

Frage:

Was ist der Unterschied zwischen zwei positiv korrelierten Merkmalen und 2 Merkmalen, die ein konkordantes Paar bilden?

Antwort:**Korrelation:**

Eine positive Korrelation bedeutet, dass für wachsende Werte des Merkmals X tendenziell auch das Merkmal Y grössere Werte aufweist. Dies ist eine globale Sicht auf den Zusammenhang der beiden Merkmale.

Konkordanz

Ein Beobachtungspaar $(x_1; y_1)$ und $(x_2; y_2)$ der Merkmale X und Y wird als konkordant bezeichnet wenn $x_1 < x_2$ und $y_1 < y_2$ ist. Beispielsweise ist das Paar, $(x_1; y_1) = (2, 3)$ und $(x_2; y_2) = (4, 4)$ konkordant, bei diskordanten Paaren besteht Beziehung: $x_1 > x_2$ und $y_1 > y_2$

Hierbei handelt es sich um **ein einzelnes** Beobachtungspaar der beiden Merkmale. Um eine Aussage über den globalen Zusammenhang der beiden Merkmale treffen zu können muss jedes Paar mit jedem verglichen werden und die Anzahl an Konkordanten (K) und Diskordanten (D) Paaren in z.B. in das Zusammenhangsmaß Gamma nach Goodman und Kruskal eingesetzt werden.

Frage:

Sind Konkordanz / Diskordanz sowie Reihen nur für ordinale Merkmale bildbar?

Antwort:

Nein, hierfür sind auch metrisch skalierte Merkmale geeignet, wichtig ist nur, dass die Merkmalsausprägungen der Merkmale X und Y sinnvoll der Größe nach geordnet werden können. Allerdings nutzt man dann nicht die gesamte Information, die in metrischen Merkmalen eigentlich enthalten ist.

Frage:

Was ist denn der Unterschied zwischen Paaren, die konkordant / diskordant sind und Paaren, die Bindungen aufweisen?

Antwort:

Konkordanz und Diskordanz, siehe obere Frage.

Paare mit Bindungen (T_x oder T_y) bedeutet, daß entweder $x_1 = x_2$ und / oder $y_1 = y_2$ ist. Hierbei lassen sich keine Aussagen über den Zusammenhang der einzelnen Paare treffen.

Beispielsweise kann man sich einen Datensatz vorstellen bei dem der Wert für x_i , $i = 1, \dots, n$ konstant ist, z.B. $x = 5$. Nur die Werte für y_i sind unterschiedlich. Besteht der gesamte Datensatz aus Paaren die Bindungen aufweisen ist keine Aussage über den Zusammenhang der beiden Merkmale möglich.

Frage:

Warum ist bei Übungsblatt 7 Aufgabe 2 das Gamma so groß?
Inwiefern werden bei der Berechnung Bindungen ignoriert bzw. wie beeinflusst das das Gamma?

Antwort:

Definition: Gamma nach Goodman und Kruskal

$$\gamma = \frac{K - D}{K + D} \quad (1)$$

Definition: Kendalls τ_b

$$\tau_b = \frac{K - D}{\sqrt{(K + D + T_x)(K + D + T_y)}} \quad (2)$$

Bei der Berechnung des Zusammenhangsmaßes γ werden die Bindungen ignoriert, T_x oder T_y tauchen in der Formel nicht auf. Zur Berechnung werden lediglich die Konkordanten und Diskordanten Paare herangezogen.

Im Beispiel herrscht eine hohe Anzahl an Bindungen in der Warenrücksendung (975) und / oder in der Pünktlichkeit (300), im Vergleich zur Anzahl der Konkordanten (400) und Diskordanten Paare (36) vor. Diese werden aber nicht in die Berechnung der Gesamtanzahl der Beobachtungspaare, im Nenner, einbezogen. Daher fällt das Ergebnis im Vergleich zu, z.B. Kendalls τ_b groß aus.

Bei Kendalls τ_b werden die Paare mit Bindungen im Nenner berücksichtigt, dadurch wird der Zusammenhang der beiden Merkmale als geringer eingeschätzt. Vergleiche dazu auch Blatt7/Aufgabe3 zum allgemeinen Zusammenhang zwischen γ und τ_b .

Frage:

Zur selben Aufgabe: Woher weiß ich beim Odds Ratio, welches der beiden Merkmale das Schichtungs- und welches das Risikomerkmale ist?

Antwort:

Vergleiche hierzu die Antworten zu:

- Frage Odds Ratio,
- Chance und Relatives Risiko,
- Blatt 7 Aufgabe 2 a.

Frage:

Ist bei Übung6 A1 die Arbeitslosigkeit das Risikomerkmal, das man die Risiken für KA bzw. LA berechnet?

Antwort:

X : Schichtungsmerkmal = Geschlecht

Y : Risikomerkmal = Arbeitslosigkeit

Andernfalls würde man ja sagen, das Risiko männlich (oder weiblich) zu sein hängt (wie auch immer) von der Arbeitslosigkeit ab. Das ist aber, falls überhaupt ein Zusammenhang besteht, nicht die richtige Wirkungsrichtung. Falls ein Zusammenhang besteht, dann hängt die Arbeitslosigkeit vom Geschlecht ab und nicht umgekehrt.

Frage:

Kendalls τ_b und τ_c bewegen sich beide zwischen -1 und 1, inwiefern ist τ_b nicht normiert?

Antwort:

τ_b gilt als Normiert, da der Nenner der Formel für τ_b sicher stellt dass das Ergebnis immer innerhalb des Intervalls $[-1; 1]$ liegt.

Frage:

Was macht τ_c zu einem besseren Zusammenhangsmaßzahl als τ_b ?

Antwort:

τ_b wird verwendet wenn Bindungen in den Daten vorliegen, jedoch nimmt dieses Maß nur im Falle einer quadratischen Kontingenztafel Werte im ganzen Intervall $[-1; 1]$ an.

τ_c gleicht diesen Nachteil durch Berücksichtigung der Dimensionen der Kontingenztafel aus.