

## Exercise 1 - R: Glutaminsäure bei Säuglingen

Laden Sie den Datensatz (08\_Glu\_Levels\_In\_Infants.txt) herunter und lesen Sie diesen mittels R ein. Bei den Daten handelt es sich um eine Untersuchung von Säuglingen, bei der die Konzentration von Glutaminsäure im Blut vor (pre) und nach (post) der Gabe eines Medikaments bestimmt wurde. Eine Kontrollgruppe von Säuglingen erhielt keine Behandlung. In der Tabelle sind die Glutaminsäure-Werte (in  $\mu\text{mol/ml}$ ) für Säuglinge aus der Versuchsgruppe (T für treatment) und der Kontrollgruppe (C für control) angegeben. Bearbeiten Sie folgende Fragestellungen/Aufgaben unter Verwendung verschiedener Funktionen:

- (a) Erklären Sie anhand des Datensatzes die Begriffe Statistik, Parameter, Population und Stichprobe. Gehen Sie zudem darauf ein, ob die Stichprobe *repräsentativ* ist oder nicht.
- (b) Welche Struktur hat der Datensatz? (Anzahl Variablen, Skalenniveaus, Datentypen)
- (c) Wieviele fehlende Werte beinhaltet der Datensatz in den pre und post Werten? Wie könnte man mit diesen Werten umgehen?
- (d) Berechnung von Mittelwert und Standardabweichung pro Gruppe. Lässt sich ein Unterschied erkennen?
- (e) Geben Sie die fünf Zahlenzusammenfassungen (Minimum, erstes Quartil, Median, drittes Quartil, Maximum) der pre und post Werte pro Gruppe an. Lässt sich ein Unterschied erkennen?
- (f) Visualisieren Sie die Daten durch Wahl (einer) geeigneter grafischer Darstellungsmöglichkeiten. (Line-, Scatter-, Barplots, etc.). Was fällt auf?

## Exercise 2 - Mittelwert oder Median?

- (a) Berechnen Sie bei den beiden folgenden Aufgaben entweder den Median oder den Mittelwert und ein dazugehöriges Streumaß und begründen Sie Ihre Entscheidung kurz!

- Gegeben seien die folgenden Netto-Monatsgehälter (in Euro) der Mitarbeiter/-innen einer Firma:

1200, 1100, 1350, 1300, 1340, 1270, 1600, 1400, 3500, 1900, 1700, 2900.

- Der FGA (Functional Gait Assessment)-Score dient zur Beurteilung der Gangsicherheit von Personen auf einer Skala von 0 bis 30. Dieser Score ist definiert als die Summe von Scores in 10 Teilbereichen (z.B. Gehen mit geschlossenen Augen), die jeweils von 0 (schwer beeinträchtigt) bis 3 (normal) bewertet werden. Folgende Scores wurden in einer Gruppe von Patientinnen und Patienten erhoben:

14, 9, 17, 28, 4, 12, 6, 13, 4, 14, 29, 23, 28.

## Exercise 3 - Skalenniveau und Listen

- (a) Geben Sie die Skalierung der folgenden Merkmale an:
- Schwangerschaftstest
  - Wohnform (bspw. WG, Einzelwohnung)
  - Entfernung der Wohnung von der Universität (drei Ausprägungen: „nah“, „weit“, „sehr weit“)
  - Wohnungsgröße (in  $m^2$ )
  - Temperatur in Celsius
  - Schlafverhalten (1: very bad, 2: bad, 3: ok, 4: good, 5: very good);
  - Durchschnittliche Stunden Schlaf pro Nacht (Einheiten: Stunden);
  - Tumorgöße (in mm);
  - Sauerstoffsättigung (in %);
  - Nasenspray (Nasivin, Fentrinol, Otrivin etc.)
- (b) Recherchieren Sie die Begriffe Verhältnis- und Intervallskala und erkläre die Begriffe. Denken Sie, dass jene Unterscheidung (statistisch, methodisch) sinnvoll ist?
- (c) Bei der Berechnung des arithmetischen Mittels zweier vorgegebener Listen erhält man jeweils den Wert 5. Die Standardabweichung der ersten Liste beträgt 1, die der zweiten Liste 2. Welche Aussagen sind zutreffend und welche falsch? Begründen Sie jeweils!
- Beide Listen haben auch den gleichen Median.
  - Der Modus beider Listen ist ebenso 5.
  - Die Spannweite der zweiten Liste ist größer als die der ersten Liste.

## Exercise 4 - R: Blütenblätter der Schwertlilien

In R ist der bekannte Datensatz von Fisher über Blütenblätter von Schwertlilien (Iris sp.) verfügbar. Dieser Datensatz beinhaltet die Spalten Species, Sepal.Length, Sepal.Width, Petal.Length und Petal.Width. Den Datensatz ist in R unter "iris" gespeichert.

- (a) Informieren Sie sich über den Befehl subset und erzeugen Sie sich damit drei Variable virginica, setosa und versicolor die jeweils die Daten dieser Art enthalten.
- (b) Berechnen Sie für jede Art den Mittelwert für die vier Variablen Sepal.Length, Sepal.Width, Petal.Length und Petal.Width
- (c) Zusatz: Plotten Sie Petal.Length und Petal.Width für alle Spezies in unterschiedlicher Farbe in einer Grafik. Vergessen Sie nicht, xlim und ylim in der plot und points Funktion anzugeben. Siehe als Beispiel Grafik 1. Interpretieren Sie auch die Grafik insbesondere im Hinblick auf eine (eindeutige) Charakterisierung der Arten.

