Serie 2

1. Bei einer Untersuchung der Beziehung zwischen dem Verkehrsfluss x (in 1000 Autos pro 24 Stunden) und dem Bleigehalt y ($\mu g/(g$ Trockengewicht)) von Baumrinden in der Nähe der Schnellstrassen wurden folgende Daten erhoben:

x_i	8.3	8.3	12.1	12.1	17.0	17.0	17.0	24.3	24.3	24.3	33.6
y_i	227	312	362	521	640	539	728	945	738	759	1263

Die Daten stehen im Datensatz `strasse.dat` mit der erklärenden Variable `verkehr` und mit der Zielvariable `blei` zur Verfügung.

a) Zeichnen Sie ein Streudiagramm dieser beiden Grössen.

b) Berechnen Sie die Regressionsgerade `blei` gegen `verkehr`, und zeichnen Sie diese ins obige Streudiagramm ein. Scheint diese Lösung sinnvoll zu sein? Sind irgendwelche Voraussetzungen für die Verwendung der KQ (Methode der Kleinsten Quadrate) verletzt?

c) Testen Sie einerseits die Nullhypothese $\beta = 0$ gegen die Alternative $\beta \neq 0$ und andererseits die Nullhypothese $\beta = 0$ gegen die Alternative $\beta > 0$, jeweils auf dem 1%- und 5%-Niveau. Geben Sie die zugehörigen p-Werte an. Geben Sie an, inwiefern sich die beiden Tests in der Interpretation unterscheiden.

d) Wie gross ist $\hat{\alpha}$? Testen Sie $H_o : \alpha \geq 0$ gegen die Alternative $H_A : \alpha < 0$. Was würde ein signifikantes Testergebnis bedeuten?

e) Geben Sie für $\hat{\alpha}$ resp. $\hat{\beta}$ ein zweiseitiges 95% Konfidenzintervall an. Gibt es einen Zusammenhang zu den obigen Tests?

f) Es existieren die folgenden drei Expertenmeinungen: Pro 1000 Autos nimmt der Bleigehalt um a) 25, b) 40, und c) 50 $\mu g/g$ zu. Beziehen Sie Stellung!

R-Anleitung (unvollständig!):

Datensatz einlesen:
```
strasse <- read.table.url(
        "http://stat.ethz.ch/Teaching/Datasets/strasse.dat", header=T)
> x11()
> plot(strasse[,"verkehr"], strasse[,"blei"], main="Streudiagramm")
> strasse.lm <- lm(blei ~ verkehr, data=strasse)     # Anpassen eines linearen Modelles
> abline(coef(strasse.lm))                            # Gerade in das Diagramm einzeichnen
> summary(strasse.lm)                                 # Zusammenfassung der Anpassung
> qt(α,n)                                             # Gibt den Wert des t_{n,α}-Quantils
```

Anhang: Lesen eines Computeroutputs zur Regression von R.

Mit dem Befehl `summary(strasse.lm)` erhält man

Coefficients	Estimate	Std. Error	t value	Pr(> $\|t\|$)	Signif
(Intercept)	$\widehat{\alpha}$	$\widehat{\sigma}(\widehat{\alpha})$	$\widehat{\alpha}/\widehat{\sigma}(\widehat{\alpha})$	P-Wert	
X	$\widehat{\beta}$	$\widehat{\sigma}(\widehat{\beta})$	$\widehat{\beta}/\widehat{\sigma}(\widehat{\beta})$	P-Wert	

`Residual standard error:` $\widehat{\sigma}$ `on` $n-2$ `degrees of freedom`
`Multiple R-Squared:` R^2
`F-statistic: ...`

Bemerkungen:

i) t `value` liefert den Wert der Teststatistik T für $\alpha_0 = 0$ bzw. $\beta_0 = 0$.

ii) Pr(> $|t|$) liefert den P-Wert des t-Test $H_o : \alpha = 0$ gegen $H_A : \alpha \neq 0$ bzw. $H_o : \beta = 0$ gegen $H_A : \beta \neq 0$.

2. In Indien behindern basische Böden, also tiefe Säurewerte oder hohe pH-Werte, Pflanzen beim Wachstum. Es werden daher Baumarten gesucht, die eine hohe Toleranz gegen solche Umweltbedingungen haben. In einem Freilandversuch wurden auf einem Feld mit grossen lokalen Schwankungen des pH-Wertes 120 Bäume gepflanzt und ihre Höhe nach 3 Jahren gemessen. Zusätzlich wurde eine Variable `l.sar` = log(SAR), (SAR=sodium absorption ratio), gemessen, die einen etwas anderen Aspekt der "Basizität" erfasst.

Die Daten stehen im Datasatz `basisch.dat` zur Verfügung.

Die quadrierte Höhe (Variable `h.quad`) wird als Zielvariable betrachtet. Die beiden erklärenden Variablen sind `ph` und `l.sar`.

Es soll untersucht werden, ob die Messung der zusätzlichen Grösse SAR „etwas bringt".

a) Erstellen Sie die Streudiagramm-Matrix.
R-Anleitung (unvollständig!): `> ?pairs` # *Dokumentation des Befehls pairs*

b) Berechnen Sie das multiple Regressionsmodell.
R-Anleitung: Das allgemeine lineare Modell schreibt man mittels
`lm` $(Y \sim x_1 + x_2 + \ldots + x_p,$ `data=` $\ldots)$.

c) Testen Sie, ob der Koeffizient von `l.sar` sich signifikant von 0 unterscheidet!

d) Berechnen Sie ein 95%−Vertrauensintervall für die erwartete quadrierte Höhe eines Baumes auf einem Boden mit `ph=8` und `l.sar` = 1, sowie ein 95%−Prognoseintervall für die quadrierte Höhe auf diesem Boden. Was würden Sie tun, wenn Sie an der Höhe und nicht an der quadrierten Höhe interessiert sind?
R-Anleitung (unvollständig!):

```
> ?predict.lm                                # Dokumentation des Befehles predict.lm
> new <- data.frame(ph=8, l.sar=1)
  ⋮
```

3. Betrachten Sie die folgenden drei Modelle:

$$\begin{aligned} x_i &= \alpha_1 + \beta_1 u_i + \varepsilon_i, \quad (i = 1, \ldots, l); \\ y_j &= \alpha_2 + \beta_2 v_j + \eta_j, \quad (j = 1, \ldots, m); \\ z_k &= \alpha_3 + \beta_3 w_k + \tau_k, \quad (k = 1, \ldots, n). \end{aligned}$$

Die ε_i's, die η_j's und die τ_k's sind alle unabhängig voneinander normalverteilt mit Erwartungswert 0 und gleicher Varianz σ^2. Wir wollen einen Test für die Nullhypothese $H_o : \beta_1 = \beta_2 = \beta_3$ (die drei Geraden sind parallel) konstruieren.
Stellen Sie die Modellgleichungen für das volle Modell und für das reduzierte Modell auf und geben Sie die Nebenbedingungen an, die vom vollen zum reduzierten Modell führen.

4. Wir betrachten das folgende Regressionsmodell:

$$y_i = \theta \cdot f(i) + \epsilon_i \qquad (i = 1, 2, \ldots)$$

mit ϵ_i i.i.d. , $E[\epsilon_i] = 0$, $E[\epsilon_i^2] = \sigma^2 < \infty$ und $E[\epsilon_i^4] < \infty$ und einer festen bekannten Funktion $f : \mathbb{R}^+ \to \mathbb{R}$.
Prüfe nach, ob die Bedingungen im Skript für Konsistenz und asymptotische Normalität des Kleinste-Quadrate-Schätzers $\widehat{\theta}_n$ erfüllt sind in den beiden folgenden Fällen

a) $f(x) = x^\gamma$ ($\gamma \in \mathbb{R}$, beliebig, aber fest)

b) $f(x) = 2^x$.

Vorbesprechung : Freitag 28.4. 13.15 im HG D 1.1.

Abgabe: Mittwoch, 10. Mai 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.