

Übungen zur Vorlesung
Datenanalyse - Dr. Terveer, Vogt, Pohl
Sommersemester 2022 Blatt 8 14.06.2022

Aufgabe 21 (Chi-Quadrat-Anpassungstest) Im Lotto 6 aus 49 kann man an den Annahmestellen den Tipp mit Hilfe eines Zufallszahlengenerators online, d.h. im Rechenzentrum der Lotteriegesellschaft erzeugen. Bei einer turnusmäßigen Prüfung des Generators ergaben sich aus 40003600 generierten Tipps folgende Häufigkeits-Verteilung der Sequenzlängen⁵ der Tipps

Sequenzlänge	1	2	3	4	5	6
Häufigkeit	20190890	17904796	1782830	119558	5384	142

Kann man anhand der obigen Daten davon ausgehen, dass der Generator ordnungsgemäß arbeitet? Verwenden Sie den Chi-Quadrat-Test.

Hinweis: Die hypothetische Sequenzlängenverteilung für Lotto 6 aus 49 lautet (bei $\binom{49}{6}$ möglichen Tipps)

Sequenzlänge	1	2	3	4	5	6	Gesamt
Günstige Fälle	7059052	6257790	623414	41624	1892	44	13983816

Aufgabe 22 (Gurtstatistik) 86769 Unfallberichte aus der kanadischen Provinz Alberta sind in folgender Tabelle zusammengefasst.

Gurt benutzt	Art der Verletzung				Gesamt
	keine	leicht	mäßig	schwer	
nein	65963	4000	2642	303	72908
ja	12813	647	359	42	13861
Gesamt	78776	4647	3001	345	86769

Prüfen Sie mit dem Chi-Quadrat-Unabhängigkeitstest, ob es einen Zusammenhang zwischen dem Tragen des Sicherheitsgurtes und der Verletzungsschwere bei Unfällen gibt. Verwenden Sie z.B. die Tabelle der Quantile der $\chi^2(n)$ -Verteilung aus der Vorlesung.

Aufgabe 23 (Kolmogoroff-Smirnoff-Test) Es soll geprüft werden, ob der Datensatz $-1, -0.2, 0.45, 1.05, 1.69$ von einer Normal-Verteilung stammt.

- Führen Sie den KS-Test mit $\alpha = 0.05$ aus. Verwenden Sie die von Lilliefors modifizierte Quantiltabelle der KS-Verteilung aus der Vorlesung.
- Erläutern Sie anhand des vorliegenden Beispiels, weshalb die Kolmogorov-Smirnov-Statistik zum Test auf Vorliegen der (stetigen) Verteilung F die folgende Darstellung hat (dabei seien $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ die geordneten Daten):

$$\sup\{|F(x) - \hat{F}(x)| : x \in \mathbb{R}\} = \max\{|F(x_{(j)}) - \frac{j-i}{n}| : i \in \{0, 1\}, j \in \{1, \dots, n\}\}$$

⁵Die Sequenzlänge eines Tipps ist die maximale Länge einer ununterbrochenen Sequenz benachbarter Zahlen im (aufsteigend angeordneten) Tipp. Z.B. hat der Tipp 4, 5, 17, 18, 19, 43 die Sequenzlänge 3 und in einem Tipp der Sequenzlänge 1 treten keine benachbarten Zahlen auf.