

Stat 471: Lecture 17

Bootstrap III: Regression

Moo K. Chung
mchung@stat.wisc.edu

October 17, 2003

- Continuing lecture 16, if f is not normal, T will be still a pivot but we do not know its distribution. So we generate 1000 bootstrap replicates of $T^* = \frac{\bar{X}^* - \bar{X}}{S^*/\sqrt{n}}$ of T and get the bootstrap CI

$$(\bar{x} - t_{1-\alpha/2}^* S/\sqrt{n}, \bar{x} - t_{\alpha/2}^* S/\sqrt{n}).$$

```
t=inline('(mean(x)-u)/(std(x)/
length(x))','x','u')
for i=1:1000;
    trep(i)=t(boot(arm,1),mean(arm))
end;
>> quantile(trep,[0.05 0.95])
-19.8246    22.0677
```

- For given bivariate data $(x_i, y_i), i = 1, \dots, n$ suppose we have model

$$Y_i = g(x_i; \beta) + \epsilon_i.$$

Let $g(x; \beta) = \beta_0 + \beta_1 x$. Fitting a somewhat crazy model like $g(x; \beta) = \frac{\beta_0}{\beta_1 - \beta_2 e^{-\beta_3 x}}$ is also possible. The least squares estimation (LSE) is given by

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n (y_i - g(x_i; \beta))^2$$

Based on LSE, the observed residual is

$$\hat{\epsilon}_i = y_i - g(x_i; \hat{\beta}). \quad (1)$$

A bootstrap dataset is then

$$Y_i^* = g(x_i; \hat{\beta}) + \epsilon_i^*. \quad (2)$$

The bootstrap estimate of β is then

$$\hat{\beta}^* = \arg \min_{\beta} \sum_{i=1}^n (y_i^* - g(x_i; \beta))^2. \quad (3)$$

- Let's find CI for β_1 for model

$$\text{grip} = \beta_0 + \beta_1 \text{arm}.$$

The pivot for β_1 is $T = \frac{\hat{\beta}_1 - \beta_1}{SE\hat{\beta}_1}$, where $SE\hat{\beta}_1 = \sqrt{\text{Var}\hat{\beta}_1}$. If you don't know the standard error SE , we estimate it, i.e. $T = (\hat{\beta}_1 - \beta_1)/S_{\hat{\beta}_1}$. Following eq.1,2 and 3 we generate 1000 bootstrap replicates of $T^* = \frac{\hat{\beta}_1^* - \hat{\beta}_1}{SE\hat{\beta}_1^*}$. We need to estimate $SE\hat{\beta}_1^*$ via bootstrap which requires bootstrapping from bootstrapped residual e_i^* (there is a formula for computing the standard error but well let's forget that). Then $100(1 - \alpha)\%$ CI for β_1 is

$$(\hat{\beta}_1 - t_{1-\alpha/2}^* S_{\hat{\beta}_1}, \hat{\beta}_1 - t_{\alpha/2}^* S_{\hat{\beta}_1}).$$

```
beta=inline('pinv([ones(147,1) x])*y')
>>b=beta(arm,grip)
b=
    54.7081
     0.7050
e=grip-b(1)-b(2)*arm;
for j=1:1000
    bse=boot(e,1);
    bsgr1p = bse + b(1) + b(2)*arm;
    bsb=beta(arm,bsgr1p);
    for i=1:50
        bs2e=boot(bse,1);
        bs2gr1p = bs2e + b(1) + b(2)*arm;
        bs2b=beta(arm,bs2gr1p);
        bsbeta(i)=bs2b(2);
    end;
    trep(j)=(bsb(2)-0.7050)/std(bsbeta);
end;
>>quantile(trep,[0.05 0.95])
-1.5308    1.6580
```