

Die Beurteilung wissenschaftlicher Leistungen durch vergleichende Begutachtung

von Udo Rempe, Kiel

Beurteilungen sind insbesondere wichtig bei:

1. der Auswahl für Juniorprofessuren oder Habilitationsstellen,
2. Zwischenevaluationen oder Habilitationen,
3. der Festsetzung der Höhe von Leistungsentgelten oder Leistungsbezügen durch Leistungsfeststellung für Forschung, Lehre und "Sonstiges",
4. Programmakkreditierung und
5. Prozessakkreditierung.

Artikel 33 Abs. 2 des Grundgesetzes bestimmt: "Jeder Deutscher hat nach Eignung, Befähigung und fachlicher Leistung gleichen Zugang zu jedem öffentlichen Amte."

Tritt bei n Beurteilungs-Stufen jede gleich häufig auf, beträgt die „optimale Varianz“ der Bewertungen $(n \cdot n - 1) / 12$.

Rechnet man mit den tatsächlich aufgetretenen Häufigkeiten, erhält man eine „tatsächliche Varianz“.

Ein einfaches Gütemaß für ein Beurteilungssystem und seine Umsetzung ist die „tatsächliche Varianz“ in Prozent der „optimalen Varianz“.

Selbermachen fördert das Verstehen!

1. Experiment und Übung:

Bitte beurteilen Sie spontan die Körperhöhe Ihres linken Nachbarn oder Ihrer linken Nachbarin als

sehr klein,

klein,

normal,

groß oder

sehr groß.

Schreiben Sie sich bitte das Ergebnis auf und korrigieren Sie es nicht im Laufe des Vortrags.

Falls Sie am linken Ende einer Reihe sitzen, beurteilen Sie bitte nach rechts hin.

Selbermachen fördert das Verstehen!

2. Experiment und Übung:

Bitte beurteilen sie die Körperhöhen ihrer beiden linken Nachbarinnen oder Nachbarn vergleichend!

Beurteilen Sie ohne Kenntnis der in den Personalausweisen ausgewiesenen Werte in cm, wer größer und kleiner ist!

Gegebenenfalls nach rechts oder nach links und rechts hin urteilen. Im Ernstfall muss ein Beurteilungsring gebildet werden (auf den Sitzplatz ganz links folgt der Sitzplatz ganz rechts).

1. Experiment mit I-Beurteilung: 18,42 %
 2. Experiment mit X-Beurteilung: 73,63 %
- BURL-Regelbeurteilungen (9 Stufen) in
Schleswig-Holstein:
insgesamt
2003 16,66 %,
2005 14,79 %,
- Staatskanzlei
2003 9,03 %,
2005 8,24 %

I-Beurteilung

Beurteilender
beurteilt
zu Beurteilenden

Petrus

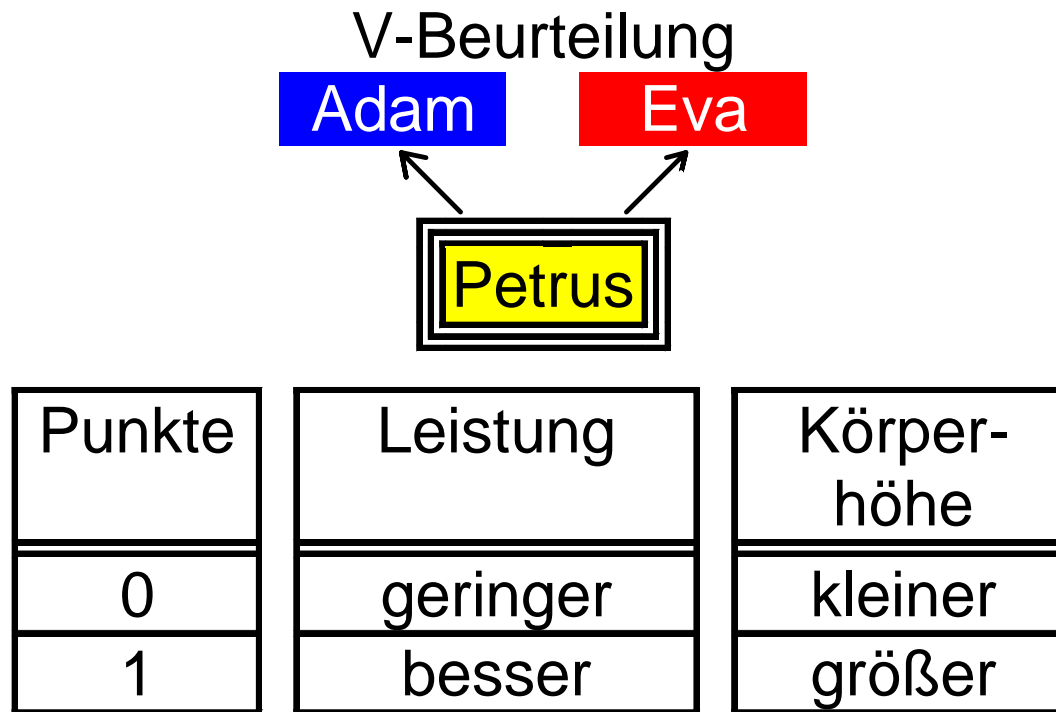


Adam

Punkte
0
1
2
3
4

Leistung
sehr schwach
schwach
normal
gut
sehr gut

Körper- höhe
sehr klein
klein
normal
groß
sehr groß



Beispiele: Beurteilung

Vorträge von	Körperhöhe von
1. Herrn Pätzold,	1. Herrn Soll
2. Herrn Schulze und	2. Frau Mehl
3. Herrn Rempe.	3. Herrn Ober

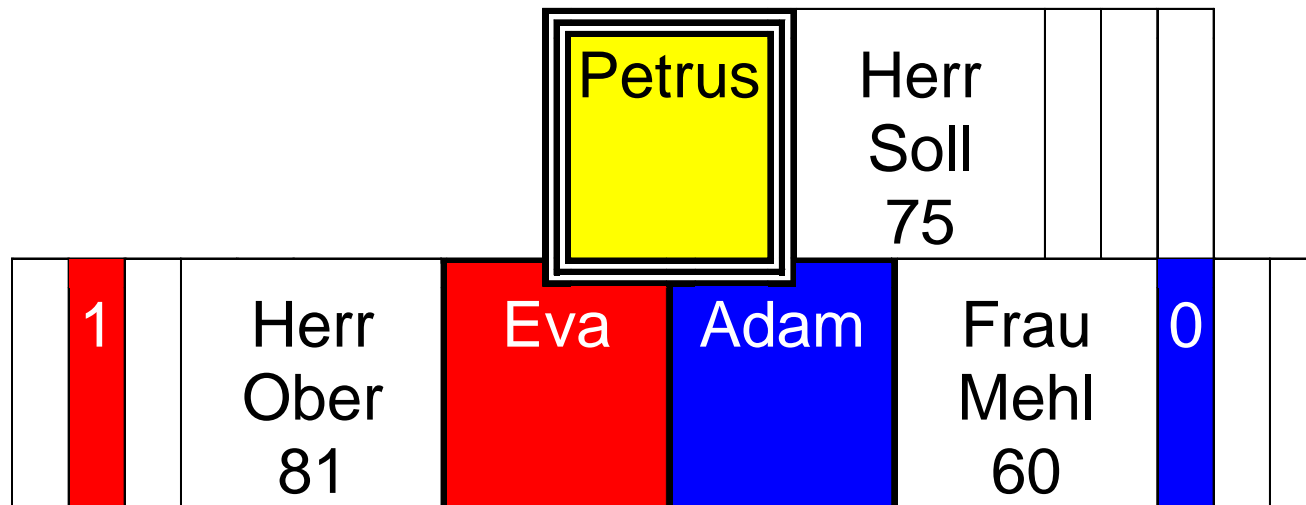
Sollten Herr Pätzold, Herr Schulze und Herr Rempe Leistungsentgelte oder Leistungsbezüge wünschen, sollten Sie nicht andere mit den erforderlichen Leistungsfeststellungen behelligen, sondern selber gutachten

Dazu bilden Sie einen Kreis von Gutachtern und zu Beurteilenden.

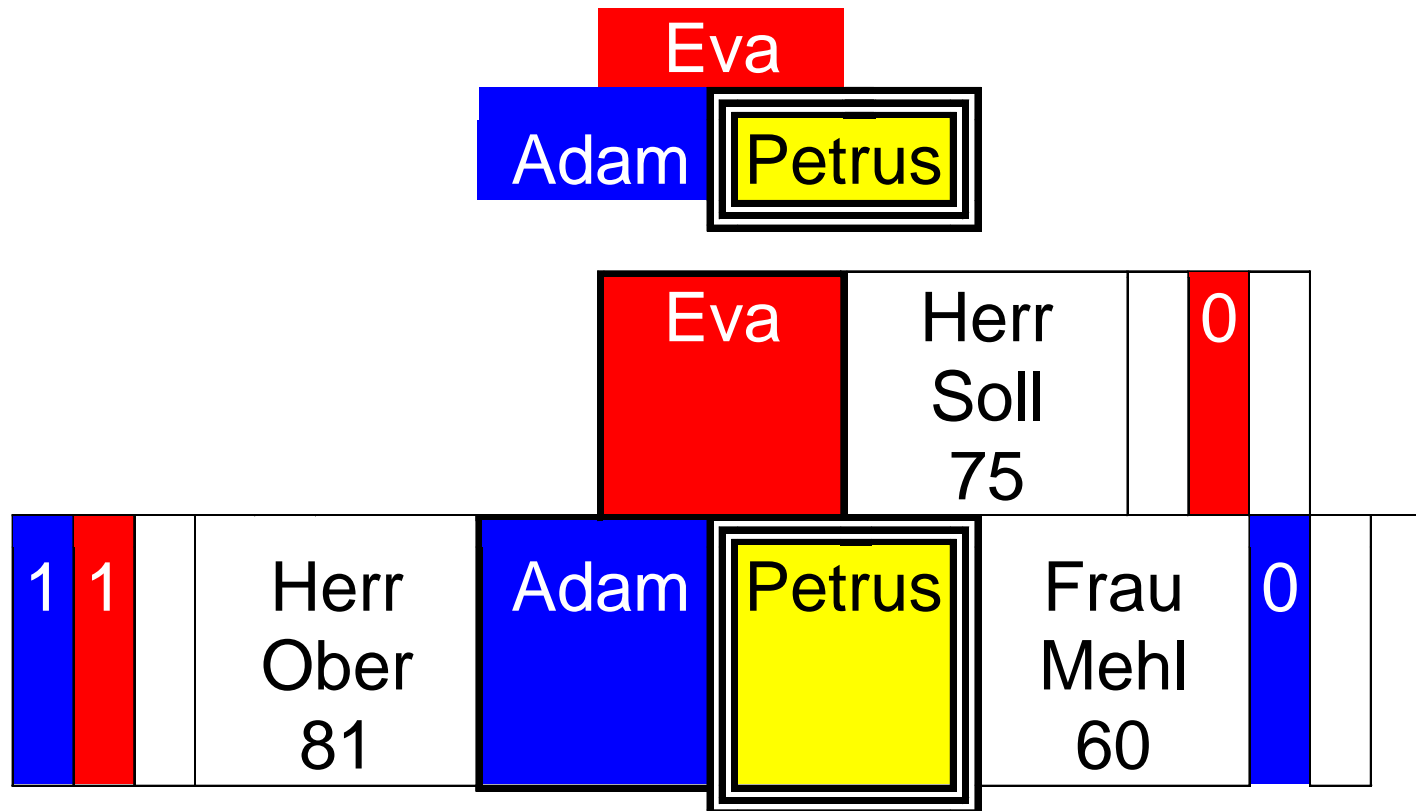
Wir stellen uns vor, dass die drei Personen tatsächlich im Kreis um einen Tisch sitzen, auf dem ein gelbes Set für die Petrus-Funktion, ein blaues Set für die Adam-Funktion und ein rotes Set für die Eva-Funktion liegt.

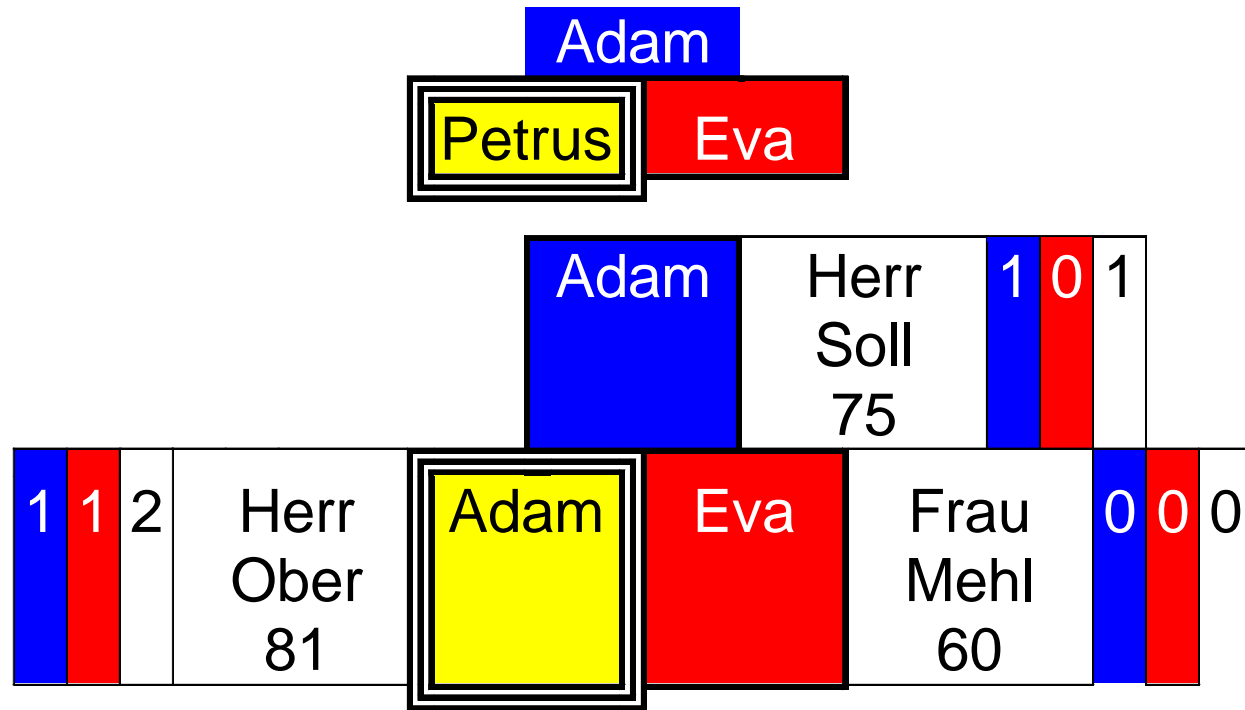


Leichter als die der Vortragsleistung ist die Körperhöhe
-100 cm beurteilbar



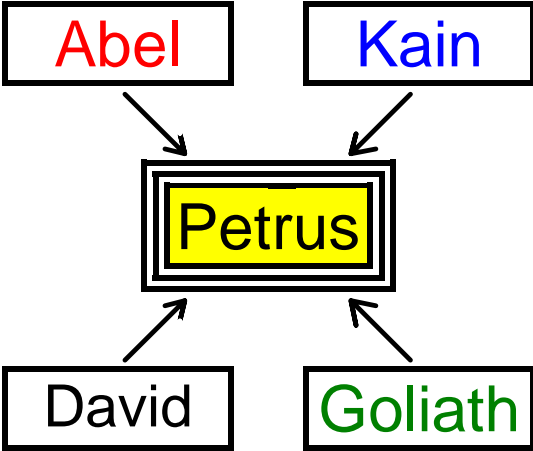
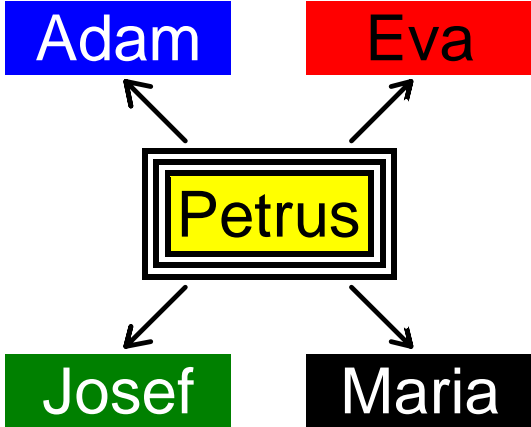
Die Funktionen werden wie „Sets“ auf einem Tisch im Uhrzeigersinn verschoben





Nachdem alle einmal geurteilt haben, sind alle zweimal beurteilt. Die Punktesummen reihen sie.

X-Beurteilung



13

Petrus	
Kain	Adam
A	B
Abel	Eva
David	Maria
Josef	Goliath
C	D

14

Petrus	
Kain	Adam
A	B
Abel	Eva
David	Maria
C	D
Josef	Goliath
E	

						Petrus	Soll 75							
				Mehl 60		Adam	Gerd 74	2						
				Ober 81			Franz 80							
				Vogt 80		Eva	Meier 82	3						
				Paul 74		Maria	Holz 73							
			0	Beer 71	Josef		Tann 74				1			
				Eisen 86			Hase 55							

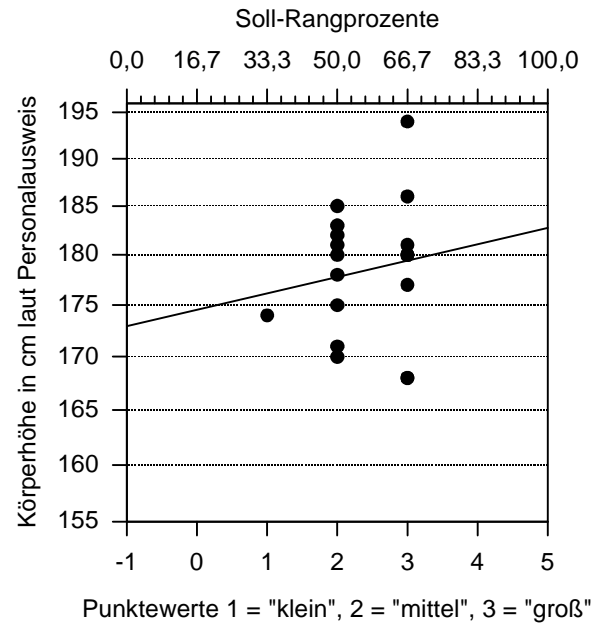
Petrus	Adam		Eva		Maria		Josef	
75 Soll	74 Gerd	2	82 Meier	3	73 Holz	1	71 Beer	0
74 Gerd	80 Franz	3	73 Holz	0	74 ⁻ Tann	1	74 ⁺ Paul	2
80 Franz	82 Meier	3	74 Tann	1	55 Hase	0	80 Vogt	2
82 Meier	73 Holz	1	55 Hase	0	86 Eisen	3	81 Ober	2
73 Holz	74 Tann	2	86 Eisen	3	71 Beer	1	60 Mehl	0
74 Tann	55 Hase	0	71 Beer	1	74 Paul	2	75 Soll	3
55 Hase	86 Eisen	3	74 ⁻ Paul	0	80 Vogt	2	74 ⁺ Gerd	1
86 Eisen	71 Beer	0	80 ⁺ Vogt	2	81 Ober	3	80 ⁻ Franz	1
71 Beer	74 Paul	1	81 Ober	2	60 Mehl	0	82 Meier	3
74 Paul	80 Vogt	3	60 Mehl	0	75 Soll	2	73 Holz	1
80 Vogt	81 Ober	3	75 Soll	2	74 ⁺ Gerd	1	74 ⁻ Tann	0
81 Ober	60 Mehl	1	74 Gerd	2	80 Franz	3	55 Hase	0
60 Mehl	75 Soll	0	80 Franz	1	82 Meier	2	86 Eisen	3

Herr Soll: 75	0	2	2	3	7
Frau Gerd: 74	2	2	1	1	6
Herr Franz: 80	3	1	3	1	8
Herr Meier: 82	3	3	2	3	11
Frau Holz: 73	1	0	1	1	3
Frau Tann: 74	2	1	1	0	4
Frau Hase: 55	0	0	0	0	0
Herr Eisen: 86	3	3	3	3	12
Frau Beer: 71	0	1	1	0	2
Frau Paul: 74	1	0	2	2	5
Herr Vogt: 80	3	2	2	2	9
Herr Ober: 81	3	2	3	2	10
Frau Mehl: 60	1	0	0	0	1

Hase	Mehl	Beer	Holz	Tann	Paul	Gerd	Soll	Franz	Vogt	Ober	Meier	Eisen
55	70	71	73	74 ⁻	74	74 ⁺	75	80 ⁻	80 ⁺	81	82	86
0	1	2	3	4	5	6	7	8	9	10	11	12
1	2	3	4	5	6	7	8	9	10	11	12	13
7	14	21	29	36	43	50	57	64	71	79	86	93
0	1	4	6	8	5	9	3	8	8	7	7	12
7	14	36	50	64	43	71	29	64	64	57	57	93

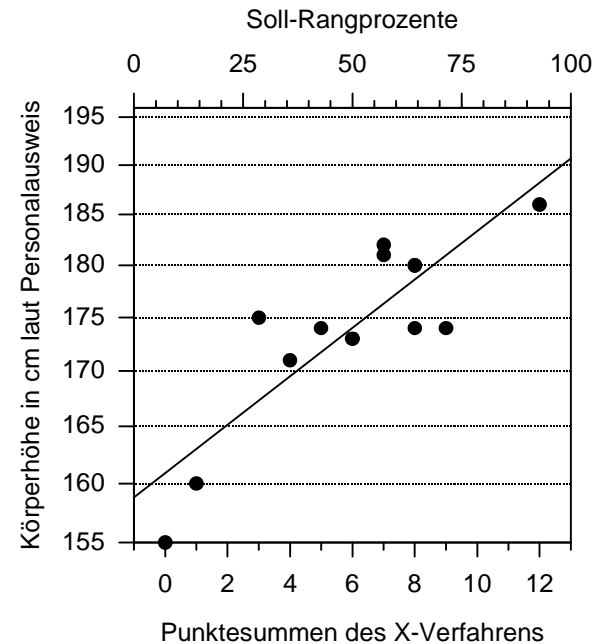
Körperhöhe in cm -100 cm	$\text{Rangprozent} = 100 \cdot \text{Rangzahl} / (\text{Anzahl} + 1)$
Punktesumme nach bekannten Messwerten	Punktesumme nach reiner Beurteilung
Rangzahl	$\text{Punktprozent} = 100 \cdot (\text{Punktesumme} + 1) / (\text{Anzahl} + 1)$

1. Experiment: I-Beurteilung: Ein Beurteiler urteilt ohne zu vergleichen



Das Bestimmtheitsmaß beträgt 0,0230.

2. Experiment: X-Beurteilung: Vier Beurteiler urteilen durch Vergleichen



Das Bestimmtheitsmaß beträgt 0,7357.

Rangkorrelation zwischen

den richtigen Rangzahlen

und

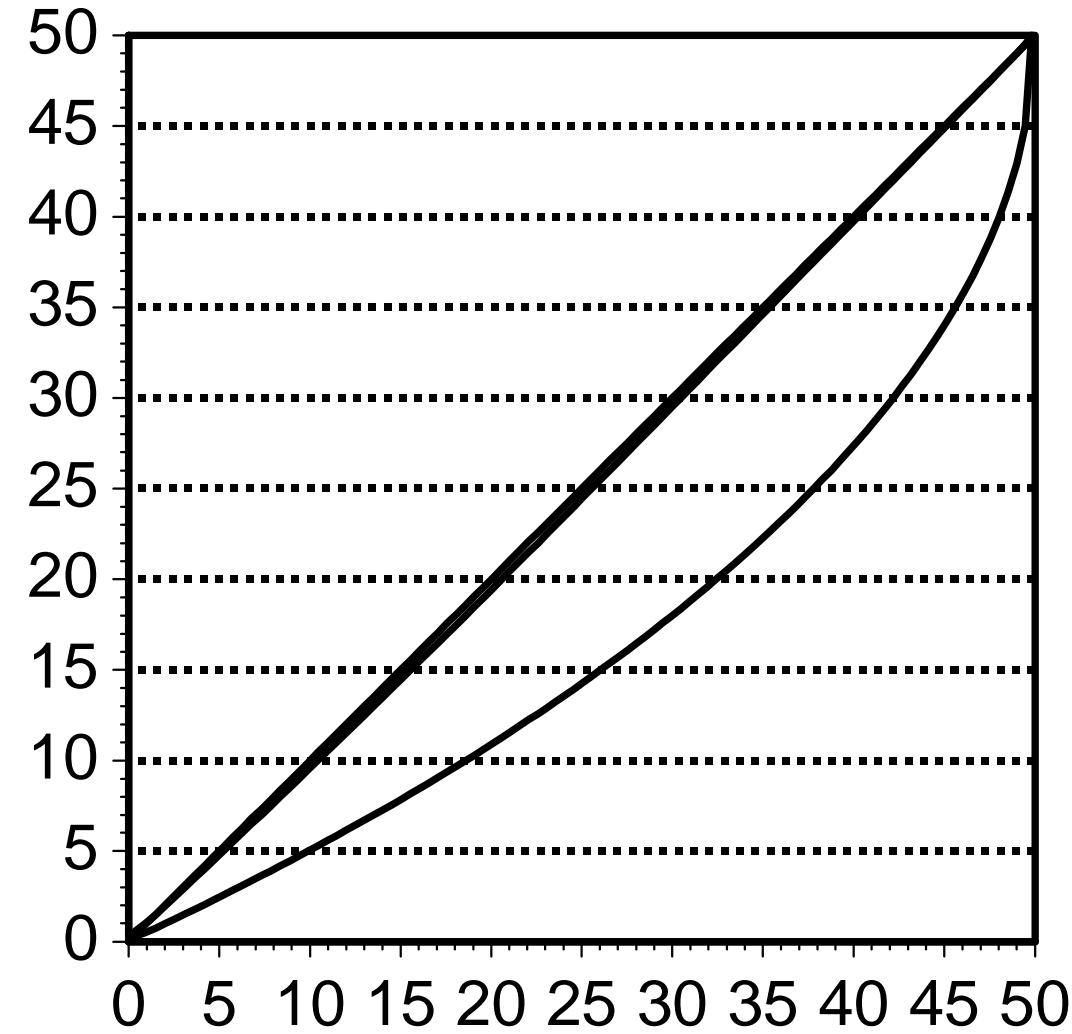
den nach Beurteilungsergebnissen ermittelten
Rangzahlen:

I-Beurteilung: 0,0754,

X-Beurteilung: 0,6676.

FAZIT: I-Beurteilung nahezu unbrauchbar.

$\arccos (0.222222 \cdot \text{Varianz} - 2.111111) \cdot 50 / \pi,$
Prozent Diskordanz, Prozent Fehlurteile



$\arccos (0.222222 \cdot \text{Varianz} - 2.111111)$
 $\cdot 50 / \pi$

Im vorliegenden Fall wird eine Varianz von 10,3077 erhalten. Die Diskordanz dazu beträgt 22,13 % und der Prozentsatz der Fehlerurteile 12,67 %. Etwa in einem Achtel der Fälle wird die größere Person für die kleinere gehalten.

Bei Leistungsbeurteilungen wird das ähnlich sein, wobei die Diskordanz aber zwischen den Fachgebieten schwankt.

Bonn	
Bielefe	Bonn
Bielefe	Bonn
Bielefe	Köln
Düssel	Köln
Düssel	Köln
Düssel	Münste
Duisbu	Münste
Duisbu	Münste
Duisbu	Wuppe
Dortmu	Wuppe
Dortmu	Wuppe
Dortmu	Aachen
Aachen	Aachen

Petrus	
Q	R
O	P
Kain	Adam
M	N
I	J
E	F
Abel	Eva
Josef	Goliath
K	L
G	H
A	B
C	D
Maria	David

Mögliche Pilotprojekte mit dem X-Verfahren:

1. Herr Rudinger organisiert die Beurteilung von Forschung, Lehre und Sonstigem bei den Psychologen in NRW.
2. Herr Hornbostel organisiert die Beurteilung von Forschung, Lehre und Sonstigem bei den Soziologen in NRW.
3. Herr Grüneberg setzt sich im Ministerium in NRW für eine Förderung ein.

Organisation über Ländergrenzen hinweg durch Fachbereichs- und Fakultätentage:

1. Vereinbarung fachgebundener Qualifikationsrahmen (BARTOSCH, RESCHAUER),
2. fachbezogene Programmakkreditierungen mit Y-Verfahren,
3. X-Beurteilung von Forschung, Lehre und Sonstigem der in den Studiengängen Tätigen.

Hinweis: Bis zu 80 Fachbereiche!

