Anacapa_Island_story

Jackson Hoeke

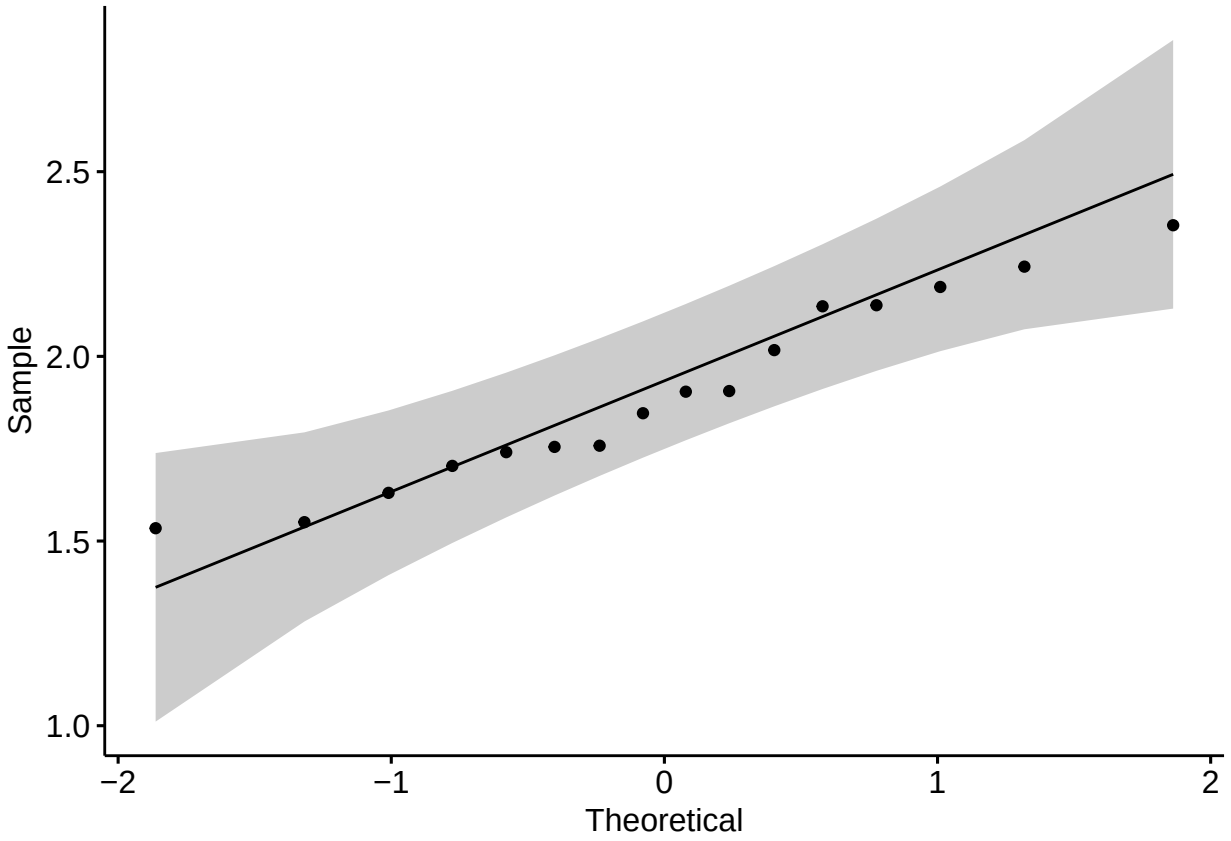
4/10/2021

First, subset data and calculate diversity (with and without white urchins)

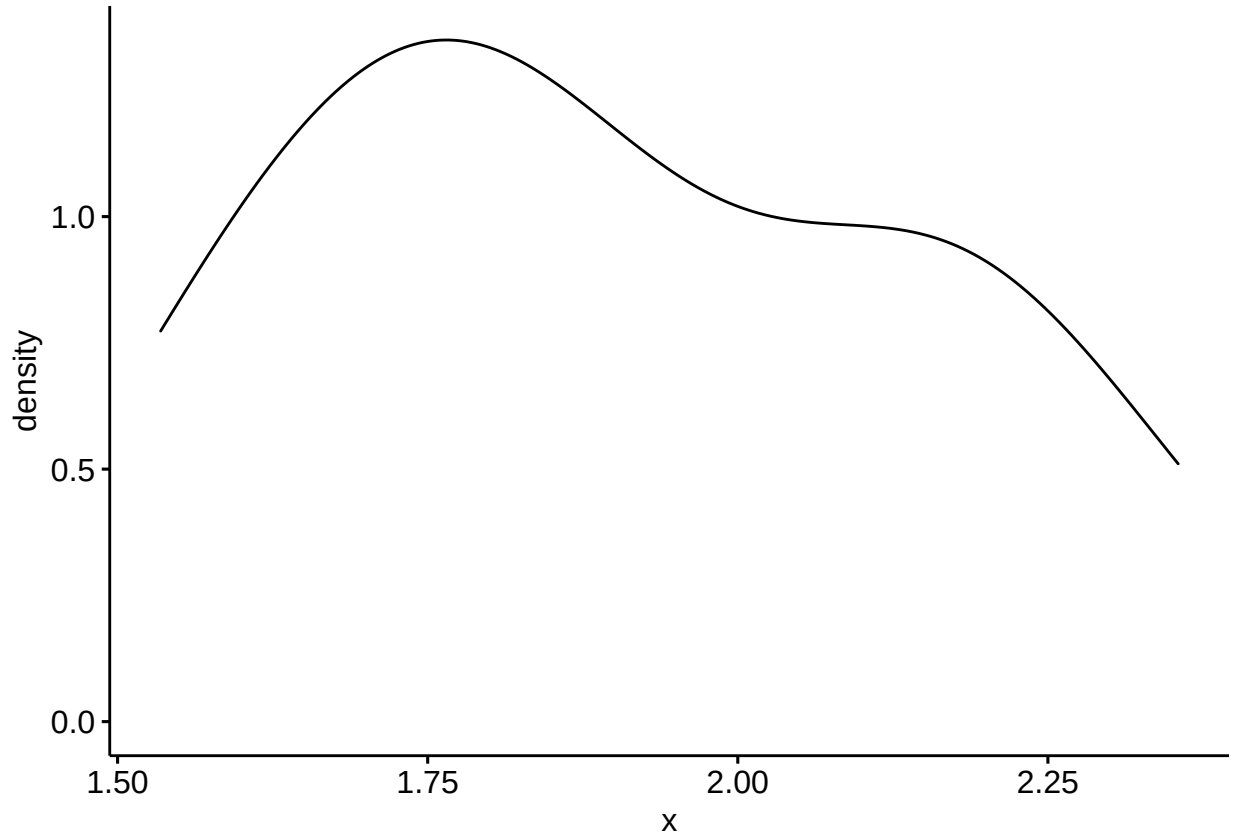
```
# Bring in data
Density <- data.frame(read.csv("Density_table_2005-2019.csv"))
Count <- data.frame(read.csv("Count_table_2005-2019.csv"))
Anacapa.d <- subset(Density, MPA_Group == "Anacapa Island")
Anacapa.c <- subset(Count, MPA_Group == "Anacapa Island")

# create a shannon diversity column using the vegan package

Anacapa.c <- mutate(Anacapa.c, Diversity = diversity(Anacapa.c[,7:163]))

# create a new data frame without white urchins

Anacapa.c.m <- subset(Anacapa.c,select=-White.sea.urchin)

# and make a shannon diversity column from this new data frame : Modified diversity

Anacapa.c.m <- mutate(Anacapa.c.m, Diversity = diversity(Anacapa.c.m[,7:163]))

# Combine columns from multiple data frames to make one simple, straightforward frame

Div.data <- data.frame(Year=Anacapa.c.m$Year,Type=Anacapa.c.m$Type,
                       Diversity=Anacapa.c$Diversity,
                       Modified.Diversity=Anacapa.c.m$Diversity)
```

Next, test if there is a difference in diversity between the SMR and the SMCA

```
# Is the (modified) diversity data normal?

# Run ggqplot to look at normality and ggdensity to examine the distribution
ggqqplot(Div.data$Modified.Diversity)
```




```
ggdensity(Div.data$Modified.Diversity)
```




```
# A shapiro test can test for normality
shapiro.test(Div.data$Modified.Diversity)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Div.data$Modified.Diversity
## W = 0.95191, p-value = 0.5205
```

```
# the var.test examines if the variance between the two data sets is equal:
# in this case it is modified diversity in the SMCA and SMR
var.test(Modified.Diversity~Type, data = Div.data)
```

```
##
##  F test to compare two variances
##
## data:  Modified.Diversity by Type
## F = 0.89151, num df = 7, denom df = 7, p-value = 0.8835
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1784829 4.4529921
## sample estimates:
## ratio of variances
##          0.8915061
```

```
# The data is normal. So a paired t-test can be used to test the effects of
# MPA type on (modified) Shannon diversity, while controlling for the effects
# of individual years.
```

```
t.test(Modified.Diversity~Type, data=Div.data,
       alternative = "two.sided", paired=TRUE,var.equal=TRUE)
```

```
##
##  Paired t-test
##
## data:  Modified.Diversity by Type
## t = -5.8635, df = 7, p-value = 0.000622
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3664202 -0.1558144
## sample estimates:
## mean of the differences
##               -0.2611173
```

```
# Null hypothesis rejected. There's definitely a significant difference
# in (modified) diversity when
# year and white urchin outbreaks are controlled for.
# (n = 8, df = 7, t = -5.8635, p = 0.000622)
```

The next step is to look at the lobsters

```
# Add spiny lobsters to data

Div.data <- mutate(Div.data, lobster=Anacapa.d$California.spiny.lobster)

# This lobster density data changes from 0 to slight to high, so even without
# a Shapiro test I know this isn't a normal distribution.
# Besides, If I just want to look at a correlation between lobster density and
# diversity, I just need a correlation test that doesn't assume normality.
# Run a spearman's rank correlation

lobs.sp = cor.test(Div.data$lobster,Div.data$Diversity, method="spearman")
lobs.sp
```

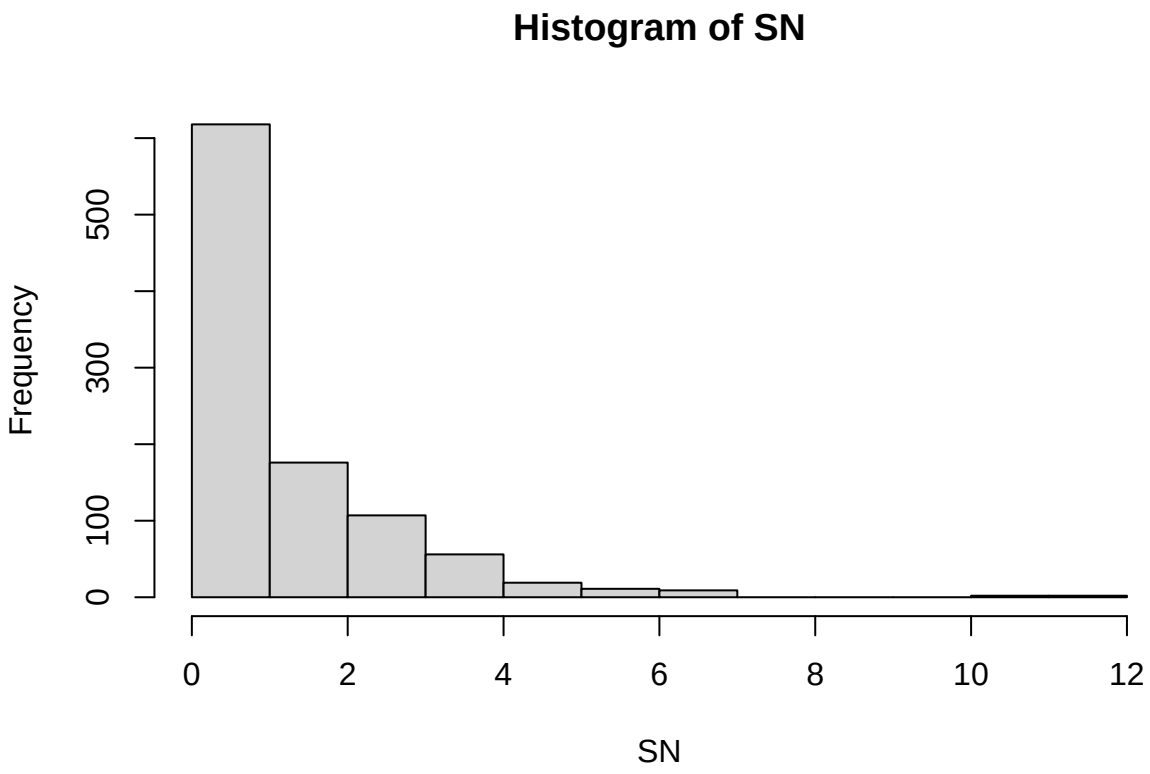
```
##
##  Spearman's rank correlation rho
##
## data:  Div.data$lobster and Div.data$Diversity
## S = 692.53, p-value = 0.946
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.0184219
```

```
# There is no correlation, (r = -0.0184219, p = 0.946) but white urchin density
# is EXTREMELY high, in 2014 and 2015, so that may be throwing it off.
# What about the (modified) diversity?

lobs.sp2 = cor.test(Div.data$lobster,Div.data$Modified.Diversity,
                    method="spearman")
lobs.sp2
```

```
##
```

```
##  Spearman's rank correlation rho
##
## data:  Div.data$lobster and Div.data$Modified.Diversity
## S = 174.75, p-value = 0.0009741
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.7430168
```

```r
# Here, the null hypothesis is rejected and there IS a correlation between
# lobster density and diversity (r = 0.7430168. p = 0.0009741)

# Does SMR or SMCA matter (i.e. does fishing pressure matter?)
# In this case, I want to see if the independent variable (MPA type) is
# affecting lobster density, so I need at t-test to compare population means in
# each MPA type. However, the data isn't normal, so I'll have to bootstrap the
# test.

# Calculate Signal:Noise (SN)
p1 = subset(Div.data, Type == "SMR")[,5]
p2 = subset(Div.data, Type == "SMCA")[,5]
b=c(p1,p2)
n1 = length(p1)
n2 = length(p2)
n =length(b)
s1 = sd(p1)
s2 = sd(p2)
pooledVar=(((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2))
signal = (1/((1/n1)+(1/n2)))*(mean(p1)-mean(p2))^2
calcSN =  signal/pooledVar

# Bootstrap t-test

trials = 1000
SN <- as.vector(NULL) # vector to hold SN
for (i in 1:trials) { #bootstrap resamples
  bs.H <- sample(b, n1, replace = TRUE) #resample H from all
  bs.C <- sample(b, n2, replace = TRUE) #resample C from all
  bs.s1 = sd(bs.H)
  bs.s2 = sd(bs.C)
  bs.pooledVar=(((n1-1)*bs.s1^2+(n2-1)*bs.s2^2)/(n1+n2-2))
  bs.signal = (1/((1/n1)+(1/n2)))*(mean(bs.H)-mean(bs.C))^2
  SN[i] = bs.signal/bs.pooledVar}
hist(SN) # plot (good habit)
```




```
bs.p = (sum(SN > calcSN))/trials # p-value.
label0="S:N:"
label2="P-value from bootstrap_m1:"
sprintf("%s %f", label0,calcSN)
```

```
## [1] "S:N: 3.701043"
```

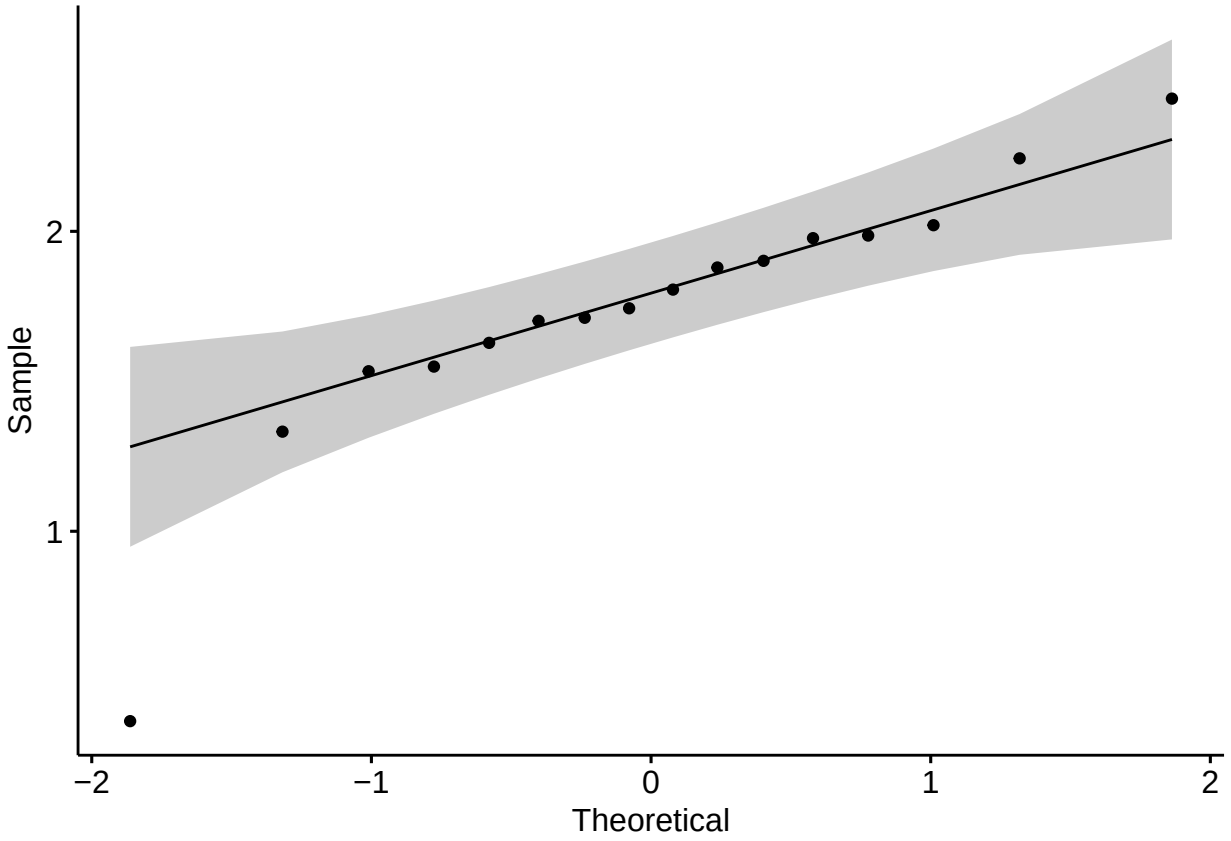
```
sprintf("%s %f", label2,bs.p)
```

```
## [1] "P-value from bootstrap_m1: 0.062000"
```

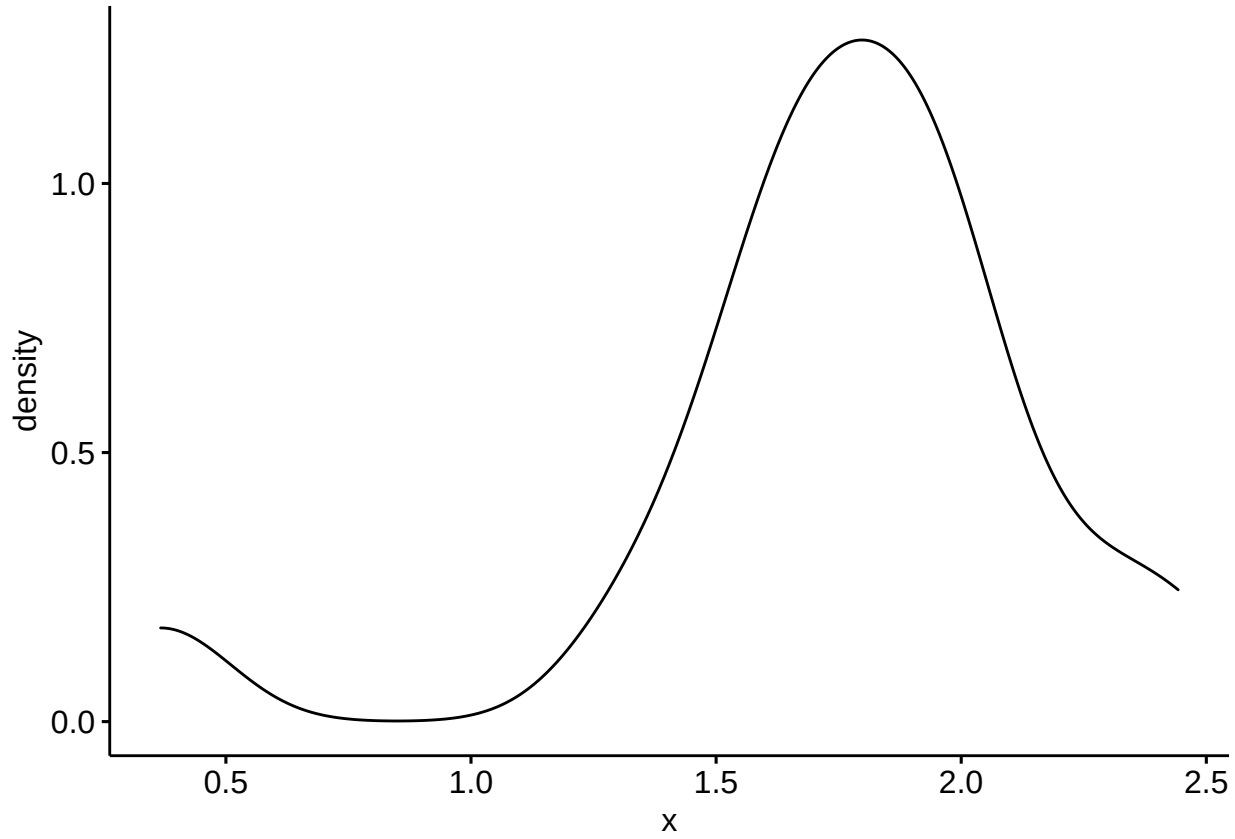
```
# Yes! The lack of fishing pressure has a significant effect!
# (S:N = 3.701043, p = 0.049)

# If white urchins are dropping diversity this much, there should be a
# correlation...
```

```
# Add white urchins to data
Div.data <- mutate(Div.data, urchin=Anacapa.d$White.sea.urchin)

# Because it is hypothesized the urchins are causing diversity to drop, a t-test
# will be used with white urchins as the independent variable.
# So is the data normal?

## visualize and test for normality
```

```
ggqqplot(Div.data$Diversity)
```




```
ggdensity(Div.data$Diversity)
```




```
shapiro.test(Div.data$Diversity)
```

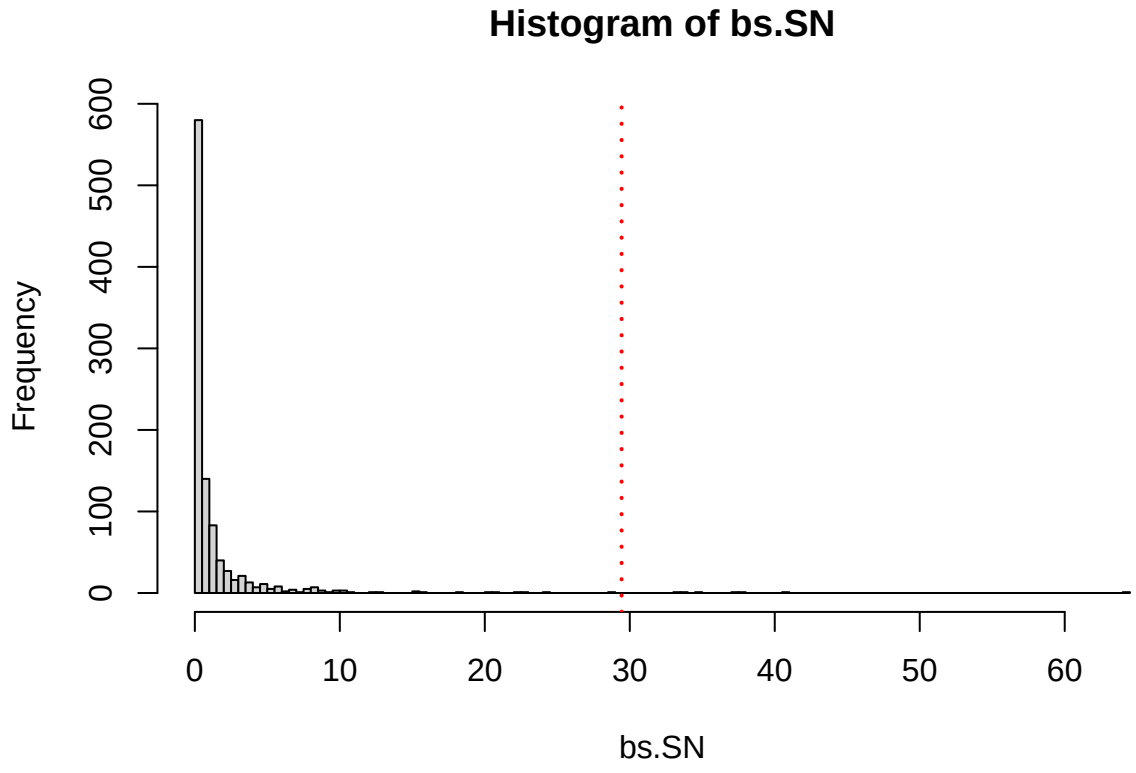
```
##
##  Shapiro-Wilk normality test
##
## data:  Div.data$Diversity
## W = 0.85571, p-value = 0.01655
```

```
# No, it is not normal. So it'll be bootstrapped.

# Calculate signal to noise
n = 16
m_x = mean(Div.data$urchin)
m_y = mean(Div.data$Diversity)
b1_num = sum((Div.data$urchin-m_x)*(Div.data$Diversity-m_y))
b1_den = sum((Div.data$urchin-m_x)^2)
b1=b1_num/b1_den
b0 = m_y-b1*m_x
SSE_R = sum((Div.data$Diversity-m_y)^2)
SSE_F = sum((Div.data$Diversity - (b0+b1*Div.data$urchin))^2)
Signal = SSE_R-SSE_F
Noise = SSE_F/(n-2)
SN = Signal/Noise

# Bootstrap t-test

trials = 1000
```

```
urn.x = Div.data$urchin
urn.y = Div.data$Diversity
bs.SN = NULL # vector to hold SN
for (i in 1:trials) {
  bs.x = sample(urn.x, n, replace = T)
  bs.y = sample(urn.y, n, replace = T)
  bs.m = lm(formula = bs.y~bs.x) #runs regression with bootstrapped data
  bs.ms = summary(bs.m)
  f = bs.ms$fstatistic[1] #extracts the f statistic from the model
  bs.SN[i] = f}
bs.p=(sum(bs.SN > SN))/trials #SN is from previous steps. If cleared workspace will need to calculate a

cat("P-value for bootstrap is", bs.p, "\n")
```

```
## P-value for bootstrap is 0.007
```

```
hist(bs.SN, breaks = 200)
abline(v=SN, col="red", lty="dotted", lwd=2)
```




```
# The null hypothesis is rejected. White urchins and Shannon diversity are
# negatively correlated in the Anacapa MPAs (n = 16, p = 0.003).

# So, are white urchins and spiny lobsters correlated too?
# Just a correlation, no assumptions on normality.
# Spearman's rank test will be used.

lo.ur.sp = cor.test(Div.data$urchin,Div.data$lobster, method="spearman")
```

```
lo.ur.sp
```

```
##
##  Spearman's rank correlation rho
##
## data:  Div.data$urchin and Div.data$lobster
## S = 304.9, p-value = 0.02675
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.5516187
```

```
# Apparently they are (Spearman's r = 0.5516187, p = 0.02675)
# What are the interactions? Run a two-way ANOVA to find out.

div.aov <- aov(Diversity ~ urchin * lobster, data = Div.data)
summary(div.aov)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## urchin          1 2.1266  2.1266  32.812 9.48e-05 ***
## lobster         1 0.1550  0.1550   2.391    0.148
## urchin:lobster  1 0.0788  0.0788   1.216    0.292
## Residuals      12 0.7778  0.0648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# There appears to be no interaction between white urchins and spiny lobsters
# in terms of diversity. The only significant correlation is that white urchins
# decrease diversity (urchin df = 1,12 F = 32.812, p = 9.48e-05).
```

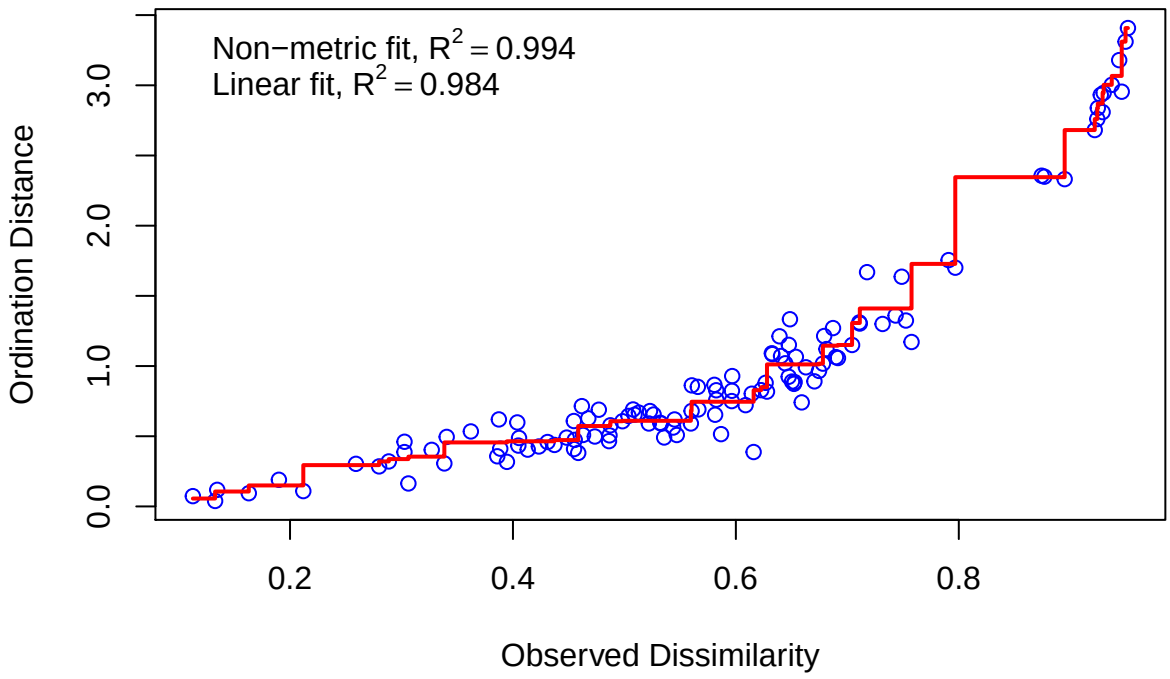
If we use an NMDS and species fit, we can characterize Anacapa's SMR and SMCA

```
# Create an NMDS for Anacapa, using SMCA and SMR each year as the different
# communities.
yr <- c("2005","2005","2006","2006","2007","2007","2008","2008","2009",
        "2009","2014","2014","2015","2015","2019","2019")
Anacapa.c.labels <- Anacapa.c[,2:6]
Anacapa.c.data <- Anacapa.c[,7:163]
Anacapa.mds <- metaMDS(Anacapa.c.data, distance = "bray", k = 2, trymax = 100,
                       autotransform = FALSE)
```

```
## Run 0 stress 0.08444952
## Run 1 stress 0.0823366
## ... New best solution
## ... Procrustes: rmse 0.1021371  max resid 0.2228917
## Run 2 stress 0.08256438
## ... Procrustes: rmse 0.02281847  max resid 0.06927003
## Run 3 stress 0.07928072
## ... New best solution
## ... Procrustes: rmse 0.04151066  max resid 0.1366239
## Run 4 stress 0.08256443
## Run 5 stress 0.08520438
## Run 6 stress 0.08236635
## Run 7 stress 0.08520424
```

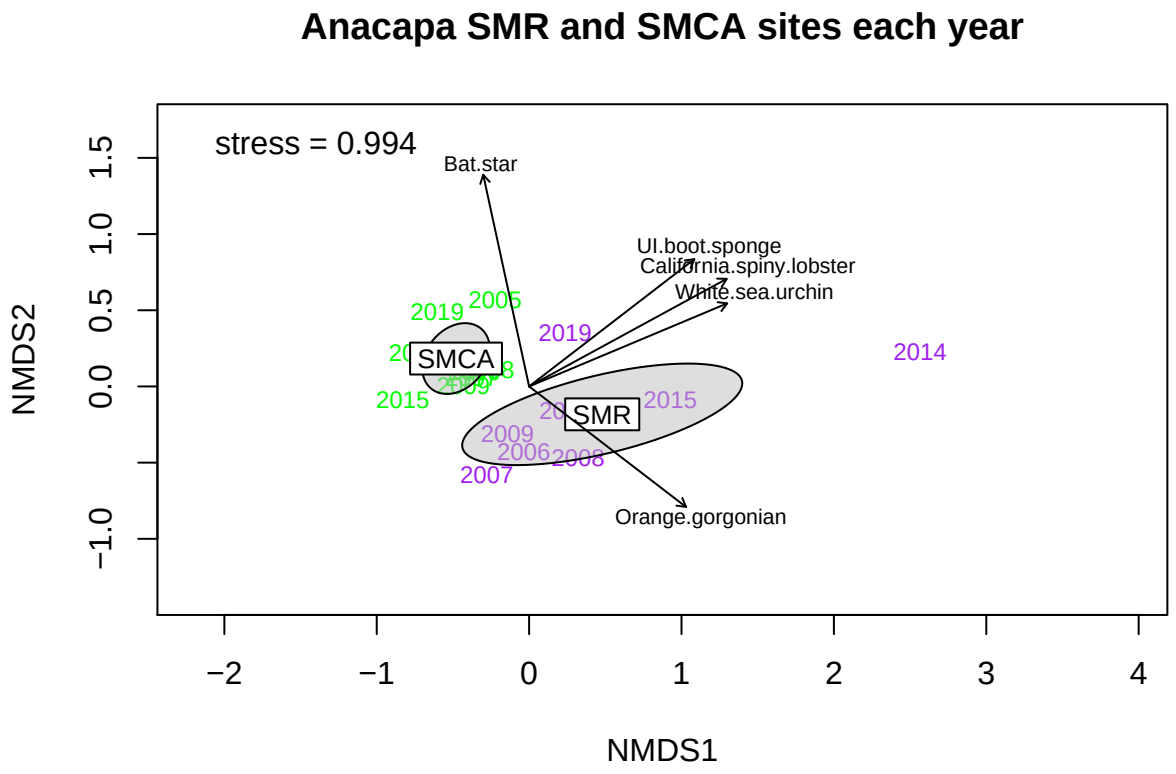
```
## Run 8 stress 0.07928073
## ... Procrustes: rmse 7.042254e-05  max resid 0.0001631998
## ... Similar to previous best
## Run 9 stress 0.08520429
## Run 10 stress 0.117302
## Run 11 stress 0.08444949
## Run 12 stress 0.08520427
## Run 13 stress 0.08256426
## Run 14 stress 0.08444971
## Run 15 stress 0.08236697
## Run 16 stress 0.08236707
## Run 17 stress 0.08444947
## Run 18 stress 0.08444945
## Run 19 stress 0.08444942
## Run 20 stress 0.07928081
## ... Procrustes: rmse 0.0002090699  max resid 0.0004965121
## ... Similar to previous best
## *** Solution reached
```

```r
# Run stressplot
stressplot(Anacapa.mds)
```




```r
MDS <- plot(Anacapa.mds, type = "n",
     main = "Anacapa SMR and SMCA sites each year")
orditorp(Anacapa.mds, display="sites", label = yr,
         col=c("green","purple","green","purple","green","purple","green",
               "purple","green","purple","green","purple","green","purple",
```

```
              "green","purple"), air=0.01, cex=0.75)
ordiellipse(Anacapa.mds, groups=Anacapa.c.labels$Type, draw="polygon",
            col=c("grey"), label=T)
# Create a species fit
Anacapa.spp.fit <- envfit(Anacapa.mds, Anacapa.c.data, permutations = 10000)
plot(Anacapa.spp.fit, p.max = 0.005, col="black", cex=0.7)
legend("topleft", "stress = 0.994", bty = "n", cex = 1)
```




```
## Very cool. And there's the spiny lobsters and White urchins, so they
## are obviously having an effect on characterizing the SMR.
```

Now what about sea cucumbers, which are fished in the SMCA?

```
# Add CA sea cucumber to data
Div.data <- mutate(Div.data, CA.sea.cucumber=Anacapa.d$California.sea.cucumber)

# Is there a correlation between California sea cucumbers and modified
# diversity?
# A correlation can be tested using Spearman's

cumber.sp <- cor.test(Div.data$CA.sea.cucumber,Div.data$Modified.Diversity,
                      method="spearman")
cumber.sp
```

```
##
##  Spearman's rank correlation rho
```

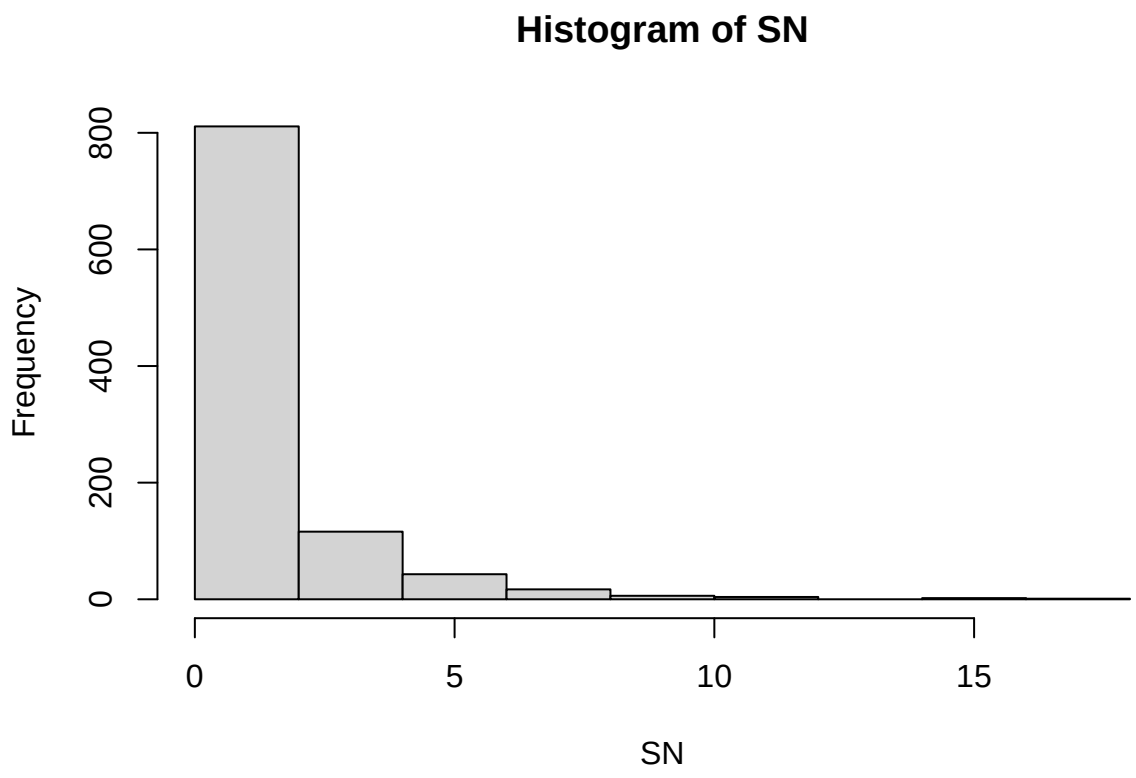
```
##
## data:  Div.data$CA.sea.cucumber and Div.data$Modified.Diversity
## S = 520, p-value = 0.379
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.2352941
```

```
# There is no correlation (r = 0.2352941, p = 0.379)

# How about... with unmodified diversity?

cumber.sp2 <- cor.test(Div.data$CA.sea.cucumber,Div.data$Diversity,
                       method="spearman")
cumber.sp2
```

```
##
##  Spearman's rank correlation rho
##
## data:  Div.data$CA.sea.cucumber and Div.data$Diversity
## S = 830, p-value = 0.4103
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.2205882
```

```
# No, no there is not a correlation here either
# (Spearman's r = -0.2205882, p = 0.4103)

# Well, is it at least doing better in the SMR where it isn't fished?
# Like with the lobsters, a bootstrapped t-test will be used to test for
# a difference between 'cumbers in the SMR vs. the SMCA.

# Calculate Signal:Noise

p1 = subset(Div.data, Type == "SMR")[,7]
p2 = subset(Div.data, Type == "SMCA")[,7]
b=c(p1,p2)
n1 = length(p1)
n2 = length(p2)
n =length(b)
s1 = sd(p1)
s2 = sd(p2)
pooledVar=(((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2))
signal = (1/((1/n1)+(1/n2)))*(mean(p1)-mean(p2))^2
calcSN =  signal/pooledVar

# Bootstrap t-test

trials = 1000
SN <- as.vector(NULL) # vector to hold SN
for (i in 1:trials) { #bootstrap resamples
  bs.H <- sample(b, n1, replace = TRUE)
  bs.C <- sample(b, n2, replace = TRUE)
  bs.s1 = sd(bs.H)
```

```
  bs.s2 = sd(bs.C)
  bs.pooledVar=(((n1-1)*bs.s1^2+(n2-1)*bs.s2^2)/(n1+n2-2))
  bs.signal = (1/((1/n1)+(1/n2)))*(mean(bs.H)-mean(bs.C))^2
  SN[i] = bs.signal/bs.pooledVar}
hist(SN) # plot (good habit)
```




```
bs.p = (sum(SN > calcSN))/trials # p-value.
label0="S:N:"
label2="P-value from bootstrap_m1:"
sprintf("%s %f", label0,calcSN)
```

```
## [1] "S:N: 1.737870"
```

```
sprintf("%s %f", label2,bs.p)
```

```
## [1] "P-value from bootstrap_m1: 0.229000"
```

```
# No, apparently the SMR doesn't have a significant effect (n = 16, p = 0.227)

# Well, how about the other abundant sea cucumber? The warty sea cucumber...

# Add to data

Div.data <- mutate(Div.data, Warty.sea.cucumber=Anacapa.d$Warty.sea.cucumber)

# And run a spearman's rank test with modified diversity
# No need to run unmodified. It's pretty clear that the urchins skew the
```

```r
# diversity index way to much for it to be useful outside urchin analysis.

w.cumber.sp <- cor.test(Div.data$Warty.sea.cucumber,Div.data$Modified.Diversity,
                        method="spearman")
w.cumber.sp
```

```
##
##  Spearman's rank correlation rho
##
## data:  Div.data$Warty.sea.cucumber and Div.data$Modified.Diversity
## S = 302, p-value = 0.02762
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.5558824
```

```r
# Warty sea cucumbers are correlated with modified diversity
# (Spearman's r = 0.5558824, p = 0.02762)

# Based on the fact that only invertebrates on the NMDS are different
# between the SMR and SMCA, it seems silly to look at those differences
# for Warty sea cucumbers.
# BUT, maybe it wouldn't hurt to run another organism on that NMDS...
```

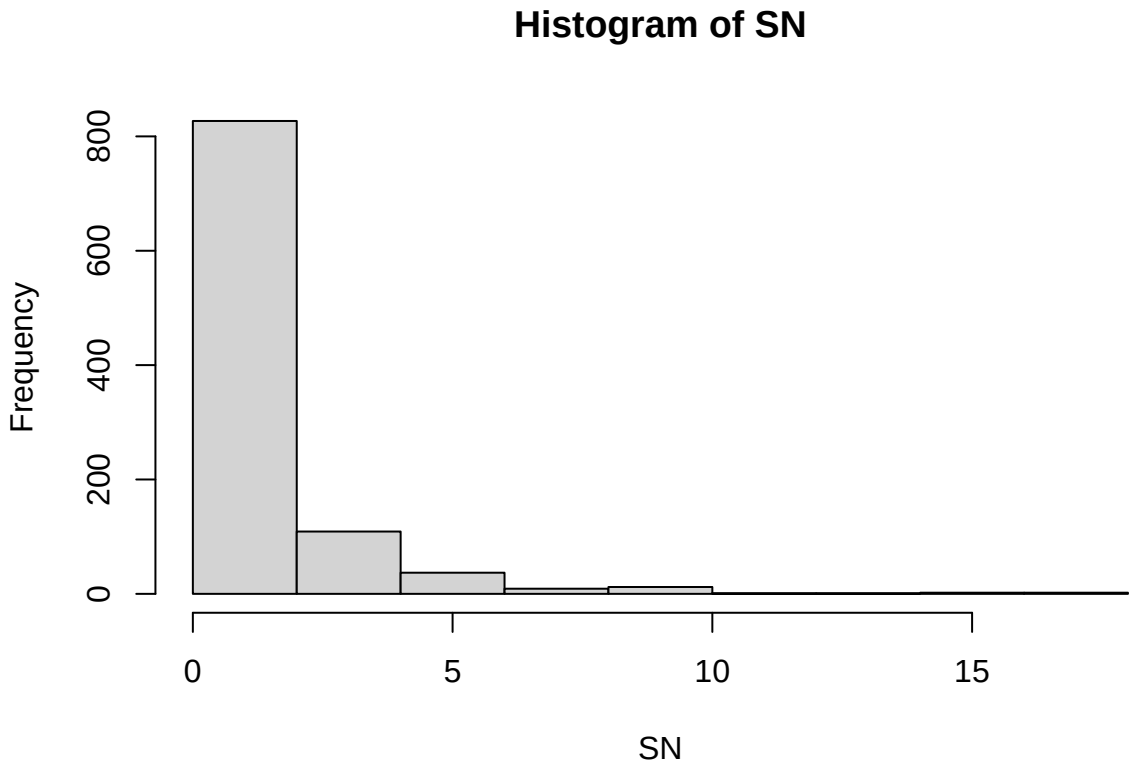
Let's run orange gorgonians!

```r
# Add orange gorgonians to the data
Div.data <- mutate(Div.data, Orange.gorgonian=Anacapa.d$Orange.gorgonian)

# Check correlation with (non-urchin) diversity, just out of curiosity.
orange.sp <- cor.test(Div.data$Orange.gorgonian,Div.data$Modified.Diversity,
                      method="spearman")
orange.sp
```

```
##
##  Spearman's rank correlation rho
##
## data:  Div.data$Orange.gorgonian and Div.data$Modified.Diversity
## S = 224.66, p-value = 0.004547
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.6696194
```

```r
# Orange gorgonians appear to have a significant correlation with diversity.
# (Spearman's r = 0.6696194, p = 0.004547)

# Now to see how associated it is with the SMR
# Use the usual bootstrapped t-test

p1 = subset(Div.data, Type == "SMR")[,9]
p2 = subset(Div.data, Type == "SMCA")[,9]
b=c(p1,p2)
n1 = length(p1)
n2 = length(p2)
```

```
n =length(b)
s1 = sd(p1)
s2 = sd(p2)
pooledVar=(((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2))
signal = (1/((1/n1)+(1/n2)))*(mean(p1)-mean(p2))^2
calcSN =  signal/pooledVar
trials = 1000
SN <- as.vector(NULL) # vector to hold SN
for (i in 1:trials) { #bootstrap resamples
  bs.H <- sample(b, n1, replace = TRUE) #resample H from all
  bs.C <- sample(b, n2, replace = TRUE) #resample C from all
  bs.s1 = sd(bs.H)
  bs.s2 = sd(bs.C)
  bs.pooledVar=(((n1-1)*bs.s1^2+(n2-1)*bs.s2^2)/(n1+n2-2))
  bs.signal = (1/((1/n1)+(1/n2)))*(mean(bs.H)-mean(bs.C))^2
  SN[i] = bs.signal/bs.pooledVar}
hist(SN) # plot (good habit)
```




```
bs.p = (sum(SN > calcSN))/trials # p-value.
label0="S:N:"
label2="P-value from bootstrap_m1:"
sprintf("%s %f", label0,calcSN)
```

```
## [1] "S:N: 25.162747"
```

```
sprintf("%s %f", label2,bs.p)
```

```
## [1] "P-value from bootstrap_m1: 0.000000"
```

```
# Well alright, seems like it has a close association with the SMR
# (S:N = 25.162747, p = 0.001)

# So... that begs the question: Is there an interaction between diversity, and
# this structure-forming invertebrate found far more in the SMR?
# Using (non-urchin) diversity of course. No need to get that unpleasant data
# skewing everything.

# A two-way ANOVA can test for this.

or.aov <- aov(Modified.Diversity ~  Type * Orange.gorgonian, data = Div.data)
summary(or.aov)
```

```
##                       Df Sum Sq Mean Sq F value Pr(>F)
## Type                   1 0.2727 0.27273   5.904 0.0317 *
## Orange.gorgonian       1 0.0662 0.06618   1.433 0.2544
## Type:Orange.gorgonian  1 0.0763 0.07626   1.651 0.2231
## Residuals             12 0.5543 0.04619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(or.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Modified.Diversity ~ Type * Orange.gorgonian, data = Div.data)
##
## $Type
##               diff        lwr       upr     p adj
## SMR-SMCA 0.2611173 0.02697103 0.4952636 0.0317448
```

```
# No... no there is no interaction between orange gorgonian density and
# modified diversity.
```

Now the fun part, graphing all this stuff!

```
M.div.plot <- ggplot(Div.data, aes(x=Year, y=Modified.Diversity, fill=Type)) +
  geom_bar(position="dodge",stat="identity") +
  ggtitle("Diversity (without white urchins) off Anacapa Island") +
  ylab("Shannon diversity")

Div.plot <- ggplot(Div.data, aes(x=Year, y=Diversity, fill=Type)) +
  geom_bar(position="dodge",stat="identity") +
  ggtitle("Shannon diversity over time off Anacapa Island") +
  ylab("Shannon diversity")

Lobs.plot <- ggplot(Div.data, aes(x=Year, y=lobster, fill=Type)) +
  geom_bar(position="dodge",stat="identity") +
  ggtitle("Spiny lobster density off Anacapa Island") +
  ylab("Density (m^2)")

Urch.plot <- ggplot(Div.data, aes(x=Year, y=urchin, fill=Type)) +
```

```
  geom_bar(position="dodge",stat="identity") +
  ggtitle("White urchin density off Anacapa Island") +
  ylab("Density (m^2)")

CA.cumber.plot <- ggplot(Div.data, aes(x=Year, y=CA.sea.cucumber, fill=Type)) +
  geom_bar(position="dodge",stat="identity") +
  ggtitle("California sea cucumber density off Anacapa Island") +
  ylab("Density (m^2)")

W.cumber.plot <- ggplot(Div.data, aes(x=Year, y=Warty.sea.cucumber, fill=Type)) +
  geom_bar(position="dodge",stat="identity") +
  ggtitle("Warty sea cucumber density off Anacapa Island") +
  ylab("Density (m^2)")

orange.plot <- ggplot(Div.data, aes(x=Year, y=Orange.gorgonian, fill=Type)) +
  geom_bar(position="dodge",stat="identity") +
  ggtitle("Orange gorgonian density off Anacapa Island") +
  ylab("Density (m^2)")

plot_grid(M.div.plot,Div.plot,Lobs.plot,Urch.plot,CA.cumber.plot,W.cumber.plot,
          orange.plot,MDS)
```