

Aufgaben zur Vorlesung  
**Multivariate Verfahren**  
Übungsblatt III

---

1. Beantworten Sie folgende Fragen bzw. bearbeiten Sie folgende Arbeitsanweisungen:
  - (a) Was muss für eine Kovarianzmatrix gelten?
  - (b) Wie berechnet sich die empirische Kovarianzmatrix? Verwenden Sie auch Matrixschreibweise.
  - (c) Was drückt der Korrelationskoeffizient aus?
  - (d) Wie hängen die Eigenwerte und die Definitheit einer Matrix zusammen?
2. Stellen Sie folgende Summen als Matrizenprodukte dar und geben Sie die Dimension der verwendeten Matrizen an:
  - (a)  $\sum_{i=1}^n \beta_i$
  - (b)  $\sum_{i=1}^n \beta_i^2$
  - (c)  $\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2$
3. Im Folgenden soll das jährliche Einkommen einer Person mit Hilfe einer Multiplen Linearen Regression modelliert werden. Geben Sie für das Vorliegen der angegebenen Variablen jeweils ein Modell an.
  - (a) Wöchentliche Arbeitszeit, Alter und Masterabschluss (ja, nein).
  - (b) Alter und Schulabschluss (Hauptschule, Realschule, Abitur).
4. Einer Ihrer Freunde hat für seine Masterarbeit den Einfluss von Neupreis in EUR, Anzahl Vorbesitzer und Anzahl Neuzulassungen des jeweiligen Modells pro Jahr auf den Preis von 200 Gebrauchtwagen für 2016 untersucht. Eine Variable für den Achsenabschnitt hat er nicht berücksichtigt. Aus seiner Statistik-Software erhält er folgenden Output:

$$\begin{array}{rcccl} Preis_t & = & 0.33 \cdot Neupreis_t & - 2501 \cdot Vorbesitzer_t & + 0.01 \cdot Neuzulassungen_t \\ & & (0.0033) & (1000) & (0.1) \end{array}$$

Da er leider die Grundlagen der Regressionsanalyse aus dem Grundstudium vergessen hat, kann er die Ausgabe des Programms nicht interpretieren und bittet Sie daher um Hilfe.

- (a) Welche der Koeffizienten sind signifikant zum Niveau  $\alpha = 5\%$ ? (Hinweis: Die geschätzten Standardfehler der geschätzten Koeffizienten sind jeweils in den Klammern angegeben.)
- (b) Interpretieren Sie die Koeffizienten.

- (c) Wie würde sich ganz allgemein die Interpretation verändern, wenn der Regressand logarithmiert wäre?
  - (d) Wie würde sich ganz allgemein die Interpretation verändern, wenn zusätzlich die Variable Neupreis logarithmiert wäre?
  - (e) Wie beurteilen Sie seine Entscheidung, eine Variable für den Achsenabschnitt nicht zu berücksichtigen?
5. Rechnerübung: Verwenden Sie den Datensatz *iris* aus Ilias. Dieser enthält vier Merkmale von Blütenblättern unterschiedlicher Schwertlilienarten (jeweils Länge und Breite des Sepalum (Kelchblatt) und Petalum(Kronblatt)). Führen Sie eine Multiple Lineare Regression durch, um aus Länge und Breite des Kelchblatts auf die Länge des Kronblatts schließen zu können. Interpretieren Sie die Koeffizienten. Sind diese signifikant? Wie beurteilen Sie die Güte der Regression insgesamt?