

회 귀 분 석 II

이재용, 임요한

서울대학교
통계학과

2017년 8월

라소에서 유의성 검정:붓스트랩

```
library(glmnet)
library(ISLR)
Hitters=na.omit(Hitters)
dim(Hitters)

attach(Hitters)
x=model.matrix(Salary~.,Hitters)[,-1]
y=Hitters$Salary

#
# dimension of x 263 X 19
#

n=263
B=1000
best=matrix(0,B,20)
```

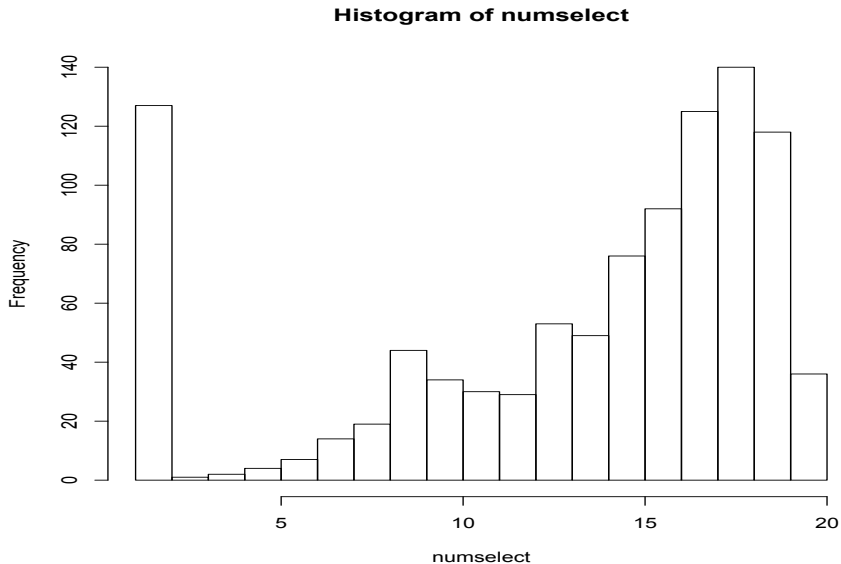
```
for(b in (1:B)){  
  
  bid=sample(n,n,replace=T)  
  bx=x[bid,]  
  by=y[bid]  
  grid=10^seq(4,-1,length=100)  
  cv.out=cv.glmnet(bx,by,alpha=1,lambda=grid)  
  blamb=cv.out$lambda.min  
  lasso.mod=glmnet(bx,by,alpha=1,lambda=exp(blamb))  
  best[b,]=as.vector(coef(lasso.mod))  
  
  cat("\t b=")  
  cat(b)  
}
```

```
lasso.mod=glmnet(x,y,alpha=1,lambda=exp(blamb))  
est=as.vector(coef(lasso.mod))  
se=sqrt(apply(best,2,var))  
tstat=est/se  
tstat
```

```
pvalue=2*(1-pnorm(abs(tstat)))  
pvalue
```

```
select=(best!=0)  
stab=apply(select,2,sum)/B  
numselect=apply(select,1,sum)  
hist(numselect)
```

선택변수



t-통계량, p-value, stability

```
>
> tstat
[1] 0.4700554 -1.2941157 1.6552804 0.0000000 0.0000000
[6] 0.0000000 1.7561824 -0.8324965 0.0000000 0.0000000
[11] 0.3713004 1.0455094 1.1459485 -1.0241761 0.6484224
[16] -2.2594369 1.9907399 0.4075420 -0.4004371 0.0000000
>
> pvalue
[1] 0.63831546 0.19562546 0.09786763 1.00000000 1.00000000
[6] 1.00000000 0.07905726 0.40512877 1.00000000 1.00000000
[11] 0.71041382 0.29578760 0.25181646 0.30575210 0.51671177
[16] 0.02385622 0.04650949 0.68360996 0.68883459 1.00000000
>
>
> stab
[1] 1.000 0.703 0.859 0.585 0.520 0.544 0.854 0.622 0.360 0.454
[11] 0.613 0.763 0.607 0.658 0.676 0.855 0.859 0.691 0.669 0.568
>
```