Linear Regression (NMSA407)
Written Exam **00** – January 1, 2022

General instructions

- Solutions can be worked out in either of languages: English, Czech, Slovak.
- Allowed time: 120 minutes.
- Use of any electronic devices is strictly prohibited.
- Leaving the lecture room is only possible after handing in the solutions to the examiner. It is then not possible to return to the lecture room during the exam.

Assignments

Problem 1

We are interested in knowing on how two factors: *season* (roční období) and *soil type* (typ půdy) influence concentration of *nitrogen* (dusík) in a soil. Certain amount n of soil samples obtained either in *summer* or *winter* was analyzed, the samples were taken in a soil of either of three types – A, B, C.

Provide a probabilistic representation of obtained data. Propose a reasonable linear model that would allow to quantify an influence of *season* and *soil type* on the *nitrogen concentration.* Explicitely write down (at least as a scheme) the form of the model matrix. Further, let

- $\overline{m}$ be the mean of the expected nitrogen concentrations once the mean is calculated over the six combinations of *season* and *soil type*;
- $\overline{m}_S$ be the mean of the *summer* expected nitrogen concentrations once the mean is calculated over the three *soil types*;
- $\overline{m}_W$ be the mean of the *winter* expected nitrogen concentrations once the mean is calculated over the three *soil types*.

Specify the model such that one of the regression parameters (which one?) is equal to $\overline{m}$.

Problem 2

Use a linear model specified in Problem 1 and derive (including a proper proof of all statements used in the derivations) a confidence interval for parameter $\theta = \overline{m}_W - \overline{m}_S$.

Do not forget to specify all (additional) assumptions needed for your derivations. In the proof, all properties of projections conducted in context of the least squares estimation and properties of related projection matrices can be used without explicit derivation of such properties.

Problem 3

Consider a homoscedastic linear model. **State and proof the theorem on a consistent estimator of the residual variance σ^2 of such a model.** In the proof, the following knowledge can be used without explicit justification:

- All knowledge related to a linear model based on data with a finite and fixed sample size n;
- Knowledge of consistent estimators for quantities other than σ^2.