

Übungen

Die Aufgaben können in **2er Gruppen** bearbeitet werden. Für jede Aufgabe sollen Sie ein R Skript erstellen, das Sie als „Aufgabennr.Vorname1.Vorname2.R“ speichern, also z.B. „3.2.Bernd.Ute.R“ für das R Skript zur Aufgabe 3.2.

Abgabe: Dienstag, 28.09.2011; 17:00 Uhr per Email

5 Grundlegende Definitionen, Merkmalstypen

Grundlegende Definitionen

Aufgabe 5.1. (4 Punkte)

Füllen Sie den Fragebogen aus und überlegen Sie sich:

- (a) Was sind hier die statistische Einheiten?
- (b) Was ist die Grundgesamtheit?
- (c) Welche Stichprobe wird befragt?
- (d) Welche Merkmale und zugehörige Merkmalsausprägungen gibt es?

Merkmalstypen

Aufgabe 5.2. (4 Punkte)

Diskutieren Sie die im Rahmen des Münchener Mietspiegel erhobenen Merkmale *Nettomiete*, *Wohnfläche*, *Baualter*, *Gebäudetyp* (Ausprägungen: *Hochhaus/Wohnblock*), *Zentralheizung* (*dezentral betriebene Strom- oder Gasheizungen/Einzelöfen/keine Heizung*), *Warmwasserversorgung* (*einfache/keine*), *Lage der Wohnung* (*einfache/durchschnittliche/gute/beste*) und *Ausstattung des Bads* (*keins/nicht gekachelt/zweites vollständiges Bad vorhanden/besondere Zusatzausstattung*) hinsichtlich ihres jeweiligen Skalenniveaus. Entscheiden Sie zudem, ob es sich um diskrete oder stetige, bzw. quantitative oder qualitative Merkmale handelt.

(Bitte wenden!)

6 Funktionsverläufe

Aufgabe 6.1. (4 Punkte)

Skizzieren Sie die folgenden Verteilungen:

- Gamma-Verteilung
- Exponentialverteilung
- Binomialverteilung
- Poissonverteilung

Aufgabe 6.2. (4 Punkte)

Erstellen Sie eine Zeichnung, die den Funktionsverlauf der Verteilungsfunktion Standardnormalverteilung zeigt. Fügen Sie dieser Zeichnung die empirische Verteilungsfunktion von 100 Zufallsdaten hinzu, die mittels der Standardnormalverteilung generiert sind.

- Generieren Sie einen Vektor mit den Zufallsdaten und sortieren Sie diesen.
- Plotten Sie die empirische Verteilungsfunktion
- Fügen Sie die Verteilungsfunktion der Normalverteilung mit ins Diagramm ein

Wiederholen Sie den Vorgang, wobei diesmal 1000 Zufallsdaten zu generieren sind.

(Bitte wenden!)

7 Darstellung univariater Daten

Grafische Darstellungen

Aufgabe 7.1. (3 Punkte)

Bei der letzten Landtagswahl in Berlin (im September 2011) ergab sich folgende Stimmverteilung beim Merkmal *Zweitstimme* (bei 1.460.012 gültigen Stimmen):

SPD	CDU/CSU	Grüne	Piraten	Die Linke	Andere
413.124	340.992	256.940	129.795	170.829	138.332

- Geben Sie die Daten als Vektor ein und ordnen Sie den Vektor absteigend. Berechnen Sie die zugehörigen prozentualen Anteile an den abgegebenen (und gültigen) Stimmen auf eine Nachkommastelle genau.
- Erzeugen Sie mit den Daten aus (a) ein mit den Parteinamen und den zugehörigen Prozentzahlen beschriftetes Kreissektorendiagramm (in den entsprechenden Parteifarben).
- Erstellen Sie ein geordnetes Säulendiagramm in den entsprechenden Parteifarben.

Darstellung quantitativer Merkmale

Aufgabe 7.2. (4 Punkte)

Lesen Sie den Datensatz `nettomieten.csv` ein. Die erste Spalte mit dem Namen `nm` gibt das Merkmal "Nettomiete" wieder. Lesen Sie diese in einen Vektor ein.

- Nehmen Sie eine geeignete Klasseneinteilung vor. Zeichnen Sie anschließend ein Säulendiagramm des Datensatzes.
- Listen Sie die absoluten Häufigkeiten auf (Tipp: Nutzen Sie den Befehl `table`)
- Zeichnen Sie ein Histogramm.
- Beurteilen Sie die Schiefe der Verteilung.

Kenngößen quantitativer Merkmale

Aufgabe 7.3. (3 Punkte)

Wir betrachten wieder den Datensatz `nettomieten.csv`

- Berechnen Sie arithmetisches Mittel und den Median der Spalte mit den Nettomieten.
- Berechnen Sie das arithmetische Mittel, den Median und den Modus der gruppierten Daten. Was fällt Ihnen auf?
- Berechnen Sie die Varianz und die Spannweite der Daten.

Quantile, Boxplots und Normal-Quantil-Plots

Aufgabe 7.4. (5 Punkte)

- (a) Zeichnen Sie den NQ-Plot des Nettomietendatensatzes. Zeichnen Sie anschließend zum Vergleich einige NQ-Plots eines mit der Normalverteilung generierten Datensatzes. Wählen Sie dazu Anzahl der Zufallsdaten, sowie Mittelwert und Varianz entsprechend zu den Daten aus dem Nettomietendatensatz.
- (b) Erzeugen Sie einen NQ-Plot jeweils eines Datensatzes mit 1000 Zufallswerten, der mit der $B(0.1, 10)$ -, der $B(0.9, 10)$ - bzw. der $\Gamma(5, 1)$ - Verteilung generiert werde.
- (c) Man verwende den Befehl `boxplot`, um mit R ein Boxplot des Nettomietendatensatzes zu erzeugen.
- (d) Was lässt sich aus dem NQ-Plot, bzw. dem Boxplot hinsichtlich der Schiefe des Nettomietendatensatzes schließen? Vergleichen Sie insbesondere den NQ-Plot des Nettomietensatzes mit dem der Zufallsdaten, die mit Hilfe der $\Gamma(5, 1)$ -Verteilung generiert wurden.