Institut für Mathematik
Abteilung Numerik
Jana de Wiljes

3. Problem sheet for **Statistical Data Analysis**

Exercise 1 (6 Points)
Consider the following linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \cdots + \beta_p X_{i,p} + \epsilon_i, \quad i = 1, \ldots, n \tag{1}$$

with $p = 3$ and $n = 201$ where $\epsilon_i \sim \mathcal{N}(0, 1)$ are iid. Furthermore, an associated realisation contained in the data set $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{X} \in \mathbb{R}^{n \times p}$ is available (see Moodle). Implement a routine that computes an estimate of $\hat{\beta}$ of the ML estimator, of $\hat{\sigma}^2$ given $\hat{\beta}$ and of the adjust $\hat{\sigma}^2_{ad}$. Comment on all your computes results.

Exercise 2 (4 Points)
Let $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ be the LS-estimator and $\hat{\sigma}^2_{ad} = \frac{1}{n-p-1} \hat{\epsilon}^\top \hat{\epsilon}$ the REML-estimator. Show that the following properties hold:

1. $\text{Cov}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$
2. $\mathbb{E}[\hat{\sigma}^2_{ad}] = \sigma^2$

Exercise 3 (6 Points)
Every human is a carrier of one of the three genotypes AA, Aa, or aa. The genotypes are occurring with the probabilities $(1-p)^2$, $2p(1-p)$ and p^2 whereas $0 < p < 1$ and testing of n persons yielded

- x persons had the genotype AA
- y persons had the genotype Aa
- z persons had the genotype aa

Describe the corresponding statistical model and determine the Maximum Likelihood Estimator for p.