How do I write my own bootstrap program?

Authored by
stats writer

July 1, 2024

A bootstrap program is a set of instructions that allow a computer to load and initialize its operating system. To write your own bootstrap program, you will need a basic understanding of computer architecture and assembly language. First, you will need to determine the specific hardware and processor specifications of your computer. Then, you will need to write the necessary code in assembly language to initialize the hardware and load the operating system into memory. This can be a complex process, so it is recommended to consult official documentation and resources for your specific hardware and programming language. Once the bootstrap program is completed, it can be loaded onto a bootable device, such as a USB drive, and used to start up the computer.

How do I write my own bootstrap program? | Stata FAQ

Stata has the convenient feature of having a bootstrap prefix command which can be seamlessly incorporated with estimation commands (e.g., logistic regression or OLS regression) and non-estimation commands (e.g., summarize). The bootstrap command automates the bootstrap process for the statistic of interest and computes relevant summary measures (i.e., bias and confidence intervals). As convenient as this command is, however, there are instances when the statistic you want to bootstrap does not work within the command. For such instances, you need to write your own bootstrap program.

This Stata FAQ shows how to write your own bootstrap

program. For the first example, we match results from the bootstrap command with results from writing a bootstrap program. Ideally, this should reveal how simple it is to write your own bootstrap program. This is followed by an example in which the statistic you want to bootstrap does not work within the bootstrap command, and therefore, requires you to write your own bootstrap program.

Example 1

This example we use the bootstrap command and replicate the results by writing our own bootstrap program. We use the High School and Beyond dataset from which we are going to regress gender (female), math score (math), writing score (write) and socio-economic status (ses) on reading score (read) and bootstrap the root mean squared error (rmse). For the bootstrap we do 100 replications and specify the seed so that we can replicate the results.

use http://statistics.ats.ucla.edu/stat/stata/notes/hsb2, clear
(highschool and beyond (200 cases))

regress read female math write ses

Source | SS df MS Number of obs = 200
-------------+------------------------------ F(4, 195) = 52.58
Model | 10854.9318 4 2713.73294 Prob > F = 0.0000
Residual | 10064.4882 195 51.6127602 R-squared = 0.5189
-------------+------------------------------ Adj R-squared = 0.5090
Total | 20919.42 199 105.122714 Root MSE = 7.1842

----
read | Coef. Std. Err. t P>|t|
-------------+--
female | -2.450171 1.101524 -2.22 0.027 -4.622602 -.2777409
math | .4565641 .0721114 6.33 0.000 .3143457 .5987825
write | .3793564 .0732728 5.18 0.000 .2348475 .5238653
ses | 1.301982 .7400719 1.76 0.080 -.1575905 2.761555

_cons | 6.833418 3.279371 2.08 0.038 .3658287 13.30101

----

bootstrap rmse=e(rmse), reps(100) seed(12345): regress read female math write ses

(running regress on estimation sample)

Bootstrap replications (100)

----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5

.... 50

.... 100

Linear regression Number of obs = 200

Replications = 100

command: regress read female math write ses

rmse: e(rmse)

| Observed Bootstrap Normal-based

| Coef. Std. Err. z P>|z|

-------------+--

rmse | 7.184202 .2594069 27.69 0.000 6.675774 7.69263

----

estat bootstrap, all

Linear regression Number of obs = 200

Replications = 100

command: regress read female math write ses

rmse: e(rmse)

--

| Observed Bootstrap

| Coef. Bias Std. Err.

-------------+--

rmse | 7.1842021 -.1006956 .25940687 6.675774 7.69263 (N)

| 6.559784 7.636096 (P)

| 6.778425 7.741319 (BC)

--

(N) normal confidence interval

(P) percentile confidence interval

(BC) bias-corrected confidence interval

Writing our own bootstrap program requires four steps.

use https://stats.idre.ucla.edu/stat/stata/notes3/hsb2, clear

(highschool and beyond (200 cases))

```
*Step 1
quietly regress read female math write ses
matrix observe = e(rmse)

*Step 2
capture program drop myboot
program define myboot, rclass
preserve
bsample
regress read female math write ses
return scalar rmse = e(rmse)
restore
end*Step 3
simulate rmse=r(rmse), reps(100) seed(12345): myboot

command: myboot
rmse: r(rmse)

Simulations (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.................................................. 50
.................................................. 100

*Step 4
bstat, stat(observe) n(200)
```

Bootstrap results Number of obs = 200

Replications = 100

--

| Observed Bootstrap Normal-based

| Coef. Std. Err. z P>|z|

-------------+--

rmse | 7.184202 .2594069 27.69 0.000 6.675774 7.69263

--

estat bootstrap, all

Bootstrap results Number of obs = 200

Replications = 100

--

| Observed Bootstrap

| Coef. Bias Std. Err.

-------------+--

rmse | 7.1842021 -.1006956 .25940687 6.675774 7.69263 (N)

| 6.559784 7.636096 (P)

| 6.778425 7.741319 (BC)

--

(N) normal confidence interval

(P) percentile confidence interval
(BC) bias-corrected confidence interval

The results from Step 4 match the results from the bootstrap command in the example above.

Example 2

In this example we write a bootstrap program where the usual bootstrap command does not accommodate the statistic we want to bootstrap. The reason why the bootstrap command does not accommodate all situations is because the bootstrap command requires a statistic that falls directly out of the "analysis" command. To see what statistics are accommodated, use either the ereturn list or return list command following the "analysis" command. The distinction between ereturn list or return list depends whether the "analysis" command is an estimation

command or not.

Suppose we want to bootstrap the variance inflation factor (vif),
which requires us to run regress and then estat vif.
In such a situation, the statistic to bootstrap falls out from a post
estimation command, which is not obtainable from regress and therefore not accommodated by
the bootstrap command. Hence, we must
write our own bootstrap program to get a bootstrap estimate of the vif.

use https://stats.idre.ucla.edu/stat/stata/notes3/hsb2, clear
(highschool and beyond (200 cases))

*Step 1
quietly regress read female math write sesestat vif

Variable | VIF 1/VIF
-------------+----------------------
write | 1.86 0.537690
math | 1.76 0.568278
female | 1.17 0.857692

ses | 1.11 0.902671

-------------+----------------------

Mean VIF | 1.47

return list

scalars:

r(vif_4) = 1.107823014259338

r(vif_3) = 1.165920257568359

r(vif_2) = 1.759701371192932

r(vif_1) = 1.859809398651123

macros:

r(name_4) : "ses"

r(name_3) : "female"

r(name_2) : "math"

r(name_1) : "write"

matrix vif = (r(vif_4), r(vif_3), r(vif_2), r(vif_1))

matrix list vif

vif

c1 c2 c3 c4

r1 1.107823 1.1659203 1.7597014 1.8598094

*Step 2

```
capture program drop myboot2
program define myboot2, rclass
preserve
bsample
regress read female math write ses
estat vif
return scalar vif_4 = r(vif_4)
return scalar vif_3 = r(vif_3)
return scalar vif_2 = r(vif_2)
return scalar vif_1 = r(vif_1)
restore
end

*Step 3
simulate vif_4=r(vif_4) vif_3=r(vif_3) vif_2=r(vif_2) vif_1=r(vif_1), ///
reps(100) seed(12345): myboot2

command: myboot2
vif_4: r(vif_4)
vif_3: r(vif_3)
vif_2: r(vif_2)
vif_1: r(vif_1)

Simulations (100)
```

----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5

.... 50

.... 100

bstat, stat(vif) n(200)

Bootstrap results Number of obs = 200

Replications = 100

--

| Observed Bootstrap Normal-based

| Coef. Std. Err. z P>|z|

-------------+--

vif_4 | 1.107823 .0344814 32.13 0.000 1.040241 1.175405

vif_3 | 1.16592 .0524449 22.23 0.000 1.06313 1.26871

vif_2 | 1.759701 .1349314 13.04 0.000 1.495241 2.024162

vif_1 | 1.859809 .1467453 12.67 0.000 1.572194 2.147425

--

estat bootstrap, all

Bootstrap results Number of obs = 200

Replications = 100

--

| Observed Bootstrap

| Coef. Bias Std. Err.

-------------+--

vif_4 | 1.107823 .0127056 .03448139 1.040241 1.175405 (N)

| 1.061917 1.197667 (P)

| 1.058617 1.172653 (BC)

vif_3 | 1.1659203 .0285308 .0524449 1.06313 1.26871 (N)

| 1.10246 1.30328 (P)

| 1.08448 1.255424 (BC)

vif_2 | 1.7597014 .0305828 .13493139 1.495241 2.024162 (N)

| 1.552449 2.081403 (P)

| 1.510279 2.026165 (BC)

vif_1 | 1.8598094 .0389828 .14674526 1.572194 2.147425 (N)

| 1.665374 2.196174 (P)

| 1.633619 2.160758 (BC)

--

(N) normal confidence interval

(P) percentile confidence interval

(BC) bias-corrected confidence interval