

Jahiya Clark  
Prof. Nelson  
Lab 8: Bootstrapping  
11/7/2021

## Lab 8: Bootstrapping

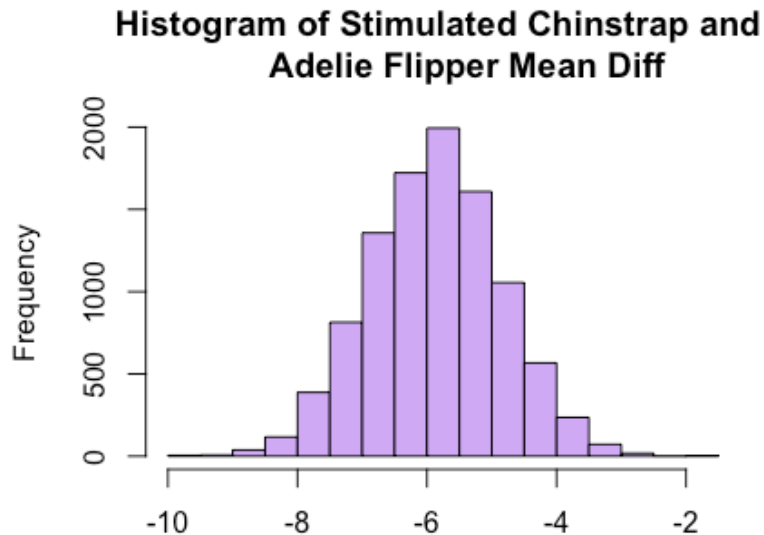
1. Mean is 5.84945

```
require(palmerpenguins)
penguin_dat = droplevels(subset(penguins, species != "Gentoo"))

pen_boot=two.boot(
  subset(penguin_dat, species == "Adelie")$flipper_length_mm,
  subset(penguin_dat, species == "Chinstrap")$flipper_length_mm,
  FUN = mean,
  R = 10000,
  na.rm = TRUE)

boot.ci(pen_boot)
mean(pen_boot$t, na.rm = TRUE)
```

- 2.



3. The 95% bootstrap CI is (-7.839, -3.899 )  

```
boot.ci(pen_boot)
quantile(pen_boot$t, 0.025)
```
4. No, I do not think the resampled differences in means follow a skewed distribution because the histogram is bell shaped. In addition, the actual mean of the data is 5.869887 which is

still quite close to my stimulated mean. The median of the stimulated data is 5.862229 which is also very close to the mean which shows there is not much variance between the samples.

5.

```
pen_ecdf <- ecdf(pen_boot$t)
```

6. 8.95% probability of observing a mean difference of -4.5 or greater  
`1 - pen_ecdf[-4.5]`

7. 1.67% probability of observing a mean difference of -8 or smaller  
`pen_ecdf[-8]`

8. Alternative Hypothesis: There is a difference of mean flipper length between Adelie penguins and Chinstrap penguins.

Null Hypothesis: There is no difference of mean flipper length between Adelie penguins and Chinstrap penguins.

9. p-value = 0.1005

```
veg = read.csv(here["data", "vegdata.csv"])
```

```
dat_veg = droplevels(subset(veg, treatment %in% c("control", "clipped")))
```

```
wilcox.test(pine ~ treatment, data = dat_veg, alternative = "two.sided")
```

10. 4.125 to 29.750

```
tree_boot =
```

```
  two.boot(  
    subset(dat_tree, treatment == "clipped")$pine,  
    subset(dat_tree, treatment == "control")$pine,  
    FUN = mean,  
    R = 10000,  
    na.rm = TRUE  
  )
```

```
quantile(tree_boot$t, c(0.025, 0.975))
```

11. 16

```
clip_mean <- mean(subset(dat_tree, treatment == "clipped")$pine)
```

```
con_mean <- mean(subset(dat_tree, treatment == "control")$pine)
```

```
clip_mean - con_mean
```

12. The Simpson's diversity index measures diversity with the parameters around species richness and relative abundance of each species. This index produces 1 to fit infinite diversity and 0 to fit no diversity, so we aim for values closer to 1.

13.

```
dat_bird = read.csv(here("data", "bird.sub.csv"))
dat_hab = read.csv(here("data", "hab.sub.csv"))

dat_all = merge(
  dat_bird,
  dat_hab,
  by = c("basin", "sub"))

s_sidi_mean <- mean(dat_all$s.sidi, na.rm = TRUE)
s_sidi_sd <- sd(dat_all$s.sidi, na.rm = TRUE)

dat_all$s.sidi.standardized <- (dat_all$s.sidi - s_sidi_mean)/s_sidi_sd
```

14.

```
dat_1 =
  subset(
    dat_all,
    select = c(b.sidi, s.sidi))
index_1 = sample(nrow(dat_1), replace = TRUE)
index_2 = sample(nrow(dat_1), replace = TRUE)

m = 10000
result = numeric(m)

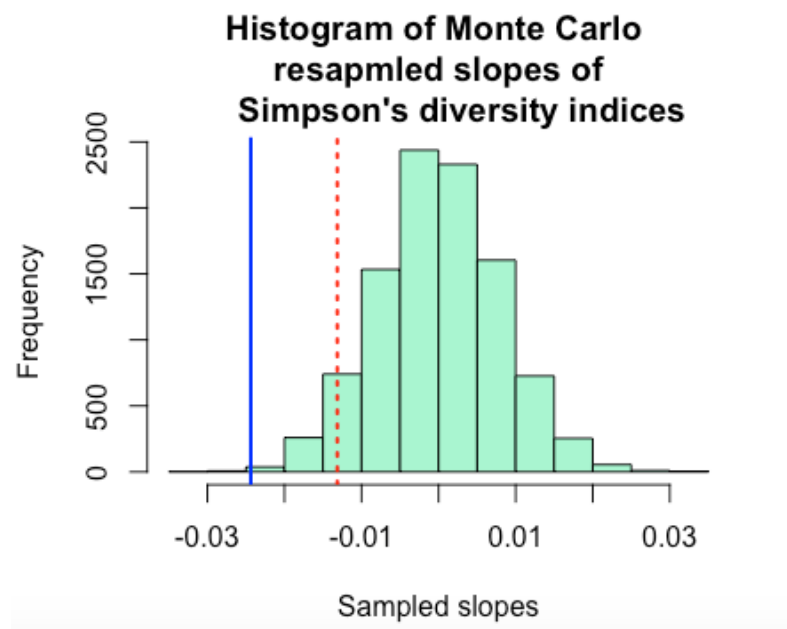
for(i in 1:m)
{
  index_1 = sample(nrow(dat_1), replace = TRUE)
  index_2 = sample(nrow(dat_1), replace = TRUE)

  dat_resampled_i =
    data.frame(
      b.sidi = dat_1$b.sidi[index_1],
      s.sidi = dat_1$s.sidi[index_2]
    )

  fit_resampled_i = lm(b.sidi ~ s.sidi, data = dat_resampled_i)

  result[i] = coef(fit_resampled_i)[2]
}
```

15.



16. My critical value was -0.01314 and yes it was bigger than my observed slope of -0.02437.
17. My conclusion about the negative relationship between vegetation cover diversity and bird diversity might be explained by habitat capacity. Birds are preferential feeders so birds might prefer to be present in areas that only have the plant species they want.