

Übungen zur Vorlesung Wahrscheinlichkeitsrechnung und Mathematische Statistik (für Informatiker)

Blatt 1

Aufgabe 0:

Installieren Sie die Statistische Software R vom CRAN (Comprehensive R Archive Network) von <http://CRAN.R-project.org>. Machen Sie sich mit den Grundlagen der Programmiersprache R vertraut. Im Moodle finden Sie ein Tutorial-Blatt, auf dem gängige Befehle aufgelistet sowie umfangreiche Einführungen verlinkt sind.

Aufgabe 1: (per Hand)

- (a) Betrachten Sie einen fiktiven medizinischen Datensatz mit folgenden Angaben:

Patientennummer, Vorname, Nachname, Geschlecht, Alter in Jahren, Blutgruppe, Blutdruck, Herzfrequenz, Anzahl Vorerkrankungen.

Geben Sie jeweils das Skalenniveau der Variablen an.

- (b) Betrachten Sie einen fiktiven Fußballdatensatz mit folgenden Angaben:

Datum des Spiels, Namen der Mannschaften, Tore der Mannschaften, Exakte Uhrzeit des Anstoßes, Tabellenplätze der Mannschaften, Anzahl Zuschauer im Stadion.

Geben Sie jeweils das Skalenniveau der Variablen an.

Aufgabe 2: (mit R)

Betrachten Sie den Datensatz `airquality`, welcher in der Basisversion von R verfügbar ist und sich mit `data(airquality)` aufrufen lässt.

- (a) Besuchen Sie die Hilfsseiten des Datensatzes über R und finden Sie heraus, in welchen Einheiten und Skalen die sechs Variablen `Ozone`, `Solar.R`, `Wind`, `Temp`, `Month`, `Day` angegeben sind.
- (b) Entfernen Sie alle Einträge ($\hat{=}$ Zeilen), die fehlende Werte `NA` enthalten.
- (c) Erstellen Sie ein Histogramm für die maximalen Tagestemperaturen `Temp`. Wählen Sie dabei die Klassen < 70 , $70-75$, $75-85$, $85-90$, > 90 .

Hinweise: Hilfeseiten von Funktionen oder Objekten werden generell über `?_____` angefordert, z. B. `?hist`. Dort finden Sie auch optionale Argumente für die Funktion `hist` für Aufgabe (c). Fehlende Einträge `NA` können über `na.omit()` aus Objekten entfernt werden. Alternativ kann über `is.na()` abgefragt werden, ob und wo Objekte NAs enthalten.

Aufgabe 3: (mit R)

Betrachten Sie zuletzt den folgenden "Datensatz". (Lassen Sie dazu folgenden R-Code durchlaufen:)

```
ID <- 1:150
set.seed(1)
location <- sample(c(rep("A",70), rep("B",60), rep("C",20)))
result <- c()
result[location=="A"] <- rnorm(70, mean=10, sd=3)
result[location=="B"] <- rnorm(60, mean=12, sd=6)
result[location=="C"] <- rnorm(20, mean=9, sd=2)
dat <- data.frame(ID=ID, location=location, result=result)
```

Es handelt sich ganz offensichtlich um eine professionelle medizinische Studie, zu der Patienten (**ID**) an verschiedenen Orten (**location**) teilgenommen haben. Die resultierenden, hochkomplexen Ergebnisse (**result**) sind für die Studienauswertung von Interesse.

- (a) Bestimmen Sie die absoluten und relativen Häufigkeiten der Studienorte **A**, **B** und **C**. Erstellen Sie ein geeignetes Stabdiagramm für die relativen Häufigkeiten.
- (b) Zeichnen Sie die empirischen Verteilungsfunktionen der Variable **result** getrennt nach den Ausprägungen für **location**. Achten Sie dabei auf Vergleichbarkeit der Grafiken.
- (c) Beurteilen Sie anhand der empirischen Verteilungsfunktionen aus (b), welche Gruppe **location** (**A**, **B**, **C**) sich bezüglich **result** von den anderen beiden Gruppen stark unterscheidet. Zeichnen Sie anschließend die empirische Verteilungsfunktion der „Ausreißergruppe“ im Vergleich zur empirischen Verteilungsfunktion aller anderen Gruppen zusammen in eine Grafik. Denken Sie dabei an Unterscheidbarkeit der beiden Kurven durch die Wahl von Farbe oder Form. Fügen Sie eine Legende hinzu.

Hinweise: Die Funktion `plot.ecdf()` wird Ihnen helfen. Für allgemeine Plot-Funktionen in R stehen Ihnen viele Optionen zur Verfügung. Ein paar Beispiele sind `xlim`, `xlab` (respektive `ylim` und `ylab`), `add`, `col`, `lwd`, `lty` uvm.

Die Besprechung dieses Übungsblatts findet an den Donnerstagen **18.10.2018** und **25.10.2018** jeweils 18:00 Uhr im HS1 EF50 statt. Die Teilnahme an einem oder gar beiden Terminen steht Ihnen frei. Es wird jedoch als Vorbereitung für die Klausur eindringlich empfohlen, dem Übungsbetrieb während des Semesters zu folgen.