Prof. Dr. Volker Schmidt
Björn Kriesche, M.Sc.

SS 2014
29. 04. 2014

# **Stochastik I** - Übungsblatt 1

Abgabe am 06. 05. **vor** Beginn der Übung

- Übungsblätter müssen vor Beginn der Übung abgegeben werden, nach 16:15 Uhr können keine Lösungen mehr angenommen werden
- bitte Namen und SCL-Logins deutlich aufs Blatt schreiben
- es ist eine Anmeldung zur Vorlesung im SLC-Portal notwendig
- Übungsblätter sollen zu zweit abgegeben werden, stehen mehr als zwei Namen auf dem Blatt, können leider keine Punkte vergeben werden
- mehrere Blätter bitte tackern
- um zur Klausur zugelassen zu werden müssen insgesamt auf allen Übungsblättern mindestens 50% der Übungspunkte erreicht werden
- bei **R**-Aufgaben immer Quelltext und Ausgabe des Programms ausdrucken und abgeben; bei Aufgaben mit Grafikausgabe diese ebenso ausdrucken und abgeben

**Aufgabe 1** (4+2,5+2,5+2 Punkte)
Es soll der Datensatz miete03.asc (steht auf der Homepage zum Download bereit) mit **R** untersucht werden. Dieser enthält Daten über 2053 Wohnungen in München aus dem Jahr 2003. Hierbei sind die Daten wie folgt bezeichnet: **nm**: Nettomiete in EUR; **nmqm**: Nettomiete pro $m^2$ in EUR; **wfl**: Wohnfläche in $m^2$; **rooms**: Anzahl der Zimmer in der Wohnung; **bj**: Baujahr der Wohnung; **bez**: Stadtbezirk; **wohngut**: Gute Wohnlage? (J=1, N=0); **wohnbest**: Beste Wohnlage? (J=1, N=0); **ww0**: Warmwasserversorgung vorhanden? (J=0, N=1); **zh0**: Zentralheizung vorhanden? (J=0, N=1); **badkach0**: Gekacheltes Badezimmer? (J=0, N=1); **badextra**: Besondere Zusatzausstattung im Bad? (J=1, N=0); **kueche**: Gehobene Küche? (J=1, N=0).

(a) Lies die Datei miete03.asc ein und erstelle einen Data Frame. Gib folgende Informationen aus dem erstellten Data Frame aus:
   (i) Wie viele Wohnungen verfügen über eine Zentralheizung?
   (ii) In welchem Stadtbezirk liegt die Wohnung mit der geringsten Nettomiete pro $m^2$?
   (iii) Ermittle die maximale Anzahl $n$ an Zimmern in einer Wohnung? Gib an, wie viele der Wohnungen mit $n$ Zimmern in Stadtbezirk 1 oder 2 liegen.

(b) Für jede Wohnung wird nun eine Bewertungszahl berechnet. Eine Wohnung erhält jeweils einen Punkt für folgende Merkmale: Warmwasserversorgung, Zentralheizung, gekacheltes Badezimmer und gehobene Küche. Sollte sich die Wohnung nicht in guter Wohnlage befinden, gibt es einen Minuspunkt. Außerdem erhalten alle Wohnungen in den Bezirken 3 und 4 jeweils zwei Zusatzpunkte. Erstelle eine weitere Spalte **bew** im Data Frame, die die Bewertung für alle Wohnungen enthält.

(c) Zeichne einen Plot, in dem für alle Wohnungen mit Baujahr später als 1990 und nicht guter Wohnlage der Mietpreis in Abhängigkeit von der Wohnfläche dargestellt ist. Beschrifte sowohl den Plot als auch die Achsen.

(d) Schreibe eine Funktion, die einen Data Frame und eine Zahl $b$ als Eingabe hat und zurück gibt, wie viele der Wohnungen in Bezirk $b$ liegen. Wende die Funktion auf die Wohnungsdaten mit Bezirk 3 an.

**Aufgabe 2** (6 Punkte)
Simuliere jeweils 1000 Realisierungen einer Gleichverteilung und einer Exponentialverteilung mit Erwartungswert 1 und Varianz 1. Plotte jeweils nebeneinander ein Histogramm und einen Boxplot der Daten. Füge dem Histogramm jeweils auch die Dichte der zugehörigen Verteilung hinzu. Beschrifte alle Plots mit einem geeigneten Titel.
Hinweis: Achte darauf, dass ein relatives Histogramm gezeichnet wird. Dies kann mit der Option 'freq' eingestellt werden (siehe auch ?plot)

**Aufgabe 3** (2,5+2,5 Punkte)

(a) Gib alle Zahlen zwischen 1 und 1000 aus, die durch 7 teilbar sind.

(b) Definiere eine Funktion, die für $x > 1$ den Wert $x^2$ zurückliefert und sonst -1. Plotte die Funktion auf dem Intervall $[-1, 5]$.

Hinweis: Soll eine selbst definierte Funktion geplottet werden, geht **R** davon aus, dass der Funktion ein Vektor $x$ als Argument übergeben wird. Daher sollte man für eine Fallunterscheidung von $x$ keine if-Abfrage verwenden. Besser ist es z.B. Vektoren der Art $(x <= c)$ zu betrachten. Diese haben die gleiche Länge wie $x$ und die einzelnen Einträge sind eins wenn der entsprechende Eintrag von $x$ kleiner gleich c ist und null sonst.

**Aufgabe 4** (5 Punkte)
Nach dem zentralen Grenzwertsatz ist bekannt, dass für $X_n \sim \text{Bin}(n, p)$ die Zufallsvariable

$$Y_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$

für $n \longrightarrow \infty$ in Verteilung gegen die Standardnormalverteilung konvergiert. Simuliere für $p = 0.7$ und $n = 10, 100, 1000$ und 10000 jeweils 10000 Realisierungen von $Y_n$ und plotte die Histogramme der simulierten Daten jeweils gemeinsam mit der Dichte der Standardnormalverteilung und einer geeigneten Beschriftung in ein gemeinsames Grafikfenster. Was ist zu beobachten?