

Linear Models in Animal Breeding

Introduction

Julius van der Werf

Linear Models are the core of almost all analysis of animal breeding data. They form a very powerful tool for the estimation and testing of effects and for the prediction of genetic merit of animals. Much of animal breeding research is also focused on estimation of variance components and genetic parameters, and dissecting the variation existing in animal breeding data into meaningful genetic and non-genetic components has evolved into an advanced science.

The application of linear models in animal breeding requires both statistical and genetic insight and to some extent is therefore somewhat unique. Animal breeders have enormously helped advance the general area of linear model applications, with Dr. C.R. Henderson being one of the main forces in the 20th century. His work in relation to mixed models and variance component estimation started toward the end of the 1940ties and he has built a tremendous fundament since. Development of the animal model and insight into the various genetic properties of the BLUP methodology did not develop until the 1980ties. Now, two decades later, application of mixed models in animal breeding has matured and much more advanced models are being used, not in the least also due to the increase of computing power. Models with various random effects are being used, and more complex variances structures are being applied, e.g. for the analysis of genomic data or of longitudinal data, recorded as repeated measurement in time or across environments. Also non-linear models are being used, with various degrees of success.

More complex models can be very useful tools to obtain more advanced insight and hypothesis testing, and more accurate prediction of genetic merit, based on data that originates from a very complex process called 'life'. The art of modeling is to simplify in a responsible and sensible manner because, as Bertrand Russel said: "Although this may seem a paradox, but all exact science is dominated by the idea of approximation". Complex models also require more knowledge from the user, not in the least to be able to judge whether the data structure he or she deals with allows the estimation of the parameters in the model. The danger nowadays maybe that it is so easy to use software packages that allow very sophisticated modeling, yet the data may not at all.

This course will introduce you to linear models in animal breeding from a rather basic level. The emphasis of the course is on understanding the principles of linear models, and starting with very simple examples is the best way to go. Doing the basic matrix calculations by hand helps to understand what software packages are actually doing. We will use some software packages and try to understand the logic behind them and the choices they provide. We will slowly build more advanced models. The emphasis will be on application, and on understanding data structure in relation to the estimation problem.

In this course you will not find much emphasis on formal proofs, and formal statistical and mathematical properties of the various methods and models. Experienced users of linear models will know that it takes some time to gain a good insight in linear models but a basic understanding of the principles and the rationale behind the various ways to build a model is a good start.