

3. Übungsblatt
Abgabe: 18. November, 16:15

Einlesen von Daten in R

Dateien, die Daten enthalten, können Sie mittels des Befehls `read.table()` einlesen (siehe Abschnitt 7.3 im R-Skript). Speichern Sie dazu die Dateien auf ihrem Computer. Sie können R mitteilen, in welchem Ordner Sie die Dateien gespeichert haben, indem Sie den Menüpunkt „Datei“ - „Vorzeichnis wechseln“ aufrufen. Nach dem Einlesen sind die Daten als data frame (siehe Abschnitt 2.5 im R-Skript) verfügbar.

Aufgabe 1: Zweidimensionale Daten
(1+1+1+3+1=7 Punkte)

In einem Versuch wurde der Widerstand von Kupferdraht in unterschiedlichen Dicken untersucht. Dazu wurde jeweils eine Spannung an 1m lange Stücke angelegt. In der Datei **widerstand.txt** sind die Radien in mm und die Widerstände in Ω aufgezeichnet.

- a) Zeichnen Sie mittels `plot()` ein Streudiagramm von Widerstand und Radius. Was kann man über die Korrelation der beiden Größen aussagen?
- b) Berechnen Sie die empirische Kovarianz.
- c) Berechnen Sie mittels der vordefinierten Funktion `cor()` den empirischen Korrelationskoeffizienten (Bravais-Pearson) und den Rang-Korrelationskoeffizienten (Spearman).
- d) Überprüfen Sie den Rangkorrelationskoeffizienten, indem Sie diesen ohne Verwendung von `cor()` berechnen.
- e) Finden Sie ein einfaches Daten-Beispiel, bei dem der Rang-Korrelationskoeffizient negativ ist und der empirische Korrelationskoeffizient positiv.

Aufgabe 2: Quantilplots
(3+2+1+1=7 Punkte)

Um die Klimaanlage zu überprüfen, wurde in einer Firma an 20 Arbeitstagen je dreimal die Temperatur gemessen (Datensatz **temperatur.txt**). Der Sollwert liegt bei 21 Grad. Entscheiden Sie anhand des Quantilplots, welche Verteilung die Temperatur am besten annähert, wobei folgende Möglichkeiten in Frage kommen:

- $X \sim \text{Exp}(0.05)$,
- $X \sim U(20, 22)$,
- $X \sim N(21, 1)$,
- $X \sim \text{Wei}(1, 1/21)$ (siehe Seite 2).

- a) Berechnen Sie die Quantilfunktionen analytisch (außer bei der Normalverteilung - hier können Sie in Teil b) mit `qnorm()` arbeiten).
- b) Erstellen Sie für die 4 Verteilungen jeweils Quantilplots mit den Temperatur-Daten. Hier können Sie mit `plot()` arbeiten, indem Sie als 1. Argument die x -Werte und als zweites die y -Werte der Punkte jeweils als Vektor übergeben. Mit `abline(0,1)` können Sie noch die Diagonale einzeichnen.
- c) Welche der 4 Verteilungen passt am besten zu den Daten?
- d) Erstellen Sie mittels `qqnorm()` und `qqline()` einen weiteren Quantilplot. Worin liegen die Unterschiede zu dem in Teil b) für die Normalverteilung erzeugten Plot?

Anmerkung: Die Weibullverteilung wird meistens für Ausfallwahrscheinlichkeiten von technischen Geräten oder Ähnlichem verwendet. Die Verteilungsfunktion einer $\text{Wei}(r, c)$ -verteilten Zufallsvariable ist gegeben durch:

$$F(x) = 1 - e^{-cx^r}.$$

Aufgabe 3: Erwartungswert der Stichprobenvarianz

(1+3=4 Punkte)

Es sei (X_1, \dots, X_n) eine Stichprobe aus unabhängigen identisch verteilten Zufallsvariablen mit $\mathbb{E}X_i = \mu$ und $\text{Var } X_i = \sigma^2$. Zeigen Sie, dass

a) $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$

b) $\mathbb{E}(S_n^2) = \sigma^2$.

Hinweis zu Teil b): Schreiben Sie $X_i - \bar{X}_n$ als $(X_i - \mu) + (\mu - \bar{X}_n)$.