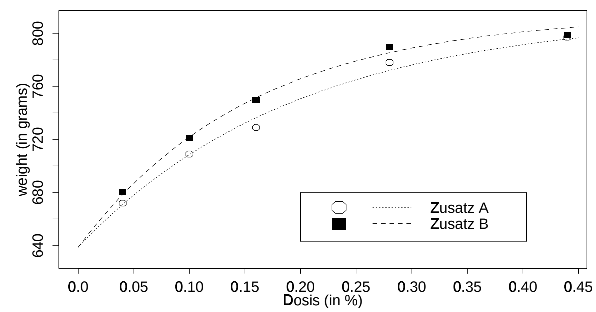
Serie 6

1. Beweise die Formeln für $\tilde{\theta}^{(-k)}(\theta_k)$ und $\tilde{S}_k(\theta_k)$, die im Skript angegeben sind (p. 47), im Fall des linearen Modells. Hinweis: Verwende einen Lagrangemultiplikator.

2. Die Daten, welche im Datensatz `body.dat` gegeben sind, stammen aus einem Wachstumsexperiment, in dem neugeborene Truthennen mit zwei verschiedenen Methionin-Nahrungsmittelzusätzen gefüttert wurden (siehe Noll et al, 1984). Die Truthennen wurden in 10 Versuchseinheiten zu je 15 Tiere unterteilt. Je 5 Versuchseinheiten erhielten den gleichen Methionin-Nahrungsmittelzusatz, deren Dosen zwischen 0.04 und 0.44% der totalen Nahrungsmittelaufnahme liegen. Die Zielvariable ist das mittlere Körpergewicht der Truthennen in einer Versuchseinheit.



Das Wachstum der Truthennen, Y_i (Variable `weight`), in Abhängigkeit der Dosis $x_{i,k}$ für die zwei verschiedenen Methionin-Nahrungsmittelzusätze, $k = 1, 2$, wird mit der nichtlinearen Funktion

$$Y_i = f(x_{i,1}, x_{i,2}) = \theta_1 + \theta_2\,(1 - \exp(\theta_3(\theta_4 x_{i,1} + x_{i,2}))) + \epsilon_i$$

beschrieben ($x_{i,1}$ = Dosis aus Zusatz A (Variable `sourceA`); $x_{i,2}$ = Dosis aus Zusatz B (Variable `sourceB`)).

a) Interpretiere dieses Modell: Bedeutung der Parameter? Übersetze die Nullhypothese "Kein Wachstumsunteschied zwischen den Gruppen".

b) Gib die Schätzwerte für die Parameter $\theta_1, \ldots, \theta_4$ an, unter Verwendung der Startwerte $\theta_1 = 640$, $\theta_2 = 160$, $\theta_3 = -7.2$, $\theta_4 = 0.90$.

c) Teste die in a) beschriebene Nullhypothese.

d) Zeichne und kommentiere die Profil-t-Funktionen und die Likelihood-Profilspuren.

R-Anleitung:

```
> truthennen <- read.table.url("....../body.dat",header=T)
> library(nls)                                              # Library "nls" holen
> truthennen.nls <- nls(weight~T1+T2*(1-exp(T3*(T4*sourceA + sourceB))),
                   data=truthennen,start=list(T1=640,T2=160,T3=-7.2,T4=0.9))
                                                            # Nichtlineare Regression
> summary(...)
> qt(0.975,6)                                               # gibt das t_{0.975,6}-Quantil
```

Auf den Studentenkontos:

```
> library(SfS)
```

Studenten, die zu Hause arbeiten:

Unter "http://stat.ethz.ch/Teaching/kuensch/Regression" kann das File 'p.profileTraces.R' heruntergeladen werden. Dieses File abspeichern (Endung .R ist wichtig!) und anschliessend im R schreiben:

```
> source(``/`Pfadname des Files'/p.profileTraces.R'')
```

Dann (für alle):

```
> truthennen.profile <- profile(truthennen.nls)
> p.profileTraces(truthennen.profile)            # Zeichnet die Profile-t-Funktionen und die
Likelihood-Profilspuren
```

3. Der Datensatz `heart.dat` enthält die nach Alter sortierten Daten von 99 Personen. Für jede Altersgruppe ist die totale Anzahl Personen gegeben (n_i) und die Anzahl Personen mit Symptomen einer Herzkrankheit (y_i).

a) Schätzen Sie die Parameter einer einfachen logistischen Regression, welche die Wahrscheinlichkeit, Symptome zu zeigen, mit dem Alter in Beziehung setzt. Testen Sie die Hypothese, dass das Alter (`age`) die Wahrscheinlichkeit, Symptome zu zeigen, beeinflusst.

R-Anleitung:

```
heart.glm <- glm(cbind(y,n-y) ~ age, family=binomial, data=heart)
```

Falls ($Y_i \sim \mathcal{B}(n_i, \pi_i)$) ist mit $n_i > 1$, wird die Zielvariable als Matrix (mit 2 Spalten) eingegeben, wobei in der ersten Spalte die Anzahl "Erfolge" (Y_i) und in der zweiten Spalte die Anzahl "Misserfolge" ($n_i - Y_i$) stehen.

b) Zeichnen Sie die logistische Regressions-Kurve und schätzen Sie das Alter, bei welchem Sie erwarten würden, dass 10%, 20%, ..., 90% der Personen Symptome zeigen. Diskutieren Sie diese Resultate.

R-Anleitung:

```
age.neu <- 0:100
heart.pred <- predict(heart.glm, newdata=data.frame(age=age.neu),
                      type="response")     # π̂_i an der Stelle age.neu_i
plot(heart$age, heart$y/heart$n, xlim=c(0,100), ylim=c(0,1))
lines(age.neu, heart.pred)
```

4. In dieser Studie wurden die Überlebenszeiten von 66 Patienten nach einer Operation beobachtet. Der Datensatz `diabetes.dat` enthält für jeden Patienten die Variablen

Variablenname	Bedeutung
`lzeit`	**Überlebenszeit in Tagen nach Operation**
`cens`	0=(Studie überlebt, zensuriert), 1=(gestorben),
`sex`	Geschlecht (m=0, w=1)
`diab`	Diabetiker (ja=1, nein=0)
`alter`	Alter in Jahren

a) Passe das Cox-Modell

$$\lambda_i(t) = \lambda_0(t) \cdot \exp(\beta_1 \texttt{sex}_i + \beta_2 \texttt{diab}_i + \beta_3 \texttt{alter}_i)$$

an und kommentiere die geschätzten Koeffizienten.

b) Führe die zusätzliche Variable `alter·diab` ein. Erkläre, was dies bedeutet. Schätze die Koeffizienten in diesem Modell und kommentiere die Resultate.

R-Anleitung:

```
> diabetes <- read.table.url(``...../diabetes.dat'',header=TRUE)
> library(survival5)
> diabetes.cox <- coxph(Surv(lzeit,cens)~ sex + diab + alter, data=diabetes)
```

Vorbesprechung : Freitag 23.6. 13.15 im HG D 1.1.

Abgabe: Mittwoch 28. Juni 2000 vor der Vorlesung.

Präsenz: Jeweils Donnerstag, 12.00 bis 13.00 Uhr im LEO C12.1, Leonhardstr. 27, oder nach Vereinbarung: Marcel Wolbers (`wolbers@stat.math.ethz.ch`), LEO C14, Tel. 632 22 52 und Isabelle Flückiger (`isabelle@stat.math.ethz.ch`), LEO C13, Tel. 632 42 76.