

Indian Statistical Institute

Applied Statistics Unit

SEMINAR NOTICE

Speaker: Saurabh Ghosh

Title: Association Mapping Of Multivariate Phenotypes In The Presence Of Missing Data

Date: 13 September, 2022

Time: 16:15 PM

Venue: ASU Seminar Room

Online Platform: Google Meet (meet.google.com/qkh-wczx-zbs)

Abstract:

Clinical end-point traits are often characterized by quantitative and/or qualitative precursors and it has been argued that it may be statistically a more powerful strategy to analyze a multivariate phenotype comprising these precursor traits to decipher the genetic architecture of the underlying complex end-point trait. While various genotype-level methods of association, such as MultiPhen (O'Reilly et al, 2012) have been developed for association mapping of multivariate phenotypes, allele-level tests (Lee et al. 2013) are known to yield more power than genotype-level tests in case-control association analyses. In this study, we explore two allele-level tests of association for analyzing multivariate phenotypes: one based on a Binomial regression model in the framework of inverted regression of genotype on phenotype and the other based on the Mahalanobis distance between the two sample means of vectors of the multivariate phenotype corresponding to the two alleles at a SNP. Both the methods inherit the flexibility of incorporating both discrete as well as continuous traits in the multivariate phenotype vector. Using extensive simulations, the potential of the methods in enhancing the power of detecting pleiotropic association is evaluated in comparison with MultiPhen, which is based on a genotype-level test. Moreover, it may often arise in practice that data may not be available on all phenotypes for a particular individual. We explore methodologies to estimate missing phenotypes conditioned on the available ones and carry out the Binomial Regression based test for association on the "complete" data. We partition the vector of phenotypes into three subsets: continuous, count and categorical phenotypes. For each missing continuous phenotype, the trait value is estimated using a conditional normal model. For each missing count phenotype, the trait value is estimated using a conditional Poisson model. For each missing categorical phenotype, the risk of the phenotype status is estimated using a conditional logistic model. We carry out simulations under a wide spectrum of multivariate phenotype models and assess the effect of the proposed imputation strategy on the power of the

association test vis-a-vis the ideal situation with no missing data as well as analyses based only on individuals with complete data. We illustrate an application of our method using data on Coronary Artery Disease.

(This is a joint work with Kalins Banerjee)

All are invited to attend.