Institut für Angewandte Mathematik
Sommersemester 2024
Andreas Eberle, Stefan Oberdörster

10. Übungsblatt „Einführung in die Statistik"

Abgabe bis Dienstag 25.6.

---

1. (Bestzeiten von Sprintern) Die nachfolgende Tabelle enthält die Bestzeiten (in Sekunden) von zehn Sprintern aus Großbritannien über 200 m bzw. 100 m für das Jahr 1988. Es handelt sich um alle Sprinter, die 1988 die 200 m in weniger als 21.20 Sekunden liefen und außerdem eine Bestzeit über 100 m angaben. Wir vermuten, dass sich die Durchschnittsgeschwindigkeit der Sprinter über 200 m von jener der Sprinter über 100 m systematisch unterscheidet, wir wissen jedoch nicht in welche Richtung: Man könnte vermuten, dass die Durchschnittsgeschwindigkeit der Sprinter über 200 m aufgrund des geringeren Einflusses der Start- und Beschleunigungsphase höher ist als jene der Sprinter über 100 m. Denkbar wäre auch, dass die längere Distanz zu einer niedrigeren Durchschnittsgeschwindigkeit führt. Testen Sie (zum Beispiel unter Verwendung von R) die Nullhypothese, dass es keine systematischen Unterschiede gibt, zum Niveau $\alpha = 0.05$ auf zwei verschiedene Arten.

Athlet	Bestz. 200 m	Bestz. 100 m
L. Christie	20.09	9.97
J. Regis	20.32	10.31
M. Rosswess	20.51	10.40
A. Carrott	20.76	10.56
T. Bennett	20.90	10.92
A. Mafe	20.94	10.64
D. Reid	21.00	10.54
P. Snoddy	21.14	10.85
L. Stapleton	21.17	10.71
C. Jackson	21.19	10.56

2. (Kolmogorov-Smirnov-Bänder und Quantile) Mit einer Sicherheit von $1-\alpha$ können wir davon ausgehen, dass für die empirische Verteilungsfunktion $\|F_n - F\|_{\sup} \leq \kappa_{n,\alpha}$ für einen gewissen kritischen Wert $\kappa_{n,\alpha} > 0$ gilt. Wie die folgenden Betrachtungen zeigen, ergeben sich hieraus automatisch Konfidenzschranken für beliebige Quantile q_γ, $\gamma \in (0,1)$:

a) Bestimmen Sie für $\kappa > 0$ und $\gamma \in (0,1)$ einen möglichst großen Index $k = k(n,\gamma,\kappa) \in \{0,1,\ldots,n\}$ und einen möglichst kleinen Index $\ell = \ell(n,\gamma,\kappa)$, so dass gilt:

$$q_\gamma \ \in \ [X_{(k)}, X_{(\ell)}] \quad \text{falls } \|F_n - F\|_{\sup} \leq \kappa.$$

Illustrieren und prüfen Sie Ihr Ergebnis anhand einer Skizze mit $n = 4$ Beobachtungen, wobei $X_{(1)} < X_{(2)} < X_{(3)} < X_{(4)}$, $\kappa = 0.2$ und $\gamma \in \{0.1, 0.5, 0.7\}$.

b) Bestimmen Sie einen Stichprobenumfang n mit $\mathbb{P}\left[\|F_n - F\|_{\sup} \geq 0.01\right] \leq 0.01$. Welches Konfidenzintervall ergibt sich auf diese Weise für den Median ?

3. (Konfidenzintervalle für den Median)

a) Angenommen man möchte für einen Datensatz mit $n = 30$ Beobachtungen ein Konfidenzintervall für den zugrundeliegenden Median $q_{0.5}$ berechnen. Bestimmen Sie *alle* minimalen Indexpaare (k, ℓ) derart, dass $[X_{(k)}, X_{(\ell)}]$ ein 95%-Konfidenzintervall für $q_{0.5}$ ist. Dabei bedeutet „minimal", dass weder $[X_{(k+1)}, X_{(\ell)}]$ noch $[X_{(k)}, X_{(\ell-1)}]$ ein 95%-Konfidenzintervall für $q_{0.5}$ wäre. Verwenden Sie hierzu die nachfolgende Tabelle der Verteilungsfunktion der Binomialverteilung oder ein eigenes Programm.

x	$F_{30,0.5}(x)$	x	$F_{30,0.5}(x)$	x	$F_{30,0.5}(x)$	x	$F_{30,0.5}(x)$
0	0.0000	8	0.0081	16	0.7077	24	0.9998
1	0.0000	9	0.0214	17	0.8192	25	1.0000
2	0.0000	10	0.0494	18	0.8998	26	1.0000
3	0.0000	11	0.1002	19	0.9506	27	1.0000
4	0.0000	12	0.1808	20	0.9786	28	1.0000
5	0.0002	13	0.2923	21	0.9919	29	1.0000
6	0.0007	14	0.4278	22	0.9974	30	1.0000
7	0.0026	15	0.5722	23	0.9993		

b) Berechnen Sie eine untere 95%-Vertrauensschranke für den Median der Lebensdauer in Monaten von Hauskatzen für den folgenden Datensatz. Formulieren Sie das Ergebnis auch in Worten.

145.4	71.7	147.9	104.9	118.1	105.6	139.3	119.0	138.9	147.7
145.1	80.2	125.6	188.2	169.6	114.2	198.0	96.4	112.6	94.9
167.2	216.6	162.8	165.9	225.7	162.0	89.7	107.0	172.9	66.6

4. (Ordnungsstatistiken) Seien $U_1, U_2, \ldots, U_n$ unabhängige, auf $[0,1]$ gleichverteilte Zufallsvariablen, und seien $U_{(1)}, U_{(2)} \cdots, U_{(n)}$ die entsprechenden Ordnungsstatistiken.

a) Zeigen Sie, dass die Verteilung von $U_{(k)}$ absolutstetig ist mit Dichte

$$f_{U_{(k)}}(u) \;=\; n\binom{n-1}{k-1} u^{k-1}(1-u)^{n-k}\, 1_{[0,1]}(u)\,.$$

b) Begründen Sie ohne lange Rechnung, dass

$$\int_0^1 u^a(1-u)^b du \;=\; (a+b+1)^{-1}\binom{a+b}{a}^{-1} \qquad \text{für alle } a, b \in \mathbb{N}_0.$$

c) Folgern Sie, dass $\mathbb{E}\left[U_{(k)}\right] = \frac{k}{n+1} =: t_k$ und $\operatorname{Var}\left[U_{(k)}\right] = \frac{t_k(1-t_k)}{n+2}$.

5. (Zusatzaufgabe: Asymptotik der G-Statistik) Sei L_n die empirische Verteilung von $X_1, X_2, \ldots, X_n$, wobei X_i ($i \in \mathbb{N}$) unabhängige, identisch verteilte Zufallsvariablen mit Werten in einer endlichen Menge $S = \{a_1, \ldots a_k\}$ und Verteilung μ sind.

a) Zeigen Sie, dass $2nH(L_n|\mu) - nD_2(L_n|\mu)$ für $n \to \infty$ stochastisch gegen Null konvergiert.

b) Folgern Sie, dass die Verteilung von $G = 2nH(L_n|\mu)$ für $n \to \infty$ schwach gegen eine Chiquadrat-Verteilung mit $k-1$ Freiheitsgraden konvergiert.