

Vier-Feldertafeln, Test nach Fisher

Es geht darum, ob zwei verschiedene Gruppen, hier Typ A und Typ B genannt, bezüglich eines ja/nein-Merkmals E unterschiedlich sind. Der Einfachheit halber wird hier von Personen gesprochen.

Es werden n Personen untersucht, von denen a vom Typ A und b vom Typ B sind.

Nach der Befragung haben e Personen das Merkmal E und ne haben es nicht.

	E	nE	
A	ea	na	a
B	eb	nb	b
	e	ne	n

zum Beispiel

	E	nE	
A	3	2	5
B	1	6	7
	4	8	12

Interessant ist nun, dass die Aufteilung der e Personen auf Typ A und B durchaus verschieden sein kann, auch wenn die Gruppen bzgl. E gleich sind.

Null-Hypothese H_0 : Die Gruppen unterscheiden sich nicht bzgl. Merkmal E.
Forschungshypothese H_1 : Typ A weist weniger E auf, also Typ B mehr (oder umgekehrt).

Man beobachtet eine Aufteilung von e.

Ist nun $\frac{e}{n} \approx \frac{ea}{a} \approx \frac{eb}{b}$, so ist gar nichts mehr zu untersuchen, das gehört sicher zu H_0 . Weicht die Beobachtung deutlicher davon ab, so ist zu fragen, mit welcher Wahrscheinlichkeit unter H_0 diese Tafel und weitere, die H_1 noch mehr stützen, auftreten.

$$\alpha = P\left(\frac{ea}{eb} \text{ oder } \dots\right)$$

Es $\binom{a}{ea}$ gibt Möglichkeiten, ea Personen mit Merkmal E unter den a Personen vom Typ A "anzukeuzen".

Es $\binom{b}{eb}$ gibt Möglichkeiten, eb Personen mit Merkmal E unter den b Personen vom Typ B "anzukeuzen".

Es $\binom{n}{e}$ gibt Möglichkeiten, e Personen mit Merkmal E unter den n Personen überhaupt "anzukeuzen".

$$P\left(\frac{ea}{eb} \text{ oder } \dots\right) = \frac{\binom{a}{ea} \binom{b}{eb}}{\binom{n}{e}}$$

Der Zähler ist also die Zahl der Möglichkeiten, die die beobachtete Aufteilung von e in A und B aufweisen. Damit ist dies die Wahrscheinlichkeit für die beobachtete Tafel.

Im Beispiel ist dies 14%. Es kann sich keine Signifikanz mehr ergeben, auch wenn man die Wahrscheinlichkeit für die noch günstigere Tafel berechnet.

	E	nE	
A	4	15	19
B	0	7	7
	4	8	12

Dies ist die Tafel, ihre Wahrscheinlichkeit ist 1%. Damit ist klar, dass die Irrtumswahrscheinlichkeit 15% wäre, wenn man behaupten würde, bei Typ B käme E seltener vor. Man behauptet also gar nichts.
 Wenn man aber diese Tafel hier unten beobachtet hätte, hätte man H_1 annehmen können.

Weitere Überlegungen kann man in der zugehörigen MuPAD-Datei nachlesen.

Zusammenfassend ergibt sich: Man kann Zeilen und Spalten vertauschen, auch E und Nicht-E als Typen und A und B als alternative Merkmale ansehen, stets erhält man dieselben Wahrscheinlichkeits-Werte. Man muss nur darauf achten, dass man die Binomialterme entweder nur aus Zeilenelementen oder nur aus Spaltenelementen baut.

Aus didaktischer Sicht ist die oben verfolgte Begründung sinnvoller (als den Term aus Spalten zu bauen), da die Aufteilung in die Typen meist "vorgegeben" ist. Bücher sind da nicht immer geschickt.