LD-Based Mapping by Cluster Analysis of the GWAS To Determine a Single Variant Most Associated With Alzheimer's Disease

Gutman, Lucas (School: iPrep Academy North)

My novel method of machine learning that I utilized with cluster analysis can be done on any computer and is fast and easy to learn. My novel method investigated a genome-wide association study (GWAS). The method involves scanning the genomes from many different people and looking for genetic markers that can be used to predict the presence of a disease. It was hypothesized that if the variants that are highly associated (r>.80) with Alzheimer's disease (from the GWAS database) are placed in a hierarchical cluster analysis algorithm, then the number of variants can be greatly reduced and linear regression of medoid analysis will further reduce that number to a single variant that is most likely to cause Alzheimer's disease thereby creating a novel method of machine learning. I used variants that have at least nine other variants associated with Alzheimer's disease, placed these variants into a hierarchical cluster analysis algorithm that reduced that number to a few clusters. The cluster's dendogram showed two distinct, significant clusters of the variants. This dendogram reduced the variants responsible for Alzheimer's disease from 16 to 2. Then, placed these clusters into a medoid partitioning algorithm to find the centrally located variants. Medoid cluster 2 showed that LD1 has the higher location. Using a linear regression model of those medoids, the highest r2 value identified the most likely variant that causes Alzheimer's disease: r2 for Variant rs290358 =.9060 r2 for Variant rs573521= .9693 Therefore, Variant rs573521 is the best variant to study for the most likely cause of Alzheimer's disease because of its higher r2 value. My hypothesis was substantiated since my novel method of machine learning narrowed down a large number of variants to a single variant.

Awards Won:

American Statistical Association: In-Kind membership to ASA for all winners, including honorable mentions

American Statistical Association: Honorable Mention