

Einführung in die Datenanalyse mit R - Aufgabenblätter

Jan-Philipp Kolb

11 Februar 2016

Aufgabe - Vorbereitung

- Prüfen Sie, ob eine Version von R auf Rechner installiert ist.
- Falls dies nicht der Fall ist, laden Sie [R](#) runter und installieren Sie R.
- Prüfen Sie, ob Rstudio installiert ist.
- Falls nicht - [Installieren](#) sie Rstudio.
- Laden Sie die R-Skripte von meinem GitHub-Account
- Erstellen Sie ein erstes Script und finden Sie das Datum mit dem Befehl `date()` und die R-version mit `sessionInfo()` heraus.

Aufgabe - Zuweisungen und Funktionen

Erzeugen Sie einen Vektor `b` mit den Zahlen von 1 bis 5 und berechnen Sie...

1. den Mittelwert
2. die Varianz
3. die Standardabweichung
4. die quadratische Wurzel aus dem Mittelwert

Aufgabe - Zusatzpakete

Gehen Sie auf und suchen Sie in dem Bereich, wo die Pakete vorgestellt werden, nach Paketen,...

- die für die deskriptive Datenanalyse geeignet sind.
- um Regressionen zu berechnen
- um fremde Datensätze einzulesen (z.B. SPSS-Daten)
- um mit großen Datenmengen umzugehen

Aufgabe - Datenimport

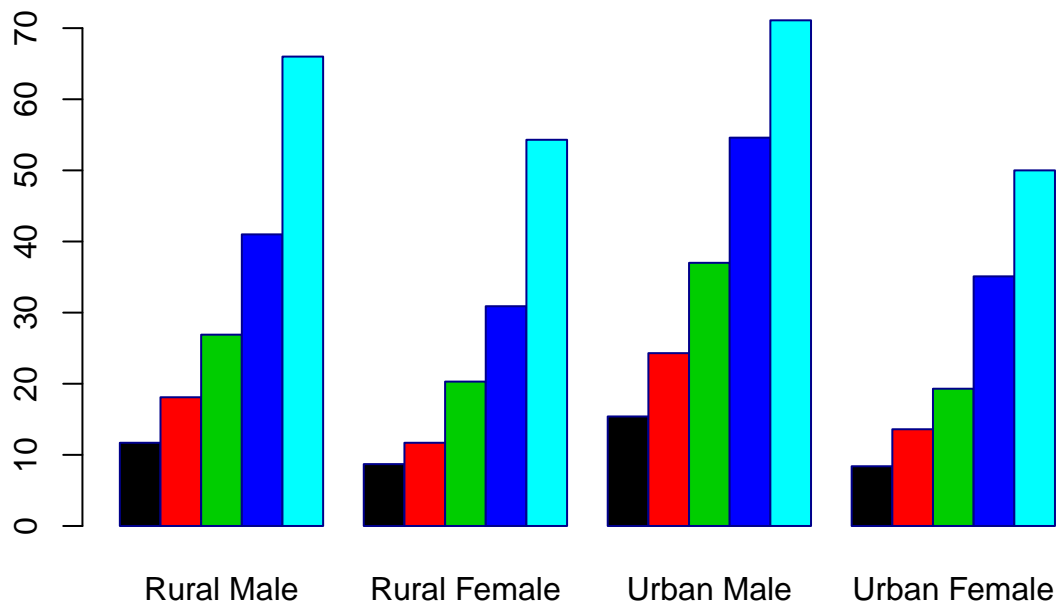
- Gehen Sie auf die Seite des Forschungsdatenzentrums und laden Sie den Campusfile des Mikrozensus 2002 herunter
- Laden Sie den Datensatz mit einer geeigneten Funktion in Ihren Workspace.
- Finden Sie heraus, wieviele Beobachtungen und Variablen der Datensatz umfasst.

Aufgabe - Apply Funktion anwenden

- Erstellen Sie eine Matrix A mit 4 Zeilen und 25 Spalten, die die Werte 1 bis 100 enthält. Analog dazu erstellen Sie eine Matrix B mit 25 Zeilen und 4 Spalten, die die Werte 1 bis 100 enthält.
- Berechnen Sie mittels dem `apply()`-Befehl den Mittelwert und die Varianz für jede Zeile von A bzw. B.
- Berechnen Sie mittels dem `apply()`-Befehl den Mittelwert und die Varianz für jede Spalte von A bzw. B.
- Standardisieren ist eine häufige Transformation von Daten; dafür wird der Mittelwert von der entsprechenden Zeile o der Spalte abgezogen und durch die entsprechende Standardabweichung geteilt. Somit besitzen die Daten einen Mittelwert von 0 und eine Standardabweichung von 1. Standardisieren Sie die Spalten der Matrix A .

Aufgabe - einfache Grafiken

- Laden Sie den Datensatz VADeaths und erzeugen Sie den folgenden plot:



Aufgabe - OECD Datensatz

- Laden Sie den oecd-Datensatz herunter und lesen Sie ihn mit folgender Funktion ein:

```
data <- read.csv("oecd.csv", header = TRUE)
```

- Überprüfen Sie die Dimension der OECD-Daten.

- Berechnen Sie die Mittelwerte und Varianzen der einzelnen Variablen mit einem geeigneten apply Befehl.
- In welchem Land waren die meisten Jugendlichen mindestens zweimal betrunken? Wie hoch ist der maximale Prozentsatz?
- In welchem Land ist die Sterblichkeit am geringsten? Wie hoch ist sie in diesem Land?
- Erstellen Sie einen neuen Datensatz, der aufsteigend nach dem Einkommen geordnet ist. Speichern Sie diesen in einer neuen .csv Datei

Aufgabe - lineare Regression

Datensatz toycars - Paket DAAG

Beschrieben wird Wegstrecke, dreier Spielzeugautos die in unterschiedlichen Winkeln Rampe herunterfahren.

- angle: Winkel der Rampe
 - distance: Zurückgelegte Strecke des Spielzeugautos
 - car: Autotyp (1, 2 oder 3)
- Lesen Sie den Datensatz toycars in einen dataframe data ein und wandeln Sie die Variable car des Datensatzes in einen Faktor (as.factor) um.
 - Erstellen Sie drei Boxplots, die die zurückgelegte Strecke getrennt nach dem Faktor car darstellen.
 - Schätzen Sie für jedes der 3 Autos separat die Parameter des folgenden linearen Modells mit Hilfe der Funktion lm()

$$distance_i = \beta_0 + \beta_1 \cdot angle_i + \epsilon_i$$

- Überprüfen Sie deskriptiv die Anpassung der drei Modelle, indem Sie die Regressionsergebnisse in einen Plot von distance gegen angle einfügen. Deutet das R^2 jeweils auf eine gute Modellanpassung hin?

Aufgabe - Datenanalyse

- Laden Sie einen Datensatz Ihrer Wahl - entweder einen eigenen oder einen der vorgestellten Datensätze
- Berechnen Sie einfache Statistiken auf den wichtigsten Variablen (Mittelwert, Median, Standardabweichung)
- Erzeugen Sie eine zweidimensionale Häufigkeitstabelle
- Führen Sie eine Regression mit sinnvoll gewählten abhängiger und unabhängiger Variablen auf den Daten durch
- Erzeugen Sie einen Lattice-plot