
SEMINAR

Sleep Hypnograms, Insurance Claims, and Stroke Kinematics: Big Data and Potentially Many Weak, Predictive Signals

Bruce J. Swihart, Ph.D.

**Postdoctoral Fellow, Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health**

ABSTRACT

The increasing ease and decreasing cost for collecting complex and structured data has ushered in the Big Data era. Three big data sets are discussed in terms of their size and modeling approaches, with a primary focus on sleep hypnogram data. 5,598 sleep hypnograms collected from the Sleep Heart Health Study are analyzed in the context of sleep-disordered breathing and fragmented sleep with scalable GEE log-linear models as well as Multi-state Survival models. The information loss between 5-state hypnograms and 3-state hypnograms is explored. In the second dataset of Insurance Claims, the task of predicting the Days in Hospital of the following year is undertaken as issued in the challenge of the \$3 million Heritage Health Prize. Hand Movements After Stroke (Stroke Kinematics) are highly structured and complex data that contain features that are a strong classifier of Stroke-status.

12:00 Noon - Wednesday, November 20, 2013
LEPH 216, 60 College Street
