Übung 7

1. Das File `/afs/ethz.ch/users/s/sfs/datasets/bremsweg.dat` enthält Messungen des Bremswegs W (in feet) bei bestimmten Anfangsgeschwindigkeiten V (in miles/hour).

a) Zeichnen Sie in einem Streudiagramm W gegen V.

b) Betrachten Sie das Modell $W_i = \beta_1 v_i + \beta_2 v_i^2 + \epsilon_i$. Berechnen Sie die LS-Schätzungen für β_1, β_2 und σ^2. Wie gross ist R^2?
(> `lm(W ~ V + V^2 -1)` , "−1" = kein Achsenabschnitt)

c) Was zeigt die Residuenanalyse?

d) Rechnen Sie eine gewichtete Regression, um die nichtkonstante Streuung zu korrigieren ($Var(\epsilon_i) \propto v_i^2$). Benützen Sie dazu die Option `weights=`$\frac{1}{v^2}$ im Befehl `lm`.

e) Berechnen Sie auch eine Regression mit den Gewichten $\frac{1}{v}$ und vergleichen Sie die Schätzungen.

f) Rechnen Sie eine Regression für das transformierte Modell mit $Z_i := W_i/v_i$

2. Die Datei `/afs/ethz.ch/users/s/sfs/datasets/chicago.dat` enthält Daten zu Reisezeiten mit öffentlichen Verkehrsmitteln in Chicago. Jede der 33 Beobachtungen entspricht einer Route. Die zweite Spalte `x` gibt die erwartete Reisezeit an, zusammengesetzt aus Fussmarsch zur Busstation (Geschwindigkeit 3 miles/hour), Wartezeit und Fahrzeit nach Fahrplan. Die dritte Spalte `y` ist die mittlere Reisezeit von `n` (erste Spalte) befragten Passagieren.

a) Rechnen Sie eine einfache lineare Regression von y auf x (ohne n zu berücksichtigen). Zeigen die Residuenplot etwas Auffälliges?

b) Gibt es einflussreiche Beobachtungen und/oder Ausreisser? Benutzen Sie die Funktion `lmdiag()`, um leverage, standardisierte Residuen und Cook's distance zu berechnen. Betrachten Sie die entsprechenden graphischen Darstellungen. Mit
`source("/afs/ethz.ch/users/s/sfs/stat0/StatMeth/lmdiag.txt")`
machen Sie sich die Funktion `lmdiag()` verfügbar (das file `lmdiag.txt` steht auch unter www zur Verfügung). Als Funktionsargument geben Sie den Namen des Models an, das Sie mit `lm` berechnet haben.

c) Natürlich sollte n nicht vernachlässigt werden. Rechnen Sie eine gewichtete Regression, um die unterschiedlichen Varianzen in den Griff zu bekommen. Die entsprechende Option im `lm`-Befehl ist `weights=n`. Vergleichen Sie die Schätzungen und die "diagnostics".

d) Rechnen Sie eine gewichtete Regression ohne die nach c) einflussreichste Beobachtung. Was ändert sich in den Schätzungen und Plots?

e) Bei genauerer Überprüfung hat sich herausgestellt, dass Beobachtung Nr. 33 von einem übereifrigen ÖV-Angestellten erfunden worden ist. Entfernen Sie diese Beobachtung. Bei Beobachtung Nr. 2, 23 und derjenigen mit dem grössten Einfluss ist ein zusätzliches Umsteigen nötig. Erzeugen Sie eine neue Variable `ch`, die den Wert 1 annimmt für die drei obengenannten Fälle und 0 sonst. Rechnen Sie eine gewichtete Regression von y auf x und ch und vergleichen Sie die Resultate.

Abgabe: Montag, 20.12.97