Statistik I für Human- und Sozialwissenschaften

12. Übung

Lösungsvorschlag

Gruppenübung

G 40 Ein Friseur fragt sich, ob ein Zusammenhang zwischen Augenfarbe und Haarfarbe von Erwachsenen besteht. Er notiert sich dafür von 40 etwa gleichalten Kunden und Kundinnen seines Friseursalons deren Augen- und Haarfarbe. Es ergibt sich folgende Kontingenztafel für die absoluten Häufigkeiten:

Haarfarbe / Augenfarbe	blau	braun	grün
blond	16	5	2
dunkel	3	3	1
rot	1	2	7

Es wird angenommen, dass sich die Haarfarbe durch i.i.d. Zufallsvariablen $X_1, \ldots, X_n$ und die Augenfarbe durch i.i.d. Zufallsvariablen $Y_1, \ldots, Y_n$ beschreiben lassen.

(a) Formulieren Sie Nullhypothese und Alternativhypothese zur Untersuchung, ob Augenfarbe und Haarfarbe voneinander unabhängige Merkmale sind.

(b) Benennen Sie einen geeigneten Statistischen Test zur Überprüfung der aufgestellten Hypothesen.

(c) Führen Sie den in (b) genannten Statistischen Test auf dem Signifikanzniveau $\alpha = 0,01$ durch und interpretieren Sie das Ergebnis.

<u>*Nullhypothese:*</u> *H_0: Haarfarbe und Augenfarbe sind voneinander unabhängige Merkmale;*
Alternative H_1: Haarfarbe und Augenfarbe sind voneinander abhängige Merkmale

<u>*Geeignetes Testverfahren:*</u> *χ^2-Unabhängigkeitstest.*

<u>*Testgröße:*</u>

$$T(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \quad = \quad \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(n_{i,j} - \frac{n_{i,\cdot} n_{\cdot,j}}{n}\right)^2}{\frac{n_{i,\cdot} n_{\cdot,j}}{n}}$$

<u>*Kritischer Bereich:*</u> *Wir lehnen H_0 ab, falls gilt:*
$T(x_1, \ldots, x_n, y_1, \ldots, y_n) > \chi^2_{(k-1)(l-1);1-\alpha}$

Kontingenztafel für die absoluten Häufigkeiten

Haar / Augen	*blau*	*braun*	*grün*	Σ
blond	*16*	*5*	*2*	*23*
dunkel	*3*	*3*	*1*	*7*
rot	*1*	*2*	*7*	*10*
Σ	*20*	*10*	*10*	*40*

Für die erwarteten Häufigkeiten $\frac{n_{i,\cdot} n_{\cdot,j}}{n}$ ergibt sich:

Haar / Augen	*blau*	*braun*	*grün*
blond	*11.5*	*5.75*	*5.75*
dunkel	*3.5*	*1.75*	*1.75*
rot	*5*	*2.5*	*2.5*

Mit diesen beiden Tabellen berechnet sich die Testgröße wie folgt:

$$\begin{aligned} T(x_1,\ldots,x_n,y_1,\ldots,y_n) &= \frac{(16-11.5)^2}{11.5} + \frac{(5-5.75)^2}{5.75} + \frac{(2-5.75)^2}{5.75} \\ &+ \frac{(3-3.5)^2}{3.5} + \frac{(3-1.75)^2}{1.75} + \frac{(1-1.75)^2}{1.75} \\ &+ \frac{(1-5)^2}{5} + \frac{(2-2.5)^2}{2.5} + \frac{(7-2.5)^2}{2.5} \\ &\approx 16.99. \end{aligned}$$

Mit $k = l = 3$ und $\alpha = 0.1$ lesen wir aus der Tabelle ab:

$$\chi^2_{(k-1)(l-1);1-\alpha} = \chi^2_{2\cdot 2;1-0.01} = \chi^2_{4;0.99} = 13.28$$

Wegen $16.99 > 13.28$ lehnen wir H_0 ab, d.h. wir haben einen Zusammenhang zwischen Augenfarbe und Haarfarbe.

G 41 Gegeben sei folgende geordnete Messreihe: 0.1; 0.7; 0.7; 2.1; 2.6; 2.6; 2.6; 2.6; 2.7; 7.4 7.4; 7.4; 7.4; 7.4; 8.2; 8.6; 9.0; 9.0; 9.3; 9.7.

Testen Sie zu einem Signifikanzniveau $\alpha = 0.1$, ob die Daten als gleichverteilt auf dem Intervall $[0; 10]$ aufgefasst werden können. Eine solche Gleichverteilung besitzt die Verteilungsfunktion

$$F_0(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{10}x & 0 \le x \le 10 \\ 1 & 1 < x. \end{cases}$$

(a) Mithilfe des Kolmogorow Smirnow Test. (Das 0.9-Quantil der Kolmogorow-Smirnow-Verteilung $k^{20}_{0.9} = 0.26473$.)

(b) Mithilfe des χ^2-Anpassungstest unter Benutzung der Klasseneinteilung $[0; 2.5)$, $[2.5; 5)$, $[5; 7.5)$ und $[7.5; 10]$.

Nullhypothese H_0: Daten sind gleichverteilt. Alternativhypothese H_1: Daten sind nicht gleichverteilt.

(a) *Die Testgröße berechnet sich durch den größten Abstand zwischen der empirischen Verteilungsfunktion der Messdaten F_n und der vermuteten Verteilungsfunktion F_0.*

$$T(X_1,\ldots,X_n) = \max_{i=1,\ldots,n}\left(\max\left\{\left|F_0\left(x_{(i)}\right) - F_n\left(x_{(i)}\right)\right|; \left|F_0\left(x_{(i)}\right) - F_n\left(x_{(i-1)}\right)\right|\right\}\right).$$

$$F_{20}(x) := \begin{cases} 0 & \text{für } x < 0.1 \\ 0.05 & \text{für } 0.1 \le x < 0.7 \\ 0.15 & \text{für } 0.7 \le x < 2.1 \\ 0.2 & \text{für } 2.1 \le x < 2.6 \\ 0.4 & \text{für } 2.6 \le x < 2.7 \\ 0.45 & \text{für } 2.7 \le x < 7.4 \\ 0.7 & \text{für } 7.4 \le x < 8.2 \\ 0.75 & \text{für } 8.2 \le x < 8.5 \\ 0.8 & \text{für } 8.5 \le x < 9.0 \\ 0.9 & \text{für } 9.0 \le x < 9.3 \\ 0.95 & \text{für } 9.3 \le x < 9.7 \\ 1 & \text{für } 9.7 \le x. \end{cases}$$

$x_{(i)}$	$F_0\left(x_{(i)}\right)$	$F_n\left(x_{(i)}\right)$	$\left\|F_0\left(x_{(i)}\right)-F_n\left(x_{(i)}\right)\right\|$	$\left\|F_0\left(x_{(i)}\right)-F_n\left(x_{(i-1)}\right)\right\|$
0.1	0.01	0.05	0.04	0.01
0.7	0.07	0.15	0.08	0.02
2.1	0.21	0.2	0.01	0.06
2.6	0.26	0.4	0.14	0.06
2.7	0.27	0.45	0.18	0.13
7.4	0.74	0.7	0.04	0.29
8.2	0.82	0.75	0.07	0.12
8.5	0.85	0.8	0.05	0.1
9.0	0.9	0.9	0	0.1
9.3	0.93	0.95	0.02	0.03
9.7	0.97	1	0.03	0.02

Der Wert der Testgröße ergibt sich als das Maximum der letzten beiden Spalten der Tabelle. $T(x_1,\ldots,x_{20})=0.29$.
H_0 *wird abgelehnt, falls* $T(x_1,\ldots,x_n)\geq k^n_{1-\alpha}$ *ist. Wegen*

$$T(x_1,\ldots,x_{20})=0.29>0.26473=k^{20}_{0.9}=k^{20}_{1-0.1}$$

lehnen wir H_0 *ab.*

(b) *Testgröße:* $$T(X_1,\ldots,X_n)=\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i}$$

Ablehung von H_0*:* $$T(x_1,\ldots,x_n)>\chi^2_{k-1;1-\alpha}.$$

Es liegen 4 Klassen mit den Häufigkeiten $n_1=4$, $n_2=5$, $n_3=5$ *und* $n_4=6$ *vor. Die erwarteten relativen Anzahlen der Klassen berechnen sich durch* $p_i=P(Y\in \textit{Klasse } i)$, *wobei* Y *eine Zufallsvariable mit Verteilung* F_0 *sei. Daraus folgt:*

$$p_1=P(Y\in \textit{Klasse 1})=P(0\leq Y<2.5)=F_0(2.5)-F_0(0)=\frac{1}{10}2.5-\frac{1}{10}0=0.25.$$

Analog erhält man $p_1=p_2=p_3=p_4=0.25$. *Damit ergibt sich als Wert der Testgröße*

$$\begin{aligned}T(x_1,\ldots,x_{20})&=\frac{(4-20\cdot0.25)^2}{20\cdot0.25}+\frac{(5-20\cdot0.25)^2}{20\cdot0.25}\\&\quad+\frac{(5-20\cdot0.25)^2}{20\cdot0.25}+\frac{(6-20\cdot0.25)^2}{20\cdot0.25}=\frac{2}{5}=0.4\end{aligned}$$

Wegen

$$T(x_1,\ldots,x_{20})=0.4<6.25=\chi^2_{3;0.9}=\chi^2_{k-1;1-\alpha}$$

kann H_0 *nicht abgelehnt werden.*