Bootstrapping in R

Dave Evans
Dave.Evans.rv2l@statefarm.com

Bootstrapping

- Bootstrap:
  - Start with a sample of data and an estimator
  - Sample from original data with replacement
  - Calculate estimator
  - Repeat
- Obtains a sampling distribution of the estimator
- Useful when the theoretical distribution is unknown, or complicated to estimate.

The boot function

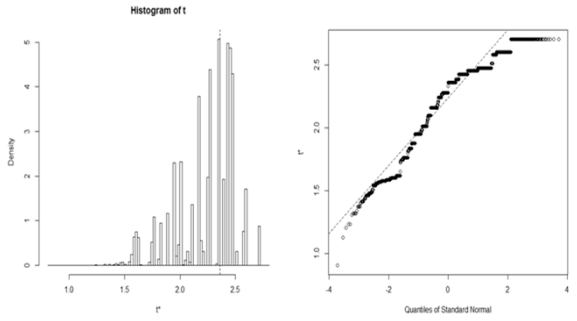
- Syntax in its simplest form is:
  - boot(data, statistic, R , ….)
  - statistic must be a function of at least 2 arguments, data and indices
  - R is the number of bootstrap replicates
  - … is where additional arguments to the statistic function are passed
- Must download the boot package and call library(boot)

Boot example

```
> library(boot)
> x <- rexp(50)
> bootpercentile <- function(x,i,p) quantile(x[i],p)
> b <- boot(x, bootpercentile, 10000, p=.95)
> b

ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = x, statistic = bootpercentile, R = 10000, p = 0.95)


Bootstrap Statistics :
    original     bias    std. error
t1* 2.361791 -0.1208445    0.269644
```

Boot example

> Plot(b)

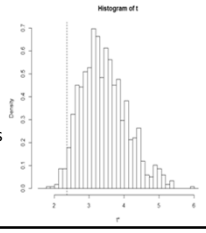


Boot confidence intervals

```
> boot.ci(b, .95)
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = b, conf = 0.95)

Intervals :
Level      Normal              Basic
95%   ( 1.954,  3.011 )   ( 2.122,  3.136 )

Level     Percentile            BCa
95%   ( 1.588,  2.602 )   ( 1.620,  2.706 )
Calculations and Intervals on Original Scale
```

Parametric Bootstrap

```
> y <- data.frame(list(sample = x))
> expboot <- function(x,B) {
+ y <- x
+ y$sample <- rexp(nrow(y),B)
+ y }
> bootpercentile2 <- function(y,p) {
+ quantile(y$sample,p) }
> b2 <- boot(y, bootpercentile2, R = 999,
+ sim = "parametric", ran.gen = expboot,
+ mle = mean(y$sample), p=.95)
> plot(b2)
```



Creates bootstrap samples using an assumed distribution

Must include a function to estimate parameters

Can provide a more complete distribution, especially for percentile type measurements

Other features of boot

- Boot contains many outputs
  - `> b$t[1:5,]`
    `[1] 2.064531 2.361791 2.241798 2.160639 1.762839`
  - `> b$seed[1]`
    `[1] 403`
- Multivariate bootstrapping –R will treat each row as a multivariate observation when using matrices or data frames
- Balanced, permutation and antithetic simulations
- Importance sampling

- Boot documentation: http://cran.r-project.org/web/packages/boot/boot.pdf
- Resampling Methods in R: The boot Package: http://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf
  - By Angelo J Canty author of the original boot Package for S

Authors of boot:

Angelo Canty and Brian Ripley (2010). boot: Bootstrap R (S-Plus) Functions. R package version 1.2-43.