## 14.4. Chi-Quadrat-Anpassungstest

17.06.19

$H_0$: X hat die Verteilung F, die durch die Wahrscheinlichkeiten $p_1, p_2, \ldots, p_I$ vollständig festgelegt ist.

$H_1$: X hat nicht die Verteilung F aus der Nullhypothese.

Ablehnung von $H_0 \Leftrightarrow p\text{-Wert} \leq 0{,}05$

Ablehnung von $H_0 \Leftrightarrow$
$\chi^2_{emp.} \geq$ oberer 5%-Punkt der Chi-Quadrat-Verteilung mit $\boxed{df = I - 1}$

Faustregel: $n \cdot p_i \geq 5$ für $i = 1, 2, 3, \ldots I$

Bsp: Im Jahr 1881 fiel dem Mathematiker Simon Newcomb auf, dass bei der Veröffentlichung von Datenmaterial wie z.B.:

- Länge von Flüssen
- Hausnummern einer Straße
- Einwohner einer Stadt

die führende Ziffer Eins wesentlich häufiger vorkommt als die Ziffern 2, 3, 4, 5, 6, 7, 8, 9.

17.06.19

Rhein $\boxed{1}$ 200 km
Elbe $\boxed{1}$.094 km
Mosel $\boxed{5}$44 km
Havel $\boxed{3}$34 km

X = führende Ziffer

| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $p_i$ | 0,301 | 0,176 | 0,125 | 0,097 | 0,079 | 0,067 | 0,058 | 0,051 | 0,046 |

Benford-Verteilung $\sum = 1$

Finanzgerichte nutzen das Benford-Gesetz, um Steuerhinterziehungen aufzudecken.

Frage: Gehorchen Beschäftigungszahlen dem Benford-Gesetz?
Stichprobe Beschäftigungszahlen u.v. führende Ziffern:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\sum$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 233 | 102 | 96 | 62 | 54 | 38 | 43 | 30 | 36 | n = 694 |
| $n \cdot p_i$ | 208,9 | 122,1 | 86,8 | 67,3 | 54,8 | 46,5 | 40,3 | 35,4 | 31,9 | |

↳ erwartete Häufigkeiten unter $H_0$

Faustregel: $n \cdot p_i \geq 5$ erfüllt

$$\chi^2_{emp.} = \frac{(233-208,9)^2}{208,9} + \frac{(102-122,1)^2}{122,1} + \frac{(96-86,8)^2}{86,8} + \frac{(62-67,3)^2}{67,3} + \frac{(54-54,8)^2}{54,8} + \frac{(38-46,5)^2}{46,5} + \frac{(43-40,3)^2}{40,3} + \frac{(30-35,4)^2}{35,4} + \frac{(36-31,9)^2}{31,9}$$

$= 10,57896$

$\boxed{df = 8}$ 17.06.19

$10{,}57896 < 15{,}507$

= oberes 5% der Chi-Quadrat-Verteilung mit $df = 8$.

d.h. $H_0$ wird nicht abgelehnt, d.h. Beschäftigungszahlen gehorchen dem Benfordgesetz.

$H_0$: Die führende Ziffer gehorcht dem Benford-Gesetz mit

$$I = 9 \left\{ \begin{array}{ll} p_1 = 0{,}301 & \text{Wht. für Ziffer } 1 \\ p_2 = 0{,}176 & \text{Wht. für Ziffer } 2 \\ p_3 = 0{,}125 & \text{Wht. für Ziffer } 3 \\ p_4 = 0{,}097 & \text{Wht. für Ziffer } 4 \\ p_5 = 0{,}079 & \text{Wht. für Ziffer } 5 \\ p_6 = 0{,}067 & \text{Wht. für Ziffer } 6 \\ p_7 = 0{,}058 & \text{Wht. für Ziffer } 7 \\ p_8 = 0{,}051 & \text{Wht. für Ziffer } 8 \\ p_9 = 0{,}046 & \text{Wht. für Ziffer } 9 \end{array} \right.$$

$df = I - 1 = 9 - 1 = 8$

17.06.19

Bsp:

Werden in einem Studiengang etwa gleich viel Frauen und Männer zugelassen?

Chi-Quadrat-Anpassungstest zum Niveau 0,05 anhand der folgenden Stichprobe:

| Geschlecht | Frauen | Männer | $\Sigma$ |
|---|---|---|---|
| $n_i$ | 44 | 56 | $n = 100$ |
| $n \cdot p_i$ | 50 | 50 | |

$H_0$: X = Geschlecht

$p_1 = \cancel{\text{Anteil}}\, P(X = \text{Frau}) = 0{,}5$

$p_2 = P(X = \text{Mann}) = 0{,}5$

$H_1$: nicht $H_0$

Faustregel: $n \cdot p_i \geq 5$ ok.

$$\chi^2_{emp.} = \frac{(44-50)^2}{50} + \frac{(56-50)^2}{50} = 1{,}44$$

$$1{,}44 < 3{,}841 \quad (df = 1)$$

d.h. $H_0$ wird nicht abgelehnt; d.h. es werden etwa gleich viel Frauen und Männer im Studiengang zugelassen.

**Bsp:**

Werden [illegible] jedes Jahr gleichviel Jungen wie Mädchen geboren?

Test zum Niveau 0,05

Stichprobe : Geburten 2015

35.935 Mädchen

36.466 Jungen

| Geschlecht | Mädchen | Jungen | $\Sigma$ |
|---|---|---|---|
| $n_i$ | 35.935 | 36.466 | 72.401 = n |
| $n \cdot p_i$ | 36.200,5 | 36.200,5 | |

$H_0$: X = Geschlecht

$p_1 = p(X = \text{Mädchen}) = 0{,}5$

$p_2 = p(X = \text{Jungen}) = 0{,}5$

$H_1$: nicht $H_0$

Faustregel: $n \cdot p_i > 5$ ok.

$$\chi^2_{emp} = \frac{(35.935 - 36.200{,}5}{36.200{,}5} + \frac{(36.466 - 36.200{,}5}{36.200{,}5} =$$

$= 3{,}89 > 3{,}841$ = oberer 5%-Punkt mit $df = 1$

d.h. Ablehnung von $H_0$; d.h. es werden nicht gleich viel Jungen wie Mädchen geboren.