STAT 400 Data Analysis Nov 2, 2005

Today we reviewed the univariate bootstrap – why the bootstrap was actually like sampling from the ecdf, and why smoothing the ecdf did not usually make much difference. We explained how the exercise related to the bootstrap idea, and showed how the "step" function in R worked on the output of the lm funtion.

No further notes are being posted, except the following e-mails about the exercise:

STAT 400 students

This is the assignment to be submitted by Wed, Nov 9, at class.

Stat 400 students are assigned a data set according to the following table: you are either group 1,2, or 3.

1	Au,Wai
1	Chan,Hok-Leung
1	Chau,Fong
1	Dai,Miao
1	Hui,Alvin
1	Liao,Chun-Kai
2	Lu,Monica
2	Poon,Danny
2	Schmidt,Nicole
2	Steininger,Gavin
2	Tong,Yee
3	Vaughan,Paul
3	Voo,Jessie
3	Wong,Kirby
3	Wu,Huanhuan
3	Yip,Yu
3	Zhang,Jinning

The data sets are attached. Just use read.table("men1") to get it once the file is in the R working directory. (Actually, data sets omitted in this posting).

The task is to use the bootstrap to study the stability (or variability) of the result

produced by the step procedure, using the men data. The dependent variable is Density and there are 14 predictor variables which could be used to predict Density.

STAT 400

I was asked for clarification of the exercise ....

1. Use Step on your original n=100 sample. Record subset of Variables chosen for model.
2. Generate a bootstrap sample (again n=100) from your original sample (n=100).
3. Use Step on the bootstrap sample. Record subset of variables chosen for the model.
4. Repeat 2. and 3. approx 25 times, always using the same original sample on which to base your bootstrap sample.

Compare your 25 models with the one you got in 1. How reliable was your original estimate of the model?

Note: Step operates on the output of the "lm" command, so you have to do that first, each time you use step.

KLW