University of Wrocław: Data Science 16.09.2019

Entrance exam: Linear Models

First name: .... Last name: ....

1. Consider the data set with $n = 20$ observations and three variables: Y, X_1, X_2. Assume that the relation between values of the response variable Y and explanatory variables X_1 and X_2 is provided by the multiple linear regression model

$$\text{for} i \in \{1, \ldots, n\} \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad , \tag{1}$$

where $\epsilon_1, \ldots, \epsilon_n$ are iid random variables from the standard normal distribution $N(0, \sigma^2)$.

a) Derive (5pt) (or provide, 2pt) the formulas for the maximum likelihood estimators $\hat{\beta}$ and $\hat{\sigma}^2$ of the vector of parameters $\beta = (\beta_0, \beta_1, \beta_2)$ and σ^2.

b) Derive (2pt) (or provide, 1pt) the formula for the distribution of $\hat{\beta}$.

c) (3pt) Let us denote by X the design matrix $X = [1|X_1|X_2]$ of the dimension 20×3, with the first column of ones corresponding to the intercept term. Let

$$X^T X = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 4 \\ 0 & 4 & 5 \end{bmatrix} \quad .$$

and $\sigma = 1$. Find the variance of $\hat{\beta}_1$.

d) (4pt) Let's now assume that σ is known and is equal to 1. We perform the regular z-test test (with known σ) of the hypothesis $H_0 : \beta_1 = 0$ vs alternative $H_A : \beta_1 \neq 0$. Calculate the power of this test in the multiple linear regression model (1) when the true (but unknown) $\beta_1 = 1$.

e) (2pt) Find the variance of the maximum likelihood estimator $\hat{\beta}_1$ in the simple regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i \quad . \tag{2}$$

f) (4pt) Calculate the power of the z-test test for the hypothesis $H_0 : \beta_1 = 0$ vs alternative $H_A : \beta_1 \neq 0$ in the simple regression model (2) when σ is known and is equal to 1 and the true (but unknown) $\beta_1 = 1$.