

Blatt 2

Hintergrund:

Jede Datei mit dem Namen “[Rohstoff].txt” enthält Rohstoffpreise von 1.1.15-19.9.16.

Silber, Gold in USD / Feinunze

Öl in USD / Barrel

Eisenerz in USD / tonne

Kaffee in USD / lb

Kakao in GBP / tonne

Milch in USD / 50l

Osaft in USD / lb

Aufgabe 1 (*Daten aus Textfiles einlesen und bereinigen*)

1. Setzt den Pfad des working directory wie in Blatt 1. Ladet den Workspace, den ihr in der letzte Sitzung gespeichert habt. (`setwd`, `load`)
2. Lest den Datensatz `eisenerz.txt` ein mit `read.table`. Wie müssen die Argumente eingestellt werden, um die Daten fehlerfrei einzulesen? (Hinweis: Schaut in die Textdatei rein, was fällt auf? Nutzt R table import error als Suchbegriff.)
3. Macht euch mit der Struktur der Daten vertraut. Wie viele Variablen/Merkmale gibt es? Sind die Variablen vom richtigen Datentyp? Wie viele Fälle/Individuen gibt es?
4. Überprüft stichprobenweise, ob die Daten richtig eingelesen wurden, indem ihr die 10. bis 20. Reihen in die Konsole ausgeben lasst. (`[]`)
5. Nennt die 2. Spalte um als “Eisenerzpreise”. (`names`)

Aufgabe 2 (*Wiederholung und Vertiefung*)

1. Lest die Datensätze `silber.txt`, `oel.txt`, `kaffee.txt`, `milch.txt` und `osaft.txt` ein mit `read.table`. Macht euch mit der Struktur der Daten vertraut wie bei Teilaufgaben 1.3 und 1.4.
2. (Für Freaks) Lest die Datensätze `silber.txt`, `oel.txt`, `kaffee.txt`, `milch.txt` und `osaft.txt` ein mit `read.table`. Untersucht die eingelesenen Daten. Was fällt auf? Wie kann man das Problem beseitigen?

Aufgabe 3 (*Deskriptive Statistik*)

1. Sucht euch einen Datensatz von `silber.txt`, `oel.txt`, `kaffee.txt`, `milch.txt` und `osaft.txt` aus. Macht euch mit den grundlegenden statistischen Eigenschaften des Datensatzes vertraut. (Hinweis: sämtliche Methoden der deskriptiven Statistik; `apply`)

2. Exportiert den Datensatz anhand einfacher Visualisierung. (`plot`, `hist`, `matplot`, `boxplot`)
3. Was könnt ihr über den Datensatz aussagen? Tauscht euch miteinander aus.

Aufgabe 4 (*Daten aufbereiten*)

1. Untersucht die Zusammenhänge zwischen den Spalten (bis auf die erste Spalte) paarweise anhand `plot`, `matplot`.
2. Berechnet die paarweisen Korrelationen und interpretiert sie. Habt ihr eine Vermutung, warum (meistens) vier Werte einem Datum zugeordnet sind?
3. Was wäre in diesem Kontext sinnvoll, wenn man nur einen repräsentativen Wert pro Tag haben will?

Aufgabe 5 (*Workspace speichern*)

Speichert den aktuellen Workspace in `praxiskurs02.RData`.