

(11) Programmieraufgabe

Bearbeiten Sie die folgenden Programmieraufgaben. Die Zwischenüberschriften geben an, auf die Inhalte welcher Einheit des Vorkurses sich Teile der Programmieraufgabe beziehen.

(2) R und VS Code

1. Über die Kurswebseite wird die Datei `Salaries.csv` bereitgestellt. Laden Sie diese Datei herunter und speichern Sie sie auf Ihrem PC.
2. Platzieren Sie die Datei in einem Ordner auf Ihrem PC und legen Sie das R-Skript `Programmieraufgabe.r` an. Öffnen Sie es in VS Code.

(6) Datenmanagement

3. Laden Sie den Inhalt der CSV-Datei mit `read.csv` in einem Dataframe D1. Speichern Sie die Anzahl der Zeilen von D1 in die Variable `n`.

(5) Listen und Dataframes

4. Erzeugen Sie einen neuen Dataframe D2, indem Sie aus D1 die Spalten `salary`, `yrs.since.phd` und `yrs.service` extrahieren. Geben Sie die ersten 10 Zeilen von D2 aus.

(8) Berechnung von Deskriptivstatistiken

5. Ermitteln Sie Stichprobenmittel, Stichprobenvarianz und Stichprobenstandardabweichung aller Variablen in D2 und speichern Sie diese in einen neuen Dataframe `D2_var`.

(9) Häufigkeitsverteilungen und Verteilungsfunktionen

6. Erstellen Sie ein Histogramm der Variable `salary` mit Klassen zwischen 50.000 \$ und 250.000 \$ und einer Klassenbreite von 10.000 \$. Speichern Sie die Abbildung als `Histogramm_1.pdf`.
7. Erstellen Sie ein Histogramm der Variable `yrs.since.phd` mit Klassenzentren von 0 bis 60 Jahren in Schritten von 5 Jahren. Speichern Sie die Abbildung als `Histogramm_2.pdf`.

(10) Maße der zentralen Tendenz und der Datenvariabilität

8. Ermitteln Sie Stichprobenmittel, Stichprobenmedian und Stichprobenmodalwert aller Variablen in `D2` und speichern Sie diese in einen neuen Dataframe `D2_mean`.

(7) Strukturiertes Programmieren, (3) Arithmetik und Variablen

9. Erzeugen Sie einen n -dimensionalen Vektor `z`, bei dem alle Einträge zunächst mit `NA` befüllt werden.
10. Benutzen Sie eine `for`-Schleife und `if-else`-Klauseln, um das i -te Element von `z` auf 1 zu setzen, wenn der i -te Eintrag von `D1$sex` gleich "Female" ist, und auf 0, wenn dieser gleich "Male" ist.

(4) Vektoren und Matrizen, (3) Arithmetik und Variablen

11. Extrahieren Sie die Variablen `yrs.since.phd` und `salary` in die n -dimensionalen Vektoren `x` und `y`. Erzeugen Sie eine $n \times 3$ Matrix `X`, deren erste Spalte ein Einsvektor ist, deren zweite Spalte mit `z` befüllt wird und deren dritte Spalte der Variable `x` entspricht.
12. Berechnen Sie für ein lineares Regressionsmodell mit dem Datenvektor `y` und der Designmatrix `X` den Betaparameterschätzer mithilfe der Formel $\hat{\beta} = (X^T X)^{-1} X^T y$. Sie sollten folgendes Ergebnis erhalten:

Betaparameterschätzer : 93105.44 -7923.624 958.079