



Colloquium Series Seminar

DATE:

Thursday,
April 3, 2008

LOCATION:

Kresge Building G1
Snyder Auditorium
(please note location)

TIME:

4:00 pm-5:00 pm

Coffee Reception at
3:45 pm

HARVARD SCHOOL OF PUBLIC
HEALTH
DEPARTMENT OF BIostatISTICS

655 Huntington Avenue
Building 2, 4th Floor

Phone: 617-432-1056

Fax: 617-432-5619

Email: dept@hsph.harvard.edu

Bradley Efron, Ph.D.

Department of Statistics
Stanford University

will present a lecture on

“Are a Set of Microarrays Independent of Each Other?”

Abstract:

A typical microarray study might involve two groups of subjects, Controls and Treatments. Each subject provides material for his or her individual microarray, reporting some large number of genetic expressions at the same time, and yielding an m by n data matrix "X", with m genes and n subjects. We expect the measurements down any one column to be correlated, since genes act in concert, making the rows of X correlated. However the columns, that is the microarrays, are usually assumed to be independent. If, for example, we form two-sample t-statistics for each gene's data, the standard Student's t null hypothesis with $n-2$ degrees of freedom requires independence, as do familiar techniques such as cross-validation and permutation testing. This talk concerns testing a matrix X for column-wise independence when the rows may be highly correlated. The effect of row-wise correlation is to greatly reduce the power of standard tests. In my main example, row-wise correlation will be shown to reduce the effective sample size from $m=20426$ to 17.

