STATISTIK FÜR DIE SOZIALWISSENSCHAFTEN

Zusammenhänge ordinaler Variablen

Zusammenhangsmaße über Ränge und Paarvergleiche

Meine Notizen:

Bivariate Zusammenhänge nach Skalenniveaus

Symmetrisch

	Dichotome	Nominalskala	Ordinalskala	Metrische Skala
Metrische Skala			heute	Korrelation
Ordinalskala	heute		heute	
Nominalskala	Cramers V	Cramers V		
Dichotome V.	Phi, α			

Asym-metrisch

AV \ UV	Dichotome	Nominalskala	Ordinalskala	Metrische Skala
Metrische Skala	Mittelwert-vergleich, Regression		heute	Regression
Ordinalskala	heute		heute	heute
Nominalskala	Odds, λ, R´	Odds, λ, R´		
Dichotome V.	Prozentsatz-differenz, α	Odds, λ, R´	heute	

- Asymmetrische Zusammenhänge gehen von einer Wirkung einer unabhängigen Variablen (UV) auf eine abhängige Variable (AV) aus.
- Symmetrische Zusammenhänge sind geeignet, wenn keine (einseitige) kausale Wirkung unterstellt werden soll oder Wirkung in beide Richtungen angenommen werden muss.

Meine Notizen:

Bivariate Zusammenhänge bei ordinalen Daten

- Bei ordinalen Variablen liegt eine begrenzte Zahl von Antwortkategorien vor, mindestens jedoch gibt es drei Kategorien. (wie Nominalskalenniveau)
- Zusätzlich zum Nominalskalenniveau sind die Antwortkategorien in eine logische Reihenfolge/Anordnung zu bringen.
- Anders als auf metrischen Skalenniveaus sind die Distanzen zwischen den gereihten Antwortkategorien aber völlig offen.
- Daraus folgt, dass metrische Zusammenhangsmaße nicht berechnet werden sollten, da sie feste Distanzen benötigen.
- Nominale Zusammenhangsmaße können berechnet werden. Diese ignorieren aber die Information der Reihung und sind deshalb nicht informationsstark.

Meine Notizen:

Beispiel: Ordinale Daten mit nominalen Maßen

	Hauptschule	Mittlere Reife	Fachhochschulreife	Abitur	Gesamt
Hauptschule	732	59	3	16	810
Mittlere Reife	534	192	19	83	828
Fachhochschulreife	82	35	9	26	152
Abitur	223	132	33	239	627
Gesamt	1571	418	64	364	2417

	Hauptschule	Mittlere Reife	Fachhochschulreife	Abitur	Gesamt
Hauptschule	223	132	33	239	627
Mittlere Reife	82	35	9	26	152
Fachhochschulreife	534	192	19	83	828
Abitur	732	59	3	16	810
Gesamt	1571	418	64	364	2417

Da das nominale Niveau keine Reihenfolge kennt, produzieren beide Tabellen die gleichen nominalen Zusammenhangsmaße/ beide sind gleich in Cramers V, λ und R´.

Meine Notizen:

Beispiel: Ordinale Daten mit nominalen Maßen

	A	B	C	D	Gesamt
L	10	0	0	0	10
M	0	10	0	0	10
N	0	0	10	0	10
O	0	0	0	10	10
Gesamt	10	10	10	10	40

	A	B	C	D	Gesamt
L	0	0	0	10	10
M	0	0	10	0	10
N	0	10	0	0	10
O	10	0	0	0	10
Gesamt	10	10	10	10	40

	A	B	C	D	Gesamt
L	10	0	0	0	10
M	0	0	10	0	10
N	0	10	0	0	10
O	0	0	0	10	10
Gesamt	10	10	10	10	40

	A	B	C	D	Gesamt
L	0	0	0	10	10
M	0	10	0	0	10
N	0	0	10	0	10
O	10	0	0	0	10
Gesamt	10	10	10	10	40

Da das nominale Niveau keine Reihenfolge kennt, produzieren beide Tabellen die gleichen nominalen Zusammenhangsmaße/ beide sind gleich in Cramers V, λ und R´.

Meine Notizen:

Beispiel: Ordinale Daten mit metrischen Maßen

	A	B	C	D	Gesamt
L	10	0	0	0	10
M	0	10	0	0	10
N	0	0	10	0	10
O	0	0	0	10	10
Gesamt	10	10	10	10	40

Da das ordinale Niveau keine Abstände kennt, können dieselben Daten unter verschiedenen Abstandsannahmen zu verschiedenen Geraden führen.

Meine Notizen:

Bivariate Zusammenhänge bei ordinalen Daten

- Grundproblem also:
  - Ordinale Variablen enthalten Ranginformationen, aber keine Abstandsinformationen!
- Möglichkeit 1:
  - Behandeln als nominalskalierte Variablen = Ignorieren der ordinalen Information
  - Ist methodisch korrekt, setzt aber durch seine Ignoranz verschiedene Fälle gleich.
- Möglichkeit 2:
  - Behandeln als metrische Variablen = Ignorieren, dass keine metrischen Informationen vorliegen
  - Ist methodisch nicht korrekt, aber leicht (interpretierbar).
- Folgeproblem:
  - Zusammenhangsmaße, die über das nominale Mess-niveau hinausgehen sollen, brauchen immer eine Annahme über Abstände.

Meine Notizen:

Abstände bei ordinalen Daten

- Grundproblem:
  - Eigentlich liegt es in der Natur des Skalenniveaus, dass die Abstände der Ausprägungen unbekannt sind.
  - Für Zusammenhangsmaße, die über das nominale Messniveau hinausgehen sollen, brauchen wir immer eine Annahme über Abstände.
- Ansatz 1: Behandeln als metrische Variable; Bedeutet letztlich die Annahme:
  - Benachbarte Ausprägungen haben immer gleiche Abstände (Äquidistanz)
  - Wird bei ordinaler Messung latenter Konstrukte oft als sinnvoll angenommen.
  - Auch für Zustimmungs-Items der meist gewählte Weg.
- Ansatz 2: Aufbrechen des Zusammenhangs in Vielzahl von Paarvergleichen
- Ansatz 3: Analyse von Rangplätzen statt von Ausprägungen
  - Statt gleichen Abständen zwischen den Ausprägungen wird gleicher Abstand zwischen den Rangplätzen angenommen (kennen Sie vom IQR).

Meine Notizen:

Analyse von Rangplätzen statt von Ausprägungen

- Lösung des Grundproblems:
  - Umrechnung ordinaler Ausprägungen in Rangwerte und Behandeln von Rangwerten wie metrische Daten!
- Berechnung:
  - Schritt 1: Aufsteigende Sortierung aller Fälle;
  - Schritt 2:

kleinster Wert	→ Rangplatz 1:	$x_{(1)}$	→ $r_X = 1$
zweitkleinster Wert	→ Rangplatz 2:	$x_{(2)}$	→ $r_X = 2$
...		...	
zweitgrößter Wert	→ Rangplatz n-1:	$x_{(n-1)}$	→ $r_X = n-1$
größter Wert	→ Rangplatz n:	$x_{(n)}$	→ $r_X = n$

  - Mögliches Problem: Fälle mit gleicher Ausprägung:
    → Verwendung mittlerer Ränge bei gleichen Ausprägungen

Meine Notizen:

Beispiel der Berechnung von Rangplätzen

- Berechnung:
  - Schritt 1: Aufsteigende Sortierung aller Fälle;
  - Schritt 2: mit Rangplätzen belegen
  - Mögliches Problem: Fälle mit gleicher Ausprägung:
    → Verwendung mittlerer Ränge bei gleichen Ausprägungen
- Beispiel: {1,2,3,4,4,5,4,1,4,1,3}
  - Schritt 1: sortieren: {1,1,1,2,3,3,4,4,4,4,5}
  - Schritt 2: Rangwerte 1,2,3,4,5,6,7,8,9,10,11
    Fall 1-3 haben nach Sortierung gleichen Rangplatz: (1+2+3)/3 = 2;
    Fall 5-6 haben nach Sortierung gleichen Rangplatz: (5+6)/2 = 5.5;
    Fall 7-10 haben nach Sortierung gleichen
    Rangplatz: (7+8+9+10)/4= 8,5
- Resultierende sortierte Rangwerte der 11 Fälle:

  {2,2,2,4,5.5,5.5,8.5,8.5,8.5,8.5,11}

Meine Notizen:

Bivariate Zusammenhänge ordinaler Daten

- Eine Möglichkeit besteht in der Umrechnung ordinaler Ausprägungen in Rangwerte und der Behandlung von Rangwerten als metrische Daten. (Ansatz 3)
- Wir können also die bekannten metrischen Maße (Korrelation, Regression) berechnen und speisen nur Rangplätze statt Ausprägungen ein.
- Vorteil: Wir finden eine Abstandsberechnung, die in der Reihungseigenschaft begründet ist.
- Nachteil: Der Abstand zweier Ausprägungen hängt von meiner Stichprobe ab.
- Für die Korrelation nennt sich dies Verfahren Spearmans Rang-Korrelation
  = Produktmomentkorrelation zwischen zwei Variablen auf
  der Basis der Rangwerte aller Realisationen

Meine Notizen:

Spearmans Rang-Korrelation

- Für die Korrelation nennt sich das Verfahren nach Ansatz 3 Spearmans Rang-Korrelation = Produktmomentkorrelation zwischen zwei Variablen auf der Basis der Rangwerte aller Realisationen:

$$r_s = \frac{Kovariation\ nach\ Spearman}{\sqrt{Variation\ von\ X\ nach\ Spearman * Variation\ von\ Y\ nach\ Spearman}} =$$

$$\frac{\sum_{i=1}^{n} r_{X,i} * r_{Y,i} - \frac{\sum_{i=1}^{n} r_{X,i} * \sum_{i=1}^{n} r_{Y,i}}{n}}{\sqrt{\left(\sum_{i=1}^{n} r_{X,i}^2 - \frac{\left(\sum_{i=1}^{n} r_{X,i}\right)^2}{n}\right) * \left(\sum_{i=1}^{n} r_{Y,i}^2 - \frac{\left(\sum_{i=1}^{n} r_{Y,i}\right)^2}{n}\right)}}$$

wobei:

r_s = Spearmans Rang-Korrelationskeoffizient;

$r_{X,i}$ = Rangplatz von Fall i bei Variable X;

$r_{Y,i}$ = Rangplatz von Fall i bei Variable Y;

Meine Notizen:

Spearmans Rang-Korrelation: ein Beispiel

Beispiel:	Fall	X	Y	r_X	r_Y	r_X^2	r_Y^2	$r_X \cdot r_Y$
	1	1	2	1.5	2.0	2.25	4.00	3.00
	2	3	4	4.0	4.5	16.00	20.25	18.00
	3	1	4	1.5	4.5	2.25	20.25	6.75
	4	6	3	5.0	3.0	25.00	9.00	15.00
	5	2	1	3.0	1.0	9.00	1.00	3.00
Summe:				15.0	15.0	54.50	54.50	45.75

$$r_s = \frac{\sum_{i=1}^{n} r_{X,i} * r_{Y,i} - \frac{\sum_{i=1}^{n} r_{X,i} * \sum_{i=1}^{n} r_{Y,i}}{n}}{\sqrt{\left(\sum_{i=1}^{n} r_{X,i}^2 - \frac{\left(\sum_{i=1}^{n} r_{X,i}\right)^2}{n}\right) * \left(\sum_{i=1}^{n} r_{Y,i}^2 - \frac{\left(\sum_{i=1}^{n} r_{Y,i}\right)^2}{n}\right)}}$$

$$= \frac{45{,}75 - \frac{15 * 15}{5}}{\sqrt{\left(54{,}5 - \frac{15^2}{5}\right) * \left(54{,}5 - \frac{15^2}{5}\right)}} = 0{,}079$$

Meine Notizen:

Berechnung von Rangwerten bei Kreuztabellen

Eigene wirtsch. Lage	Allgem. wirtschaftliche Lage (1)	(2)	(3)	(4)	(5)	$\sum$
Sehr gut (1)	48	74	32	5	1	160
Gut (2)	73	645	526	72	6	1322
Teils,teils (3)	12	143	361	115	16	647
Schlecht (4)	5	30	68	64	11	178
Sehr schlecht (5)	1	6	14	11	6	38
$\sum$	139	898	1001	267	40	2345
Kumul. Summen	139	1037	2038	2305	2345	
Rangplatz:	70	588.5	1538	2172	2325.5	

70=(1+139)/2; 588.5=(139+1+1037)/2; 1538=(1037+1+2038)/2; 2172=(2038+1+2305)/2; 2325.5=(2305+1+2345)/2

Kumul. Summen	Rang-platz	
160	80.5	= (1+160)/2
1482	821.5	= (160+1+1482)/2
2129	1806	= (1482+1+212)/2
2307	2218.5	= (2129+1+2307)/2
2345	2326.5	= (2307+1+2345)/2

1. Kumulieren der Häufigkeiten
2. Rangplatz einer Kategorie = (Anzahl kumul. Fälle bis zur vorhergehenden Kateg.+ 1 + Anzahl kumul. Fälle bis Kateg.) / 2

Meine Notizen:

Spearmans Rang-Korrelation an größeren Kreuztabellen

Eigene wirtsch. Lage	Allgem. wirtschaftliche Lage (1)	(2)	(3)	(4)	(5)	$\sum$	Rang-Platz	$\sum r_Y$	$\sum r^2_Y$
Sehr gut (1)	48	74	32	5	1	160	80.5	12880	1036840
Gut (2)	73	645	526	72	6	1322	821.5	1086023	892167894.5
Teils,teils (3)	12	143	361	115	16	647	1806	1168482	2110278492.
Schlecht (4)	5	30	68	64	11	178	2218.5	394893	876070120.5
Sehr schlecht (5)	1	6	14	11	6	38	2326.5	88407	205678885.5
$\sum$	139	898	1001	267	40	2345	Summe:	2750685	4085232233.
Rangplatz	70	588.5	1538	2172	2325.5				

$\sum r_X$ = 2750685 = $n\cdot(n+1)/2$

$\sum r^2_X$ = 4155409843 = $139\cdot70^2 + \ldots + 40\cdot2325.5^2$

$\sum r_X\cdot r_Y$ = 3587487381 = $48\cdot80.5\cdot70 + \ldots + 6\cdot2326.5\cdot2325.5$

$$r_s = \frac{Kovariation\ nach\ Spearman}{\sqrt{Variation\ X\ nach\ Spearman\ * Variation\ Y\ nach\ Spearman}} = \frac{\sum_{i=1}^{n} r_{X,i} * r_{Y,i} - \frac{\sum_{i=1}^{n} r_{X,i} * \sum_{i=1}^{n} r_{Y,i}}{n}}{\sqrt{\left(\sum_{i=1}^{n} r^2_{X,i} - \frac{\left(\sum_{i=1}^{n} r_{X,i}\right)^2}{n}\right) * \left(\sum_{i=1}^{n} r^2_{Y,i} - \frac{\left(\sum_{i=1}^{n} r_{Y,i}\right)^2}{n}\right)}}$$

$$= \frac{\sum_{i=1}^{n} r_{X,i} * r_{Y,i} - \frac{n*(n+1)^2}{4}}{\sqrt{\left(\sum_{i=1}^{n} r^2_{X,i} - \frac{n*(n+1)^2}{4}\right) * \left(\sum_{i=1}^{n} r^2_{Y,i} - \frac{n*(n+1)^2}{4}\right)}}$$

$$= \frac{3587487381 - \frac{2750685 * 2750685}{2345}}{\sqrt{\left(4155409843 - \frac{2750685^2}{2345}\right) * \left(4085232233 - \frac{2750685^2}{2345}\right)}} = 0{,}404$$

Meine Notizen:

Bivariate Zusammenhänge ordinaler Daten: Ansatz 3

- Eine Möglichkeit besteht in der Umrechnung ordinaler Ausprägungen in Rangwerte und der Behandlung von Rangwerten als metrische Daten. (Ansatz 3)
- Wir können also die bekannten metrischen Maße (Korrelation, Regression) berechnen und speisen nur Rangplätze statt Ausprägungen ein.
- Vorteil: Wir finden eine Abstandsberechnung, die in der Reihungseigenschaft begründet ist.
- Nachteil: Der Abstand zweier Ausprägungen hängt von meiner Stichprobe ab.
- Dieselbe Idee lässt sich auch auf einen T-Test übertragen, der zwei Gruppen in einer ordinalen Variable vergleichen möchte. Verfahren dazu sind bekannt als U-Test, Wilcoxon-Test oder Mann-Whitney-Test.

Meine Notizen:

Abstände bei ordinalen Daten

- Grundproblem:
  - Eigentlich liegt es in der Natur des Skalenniveaus, dass die Abstände der Ausprägungen unbekannt sind.
  - Für Zusammenhangsmaße, die über das nominale Messniveau hinausgehen sollen, brauchen wir immer eine Annahme über Abstände.
- Ansatz 1: Behandeln als metrische Variable; Bedeutet letztlich die Annahme:
  - Benachbarte Ausprägungen haben immer gleiche Abstände (Äquidistanz)
  - Wird bei ordinaler Messung latenter Konstrukte oft als sinnvoll angenommen.
  - Auch für Zustimmungs-Items der meist gewählte Weg.
- Ansatz 2: Aufbrechen des Zusammenhangs in Vielzahl von Paarvergleichen
- Ansatz 3: Analyse von Rangplätzen statt von Ausprägungen
  - Statt gleichen Abständen zwischen den Ausprägungen wird gleicher Abstand zwischen den Rangplätzen angenommen (kennen Sie vom IQR).

Meine Notizen:

Ansatz 2: Aufbrechen des Zusammenhangs in Vielzahl von Paarvergleichen

- Wir gehen von einem Datensatz mit n Befragten aus.
- Wir interessieren uns für den Zusammenhang zwischen zwei ordinalen Variablen.
- Wir wählen 2 Befragte aus und vergleichen die Ausprägungen in X:

  Ist: $x_i > x_j$ oder $x_i < x_j$ oder $x_i = x_j$?

- Wir schauen uns die Ausprägungen in Y an. Es liegt einer von 5 möglichen Paartypen vor:
  - $x_i > x_j$ & $y_i > y_j$ $\rightarrow$ ***konkordantes*** (gleichgerichtetes) Paar;
  - $x_i > x_j$ & $y_i < y_j$ $\rightarrow$ ***diskordantes*** (gegengerichtetes) Paar;
  - $x_i = x_j$ & $y_i \neq y_j$ $\rightarrow$ ***x-verbundenes*** (engl.: tied x) Paar;
  - $x_i \neq x_j$ & $y_i = y_j$ $\rightarrow$ ***y-verbundenes*** (engl.: tied y) Paar;
  - $x_i = x_j$ & $y_i = y_j$ $\rightarrow$ ***x,y-verbundenes*** (engl.: tied x,y) Paar.

Wir müssen jedes mögliche Paar (es gibt 0,5*(n-1)*n) einem der 5 Typen zuordnen!

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

Eigene wirtsch. Lage (Y)	Allg. wirtschaftliche Lage (X)				
	(1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Bei gleicher Anordnung beider Variablen:

Konkordante Paare:	Fälle in Zellen rechts unterhalb
Diskordante Paare:	Fälle in Zellen links unterhalb
X-verbundene P.:	Fälle in Zellen gleicher Spalte
Y-verbundene P.:	Fälle in Zellen gleicher Zeile
X,Y-verbundene P.:	Fälle in gleicher Zelle

Fall i: X=2 Y=1
Fall j: X=4 Y=3
→ konkordant

Fall i: X=2 Y=1
Fall j: X=1 Y=3
→ diskordant

Fall i: X=2 Y=1
Fall j: X=2 Y=4
→ x-verbunden

Fall i: X=2 Y=1
Fall j: X=5 Y=1
→ y-verbunden

Fall i: X=2 Y=1
Fall j: X=2 Y=1
→ x,y-verbunden

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

Eigene wirtsch. Lage (Y)	Allg. wirtschaftliche Lage (X) (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Bei gleicher Anordnung beider Variablen:

Konkordante Paare:	Fälle in Zellen rechts unterhalb
Diskordante Paare:	Fälle in Zellen links unterhalb
X-verbundene P.:	Fälle in Zellen gleicher Spalte
Y-verbundene P.:	Fälle in Zellen gleicher Zeile
X,Y-verbundene P.:	Fälle in gleicher Zelle

Anzahl konkordanter Paare C:

$$\begin{aligned}
C = {} & 48\cdot(645+526+72+6 +143+361+115+16 +30+68+64+11 +6+14+11+6) \\
& + 74\cdot(526+72+6 +361+115+16 +68+64+11 +14+11+6) \\
& + 32\cdot(72+6 +115+16 +64+11+11+6) \quad + 5\cdot(6+16+11+6) \\
& + 73\cdot(143+361+115+16 +30+68+64+11 +6+14+11+6) \\
& + 645\cdot(361+115+16 +68+64+11 +14+11+6) \\
& + 526\cdot(115+16+64+11+11+6) \quad + 72\cdot(16 +11+6) \\
& + 12\cdot(30+68+64+11 +6+14+11+6) +143\cdot(68+64+11 +14+11+6) \\
& + 361\cdot(64+11+11+6) + \quad 115\cdot(11+6) \\
& +5\cdot(6+14+11+6) +30\cdot(14+11+6) +68\cdot(11+6) +64\cdot 6 \\
= {} & 880472
\end{aligned}$$

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

Eigene wirtsch. Lage (Y)	Allg. wirtschaftliche Lage (X) (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Bei gleicher Anordnung beider Variablen:

Konkordante Paare:	Fälle in Zellen rechts unterhalb
Diskordante Paare:	Fälle in Zellen links unterhalb
X-verbundene P.:	Fälle in Zellen gleicher Spalte
Y-verbundene P.:	Fälle in Zellen gleicher Zeile
X,Y-verbundene P.:	Fälle in gleicher Zelle

Anzahl diskordanter Paare D:

$$
\begin{aligned}
D = {} & 74\cdot(73+12+5+1) + 32\cdot(73+645+12+143+5+30+1+6) \\
& + 5\cdot(73+645+526+12+143+361+5+30+68+1+6+14) \\
& + 1\cdot(73+645+526+72+12+143+361+115+5+30+68+64+1+6+14+11) \\
& + 645\cdot(12+5+1) + 526\cdot(12+143+5+30+1+6) + 72\cdot(12+143+361+5+30+68+1+6+14) \\
& + 6\cdot(12+143+361+115+5+30+68+64+1+6+14+11) \\
& + 143\cdot(5+1) + 361\cdot(5+30+1+6) + 115\cdot(5+30+68+1+6+14) \\
& + 16\cdot(5+30+68+64+1+6+14+11) \\
& + 30\cdot 1 + 68\cdot(1+6) + 64\cdot(1+6+14) + 11\cdot(1+6+14+11) \\
= {} & 249538
\end{aligned}
$$

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

Eigene wirtsch. Lage (Y)	Allg. wirtschaftliche Lage (X) (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Bei gleicher Anordnung beider Variablen:

Konkordante Paare:	Fälle in Zellen rechts unterhalb
Diskordante Paare:	Fälle in Zellen links unterhalb
X-verbundene P.:	Fälle in Zellen gleicher Spalte
Y-verbundene P.:	Fälle in Zellen gleicher Zeile
X,Y-verbundene P.:	Fälle in gleicher Zelle

Anzahl x-verbundener Paare Tx:

$$
\begin{aligned}
T_X = \; & 48\cdot(73+12+5+1) + 73\cdot(12+5+1) + 12\cdot(5+1) + 5\cdot 1 \\
& + 74\cdot(645+143+30+6) + 645\cdot(143+30+6) + 143\cdot(30+6) + 30\cdot 6 \\
& + 32\cdot(526+361+68+14) + 526\cdot(361+68+14) + 361\cdot(68+14) + 68\cdot 14 \\
& + 5\cdot(72+115+64+11) + 72\cdot(115+64+11) + 115\cdot(64+11) + 64\cdot 11 \\
& + 1\cdot(6+16+11+6) + 6\cdot(16+11+6) + 16\cdot(11+6) + 11\cdot 6 \\
= \; & 506992
\end{aligned}
$$

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

Eigene wirtsch. Lage (Y)	Allg. wirtschaftliche Lage (X) (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Bei gleicher Anordnung beider Variablen:

Konkordante Paare:	Fälle in Zellen rechts unterhalb
Diskordante Paare:	Fälle in Zellen links unterhalb
X-verbundene P.:	Fälle in Zellen gleicher Spalte
Y-verbundene P.:	Fälle in Zellen gleicher Zeile
X,Y-verbundene P.:	Fälle in gleicher Zelle

Anzahl x,y-verbundener Paare Txy:

$$
\begin{aligned}
T_{XY} = {} & 48\cdot(48-1)/2 + 74\cdot(74-1)/2 + 32\cdot(32-1)/2 + 5\cdot(5-1)/2 + 1\cdot(1-1)/2 \\
& + 73\cdot(73-1)/2 + 645\cdot(645-1)/2 + 526\cdot(526-1)/2 + 72\cdot(72-1)/2 + 6\cdot(6-1)/2 \\
& + 12\cdot(12-1)/2 + 143\cdot(143-1)/2 + 361\cdot(361-1)/2 + 115\cdot(115-1)/2 + 16\cdot(16-1)/2 \\
& + 5\cdot(5-1)/2 + 30\cdot(30-1)/2 + 68\cdot(68-1)/2 + 64\cdot(64-1)/2 + 11\cdot(11-1)/2 \\
& + 1\cdot(1-1)/2 + 6\cdot(6-1)/2 + 14\cdot(14-1)/2 + 11\cdot(11-1)/2 + 6\cdot(6-1)/2 \\
= {} & 442143
\end{aligned}
$$

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

Eigene wirtsch. Lage (Y)	Allg. wirtschaftliche Lage (X) (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Bei gleicher Anordnung beider Variablen:

Konkordante Paare:	Fälle in Zellen rechts unterhalb
Diskordante Paare:	Fälle in Zellen links unterhalb
X-verbundene P.:	Fälle in Zellen gleicher Spalte
Y-verbundene P.:	Fälle in Zellen gleicher Zeile
X,Y-verbundene P.:	Fälle in gleicher Zelle

Anzahl Paarvergleiche insgesamt:

Anzahl konkordanter Paare	C	=	880472
Anzahl diskordanter Paare	D	=	249538
Anzahl x-verbundener Paare	T_X	=	506992
Anzahl y-verbundener Paare	T_Y	=	669195
Anzahl x,y-verbundener Paare	T_{XY}	=	442143
	Summe	=	2748340

$= n \cdot (n-1)/2 = 2345 \cdot 2344/2 = 2748340$

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

Eigene wirtsch. Lage (Y)	Allg. wirtschaftliche Lage (X) (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Bei gleicher Anordnung beider Variablen:

Konkordante Paare:	Fälle in Zellen rechts unterhalb
Diskordante Paare:	Fälle in Zellen links unterhalb
X-verbundene P.:	Fälle in Zellen gleicher Spalte
Y-verbundene P.:	Fälle in Zellen gleicher Zeile
X,Y-verbundene P.:	Fälle in gleicher Zelle

Berechnungsformeln:

$$C=\sum_{i=1}^{I-1}\sum_{j=1}^{J-1} n_{i,j}\cdot\left(\sum_{k=i+1}^{I}\sum_{m=j+1}^{J} n_{k,m}\right);\ D=\sum_{i=1}^{I-1}\sum_{j=2}^{J} n_{i,j}\cdot\left(\sum_{k=i+1}^{I}\sum_{m=1}^{j-1} n_{k,m}\right)$$

$$T_X=\sum_{i=1}^{I}\sum_{j=1}^{J-1} n_{i,j}\cdot\left(\sum_{k=j+1}^{J} n_{i,k}\right);\ T_Y=\sum_{j=1}^{J}\sum_{i=1}^{I-1} n_{i,j}\cdot\left(\sum_{k=i+1}^{I} n_{k,j}\right);\ T_{X,Y}=\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{i,j}\cdot\left(n_{i,j}-1\right)}{2}$$

Aus all diesen Paarvergleichen lassen sich Zusammenhangsmaße ableiten:

Idee:

Spricht ein Paarvergleich eher für einen positiven Zusammenhang, eher für einen negativen Zusammenhang oder gegen einen monotonen Zusammenhang?

Meine Notizen:

Ansatz 2: Paarvergleiche am Beispiel

- Aus all diesen Paarvergleichen lassen sich Zusammenhangsmaße ableiten:

  Idee:

  Spricht ein Paarvergleich eher für einen positiven Zusammenhang, eher für einen negativen Zusammenhang oder gegen einen monotonen Zusammenhang?

- Bei positive Beziehung: viele konkordante Paare!

  Bei negativer Beziehung: viele diskordante Paare!

  $C - D > 0$ → eher positive Beziehung; $C - D < 0$ → eher negative Beziehung;

  $C = D$ → keine (monotone) Beziehung

- Welche Rolle spielen verbundene Paare?

  verbundene Paare sprechen gegen gerichtete Beziehung von Y auf X:

Meine Notizen:

Paarvergleiche am Beispiel: Sommers d

Eigene wirtsch. Lage	Allgem. wirtschaftliche Lage (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Anzahl konkordanter Paare	C	=	880472
Anzahl diskordanter Paare	D	=	249538
Anzahl x-verbundener Paare	T_X	=	506992
Anzahl y-verbundener Paare	T_Y	=	669195
Anzahl x,y-verbundener Paare	T_{XY}	=	442143
	Summe	=	2748340

Asymmetrische Beziehung von X auf Y: Somers d_{YX}

- C spricht für positive, D für negative und T_Y gegen Beziehung von X auf Y

$$\rightarrow d_{YX} = \frac{C-D}{C+D+T_Y} \qquad d_{YX} = \frac{C-D}{C+D+T_Y} = \frac{880472-249538}{880472+249538+669195} = 0.351$$

Asymmetrische Beziehung von Y auf X: Somers d_{XY}

- C spricht für positive, D für negative und T_X gegen Beziehung von Y auf X

$$\rightarrow d_{XY} = \frac{C-D}{C+D+T_X} \qquad d_{XY} = \frac{C-D}{C+D+T_X} = \frac{880472-249538}{880472+249538+506992} = 0.385$$

Meine Notizen:

Asymmetrische Zusammenhänge: Sommers d

- Sommers d ist ein Maß zur Messung asymmetrischer Zusammenhänge zwischen ordinalen Variablen.
- Es teilt die Differenz von konkordanten und diskordanten (positiven und negativen) Paaren durch die Summe gerichteter Paare im Sinne der Asymmetrie:

$$\rightarrow d_{YX} = \frac{C - D}{C + D + T_Y} \qquad \rightarrow d_{XY} = \frac{C - D}{C + D + T_X}$$

- Sommers d kann als Verallgemeinerung der Prozentsatzdifferenz verstanden und entsprechend interpretiert werden.

Meine Notizen:

Symmetrische Zusammenhänge: Tau-b

- Das symmetrische Zusammenhangsmaß für Dichotome Variablen Phi kann betrachtet werden als geometrisches Mittel der Prozentsatzdifferenzen, also als Wurzel des Produkts.
- Auf die gleiche Weise kann auch ein symmetrisches Zusammenhangsmaß für ordinale Daten entstehen, das Tau-b genannt wird:

$$\rightarrow \left|\tau_b\right| = \sqrt{d_{YX} \cdot d_{XY}} = \sqrt{\frac{C-D}{C+D+T_X} \cdot \frac{C-D}{C+D+T_Y}}; \quad \tau_b = \frac{C-D}{\sqrt{\left(C+D+T_X\right) \cdot \left(C+D+T_Y\right)}}$$

Meine Notizen:

Paarvergleiche am Beispiel: Tau-b

Eigene wirtsch. Lage	Allgem. wirtschaftliche Lage (1)	(2)	(3)	(4)	(5)
Sehr gut (1)	48	74	32	5	1
Gut (2)	73	645	526	72	6
Teils,teils (3)	12	143	361	115	16
Schlecht (4)	5	30	68	64	11
Sehr schlecht (5)	1	6	14	11	6

Anzahl konkordanter Paare	C	=	880472
Anzahl diskordanter Paare	D	=	249538
Anzahl x-verbundener Paare	T_X	=	506992
Anzahl y-verbundener Paare	T_Y	=	669195
Anzahl x,y-verbundener Paare	T_{XY}	=	442143
	Summe	=	2748340

$$\rightarrow \left|\tau_b\right| = \sqrt{d_{YX} \cdot d_{XY}} = \sqrt{\frac{C-D}{C+D+T_X} \cdot \frac{C-D}{C+D+T_Y}}; \quad \tau_b = \frac{C-D}{\sqrt{\left(C+D+T_X\right)\cdot\left(C+D+T_Y\right)}}$$

$$\tau_b = \frac{C-D}{\sqrt{\left(C+D+T_X\right)\cdot\left(C+D+T_y\right)}} = \frac{880472-249538}{\sqrt{\left(880472+249538+669195\right)\cdot\left(880472+249538+506992\right)}} = 0.368$$

Meine Notizen:

Symmetrische Zusammenhänge: Alternativen

- Bei empirisch möglichen vollständigen Rangreihen sprechen T_X, T_Y und $T_{X,Y}$ gegen eine monotone Beziehung. Das berücksichtigt Tau-a:

$$\rightarrow \tau_a = \frac{C-D}{C+D+T_Y+T_X+T_{X,Y}} = \frac{C-D}{n\cdot(n-1)/2} \qquad \tau_a = \frac{C-D}{n\cdot(n-1)/2} = \frac{880472-249538}{2748340} = 0.230$$

- Gamma vergleicht nur konkordante und diskordante Paare:

$$\rightarrow \gamma = \frac{C-D}{C+D} \qquad \gamma = \frac{C-D}{C+D} = \frac{880472-249538}{880472+249538} = 0.558$$

- Als Regeln gelten:

$$1 \geq |\gamma| \geq |\tau_b| \geq |\tau_a| \geq 0 \text{ und } |\tau_b| = \sqrt{d_{YX}\cdot d_{XY}}$$

Meine Notizen:

Bivariate Zusammenhänge nach Skalenniveaus

Symmetrisch

	Dichotome	Nominalskala	Ordinalskala	Metrische Skala
Metrische Skala			τ_a, τ_b, γ, r_s	Korrelation
Ordinalskala	τ_a, τ_b, γ, r_s		τ_a, τ_b, γ, r_s	
Nominalskala	Cramers V	Cramers V		
Dichotome V.	Phi, α			

Asym-metrisch

AV \ UV	Dichotome	Nominalskala	Ordinalskala	Metrische Skala
Metrische Skala	Mittelwert-vergleich, Regression		Sommers d	Regression
Ordinalskala	Sommers d, Tests		Sommers d	Sommers d
Nominalskala	Odds, λ, R´	Odds, λ, R´		
Dichotome V.	Prozentsatz-differenz, α	Odds, λ, R´	Sommers d	

- Asymmetrische Zusammenhänge gehen von einer Wirkung einer unabhängigen Variablen (UV) auf eine abhängige Variable (AV) aus.
- Symmetrische Zusammenhänge sind geeignet, wenn keine (einseitige) kausale Wirkung unterstellt werden soll oder Wirkung in beide Richtungen angenommen werden muss.

Meine Notizen:

Was Sie am Ende der Woche können sollten

- **Kern**: Sie analysieren Zusammenhänge ordinaler Variablen unter Berücksichtigung der Messprobleme dieser Zusammenhänge.
- Sie bewerten das Problem in der Messung ordinaler Zusammenhänge.
- Sie beschreiben drei Lösungsansätze für das Problem.
- Sie bestimmen die Rangplätze ordinaler Daten (in Kreuztabellen).
- Sie berechnen Spearmans Rangkorrelation.
- Sie erklären die Logik von Paarvergleichen und dabei insbesondere die 5 möglichen Ergebnisse eines Paarvergleichs.
- Sie berechnen die Summen von Paarvergleichen der 5 Typen in Kreuztabellen und nutzen diese zur Bestimmung und Interpretation von Sommers d, Tau-b, Tau-a und Gamma.

Meine Notizen: