Serie 1

1. Sei $(Y, X_1, \ldots, X_p)$ ein $(p+1)$-dimensionaler Zufallsvektor. Wir suchen die Koeffizienten $\theta_0, \theta_1, \ldots, \theta_p$ derart, dass

$$\mathcal{E}[(Y - \theta_0 - \sum_{j=1}^{p} \theta_j X_j)^2]$$

minimal ist ($\mathcal{E}[Y^2] < \infty$, $\mathcal{E}[X_j^2] < \infty$ vorausgesetzt).

a) Stelle das zugehörige Gleichungssystem auf (das Analogon zu den Normalengleichungen).

b) Zeige, dass $\theta_0^{opt} = \mathcal{E}[Y] - \sum_{j=1}^{p} \theta_j^{opt} \mathcal{E}[X_j]$ ist und stelle ein Gleichungssystem für $\theta_1, \ldots, \theta_p$ auf.

2. Betrachte die zweidimensionale Normalverteilung $\mathcal{N}(\mu = \binom{\mu_x}{\mu_y}, \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix})$.

a) Zeige, dass

$$(x - \mu_x, y - \mu_y)\Sigma^{-1}\begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} = (y - \mu_y - \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x))^2 \frac{1}{\sigma_y^2(1-\rho^2)} + \frac{(x-\mu_x)^2}{\sigma_x^2}$$

gilt.

b) Es sei (X, Y) zweidimensional normalverteilt. Schliesse aus a), dass die Randverteilung von X eine Normalverteilung ist.

c) Berechne $\lim_{h \downarrow 0} P[Y \leq y | x \leq X \leq x + h]$.

3. Linearisierbare Funktionen

Die folgende Tabelle enthält einige Funktionen, die sich durch Transformation linearisieren lassen. Vervollständige die Tabelle, indem Du die zu verwendende Transformation und die lineare Form einträgst.

Funktion	Transformation	Lineare Form
$y = \alpha x^{\beta}$	$y' = \log(y),\ x' = \log(x)$	$y' = \log(\alpha) + \beta \cdot x'$
$y = \alpha e^{\beta \cdot x}$	....	....
$y = \alpha + \beta \cdot \log(x)$	....	....
$y = x/(\alpha \cdot x - \beta)$	....	....
$y = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta \cdot x}}$	....	....
$y = \alpha e^{\beta/x}$	....	....
$y = 1/(\alpha + \beta e^{-x})$	....	....

4. In dieser Aufgabe betrachten wir 4 Datensätze, die von Anscombe konstruiert wurden. In jedem der Datensätze gibt es eine Zielvariable Y und eine unabhängige Variable X.
Stelle jeden der 4 Datensätze als Streudiagramm dar, zeichne die Regressionsgerade ein und kommentiere die Ergebnisse.

'R'-Hinweis: Datensatz einlesen:

```
anscombe <- read.table.url(
  "http://stat.ethz.ch/Teaching/Datasets/anscombe.dat", header=T)
```

Wenn das zu Hause nicht geht: Datensatz speichern und dann mit

```
anscome <- read.table("filename",header=T)
```

einlesen. Die x/y-Werte sind `X1.3` für `Y1,Y2,Y3`, und `X4,Y4`.
Die Regression kann man mit

```
ans1.lm <- lm(Y1~X1.3,data=anscombe)
```
oder
```
ans1.lm <- lm(anscombe$Y1~anscombe$X1.3)
summary(ans1.lm)
```

berechnen und numerisch auswerten Mit `par(mfrow=c(2,2))` wird das Grafikfenster so eingeteilt, dass alle 4 Bilder nebeneinander passen.
Den Scatterplot und die Regressionsgerade erhält man mit

```
plot(anscombe$X1.3,anscombe$Y1)
abline(ans1.lm)
```

5. Der Datensatz `airline` enthält die monatliche Anzahl abfliegende Flugpassagiere in den USA für die Jahre 1949-1960.

a) Plotte die Daten gegen die Zeit und beschreibe in Worten, welche Phänomene Du siehst.

b) Plotte als nächstes die logarithmierten Daten gegen die Zeit und kommentiere die Unterschiede.

c) Überlege, wie man die Funktionen f_j der Zeit wählen könnte, damit ein lineares Modell der Form

$$\log(y_t) = \sum_{j=1}^{p} \theta_j f_j(t) + \epsilon_t$$

plausibel ist.

d) Passe das in c) gewählte Modell an und plotte die angepassten Werte und die Residuen gegen die Zeit. Kommentar?

'R'-Hinweis: Datensatz einlesen:

```
airline <- scan.url("http://stat.ethz.ch/Teaching/Datasets/airline.dat")
```

Die Plots werden schöner, falls `airline` eine Zeitreihe ist. Dies ist aber nicht unbedingt nötig:

```
library(ts); airline <- ts(airline,start=c(1949,1),freq=12)
```

Die Regression mittels

```
reg <- lm(airline ~ f1+...+fp-1)
```

anpassen, wobei `fj` ein Vektor der Länge 144 ist, der die Werte $f_j(t)$ enthält. Die `-1` bedeutet, dass kein Achsenabschnitt ins Modell genommen wird. Zur Erzeugung der `fj` ist ev. der Befehl `rep` nützlich. Schau Dir an, was z.B. `c(1,rep(0,11))` liefert. Die angepassten Werte und die Residuen erhält man mittels `fitted(reg)` und `resid(reg)`.

Vorbesprechung : Freitag 7.4. 13.15 im HG D 1.1, mit anschliessender 'R'-Einführung im HG E26.1.

Abgabe: Mittwoch, 26. April 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.