Regression I — Kontrollaufgaben

Name:

Vorname:

Falls Sie etwas nicht verstehen oder Probleme mit R haben, fragen Sie uns, bevor Sie zuviel Zeit verlieren.

Es sollen keine Computeroutputs abgegeben werden. Notieren Sie bitte Ihre Lösungen auf eigene Blätter.

Zur Erinnerung: Die Funktionen `regr` usw. aus dem Skript können folgendermassen geladen werden:

```
> source("ftp://stat.ethz.ch/NDK/Source-NDK-9/R/regr.R")
```

1. Treibstoffverbrauch (8 Punkte)

Der Datensatz `auto.dat` enthält den Verbrauch und das Gewicht von verschiedenen Autotypen aus dem Jahre 1989. Es soll eine Regression der Zielgrösse Verbrauch auf die Ausgangsgrösse Gewicht durchgeführt werden.

Den Datensatz können Sie mit

```
> t.url <- "http://stat.ethz.ch/Teaching/Datasets/NDK/auto.dat"
> d.auto <- read.table(t.url, header=TRUE)
```

einlesen.

a) Führen Sie die einfache lineare Regression mit den richtigen Transformationen durch, mit der Zielgrösse `Verbrauch` und der Ausgangsgrösse `Gewicht`! Geben Sie die Schätzwerte für die Parameter β_0, β_1 und σ an!

b) Was bedeutet der gefundene Zusammenhang für die untransformierten Daten? Schreiben Sie die Gleichung dafür auf! Setzen Sie für die Koeffizienten die in a) erhaltenen Schätzungen ein!

c) Geben Sie ein 95-% Vertrauensintervall für den Parameter β_1 an!

d) Im Steudiagramm der untransformierten Daten scheint der Verbrauch proportional vom Gewicht abzuhängen. Falls das stimmen sollte, was kann man daraus für den Parameter β_1 ableiten? Ist diese Aussage mit den Daten auf dem 95%-Niveau verträglich?

2. Modellsuche (10 Punkte)

Der künstlich erzeugte Datensatz `newdata.dat` enthält 50 Beobachtungen der Variablen y (Zielgrösse) und x_1, x_2, x_3, x_4 (Ausgangsgrössen). Die Daten wurden simuliert, beruhen also auf einem bekannten Modell. Ziel dieser Aufgabe ist es, das wahre (ist meistens relativ schwierig) oder ein ähnlich gut zu den Daten passendes Modell zu finden.

Dabei müssen Sie vielleicht Variable transformieren, spezielle Punkte finden und unnötige Variable aus dem Modell nehmen.

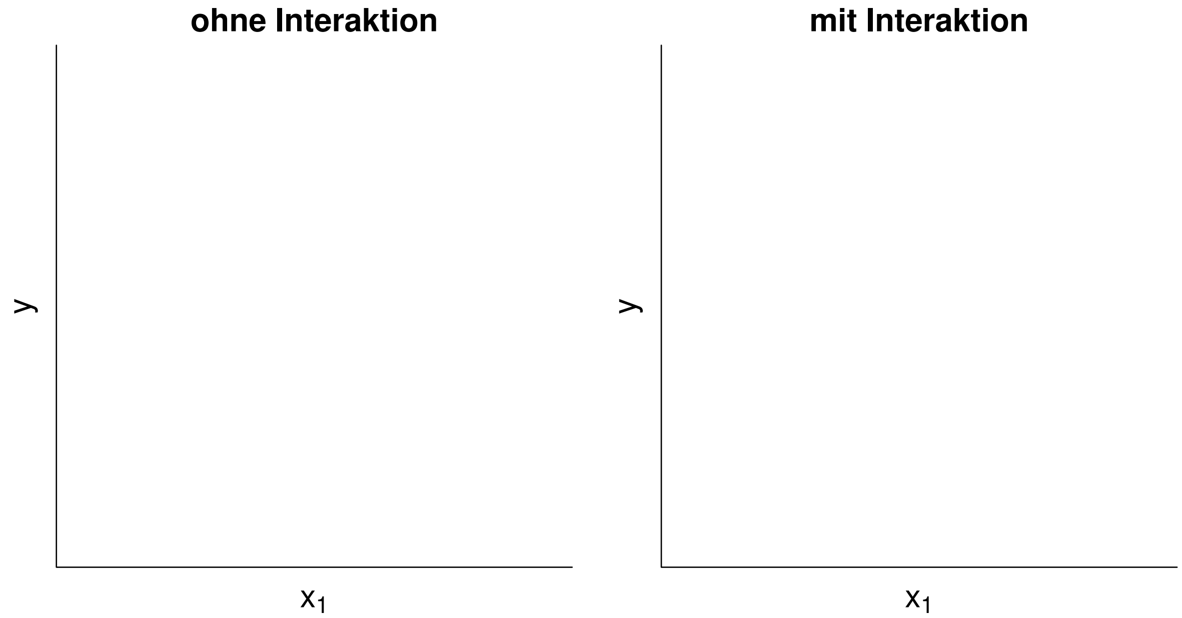
Schreiben Sie möglichst genau auf, weshalb Sie die vorgenommenen Schritte durchgeführt haben. Zeichnen Sie Skizzen von interessanten Plots und geben Sie Kommentare dazu ab!

Hinweise:

- Benutzen Sie die Funktionen `regr` und `plot("regr-Objekt")`, die sehr nützliche Plots für die Modellsuche liefert.
- Prüfen Sie am Schluss, ob Ihr gefundenes Modell keine Voraussetzungen mehr verletzt.

3. Gemischte Aufgaben (5 Punkte)

a) Gegeben sei eine Zielvariable z und eine erklärende Variable x_1, welche beide kontinuierlich sind. Weiter gebe es eine nominale Grösse x_2 mit drei Stufen, d.h. x_2 kann nur drei verschiedene Werte annehmen (z.B. 3 Klassen). Veranschaulichen Sie graphisch den Zusammenhang zwischen y und x_1, x_2 bei einem linearen Modell ohne Interaktion zwischen x_1 und x_2 und bei einem linearen Modell mit Interaktion!



b) Welche Bedeutung hat ein Vorhersageintervall?

c) Weshalb bezeichnen Mathematiker die lineare Regression als linear, obwohl man mit Transformation auch nicht-lineare Zusammenhänge modellieren kann?
Wieso ist für Mathematiker diese Art von Linearität wichtig?