Analysis of Environmental Data – Lab 8
**Olivia Dinkelacker**

Q1 (1 pt.): Calculate the standard deviation of the differences in mean flipper length from your bootstrap simulation. Show the R-code you used to find do the calculation.

sd(pen_boot$t)
1.027241

Q2 (2 pts.): Include your histogram of bootstrapped differences in your lab report (you don't need to show the R-code but make sure your plot includes appropriate title, axes, etc.).

flipper length differences Adelie and Chinstrap
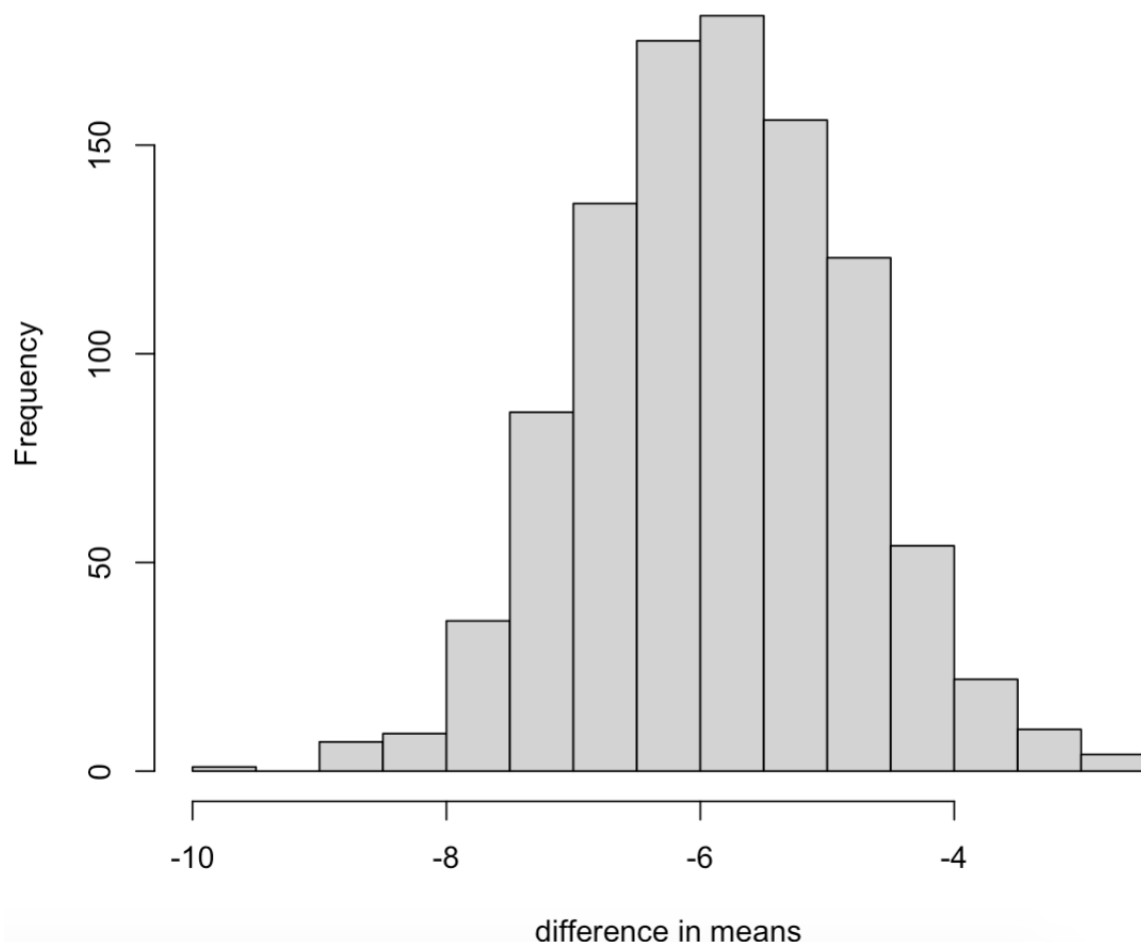
Q3 (2 pts.): What was the 95% bootstrap CI you calculated using `quantile()`? Show the R-code you used to answer the question.

boot.ci(pen_boot)
quantile(pen_boot$t, c(0.025, 0.975))
   2.5%   97.5%
3.921575 7.910056

Q4 (4 pts.): Do you think the resampled differences in means follow a skewed distribution? Your answer should make reference to the *mean*, *median*, **and** *histogram* of the differences in means.

The differences in means roughly follow a normal distribution since the histogram is almost bell shaped, and also the values of the mean and median are almost the same.

mean(pen_boot$t)
-5.862947
median(pen_boot$t)
-5.847792

Q5 (2 pts.): Show the R-code you used to create `pen_ecdf()`

pen_edcf = ecdf(pen_boot$t)

Q6 (2 pts.): What is the probability, according to the empirical distribution function, of observing a mean difference of -4.5 *or greater*? Show the R code you used to perform the calculation.

1-pen_edcf(-4.5)
0.09

Q7 (2 pts.): What is the probability, according to the empirical distribution function, of observing a mean difference of -8 *or smaller*? Show the R code you used to perform the calculation.

pen_ecdf(-8)
0.017

Q8 (3 pts.): State the null and alternative hypotheses of a *two-sample*, *two-tailed* test for the difference in mean flipper lengths between the two penguin species.

Null hypothesis: There is no difference in the mean flipper length between the two species.
Alternative: There is a difference in the mean flipper length between the two species.

Q9 (2 pts.): What was the p-value? Show the R-code you used to find out.

wilcox.test(pine~treatment,data = dat_tree, alternative = "two.sided")
p-value = 0.1005

Q10 (1 pt.): What were the endpoints of your bootstrap CI? Show the R-code you used to find out.

```
boot.ci(tree_boot, na.rm=TRUE)
quantile(tree_boot$t, c(0.025, 0.95))
 2.5%   95%
 3.875 27.375
```

Q11 (1 pt.): What is the observed difference in mean tree counts and does it fall within the 95% bootstrap CI?

```
mean(tree_boot$t)
16.03013
```

The value falls into the 95% confidence interval.

Q12 (2 pts.): Briefly describe the Simpson diversity index, and explain what it quantifies.

The Simpson diversity index is a measure of diversity taking into account the number of species present.

Q13 (2 pts.): Show the code you used to z-standardize the `s.sidi` column

```
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)
```

Q14 (6 pts.): Show the code for your completed Monte Carlo simulation loop.

```
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]
}
```

Q15 (2 pts.): In your report, include a plot of your histogram of Monte Carlo resampled slopes. Include vertical lines showing the observed slope and the critical value from the resampled MC slopes.
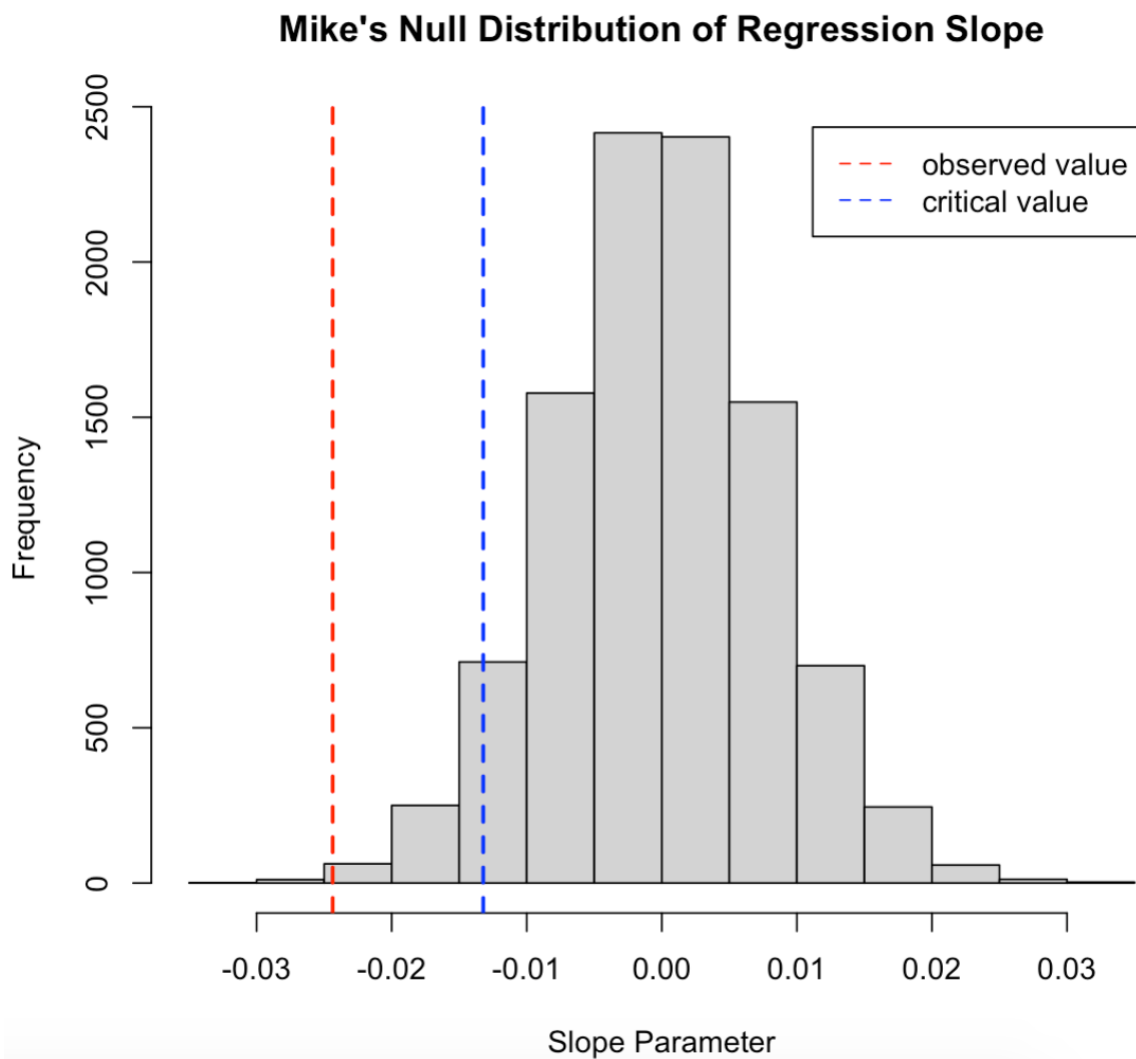
Q16 (2 pts.): What was your critical value? Was the observed slope less than the critical value?
quantile(result_mc, c(.05))
        5%
-0.01322205

The observed slope value (-0.02437131) is not less than the critical value.

Q17 (3 pts.): What is your conclusion regarding the evidence of a negative relationship between vegetation cover diversity and bird diversity? Make sure to justify your conclusions using the results of your simulation analysis.

There is strong evidence of a strong negative relationship between vegetation cover diversity and bird diversity, because the observed slope is much lower than the critical value.

Q18 (2 pts.): Show the code you used in your bootstrap loop.
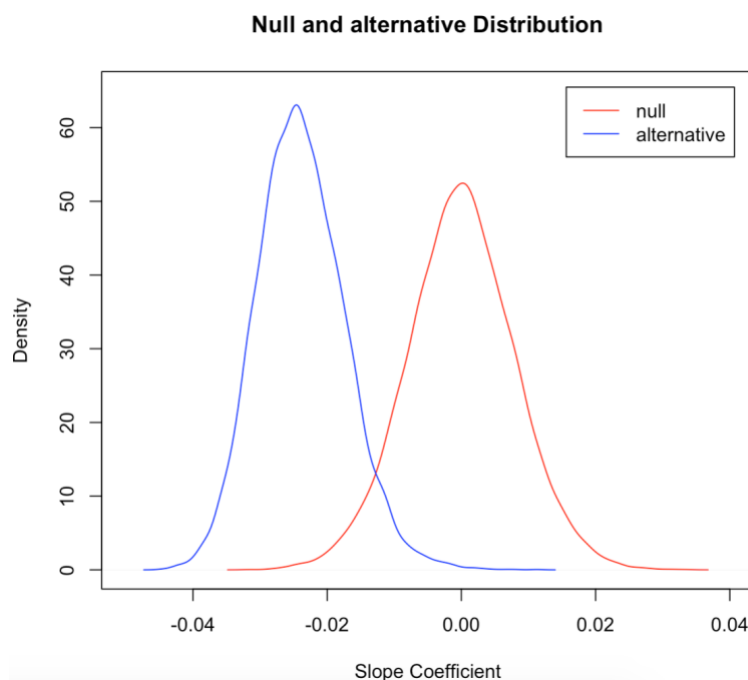
```
m = 10000
result_boot = numeric(m)

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]
```

Q19 (4 pts.): Include your double density plot. For full credit your plot must include: a legend, the two density curves in different colors, appropriate axis labels and title.

Q20 (2 pts.): Recall that the bootstrap curve shows the distribution of plausible values for the slope coefficient if we could resample the original data. The Monte Carlo curve shows the distribution of plausible values for the slope coefficient if the null hypothesis were true. How can you interpret the region that falls under both curves?

The region that falls under both of the curves represents ambiguous values. We can't be sure if the value rejects or accepts the null hypothesis when it falls within this region.