

Blatt 8

ALM und zufällige Effekte

Aufgabe 8.1 Gewichtetes Mittel

Betrachten Sie das Modell

$$y_i = \mu + \gamma_i \eta_i, \quad i = 1, \dots, n,$$

mit bekannten und konstanten $\gamma_i > 0$, wobei die Fehler η_i unkorreliert mit Erwartungswert 0 und Varianz 1 sind, μ eine Konstante, und bringen Sie es auf die Form eines ALM. Ist μ schätzbar? Falls ja, bestimmen Sie den BLUE für μ . Was bedeutet es, wenn die Kovarianzmatrix der Fehler nicht vollen Rang hat?

Aufgabe 8.2 Stratified Data

Sie sollen ein gewisses Merkmal einer Population erheben, welche aus K „Schichten“ (Strata) besteht. Für jedes Stratum, $k = 1, \dots, K$, gehen Sie davon aus, dass die Merkmale bei den m_k Individuen dieses Stratums mit Varianz σ_k^2 um den Mittelwert μ_k dieses Stratums streuen.

Sie sollen nun eine Stichprobe erheben, um den Mittelwert μ für die gesamte Bevölkerung zu bestimmen; der Gesamtumfang der Stichprobe ist aus Kostengründen allerdings auf n beschränkt.

- Wie groß sollte der Stichprobenumfang je Stratum sein, um μ möglichst genau bestimmen zu können?
- Man könnte die Strata ignorieren und die Stichprobe einfach zufällig und gleichverteilt (mit oder ohne Zurücklegen?) aus der Gesamtbevölkerung ziehen und deren Mittelwert als Schätzer für μ verwenden. Können Sie das als ALM formulieren? Wie genau wird μ dann bestimmt?
- Welches Vorgehen ist vorzuziehen?

Stratified sampling wird insbesondere bei Meinungsumfragen oft angewandt.

Aufgabe 8.3 Erdhölraffination

Laden Sie den Datensatz `petrol` aus `library(MASS)`. Er enthält Messdaten bezüglich 10 verschiedener Rohölsorten (No). Gemessen wurden das spezifische Gewicht (SG), der Dampfdruck (VP), die Flüchtigkeit des Rohöls (V10), die Flüchtigkeit des raffinierten Endproduktes (EP) und der Ertrag (Y). Für jedes Rohöl wurden Messreihen zu verschiedenen gewünschten Eigenschaften des Endproduktes durchgeführt. Untersuchen Sie den Zusammenhang zwischen Ertrages und den übrigen Parametern.

- Stellen Sie den Ertrag in Abhängigkeit von der Flüchtigkeit des Endproduktes dar und interpretieren Sie die Abbildungen.
- Zentrieren Sie die Messdaten und arbeiten Sie mit dem Modell $Y_{ij} = \mu_j + \beta_j \text{EP}_{ij} + \varepsilon_{ij}$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. Dabei beschreibt j die Probe und i die zugehörige Beobachtung. Bestimmen Sie jeweils den Anstieg und den Achsenabschnitt.

- (c) Modellieren Sie nun mit einem gemeinsamen Anstieg, $Y_{ij} = \mu_j + \beta_1 \mathbf{EP}_{ij} + \varepsilon_{ij}$, und vergleichen Sie die Modelle.
- (d) In (b) haben wir gesehen, dass die Achsenabschnitte stark variieren. Wir wollen dieses Verhalten versuchen zu erklären. Betrachten Sie dazu das Modell $Y_{ij} = \mu + \beta_1 \mathbf{SG}_{ij} + \beta_2 \mathbf{VP}_{ij} + \beta_3 \mathbf{V10}_{ij} + \beta_4 \mathbf{EP}_{ij} + \varepsilon_{ij}$. Ist dieses Modell in dem Modell in (c) enthalten? Erklären Sie! Vergleichen Sie weiterhin beide Modelle.
- (e) Wir nehmen nun an, dass die jeweiligen Probe zufällige Stichproben einer Population darstellen. Modellieren Sie die Achsenabschnitte als zufällig in Abhängigkeit von der Probe, $Y_{ij} = \mu + e_i + \beta_1 \mathbf{SG}_{ij} + \beta_2 \mathbf{VP}_{ij} + \beta_3 \mathbf{V10}_{ij} + \beta_4 \mathbf{EP}_{ij} + \varepsilon_{ij}$. Hierbei sei $e_i \sim \mathcal{N}(0, \tau^2)$ mit $\text{Cov}(e) = \tau^2 I$, $\text{Cov}(e_i, \varepsilon_{ij}) = 0$. Können Regressoren entfernt werden?
- (f) Reduzieren das Modell zu $Y_{ij} = \mu + e_i + \beta_3 \mathbf{V10}_{ij} + \beta_4 \mathbf{EP}_{ij} + \varepsilon_{ij}$. Und vergleichen Sie mit (e).
- (g) Prüfen Sie nun, ob außerdem noch der Anstieg für EP zufällig modelliert werden muss. Betrachten Sie dazu das Modell $Y_{ij} = \mu + e_i + \beta_3 \mathbf{V10}_{ij} + (\beta_4 + \eta_i) \mathbf{EP}_{ij} + \varepsilon_{ij}$. Hierbei seien (e_i, η_i) unabhängig von ε_{ij} , können jedoch untereinander korreliert sein.

Als Grundlage kann das folgende Kode-Skelett verwendet werden. https://www.tu-ilmenau.de/fileadmin/media/analysis/schmitz/Lehre/stat_ana_SS17/R_petroldummy.R

Nächste Übung: Donnerstag 13. Juli 2017 um 13:00 Uhr im Sr C 115