Prof. Dr. E. Cramer, Prof. Dr. M. Kateri RWTH Aachen, SS 2021
T. v. Bentum, L. Kaufmann, M.Sc.
publication: May 14th solutions: May 20th

Applied Data Analysis

Exercise Sheet 3

Exercise 11

For a linear model, write $B\boldsymbol{\beta} = \sum_i B_i\beta_i$, where B_i, $i \in \{1, \ldots, p\}$, are the column vectors of $B = [B_1|\ldots|B_p]$. Show that an individual parameter β_i is identifiable if and only if $\{B_j, j \in \{1, \ldots, p\}\backslash\{i\}\}$ is linearly independent from B_i, $i \in \{1, \ldots, p\}$.

Definition: Let $A \in \mathbb{R}^{n\times n}$, $n \in \mathbb{N}$, be a real square matrix. A representation of A as a matrix product $A = QR$ with $Q \in \mathbb{R}^{n\times n}$ an orthogonal matrix and $R \in \mathbb{R}^{n\times n}$ an upper triangular matrix is called a *QR-decomposition* (of A).

More generally, if $A \in \mathbb{R}^{m\times n}$, $m, n \in \mathbb{N}$, is a rectangular matrix, (thus, by neccessity $m \geq n$), then, a factorization

$$A = QR$$

with $Q \in \mathbb{R}^{n\times m}$ a matrix with orthogonal column vectors and $R \in \mathbb{R}^{m\times m}$ an upper triangular matrix is called *QR-decomposition* of A.

QR-decompositions can be used for implementations that yield greater numerical stability than direct application of algorithms to A (see e.g. R-Lab, Task 17).

Remark: Note that some computer algebra programs use the convention

$$A = QR = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0_{(m-n)\times n} \end{pmatrix} = Q_1R_1$$

where $Q \in \mathbb{R}^{m\times m}$ is an orthogonal matrix and $R_1 \in \mathbb{R}^{n\times n}$ is a an upper triangular matrix. The submatrix Q_2 is, in general, not unique and completes the column vectors of Q_1 to an orthonormal basis of $\mathbb{R}^n$.

One procedure to derive the QR-decomposition of a matrix is called the *Gram-Schmidt process.*

Definition: Consider a matrix

$$A = [a_1|\ldots|a_n] \in \mathbb{R}^{m\times n}$$

with full column rank. Let

$$u_1 = a_1, \qquad e_1 = \frac{u_1}{||u_1||}$$

$$u_2 = a_2 - \langle a_2, e_1\rangle e_1 \qquad e_2 = \frac{u_2}{||u_2|}$$

$$\dots$$

$$u_k = a_k - \langle a_k, e_1\rangle e_1 - \dots - \langle a_k, e_{k-1}\rangle e_{k-1}, \qquad e_k = \frac{u_k}{||u_k||}$$

Then, the set of vectors $\{e_1, \dots, e_n\}$ is orthonormal and for each subset $I \subseteq \{1, \dots, n\}$ the set $\{e_i, i \in I\}$ has the same span as the corresponding set $\{a_i, i \in I\}$. The above algorithm is called the *Gram-Schmidt procedure.*

Remark: The Gram-Schmidt procedure results in the QR-decomposition

$$A = QR = [e_1|\dots|e_n] \begin{pmatrix} \langle a_1, e_1\rangle & \langle a_2, e_1\rangle & \dots & \langle a_n, e_1\rangle \\ 0 & \langle a_2, e_2\rangle & \dots & \langle a_n, e_2\rangle \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \langle a_n, e_n\rangle \end{pmatrix}$$

The procedure can be understood as first normalizing the vector a_1 to a unit vector e_1 and then, for each $k \in \{2, \dots, n\}$, successively first eliminating the components in the directions of the previous vectors $a_1, \dots a_{k-1}$ (thus ensuring orthogonality) before again normalizing to a unit vector u_k. In particular, this yields the representations

$$\begin{aligned} a_1 &= \langle e_1, a_1\rangle e_1 \\ a_2 &= \langle e_1, a_2\rangle e_1 + \langle e_2, a_2\rangle e_2 \\ &\dots \\ a_k &= \sum_{j=1}^{k} \langle e_j, a_k\rangle e_j \end{aligned}$$

with respect to the new orthonormal basis of the span of the column vectors.

Exercise 12

Consider a linear model

$$\boldsymbol{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

as in Definition I.4.2. with $p < n$ and $rank(B) = p$.

(a) Let $B = QR$ be the QR-decomposition of B. Show that the LSE of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = R^{-1}Q'\boldsymbol{Y}.$$

(b) Let

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

Use the Gram-Schmidt procedure to derive a QR-decomposition of A.

(c) Assume that the linear model
$$\boldsymbol{Y} = A\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
is given. Use the results from (b) to derive $\hat{\boldsymbol{\beta}}$ and calculate the estimator for the observation $(y_1, y_2, y_3)' = (1, 1, 1)'$.

In Definition I.4.2 it is assumed that $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 I_p$, $\sigma^2 > 0$, that is, the observations have the same finite variance (*homoscedasticity*). In practice, this assumption is often violated. If, for example, one assumes a regression model for a time series, one may see that the scatter around the regression line increases with time. At other times, the variance may increase with the magnitude of the observations.

In these cases, the method of ordinary least squares (OLS) may lead to statistically inefficient procedures.

Definition: Let
$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$
with $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \Theta \subseteq \mathbb{R}^p$ a parameter space, $X \in \mathbb{R}^{n\times p}$ a known design matrix and $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\Sigma$, with $\Sigma \in \mathbb{R}^{n\times n}_{\geq 0}$, a known variance-covariance matrix and $\sigma^2 > 0$ unknown.
Then, if Σ is regular, the method of generalized least squares (GLS) estimates the parameter vector $\boldsymbol{\beta}$ by minimizing
$$\hat{\boldsymbol{\beta}} = \text{argmin}_{\beta\in\Theta}||\boldsymbol{Y} - X\boldsymbol{\beta}||^2_\Sigma = \text{argmin}_{\beta\in\Theta}(\boldsymbol{Y} - X\boldsymbol{\beta})'\Sigma^{-1}(\boldsymbol{Y} - X\boldsymbol{\beta})$$
the squared Mahalanobis distance of the residual vector.

Exercise 13

Consider the model in (1) and denote by $\hat{\boldsymbol{\beta}}$ the GLS estimator of $\boldsymbol{\beta}$.

(a) Derive the normal equations for $\hat{\boldsymbol{\beta}}$.

In the following, assume that X is of full rank and Σ is regular.

(b) Show that
$$\hat{\boldsymbol{\beta}} = (X'\Sigma^{-1}X)^{-1}X\Sigma^{-1}\boldsymbol{Y}.$$

(c) Calculate $\text{E}(\hat{\boldsymbol{\beta}})$ and $\text{Cov}(\hat{\boldsymbol{\beta}})$.

(d) Find a transformation of the data such that
$$\boldsymbol{Z} = B\boldsymbol{\beta} + \boldsymbol{\eta} \tag{2}$$
is a LM and the OLS estimator with respect to the model (2) is equivalent to the GLS estimator with respect to the model (1).

Hint: Argue that $\Sigma = A'A$ for some $A \in \mathbb{R}^{n\times n}_{>0}$ and use this regular square root of A to find a suitable transformation.

(e) Further assume that $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2\Sigma)$. Find the likelihood equations for $\boldsymbol{\beta}$ and use the results from the lectures and (d) to show that the MLE coincides with the GLS estimator.

(f) Under the assumptions of (e) derive a testing procedure for the model (1) analogous to I.4.40.