Technische Hochschule Köln
Fakultät für Wirtschafts- und Rechtswissenschaften
Prof. Dr. Arrenberg
Raum 221, Tel. 39 14
jutta.arrenberg@th-koeln.de

Übungen zur Vorlesung QM III

Chi-Quadrat-Unabängigkeitstest

Aufgabe 16.1

Werden Frauen bei der Zulassung zu einem Studium bevorzugt? Um diese Frage zu beantworten, wurden an der University of California in Berkeley (USA) 1 518 Bewerber (w,m) auf die beiden Studiengänge A und B befragt. Es ergaben sich folgende Daten:

Studiengang A

	Bewerber		$\sum$
	nicht zugelassen	zugelassen	
weiblich	21	87	
männlich	313	512	
$\sum$			

Studiengang B

	Bewerber		$\sum$
	nicht zugelassen	zugelassen	
weiblich	8	17	
männlich	207	353	
$\sum$			

Prüfen Sie mit dem Chi-Quadrat-Unabhängigkeitstest die Hypothese, dass Zulassung (ja/nein) und Geschlecht (w/m) stochastisch unabhängig voneinander sind. Ziehen Sie zum Testen die Stichprobe

a) nur für den Studiengang A heran, also $n = 933$.

b) nur für den Studiengang B heran, also $n = 585$.

c) beide Studiengänge A,B heran, also $n = 1\,518$.

Lösung zu Aufgabe 16.1
H_0: „Zulassung und Geschlecht sind stochastisch unabhängig“ gegen H_1: „nicht H_0“

a) Studiengang A
Erwartete Häufigkeiten unter H_0:

Studiengang A

	Bewerber nicht zugelassen	Bewerber zugelassen	$\sum$
weiblich	$\frac{108 \cdot 334}{933} = 38{,}7$	$\frac{108 \cdot 599}{933} = 69{,}3$	108
männlich	$\frac{825 \cdot 334}{933} = 295{,}3$	$\frac{825 \cdot 599}{933} = 529{,}7$	825
	334	599	933

D.h. die minimale erwartete Häufigkeit beträgt 38,7 und ist somit größer gleich eins. Ferner hat keine der vier Zellen eine erwartete Häufigkeit von höchstens fünf. Der Freiheitsgrad beträgt eins. Somit ist die Faustregel erfüllt und der p-Wert muss gemäß der Kontinuitätskorrektur nach Yates bestimmt werden:

$$\chi^2_{emp.} = \frac{(|21-38{,}7|-0{,}5)^2}{38{,}7} + \frac{(|87-69{,}3|-0{,}5)^2}{69{,}3} + \frac{(|313-295{,}3|-0{,}5)^2}{295{,}3} + \frac{(|512-529{,}7|-0{,}5)^2}{529{,}7} = 7{,}6\overline{4} + 4{,}2689755 + 1{,}0018286 + 0{,}5585048 = 13{,}47375$$

Der obere 5%-Punkt der Chi-Quadrat-Verteilung mit einem Freiheitsgrad beträgt gemäß der Formelsammlung 3,841. Da $13{,}47375 \geq 3{,}841$ gilt, wird H_0 abgelehnt; d.h. die Zulassung hängt stochastisch ab vom Geschlecht.

b) Studiengang B
Erwartete Häufigkeiten unter H_0:

Studiengang B

	Bewerber nicht zugelassen	Bewerber zugelassen	$\sum$
weiblich	$\frac{25 \cdot 215}{585} = 9{,}2$	$\frac{25 \cdot 370}{585} = 15{,}8$	25
männlich	$\frac{560 \cdot 215}{585} = 205{,}8$	$\frac{560 \cdot 370}{585} = 354{,}2$	560
	215	370	585

D.h. die minimale erwartete Häufigkeit beträgt 9,2 und ist somit größer gleich eins. Ferner hat keine der vier Zellen eine erwartete Häufigkeit von höchstens fünf. Der Freiheitsgrad beträgt eins. Somit ist die Faustregel erfüllt und der p-Wert muss gemäß der Kontinuitätskorrektur nach Yates bestimmt werden:

$$\chi^2_{emp.} = \frac{(|8-9{,}2|-0{,}5)^2}{9{,}2} + \frac{(|17-15{,}8|-0{,}5)^2}{15{,}8} + \frac{(|207-205{,}8|-0{,}5)^2}{205{,}8} + \frac{(|353-354{,}2|-0{,}5)^2}{354{,}2} = 0{,}53260870 + 0{,}031012658 + 0{,}002380952 + 0{,}0013883399 = 0{,}08803788$$

Der obere 5%-Punkt der Chi-Quadrat-Verteilung mit einem Freiheitsgrad beträgt gemäß der Formelsammlung 3,841. Da $0{,}08803788 < 3{,}841$ gilt, wird H_0 nicht abgelehnt; d.h. die Zulassung hängt nicht stochastisch vom Geschlecht ab.

c) Studiengänge A,B zusammen

Beobachtete Häufigkeiten:

Studiengänge A,B

	Bewerber		$\sum$
	nicht zugelassen	zugelassen	
weiblich	29	104	133
männlich	520	865	1 385
	549	969	1 518

Erwartete Häufigkeiten unter H_0:

Studiengänge A,B

	Bewerber		$\sum$
	nicht zugelassen	zugelassen	
weiblich	$\frac{133 \cdot 549}{1\,518} = 48{,}1$	$\frac{133 \cdot 969}{1\,518} = 84{,}9$	133
männlich	$\frac{1\,385 \cdot 549}{1\,518} = 500{,}9$	$\frac{1\,385 \cdot 969}{1\,518} = 884{,}1$	1 385
	549	969	1 518

D.h. die minimale erwartete Häufigkeit beträgt 48,1 und ist somit größer gleich eins. Ferner hat keine der vier Zellen eine erwartete Häufigkeit von höchstens fünf. Der Freiheitsgrad beträgt eins. Somit ist die Faustregel erfüllt und der p-Wert muss gemäß der Kontinuitätskorrektur nach Yates bestimmt werden:

$$\chi^2_{emp.} = \frac{(|29-48{,}1|-0{,}5)^2}{48{,}1} + \frac{(|104-84{,}9|-0{,}5)^2}{84{,}9} + \frac{(|520-500{,}9|-0{,}5)^2}{500{,}9} + \frac{(|865-884{,}1|-0{,}5)^2}{884{,}1} = 7{,}1925156 + 4{,}0749117 + 0{,}6906768 + 0{,}3913132 = 12{,}34942$$

Der obere 5%-Punkt der Chi-Quadrat-Verteilung mit einem Freiheitsgrad beträgt gemäß der Formelsammlung 3,841. Da $12{,}34942 \geq 3{,}841$ gilt, wird H_0 abgelehnt; d.h. die Zulassung hängt stochastisch vom Geschlecht ab.

Die Frage, welche der beiden Testentscheidungen in der Realität zutrifft, hängt davon ab, welche Stichproben als repräsentativ angesehen werden können. Da wir nichts über das Zustande kommen der Stichproben wissen, können wir auch nicht beurteilen, ob eine der drei Stichproben als repräsentativ angesehen werden kann.

Um zu entscheiden, welches Geschlecht aufgrund der Stichproben unter a) und unter c) bevorzugt wird, könnte das Assoziationsmaß γ (kennen wir jedoch nicht!) herangezogen werden:

$$\gamma = \frac{21 \cdot 512 - 87 \cdot 313}{21 \cdot 512 + 87 \cdot 313} = -0{,}434$$

; d.h. in der Stichprobe unter a) gibt es eine schwache Tendenz dafür, dass Frauen bevorzugt zugelassen werden.

$$\gamma = \frac{29 \cdot 865 - 104 \cdot 520}{29 \cdot 865 + 104 \cdot 520} = -0{,}366$$

; d.h. in der Stichprobe unter c) gibt es eine schwache Tendenz dafür, dass Frauen bevorzugt zugelassen werden.