THE UNIVERSITY OF CHICAGO

Department of Statistics
MASTER'S THESIS PRESENTATION

YUFAN HU

Department of Statistics
The University of Chicago

Effectiveness of Addressing Cofounders in Large Scale Genetic Association Studies

TUESDAY, May 9, 2017, at 10:30 AM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

This paper examines the means that were taken to adjust for confounding due to population substructure in genetic association studies with social-scientific outcomes. As population substructure may lead to biased test statistics and inflated variances, the failure to remove it from association studies often results in increases in false positive rates. Focusing on "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment" (Rietveld et al., 2013), this paper further evaluates whether the methods used in the Genome-wide Association Study (GWAS) were sufficient and appropriate. The GWAS claims to find three genome-wide significant SNPs that relate to educational attainment, after including the first four principal components as controls in all cohort-level analyses on Caucasian samples to guard against population substructure. The fixation index, F_{ST}, is calculated using allele frequencies of different populations to quantify the genetic differences between these populations and is used to assess whether a connection remains between population differences and the p-value data from regression analyses in the original GWAS. After analyzing the distribution of these p-values and performing relative regression analyses, a statistically significant relationship is found between F_{ST} and the p-values, suggesting that gene discoveries made in the Education Attainment GWAS are still likely to be influenced by population substructure.