Universität Konstanz
Fachbereich Mathematik und Statistik
Prof. Dr. J. Beran
WS 2007/08
29. Oktober 2007

# 2. Computerübung zur Mathematischen Statistik

**Aufgabe 1 (Einfache Lineare Regression)**

(a) Lesen Sie den Datensatz miete.sav in einen Data-Frame miete ein. Schauen Sie sich den Datensatz mit dem Befehl brush(miete) an.

(b) Stellen Sie für die stetigen Variablen den Zusammenhang zur Nettomiete pro Quadratmeter jeweils in einem Streudiagramm dar.

(c) Testen Sie, ob sich die Mittelwerte der Nettomiete pro Quadratmeter in den beiden Kategorien der binären Variablen unterscheiden. Hinweis: Verwenden Sie die Funktion t.test()

(d) Betrachten Sie bei den Mietspiegeldaten die Einfachregression

$$NMQM = \alpha + \beta * FLAECHE$$

Schätzen Sie die Koeffizienten der Regression und die Varianz.

(e) Verwenden Sie die erhaltenen Schätzungen $\hat{\alpha}$, $\hat{\beta}$ und $\hat{\sigma}^2$ um neue Daten nmqm.sim zu simulieren:
nmqm.sim$_i \approx N(\hat{\alpha} + \hat{\beta}*$ FLAECHE$_i$,$\hat{\sigma}^2)$

(f) Berechnen Sie anschließend die Regression

$$nmqm.sim = \alpha^* + \beta^* * FLAECHE$$

und vergleichen Sie die Regressionsergebnisse des simulierten Datensatzes mit denen der Originaldaten.

**Aufgabe 2 (Multiple Regression)** Verwenden Sie die Daten aus Aufgabe 1!

(a) Schätzen Sie ein multiples Regressionsmodell für die Nettomiete pro Quadratmeter (ohne die Variable MIETE)

(b) Untersuchen und interpretieren Sie die Ergebnisse mit den Funktionen summary() und plot(). Welche Variablen sind als Einflussgrößen sinnvoll?

**Aufgabe 3 (MSE Kriterium von Schätzern)**
Gegeben sei eine Stichprobe vom Umfang n einer Binomialverteilung B(n,p) mit unbekanntem Parameter $0 < p < 1$. Der Maximum-Likelihood-Schätzer ist gegeben durch $\hat{p}_{ML} := \overline{X}_n$. Als alternativen Schätzer betrachten wir

$$\hat{p}_a := \frac{\sum_{i=1}^n x_i + 1}{n+2}$$

(a) Ist $\hat{p}_a$ erwartungstreu?

(b) Bestimmen Sie die mittlere quadratische Abweichung (MSE) von $\hat{p}_a$.
[MSE($\hat{p}_a$)=Var($\hat{p}_a$) + Bias$^2$, $Bias = E(\hat{p}_a - p)$]

(c) Vergleichen Sie Ihr Ergebnis aus (b) mit der mittleren quadratischen Abweichung von $\hat{p}_{ML}$ für n=8 und n=40.Stellen Sie dazu den MSE als Funktion des wahren Parameters p dar.Was schließen Sie aus dem Vergleich?

Die Datensätze finden sie auf http://www.math.uni-konstanz.de/~beran/MathstatistikWS07.html.