

## Workshop

# Latent Variables in High Dimensional Data

**William R. Atchley and Lisa McFerrin**

North Carolina State University

---

This workshop introduces the concept and relevant statistical methodology for “Latent Variables and Latent Structure” as they relate to analyses of high dimensional molecular data (HDMD). A latent variable model relates a set of observed or *manifest variables* to a set of underlying latent variables. Latent variables are not directly observed but are rather inferred through a mathematical model constructed from variables that are observed and directly measured.

HDMD typically contain thousands of data points arising from a substantially smaller number of sampling units. Such data present many complexities. They are typically highly interdependent, exhibit complex underlying components of variability, and meaningful replication is rare. The workshop will focus on three multivariate statistical methods that can facilitate description and analyses of latent variables and latent structure inherent to HDMD. These statistical methods are intended to reduce the dimensionality of HDMD such that patterns of biologically meaningful patterns of multidimensional covariability are exposed so that relevant biological questions can be explored.

The workshop will introduce principal components analysis, common factor analysis and discriminant analysis and, in the time available, discuss their relative suitability for exploring a series of important biological questions. These three methods will be demonstrated using a set of R programs together with annotated output and documentation.