Matthew Jusino

Lab 7

- **Q1 (1 pt.):** What is the sample size, n? Show the code you used for the calculation and remember to check for missing data.

Sample size n is equal to 342.

```
dat_gentoo = droplevels(subset(penguins, species = "Gentoo"))

dat_gentoo$bill_length_mm = !is.na(dat_gentoo$bill_length_mm)

n = sum(dat_gentoo$bill_length_mm, na.rm = TRUE)
```

- **Q2 (1 pt.):** What is the sample standard deviation? Show the code you used for the calculation.

Sd is equal to 0.07613805

```
sd(dat_gentoo$bill_length_mm, na.rm = TRUE
```

- **Q3 (2 pts.):** What are the critical t-values? Show the R code you used for the calculation.

t_crit is equal to 1.966945

```
t_crit = abs(qt(alpha / 2, df = n - 1))
```

- **Q4 (1 pt.):** What is the sample standard error? Show the R code you used for the calculation.

The sample standard error is equal to 0.004117074

```
sse = sd(dat_gentoo$bill_length_mm, na.rm = TRUE) / sqrt(n)
```

- **Q5 (2 pts.):** Finally, construct the CI and show the R code you used for the calculation.

```
alpha = 0.05

dat_gentoo$bill_length_mm = !is.na(dat_gentoo$bill_length_mm)

n = sum(dat_gentoo$bill_length_mm, na.rm = TRUE)
```

```
sse = sd(dat_gentoo$bill_length_mm, na.rm = TRUE) / sqrt(n)
t_crit = abs(qt(alpha / 2, df = n - 1))
ci_radius = sse * t_crit
anst_ci = c(
  lower = mean(dat_gentoo$bill_length_mm, na.rm = TRUE) - ci_radius,
  upper = mean(dat_gentoo$bill_length_mm, na.rm = TRUE) + ci_radius)
print(round(anst_ci, 4))
```

The CI is 0.9861 to 1.0023

Lower Upper

- **Q6 (1 pt.):** What is the CI?

The CI is

2.5% 97.5%

0.985 1.00

- **Q7 (1 pt.):** Show the r code you used to call the `boot()` function.

```
boot_mean = function(x,i)
{
  return(mean(x[i], na.rm = TRUE))
}

gentboot =
  boot(
    data = dat_gentoo$bill_length_mm,
    statistic = boot_mean,
```

```
    R = 342)
print(gentboot)
```

- **Q8 (2 pts.):** Show the r code you used to calculate the upper and lower 2.5% quantiles.

```
str(gentboot)
mean(dat_gentoo$bill_length_mm)
gentboot$t0
mean(gentboot$t) - gentboot$t0
sd(gentboot$t)
quantile(
  gentboot$t,
  c(0.025, 0.975))
```

- **Q9 (5 pts.):** Show your completed `rarefaction_sampler()` function.

```
rarefaction_sampler = function(input_dat, n_iterations)
{
  n_input_rows = nrow(input_dat)

  results_out = matrix(
    nrow = n_iterations,
    ncol = n_input_rows)

  for(i in 1:n_iterations)
  {
```

```
    for(j in 1:n_input_rows)
    {
      rows_j = sample(n_input_rows, size = j, replace=TRUE)

      t1 = input_dat[rows_j, ]
      t2 = apply(t1, 2, sum)
      results_out[i, j] = sum(t2 > 0)
    }
  }
  return(results_out)
}
```

- **Q10 (1 pt.):** What did you find most difficult about building the function?

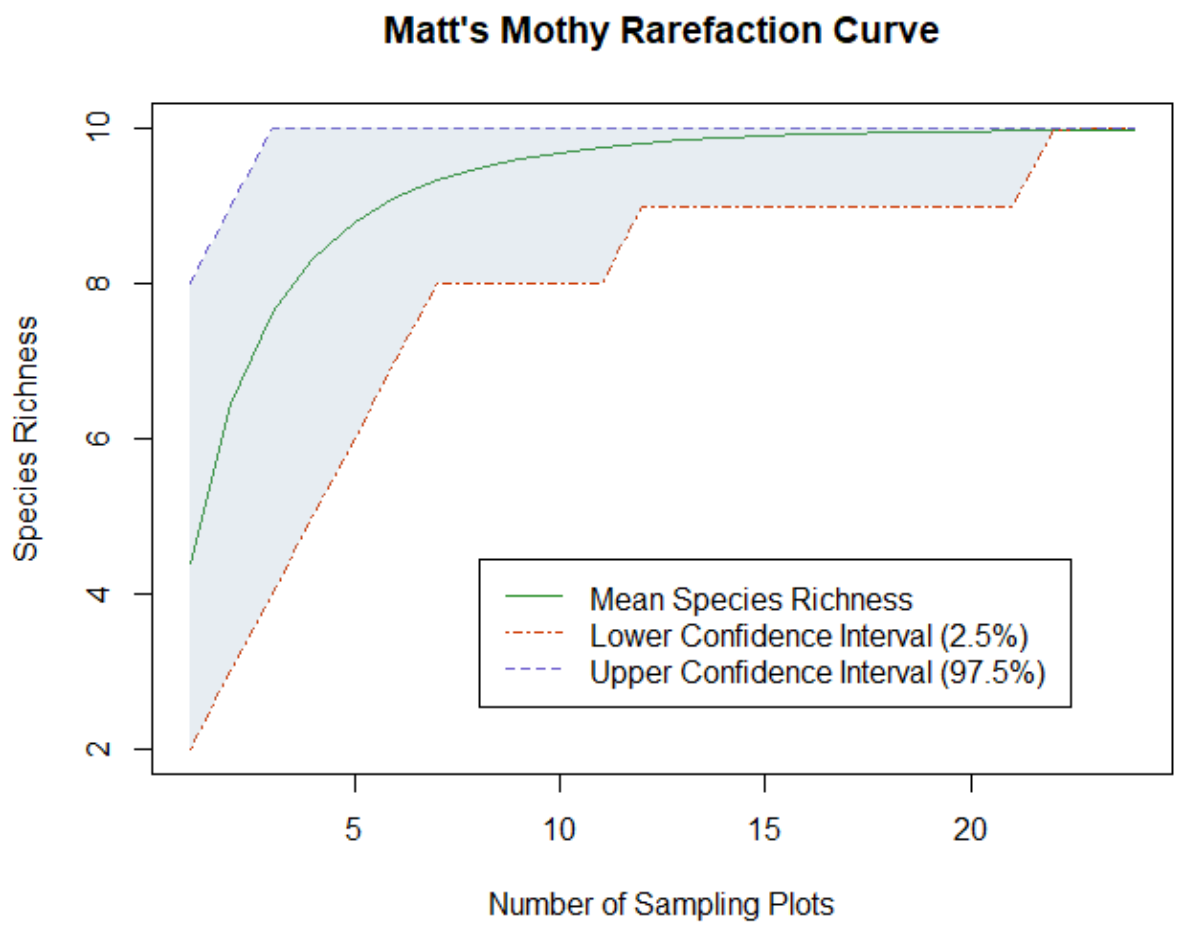
The most difficult part of building the function was simply trying to figure out the subsetting of a loop within a loop.

- **Q11 (4 pts.):** Show the code you used to perform the simulations and construct the curve.

```
rarefact = rarefaction_sampler(moths[, -1], 10000)

rare_mean = apply(rarefact, 2, mean)
rare_quant = apply(rarefact, 2, quantile, probs=c(0.025, 0.975))
rare = t(rbind(rare_mean, rare_quant))
```

- **Q12 (4 pts.):** Include your rarefaction curve plot in your report. Show the R-code you used to create your plot.




```
matplot(
  rare,
  type='l',
  xlab='Number of Sampling Plots',
  ylab='Species Richness',
  main = "Matt's Mothy Rarefaction Curve",
  lty = c(1,4,8), col = c("forestgreen","orangered3","slateblue3"))

legend(
  'bottomright',
```

```
  legend=c('Mean Species Richness', 'Lower Confidence Interval (2.5%)', 'Upper Confidence Interval
(97.5%)'),

  lty=c(1,4,8),col=c("forestgreen","orangered3","slateblue3"),inset=c(.1,.1))



polygon(x = c(1:24, 24:1), y = c(rare[,2], rev(rare[,3])) , col = adjustcolor("slategray3", 0.25),
border = NA)
```

- **Q13 (2 pts.):** About how many sites should you visit if you want to see all of the moth species? Explain your reasoning using your rarefaction curve figure.

You will most likely see all species if you visit at least 10 sampling plots, as both the mean and the 97.5% curve have reached 10 species richness by that point, and the 2.5% curve has reached 8 species richness. But to be absolutely certain of seeing all species, you must visit 23 sampling plots, as the 2.5% curve is still at 9 species until 22 plots, where it jumps up to 9.975, and then 10 at 23 plots.