Rangkorrelationskoeffizient nach Spearman r_{sp}

Zusammenhangsmaß für zwei (mindestens) ordinalskalierte abh. Größen

$$\boxed{r_{sp} = 1 - \frac{6 \sum (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1)}}$$

n: Stichprobenumfang

Anzahl der Datenpaare

R_{x_i}: Rangziffer der i-ten Beobachtung bzgl. X

R_{y_i}: analog

$$-1 \leq r_{sp} \leq 1$$

Interpretationen analog Bravais-Pearson

$r = 1$	perfekt	positiv	linear
$0,7 < r < 1$	stark	"	"
$0,3 \leq r \leq 0,7$	mäßig	"	"
$0 < r < 0,3$	schwach	"	"
$r = 0$	kein	linearer	

Extra: zur Vergabe von Rängen

Bsp. 1

x_i	25	36	18	9	23	47	26	5	21
R_{x_i}	6	8	3	2	5	9	7	1	4

Wir wollen die Ränge von den kleinen zu den großen Werten verteilen

Bsp. 2: Kommen bei X und/oder Y Ausprägungen mehrfach vor, spricht man so genannten Bindungen.

Die Rangziffer ergibt sich als arithmetischen Mittel der belegten Plätze.

x_i	22	23	20	20	21	21	20	21	23	21
R_{x_i}	8	9,5	2	2	5,5	5,5	2	5,5	9,5	5,5

i	1	2	3	4	5	6	7	8	9	10
x_i sort	20	20	20	21	21	21	21	22	23	23

R_{x_i}: $\frac{1+2+3}{3} = \underline{\underline{2}}$ $\quad \frac{4+5+6+7}{4} = \underline{\underline{5,5}}$ $\quad \frac{8}{1} = \underline{\underline{8}}$ $\quad \frac{9+10}{2} = \underline{\underline{9,5}}$

Diese Rangziffernvergabe bewirkt:

1) Die Rangziffer ist eine ganze Zahl oder eine ...,5 - Zahl !!!

2) $\sum R_{x_i} = \sum R_{y_i} = \frac{n}{2} \cdot (n+1)$

Aufgabe 5

X: Schriftliche Note in VWL

Y: Mündliche Note in VWL

Standardtabelle

x_i	y_i	R_{x_i}	R_{y_i}	$(R_{x_i} - R_{y_i})$	$(R_{x_i} - R_{y_i})^2$
1	1	1	2	-1	1
2	1	3	2	1	1
2	2	3	4,5	usw.	usw.
2	2	3	4,5		
3	1	6	2		
3	4	6	7		
3	5	6	9		
4	4	8	7		
5	4	9	7		
				$\underline{\underline{0}}$	$\sum (R_{x_i} - R_{y_i})^2$
					37,5

NIE NEGATIV

Bearbeitung der Ränge für X

i	1	2	3	4	5	6	7	8	9
x_i^*	1	2	2	2	3	3	3	4	5
sort	⋮								

$$\underbrace{2 \quad 2 \quad 2}_{\frac{2+3+4}{3} = \underline{\underline{3}}} \qquad \underbrace{3 \quad 3 \quad 3}_{\frac{5+6+7}{3} = \underline{\underline{6}}}$$

R_{x_i}: $\underline{\underline{1}}$, $\underline{\underline{3}}$, $\underline{\underline{6}}$, $\underline{\underline{8}}$, $\underline{\underline{9}}$

$$r_{sp} = 1 - \frac{6 \sum (R_{x_i} - R_{y_i})^2}{n(n^2-1)}$$

1

$$r_{sp} = 1 - \frac{6 \cdot 37.5}{9 \cdot (9^2-1)} = 0{,}6875$$

In der Stichprobe liegt ein mäßiger positiver linearer Zusammenhang zwischen X und Y vor.

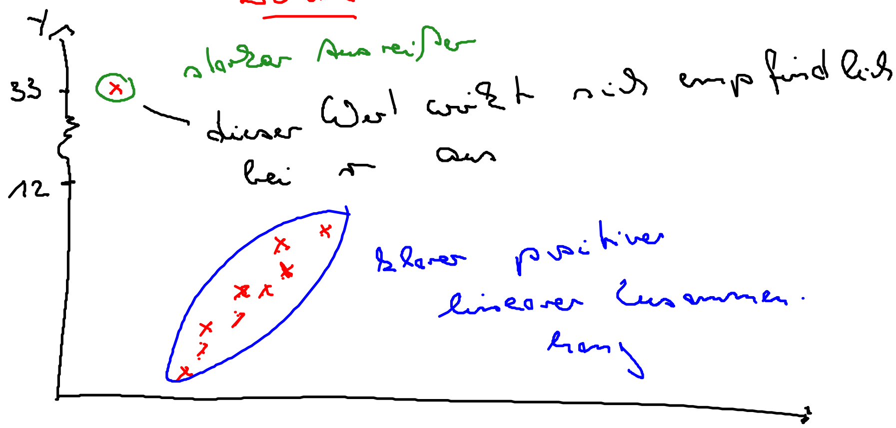
Aufgabe 4

1. Wert wird geändert

Tag	1	...
A-Bst	1	...
B-Bst	33	...

1) Streuungsdiagramm

Schema:



2) Standardtabelle

x_i	y_i	$x_i y_i$	x_i^2	y_i^2	R_{x_i}	R_{y_i}	$(R_{x_i}-R_{y_i})$	$(R_{x_i}-R_{y_i})^2$
		Bravais-Pearson						
							Spearman	

Bravais-Pearson

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}}$$

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$\bar{y} = \frac{1}{n}\sum y_i$$

Spearman

$$r_{sp} = 1 - \frac{6 \sum (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1)}$$

Musteraufgabe 3

erster Wert verändert

x_i	y_i	$x_i y_i$	x_i^2	y_i^2	R_{x_i}	R_{y_i}	$(R_{x_i} - R_{y_i})$	$(R_{x_i} - R_{y_i})^2$
1	33	33	1	1089	1	10	−9	81
2	1	2	4	1	2	1	1	1
4	2	8	16	4	3	2	1	1
5	4	20	25	16	4	3	1	1
6	5	30	36	25	5	4	1	1
8	6	48	64	36	6	5	1	1
9	8	72	81	64	7	6	1	1
10	9	90	100	81	8	7	1	1
12	10	120	144	100	9	8	1	1
13	12	156	169	144	10	9	1	1
70	90	579	640	1560			0	90
		3	4	5				

$n = 10$

$\bar{x} = \frac{1}{10} \cdot 70 = 7$

$\bar{y} = \frac{1}{10} \cdot 90 = 9$

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}}$$

3 4 5

$$r = \frac{579 - 10 \cdot 7 \cdot 9}{\sqrt{(640 - 10 \cdot 7^2)(1560 - 10 \cdot 9^2)}}$$

$n = 10$

$$r_{sp} = 1 - \frac{6 \cdot \sum (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1)}$$

$$r_{sp} = 1 - \frac{6 \cdot 90}{10 \cdot (10^2 - 1)}$$

$= 0{,}45$

$= \underline{\underline{-0{,}152}}$

In der Stichprobe liegt ein schwacher negativer linearer Zusammenhang zwischen X und Y vor.

In der Stichprobe liegt ein mäßiger positiver linearer Zusammenhang zwischen X und Y vor.

Ein Maß für nominalskalierte stat. Größen

→ normierter Kontingenzkoeffizient C_{korr}

korrigierter

in der nächsten Veranstaltung

Das ist nicht mehr für die Klausur relevant!

Ein Maß für <u>dichotome Merkmale</u>

Merkmale mit zwei Ausprägungen

ja / nein

vorhanden / nicht vorhanden

Vier-Felder-Phi (Φ) - Koeffizient

X \ Y	ja	nein	
ja	h_{11}	h_{12}	$h_{1\cdot}$
nein	h_{21}	h_{22}	$h_{2\cdot}$
	$h_{\cdot 1}$	$h_{\cdot 2}$	n

$$\Phi = \frac{h_{11} \cdot h_{22} - h_{21} \cdot h_{12}}{\sqrt{h_{1.} \cdot h_{2.} \cdot h_{.1} \cdot h_{.2}}}$$

X: Vorschule ja / Nein

Y: Kind Knabe / Mädchen

Ja / Nein

Mädchen / Knabe

X \ Y	Knabe	Mädchen	
Ja	20	10	30
Nein	30	40	70
	50	50	100

$$\Phi = \frac{20 \cdot 40 - 30 \cdot 10}{\sqrt{30 \cdot 70 \cdot 50 \cdot 50}}$$

$$\approx \underline{\underline{0{,}22}}$$

X \ Y	Mädchen	Knabe	
Ja	10	20	30
Nein	40	30	70
	50	50	100

$$\Phi = \frac{10 \cdot 30 - 40 \cdot 20}{\sqrt{30 \cdot 70 \cdot 50 \cdot 50}}$$

$$\approx \underline{\underline{-0{,}22}}$$

Nur der Betrag von Φ darf interpretiert werden

$$\underline{\underline{|\Phi| = 0{,}22}}$$

In der Stichprobe liegt zwischen X und Y nur ein schwacher Zusammenhang vor.