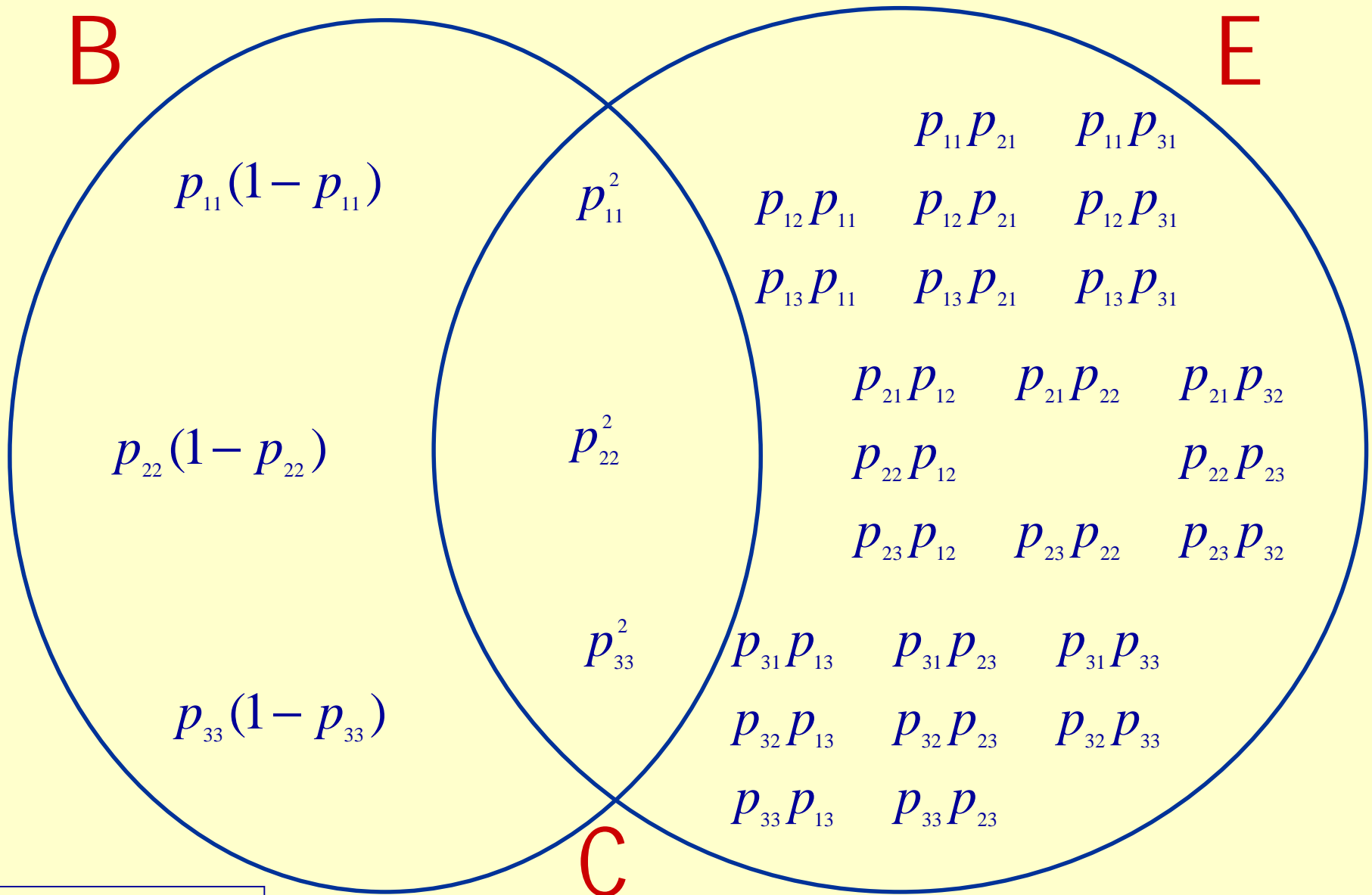
$$B = \{ p_{11}, p_{22}, p_{33} \}$$

$$E = \{ p_{11}^2, p_{11}p_{21}, \dots, p_{13}p_{31}, \\ p_{21}p_{12}, p_{21}p_{22}, \dots, p_{23}p_{32}, \\ p_{31}p_{13}, p_{31}p_{23}, \dots, p_{33}^2 \}$$

$$C = \{ p_{11}^2, p_{22}^2, p_{33}^2 \}$$

$$E = p_{11}p_{11} + p_{22}p_{22} + p_{33}p_{33} \\ = (p_{11} + p_{12} + p_{13}) \cdot (p_{11} + p_{21} + p_{31}) + \\ (p_{21} + p_{22} + p_{23}) \cdot (p_{12} + p_{22} + p_{32}) + \\ (p_{31} + p_{32} + p_{33}) \cdot (p_{13} + p_{23} + p_{33})$$

Mengendarstellung:



Beispiele (1):

Gleiches **B**, unterschiedliches **Kappa**:

$\frac{3}{9}$	0	0	$\frac{3}{9}$
0	0	0	0
$\frac{1}{9}$	$\frac{5}{9}$	0	$\frac{6}{9}$
$\frac{4}{9}$	$\frac{5}{9}$	0	1

$\frac{3}{9}$	0	$\frac{3}{9}$	$\frac{6}{9}$
0	0	0	0
$\frac{3}{9}$	0	0	$\frac{3}{9}$
$\frac{6}{9}$	0	$\frac{3}{9}$	1

$$\kappa = 0,21$$

$$\kappa_{ney} = 0,25$$

$$\kappa = -0,50$$

$$\kappa_{ney} = 0,25$$

Unterschiedliches **B**, gleiches **Kappa**:

$\frac{6}{9}$	0	$\frac{3}{9}$	1
0	0	0	0
0	0	0	0
$\frac{6}{9}$	0	$\frac{3}{9}$	1

$\frac{2}{9}$	0	$\frac{7}{9}$	1
0	0	0	0
0	0	0	0
$\frac{2}{9}$	0	$\frac{7}{9}$	1

$$\kappa = 0$$

$$\kappa_{ney} = 0,41$$

$$\kappa = 0$$

$$\kappa_{ney} = 0,18$$

Beispiele (2):

1. Kappa-Paradoxon:

$\frac{2}{9}$	0	0	$\frac{2}{9}$
$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{4}{9}$
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$
$\frac{3}{9}$	$\frac{3}{9}$	$\frac{3}{9}$	1

$$\kappa = 0,51$$

$$\kappa_{ney} = 0,61$$

$\frac{6}{9}$	0	0	$\frac{6}{9}$
$\frac{1}{9}$	0	$\frac{1}{9}$	$\frac{2}{9}$
0	$\frac{1}{9}$	0	$\frac{1}{9}$
$\frac{7}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	1

$$\kappa = 0,27$$

$$\kappa_{ney} = 0,41$$

2. Kappa-Paradoxon:

$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{5}{9}$
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$
0	0	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{1}{9}$	$\frac{3}{9}$	$\frac{5}{9}$	1

$$\kappa = 0,13$$

$$\kappa_{ney} = 0,30$$

$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{5}{9}$
$\frac{2}{9}$	$\frac{1}{9}$	0	$\frac{3}{9}$
$\frac{1}{9}$	0	0	$\frac{1}{9}$
$\frac{5}{9}$	$\frac{3}{9}$	$\frac{1}{9}$	1

$$\kappa = -0,18$$

$$\kappa_{ney} = 0,29$$

3. Kappa-Paradoxon:

0	0	0	0
$\frac{2}{9}$	0	0	$\frac{2}{9}$
$\frac{5}{9}$	$\frac{2}{9}$	0	$\frac{7}{9}$
$\frac{7}{9}$	$\frac{2}{9}$	0	1

$$\kappa = -0,04$$

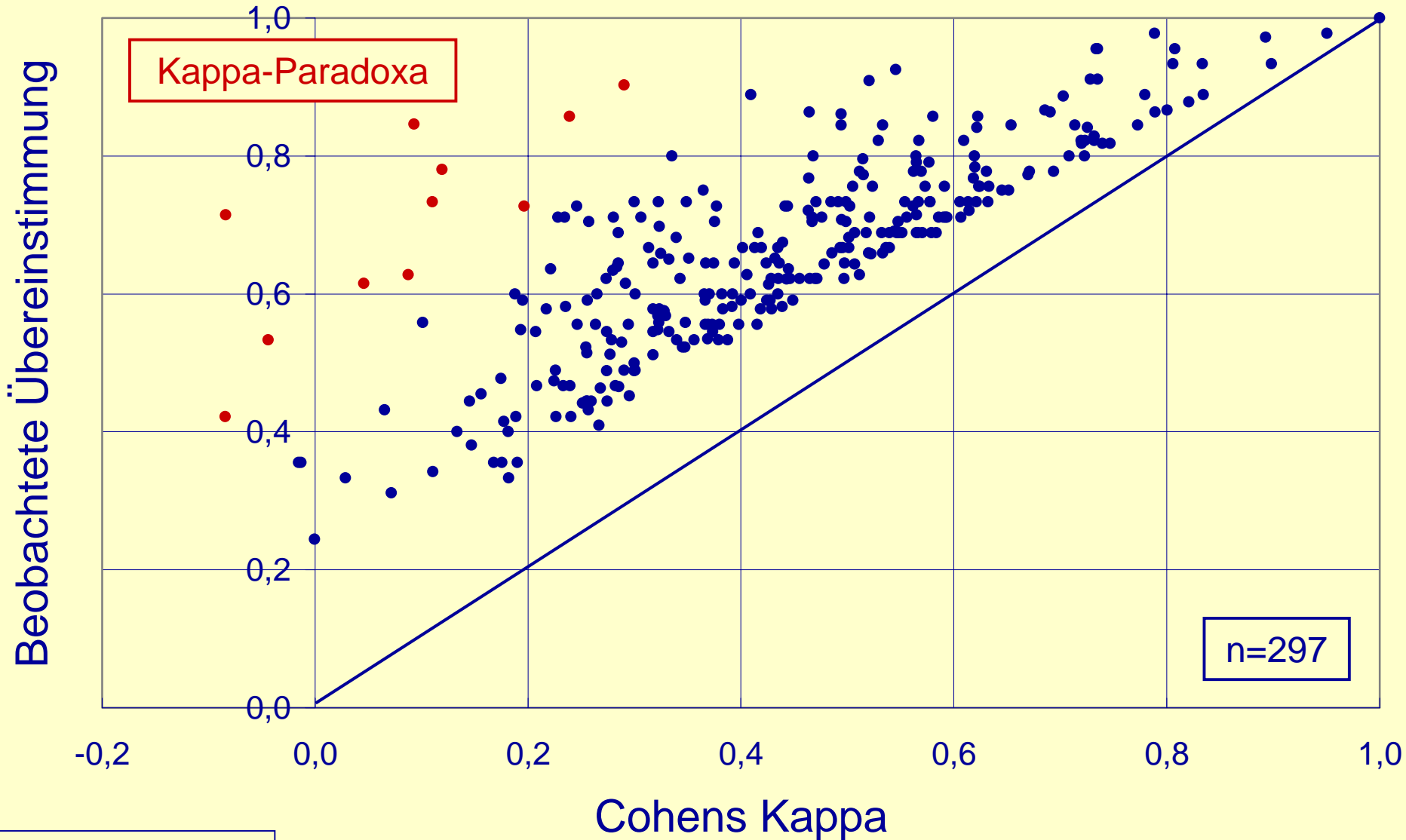
$$\kappa_{ney} = 0$$

0	0	$\frac{4}{9}$	$\frac{4}{9}$
0	0	0	0
$\frac{5}{9}$	0	0	$\frac{5}{9}$
$\frac{5}{9}$	0	$\frac{4}{9}$	1

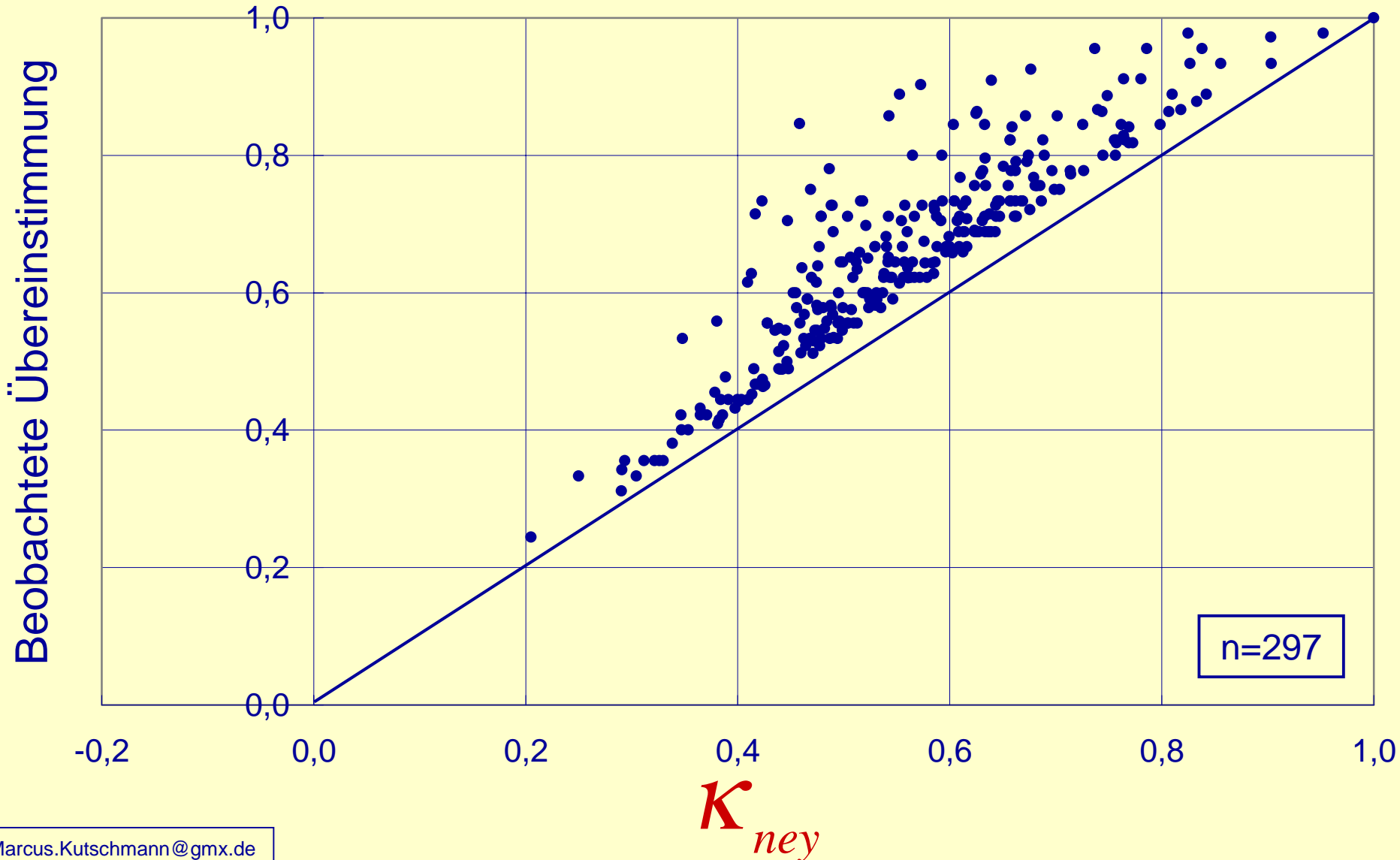
$$\kappa = -0,98$$

$$\kappa_{ney} = 0$$

Scatterplot B vs. Cohens Kappa (Beispiel):



Scatterplot B vs. K_{ney} (Beispiel):



- K_{ney} berücksichtigt nur den Anteil der zufallsmäßig zu erwartenden Übereinstimmung, der gleichzeitig auch Anteil der beobachteten Übereinstimmung ist.
- K_{ney} kann nicht negativ werden: Minimum ist 0, Maximum ist 1.
- K_{ney} ist (ebenso wie Cohens Kappa) nicht definiert, falls nur eine Zelle der Agreement-Matrix besetzt (mit 1).
- Ist keine Übereinstimmung zu beobachten, findet bei K_{ney} auch keine Zufallskorrektur statt.
- K_{ney} ist erweiterbar (multiple Ratings, Gewichtung, mehrere Kategorien).
- K_{ney} ist Paradoxa-frei.

Andere Ansätze zur Lösung der Paradoxa:

- Brennan & Prediger (1981):
Coefficient kappa: Some uses, misuses, and alternatives.
- Cicchetti & Feinstein (1990):
High agreement but low kappa: II. Resolving the paradoxes.
- Byrt, Bishop & Carlin (1993):
Bias, prevalence and kappa.
- Lantz & Nebenzahl (1996):
Behavior and interpretation of the kappa statistic: Resolution of the two paradoxes.

Die semantische Lösung:

Betrachte den Kappa-Koeffizienten nur als das, was er tatsächlich ist: Ein Koeffizient, der (die adjustierte) Differenz zwischen beobachteter Übereinstimmung und erwarteter Übereinstimmung misst.

Misst man Kappa also **nicht mehr die Bedeutung eines zufallskorrigierten Übereinstimmungskoeffizienten** bei, ergeben sich auch keine paradoxen Ergebnisse.

Ich bedanke mich für Ihre Aufmerksamkeit!