12. Übungsblatt

Aufgabe 1 (2 + 2 + 1 Puntkte).

Wir betrachten das Beispiel 79 aus der Vorlesung und beobachten, wie oft die Zahl „Vier“ fällt, wenn wir einen Würfel 100 mal werfen.

(a) Bestimmen Sie das (approximative) 0.95-Konfidenzintervall für θ aus Beispiel 60.

(b) Benutzen Sie Ihre Lösung für (a), um einen (approximativen) Test für das Testproblem

$$H_0 : \text{Wahrscheinlichkeit für „Vier“ ist } \frac{1}{6}$$
$$\text{vs. } H_1 : \text{Wahrscheinlichkeit für „Vier“ ist ungleich } \frac{1}{6}$$

zu finden.

(c) Nun zählen wir in unserem Experiment 19 mal die „Vier“. Werten Sie den Test aus (b) aus.

Aufgabe 2 (2 + 2 + 2 + 2 + 2 + 1 Punkte).

Wir untersuchen eine Maschine, die Schrauben herstellt. Uns ist bekannt, dass die Länge der Schrauben im Mittel $\mu = 5$ beträgt (Angabe in cm). Der Hersteller behauptet, dass die Standardabweichung der Länge der Schrauben höchstens $\sigma_0 = 0.3$ cm beträgt. Wir vermuten jedoch, dass die Standardabweichung tatsächlich größer ist. Für eine statistische Untersuchung nehmen wir an, dass die Länge der Schrauben normalverteilt ist mit Mittelwert $\mu = 5$ und (unbekannter) Standardabweichung $\sigma > 0$. Dabei vernachlässigen wir, dass die Schrauben eigentlich keine negative Länge haben können.
Für $n = 10$ Schrauben beobachten wir folgende Längen (in cm):

5.6	5.2	4.1	3.7	6.5	3.6	6.0	6.1	5.2	4.2

(a) Wir betrachten $X_1, ..., X_n \sim N(\mu, \sigma^2)$ unabhängig und identisch verteilt. Stellen Sie zu $\sigma_1, \sigma_0 > 0$, den Likelihoodquotienten $L_x(\sigma_0, \sigma_1)$ auf.

(b) Sei $\alpha \in (0, 1)$. Wir betrachten die einfachen Hypothesen

$$H_0 : \sigma = \sigma_0 \quad \text{gegen} \quad H_1 : \sigma = \sigma_1$$

wobei $\sigma_1 > \sigma_0$ angenommen sei. Zeigen Sie, dass ein $d^* \in \mathbb{R}$ existiert, sodass ein bester Test δ^* für H_0 vs. H_1 gegeben ist durch

$$\delta^*(x) = \begin{cases} 1, & T(x) \geq d^*, \\ 0, & T(x) < d^* \end{cases}$$

mit $T : \mathbb{R}^n \to \mathbb{R}$, $T(x) = \sum_{i=1}^n (x_i - \mu)^2$.

Hinweis: Nutzen Sie Neyman-Pearson. Außerdem dürfen Sie ohne Beweis verwenden, dass $T(X_1, \ldots, X_n)$ eine Dichte besitzt.

(c) Zeigen Sie, dass δ^* sogar gleichmäßig bester Test ist für

$$H_0 : \sigma = \sigma_0 \quad \text{gegen} \quad H_1' : \sigma > \sigma_0.$$

(d) Zeigen Sie, dass δ^* sogar gleichmäßig bester Test ist für

$$H_0' : \sigma \leq \sigma_0 \quad \text{gegen} \quad H_1' : \sigma > \sigma_0.$$

Hinweis: Definieren Sie $Z_n := \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2$ und überlegen Sie sich, dass die Verteilung von Z_n nicht von σ abhängt.

(e) Es ist bekannt, dass $Z_n \sim \chi_n^2$, wobei χ_n^2 die sogenannte zentrale Chi-Quadrat-Verteilung mit n Freiheitsgraden bezeichnet. Es bezeichne $\chi_{n,1-\alpha}^2$ das $(1 - \alpha)$-Quantil dieser Verteilung. Zeigen Sie, dass $d^* = \sigma_0^2 \cdot \chi_{n,1-\alpha}^2$.

(f) Sei nun $\alpha = 0.05$. Werden Sie dem Hersteller der Maschine auf Basis unserer Beobachtungen und des Tests δ^* für die Hypothese H_0' vs. H_1' vorwerfen, dass die Angabe der Standardabweichung falsch ist? Die folgende Tabelle zeigt einige Quantile einer zentralen Chi-Quadrat-Verteilung mit verschiedenen Freiheitsgraden.

	Quantiltabelle der zentralen χ^2 – Verteilungen mit 1,…,10 Freiheitsgraden									
P(X≤x)↓\ dof→	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0.05	0.004	0.1	0.35	0.71	1.14	1.63	2.17	2.73	3.32	3.94
0.1	0.02	0.21	0.58	1.06	1.61	2.2	2.83	3.49	4.17	4.87
0.2	0.06	0.45	1.01	1.65	2.34	3.07	3.82	4.59	5.38	6.18
0.3	0.15	0.71	1.42	2.2	3	3.83	4.67	5.53	6.39	7.27
0.5	0.46	1.39	2.37	3.36	4.35	5.35	6.35	7.34	8.34	9.34
0.7	1.07	2.41	3.66	4.88	6.06	7.23	8.38	9.52	10.66	11.78
0.8	1.64	3.22	4.64	5.99	7.29	8.56	9.8	11.03	12.24	13.44
0.9	2.71	4.61	6.25	7.78	9.24	10.64	12.02	13.36	14.68	15.99
0.95	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31
0.99	6.63	9.21	11.34	13.28	15.09	16.81	18.48	20.09	21.67	23.21
0.999	10.83	13.82	16.27	18.47	20.52	22.46	24.32	26.12	27.88	29.59

Abbildung 1: Quantiltabelle verschiedener Chi-Quadrat-Verteilungen

Aufgabe 3 (t-Test, 1 + 1 + 1 + 1 Punkte).

In dieser Aufgabe wollen wir anhand von Klimadaten des deutschen Wetterdienstes aus den Jahren 1971-2023 überprüfen, ob sich die Jahresdurchschnittstemperatur in Deutschland in den letzten Jahren erhöht hat. Auf der Seite des deutschen Wetterdienstes (`https://www.dwd.de/DE/leistungen/cdc/cdc_ueberblick-klimadaten.html`) lassen sich diverse Wetterdaten herunterladen. In der Datei „Wetter.csv“ auf Moodle finden sich die monatlichen Temperaturdaten der Wetterstation in Großenkneten in einer etwas einfacher zu handhabenden Datei als direkt vom DWD.

(a) Lesen Sie die Datei mittels „read.csv" ein. Falls Sie mit der Cloud-Version von R-Studio arbeiten müssen Sie die Datei erst hochladen. „read.csv" nutzt als Ausgangspunkt für den Namen der Datei ihr aktuelles working directory. Sie können ihr working directory unter „Session > Set Working Directory" wechseln. Wir empfehlen es auf den Ordner zu setzten in der sich die .csv Datei befindet. Die Funktion „read.csv" hat als Ausgabe ein data frame. Auf die Spalten eines Data Frames können Sie mit folgender Syntax zugreifen „frame$Spalte". Erstellen Sie nun ein Vektor mit den Jahresdurchschnittstemperaturen 1971-2000 und 2001-2023 und plotten Sie sich die Werte von 1971-2023.

(b) Fügen Sie dem plot den jeweiligen Mittelwert der Temperatur des ersten und zweiten Zeitraums hinzu.

(c) Wir nehmen nun vereinfacht an, dass die Temperaturen pro Jahr in der Zeit von 1971-2000 durch eine Zufallsvariable $X_i \sim \mathcal{N}(\mu_1, \sigma^2), i \in \{1971, 2000\}$ und in der Zeit von 2001-2023 durch eine Zufallsvariable $Y_i \sim \mathcal{N}(\mu_2, \sigma^2), i \in \{2001, 2023\}$ mit unbekannten μ_1, μ_2, σ dargestellt werden kann. Wir wollen nun die Hypothesen

$$H_0 : \mu_1 \geq \mu_2 \qquad \text{gegen} \qquad H_1 : \mu_1 < \mu_2$$

testen. Dazu führen wir einen Zwei-Stichproben t-Test durch. Berechnen Sie

$$\begin{aligned}
n &= 30, m = 23 \\
\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
\bar{Y} &= \frac{1}{m}\sum_{i=1}^{m} Y_i \\
V_X &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
V_Y &= \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2 \\
S &= \sqrt{\frac{(n-1)V_X + (m-1)V_Y}{n+m-2}} \\
T &= \sqrt{\frac{nm}{n+m}}\frac{\bar{Y} - \bar{X}}{S}.
\end{aligned}$$

Die Zufallsvariable T ist unter der Nullhypothese student-T-verteilt mit $n + m - 2$ Freiheitsgraden. Den Wert der Verteilungsfunktion einer student-T-Verteilung lässt sich in R mit „pt" ausrechnen, wobei der Parameter „df" entsprechend auf die Anzahl der Freiheitsgrade gesetzt werden muss. Berechnen Sie den Fehler 1. Art dieses T-Tests. Welchen Schluss können Sie ziehen?

(d) Wir haben in unseren Hypothesen angenommen, dass beide Stichproben die gleiche Varianz σ^2 haben. Diese Annahme ist natürlich im Allgemeinen nicht erfüllt. Die R-Funktion „t.test" führt den t-Test ohne diese Annahme durch. Nutzen Sie diese Funktion um den t-Test ohne diese Ahnahme durchzuführen.
Hinweis: Die Ergebnisse weichen nur minimal von denen davor ab.

Abgabe in Zweier- oder Dreiergruppen bis Montag, 15.07.2024, 11 Uhr in den Zettelkästen im Mathematikon.

Homepage der Vorlesung:

https://nps.math.uni-heidelberg.de/EWTS24/