Homework week 01

Daniel

2025-02-26

#read table

setwd("C:/Users/dan91/Rstudio/stat\_data/mock\_dataTS")  
cop\_compo <- read.table("Copepod\_composition.txt", header = T)  
cop\_den <- read.table("Cop\_density.txt")  
cop\_sp <- read.table("copepodSPlist.txt", fill = T, sep = "\n") #separate by different row

# 1. Calculate the copepod density for each species for each cruise-station

colnames(cop\_compo) <- c("1", "3", "4", "6", "13", "16", "19a", "21", "23a", "25a", "18", "19b", "20", "22a", "23b", "25b", "27a", "29a", "Aa", "Ba", "Ca", "Da", "E", "F", "G", "22b", "23c", "25c", "27b", "29b", "Ab", "Bb", "Cb", "Db")  
head(cop\_compo)

## 1 3 4 6 13 16 19a 21 23a 25a 18 19b 20 22a 23b 25b 27a 29a  
## 1 0 0 0 0 0 0.00 0.00 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.30 0.00 0.00  
## 2 0 0 0 0 0 0.00 0.00 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00  
## 3 0 0 0 0 0 0.22 2.34 0 2.51 1.62 0.00 0.00 0.00 0.00 0.00 0.00 1.52 0.30  
## 4 0 0 0 0 0 0.00 0.00 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00  
## 5 0 0 0 0 0 0.00 0.00 0 0.00 0.00 4.07 1.56 1.08 4.83 8.49 1.49 0.00 0.00  
## 6 0 0 0 0 0 0.00 0.00 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.76 1.51  
## Aa Ba Ca Da E F G 22b 23c 25c 27b 29b Ab Bb Cb  
## 1 0.00 0.00 0.00 0.00 0.00 0.00 0.0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00  
## 2 0.00 0.00 0.00 0.00 0.00 0.00 0.0 0.00 0.00 0.00 0.00 0.42 0.00 0.00 0.00  
## 3 3.06 1.35 1.24 0.62 2.92 0.31 1.4 0.00 0.00 0.00 0.00 0.00 0.00 2.22 1.13  
## 4 0.26 0.00 0.00 0.00 0.00 0.00 0.0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00  
## 5 0.26 0.00 0.00 0.00 0.32 1.53 0.0 19.42 51.76 2.81 3.85 6.28 0.55 0.28 0.00  
## 6 0.00 0.00 0.00 0.00 0.00 0.00 0.0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00  
## Db  
## 1 0.00  
## 2 0.00  
## 3 0.93  
## 4 0.00  
## 5 0.62  
## 6 0.00

rownames(cop\_den) <- colnames(cop\_compo)  
head(cop\_den)

## V1  
## 1 1119  
## 3 1153  
## 4 1719  
## 6 855  
## 13 1246  
## 16 2123

#multiply proportion of composition and density  
sp\_den\_cr <- matrix(0, nrow = length(cop\_sp$V1), ncol = length(cop\_den$V1))  
for(i in 1:length(cop\_sp$V1)){  
 for(j in 1:length(cop\_den$V1)){  
 sp\_den\_cr[i, j] <- cop\_compo[i, j]/100 \* cop\_den[j,]  
 }  
}  
sp\_den\_cr.df <- as.data.frame(sp\_den\_cr)  
colnames(sp\_den\_cr.df) <- colnames(cop\_compo)  
head(sp\_den\_cr.df)

## 1 3 4 6 13 16 19a 21 23a 25a 18 19b 20 22a  
## 1 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000 0.0000  
## 2 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000 0.0000  
## 3 0 0 0 0 0 4.6706 27.1206 0 33.9101 15.552 0.0000 0.00 0.0000 0.0000  
## 4 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000 0.0000  
## 5 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 119.9022 29.64 16.2864 195.2769  
## 6 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000 0.0000  
## 23b 25b 27a 29a Aa Ba Ca Da E  
## 1 0.0000 18.9960 0.0000 0.0000 0.0000 0.0000 0.000 0.0000 0.0000  
## 2 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000 0.0000 0.0000  
## 3 0.0000 0.0000 31.9352 14.4690 48.5928 39.0825 47.988 8.3948 160.3664  
## 4 0.0000 0.0000 0.0000 0.0000 4.1288 0.0000 0.000 0.0000 0.0000  
## 5 417.6231 94.3468 0.0000 0.0000 4.1288 0.0000 0.000 0.0000 17.5744  
## 6 0.0000 0.0000 15.9676 72.8273 0.0000 0.0000 0.000 0.0000 0.0000  
## F G 22b 23c 25c 27b 29b Ab Bb  
## 1 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000 0.00000  
## 2 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.175308 0.00000 0.00000  
## 3 24.5458 17.78 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000 1.10334  
## 4 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000 0.00000  
## 5 121.1454 0.00 69.54302 135.4973 2.266827 1.77331 2.621272 0.05093 0.13916  
## 6 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000 0.00000  
## Cb Db  
## 1 0.000000 0.000000  
## 2 0.000000 0.000000  
## 3 0.290862 0.434031  
## 4 0.000000 0.000000  
## 5 0.000000 0.289354  
## 6 0.000000 0.000000

#sum of the same station  
sp\_den\_cr.df$`19a` <- sp\_den\_cr.df$`19a` + sp\_den\_cr.df$`19b`  
  
sp\_den\_cr.df$`23a` <- sp\_den\_cr.df$`23a` + sp\_den\_cr.df$`23b` + sp\_den\_cr.df$`23c`  
  
sp\_den\_cr.df$`25a` <- sp\_den\_cr.df$`25a` + sp\_den\_cr.df$`25b` + sp\_den\_cr.df$`25c`  
  
sp\_den\_cr.df$`22a` <- sp\_den\_cr.df$`22a` + sp\_den\_cr.df$`22b`  
  
sp\_den\_cr.df$`27a` <- sp\_den\_cr.df$`27a` + sp\_den\_cr.df$`27b`  
  
sp\_den\_cr.df$`29a` <- sp\_den\_cr.df$`29a` + sp\_den\_cr.df$`29b`  
  
sp\_den\_cr.df$Aa <- sp\_den\_cr.df$Aa + sp\_den\_cr.df$Ab  
  
sp\_den\_cr.df$Ba <- sp\_den\_cr.df$Ba + sp\_den\_cr.df$Bb  
  
sp\_den\_cr.df$Ca <- sp\_den\_cr.df$Ca + sp\_den\_cr.df$Cb  
  
sp\_den\_cr.df$Da <- sp\_den\_cr.df$Da + sp\_den\_cr.df$Db  
  
sp\_den\_cr.df <- sp\_den\_cr.df[, -c(12, 15, 16, 26, 27, 28, 29, 30, 31,32,33,34)]  
colnames(sp\_den\_cr.df) <- c("1", "3", "4", "6", "13", "16", "19", "21", "23", "25", "18", "20", "22", "27", "29", "A", "B", "C", "D", "E", "F", "G")  
rownames(sp\_den\_cr.df) <- cop\_sp$V1  
  
#species density at each cruise-station  
head(sp\_den\_cr.df)

## 1 3 4 6 13 16 19 21 23 25 18  
## Acartia bifilosa 0 0 0 0 0 0.0000 0.0000 0 0.0000 18.99600 0.0000  
## Acartia erythraea 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.00000 0.0000  
## Acartia negligence 0 0 0 0 0 4.6706 27.1206 0 33.9101 15.55200 0.0000  
## Acartia omori 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.00000 0.0000  
## Acartia pacifica 0 0 0 0 0 0.0000 29.6400 0 553.1204 96.61363 119.9022  
## Acartia sp 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.00000 0.0000  
## 20 22 27 29 A B  
## Acartia bifilosa 0.0000 0.0000 0.00000 0.000000 0.00000 0.00000  
## Acartia erythraea 0.0000 0.0000 0.00000 0.175308 0.00000 0.00000  
## Acartia negligence 0.0000 0.0000 31.93520 14.469000 48.59280 40.18584  
## Acartia omori 0.0000 0.0000 0.00000 0.000000 4.12880 0.00000  
## Acartia pacifica 16.2864 264.8199 1.77331 2.621272 4.17973 0.13916  
## Acartia sp 0.0000 0.0000 15.96760 72.827300 0.00000 0.00000  
## C D E F G  
## Acartia bifilosa 0.00000 0.000000 0.0000 0.0000 0.00  
## Acartia erythraea 0.00000 0.000000 0.0000 0.0000 0.00  
## Acartia negligence 48.27886 8.828831 160.3664 24.5458 17.78  
## Acartia omori 0.00000 0.000000 0.0000 0.0000 0.00  
## Acartia pacifica 0.00000 0.289354 17.5744 121.1454 0.00  
## Acartia sp 0.00000 0.000000 0.0000 0.0000 0.00

# 2. For each cruise-station, calculate the species richness (number of species) and Shannon diversity index

### 2(1). species richness

sp\_rich <- ifelse(sp\_den\_cr.df > 0, 1, 0)   
sp\_rich.df <- as.data.frame(sp\_rich)  
  
#calculate species number at each cruise-station  
richness <- data.frame(colSums(sp\_rich.df))  
richness

## colSums.sp\_rich.df.  
## 1 6  
## 3 12  
## 4 8  
## 6 9  
## 13 31  
## 16 29  
## 19 65  
## 21 7  
## 23 66  
## 25 68  
## 18 39  
## 20 32  
## 22 33  
## 27 60  
## 29 39  
## A 76  
## B 83  
## C 79  
## D 72  
## E 44  
## F 38  
## G 25

### 2(2). Shannon index

#species proportion at each cruise-station  
cop\_compo$`19a` <- cop\_compo$`19a` + cop\_compo$`19b`  
  
cop\_compo$`23a` <- cop\_compo$`23a` + cop\_compo$`23b` + cop\_compo$`23c`  
  
cop\_compo$`25a` <- cop\_compo$`25a` + cop\_compo$`25b` + cop\_compo$`25c`  
  
cop\_compo$`22a` <- cop\_compo$`22a` + cop\_compo$`22b`  
  
cop\_compo$`27a` <- cop\_compo$`27a` + cop\_compo$`27b`  
  
cop\_compo$`29a` <- cop\_compo$`29a` + cop\_compo$`29b`  
  
cop\_compo$Aa <- cop\_compo$Aa + cop\_compo$Ab  
  
cop\_compo$Ba <- cop\_compo$Ba + cop\_compo$Bb  
  
cop\_compo$Ca <- cop\_compo$Ca + cop\_compo$Cb  
  
cop\_compo$Da <- cop\_compo$Da + cop\_compo$Db  
  
cop\_compo <- cop\_compo[, -c(12, 15, 16, 26, 27, 28, 29, 30, 31,32,33,34)]  
colnames(cop\_compo) <- c("1", "3", "4", "6", "13", "16", "19", "21", "23", "25", "18", "20", "22", "27", "29", "A", "B", "C", "D", "E", "F", "G")  
rownames(cop\_compo) <- cop\_sp$V1  
head(cop\_compo)

## 1 3 4 6 13 16 19 21 23 25 18 20 22 27 29  
## Acartia bifilosa 0 0 0 0 0 0.00 0.00 0 0.00 0.30 0.00 0.00 0.00 0.00 0.00  
## Acartia erythraea 0 0 0 0 0 0.00 0.00 0 0.00 0.00 0.00 0.00 0.00 0.00 0.42  
## Acartia negligence 0 0 0 0 0 0.22 2.34 0 2.51 1.62 0.00 0.00 0.00 1.52 0.30  
## Acartia omori 0 0 0 0 0 0.00 0.00 0 0.00 0.00 0.00 0.00 0.00 0.00 0.00  
## Acartia pacifica 0 0 0 0 0 0.00 1.56 0 60.25 4.30 4.07 1.08 24.25 3.85 6.28  
## Acartia sp 0 0 0 0 0 0.00 0.00 0 0.00 0.00 0.00 0.00 0.00 0.76 1.51  
## A B C D E F G  
## Acartia bifilosa 0.00 0.00 0.00 0.00 0.00 0.00 0.0  
## Acartia erythraea 0.00 0.00 0.00 0.00 0.00 0.00 0.0  
## Acartia negligence 3.06 3.57 2.37 1.55 2.92 0.31 1.4  
## Acartia omori 0.26 0.00 0.00 0.00 0.00 0.00 0.0  
## Acartia pacifica 0.81 0.28 0.00 0.62 0.32 1.53 0.0  
## Acartia sp 0.00 0.00 0.00 0.00 0.00 0.00 0.0

# calculate the index  
shannon\_index <- data.frame(-colSums(cop\_compo/100 \* log(cop\_compo / 100), na.rm = T))  
colnames(shannon\_index) <- "Shannon index"  
shannon\_index

## Shannon index  
## 1 1.080782  
## 3 1.256126  
## 4 1.045406  
## 6 1.114551  
## 13 2.144528  
## 16 1.413396  
## 19 4.595925  
## 21 1.566761  
## 23 5.902687  
## 25 5.414535  
## 18 2.841237  
## 20 2.568974  
## 22 4.003199  
## 27 4.900656  
## 29 3.988412  
## A 5.273402  
## B 5.252945  
## C 5.349971  
## D 5.261798  
## E 3.020831  
## F 2.890280  
## G 1.692271

# 3. For each of the dominant species (species >=5% of total composition in any cruise-station regardless of the seasons), calculate the average density by seasons.

setwd("C:/Users/dan91/Rstudio/stat\_data/mock\_dataTS")  
cop\_compo\_season <- read.table("Copepod\_composition.txt", header = T)  
sp\_den\_cr\_season <- matrix(0, nrow = length(cop\_sp$V1), ncol = length(cop\_den$V1))  
for(i in 1:length(cop\_sp$V1)){  
 for(j in 1:length(cop\_den$V1)){  
 sp\_den\_cr\_season[i, j] <- cop\_compo\_season[i, j]/100 \* cop\_den[j,]  
 }  
}  
sp\_den\_cr\_season.df <- as.data.frame(sp\_den\_cr\_season)  
colnames(sp\_den\_cr\_season.df) <- colnames(cop\_compo\_season)  
head(sp\_den\_cr\_season.df)

## X.p1 p3 p4 p6 p13 p16 p19 p21 p23 p25 s18 s19 s20  
## 1 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000  
## 2 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000  
## 3 0 0 0 0 0 4.6706 27.1206 0 33.9101 15.552 0.0000 0.00 0.0000  
## 4 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000  
## 5 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 119.9022 29.64 16.2864  
## 6 0 0 0 0 0 0.0000 0.0000 0 0.0000 0.000 0.0000 0.00 0.0000  
## s22 s23 s25 s27 s29 sA sB sC sD  
## 1 0.0000 0.0000 18.9960 0.0000 0.0000 0.0000 0.0000 0.000 0.0000  
## 2 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000 0.0000  
## 3 0.0000 0.0000 0.0000 31.9352 14.4690 48.5928 39.0825 47.988 8.3948  
## 4 0.0000 0.0000 0.0000 0.0000 0.0000 4.1288 0.0000 0.000 0.0000  
## 5 195.2769 417.6231 94.3468 0.0000 0.0000 4.1288 0.0000 0.000 0.0000  
## 6 0.0000 0.0000 0.0000 15.9676 72.8273 0.0000 0.0000 0.000 0.0000  
## sE sF sG w22 w23 w25 w27 w29 wA  
## 1 0.0000 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000  
## 2 0.0000 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.175308 0.00000  
## 3 160.3664 24.5458 17.78 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000  
## 4 0.0000 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000  
## 5 17.5744 121.1454 0.00 69.54302 135.4973 2.266827 1.77331 2.621272 0.05093  
## 6 0.0000 0.0000 0.00 0.00000 0.0000 0.000000 0.00000 0.000000 0.00000  
## wB wC wD  
## 1 0.00000 0.000000 0.000000  
## 2 0.00000 0.000000 0.000000  
## 3 1.10334 0.290862 0.434031  
## 4 0.00000 0.000000 0.000000  
## 5 0.13916 0.000000 0.289354  
## 6 0.00000 0.000000 0.000000

#find out dominant species  
dominant\_sp.df <- data.frame(ifelse(cop\_compo\_season >= 5, 1, 0))  
dominant\_species <- rownames(dominant\_sp.df)[rowSums(dominant\_sp.df) >= 1]  
as.numeric(dominant\_species)

## [1] 5 14 15 35 40 51 54 60 72 79 84 85 86 88 106 112 117 118 123  
## [20] 126 135 142 145 147 161 164 169

cop\_sp[dominant\_species,]

## [1] "Acartia pacifica"   
## [2] "Calanus sinicus"   
## [3] "Canthocalanus pauper"   
## [4] "Clausocalanus furcatus"   
## [5] "Clausocalanus minor"   
## [6] "Subeucalanus pileatus"   
## [7] "Subeucalanus copepodid"   
## [8] "Euchaeta copepodid"   
## [9] "Acrocalanus gibber"   
## [10] "Calocalanus pavoninus"   
## [11] "Paracalanus aculeatus"   
## [12] "Paracalanus pavus"   
## [13] "Paracalanus serrulus"   
## [14] "Parvocalanus crassirostris"   
## [15] "Scolecithricella longispinosa"   
## [16] "Temora turbinata"   
## [17] "Oithona attenuata"   
## [18] "Oithona brevicornis"   
## [19] "Oithona plumifera"   
## [20] "Oithona similis"   
## [21] "Euterpina acutifrons"   
## [22] "Corycaeus (Ditrichocorycaeus) affinis"  
## [23] "Corycaeus (Ditrichocorycaeus) dahli"   
## [24] "Corycaeus (Ditrichocorycaeus) lubbocki"  
## [25] "Corycaeidae copepodid"   
## [26] "Oncaea conifera"   
## [27] "Oncaea venusta"

#calculate the mean of three seasons respectively  
dominant\_sp\_den <- sp\_den\_cr\_season.df[dominant\_species, ]  
dominant\_sp\_den$spring\_mean <- rowMeans(dominant\_sp\_den[, 1:10])  
dominant\_sp\_den$summer\_mean <- rowMeans(dominant\_sp\_den[, 11:25])  
dominant\_sp\_den$winter\_mean <- rowMeans(dominant\_sp\_den[, 26:34])  
  
dominant\_sp\_den\_season <- dominant\_sp\_den[, 35:37]  
rownames(dominant\_sp\_den\_season) <- cop\_sp[dominant\_species,]  
dominant\_sp\_den\_season

## spring\_mean summer\_mean winter\_mean  
## Acartia pacifica 0.00000 67.728267 23.57568900  
## Calanus sinicus 86.46599 2.500233 4.60278089  
## Canthocalanus pauper 3.42582 204.692020 1.26660244  
## Clausocalanus furcatus 11.92657 63.691307 0.67185067  
## Clausocalanus minor 11.44634 23.415760 0.68068656  
## Subeucalanus pileatus 0.00000 28.013247 0.86376644  
## Subeucalanus copepodid 4.87340 40.484947 0.22946133  
## Euchaeta copepodid 28.35273 41.598293 23.15445778  
## Acrocalanus gibber 10.71096 156.406840 0.99605678  
## Calocalanus pavoninus 1.45908 26.636913 0.00000000  
## Paracalanus aculeatus 54.09300 18.917167 9.39986078  
## Paracalanus pavus 626.17119 414.204633 18.13617244  
## Paracalanus serrulus 0.00000 107.269380 0.02491267  
## Parvocalanus crassirostris 11.22283 419.718387 0.03758444  
## Scolecithricella longispinosa 0.76840 20.457953 4.46338178  
## Temora turbinata 3.69089 267.764627 0.97009244  
## Oithona attenuata 0.47348 134.511173 0.00000000  
## Oithona brevicornis 0.23674 53.421613 0.00000000  
## Oithona plumifera 28.78133 83.105013 0.99051589  
## Oithona similis 115.08926 38.426887 0.00000000  
## Euterpina acutifrons 7.36190 197.810427 0.00000000  
## Corycaeus (Ditrichocorycaeus) affinis 161.22854 6.445867 0.00000000  
## Corycaeus (Ditrichocorycaeus) dahli 1.75298 97.267093 0.08584778  
## Corycaeus (Ditrichocorycaeus) lubbocki 0.00000 75.731147 1.56444311  
## Corycaeidae copepodid 21.42939 0.000000 0.00000000  
## Oncaea conifera 1.12309 116.335620 0.95004678  
## Oncaea venusta 33.96802 341.826987 1.25436611