Kontingenzanalyse und Chi²-Test

Besteht ein Zusammenhang zwischen zwei kategorialen Merkmalen? Unterscheiden sich die Verteilungen eines kategorialen Merkmals zwischen zwei oder mehr Gruppen voneinander? Dies sind typische Fragestellungen, die eine Kontingenzanalyse beantwortet. Sie kann auch metrische Merkmale berücksichtigen, wenn die Ausprägungen zu Klassen zusammengefasst werden.

Messung der Stärke des Zusammenhangs. Ein Beispiel: Es wurde die ungestützte Erinnerung an die Werbung für eine Marke erhoben. Untersucht werden soll, ob es einen Zusammenhang zwischen der Erinnerung und dem Alter gibt, das heißt ob sich der Anteil derjenigen, die sich an die Werbung erinnern, zwischen den Altersgruppen unterscheidet. Die Kontingenztabelle zeigt in schwarz das Ergebnis einer Befragung von 500 Personen. Beispielsweise haben fünf der 18- bis 29-Jährigen die Werbung erinnert, 75 nicht. Das Verhältnis von erinnerter Werbung zu nicht erinnerter Werbung beträgt in dieser Gruppe 1:15, in der Gruppe der 30- bis 49-Jährigen 1:3 und in der Gruppe der ≥ 50-Jährigen 1:5. Die Unterschiede deuten auf eine Abhängigkeit der beiden Merkmale hin.

Beobachtete / erwartete absolute Häufigkeiten

	18–29 Jahre	30–49 Jahre	≥ 50 Jahre	Zeilensumme
Erinnerung ja	5 / 16	75 / 60	20 / 24	100
Erinnerung nein	75 / 64	225 / 240	100 / 96	400
Spaltensumme	80	300	120	500

Vollkommene Unabhängigkeit läge vor, wenn die Verhältnisse in allen Altersgruppen gleich sind. Für die in blau dargestellten Häufigkeiten ist dies der Fall. Das Verhältnis ist immer 1:4. Diese bei Unabhängigkeit zu erwartenden Häufigkeiten lassen sich je Zelle gemäß Zeilensumme · Spaltensumme / Stichprobenumfang bestimmen, so dass Zeilen- und Spaltensummen unverändert bleiben. Je größer die Abweichungen der beobachteten (f_b) von den erwarteten Häufigkeiten (f_e) sind, desto stärker ist der Zusammenhang zwischen beiden Merkmalen. Um die Stärke des Zusammenhangs zu quantifizieren, wird je Zelle $(f_b - f_e)^2 / f_e$ berechnet. Die Summe dieser Werte über alle Zellen wird zumeist als χ^2 bezeichnet und ist in dem Beispiel gleich 14,974. Auf dessen Basis sind verschiedene Zusammenhangsmaße definiert. Mit n als Stichprobenumfang ist der

$$\text{Kontingenzkoeffizient} = \sqrt{\frac{\chi^2}{\chi^2+n}}$$

In dem Beispiel beträgt er 0,171. Generell ist sein Minimum gleich 0 und sein Maximum gleich $\sqrt{\frac{\min(k,l)-1}{\min(k,l)}}$ wenn k und l der Anzahl der Kategorien der beiden Merkmale entsprechen. Damit eine Interpretation vor dem Hintergrund des gängigen Wertebereichs von 0 (≙vollkommene Unabhängigkeit) bis 1 (≙totale Abhängigkeit) möglich ist, erfolgt eine Multiplikation mit dem Kehrwert des Maximums und es ergibt sich der

$$\text{Korrigierte Kontingenzkoeffizient} = \sqrt{\frac{2}{2-1}} \cdot 0{,}171 = 0{,}241$$

Explizit als Maß für die Effektstärke (für die Relevanz eines Effekts) gilt Cohens $w = \sqrt{\frac{\chi^2}{n}}$.

Nach Cohen (1988) ist ein Effekt ab 0,1 klein, ab 0,3 mittel und ab 0,5 groß. In diesem Beispiel ist Cohens w = 0,173 und damit der Effekt eher klein.

Chi²-Test. Ob ein Zusammenhang zwischen zwei Merkmalen oder Unterschiede in den Verteilungen nicht allein auf den Zufall zurückzuführen sind, wird mit dem Chi²-Test überprüft. Dessen Name ist angelehnt an die Teststatistik χ^2, die mit steigendem Stichprobenumfang näherungsweise Chi²-verteilt ist mit (k-1)·(l-1) Freiheitsgraden. Die Freiheitsgrade entsprechen der Anzahl Zellen, die frei variierbar sind, wenn die Zeilen- und Spaltensummen feststehen. Das heißt, in einer 2x3-Kontingenztabelle können zwei Werte beliebig vorgegeben werden. Dann sind die Werte der restlichen vier Zellen eindeutig bestimmt.

Der p-Wert – die Wahrscheinlichkeit, die Hypothese „Es besteht kein Zusammenhang (zwischen Erinnerung und Alter)" fälschlicherweise abzulehnen – lässt sich beispielsweise mit Hilfe der Excel-Funktion CHIQU.VERT.RE(χ^2; (k-1)·(l-1)) bestimmen. Im Beispiel ist p = 0,0006 = 0,06 % und damit deutlich kleiner als das übliche Signifikanzniveau von α = 5 %. Insofern wird die Hypothese abgelehnt; der Zusammenhang bzw. die Unterschiede sind signifikant.

Um herauszufinden, zwischen welchen Altersgruppen die Unterschiede signifikant sind, erfolgen analog zur Varianzanalyse paarweise Chi²-Tests mit Anpassung des Signifikanzniveaus aufgrund der alpha-Fehler-Kumulierung (siehe Lüken/Schimmelpfennig 2017).

Als Faustregel für die Anwendbarkeit des Chi²-Tests gilt, dass nicht mehr als 20 % der Zellen eine erwartete absolute Häufigkeit kleiner als fünf und keine Zelle eine kleinere als eins haben sollten. Je nach Größe der Kontingenztabelle kommen ansonsten verschiedene alternative Tests in Betracht (siehe Bortz et al. 2008).

In Ausgabe 2/2018: Entscheidungsbäume – Algorithmen im Überblick

Johannes Lüken, Diplom-Psychologe, ist Leiter des Bereichs Data Sciences bei IfaD.

jlueken@ifad.de

Prof. Dr. Heiko Schimmelpfennig, Diplom-Kaufmann, ist Projektleiter für Data Sciences bei IfaD.

hschimmelpfennig@ifad.de

Literatur

Bortz, J.; Lienert, G. A.; Boehnke, K.: *Analyse von Häufigkeiten.* In: Verteilungsfreie Methoden der Biostatistik, 3. Auflage, Berlin, Heidelberg, 2008, S. 87-196.

Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences,* 2. Auflage, Hillsdale, 1988.

Lüken, J.; Schimmelpfennig, H.: *Einfaktorielle Varianzanalyse* In: planung&analyse, Nr. 2/2017, S. 73.