Institut für Angewandte Mathematik
Sommersemester 2024
Andreas Eberle, Stefan Oberdörster

6. Übungsblatt „Einführung in die Statistik"

Abgabe bis Dienstag 28.5.

---

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 unbekannter 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. (Entropiemaximierung unter Nebenbedingungen) Sei S ein endlicher Zustandsraum, und $H(\mu) = -\sum_{\mu(x)>0} \mu(x) \log \mu(x)$ das Entropiefunktional auf dem Raum

$$\mathrm{WV}(S) \;=\; \left\{ \mu \in \mathbb{R}^S \;:\; \sum_{x \in S} \mu(x) = 1,\;\; \mu(x) \geq 0 \,\forall\, x \right\}$$

der Wahrscheinlichkeitsmaße auf S.

a) Zeigen Sie, dass $\mathrm{WV}(S)$ konvex und H strikt konkav ist.

b) Sei $\nu \in \mathrm{WV}(S)$. Ist das relative Entropiefunktional $\mu \mapsto H(\mu|\nu)$ konkav oder konvex?

c) Sei $U : S \to \mathbb{R}$ eine Funktion auf S. Zeigen Sie, dass die *Boltzmann-Verteilung*

$$\mu_\beta(x) \;=\; \mathcal{Z}(\beta)^{-1}\, e^{-\beta U(x)} \quad (x \in S) \qquad \text{mit} \qquad \mathcal{Z}(\beta) \;=\; \sum\nolimits_{x \in S} e^{-\beta U(x)}$$

für $\beta \in \mathbb{R}$ die Entropie unter allen Wahrscheinlichkeitsverteilungen μ auf S mit festem Erwartungswert $\sum_{x \in S} U(x)\mu(x) = \sum_{x \in S} U(x)\mu_\beta(x) =: m(\beta)$ maximiert.