BIOSTAT 2017

21st SCHOOL OF BIOMETRICS

Snehalata Huzurbazar
University of Wyoming,
Department of Statistics,
Laramie, WY 82071, USA

lata@uwyo.edu
http://www.uwyo.edu/statistics/facultyandstaff/huzurbazar_snehalata.html

Topic:

Human Microbiome Data Analysis: Overview and Statistical Problems

Abstract:

By latest estimates, the human body contains about half human cells and half bacterial cells, though previous estimates have this ratio as skewed as 1:10. The bacteria form ecosystems in the human body and are referred to as the human microbiome, which has become increasingly important for various aspects of human health. The main source for microbiome data is from sequencing the 16S rRNA gene. The resulting data are highly sparse, high-dimensional counts usually with fewer samples than number of (bacterial) taxa, namely, $n<p$. My experience comes from working with the vaginal microbiome along with covariate data from female subjects as part of a U.S. National Institutes of Health Project. As we worked with such data, we used the opportunity to examine common first steps such as filtering and normalization, and developed some proposals for improvement. We then explored meaningful visualizations that can lend biological insights for such data. In the process, we compared various ordination methods that are commonly used by biologists for analyzing these data. Currently, we are examining methods for longitudinal analysis for such data collected over a handful of repeated visits.
Using my experience, the lectures will be structured to give the audience an overview of the methods currently used to analyze human microbiome data, and also to point towards the open problems where the statistical sciences could make contributions.