Serie 5

1. Mit `read.table("/home/sfsadm/ueb.datasets/indi.dat") -> d.indi` liest man die Kindersterblichkeitsrate (IMR) und einen Index für den Lebensstandard (PQLI) von 13 ausgewählten indischen Staaten.

Wir wollen die Sterblichkeit in Abhängigkeit des Lebensstandards modellieren.

a) Erstelle ein Streudiagramm von IMR vs. PQLI und kommentiere!

b) Führe je eine lineare Regression durch, einmal mit den Daten vom Staat Kerala und einmal ohne! Zeichne die beiden Regressionsgeraden in das Streudiagramm von IMR vs. PQLI!

c) Diskutiere die Ergebnisse:
- warum ist $\hat{\sigma}^2$ etwas kleiner, wenn der Staat Kerala einbezogen wird?
- warum nimmt die Varianz von $\hat{\beta}$ stark ab, wenn der Staat Kerala einbezogen wird?
Die Varianz von $\hat{\beta}$ wird geschätzt, indem für σ die Schätzung $\hat{\sigma}$ verwendet wird.

Hinweise:
a) Benütze `n.plot(..)`
b) Benütze `abline(..)` für das Zeichnen der Regressionsgeraden. Wenn mit `lm13 <- lm(IMR ~ PQLI)` die volle Regression berechnet wurde, kann man mit `lm12 <- update(lm13, subset = -13)` dasselbe Modell ohne die 13. Beobachtung anpassen.
c) Benütze `summary`(*Resultat von* `lm`).

Es sollen ausser (kommentierten) Graphik-Plots keine Computeroutputs abgegeben werden; schreibe die wenigen verlangten Resultate ab!

2. Die Bau-Daten von 32 Leicht-Wasser-Reaktoren (LWR) sind als Data-Frame `d.nuke` gegeben. Die Frage ist, welche Faktoren die Endkosten *wie* beeinflussen. Erklärungen zu den Variablen befinden sich auf der Datei `/home/sfsadm/ueb.datasets/nuke.doc` (z.B. mit "File-Manager" oder "Text-Edit" anschauen!)

a) Berechne ein volles Modell, d.h. die Kosten `C` in Abhängigkeit aller (10) übrigen Variablen (Abkürzung in S-plus: "." für *"alle übrigen Variabeln"*, also `lm(C ~ . , data = d.nuke)`). Diskutiere den Plot der Residuen gegen den Fit ("Tukey-Anscombe").

b) Logarithmiere die Variablen `C`, `N`, `S`, `T1`, `T2` und vergleiche mit a).

c) **freiwillig**: Teste die Hypothese *"Die Kosten hängen nicht von* `log(T1)`, `log(T2)`, `PR` *und* `BW` *ab"* gegen das alternative Modell in b).
Hinweis: Siehe Stahel-Skript 13.7.e & j
Berechne zuerst das Modell gemäss der Nullhypothese. Die Funktion `anova`(name des fits) (auch `summary.aov(..)`) liefert die Grössen, aus denen man die Teststatistik berechnen kann. Die Quantile der F-Verteilung sind mit `qf(`*p, df1, df2*`)` erhältlich.

Abgabe: Mo, 9. Dez. 96.
Hilfe auch via e-mail: To: `maechler@stat.math.ethz.ch`
oder sonst: Tel.: (63) 2-3408