

## Angewandte Statistik für Biometrie

(Abgabe: Di., 18.05.2010, 13:15 Uhr, vor den Übungen)

1. Löse Aufgabe 2b) von Blatt 3.

(4 Punkte)

2. Lade den Datensatz `airline.dat` von der Veranstaltungshomepage herunter. Dieser Datensatz enthält die monatliche Anzahl an Passagieren einer gewissen Fluggesellschaft über einen bestimmten Zeitraum. Ziel dieser Aufgabe ist es eine geeignete Modellierung mittels linearer Regression für diesen Datensatz zu finden. Im Idealfall sollten die Residuen  $\varepsilon_t \sim N(0, \sigma^2)$  iid Zufallsvariablen sein. Führe wiederholt folgende Schritte durch, um ein geeignetes Modell zu finden.

- Passe ein lineares Modell an.
- Plote die Residuen und untersuche, ob der Plot auf  $N(0, \sigma^2)$ -verteilte Zufallsvariablen hindeutet.

In Schritt 1 kann man zunächst eine polynomiale Regression versuchen, z.B.  $X_i = \beta_0 + \beta_1 t + \dots + \beta_d t^d$ . Hier kommt es auf die richtige Wahl von  $d$  an. Versuche mit Hilfe des Bestimmtheitsmaßes eine sinnvolle Wahl zu treffen. Offensichtlich enthalten die Daten eine gewisse Periodizität. Untersuche die Residuen des angepassten Modells auf Periodizitäten und versuche dein Modell mittels Termen der Form  $\alpha_1 \sin(\gamma_1 t - \gamma_2)$  und  $\alpha_2 \cos(\gamma_3 t - \gamma_4)$  zu erweitern. Welche Periodenlänge bietet sich hier an?

*Hinweis:* Eventuell ist eine `for-Schleife` hilfreich. Informationen dazu bekommt man mit dem Befehl `help('for')`.

(12 Punkte)

<http://www.uni-ulm.de/mawi/zawa/lehre/sommer2010/asb2010.html>