Heather Siart
ECO – 634 Analysis of Environmental Data
10/31/2021
Partners: Mani, Matt, John, Bonni

Lab 7: The Bootstrap

Q1: The sample size for n is 123

Q2: 3.081857
ssd = sd(dat_gentoo$bill_length_mm, na.rm = TRUE)

Q3: crit_upper is 1.9796 and crit_lower is -1.9796

crit_upper = qt(1 - (0.05/2), 122)
crit_lower = qt(0.05/2, 122)

Q4: 0.2778817
sse_mean = function(x, na.rm = TRUE) sd(x)/sqrt(lentgh(x))
sse_mean(dat_gentoo$bill_length_mm)

Q5: The parametric CI is 43.92193 ± 0.02701383
Upper: 43.94894
Lower: 43.89492

# Calculate critical t-values using alpha value (1 - 0.95)
alpha = 0.05

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

#Q5
# Calculate the CI
mean(penguins$bill_length_mm, na.rm = TRUE)

# This is the radius of the CI
ci_parametric = t_crit * sse

#Express CI as mean +/- radius
#upper
43.92193 + 0.02701383
#lower
43.92193 - 0.02701383

Q6, Q7, & Q8:

```r
pen_boot =

 boot(
   data = dat_gentoo$bill_length_mm,
   statistic = boot_mean,
   R = 10000
 )
print(pen_boot)
```

Bootstrap Statistics :
   original   bias   std. error
t1* 47.50488 -0.003338943   0.2777055

```r
quantile(
 pen_boot$t,
 c(0.025, 0.975))
   2.5%   97.5%
46.96585 48.05418
```

Q9:
```r
rm(list = ls())

moths = read.csv(here("data", "moths.csv"))
moth_dat = moths[,-1]
n = nrow(moth_dat)
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)
  {

    rows_j = sample(n, size = j, replace=TRUE)

    t1 = input_dat[rows_j, ]

    t2 = apply(t1, 2, sum)
```

```
    results_out[i, j] = sum(t2 > 0)
  }
 }
 return(results_out)
}
rarefact = rarefaction_sampler(moths[,-1], 10000)
head(rarefact)
```

Q10: The hardest part of creating this code was making sure I was calling everything in the right order so it would all flow.

Q11:

```
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))

matplot(
 rare,
 type='l',
 xlab='Number of sampling plots',
 ylab='Species richness',
 main='Convidence Intervals with 10k samples',
 )

legend(
 'bottomright',
 legend=c('mean','2.5%','97.5%'),
 cex = (0.65),
 lty=c(1,2,3),col=c(1,2,3), inset=c(.1,.1))
```
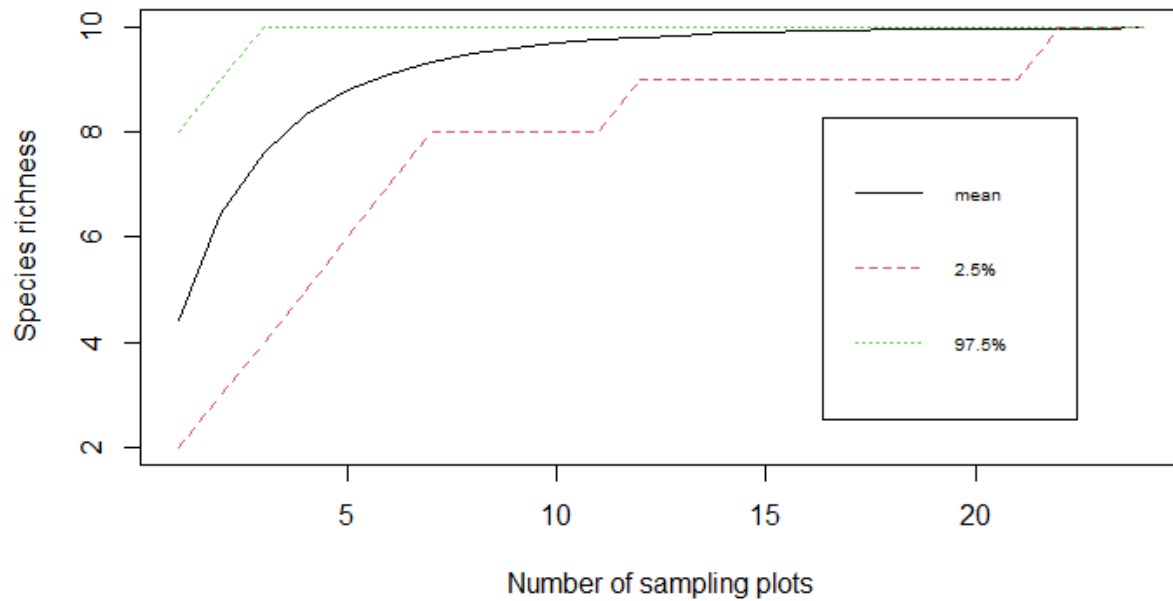
Q12:

Convidence Intervals with 10k samples

Q13: I would visit around 25 sites. That is where all three lines meet, and you aren't missing a species. You would probably be fine with 15 sample spots because at 97.5% you have a high chance of getting all 10 species and at 2.5% CI you have a chance of getting 8 of the 10 species each time.