3. Kalkül II: Statistik und arbeiten mit R (6 Punkte)

Lösung

Der Blumenmix Blütentraum enthält 12 Tulpenzwiebeln. Laut Packungsaufschrift sind im Schnitt x Zwiebeln von roten Tulpen und $y = 12 - x$ Zwiebeln von gelben Tulpen enthalten. Anneliese kauft 10 Packungen. Leider ist die Packungsaufschrift nicht lesbar, sodass sie nicht erkennen kann, was x ist. Sie pflanzt alle Zwiebeln ein und beobachtet bei allen Tulpen, welche Farbe sie haben. Nehmen Sie im Folgenden an, dass aus allen Zwiebeln eine Tulpe sprießt.

a) Geben Sie ein geeignetes statistisches Modell an.

Anneliese pflanzt 120 Zwiebeln ein und beobachtet die Farbe jeder gewachsenen Tulpe. Daher wählen wir

- $\Omega = \{0,1\}^{120}$, ✓ *wobei* 0 *für eine gelbe und* 1 *für eine rote Tulpe steht.* ✓ *Für* $\omega = (\omega_i)_{i\in\{1,\ldots,120\}}$ *kodiert* ω_i *die Farbe der* i*-ten Tulpe;*
- $\mathcal{F} = \mathcal{P}(\Omega)$; ✓
- $\mathcal{P} = (P_\theta)_{\theta\in\Theta}$ *mit* $P_\theta = \mathrm{Bernoulli}(\theta)^{\times 120}$ ✓ *und*

$$\Theta = \left\{\frac{k}{12} : k \in \{2,\ldots,11\}\right\}. \checkmark$$

Wir nehmen an, dass die Zwiebeln unabhängig voneinander mit Wahrscheinlichkeit θ *zu einer roten und mit Wahrscheinlichkeit* $1-\theta$ *zu einer gelben Tulpe gehören. Es ist anzunehmen, dass die durchschnittliche Anzahl der Zwiebeln von roten Tulpen eine natürliche Zahl zwischen* 2 *und* 11 *ist. Nur so können auch gelbe Tulpen in der Packung sein.*

b) Nach einer kurzen Unterhaltung mit einem Nachbarn vermutet Anneliese nun, dass in jeder Packung aus mehr als der Hälfte aller Zwiebeln rote Tulpen sprießen und möchte daher die Nullhypothese: „Der Anteil der Zwiebeln für rote Tulpen beträgt höchstens 1/2.“ auf einem Signifikanzniveau von 10% testen.

i) Wie lauten die Hypothesen Θ_0 und Θ_1?

Die Hypothesen lauten

$$H_0: \ \theta \in \Theta_0 = \left\{\frac{k}{12} : k \in \{2,\ldots,6\}\right\} \quad \textit{gegen}$$
$$H_1: \ \theta \in \Theta_1 = \left\{\frac{k}{12} : k \in \{7,\ldots,11\}\right\}. \checkmark$$

ii) Bestimmen Sie den Ablehnungsbereich K des Tests unter Verwendung folgender Ausgabe in R.

Der Befehl `qbinom(p, size, prob)` bestimmt die kleinste Zahl x, sodass

$$\sum_{i=0}^{x} \binom{\texttt{size}}{i} \texttt{prob}^i \cdot (1-\texttt{prob})^{\texttt{size}-i} \geq \texttt{p}.$$

An der Ausgabe lesen wir ab, dass 68 die größte Zahl ist, für die gilt:

$$\sum_{i=68}^{120} \binom{120}{i} \left(\frac{1}{2}\right)^i \cdot \left(\frac{1}{2}\right)^{120-i} \approx 0.0853 \leq 0.1 = \alpha.$$

Der Ablehnungsbereich lautet also $K = \{68 \checkmark, 69, ..., 120 \checkmark\}$. *Die Nullhypothese wird genau dann verworfen, wenn die Anzahl der Zwiebeln von roten Tulpen größer ist als 67.*

```
seq(0,1,.1)
##  [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
q = qbinom(seq(0,1,.1), size = 120, prob = 1/2)
print(q)
##  [1]   0  53  55  57  59  60  61  63  65  67 120
pbinom(q, size = 120, prob = 1/2, lower.tail = FALSE)
##  [1] 1.00000000 0.88239832 0.79428839 0.67586997 0.53634249 0.46365751
##  [7] 0.39216409 0.26149549 0.15765165 0.08532261 0.00000000
```

iii) Geben Sie im Sachzusammenhang eine Interpretation folgender Ausgabe:

```
binom.test(c(67,53), p = 1/2, alternative = "greater")
##
##  Exact binomial test
##
## data:  c(67, 53)
## number of successes = 67, number of trials = 120, p-value = 0.1176
## alternative hypothesis: true probability of success is
## greater than 0.5
## 95 percent confidence interval:
##  0.4791448 1.0000000
## sample estimates:
## probability of success
##              0.5583333
```

Bei dem vorliegenden Binomialtest gab es 120 Versuche, davon waren 67 Treffer. Es wurden also 67 roten und 53 gelbe Tulpen gezählt. ✓ *Die Alternative lautet:*

„Der Anteil der Zwiebeln für rote Tulpen beträgt mehr als 1/2." ✓

Das niedrigste Signifikanzniveau, zu dem die Nullhypothese bei dieser Beobachtung gerade noch verworfen werden kann, beträgt ca. 11% (`p-value`). Der Anteil der roten Tulpen in der Messreihe beträgt ca. 56% (`probability of success`), und mit einer Wahrscheinlichkeit von 95% ist die tatsächliche Trefferwahrscheinlichkeit größer als 48% (`95 percent confidence interval`). Anneliese kann also die Nullhypothese nicht zum Niveau $\alpha = 0.1$ *verwerfen.* ✓ ✓