Bsp: Test auf Normalverteilung

Klasseneinteilung

$A_1 =]-\infty, 0]$, $A_2 =]0, 5000]$, $A_3 \ldots$

$N_1 = 0$ $N_2 = 148$, ...

aus den ursprünglichen Daten wurden das empirische Mittel

$$\bar{\mu}_n = 6892$$

und die empirische Standardabweichung

$$\bar{\sigma}_n = 14984$$

geschätzt.

Berechnung der theoretischen Häufigkeiten unter

H_0: Daten sind normalverteilt (mit $\mu = 6892$, $\sigma_n = 14984$, $n = 197$)

$$f = 9 - 2 - 1 = 6$$

Theoretischen Häufigkeiten

$$P(X \in A_1 \mid H_0)$$

$$= P_{\mu,\sigma^2}(X \in]-\infty, 0])$$

$$= P_{\mu,\sigma^2}(X \leq 0) = P_{\mu,\sigma^2}\Big(\underbrace{\frac{X-\mu}{\sigma}}_{Y} \leq \frac{-\mu}{\sigma}\Big)$$

$$= P_{0,1}\left(Y \leq -\frac{6892}{14984}\right) = \Phi_{0,1}(-0{,}46)$$

$$\approx 0.3228$$

Tabelle der Standardnormalverteilung

theoretische Häufigkeit:

$$F_1 = n \cdot P(X \in A_1 \mid H_0)$$

$$= 197 \cdot 0.3228 = 63{,}59$$

$$F_2 = n \cdot P(X \in A_2 \mid H_0)$$

$$= n \cdot P_{\mu,\sigma^2}(0 \leq X \leq 5000)$$

$$= 197 \cdot \left(\Phi_{0,1}\left(\frac{5000-6892}{14984}\right) - \Phi_{0,1}\left(-\frac{6892}{14984}\right)\right)$$

$$\approx 25.02$$

Begründung für die Verwendung der χ^2-Verteilung.

Erinnerung: Falls $Y_1, \ldots, Y_n$ unabh. Standardnormalverteilt:

$$\Rightarrow \underbrace{Y_1^2 + Y_2^2 + \ldots + Y_n^2}_{} \text{ ist } \chi^2\text{-Vert. zum Par. } n$$

Beim χ^2-Test wurden n Messwerte in k Gruppen Klassen

Falls n groß ist, und die Klassengröße nicht zu klein, so sollte unter der Nullhyp.

N_i ungefähr binomialverteilt mit Parameter $\underline{n}$ und $\frac{F_i}{n}$ sein

(genauer ist der Vektor $N_1, \dots, N_k$ annähernd multinomialverteilt mit Par.

$$n, \frac{F_1}{n}, \frac{F_2}{n}, \dots, \frac{F_k}{n})$$

d.h. $E[N_i] = F_i$, $V(N_i) = F_i(1-\frac{F_i}{n}) \approx F_i$

d.h.
$$\underbrace{\frac{N_i - F_i}{\sqrt{F_i(1-\frac{F_i}{n})}}}_{Y_i} \approx N(0,1)$$
zentraler Grenzwertsatz (n groß)

also $\sum_{i=1}^{k} \frac{(N_i - F_i)^2}{F_i} \approx \chi^2$ mit Par. $k-1$

χ^2-Test auf Unabhängigkeit:

Zur Erinnerung: 2 Ereignisse A, B sind unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$

H_A: Häufigkeit für das Ereignis A

$$P(A) \approx \frac{H_A}{n}$$

n Messungen.

Unabhängigkeit: $\frac{H_{A\cap B}}{n} = \frac{H_A}{n} \cdot \frac{H_B}{n}$

$$\Longleftrightarrow \quad \underline{H_{A\cap B} = \frac{1}{n} \cdot H_A \cdot H_B}$$

d.h. $\boxed{F_{ij} = \frac{1}{n} \cdot \underline{F_i} \cdot \underline{F_j} \approx \frac{1}{n} \cdot N_{i*} \cdot N_{*j}}$

Bsp. Genveränderungen

$n = 100$, $\alpha = 0.05$

$f = (2-1)(2-1) = \underline{1}$

theoretische Häufigkeiten

$$F_{AB} = \frac{1}{100} \cdot N_{A*} \cdot N_{*B} = \frac{1}{100} \cdot 30 \cdot 40 = 12$$

Theoretische Häufigkeiten

	F_A	F_{A^c}	Gesamt
F_B	12	18	30
F_{B^c}	28	42	70
	40	60	

Testwert
$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(F_{ij} - N_{ij})^2}{F_{ij}}$$
$$= \frac{(F_A - N_A)^2}{F_A} + \frac{(N_B - F_B)^2}{F_B} + \frac{(N_{A^c} - F_{A^c})^2}{F_{A^c}} + \frac{(N_{B^c} - F_{B^c})^2}{F_{B^c}}$$
$$= \frac{(16-12)^2}{12} + \dots + \dots$$
$$= 3{,}17$$

Vergleichswert $\chi^2_{1,0,95} = 3{,}842 > \chi^2$

$\underline{H_0 \text{ wird nicht abgelehnt}}$