

Aufgaben zur Vorlesung „Werkzeuge der empirischen Forschung“

Aufg. 1) (5 P.) Lesen Sie folgende Dateien aus dem Verzeichnis
https://www.informatik.hu-berlin.de/~koessler/SAS_Kurs/Kursdaten/ ein:

tibetan.dat	computer.dat	darwin.dat
ttest.dat	banknote.dat	toxaemia.dat
skull.dat	synchro.dat	water.dat
heroin.dat	geyser.dat	maschinen.dat
challenger.dat		

Geben Sie die dabei die Eingabedatei jeweils direkt als URL an, d.h. mit `file="https://.../Kursdaten/Dateiname.dat"` statt sie herunterzuladen. Benennen Sie die Spalten korrekt nach der Datenbeschreibung (s.u.) und achten Sie auch auf korrekte Datentypen (d.h. Spaltenklassen). Lesen Sie die Dateien jeweils als `data.frame` ein und sichern Sie einen der `data.frames` mit der Funktion `save`.

Hinweis: Eine Beschreibung der Dateien finden Sie unter:

https://www.informatik.hu-berlin.de/~koessler/SAS_Kurs/Kursdaten.pdf

Aufg. 2) (4 P.) Berechnen Sie folgende Wahrscheinlichkeiten (mit R (1 P.) und auf Papier (3 P.)):

- $P(X > 6)$, wenn $X \sim Bi(10, 5/6)$
- $P(X > 6)$, wenn $X \sim Poi(50/6)$
- $P(X > 6)$, wenn $X \sim Geo(5/6)$
- $P(X > 6)$, wenn $X \sim N(6, 3^2)$

Aufg. 3) (4 P.) (auf Papier oder pdf-File) Berechnen Sie die Varianz einer binomialverteilten Zufallsvariablen $X \sim Bi(n, p)$.

Aufg. 4) (2 P.) (Poisson-Modelle)

- a) (1 P.) Angenommen, die Anzahl der Computerabstürze pro Tag an einem Institut sei poissonverteilt mit Parameter $\lambda = 0.01$, d.h. deren erwartete Anzahl ist gleich 0.01. Berechnen Sie die Wahrscheinlichkeit dafür, dass morgen an diesem Institut wenigstens ein Computer ausfällt.
- b) (1 P.) Angenommen, die Anzahl der Computerabstürze pro Jahr an einem Institut sei poissonverteilt mit Parameter $\lambda = 3.65$. Berechnen Sie die Wahrscheinlichkeit dafür, dass im Laufe des Jahres an diesem Institut wenigstens drei Computer ausfallen.

Folgende Aufgaben können Sie bearbeiten, wenn Sie noch Übungsbedarf haben. Sie werden nicht korrigiert.

Aufg. R01-1) (keine Abgabe, 0 P.) Schreiben Sie eine R-Funktion, die einem gegebenem Vektor $v \in \mathbb{R}^n$ seine Länge (d.h. euklidische Norm im \mathbb{R}^n) zuordnet. Testen Sie sie mit den Vektoren $a = (1, 2, 3)$, $b = (4, 5, 6)$ und $c = (7, 8, 9)$. **Hinweis:** Ohne Funktionsnamen und Zuweisung genügen 25 Zeichen für diese Funktionsdefinition.

Aufg. R01-2) (keine Abgabe, 0 P.) In der Ebene seien sechs Punkte durch ihre Koordinaten gegeben:

x	y
-1	5
2	0
-1	3
3	-2
4	-8
-8	-1

Erzeugen Sie einen `data.frame`, der neben den angegebenen Koordinaten x, y die Nummer des Quadranten von (x, y) und den Quotienten $\frac{x}{y}$ enthält und deren Einträge nach den Quadranten geordnet sind.

Hinweis: Nutzen Sie z.B. `ifelse`, `for` oder `sapply/Map`. Punkte auf den Achsen können Sie beliebig einem Quadranten zuordnen oder `NA` nutzen.