

Angewandte Regression — Serie 4

1. Der Datensatz `asphalt` enthält Daten zur Abnutzung von Strassenbelägen. In einem Experiment wurden 31 verschiedene Strassenbeläge gemischt. Als Zielvariable wurde die Tiefe der Radspur in inches gemessen, nachdem eine Million Räder über den Belag gefahren waren. Die Variablen sind:

RUT	Abnutzung des Belags in inches pro 1 Mio. Räder
VISC	Viskosität des Asphalts
ASPH	Anteil des Asphalts im Oberflächenbelag (in %)
BASE	Anteil des Asphalts im Unterbelag (in %)
FINES	Anteil der Feinteile im Oberflächenbelag (in %)
VOIDS	Anteil der Hohlräume im Oberflächenbelag (in %)
RUN	Indikatorvariable, welche die zwei Versuchsreihen unterscheidet

Quelle: R.V. Hogg and J. Ledolter, Applied Statistics for Engineers and Physical Scientists, Maxwell Macmillan International Editions, 1992, p.393

- a) Betrachten Sie die Streudiagramme von jeder Variablen gegen jede. Was fällt Ihnen auf? Welche Variablen sollten transformiert werden? Benützen Sie für die folgenden Teilaufgaben die transformierten Variablen.
- R-Hinweis:** Um festzustellen, ob die zwei Versuchsreihen sich unterscheiden, kann man die Punkte entsprechend dem RUN einfärben, indem `col=d.asphalt$RUN+2` gesetzt wird.
- b) Rechnen Sie eine Regression mit dem vollen Modell ohne Interaktionen. Geben Sie die geschätzte Standardabweichung der Fehler und R^2 an.
- c) Lassen Sie die nicht signifikanten Variablen im Modell weg und machen Sie eine neue Regression. Darf man alle signifikanten Variablen auf's mal weg lassen, oder nur eine Variable pro Modell-Schritt? Begründen Sie Ihre Wahl. Schreiben Sie Ihr finales Modell auf und tabellieren Sie die geschätzte Standardabweichung der Fehler und R^2 Ihrer Modelle.

R-Hinweis:

Die von Werner Stahel geschriebenen Funktionen `regr`, `g.res2x` und `plot.regr` können Sie mit dem Befehl

- `library(regr0)` in die R-Umgebung einbeziehen,
- `install.packages("regr0",repos="http://R-Forge.R-project.org")` zu Hause auf Ihrem Computer in Ihre R-Umgebung installieren.

- 2.** Eine Insektenkundlerin untersuchte den Einfluss verschiedener Faktoren auf die Cholinestrase-Aktivität in Grilleneiern.

Den Datensatz finden Sie in `cricket.dat`. Er enthält die folgenden Angaben:

- y Cholinestrase-Aktivität der Eier, in Prozent des normalen
Achtung: hier ist Prozent als Normierung gemeint und nicht als Anteil
- x_1 Alter der Eier in Tagen zur Zeit der Behandlung
- x_2 Alter der Eier in Tagen zur Zeit der Beobachtung
- x_3 Insektizid: 0=Carbaryl, 1=Propoxur
- x_4 Dosis der Behandlung in Mikrogramm pro Ei

- a) Betrachten Sie die zweidimensionalen Streudiagramme.
- b) Passen Sie ein Regressionsmodell der folgenden Form an

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 + \epsilon$$

und machen Sie eine Residuenanalyse.

- c) Untersuchen Sie, ob durch Transformationen (neben Logarithmus, Wurzel gibt es auch Quadrat) der erklärenden Variablen respektive der Zielvariable das Modell verbessert werden kann. Machen Sie eine Residuenanalyse.
- d) Überlegen Sie, ob durch das Hinzufügen von quadratischen Termen das Modell verbessert werden kann. Nehmen Sie nur jene quadratischen Terme ins Modell auf, die signifikant sind. Machen Sie eine Residuenanalyse.
- e) Untersuchen Sie, ob durch das Hinzufügen von Interaktionstermen das Modell weiter verbessert werden kann. Machen Sie eine Residuenanalyse.
- f) Was passiert, wenn man den Zeitraum zwischen Behandlung und Beobachtung ins Modell einbezieht? Machen Sie eine Residuenanalyse.