5.7 Chi-Quadrat Tests für diskrete Daten

Gegeben seien Zähldaten wie $X_i =$ Anzahl Fälle in der Klasse i. Wie testen wir, ob diese Daten mit einem Modell konsistent sein?

5.7.1 Beispiele

(1) Ist die Anzahl der Tore in einem Fußballspiel eine poissonverteilte ZV mit $\lambda = 3$? Daten aus 18 Spielen aus der 1. und 2. Bundesligen stehen zur Verfügung:

$$7, 9, 2, 1, 7, 3, 5, 3, 6 \quad \text{und} \quad 2, 3, 5, 3, 5, 2, 3, 0, 5$$

(2) Gibt es eine geschlechtsspezifische Abneigung gegen das Studium an der Mathematisch-Naturwissenschaftlichen Fakultät der Uni Augsburg? Im Sommersemester 1999 waren die Zahlen

	Frauen	Männer
MNF	433	693
Rest	5505	4641

Methode

Wir berechnen die unter dem Modell zu erwartenden Werten und vergleichen sie mit den beobachteten Werten.

Beispiel (1) Anzahl Tore

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Sei e_i die erwartete Anzahl Spiele mit i Toren unter den 18 Spielen und o_i die beobachtete Zahl:

Tore	0	1	2	3	4	5	6	≥ 7
o_i	1	1	3	5	0	4	1	3
e_i	0.89	2.69	4.03	4.03	3.02	1.81	0.91	0.60

Heuristische Idee. Die (quadrierten) Abweichungen

$$(o_i - e_i)^2$$

messen den Unterschied zum Modell für eine Klasse, aber sie sind nicht gut untereinander vergleichbar. Wir wollen nicht die (o_i, e_i) Paare

$$(100, 110) \quad \text{und} \quad (10, 20)$$

gleich bewerten. Deshalb arbeiten wir mit den relativen Abweichungen

$$\frac{(o_i - e_i)^2}{e_i}$$

Wir summieren sie über alle Klassen, und verwenden die Teststatistik

$$T = \sum_i \frac{(o_i - e_i)^2}{e_i}$$

Für H_0 stellt sich heraus, dass unter gewissen Annahmen

$$T \sim$$

wo die Anzahl der Klassen (”Anzahl Freiheitsgrade”) ist. Wir werden H_0 verwerfen, wenn T groß ist.

$$P(T > 15.5) = 0.05 \quad P(T > 20.1) = 0.01$$

Für die Daten finden wir $T = 16.79$, so dass wir H_0 zu einem Testniveau von 5% verwerfen können, aber nicht zu einem Niveau von 1%.

Der Durchschnitt der Tore pro Spiel war $71/18 = 3.944$. Wenn wir einen Test mit diesem λ ausführen, finden wir $T = 7.03$, aber wir müssen die Benutzung des Schätzers berücksichtigen. Da ein Parameter geschätzt worden ist, verlieren wir und nehmen $T \sim$

$$P(T > 14.1) = 0.05$$

H_0 wird in diesem Fall akzeptiert.

Beispiel (2) Fakultäten

H_0: kein Geschlechtseinfluß — der Anteil in MNF entspricht dem Gesamtanteil der Frauen an Studierenden.

Insgesamt gibt es 5938 Frauen und 5334 Männer mit einem Frauenanteil von $\frac{5938}{11272}$. Unter H_0 erwarten wir

$$\frac{5938}{11272} * 1126$$

Frauen in MNF. Die erwarteten Werte für die Tabelle sind

	Frauen	Männer
MNF	593.17	532.83
Rest	5344.83	4801.17

Aus demselben heuristischen Argument wie vorher nehmen wir als Teststatistik

$$X^2 = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Approximativ gilt

$$X^2 \sim \chi^2_{(r-1)(c-1)}$$

wo r die Anzahl von Zeilen (”rows”) sei und c die Anzahl von Spalten (”columns”). $r = 1, c = 1 \Rightarrow X^2 \sim \chi^2_1$

$$X^2 = 101.5 \Rightarrow \quad H_0 \text{ verwerfen.}$$

5.7.2 (Zweidimensionale) Kontingenztafeln im allgemeinen

Gegeben seine zwei Klassifikationen der Daten, $i = 1, \ldots, r$ und $j = 1, \ldots, c$. Diese Information kann in einer $r \times c$ Tabelle mit Einträgen

$$n_{ij} = \text{Anzahl Fälle in } i \text{ und } j$$

zusammengefasst werden. Wir interessieren uns meistens für die Hypothese

H_0: keine Assoziation zwischen den Klassifikationen.

Sei

$$p_{ij} = P(\text{Zeilenklasse} = i, \text{Spaltenklasse} = j)$$

$$p_{i.} = P(\text{Zeilenklasse} = i)$$

$$p_{.j} = P(\text{Spaltenklasse} = j)$$

$$\text{dann } H_0 \Rightarrow p_{ij} = p_{i.} p_{.j}$$

Wir schätzen $p_{i.}, p_{.j}$ mit

$$\hat{p}_{i.} = \frac{n_{i.}}{n} \qquad \hat{p}_{.j} = \frac{n_{.j}}{n}$$

$$\Rightarrow e_{ij} = n \left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.j}}{n}\right)$$

$$= \frac{(\text{Zeilensumme } i)(\text{ Spaltensumme } j)}{n}$$

Als Teststatistik haben wir

$$X^2 = \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Für $n \to \infty$ gilt $X^2 \sim$

Als Faustregel (Cochran 1952!) hat man $e_{ij} \geq 5 \quad \forall ij$, aber das ist sicherlich zu streng.

Freiheitsgrade

es sind rc Zellen (Kombinationen) rc
n wird festgelegt -1
$p_{i.}$ werden geschätzt $-(r-1)$
$p_{.j}$ werden geschätzt $-(c-1)$
Es bleiben

$$Y = X^2 \sim \chi^2_k$$

$$f(y) = \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} y^{k/2-1} e^{-y/2} \quad x \geq 0$$

$$E[Y] = k$$

$$V[Y] = 2k$$

Für $k = 1, 2$ hat die Dichte ein Maximum bei $y = 0$.
Für $k \geq 3$ ist $f(0) = 0$ und das Maximum liegt bei $y = k - 2$.

5.7.3 Interpretationen

In 1949 hatten Kieser und Schäfer aus *Who's Who* eine Tabelle für 1436 Frauen zusammengestellt, die all mindestens einmal verheiratet worden waren.

Ausbildung	Einmal Verheiratet	Mehrmals
College	550	61
Kein College	681	144

H_0: keine Assoziation

$$X^2 = 16.0 \Rightarrow \quad H_0 \text{ verwerfen, da } \chi^2_{1,0.99} = 6.63$$

Heißt das, dass Frauen ohne Ausbildung heirateten öfters $(17\%$ v. $10\%)$?

Oder, dass Frauen die öfters heirateten eher nicht zum College gingen$(30\%$ v. $45\%)$?

Erstens haben wir nur die Hypothese von keiner Assoziation verworfen, ohne Kausalität zu besprechen. Zweitens wird es andere Faktoren geben.

Die Daten sind keine zufällige Stichprobe aus einer Grundgesamtheit, kann man trotzdem den Test durchführen und die Resultate verallgemeinern?

Ein komplexeres Beispiel — Intelligenz und Kleidung

Gilbey (Biometrika **8** 94) hat die Verteilung von 1725 Kindern besprochen, die nach Intelligenz

A: geistig mangelhaft
B: langsam und dumm
C. dumm
D: langsam aber intelligent
E: ziemlich intelligent
F: ausgesprochen fähig
G: sehr fähig

und nach Kleidung

sehr gut gekleidet
gut gekleidet
schlecht, aber passend
sehr schlecht

klassifiziert worden sind.

Was kann man daraus lesen?
Wie könnten die Daten graphisch dargestellt werden?
Könnte man Untergruppen testen oder Kategorien kombinieren?

Table 2. *Distribution of* 1725 *school children according to their standard of clothing and their intelligence: Kendall & Stuart* (1967, p. 558) *after Gilby.*

	Intelligence class						
Standard of clothing	A, B	C	D	E	F	G	Total
Very well clad	33	48	113	209	194	39	636
Well clad	41	100	202	255	138	15	751
Poor but passable	39	58	70	61	33	4	265
Very badly clad	17	13	22	10	10	1	73
Total	130	219	407	535	375	59	1725