Lösungen Übungsblatt 12)

① a)+b)

	grün	gelb	rot	Σ
Länglich	6 / 3.6	6 / 10.5	$0.6\cdot 30 = 18$ / 15.9	30
blockig	6 / 8.4	29 / 24.5	$0.5 \cdot 70 = 35$ / 37.1	70
Σ	12	35	53	n=100

X: Form (Zeilen), Y: Farbe (Spalten)

erwartete Häufigkeit $E_{\text{blockig, grün}} = \frac{70 \cdot 12}{100} = 8.4$

c) Hypothesen bzgl. der Paprika-Population:

H_0: Form X und Farbe Y von Gemüsepaprika sind stochastisch unabhängig.
(Es gibt keinen Zusammenhang zwischen X, Y.)

H_1: Form X und Farbe Y sind stochastisch abhängig
(Es gibt einen Zusammenhang von Form u. Farbe.)

d) Die Verwendung einer stetigen χ^2-Verteilung als Näherung der eigentlich diskreten Prüfverteilung ist ausreichend genau genug (also nicht zu grob und fehlerhaft), wenn nicht mehr als $1/5 = 20\%$ aller berechneten erwarteten Häufigkeiten $E_{ij} < 5$ ausfallen.

Hier: $6 \times E_{ij}$ $\begin{cases} 5\times \; E_{ij} \geq 5 \\ 1\times \; E_{ij} < 5 \; (3.6) \end{cases}$ $1/6 \approx 16.7\% < 20\%$ $E_{ij} < 5$ $\Rightarrow$ Kriterium erfüllt

e) H_0-Akzeptanzbereich:

$$[0, \chi^2_{df=(2-1)\cdot(3-1), 95\%}] = [0, \chi^2_{df=2, 95\%}] = \underline{\underline{[0, 5.99]}}$$

Teststatistik:

$$\chi^2_{emp} = \sum_{j=1}^{3}\sum_{i=1}^{2} \frac{(H_{ij} - E_{ij})^2}{E_{ij}} = \frac{(6-3.6)^2}{3.6} + \frac{(6-10.5)^2}{10.5} + \frac{(18-15.9)^2}{15.9} + \frac{(6-8.4)^2}{8.4} + \frac{(29-24.5)^2}{24.5} + \frac{(35-37.1)^2}{37.1}$$

$$= \underline{\underline{5.437}}$$

f) Testentscheidung:

H_0 beibehalten, denn $\begin{cases} \chi^2_{emp} = 5.437 \in [0, 5.99] \\ \text{p-value} = 0.066 > \alpha = 0.05 \end{cases}$

Fazit: Bei $\alpha = 5\%$ reicht die in der SP vorhandene Abweichung von stoch. Unabhängigkeit nicht aus, um bei Gemüsepaprika generell die Glaubhaftigkeit der stoch. UA (H_0) von Form und Farbe zu verwerfen.

g) $5\% = \alpha = P\big(H_0$ (UA von Form X, Farbe Y) anhand analoger SP ablehnen $\mid$ Form X und Farbe Y sind bei Gemüsepaprika-Pop. tatsächlich stoch. ua.$\big)$.

$n = 100$, Rand: 30, 70; 12, 35, 53

---

③ χ^2-Anpassungstest

a) Hypothesen bzgl. der Bevölkerungs-Population:

H_0: Von 100% Bevölkerung haben 1% Mo, 2% Di, 3% Mi, 4% Do, 10% Fr, 30% Sa und 50% So als Lieblingstag.

$p_1 = 0.01,\ p_2 = 0.02,\ p_3 = 0.03,\ p_4 = 0.04,\ p_5 = 0.1,\ p_6 = 0.3,\ p_7 = 0.5$

H_1: Mindestens 1 Wochentag tritt so nicht als Lieblingstag auf.

b) Für den SP-Umfang $\boxed{n = 350}$ entsprechen $p_1, p_2, \dots p_7$ folgenden erwarteten (absoluten) Häufigkeiten $\boxed{E_i = n \cdot p_i}$

E_1	E_2	E_3	E_4	E_5	E_6	E_7
$350 \cdot 0.01$	$350 \cdot 0.02$	$350 \cdot 0.03$	$350 \cdot 0.04$	$350 \cdot 0.1$	$350 \cdot 0.3$	$350 \cdot 0.5$
$= 3.5 < 5$	$= 7 > 5$	$= 10.5 > 5$	$= 14 > 5$	$= 35 > 5$	$= 105 > 5$	$= 175 > 5$

Für $1/7 \approx 14.29\% < 20\%$ aller E_i ergibt der Wert $< 5 \Rightarrow$ Kriterium ✓

c) H_0-Akzeptanzbereich:

$$[0, \chi^2_{df=(7-1),\,95\%}] = [0, \chi^2_{df=6,\,95\%}] = [0, 12.59]$$

Teststatistik:

$$\chi^2_{emp} = \sum \frac{(H_i - E_i)^2}{E_i} = \frac{(1-3.5)^2}{3.5} + \frac{(2-7)^2}{7} + \frac{(4-10.5)^2}{10.5} + \frac{(6-14)^2}{14} + \frac{(33-35)^2}{35} + \frac{(120-105)^2}{105} + \frac{(184-175)^2}{175} = 16.672$$

d) Testentscheidung:

H_0 ablehnen, denn $\chi^2_{emp} = 16.672 \notin [0, 12.59]$ bzw. p-value $= 0.011 < \alpha = 0.05$

e) $0.011 = \text{p-value} = P\big(\chi^2_{emp} \geq 16.672$ bei analoger SP $\mid$ H_0-Hypothese bzgl. der Lieblingstage trifft in Bev. zu$\big)$

f) $H_0: p_1 = p_2 = \dots = p_7 = 1/7 \Rightarrow$ alle $E_i = n \cdot p_i = 350 \cdot 1/7 = 50 > 5$

$$\chi^2_{emp} = \frac{(1-50)^2}{50} + \dots + \frac{(184-50)^2}{50} = 638.04 \notin [0, 12.59] \Rightarrow H_0 \text{ ablehnen}$$

Lösungen Übungsblatt 13)

(1) Zsf: Die Altersgruppe beeinflusst die erfassten Entfernungen X

- Median steigt mit Ak an ($23 < 23.5 < 25 < 26.5$) $\Rightarrow$ je älter, desto größer X, wobei Ak 8 und Ak 10 recht ähnlich ausfallen)
- Streuung der X-Werte unabhängig von Ak recht ähnlich IQR ($\approx 2.4 - 3$), Datenspann ($\approx 7 - 8$)
- Asymmetrie der Entfernungswerte nimmt mit wachsender Ak an Stärke zu und ist überall linkssteil ausgeprägt

Zsf: In der betrachteten SP war die Handspanne nahezu unabhängig vom Geschlecht ausgeprägt

- gleiche Median bei ♀ und ♂ (jeweils 18 cm)
- gleiche IQR (2 cm) bei ♀ und ♂, aber $\text{Datenspanne}_{♀} = 7.5 > \text{Datenspanne}_{♂} = 5$

$\Rightarrow$ Frauen-Handspanne variiert stärker, als beim Mann

- Symmetrieverhalten der Verteilungen ist abhängig vom Geschlecht ausgeprägt

Frau: symmetrisch | Mann: linkssteil