MSA Advanced Statistical Modelling: Project assignment

Smart farming: predictive modelling to study growth process

The objective of the project is to develop statistical models and methods to characterize growth process and develop predictive models to as a function of available covariates in farming.

We have two sets of data available for this purpose.

- Pig growth modelling: predict the weight of the subject
- Onion growth modelling: predict the weight of the bulb

Warning: This dataset is strictly for the project of this course only, should not be shared with anybody, and should not be made available to public by any means.

Onion Data: Growth of the onion has been monitored from seedlings to harvest with regular measurements from leaves above the ground and the bulbs under the ground. A visit was made in every 1-2 weeks to collect data in several farms. Each time a sample of onion of size 5 (bulbs) or 10 (leaves) were selected to measure several characteristics of leaves and bulbs and this was repeated for three times. After measurements, the sampled onions were disposed so the individual unit of the onions are not followed. Also the sample unit of onion for bulbs and leaves does not match.

- xxx_underground.csv: growth measurements of bulbs
- xxx_ground.csv: growth measurements of leaves

Pig data Growth of the pig has been monitored from birth to release to the market for a sample of pigs. Measurements were made approximately once in every week in several farms. The same subjects have been followed through the study.

Analysis: There is no prescription as to what your analysis should include, though you are expected to follow 'good statistical principles'. Good ideas are exploratory analyses of the variables, and fitting various models, consistent with the data that you can interpret in non–statistical language. You may use any of the techniques presented in this course, or indeed, any other, in order to understand which, if any, of the covariates are related to the growth process, and in what way.

Some points to consider:

- Write a paragraph describing background of the study and relevance of your project in your own words and give the scientific objectives in analysing them.
- Prepare a suitable numerical and graphical representation highlighting whatever features you consider to be relevant. Don't forget to number the figures (and tables) in order and include a caption for each. They should be quoted in the main body of the text to refer to.
- Start with a simple analysis that could be used as a benchmark.

- Develop a longitudinal model for the dataset, including sufficient details of the assumptions you are making and justification of the statistical methods used.
- Consider an alternative method that you are aware of and make comparisons.
- Give interpretation of the results and make your conclusions.

Report: Your report should be no more than 10 A4 pages in length, in 12pt font and including appropriate tabular and graphical material. It should be structured as a scientific report with an introduction, sections describing your analysis of the data, a statement of your conclusions, and references to the relevant literature.

The purpose of the reference is to refer readers to the original source of the information that you may assume in your writing. So do make an appropriate selection that is relevant in the text, such as more background information that you omit, to justify your claims or to give credits to others.

Do not include any code or a screen capture graphic from R output in the report. Carefully choose graphics and tables that are of good quality and are pleasing to see on the page.

Presentation: The first presentation will be on the exploratory data analysis on 5th November for 5-10 minutes. The second presentation will be on model building and preliminary analysis on 3rd December for 5-10 minutes. The final presentation will be in January during the exam period. Each will have a total of 25 minutes (15 min. presentation + 8 min. questions + 2 min. preparation/turn-over).

Submission: The report with corresponding codes, as well as the final presentation slide should be submitted by 18th Jan. 2025.