R. Höpfner: Statistik mit Rechnerübungen für Lehramtsstudierende, SoSe19

Übungsblatt 5

Abgabe per mail an `hoepfner@mathematik.uni-mainz.de` bis **MI 10.07.19**,
Besprechung am FR 12.07.19

Aufgabe 5.1 : Eine Untersuchung des Intelligenzquotienten in einer Gruppe fünfjähriger Kinder (per Konvention ist der 'Mittelwert der Messgrösse IQ über die Gesamtheit aller Menschen' auf 100 normiert) ergab die folgenden Werte:

103	124	124	104	96	92	124	99	92	116	99	81	117	100	89	125
127	112	48	139	118	107	106	129	117	123	118	84	117	101	141	124
110	98	109	120	127	103	118	117	115	119	117	92	101	119	144	119
127	113	127	103	128	86	112	115	117	99	110	139	117	96	111	118
126	126	89	102	134	93	115	99	99	122	106	124	100	114	121	119
108	110	127	118	107	123	102	110	114	118	101	121	114			

Man nehme an, dass diese Daten als iid und normalverteilt angesehen werden dürfen.

a) Man teste die Hypothese $\mathbf{H} : \mu = 100$ gegen die Alternative $\mathbf{K} : \mu \neq 100$ zum Niveau $\alpha = 0.1$ und interpretiere das Ergebnis.

b) Man teste die Hypothese $\mathbf{H} : \mu \leq 105$ gegen die Alternative $\mathbf{K} : \mu > 105$ zum Niveau $\alpha = 0.05$.

c) Ihr Flurnachbar studiert Psychologie und meint sich zu erinnern, die Standardabweichung in IQ-Daten sei mit $\sigma_0 = 10$ anzusetzen. Prüfen Sie dies mit einem χ^2-Varianztest (vgl. 1.18 der Vorlesung) für den vorliegenden Datensatz. Am nächsten Morgen meint derselbe Flurnachbar zu Ihnen, vielleicht sei es doch eher 15. Prüfen Sie auch das.

(Abgabe: kurze Text/Formelzeilen als pdf)

Aufgabe 5.2 : Man kann annehmen, dass die Körperlängen X bzw. Y neugeborener Jungen bzw. Mädchen normalverteilte Zufallsvariablen sind. Eine Untersuchung von $n = 17$ Zwillingspärchen ergab die folgenden Werte:

i	:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Y_i	:	49	53	51	47	54	51	47	45	51	49	51	50	53	51	48	46	54
X_i	:	50	55	51	49	54	52	47	47	50	51	52	49	55	52	49	50	54

In welchem der folgenden Modelle ist die Aussage 'Jungen sind bei ihrer Geburt in Mittel schwerer als Mädchen' statistisch signifikant zum Niveau $\alpha = 0.15$:

i) Zweistichproben-Normalverteilungsmodell (ZN'), unter der Voraussetzung gleicher Varianz in beiden Teilstichproben ?

ii) Zweistichproben-Lokationsmodell ($\mathcal{Z}^*$), mit Mediantest oder Wilcoxontest ?

(Abgabe: kurze Text/Formelzeilen als pdf)

Aufgabe 5.3 (freiwillige Zusatzaufgabe: eine ungewöhnliche Likelihoodfläche): Man betrachte ein 'gestörtes' Lokations- und Skalenmodell für iid Beobachtungen, in dem die Verteilung der Einzelbeobachtung durch ihre Lebesgue-Dichte

$$f_{\mu,\sigma}(x) \;=\; (1-\alpha)\,\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \;+\; \alpha\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}(x-\mu)^2} \quad , \quad x \in I\!R$$

gegeben sei. Dabei seien $\mu \in I\!R$ und $\sigma > 0$ die unbekannten Parameter der Verteilung; der Mischungsparameter $0 < \alpha < \frac{1}{2}$ sei fest und bekannt. LeCam z.B. schlug –nicht ohne Ironie– vor, $\alpha := 10^{-10^{137}}$ zu wählen: siehe LeCam, L.: Maximum likelihood, an introduction. Intern. Statist. Review 58, 151–171 (1990).

a) Man generiere einen kleinen Datensatz von Beobachtungen $X_1, \ldots, X_n$ (etwa: $n = 5$) unter $\mu_0 = 0.5$, $\sigma_0 = 0.5$, $\alpha = 0.3$. Man visualisiere die Loglikelihoodfläche gegeben diese Beobachtungen über einem geeigneten Gitter von (μ, σ)-Werten. Man benutze dabei σ-Werte, die immer näher an 0 herankommen, um die am Rand der Loglikelihoodfläche für $(\mu, \sigma) \to (X_i, 0^+)$, $1 \le i \le n$ entstehenden Singularitäten sichtbar zu machen.

b) Man überlege sich aus der Gestalt der Produktdichten, warum in diesem Modell die Loglikelihoodfunktion basierend auf Beobachtungen $X_1, \ldots, X_n$ Singularitäten für

$$(\mu, \sigma) \;\longrightarrow\; (X_i, 0^+) \quad , \quad 1 \le i \le n$$

aufweist. Was folgt daraus für Maximum-Likelihood-Schätzverfahren?

c) In welcher Weise ändert sich die log-likelihood Fläche bei wachsender Zahl n von Beobachtungen ?

Hinweis 1: Man runde in a) die Beobachtungen $X_1, \ldots, X_n$ auf eine Nachkommastelle, und verwende ein μ-Gitter mit ebenfalls auf eine Nachkommastelle gerundeten Werten ... (warum?)

Hinweis 2: Bereits auf dem Ankündigungsblatt für diese Vorlesung haben Sie diesen Typ von Loglikelihoodfunktion im Bild gesehen ...

(freiwillige Abgabe zu a): Programm und ein guter 'persp'-plot als pdf)