Übung 8

1. In einer Studie über die Kontrolle des Infektionsrisikos in US-amerikanischen Spitälern wurden bei einer Zufallsstichprobe von 113 Spitälern folgende Variablen erfasst:

id	Identifikationsnummer des Spitals
length	mittlere Hospitalisationslänge aller PatientInnen (in Tagen)
age	mittleres Alter aller PatientInnen (in Jahren)
inf	mittleres Infektionsrisiko (in Prozent)
cult	Anzahl bakteriologischer Tests pro nichtsymptomatische PatientIn $\times$ 100
xray	Anzahl Röntgenuntersuchungen pro nichtsymptomatische PatientIn $\times$ 100
beds	Anzahl Betten
school	Universitätsklinik 1=ja 0=nein
region	geographische Region 1=NE 2=N 3=S 4=W
pat	mittlere Anzahl PatientInnen im Spital pro Tag
nurs	mittlere Anzahl vollbeschäftigter, ausgebildeter Krankenschwestern und Pfleger
serv	Prozentsatz angebotener Dienstleistungen aus insgesamt 35 möglichen

Die Daten befinden sich im File `/afs/ethz.ch/users/s/sfs/datasets/senic2.dat` (oder wie immer auf www).

a) Lesen Sie die Daten mit `read.table(...,header=T)` ein.
Definieren Sie die kategoriellen Variablen `school` und `region` als sogenannte Faktoren: z. B. `senic$school <- factor(senic$school)`.

b) Stellen Sie die einzelnen Variablen graphisch dar und beschreiben Sie sie quantitativ. Gibt es Ausreisser? schiefe Verteilungen? Betrachten Sie Streudiagramme von je zwei Variablen (`pairs()` und/oder `brush()` (`brush()` funktioniert nur in Splus!)).

c) Berechnen Sie für die Hospitalisationsdauer (Variable `length`) ein Modell, das alle erklärenden Variablen enthält.

d) Entfernen Sie die erklärende Variable mit dem grössten P-Wert (kleinstem Wert der t-Teststatistik) und berechnen Sie das reduzierte Modell. Fahren Sie so weiter, bis alle im Modell verbleibenden Variablen signifikant sind (P-Wert < 0.05).

e) Wenden Sie die S-Plus-Funktion `step()` auf das volle Modell in c) an. Erhalten Sie dasselbe Modell wie vorher in d)?

f) Berechnen Sie die PRESS–Statistik für die Modelle in d) und e) . Die Funktion `lmdiag()` liefert die quadrierten PRESS–Residuen. Für die PRESS–Statistik müssen Sie diese nur noch aufaddieren. Bitte sourcen Sie die Funktion neu (neue Version von lmdiag) mit
`source("/afs/ethz.ch/users/s/sfs/stat0/StatMeth/lmdiag.txt")`
bzw. holen Sie sich die neue Version der Funktion `lmdiag()` über das www (file `lmdiag.txt`)

g) Überprüfen Sie die Modellannahmen für die Modelle in d) und e).

h) Lassen Sie die einflussreichste Beobachtung und den Ausreisser weg und suchen Sie für die verbleibenden 111 Beobachtungen das "beste" Modell. Erhalten Sie das Gleiche wie vorher?

Abgabe: Montag, 10.01.2000