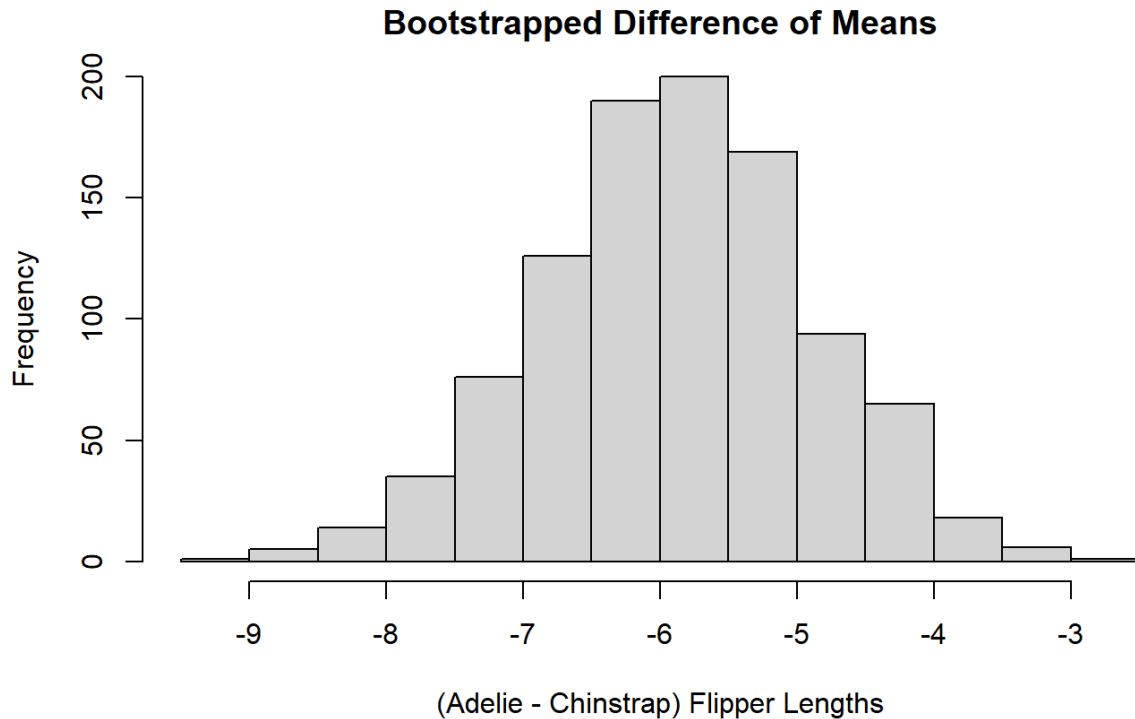


Ian Eggleston

Worked with Sarah Guitart

1. SD = 0.994549



- 2.
3. Confidence Interval: 2.5% = -7.941875 97.5% = -4.025236
`quantile(pen_boot$t, c(0.025, 0.975))`
4. The resampled differences do not suggest a skewed data set because the histogram shows a normal distribution around the median, which is also very similar to the mean.
5. `pen_ecdf = ecdf(pen_boot$t)`
6. 0.0857 8.57%
`1 - pen_ecdf(-4.5)`
7. 0.0182 1.82%
`pen_ecdf(-8)`
8. The null hypothesis states that there is no difference in the mean flipper lengths between chinstraps and adelie penguins. The alternative hypothesis states that there is a relationship between the two species' flipper lengths, but we cannot restrict it to either greater or less than.
9. P = 0.1005
`table(dat_tree$treatment)`
`dat_tree_control = na.omit(subset(dat_tree, treatment == "clipped")$pine)`
`dat_tree_clipped = na.omit(subset(dat_tree, treatment == "control")$pine)`
`wilcox.test(pine ~ treatment, data = dat_tree)`
10. Confidence interval
2.5% 3.871875 97.5% 29.75
11. 16 is the difference between means and does fall within the CI.

12. The Simpson diversity index quantifies the diversity of a group of organisms. It compares the number of one species to the total number of organisms to see the relative amounts of each species. This gives us an idea of how species diverse the area is.

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
mean(dat_all$s.sidi.standardized)
sd(dat_all$s.sidi.standardized)
```

14.

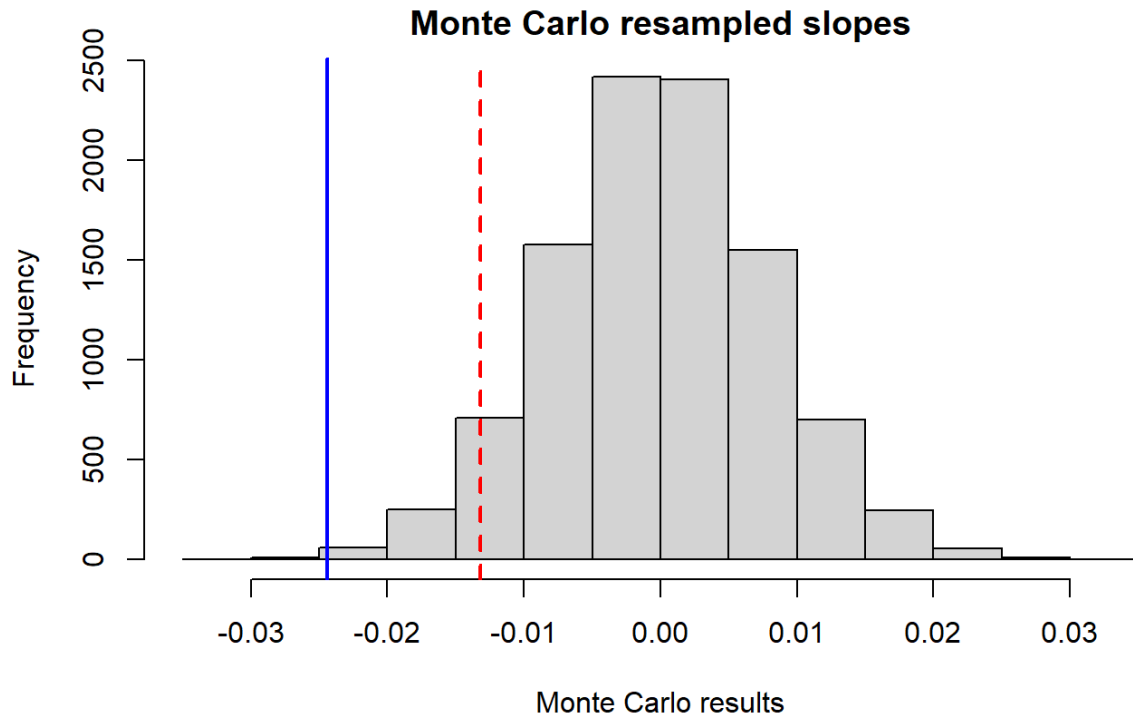
```
m = 10000
result_mc = numeric(m)

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

plot(
  b.sidi ~ s.sidi, data = dat_resampled_i,
  main = "Simpson's diversity indices (MC resampled data)",
  xlab = "Vegetation cover diversity",
  ylab = "Bird diversity")
abline(fit_resampled_i)
```



- 15.
16. Critical value = -0.01320388
The observed slope was less than the critical value.
17. The results suggest that we cannot reject the null hypothesis, that there is no relationship between vegetation and bird diversity. This would then suggest that there might be a relationship between these two variables, based on the observed slope value being outside the 95% interval. This means it is outside of the regular 95% of data points and would not likely be due to chance.
18.


```

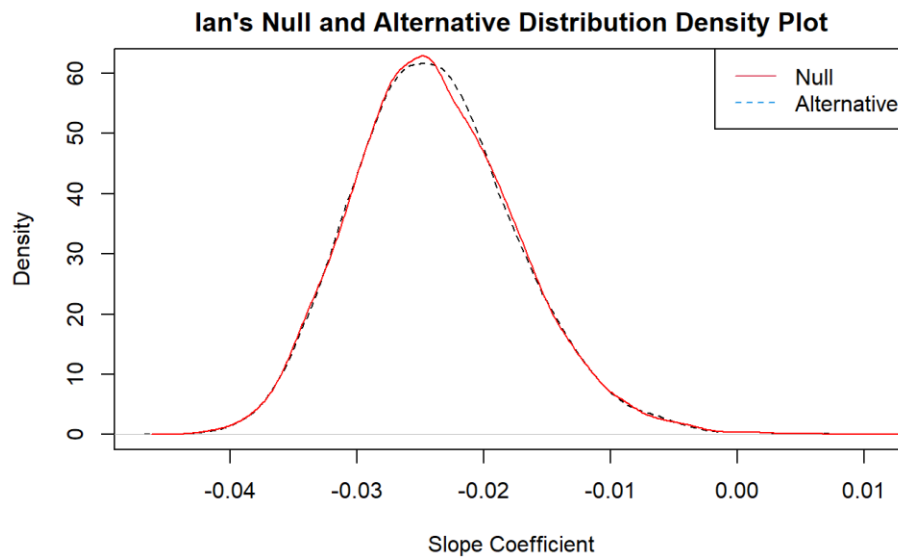
m = 10000
result_boot = numeric(m)

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

  dat_boot = dat_1[index_1, ]

  fit_bs1 = lm(b.sidi ~ s.sidi, data = dat_boot)
  result_boot[i] = coef(fit_bs1)[2]
}
hist(
  result_boot,
  main = "Ian's Alternative Distribution of Regression Slope",
  xlab = "Slope Parameter")
      
```

```
abline(v = slope_observed, lty = 2, col = "red", lwd = 2)
abline(v = 0, lty = 2, col = 1, lwd = 2)
```



- 19.
20. The region under the graph where they both fall under would be data points that satisfy both the null and alternative hypotheses. These data points are a result of the error from resampling and would be the outlying samples that we cannot reject the null hypothesis for.