

R-Einführungsübung

1. Visualisierung von Daten und Berechnung von Kenngrößen

Das Dataframe `d.fuel` enthält die Daten verschiedener Fahrzeuge aus einer amerikanischen Untersuchung der 80er-Jahre. Jede Zeile (row) enthält die Daten eines Fahrzeuges (ein Fahrzeug entspricht einer Beobachtung).

- a) Lesen Sie die Daten ein mit den folgenden R-Befehlen:

```
t.file <- "http://stat.ethz.ch/Teaching/Datasets/NDK/d.fuel.dat"
d.fuel <- read.table(t.file,header=T,sep=",")1
```

Das Argument `sep=","` braucht es, weil die Kolonnen im File `d.fuel.dat` durch Kommata getrennt sind. Sie können übrigens den Inhalt des Files `d.fuel.dat` mit einem Internetbrowser anschauen, indem Sie die obige URL in Ihren Browser eintippen.

- b) Betrachten Sie die eingelesenen Daten.

Im File `d.fuel.dat` wurden die Zeilen durchnummeriert und daher steht in der ersten Spalte die Nummer der Zeile. Die Spalten (columns) enthalten die folgenden Variablen:

```
weight: Gewicht in Pounds (1 Pound = 0.45359 kg)
mpg:    Reichweite in Miles Per Gallon (1 gallon = 3.789 l; 1 mile = 1.6093 km)
type:   Autotyp
```

- c) Wählen Sie nur die fünfte Zeile des Dataframe `d.fuel` aus. Welche Werte stehen in der fünften Zeile?
- d) Wählen Sie nun die erste bis fünfte Beobachtung des Datensatzes aus. So lässt sich übrigens bei einem unbekanntem Datensatz ein schneller Überblick über die Art des Dataframe gewinnen.
- e) Zeigen Sie gleichzeitig die 1. bis 3. und die 57. bis 60. Beobachtung des Datensatzes an.
- f) Berechnen Sie den Mittelwert der Reichweiten aller Autos in Miles/Gallon.
- g) Berechnen Sie den Mittelwert der Reichweite der Autos 7 bis 22.
- h) Erzeugen Sie einen neuen Vektor `t.kml`, der alle Reichweiten in km/l, und einen Vektor `t.kg`, der alle Gewichte in kg enthält.
- i) Berechnen Sie den Mittelwert der Reichweiten in km/l und denjenigen der Fahrzeuggewichte in kg.
- j) Zeichnen Sie ein Streudiagramm, welches den Verbrauch pro 100km als Funktion des Gewichtes in kg darstellt.
- k) Machen Sie eine Stamm-Blatt-Darstellung der Benzinverbräuche pro 100 km. Bestimmen Sie den minimalen und maximalen Verbrauch.

R-Hinweis: `stem()`

¹Alternativ können Sie den Dataframe `d.fuel.dat` von der Internetseite <http://stat.ethz.ch/Teaching/Datasets/NDK> in einen für diesen Zweck erstellten Ordner `Datasets` in Ihrem Home-Directory kopieren (speichern unter `T:/ndk../Datasets/d.fuel.txt`) und dann von dort in R einlesen mit dem Befehl `d.fuel <- read.table("T:/ndk../Datasets/d.fuel.txt",header=T,sep=",")`

- l) Zeichnen Sie zuerst ein Histogramm des Verbrauchs der Autos (pro 100km) mit den Defaulteinstellungen und dann ein Histogramm mit 15 Klassen statt nur 8, einer x-Achse von 0 bis 15 und einem Titel.
- m) Zeichnen Sie einen Boxplot der Benzinverbräuche.
R-Hinweis: `boxplot()`
- n) Vergleichen Sie die Standardabweichung und den MAD der Benzinverbräuche miteinander (vgl. Stat. Datenanalyse, Kap. 2.3).
R-Hinweis: `mad()`, `sd()`
- o) Vergleichen Sie den Mittelwert und den Median der Benzinverbräuche in l/100km.

2. Korrelationen (R-Funktion: `cor()`)

- a) Erzeugen Sie den Vektor `t.x` mit den Werten -10,-9,...,9,10 und den Vektor `t.x1` mit den Werten 0,1,...,9,10.
Erzeugen Sie dann die Vektoren `t.y` und `t.y1`, deren Elemente die Quadratwerte der entsprechenden Elemente von `t.x` bzw. `t.x1` enthalten.
- b) Zeichnen Sie die Streudiagramme `t.y` vs. `t.x` und `t.y1` vs `t.x1` .
- c) Berechnen Sie die Korrelationskoeffizienten zwischen `t.x` und `t.y` bzw. zwischen `t.x1` und `t.y1`. Warum sind die beiden Korrelationen so verschieden (vgl. Stat. Datenanalyse, Abschnitt 3.2.h)?