

ESM 206B

UCSB Bren School

April 2009

Hampton

Bootstrapping Lab

In this lab, you will do some simple bootstrapping in R to construct confidence intervals and interpret the meaning of these results.

The scenario is that a researcher has collected data on the feeding preferences of predatory aquatic insects, with the goal of determining which predator may make the most suitable biocontrol agent for mosquitoes at the larval stage. Mosquito larvae grow in aquatic environments, in tree holes, the hollows formed by discarded tires, small ponds, puddles, slow rivers – any area of slow-moving or stagnant water where adults can lay eggs. At this stage of the life cycle, the mosquitoes are aggregated and fairly vulnerable to targeted control. There is interest in the use of aquatic insects as biocontrol predators for mosquitoes – the insects are highly mobile, they can fly among different aquatic environments and then hunt for prey underwater. They can leave to find another pond when prey are depleted. They are native and abundant in many areas of the world where mosquitoes are particularly problematic in the spread of disease. Many aquatic insects (particularly the hemipterans or “true bugs”) are “wasteful killers”, meaning that they will kill far more prey than they can possibly eat, making them potentially very effective against mosquito larvae which can be extremely numerous in slow-moving water.

Standard statistics packages do not include bootstrapping, so we will work in R. You will need a special library for R called “boot” – it should be on the lab machines, but if you are working on this at home, go to the end right now and find instruction on loading new “packages” into R.

1. Place 4 files into a folder called “Mosquitoes” on your desktop: anisops.csv, diplonychus.csv, notonecta.csv, and Boot.R
2. Open R
3. Change Directory to point to the folder called “Mosquitoes”. In windows, this is File -> Change dir; (on a mac, this is Misc -> Change Working Directory) select the “Mosquitoes” folder so R knows where to look for your files.
4. File ->Open Script -> Boot (on mac, File-open document-boot)
5. Look at Boot.R – everything behind a # is just a “comment” – it’s for you to read, but R ignores it
6. Read through the first example that uses the predator Anisops, to see what you will be doing in the exercise
7. Load the “boot” library, and the data and look at the first 5 lines of data to see that it loads properly– highlight everything between the very top of the script and “head(anisops,5)” – right click and “Run” (or Copy and Paste into the Console window and hit Enter) – just be sure that you

don't get error messages, hail Nick if you do – errors may relate either to loading the “boot” library or the data

8. Highlight and run the next block of code down to “mean(anisops)”
9. Write down the mean and confidence limits calculated in the traditional (“classical”) way.
10. Highlight the next block of text and run it, to make R bootstrap the data 999 times – nothing should happen, if you do not have an error message move on. Hail Nick if you do.
11. Highlight and run the next block of text to make a histogram of your bootstrapped distribution of the mean preference of Anisops for mosquito larvae. (*On a mac, you may need to resize the window in order to see the whole histogram. Just grab the corner and resize it, larger, smaller till you see values on both axes.*) Looking at the histogram, do you think that Anisops prefers mosquito larvae over chironomid larvae?
12. Highlight and run the next block of text to look at bootstrapped confidence limits and the bootstrapped mean.
  - a. Write down the mean.
  - b. boot.ci reports 4 different methods of calculating confidence limits. We talked about the “percentile” method in class. Write down the upper and lower confidence limits derived from the percentile bootstrap approach.
13. Repeat this exercise with the other two data sets for two other predators: *diplonychus.csv* and *notonecta.csv*. We have already modified the code for these two data sets, shown in *Boot.R*, below the code you just ran.
14. Which predator is likely to be the better biocontrol agent? What makes you come to this conclusion?
15. Do the predators appear to differ significantly from each other?
16. Did the bootstrapped confidence limits differ much from the “classical” confidence intervals? Why might bootstrapping be preferred over the classical approach?

*Installing a package in R.* Use this if the command “library(boot)” doesn't work. The library/package “boot” has tools for bootstrapping that are used in the code you have for lab.

In Windows. Packages -> Install Package -> USA (CA1 or CA2) -> boot

In Mac. Packages & Data (choose USA CA1 or CA2) – search for “boot”, and Get List; select “boot”, Install Selected