Institut für Angewandte Mathematik
Sommersemester 2022
Andreas Eberle, Luis Aniello La Rocca

7. Übungsblatt „Einführung in die Statistik"

Da die Tutorien am Dies Academicus und an Christi Himmelfahrt ausfallen, kann dieses Übungsblatt bis Mittwoch 25.5. abgegeben werden.

---

1. (Qualitätskontrolle) Die Stiftung Warentest berichtet im Artikel 'Nicht ungetrübt', dass in 2 von 24 getesteten Apfelsäften das Schimmelpilztoxin Patulin oberhalb des EU-weiten Grenzwertes von $m_0 = 50\mu g/l$ nachgewiesen werden konnte.
Nehmen Sie an, dass der Gehalt von x (in $\mu g/l$) an Patulin in einer bestimmten Sorte Apfelsaft normalverteilt ist mit unbekanntem Mittelwert m und Varianz v. Vor Versand einer Lieferung testet der Produzent die Hypothese

$$H_0 : m \leq m_0 \quad \text{gegen} \quad H_1 : m > m_0$$

zum Irrtumsniveau $\alpha = 0,01$ mit einer Stichprobe der Größe $n = 5$. Die Ware wird nur ausgeliefert, wenn der Test H_0 nicht verwirft. Eine Verbraucherorganisation testet ebenfalls eine Stichprobe $n = 5$ zum Irrtumsniveau $\alpha = 0.01$ bezüglich

$$H_0 : m \geq m_1 \quad \text{gegen} \quad H_1 : m < m_1 \quad (m_1 = 40\mu g/l)$$

a) Beschreiben Sie inhaltlich die möglichen Fehlentscheidungen der beiden Tests. Skizzieren Sie die Gütefunktion der beiden Tests als Funktion von m für $v \in \{5, 10\}$.

b) Zu welchem Ergebnis kommen die beiden Tests, wenn der Stichprobenmittelwert $\bar{x} = 52,5$ und die Stichprobenvarianz $s^2 = 8,7$ beträgt.

c) Wie groß müsste n gewählt werden, damit sich die in b) erhaltene Antwort ändert?

2. (Zwei-Stichproben-t-Test; Aufgabe zählt doppelt) Seien $X_1, \ldots, X_{n_1} \sim \mathcal{N}(m_1, v)$ und $Y_1, \ldots, Y_{n_2} \sim \mathcal{N}(m_2, v)$ unter $\mathbb{P}_{m_1,m_2,v}$ unabhängige Stichproben von zwei Normalverteilungen, und sei $n = n_1 + n_2$. Wir gehen davon aus, dass die Parameter m_1, m_2 und v alle unbekannt sind. Zeigen Sie:

a) Der Maximum-Likelihood Schätzer für $\theta = (m_1, m_2, v)$ ist

$$(\bar{X}, \bar{Y}, \tilde{V}) = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} X_i \,,\, \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i \,,\, \frac{1}{n} \left(\sum_{i=1}^{n_1} \left(X_i - \bar{X}\right)^2 + \sum_{i=1}^{n_2} \left(Y_i - \bar{Y}\right)^2 \right) \right).$$

b) Die renormierte Stichprobenvarianz $V^* = \frac{n}{n-2}\tilde{V}$ ist ein erwartungstreuer Schätzer.

c) Jeder Likelihood-Quotienten-Test für $H_0 \colon m_1 = m_2$ vs. $H_1 \colon m_1 \neq m_2$ hat einen Verwerfungsbereich der Form $\{|T(X)| > c\}$ mit der *Zweistichproben-t-Statistik*

$$T(X) \;=\; \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,\frac{\bar{X} - \bar{Y}}{\sqrt{V^*}}\,.$$

d) Zeigen Sie, dass $\bar{X} - \bar{Y}$ und V^* unter $\mathbb{P}_{m,v}$ unabhängige Zufallsvariablen sind mit

$$\frac{\bar{X} - \bar{Y} - m_1 + m_2}{\sqrt{v \cdot (n_1^{-1} + n_2^{-1})}} \sim \mathcal{N}(0,1)\,, \qquad \text{und} \qquad (n-2)\frac{V^*}{v} \;\sim\; \chi^2(n-2)\,.$$

Folgern Sie, dass $T(X)$ unter der Nullhypothese $\mathrm{t}(n-2)$-verteilt ist.

e) In den Chiwondo Beds, Malawi, wurden 77 Backenzähne gefunden, welche zwei verschiedenen Arten zugeordnet werden konnten:

Hipparion africanum $\approx$ 4 Mio. Jahre, 39 Zähne
Hipparion libycum $\approx$ 2,5 Mio. Jahre, 38 Zähne

Es ist bekannt, dass das Klima vor 3,8 Millionen Jahren stark abkühlte und das Hipparion wurde vom Laubfresser zum Grasfresser. Es wird bei beiden Zahngruppen die mesiodistale Länge gemessen ($x_1^a, \ldots, x_{39}^a$ und $x_1^l, \ldots, x_{38}^l$). Es wird davon ausgegangen, dass diese in beiden Fällen normalverteilt ist mit gleicher (unbekannter) Varianz v. Untersuchen Sie, ob sich die Nullhypothese, dass die Zähne gleich sind, zum Niveau 0,01 verwerfen lässt. Welche Schlüsse lassen sich daraus ziehen?

	Anzahl	Mittelwert	Stichprobenvarianz
Hipparion africanum	39	25,9	2,2
Hipparion libycum	38	28,4	4,3

3. (Zwei Hypothesentests)

a) Es werden $n = 17$ Trauerschnäpper in Käfigen einer Beleuchtung mit blauem Licht ausgesetzt (Versuchsbedingung 1) und jeweils in mehreren Durchgängen ihre Flugrichtung ermittelt. Die Flugrichtung wird als Punkt auf einem Kreis dargestellt. Aus allen Punkten auf dem Kreis wird der Schwerpunktvektor ermittelt. Danach wird der gleiche Versuch mit grünem Licht (Bedingung 2) durchgeführt. Für jeden Vogel i = 1,...,17 bezeichnen wir mit x_i^1 die Länge des Schwerpunktvektors bei blauem Licht und mit x_i^2 die Länge des Schwerpunktvektors bei grünem Licht. Sei $x_i = x_i^2 - x_i^1$ die Differenz. Da Schwerpunktvektoren Mittelwerte vieler zufälliger Beobachtungen sind, können wir davon ausgehen, dass diese approximativ normalverteilt sind mit konstanter Varianz σ^2. Untersuchen Sie, ob sich die Hypothese, dass die Farbe des Lichtes keine Rolle für die Orientierungsgenauigkeit der Trauerschnäpper spielt, zum Niveau 0,05 verwerfen lässt. Der Stichprobenmittelwert der x_i beträgt $\bar{x} = 0.0518$ und die Stichprobenvarianz $s^2 = 0.0912$.

b) In einer Urne liegen s schwarze und $n - s$ rote Kugeln. Die Gesamtanzahl der Kugeln sei bekannt. Wir möchten die unbekannte Anzahl $\theta = s$ der schwarzen Kugeln schätzen und ziehen dazu ohne Zurücklegen N $(N < n)$ Kugeln. Bestimmen Sie einen gleichmäßig mächtigsten Test für die Hypothese $\Theta_0 = \{0, \ldots, \theta_0\}$ gegen die Alternative $\Theta_1 = \{\theta_1, \theta_1 + 1, \ldots, n\}$, mit $\theta_1 > \theta_0$.

4. (Dualität von Konfidenzbereichen und Hypothesentests)
Sei $(\Omega, \mathcal{A}, (P_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell, und $X : \Omega \to S$ die Beobachtung.

a) Zeigen Sie: Ist $x \mapsto I(x) \subseteq \Theta$ ein Konfidenzbereich für ϑ zum Niveau $1 - \alpha$ und $\vartheta_0 \in \Theta$, dann ist

$$C(\vartheta_0) \; = \; \{x \in S : \vartheta_0 \notin I(x)\}$$

der Verwerfungsbereich eines Tests zum Niveau α der Hypothese $\Theta_0 = \{\vartheta_0\}$.

b) Umgekehrt sei $C(\theta_0)$ für jedes $\vartheta_0 \in \Theta$ der Verwerfungsbereich eines Tests der Hypothese $\Theta_0 = \{\vartheta_0\}$ zum Niveau α. Konstruieren Sie einen Konfidenzbereich für ϑ zum Niveau $1 - \alpha$.

c) Illustrieren Sie die Aussagen anhand eines Beispiels.