



統計科學研究所

INSTITUTE OF  
STATISTICAL SCIENCE



統計所學術演講



中研院統計所

## 學術演講

講題：Random Integrated Subdata Ensemble (RISE) for  
Big Data Model Building

講者：Prof. Lih-Yuan Deng ( 鄧利源教授 )  
( University of Memphis, U.S.A.)

時間：2023年1月5日(星期四)，10:30-12:00

地點：統計所B1演講廳

### Abstract

We discuss our newly proposed Random Integrated Subdata Ensemble (RISE) method to build a more efficient big data model such as variable selections and/or model building. Generally speaking, large sample size tends to make some "variables" statistically significant while they may not have real "practical importance". Therefore, it is more likely to overly select such variables simply because of using (relatively) large sample size. For most "big data" applications, the number of observations for the whole data and training data can easily exceed few thousands. Consequently, most "typical" statistical variable selection procedures tends to overly select "less important" variables which are "statistically significant". RISE is a better strategy to choose and analyze various subdata of a "smaller size" that can combine or ensemble various results to build a more efficient and reliable model.

※ 實體與線上視訊同步進行。

※ 茶會：上午10：10開始。