

Computergestützte Mathematik

Aufgabe 12.1: Anteil überdurchschnittlicher Werte

Schreiben Sie eine Funktion, die für einen gegebenen Datenvektor $x = (x_1, \dots, x_n)$ den Anteil der überdurchschnittlichen Werte berechnet, also die Anzahl der Werte x_i , die größer als das arithmetische Mittel aller x_i sind, dividiert durch n .

Testen Sie ihre Funktion an einem Vektor aus $n = 10000$ standardnormalverteilten Zufallszahlen.

Berechnen Sie danach den Anteil der flächenmäßig überdurchschnittlich großen US-Bundesstaaten (eingebauter Datenvektor `state.area`).

Aufgabe 12.2: Ein t-Test

- Extrahieren Sie aus dem Iris-Datensatz `iris` (vgl. Aufg. 11.1) jeweils die Breite der Kelchblätter (`Sepal.Width`) der Schwertlilienarten (`Species`) *versicolor* und *virginica* getrennt als Daten-Vektoren A und B .
- Vergleichen Sie die Verteilungen von A und B mit einem Boxplot.
- Testen Sie die Hypothese, dass die beiden Arten im Mittel gleich breite Kelchblätter haben, mit einem t-Test.

Aufgabe 12.3: Ein lineares Modell

Der eingebaute Dataframe `ChickWeight` enthält unter anderem das Gewicht (`weight`, in Gramm) und das Alter (`Time`, in Tagen) von Küken („chicks“).

- Erstellen Sie ein Streudiagramm aller Küken mit `Time` auf der x -Achse und `weight` auf der y -Achse.
- Berechnen Sie die Koeffizienten a und b des linearen Modells

$$\text{weight} = a + b \cdot \text{Time}$$

und zeichnen Sie in das Streudiagramm die berechnete Regressionsgerade mit ein.

- Selektieren Sie nun das Gewicht aller 10 Tage alten Küken als Datenvektor $W10$ und das Gewicht aller 12 Tage alten Küken als $W12$.
Vergleichen Sie $W10$ und $W12$ mit einem Boxplot.