

SAMSI Program on Multiplicity and Reproducibility in Scientific Studies

Stan Young
April 28, 2006

Subgroup Problem

1. In a clinical trial, which subgroup responds well? Which has side effects?
2. How can you be sure it is not a false positive?
3. We have a landmark data set: BiDil works better in African-Americans. (Dave DeMets, U Wis.

Lack of Reproducibility of Non-randomized Studies

1. Work from JAMA paper.
2. New book chapter (85 pages!).
3. Samuel Shapiro epidemiology paper.
4. The “sociology” of hypothesis testing.

New Method for $n \ll p$

1. Non-negative matrix factorization
2. Variable weighted clustering
3. Inference-based, matrix factorization