Übungen zur Angewandten Stochastik

Blatt 9

---

1. Aufgabe Für zwei Objekte mit unbekannten Gewichten w_1 und w_2 werden die Gewichte der Einzelobjekte, die Summe und Differenz der Gewichte gemessen, jede der vier Messungen sei mit einem unabhängigen Fehler derselben Varianz behaftet. Es wurden die Messwerte 3.4, 2.9, 6.3, 0.7 beobachtet.
a) Stellen Sie ein lineares Modell auf.
b) Bestimmen Sie die kleinste-Quadrate-Schätzer für w_1, w_2, $w_1 - w_2$ sowie deren geschätzte Streuung.
c) Testen Sie die Hypothese $H_0 : w_1 = w_2$ gegen $H_1 : w_1 \neq w_2$ (unter der Annahme normalverteilter Messfehler).

2. Aufgabe Die Datei `Kuckuckseier.dat` enthält die Längen von 120 Kuckuckseiern und zu jeder Beobachtung jeweils die Wirtsspezies, in deren Nest das Ei gefunden wurde (aus O.H. Latter, The Egg of Cuculus Canorus ..., *Biometrika* **1**, 164–176 (1902)). Testen Sie anhand dieser Daten (und einer Normalverteilungsannahme) die Hypothese, dass die Länge nicht von der Wirtsspezies abhängt.

3. Aufgabe Folgende Tabelle zeigt Dauer des Studiums (in Semestern) und Einstiegsgehalt (in T€) der Absolventen eines Jahres am Fachbereich Mathematik und Informatik der (hypothetischen) Yule-Simpson-Paradoxie-Universität:

Semester	12	14	16	12	15	14	13	14	11	13	10	12	14	13	14	15
Gehalt	39.4	38.2	37.4	39.5	32.8	35.3	39.1	35.2	37.9	35.7	41	40.9	34.2	38.4	36.2	38.4
Semester	9	11	9	9	12	13	11	10	10	10	9	10	12	10		
Gehalt	33.7	35.9	36.1	34.2	29.9	31.9	33.3	36.2	33.8	32.9	33.3	35.1	34.2	35.3		

a) Schlägt sich (für diese Absolventen) ein längeres Studium in einem höheren Anfangsgehalt nieder? Bestimmen Sie die Regressiongerade für Studiendauer gegen Anfangsgehalt.
b) Ändert sich Ihr Befund, wenn Sie zusätzlich erfahren, dass die oberen beiden Zeilen der Tabelle sich auf die Absolventen des Fachs Informatik, die unteren beiden sich auf die Absolventen des Fachs Mathematik beziehen, und Sie dieselbe Regression jeweils innerhalb dieser beiden Gruppen durchführen?

4. Aufgabe („Problem des Post-Docs“) Sein $X_1, X_2, \dots$ u.a., $\mathcal{N}(\vartheta, 1)$-verteilt, $\vartheta \in \mathbb{R}$. Um die Hypothese $H_0 : \vartheta \leq 0$ gegen die Alternative $H_1 : \vartheta > 0$ zu gegebenem Irrtumsniveau $\alpha \in (0, 1)$ anhand der erstem m Beobachtungen $X_1, \dots, X_m$ zu testen, wird man naheliegenderweise

$$T_m := -\Phi^{-1}(1-\alpha) + \frac{1}{\sqrt{m}} \sum_{i=1}^{m} X_i$$

berechnen und H_0 genau dann verwerfen, wenn $T_m > 0$ (Φ^{-1} sei die Inverse der Verteilungsfunktion der Standardnormalverteilung).
Was kann man über das tatsächliche Niveau $\alpha_{n,k}$ des Tests mit Ablehnungsbereich $\{\max_{m=n,\dots,n+k} T_m > 0\}$ für $n, k \in \mathbb{N}$ sagen?
a) Zeigen Sie, dass $\sqrt{n}(\alpha_{n,1} - \alpha) \underset{n\to\infty}{\to} \varphi\big(\Phi^{-1}(1-\alpha)\big)/\sqrt{2\pi}$ (φ ist die Dichte der Standardnormalverteilung).
b) Bestimmen Sie $\alpha_{n,k} - \alpha_n$ numerisch durch eine Simulationsstudie (beispielsweise mit `R`) für $\alpha \in \{0.05, 0.01\}$, $n \in \{50, 100, 500\}$, $k \in \{1, 5, 20\}$.

(Hinweis: Das Problem (samt Lösung) findet sich in L. Mattner, One optional observation inflates α by $100/\sqrt{n}$ per cent, arXiv:0710.5154v2, dessen sprechender Titel Teil a) zusammenfasst.)