Homework 8

Zachary Foster

12/02/2014

```
library(boot)
library(ggplot2)
library(plyr)
library(reshape2)
```

```
# Read input data ------------------------------------------------------------
data <- read.table("emap.build08.txt", header = TRUE)
data <- data[ , !(names(data) %in% c("LAT", "LON"))]
# # Log varaibles ------------------------------------------------------------
# vars_to_log <- c("LK.HA", "POPDENKM", "TOT.RD")
# data[vars_to_log] <- lapply(data[vars_to_log], function(x) log(x + 1))
# names(data) <- ifelse(names(data) %in% vars_to_log,
#                       paste("LOG", names(data), sep = "."),
#                       names(data))
```

```
count <- 100000
get_median <- function(data, index) median(data[index])
boot_info <- boot(data$LK.HA, get_median, count)
boot_info
```

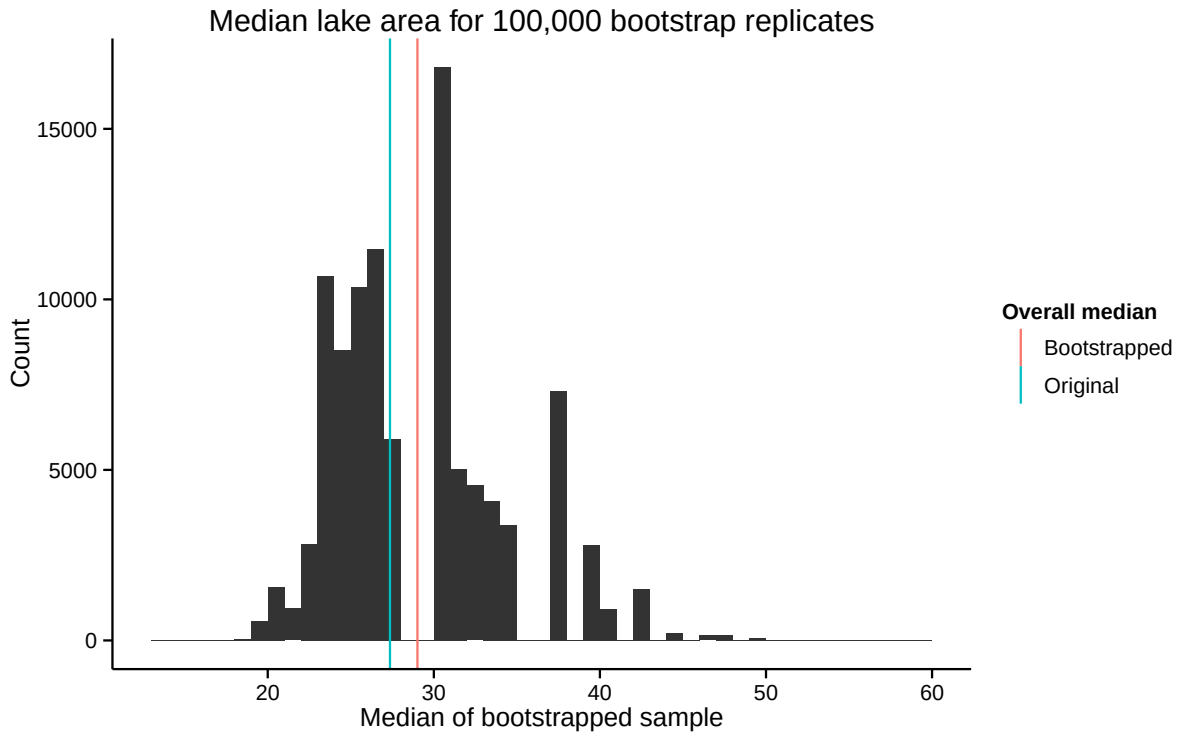
```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = data$LK.HA, statistic = get_median, R = count)
##
##
## Bootstrap Statistics :
##     original  bias    std. error
## t1*    27.36   1.654       5.196
```

```
original_median <- median(data$LK.HA)
boot_median <- mean(boot_info$t)
boot_bias <- boot_median - original_median
```

The original median is 27.36 and the mean median of 105 bootstrap replicates is 29.0141, indicating a bias of 1.6541.

```
graph_data <- data.frame(median = boot_info$t)
vline_data <- data.frame(label = c("Original", "Bootstrapped"),
                         value = c(original_median, boot_median))
ggplot(graph_data, aes(x = median)) +
  geom_histogram(binwidth = 1) +
  geom_vline(data = vline_data,
```

```
            aes(xintercept = value, color = label),
            show_guide = TRUE) +
  labs(title = "Median lake area for 100,000 bootstrap replicates",
       x = "Median of bootstrapped sample",
       y = "Count",
       color = "Overall median") +
  theme_classic()
```




```
boot.ci(boot_info, type = c("perc", "bca"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 100000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_info, type = c("perc", "bca"))
##
## Intervals :
## Level     Percentile            BCa
## 95%   (21.63, 40.01 )   (20.64, 39.49 )
## Calculations and Intervals on Original Scale
```

```
log_lake_area <- log(data$LK.HA + 1)
error <- qnorm(0.975)*sd(log_lake_area)/sqrt(length(log_lake_area))
ci <-  c(mean(log_lake_area) - error, mean(log_lake_area) + error)
exp(1)^ci - 1
```

```
## [1] 27.73 44.08
```