Keegan Moynahan

Lab 8

1) Sd of differences in mean = 1.012

install.packages("simpleboot")

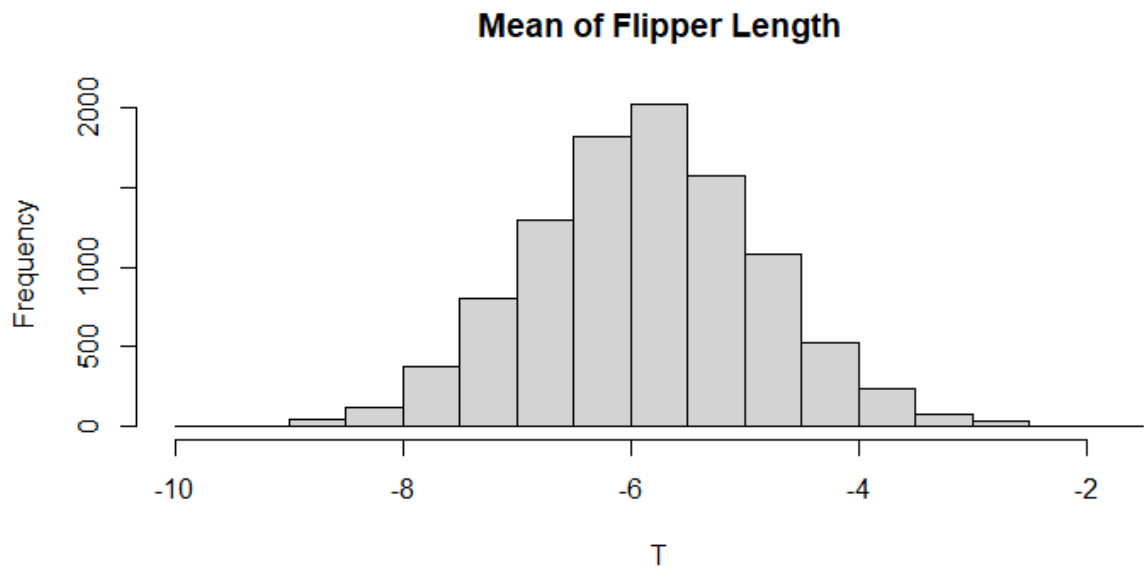library(simpleboot)

x = droplevels(subset(dat_pen, species != "Chinstrap"))

y = droplevels(subset(dat_pen, species != "Adelie"))

pen_boot <- two.boot(

 na.omit(x$flipper_length_mm), na.omit(y$flipper_length_mm), mean,

 10000)

sd(pen_boot$t)

2)

Mean of Flipper Length

3) The 95% CI is   2.5% =- 7.819376, 97.5% = -3.839638

quantile(pen_boot$t, probs = c(0.025, 0.975))

4) The mean is -5.863 and the median is 5.860. Following the histogram with these values of mean and median, they tell me the data isn't skewed one way, the data is pretty evenly distributed.
5)   pen_ecdf = ecdf(pen_boot$t)
6) The probability of -4.5 of greater is 0.0866
    1 - pen_ecdf(-4.5)
7) The proability of -8 of less is 0.0156

pen_ecdf(-8)

8) The null hypothesis would be there is no difference in flipper length between the two species of penguins. The alternative hypothesis would be that there is a difference between the flipper length of the two species

9) P-value = 0.1005

```
require(here)

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

head(veg)

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

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

10) The end points of the 95% CI are   2.5% 3.996875   97.5% 30.259375

```
x = droplevels(subset(veg_treat, treatment == "clipped"))

y = droplevels(subset(veg_treat, treatment == "control"))

tree_boot <- two.boot(x$pine, y$pine, mean, 1000)

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

11) The observed difference is 16 which falls in between the CI.

12) The Simpson diversity index has a range of 0-1, if a high score is given it indicates high diversity and a low score indicates low diversity. A score of 0 means there is only one species.

13)

```
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)

```
m = 10000

result_mc = numeric(m)


for(i in 1:m)

{

 index_1 = sample(nrow(dat_1), replace = TRUE)
```
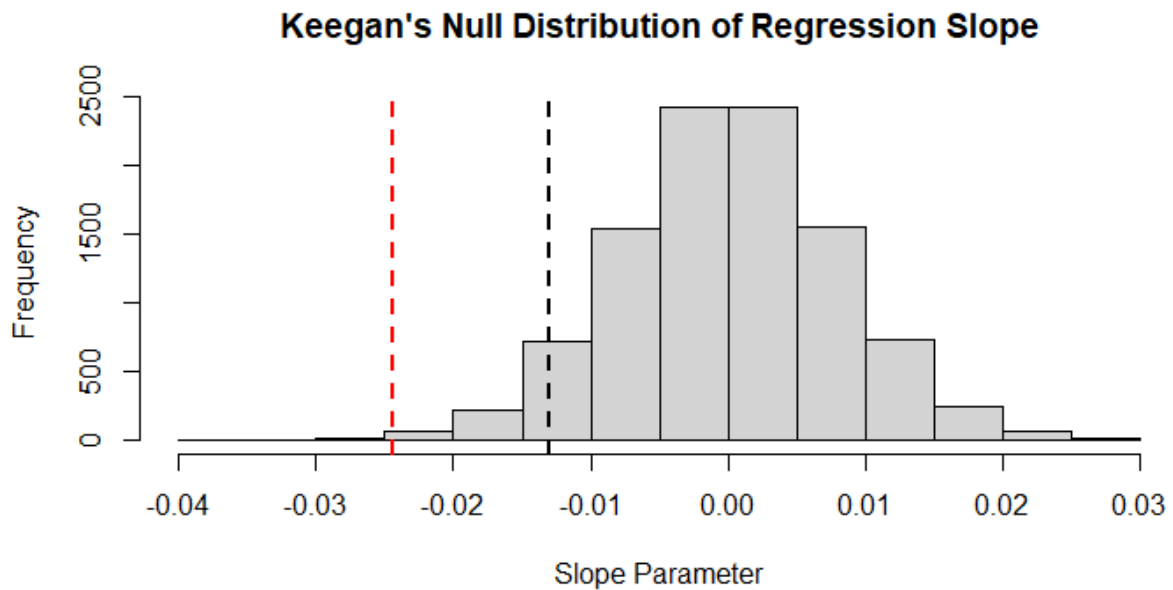
```
index_2 = sample(nrow(dat_1), replace = TRUE)


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)
 slope_resampled_i = coef(fit_resampled_i)[2]
 result_mc[i] = coef(fit_resampled_i)[2]
}
```

15)

Red = observed slop black = critical value

16) The critical value is -0.0129648 which is more than the observed slope
17) Because there is a negative observed slope and critical value and a negative relationship between critical value and slope observed that would mean that the less vegetation means less birds seen. I would also reject the null hypothesis and accept the alternative hypothesis because of the observed slope being more extreme than the critical value.
18)

```
m = 10000
result_boot = numeric(m)


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

  dat_boot = dat_1[index_1, ]
  head(dat_boot)
  fit_bs1 = lm(b.sidi ~ s.sidi, data = dat_boot)


  coef(fit_bs1)


  result_boot[i] = coef(fit_bs1)[2]
}
```
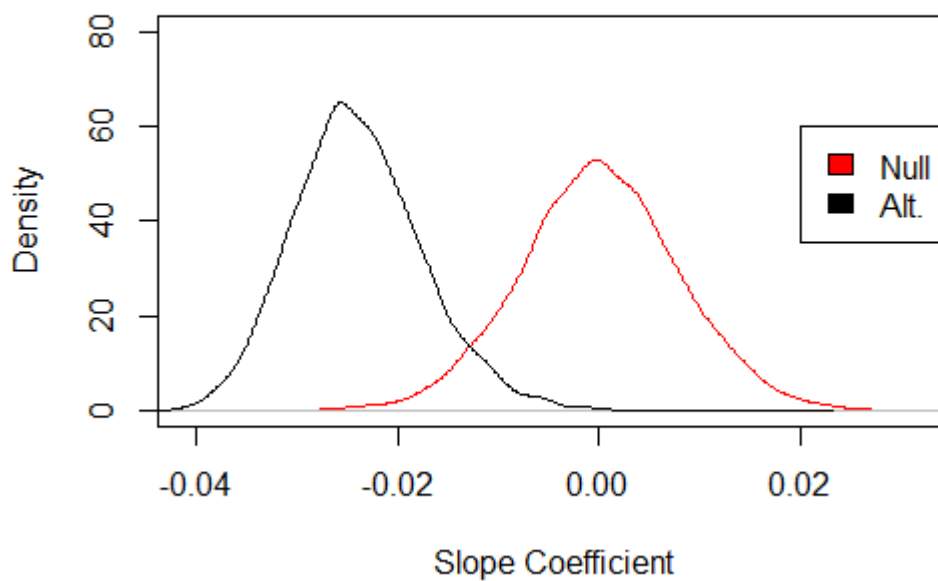
19)

20) I would reject the null hypothesis for the area that falls under both the null and alternative hypothesis. I would accept the alternative hypothesis because of the observed slope. The alternative hypothesis also creates a more normally distribution.