

CAS Datenanalyse

Probeprüfung Zeitreihen und Prognosen

Name:	Punkte:	Note:
-------	---------	-------

Bewertung

Geben sie bei quantitativen Fragen immer alle benutzten Formeln an, bevor sie die Zahlen einsetzen und Schlussresultate berechnen. Erfragte Begründungen müssen in ganzen Sätzen verständlich ausformuliert sein. Fehlt eine Begründung, bzw. ein klar ersichtlicher Lösungsweg, so wird auch bei korrekten Zahlenresultaten höchstens die Hälfte des Punktemaximums vergeben.

Erlaubte Hilfsmittel:

open book“, d.h. beliebige schriftliche Hilfsmittel sind erlaubt. Taschenrechner sind ebenfalls zugelassen. Ein Laptop darf zur Bearbeitung der Aufgaben eingesetzt werden. Auf dem Computer darf R-Studio und ein PDF-Reader betrieben werden. Bereits bestehende R-Files dürfen eingesetzt werden.

Viel Erfolg

Wir betrachten in dieser Aufgabe die **mittlere Milchleistung pro Kuh in den USA** zwischen **1994 und 2005**. Die Daten befinden sich im File **milk_data.rda**. Es handelt sich um Monatsdaten die in einem Data Frame vorliegen, vorhanden sind 12 komplette Jahre ab Januar 1994. Im Data Frame sind die folgenden Variablen enthalten:

milk: mittlere Milchleistung pro Kuh
zeitschritte: Zeitschritte
monat: Monate

- a) (1 Punkt) Definieren Sie die **Spalte milk** gemäss den Angaben von oben in R sinnvoll und korrekt **als Zeitreihe**. Geben Sie den verwendeten R-Befehl an.
- b) (2 Punkte) Schauen Sie sich die Zeitreihe in R an. Ist die Zeitreihe **stationär**? Begründen Sie ihre Antwort.
- c) Führen Sie eine **STL-Zerlegung der Zeitreihe** durch, wobei Sie annehmen können, dass die Saisonkomponente konstant ist.
 - i) (1 Punkt) **Welche der drei Komponenten der STL-Zerlegung dominiert die Struktur der Zeitreihe?**
 - ii) (1 Punkt) Ist eine **log-Transformation der Daten angebracht**?
- d) (1 Punkt) Nehmen Sie nun den stationären Restterm aus der STL-Zerlegung. Ist das Anpassen eines AR(p)-Modells angebracht? Argumentieren Sie mit Hilfe des ACF und PACF-Plot.
- e) (2 Punkte) Unabhängig von ihren Argumenten in d) sollen Sie nun aus der PACF die zwei plausibelsten Ordnungen p für ein AR(p)-Modell ablesen und angeben. Erklären Sie ihre Wahl.
- f) (2 Punkte) Passen Sie nun ein AR(p)-Modell an die Daten des Restterms an, welches zum kleinsten AIC-Wert führt. Geben Sie dessen Ordnung p an. Geben Sie den verwendeten R-Code an.
- g) Passen Sie nun ein additives Holt-Winters Modell an die in a) generierte Zeitreihe an.
 - i) (2 Punkte) Welche Bedeutung und welchen Wert hat der Parameter β ? Welchen Schluss ziehen Sie aus diesem Resultat?
 - ii) (1 Punkt) Wie gross ist die mit dem Holt-Winters Modell prognostizierte Milchleistung für den Juni 2006.
- h) Berechnen Sie das Zeitreihen-Regressionsmodell **milk ~ zeitschritte + monat** für die gegebenen Daten (eine Transformation ist nicht nötig).
 - i) (0.5 Punkte) Geben Sie den dazu verwendeten R-Befehl an.
 - ii) (0.5 Punkt) Ändert sich die durchschnittliche Milchleistung der Kühe zwischen 1994 und 2005 signifikant ($\alpha = 1\%$) aufgrund der Resultate des Regressionsmodells?
 - iii) (1 Punkte) Können Sie dem p-Wert von **milk** vertrauen? Bitte begründen Sie ihre Antwort.
- i) (2 Punkt) Welches AR(p)-Modell beschreibt die Residuen des Regressionsmodell aus h) gut?
- j) Verwenden sie nun die Funktion **glst()**, um im Zeitreihen-Regressionsmodell von Frage h) auch noch die Korrelation der Residuen zu berücksichtigen
 - i) (2 Punkte) Geben Sie die dafür die notwendigen R-Befehle an.
 - ii) (1 Punkt) Wie beantworten Sie nun die Frage h-ii)?