Multivariate Statistik, Übung 12

Henry Haustein

Aufgabe 1

Wir benutzen die Identitäten $k_{ij} = f_{ij} \cdot k_{..}$, $k_{i.} = f_{i.} \cdot k_{..}$ und $k_{.j} = f_{.j} \cdot k_{..}$. Dann ergibt sich:

$$\begin{aligned}
\sum_{i=1}^{p}\sum_{i=1}^{q} \frac{\left(k_{ij} - \frac{k_{i.}k_{.j}}{k_{..}}\right)^2}{\frac{k_{i.}k_{.j}}{k_{..}}} &= \sum_{i=1}^{p}\sum_{i=1}^{q} \frac{\left(f_{ij} \cdot k_{..} - \frac{f_{i.} \cdot k_{..} \cdot f_{.j} \cdot k_{..}}{k_{..}}\right)^2}{\frac{f_{i.} \cdot k_{..} \cdot f_{.j} \cdot k_{..}}{k_{..}}} \\
&= \sum_{i=1}^{p}\sum_{i=1}^{q} \frac{(k_{..}\left[f_{ij} - f_{i.}f_{.j}\right])^2}{f_{i.}f_{.j} \cdot k_{..}} \\
&= \sum_{i=1}^{p}\sum_{i=1}^{q} \frac{k_{..}^2 (f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j} \cdot k_{..}} \\
&= \sum_{i=1}^{p}\sum_{i=1}^{q} \frac{k_{..}(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}} \\
&= k_{..} \sum_{i=1}^{p}\sum_{i=1}^{q} \frac{(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}}
\end{aligned}$$

Aufgabe 2

Der Term $\frac{k_{i.}k_{.j}}{k_{..}}$ drückt die erwartete Anzahl bei Unabhängigkeit aus. Die Teststatistik ist also eine Art relative Abweichung zur Unabhängigkeit.

Aufgabe 3

(a) Die (unvollständige) Kontingenztafel lautet

	Zulassung	Ablehnung	Σ
Soziologie	12	88	100
Maschinenbau	x	y	$x+y$
Sportwissenschaften	25	25	50
Σ	$37+x$	$113+y$	$150+x+y$

Wir wissen, dass $\frac{37+x}{150+x+y} = 0.484$ und $\frac{x}{x+y} = 7 \cdot \frac{12}{100}$ ist. Aus der zweiten Gleichung erhalten wir

$$\begin{aligned}
\frac{x}{x+y} &= \frac{84}{100} \\
x &= \frac{84}{100}x + \frac{84}{100}y \\
\frac{16}{100}x &= \frac{84}{100}y \\
y &= \frac{16}{84}x
\end{aligned}$$

Aus der ersten Gleichung erhalten wir:

$$\begin{aligned}
\frac{37+x}{150+x+y} &= 0.484 \\
37+x &= 72.6 + 0.484x + 0.484y \\
\frac{129}{250}x &= 35.6 + 0.484y \\
\frac{129}{250}x &= 35.6 + 0.484 \cdot \frac{16}{84}x \\
\frac{89}{210}x &= 35.6 \\
x &= 84 \\
y &= 16
\end{aligned}$$

(b) Die Tabelle der relativen Häufigkeiten ist dann

	Zulassung	Ablehnung	Σ
Soziologie	0.048	0.352	0.4
Maschinenbau	0.336	0.064	0.4
Sportwissenschaften	0.1	0.1	0.2
Σ	0.484	0.516	1

(c) Wir benutzen den χ^2-Unabhängigkeitstest, weil wir nur nominale Daten haben.
H_0 : Studiengang und Zulassung sind unabhängig
H_1 : Studiengang und Zulassung sind nicht unabhängig
Die Teststatistik ergibt sich zu

$$\begin{aligned}
T &= k_{..} \sum_{i=1}^{p} \sum_{i=1}^{q} \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}} \\
&= 103.83
\end{aligned}$$

Der kritische Wert ist $\chi^2_{(k-1)(l-1);1-\alpha} = \chi^2_{2\cdot 1;1-0.05} = 5.9915$. Damit wird H_0 abgelehnt und H_1 angenommen. Es besteht also ein Zusammenhang zwischen Studiengang und Zulassung.