StatGen Seminar Series

Data Integration Via Analysis of Subspaces (DIVAS)

Jan Hannig, PhD
Professor
Department of Statistics and
Operations Research
UNC—Chapel Hill

Friday, May 3, 2019 at 10:00 am
MSRB III 1125 (Seminar Room)

A major challenge in the age of Big Data is the integration of disparate data types into a data analysis. That is tackled here in the context of data blocks measured on a common set of experimental subjects. This data structure motivates the simultaneous exploration of the joint and individual variation within each data block. This is done here in a way that scales well to large data sets (with blocks of wildly disparate size), using principal angle analysis, careful formulation of the underlying linear algebra, and differing outputs depending on the analytical goals. Ideas are illustrated using cancer and neuroimaging data sets.

Joint work with J.S. Marron, Jack Prothero, and Meilei Jiang. There is an earlier paper on this subject that will be partially discussed here: https://doi.org/10.1016/j.jmva.2018.03.008

Duke University School of Medicine

Duke Center for Statistical Genetics and Genomics