Stanford University
Department of Statistics

**DEPARTMENTAL SEMINAR**

*** Note Special Day ***
4:15pm, Thursday, February 16, 2012
Sequoia Hall Room 200

**Speaker:** Kai Zhang  
*Department of Statistics,  
The Wharton School, University of Pennsylvania*

**Title:** Valid Post-Selection Inference

**Abstract:**

It is common practice in statistical data analysis to perform data-driven model selection and derive statistical inference from the selected model as if this model was known in advance. Such inference is generally invalid. We propose to produce valid “post-selection inference” by reducing the problem to one of simultaneous inference. Simultaneity is required for all linear functions that arise as coefficient estimates in all submodels. By purchasing “simultaneity insurance” for all possible submodels, the resulting post-selection inference is rendered universally valid under all possible model selection procedures. This inference is therefore generally conservative for particular selection procedures, but it is always more precise than full Scheffe protection. We describe the structure of the simultaneous inference problem and give some asymptotic results. We also develop an algorithm for numerical computation for the width of our new confidence intervals.

This is joint work with Richard Berk, Lawrence Brown, Andreas Buja, and Linda Zhao.