Institut für Mathematik
Abteilung Numerik
Jana de Wiljes

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

Exercise 1 (9 Points)
Given data set $\mathbf{y}_{snow}$ and realisations x_1, x_2, x_3, x_4, x_5 of potential drivers of $\mathbf{y}_{snow}$ (see Moodle). Implement a linear regression and determine the associated unknown β_i values. Try all combinations of subsets of x_1, x_2, x_3, x_4, x_5 for the regression. Further compute the coefficient of determination R^2 and its corrected counter part $\bar{R}^2$ for all combinations. Argue which variables should be taken into account as drivers of $\mathbf{y}_{snow}$.

Exercise 2 (6 Points)
Show that $\mathbb{E}[\hat{\epsilon}] = 0$ and determine $Cov(\hat{\epsilon})$ for the linear regression problem with $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$.