

# Übungen zur Vorlesung Wahrscheinlichkeitsrechnung und mathematische Statistik

## Übungsblatt 1

### Aufgabe 1: Merkmale und Datentypen

- (a) Gegeben sei ein Datensatz mit folgenden Angaben von Laptops:  
Herstellungsdatum, Prozessor, Prozessornummer, Gewicht, Taktfrequenz, Anzahl der Prozessorkerne, Cachegröße, Bildschirmdiagonale, Auflösung.  
Geben Sie jeweils das Skalenniveau der Variablen an.
- (b) Gegeben sei ein Datensatz mit folgenden individuellen Angaben von Patienten:  
Name, Geburtsdatum, Alter, Geschlecht, Augenfarbe, Blutgruppe, Gewicht, Größe, Bewertung des Gesundheitszustands (1-10), Körpertemperatur.  
Geben Sie jeweils das Skalenniveau der Variablen an.

### Aufgabe 2: Statistische Software R

Installieren Sie die Statistische Software R vom Comprehensive R Archive Network (CRAN), erhältlich von der Seite <http://CRAN.R-project.org>. Machen Sie sich mit den Grundlagen der Programmiersprache R vertraut, zum Beispiel durch das Studium der Folien `WRUMS_Software.pdf` von der Internetseite der Vorlesung.

### Aufgabe 3: Tabellarische und grafische Darstellung

Laden Sie von der Seite der Veranstaltung den Datensatz 'Daten' herunter. Er liegt sowohl im *.RData* als auch im *.txt* Format vor.

Dieser Datensatz enthält folgende klinische Variablen von 100 Patienten: Patientennummer, Therapie, Geschlecht, Alter, Größe, Gewicht vor und nach der Therapie, Blutdruck, Raucherstatus, Überlebenszeit, Ereignisindikator (1 für verstorben, 0 für lebt). Bei manchen Patienten fehlen bestimmte Einträge, entfernen Sie diese Patienten komplett für die folgenden Analysen.

- (a) Geben Sie für jedes in der Tabelle dargestellte Merkmal an, um welchen Skalen- und Merkmalstyp es sich handelt.
- (b) Bestimmen Sie für die in der Tabelle aufgelisteten Patienten die absoluten und relativen Häufigkeiten der angewendeten Therapien. Erstellen Sie geeignete Stabdiagramme.

- (c) Zeichnen Sie die empirischen Verteilungsfunktionen des Blutdrucks von den rauchenden und nichtrauchenden Patienten.
- (d) Erstellen Sie ein Histogramm für die Größe der Patienten. Teilen Sie die Daten dazu in die Gruppen 100 bis 120 cm, 120 bis 180 cm und 180 bis 200 cm ein.
- (e) Erstellen Sie ein Histogramm der Variable Gewicht1 (Gewicht vor Therapie). Wählen Sie dabei eine für die Daten geeignete Klassenbreite aus.

Hinweis: Das Laden der Daten in R kann mit den Befehlen `load("Daten.RData")` für den R-Datensatz oder `read.table("Daten.txt",header=T)` für die Textdatei durchgeführt werden. Die Auswahl der Patienten ohne fehlende Beobachtungen erfolgt zum Beispiel mittels `select <- which(rowSums(is.na(Datensatz))>0|Datensatz$Raucherstatus=="")` und `Daten <- Datensatz[-select,]`. Danach sind die Funktionen `\attach{Daten}`, `table`, `plot.ecd` und `hist` nützlich, siehe jeweils die Hilfeseiten dazu.

**Besprechung** der Aufgaben: Donnerstag, 19.10.2017, 18:05 Uhr in EF 50, Hörsaal 1.