Elisabeth Aufhauser, unveröffentlichter Text
Unterrichtsmaterial Statistik-UE für Soziologie

Chi-Quadrat-Test

Chi-Quadrat-Test auf Unabhängigkeit zwischen zwei kategorialen Merkmalen

Die Beziehung zwischen einer kategorialen Variable A mit I Kategorien und einer kategorialen Variable B mit J Kategorien lässt sich in Form einer sog. Kreuztabelle (einer I x J - dimensionalen Häufigkeitstabelle) darstellen. In SPSS können wir uns Kreuztabellen mit der Prozedur Analysieren > Deskriptive Statistiken > Kreuztabellen... erstellen. Die Zeilensummen in diesen Kreuztabellen entsprechen der einfachen Häufigkeits- oder Randverteilung der Variable A, die Spaltensummen entsprechen der einfachen Häufigkeits- oder Randverteilung der Variable B. In den Zellen der Tabelle stehen die Häufigkeiten, mit denen die verschiedenen Merkmalsausprägungen der beiden Variablen bei den Individuen (Fällen) in der Stichprobe gleichzeitig auftreten.

Wenn Individuen, die auf der Variable A eine bestimmte Ausprägung besitzen, auch auf der Variable B überproportional häufig eine ganz bestimmte Merkmalsausprägung (oder einige wenige bestimmte Merkmalsausprägungen) aufweisen, so sprechen wir von einer Assoziation zwischen den beiden Variablen.[1] Umgekehrt können wir von Unabhängigkeit zwischen den beiden Merkmalen sprechen, wenn die verschiedenen Kombinationen der Merkmalsausprägungen von A und B nicht überproportional häufig gemeinsam auftreten. Was bedeutet Überproportionalität im statistischen Sinn?

Wie Sie alle wissen, beträgt die Wahrscheinlichkeit einen Sechser zu Würfeln 1/6 und die Wahrscheinlichkeit zwei Sechser gleichzeitig zu würfeln 1/6 * 1/6=1/36. Wahrscheinlichkeitstheorisch formuliert: Die Wahrscheinlichkeit für das kombinierte Auftreten zweier unabhängiger Ereignisse ist gleich dem Produkt der Wahrscheinlichkeiten für das Auftreten der Einzelereignisse. Vor diesem Hintergrund kann die Unabhängigkeit kategorialer Variablen folgendermaßen definiert werden:

> Zwei kategoriale Variablen A und B sind unabhängig voneinander, wenn die Wahrscheinlichkeiten für das kombiniere Auftreten der Merkmalsausprägungen dem Produkt aus den (Rand)-Wahrscheinlichkeiten der zugehörigen Einzelkategorien entsprechen.

Angenommen es bestehe Unabhängigkeit zwischen den beiden Variablen A und B. Über die Formel "Zeilensumme mal Spaltensumme dividiert durch Stichprobengröße" können für jede Zelle in der Tabelle die unter dieser Annahme (Hypothese) erwarteten Werte geschätzt werden.[2] In SPSS werden derartige erwartete Werte unter der Überschrift "erwartete Häufigkeiten" in der Tabelle ausgedruckt, wenn Sie im Untermenü Zellen erwartete Häufigkeiten anklicken. Die Differenzen, die sich in der Kreuztabelle zwischen beobachteten und erwarteten Werten bilden lassen, werden als Residuen bezeichnet. SPSS druckt diese Residuen unter der Überschrift Residual aus, wenn Sie im Untermenü Zellen Residuen unstandardisiert anklicken.
Sie erinnern sich: Wir arbeiten mit Stichproben. Auch wenn in der Grundgesamtheit tatsächlich Unabhängigkeit zwischen zwei Variablen besteht, kann es rein zufallsbedingt passieren, dass in der Tabelle gewisse Abweichungen zwischen beobachteten und 'unter Unabhängigkeit' zu erwartenden Werten auftreten. Kleine Abweichungen (Residuen) sollten uns daher nicht irritieren. Ob die Abwei-

[1] Der statistische Zusammenhang zwischen den Werten zweier metrischer Variablen wird Korrelation genannt. Das Konzept der Assoziation stellt daher das kategoriale Pendant zum metrischen Konzept der Korrelation dar.
[2] Die Formel läßt sich leicht ableiten, wenn Sie P(AB)=P(A)*P(B) für die verschiedenen Kombinationen von A und B mit der Stichprobengröße multiplizieren.

chung zwischen beobachteten und erwarteten Werten als "klein" oder "groß" zu bezeichnen ist, hängt natürlich von der Höhe der absoluten Häufigkeiten in der entsprechenden Zelle der Tabelle ab. Eine Abweichung von 10 ist relativ "klein" wenn sie sich aus der Differenz zwischen einem beobachteten Wert von 200 und einem erwarteten Wert von 210 ergibt, aber relativ "groß", wenn sie sich aus dem Vergleich zwischen einem beobachteten Wert von 2 und einem erwarteten Wert von 12 ergibt. Um beurteilen zu können, ob eine Abweichung "klein" oder "groß" ist, werden sog. standardisierte Residuen berechnet: Die absoluten Residuen in jeder Zelle der Tabelle werden durch die Quadratwurzel der absoluten erwarteten Häufigkeiten dividiert. In SPSS werden die standardisierten Residuen unter der Überschrift "standardisierte Residuen" ausgedruckt, wenn Sie im Untermenü Zellen Residuen standardisiert anklicken.

Mit dem sog. **Pearson Chi-Quadrat-Wert** wird getestet, ob die in einer Stichprobe beobachtete Häufigkeitsverteilung über die Zellen einer Kreuztabelle mit der Hypothese der Unabhängigkeit von Zeilen- und Spaltenvariable in der Grundgesamt "verträglich" ist. Formal lässt sich dieser Test in Form von Null- und Alternativhypothese folgendermaßen formulieren:

Nullhypothese: Die beiden Variablen A und B sind unabhängig voneinander, d.h. die Wertausprägungen der beiden Variablen treten unabhängig voneinander auf.
Alternativhypothese: Die beiden Variablen sind - in welcher Form auch immer - miteinander assoziiert, d.h. bestimmte Wertausprägungen der beiden Variablen A und B treten über-proportional häufig gemeinsam (gleichzeitig) auf

Das, was statistisch tatsächlich getestet wird, lässt sich exakter auch so formulieren:

Nullhypothese: Die Wahrscheinlichkeit, dass die gemeinsame Häufigkeitsverteilung der beiden Variablen A und B in unserer Stichprobe aus einer Grundgesamtheit stammt, in der Unabhängigkeit zwischen den beiden Variablen besteht ist hoch, d.h. zumindest höher als das von uns gewählte Niveau für die Irrtumswahrscheinlichkeit α. [3]

Alternativhypothese: Die Wahrscheinlichkeit, dass die gemeinsame Häufigkeitsverteilung der beiden Variablen A und B in unserer Stichprobe aus einer Grundgesamtheit stammt, in der Unabhängigkeit zwischen den beiden Variablen besteht, ist sehr, klein, d.h. sie unterschreitet das von uns gewählte Niveau für die Irrtumswahrscheinlichkeit α.

Wenn die Nullhypothese statistischer Unabhängigkeit zwischen den beiden Variablen A und B in der Grundgesamtheit gilt, so lässt sich zeigen, dass die standardisierten Residuen in einer Zufallsstichprobe eine Standardnormalverteilung besitzen.

Sie erinnern sich: Eine Standardnormalverteilung hat den Mittelwert 0 und die Standardabweichung 1. Werte über 2 sind in einer derartigen Verteilung relativ selten. Wenn Sie einen ersten Überblick haben wollen, ob die Häufigkeitsverteilung in Ihrer Stichprobe auf Unabhängigkeit zwischen A und B hinweist, können Sie sich die standardisierten Residuen ansehen: Sind sehr viele der standardisierten Residuen größer 2 (oder kleiner -2), so werden Sie die Nullhypothese statistischer Unabhängigkeit zwischen den beiden Variablen ablehnen müssen.

Der exakte statistische Test geht folgendermaßen: Wenn Sie die standardisierten Residuen quadrieren und diese quadrierten Werte über alle Zellen der Tabelle hinweg aufsummieren, erhalten Sie die Pearson **Chi-Quadrat-Statistik**. Falls die Nullhypothese statistischer Unabhängigkeit der zwei

[3] Typische Werte für das Irrtumswahrscheinlichkeitsniveau in sozialwissenschaftlichen Analysen sind 5% oder 10%. Das "Gegenstück" zur Irrtumswahrscheinlichkeit wird als Signifikanzniveau bezeichnet. Die typischen Werte für das Signifikanzniveau sind dann 95% bzw. 90%.

Variablen A und B in der Grundgesamtheit zutrifft, besitzt die Pearson Chi-Quadrat-Statistik eine sog. Chi-Quadrat-Verteilung mit (I-1)*(J-1) Freiheitsgraden.[4] D.h.

Zahl der **Freiheitsgrade** = Zahl der Zeilen minus 1 * Zahl der Spalten minus 1. [5]

Die folgende Abbildung zeigt die Dichte- und Verteilungsfunktion einer Chi-Quadrat-Verteilung mit 3 Freiheitsgraden. Unter der Vorraussetzung, dass die Nullhypothese statistischer Unabhängigkeit zwischen den Variablen in der Grundgesamtheit zutrifft, liegt die Wahrscheinlichkeit, in der Stichprobe einen ChiQuadratwert von 6 zu erhalten, bei etwa 5% (vgl. Dichtefunktion). Die Wahrscheinlichkeit, dass der Chi-Quadratwert größer 6.25 ist, liegt insgesamt bei 10% (vgl. Verteilungsfunktion). D.h.

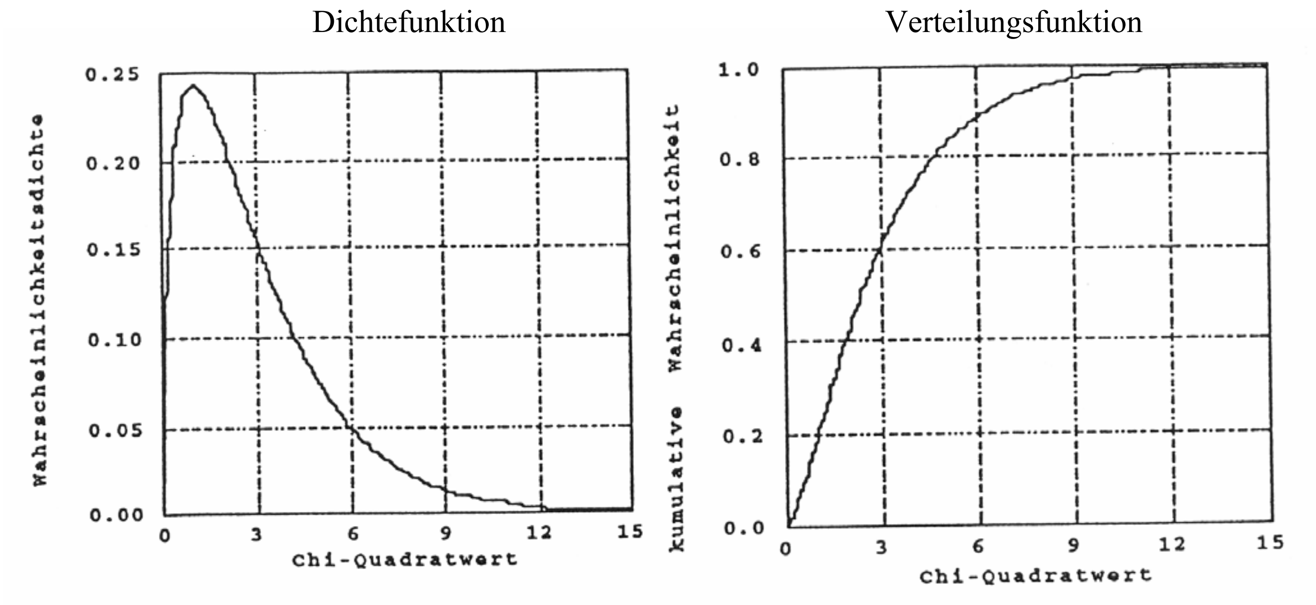

Abb.: Chi-Quadratverteilung mit 3 Freiheitsgraden

Wenn der von uns für die Stichprobe berechnete Chi-Quadratwert 6.25 überschreitet, liegt die Wahrscheinlichkeit, daß wir es mit einer Stichprobenverteilung zu tun haben, die aus einer Grundgesamtheit stammt, in der die beiden Variablen A und B unabhängig voneinander sind, bei etwa 10% - sie ist "relativ" klein. Auf einem Irrtumswahrscheinlichkeitsniveau von α =10% (bzw. dem entsprechenden Signifikanzniveau von 90%) können wir in einem derartigen Fall die Nullhypothese einer statistischen Unabhängigkeit zwischen A und B in der Grundgesamtheit ablehnen.[6] Umgekehrt gilt:

[4] Für alle die es genau wissen wollen: Die Summe der Quadrate von normalverteilten Zufallsvariablen besitzt eine Chi-Quadrat-Verteilung. Die Zahl der Freiheitsgrade einer Statistik ergibt sich generell aus der Zahl der unabhängigen Zufallsvariablen die in die Statistik eingehen, abzüglich der Zahl an unabhängigen Schätzwerten die zur Berechnung der Teststatistik verwendet werden. In unserem Fall stellen die I*J standardisierten Residuen derartige normalverteilte Zufallsvariablen dar. Bei der Berechnung (Schätzung) der erwarteten Werte verwenden wir darüberhinaus I+J Schätzwerte für die Randverteilung (die in der Stichprobe beobachteten Randverteilungswerte). Da die Stichprobengröße vorgegeben ist, sind nur I+J-1 dieser Randverteilungswerte frei zu bestimmen (unabhängig).
Die Zahl der Freiheitsgrade für eine I*J-Tabelle errechnet sich daher aus I*J (Zahl der normalverteilten Zufallsvariablen) minus (I+J-1) (Zahl der unabhängigen Schätzwerte, die in die Test-statistik eingehen). Anschaulich können Sie sich die Berechnung der Zahl an Freiheitsgraden auch so vorstellen: Wenn in einer zweidimensionalen Kreuztabelle die Randwerte vorgegeben sind, so sind genau (I-1)*(J-1) Werte frei wählbar, um die Tabelle vollständig auszufüllen.

[5] Beispielsweise hat eine 4x2 Kreuztabelle (4-1)*(2-1)=3*1=3 Freiheitsgrade.

[6] Der entsprechende kritische Wert für α =5% liegt bei 7.81 - d.h. die Ablehnung der Nullhypothese Unabhängigkeit wird umso 'schwieriger', je kleiner Sie die Irrtumswahrscheinlichkeit wählen. D.h. je niedriger Sie die Irrtumswahrscheinlichkeit ansetzen, desto 'sicherer' wollen Sie sein, daß Sie die Nullhypothese aufgrund der beobachteten Zusammenhänge in der Stichprobe tatsächlich ablehnen können.

Wenn der von uns für die Stichprobe berechnete Chi-Quadratwert unter 6.25 liegt, ist die Wahrscheinlichkeit, dass wie es mit einer Stichprobenverteilung zu tun haben, die aus einer Grundgesamtheit stammt, in der die beiden Variablen A und B unabhängig voneinander sind, größer als 10%. Auf dem Signifikanzniveau von 90% können wir in einem derartigen Fall die Nullhypothese einer statistischen Unabhängigkeit zwischen A und B in der Grundgesamtheit **nicht** ablehnen.

In der heutigen Forschungspraxis müssen wir die kritischen Werte bei verschiedenen Irrtumswahrscheinlichkeiten bzw. Freiheitsgraden nicht mehr in komplizierten statistischen Tabellen nachschlagen. Die Computerprogramme berechnen automatisch nicht nur die Chi-Quadratwerte für unsere Tabellen, sondern auch jene Wahrscheinlichkeit, mit der der berechnete Wert erreicht wird, wenn die Nullhypothese der Unabhängigkeit von A und B in der Grundgesamtheit zutrifft. In SPSS können Sie sich den Chi-Quadrat-Wert und das zugehörige Wahrscheinlichkeitsniveau ausgeben lassen, indem Sie im Untermenü Statistik Chi-Quadrat anklicken. Ausgedruckt werden Pearson-Chi-Quadrat-Wert, die Zahl an Freiheitsgraden (unter der Überschrift DF für Degrees of Freedom) sowie der zur Chi-Quadrat-Statistik gehörige Wahrscheinlichkeitswert (unter der Überschrift Signifikanz). Für Ihre zukünftige empirische Arbeit sollten Sie sich nach dieser umfangreichen Einführung vor allem folgendes merken:

Significance von Chi-Quadrat-Wert $> \alpha$: Nullhypothese der Unabhängigkeit kann aufgrund der beobachteten Verteilung in der Stichprobe nicht abgelehnt werden.

Significance von Chi-Quadrat-Wert $< \alpha$: Nullhypothese der Unabhängigkeit kann aufgrund der beobachteten Verteilung in der Stichprobe abgelehnt werden.

Zur Erinnerung: Als Nullhypothese formulieren wir immer das, von dem wir "sehr sicher" sein wollen, dass es nicht zutrifft, wenn wir es ablehnen. D.h. im konkreten Fall wollen wir nur dann von einer Assoziation zwischen zwei Variablen A und B sprechen, wenn es tatsächlich sehr unwahrscheinlich ist, dass sich in unserer Stichprobenverteilung rein "zufällig" Zusammenhänge zeigen, die es in der Grundgesamtheit gar nicht gibt.

In SPSS finden Sie im Untermenü Statistik eine ganze Reihe von Assoziationsmaßen. Während mit dem Chi-Quadrat-Wert nur getestet werden kann, ob Unabhängigkeit zwischen zwei Variablen aufgrund der Stichprobenverteilung wahrscheinlich ist oder nicht, können mit den Assoziationsmaßen Aussagen zu Stärke und Richtung von Zusammenhängen gemacht werden.

Die Berechnung derartiger Assoziationsmaße macht daher immer nur dann Sinn, wenn die Hypothese der Unabhängigkeit mit dem Chi-Quadrat-Test verworfen wird – d.h. der unter Signifikanz angeführte Wahrscheinlichkeitswert sehr klein (am besten Null) ist.

Falls "sehr viele" der erwarteten Häufigkeitswerte in der Tabelle "sehr klein" sind
(grobes Richtmaß: mehr als 20% der erwarteten Werte in der Tabelle liegen unter 5),
dann sind die unter Signifikanz ausgewiesenen Wahrscheinlichkeitswerte für den Chi-Quadrat-Wert eigentlich nicht gültig. SPSS druckt den Prozentsatz der erwarteten Werte unter 5 auch für Sie aus. Falls dieser Prozentsatz hoch ist, sollten Sie Sich auf alle Fälle überlegen, bestimmte "locker" besetzte Kategorien auf der Zeilen- und/oder Spaltenvariable zusammenzufassen.