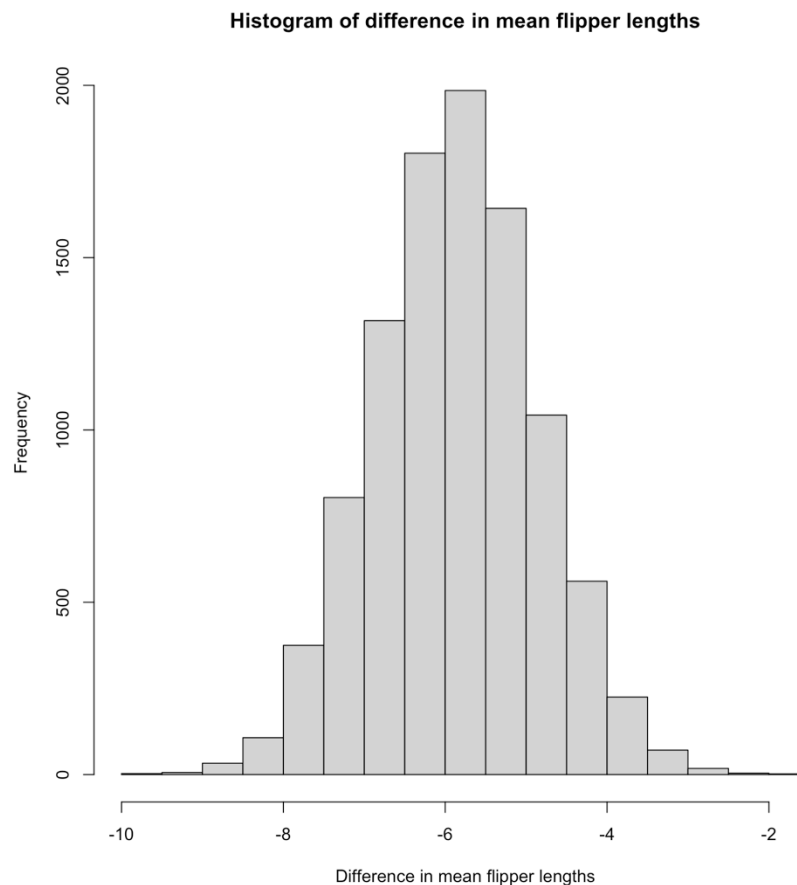Matt Fertakos w/ input from Bonnie

Question 1:
pen_boot = two.boot(sample1=adelfliplength, sample2=chinfliplength, FUN=mean,
R=10000,na.rm=TRUE)
sd(pen_boot$t)
1.007668

Question 2:

Histogram of difference in mean flipper lengths

Difference in mean flipper lengths

Question 3:
quantile(pen_boot$t,c(0.025,0.975),na.rm=TRUE)
        2.5%   97.5%
CI is -7.816329 to -3.904232

Question 4:
No, I do not think my resampled differences in means follow a skewed distribution. This is because the mean (-5.86) and median (-5.85) are close in value, and the histogram of the mean differences appears normally distributed. If this was a skewed distribution the bars of the

histogram would concentrate on one side or the other, and the median and mean would be further apart in value from each other.

Question 5:
pen_ecdf<-ecdf(pen_boot$t)

Question 6:
1-pen_ecdf(-4.5)
#0.0881

Question 7:
pen_ecdf(-8)
#0.0149

Question 8:
Null: There is no difference in mean flipper lengths between the two penguin species.
Alternative: There is a significant difference in mean flipper lengths between the two penguins.

Question 9:
veg = read.csv(here("eco_634_2021","data","vegdata.csv"))
dat_tree = droplevels(subset(veg, treatment %in% c("control", "clipped")))
wilcox.test(data=dat_tree,pine~treatment)
#p=0.1005

Question 10:
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))
#4.25 to 30.00

Question 11:
mean(tree_boot$t)
The observed difference in mean tree counts was 16.01 and it does fall within the range of the 95% bootstrap CI

Question 12:
The Simpson diversity index is a quantification of species diversity based on both the number of

species and species abundance. Outputs span from 0 to 1 in which 0 means no diversity and 1 means infinite diversity.

Question 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
```

Question 14:

```
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]
}
```
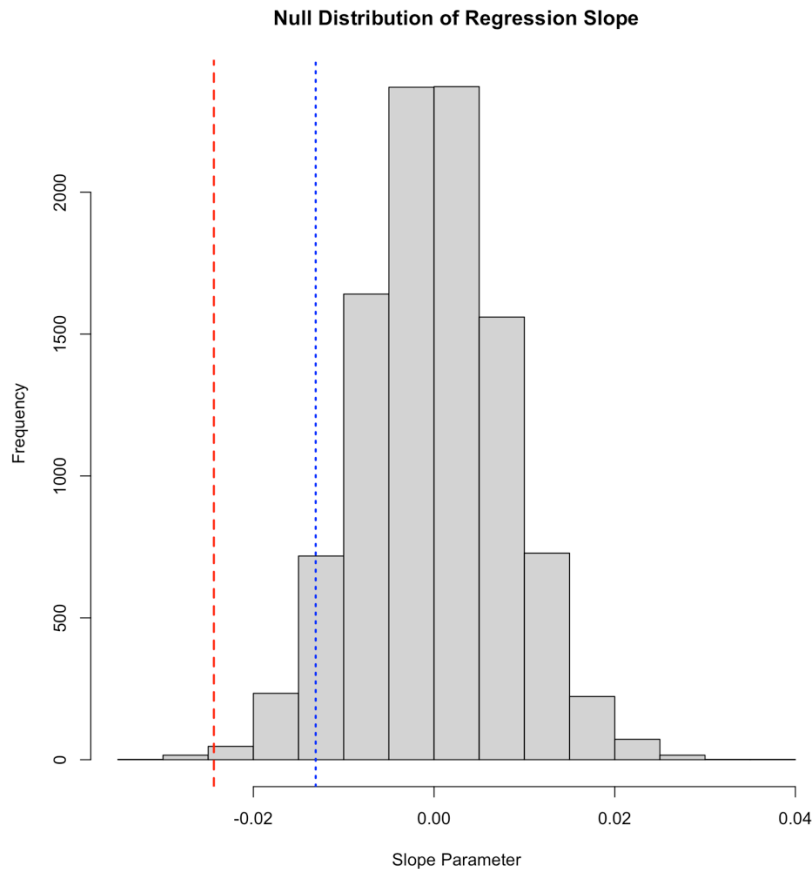
Question 15:
Blue line is the critical slope, red line is the observed slope.

Null Distribution of Regression Slope

Question 16:
critical_slope <- quantile(result, c(.05))
-0.01309726 is the critical slope, which is greater than the observed slope of -0.02437.

Question 17:
There is a significant negative relationship between vegetation cover diversity and bird diversity. This is supported as true because our observed slope (red line in histogram) is less than the critical slope (blue line in histogram) created through the randomization loop. In the slopes outputted by the 10000 replicate randomization loop, only 17 are equal to or less than the observed slope. This means there is a p value of 0.0017, which is significant. There is an extremely small chance on randomization of the data we would observe the slope we observed.