



統計科學研究所

INSTITUTE OF
STATISTICAL SCIENCE



SEMINAR



STAT.AS

Seminar

Title : Composition and sample size determination for training set in genomic prediction

Speaker : Prof. Chen-Tuo Liao (廖振鐸 教授)

(Department of Agronomy, National Taiwan University)

Time : 10:30 AM~12:00 PM, Monday, Mar. 6, 2023

Place : Auditorium, B1F, Institute of Statistical Science

Abstract

Genomic prediction (GP) is a statistical method used to select quantitative traits in animal or plant breeding. For this purpose, a GP model is first built that uses phenotype and genotype data in a training set. The trained model is then used to predict genomic estimated breeding values (GEBVs) for individuals with genotypic data along. For a specified test set, we develop a highly efficient algorithm to determine an optimal subset from a candidate set in which the individuals have been genotyped but not phenotyped yet. The chosen subset serves as the training set to be phenotyped, and then the GP model is built using its phenotype and genotype data. In this study, we propose an optimality criterion, called as r-score, to determine the required training set. The r-score criterion is derived directly from Pearson's correlation between GEBVs and phenotypic values of the test set. The proposed method is shown to be advantageous over existing ones, mainly because that it fully uses the genomic relationship between the test set and the training set by taking into account both the variance and bias for predicting the GEBVs. By applying the logistic growth curve to draw a connection between r-score and the training set size, a practical approach is proposed to determine the sample size of the optimal training set. Some real genome datasets are used to illustrate the proposed approach.

※ Online live streaming through Cisco Webex will be available.

※ The tea reception will be held at 10:10.