Aufgabe 1

384 zufällig ausgewählte Personen wurden nach ihrem Unfall in einer bestimmten Angelegenheit befragt. Zur statistischen Auswertung wurden die Urteile jeweils in eine von 6 Kategorien eingeordnet und in folgender Tabelle dargestellt:

Kategorie	*I*	*II*	*III*	*IV*	*V*	*VI*
Anzahl der Urteile	58	61	72	67	57	69

Testen Sie mit einem geeigneten Testverfahren zum Niveau $\alpha = 0{,}05$, ob in der Grundgesamtheit alle sechs Kategorien gleich wahrscheinlich sind.

Lösung 1 Chi-Quadrat-Anpassungstest

Der Chi-Quadrat-Anpassungstest vergleicht die beobachteten Häufigkeiten O_i in den Klassen mit den erwarteten Häufigkeiten E_i unter der Annahme, dass die Zahlen gleichverteilt sind.

Bei einer Gleichverteilung der Urteile über die $d = 6$ Kategorien, erwarten wir, dass jede Klasse etwa $E_i = \frac{n}{d} = \frac{384}{6} = 64$ Urteile enthält.

Damit die Ergebnisse dieses Tests zuverlässig sind, muss für mindestens 80% der $i \in \{1, \ldots, d\}$ gelten, dass $E_i \geq 5$ ist und für alle i, dass $E_i \geq 1$ ist. Da $E_i = 64$ sind beide Bedingungen für alle i erfüllt.

Die Chi-Quadrat-Teststatistik D wird wie folgt berechnet:

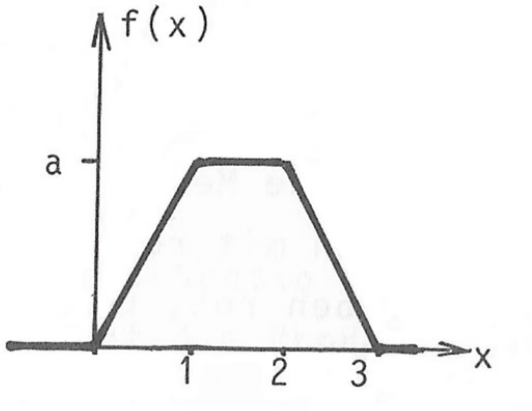
$$\begin{aligned}
D &= \sum_{i=1}^{d} \frac{(O_i - E_i)^2}{E_i} \\
&= \sum_{i=1}^{6} \frac{(O_i - 64)^2}{64} \\
&= \frac{1}{64}\left(6^2 + 3^2 + 8^2 + 3^2 + 7^2 + 5^2\right) \\
&= 3
\end{aligned}$$

Diesen Wert vergleichen wir mit dem 0,95-Quantil der Chi-Quadrat-Verteilung für $d - 1 = 5$ Freiheitsgrade.

Da $D = 3 < 11{,}07 = \chi^2_{5;\,0{,}95}$ gibt es keine ausreichenden Beweise, um die Nullhypothese abzulehnen.

Aufgabe 2

Von einer Zufallsvariablen X wird vermutet, dass sie die nebenstehende Dichte f besitzt mit $f(x) = 0$ für $x \neq [0; 3]$.



a) Bestimmen Sie die Konstante a so, dass f eine Dichte ist.

b) Testen Sie die Vermutung mit folgender Stichprobe zum Niveau $\alpha = 0{,}05$:

Klasse	abs. Häufigkeit
$[0;1]$	15
$(1;2]$	29
$(2;3]$	6

Lösung 2 Chi-Quadrat-Anpassungstest

Damit die Funktion f eine Dichtefunktion sein kann, muss $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ sein, da die Gesamtwahrscheinlichkeit für alle möglichen Ergebnisse einer Zufallsvariablen immer 1 ist. Dies ist der Fall für $a = 0{,}5$.

Bei einer Stichprobe vom Umfang $n = 50$ würde man unter Annahme der Dichte erwarten, dass Klasse A_1 im Intervall $[0;1]$ und Klasse A_3 im Intervall $(2;3]$ jeweils $E_{1,3} = 12{,}5$ Elemente, sowie Klasse A_2 im Intervall $(1;2]$ $E_2 = 25$ Elemente enthält.

Da $E_i \geq 5 \; \forall \; i \in \{1,2,3\}$ sind die Bedingungen für die Anwendung des Chi-Quadrat-Tests erfüllt.

Wir berechnen die Chi-Quadrat-Teststatistik D:

$$\begin{aligned} D &= \sum_{i=1}^{d} \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(15-12{,}5)^2}{12{,}5} + \frac{(29-25)^2}{25} + \frac{(6-12{,}5)^2}{12{,}5} \\ &= 4{,}52 \end{aligned}$$

Diesen Wert vergleichen wir mit dem 0,95-Quantil der Chi-Quadrat-Verteilung für $d-1 = 2$ Freiheitsgrade $\chi^2_{2;\,0{,}95} = 5{,}991$.

Da $D = 4{,}52 < 5{,}991 = \chi^2_{2;\,0{,}95}$ können wir die Nullhypothese nicht ablehnen. Die Vermutung, dass die Zufallsvariable X die Dichte f besitzt, lässt sich also anhand der Daten nicht auf dem 5%-Signifikanzniveau widerlegen.

Aufgabe 3

Bei der Bestimmung des Geburtsgewichts von 100 Mädchen ergaben sich folgende gerundeten Werte:

Gewicht in kg	2,7	2,8	2,9	3,0	3,1	3,2	3,3	3,4	3,5	3,6
Anzahl der Mädchen	6	8	11	13	14	11	13	8	9	7

Testen Sie zum Niveau $\alpha = 0{,}05$ die Hypothese das Geburtsgewicht folgt einer

a) Gleichverteilung in $[2{,}65;\ 3{,}65]$ mit der Klasseneinteilung

$$[2{,}65;\ 3{,}0],\ (3{,}0;\ 3{,}3],\ (3{,}3;\ 3{,}65].$$

b) Normalverteilung mit der Klasseneinteilung

$$(-\infty;\ 2{,}8],\ (2{,}8;\ 3{,}0],\ (3{,}0;\ 3{,}2],\ (3{,}2;\ 3{,}4] \text{ und } (3{,}4;\ \infty).$$

Lösung 3a

Bei einer Gleichverteilung der $n = 100$ Datenpunkte im Intervall $[2{,}65;\ 3{,}65]$ ist die erwartete Häufigkeit E_i der Klasse A_i gleich dem Produkt aus n und der Klassenbreite.

Klasse A_i	Intervall	beobachtete Häuf. O_i	erwartete Häuf. E_i
A_1	$[2{,}65;\ 3{,}0]$	38	35
A_2	$(3{,}0;\ 3{,}3]$	38	30
A_3	$(3{,}3;\ 3{,}65]$	24	35

Da für $i \in \{1, 2, 3\}$ die erwartete Häufigkeit $E_i > 5$ ist, ist der Chi-Quadrat-Anpassungstest für die Untersuchung geeignet.

Wir berechnen die Chi-Quadrat-Teststatistik D:

$$\begin{aligned} D &= \sum_{i=1}^{d} \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(38-35)^2}{35} + \frac{(38-30)^2}{30} + \frac{(24-35)^2}{35} \\ &= \frac{614}{105} \\ &= 5{,}8\overline{476190} \end{aligned}$$

Da $D \approx 5{,}848 < 5{,}991 = \chi^2_{2,0{,}95}$ ist, gibt es keine ausreichende Evidenz, um die Nullhypothese einer Gleichverteilung abzulehnen. Das bedeutet, dass die Daten nicht signifikant von einer Gleichverteilung in [2,65; 3,65] abweichen.

Lösung 3b

Für den Vergleich mit einer Normalverteilung der Datenpunkte werden zunächst der Erwartungswert $\overline{X}$ und die empirische Varianz s^2 der Stichprobe benötigt.

$$\begin{aligned}
\overline{X} &= \frac{1}{n}\sum_{i=1}^{10} h_i \cdot A_i \\
&= \frac{1}{100}(6 \cdot 2{,}7 + 8 \cdot 2{,}8 + 11 \cdot 2{,}9 + 13 \cdot 3{,}0 + 14 \cdot 3{,}1 + \\
&\quad 11 \cdot 3{,}2 + 13 \cdot 3{,}3 + 8 \cdot 3{,}4 + 9 \cdot 3{,}5 + 7 \cdot 3{,}6) \\
&= 3{,}149
\end{aligned}$$

$$\begin{aligned}
s^2 &= \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n}\left(X_i^2\right) - n \cdot \overline{X}^2\right) \\
&= \frac{1}{99} \cdot (998{,}17 - 991{,}6201) \\
&\approx 0{,}0661606
\end{aligned}$$

Für $X = \{\text{Geburtsgewicht}\} \sim \mathcal{N}(3{,}149;\ 0{,}0661606)$ werden folgende Häufigkeiten erwartet.

Klasse A_i	Intervall	beob. Häuf. O_i	erw. Häuf. E_i
A_1	$(-\infty;\ 2{,}8]$	14	8,7417
A_2	$(2{,}8;\ 3{,}0]$	24	19,3784
A_3	$(3{,}0;\ 3{,}2]$	25	29,7384
A_4	$(3{,}2;\ 3{,}4]$	21	25,6840
A_5	$(3{,}4;\ \infty)$	16	16,4574

Wobei allgemein für das Intervall $A_i = (a; b]$ die erwartete Häufigkeit wie folgt berechnet werden kann:

$$E_i = n \cdot \int_a^b \frac{1}{s \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{(x - \overline{X})^2}{2 \cdot s^2}\right)\,\mathrm{d}x$$

ist.

Wir berechnen die Chi-Quadrat-Teststatistik D:

$$\begin{aligned} D &= \sum_{i=1}^{d} \frac{(O_i - E_i)^2}{E_i} \\ &\approx 5{,}88708 \end{aligned}$$

Die Anzahl der Freiheitsgrade ist $d - 1 -$ Anzahl der zu schätzenden Parameter. Da wir hier den empirischen Erwartungswert und die empirische Varianz als Schätzer verwenden, ist die Anzahl der Freiheitsgrade 2 und wir betrachten das 95%-Quantil der χ^2-Verteilung für 2 Freiheitsgrade.

Da $D = 5{,}887 < 5{,}991 = \chi^2_{2,0{,}95}$ gibt es keinen ausreichenden Grund, die Nullhypothese zu verwerfen. Das bedeutet es gibt keine signifikanten Beweise dafür, dass das Geburtsgewicht der Mädchen nicht normalverteilt ist.