Philipps-Universität Marburg
Fachbereich Mathematik und Informatik
Prof. Dr. H. Holzmann
G. Alexandrovich

Sommersemester 2013

Übungen zur vertiefenden mathematischen Statistik

Blatt 3

Abgabe: Montag, 06.05.2013 vor der Vorlesung

Aufgabe 1 Laden Sie den Datensatz `savings` aus der R-Paket `faraway`.
i) Führen Sie eine lineare Regression durch, wobei die Variable `sr` (savings rate) als abhängige und die restlichen als erklärende Variablen verwendet werden.

ii) Untersuchen Sie den `summary`-Output des erhaltenen `lm()`-Objektes.

iii) Konstruieren Sie einen elliptischen und einen rechteckigen [1] Konfidenzbereich für den Vektor $(\beta_{pop15}, \beta_{pop75})$ und stellen Sie beide Bereiche in einem Plot dar. Sie können dabei das Paket `ellipse` benutzen.

iv) Führen Sie eine Variablenselektion anhand der p-Werte aus dem t-Test durch. Verwenden Sie 5% als Schwelle für zulässige p-Werte.

Aufgabe 2 (Konfidenz- und Prädiktionsintervalle)
Die folgende Tabelle enthält Daten über die Anzahl von Stunden, die 8 Studenten einer Vorlesung außerhalb der Vorlesungsstunden in einem Zeitraum von drei Wochen zum Lernen aufgewendet haben, sowie ihre Punktzahlen, die sie am Ende in der Prüfung erreicht haben.

Lernzeit in Stunden (x)	20	16	34	23	27	32	18	22
Punktzahl in der Prüfung (y)	64	61	84	70	88	92	72	77

Es sei das Modell der klassischen linearen Regression zugrunde gelegt, das heißt für die Fehler gelte die Normalverteilungsannahme.

1. Bestimmen Sie die Kleinste-Quadrate-Regressionsgerade des linearen Modells $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ und tragen Sie diese in ein Streudiagramm der Daten ein.

2. Testen Sie auf dem $1\%-$Signifikanzniveau die Nullhypothese, dass die Steigung der Regressionsgeraden Null ist und interpretieren Sie das Ergebnis.

3. Prognostizieren Sie mit Hilfe der Regressionsgleichung aus Teilaufgabe a) das Prüfungsergebnis eines Studenten, der $x_0 = 30$ Stunden für das Studium des Vorlesungsmaterials verwendet hat.

   Bitte wenden.

[1] nach Bonferroni-Methode

4. Bestimmen Sie „von Hand“ ein 90%−Konfidenzintervall für die mittlere Punktzahl von Studenten, die $x_0 = 30$ Stunden gelernt haben.

5. Bestimmen Sie „von Hand“ ein 90%−Prädiktionsintervall für die Punktzahl eines einzelnen Studenten, der $x_0 = 30$ Stunden für die Vorbereitung verwendet hat.

6. Versuchen Sie die Teilaufgaben c) bis e) mit Hilfe der Funktion `predict()` in R zu lösen und vergleichen Sie die Ergebnisse.

Aufgabe 3 (V) Im Folgenden betrachten wir ein klassisches lineares Model ohne Intercept mit $X^\top X = I_p$. $\hat{\beta}_{LS}$ bezeichne den Kleinste-Quadrate-Schätzer und $\gamma \in (0, \infty)$ sei fix.

i) Wir betrachten den so genannten LASSO-Schätzer (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator) für den Parameter β der linearen Regression:

$$\hat{\beta}_{lasso} := \text{argmin}_\beta ||Y - X\beta||_n^2 + \gamma \sum_{j=1}^{p} |\beta_j|,$$

Zeigen Sie: es gilt

$$\hat{\beta}_{lasso,j} = \text{sgn}(\hat{\beta}_{LS,j})(|\hat{\beta}_{LS,j}| - \frac{\gamma}{2})^+,\ 1 \leq j \leq p.$$

ii) Wir betrachten den Schätzer

$$\hat{\beta}_{th} := \text{argmin}_\beta ||Y - X\beta||_n^2 + \gamma \sum_{j=1}^{p} \mathbf{1}_{\{|\beta_j|>0\}}.$$

Zeigen Sie: es gilt

$$\hat{\beta}_{th,j} = \hat{\beta}_{LS,j} \mathbf{1}_{\{\hat{\beta}^2_{LS,j} > \gamma\}},\ 1 \leq j \leq p.$$