Stat 415, section 3. HW 1 solutions

This is written in Word, which changes characters to look pretty. You can cut and paste most of the code below into R. You can not cut and paste code with quotes. R doesn’t recognize Word’s pretty quotes.

1) Echinacea metabolites

a) Pur504 (row 30)

You are looking for the sample with the smallest distance. That’s most easily done by looking at the Pur307 row of the matrix containing the distances.

lipo.m <- as.matrix(lipo[,-(1:2)])

dimnames(lipo.m)[[1]] <- lipo$access

lipo.canb <- vegdist(lipo.m, ‘canb’)

temp <- as.matrix(lipo.canb)

temp[‘Pur307’, ]

The smallest is Pur504 (row 30) with a distance of 0.426

b) There are number of metabolites that are in Pur307 and Pur504, but not in Ang267.

lipo.m[c(1,28, 30),]

c) Pur 313

lipo.pa <- decostand(lipo.m, 'pa')

lipo.jd <- vegdist(lipo.pa, 'jacc')

temp <- as.matrix(lipo.jd)

temp[‘Pur307’, ]

The smallest is Pur313 (row 29).

d) Most metabolites are present in all 3 samples or absent in all three samples. There are two that are not. Pur307 and 313 differ in metabolite (unknown\_A2). Pur307 and 504 differ in 2 (unknown\_A2 unknown\_B7).

lipo.pa[28:30,]

e) I expect them to be quite similar because the two distance measures are highly correlated.

The plot is on the next page



f) 3 Dimensions. Stress for the nMDS 2D solution is 11.5%. That for the 3D solution is 7.9%

lipo.mds2 <- metaMDS(lipo.canb, k = 2, expand=F)

# because we’re starting with the distance matrix autotransform=F is not necessary

#or, to get the species scores, use the data matrix

lipo.mds2 <- metaMDS(lipo.m, ‘canb’, k=2, autotransform=F, expand=F)

lipo.mds2 <- metaMDS(lipo.canb, k = 2, expand=F)

g) The very obvious ones are Pur, Lae, Tenn. All three Pur samples and all Tenn samples are on the edge of the plot of samples, close to each other and relatively well separated from the rest. The three Lae samples are also close to each other in the middle of the plot but separated from the other species.

ordirgl(lipo.mds3)

Note: Pal215 is one sample that stands out from the rest, but the other Pal samples are quite distant.

h) Again, Lae, Pur, and Tenn are the species relatively well separated from the rest (plot on next page).

ordiplot(lipo.mds2,type=’t’,disp=’sites’)



i) Your explanation here is more important than your specific answer.

I would use the 2D plot because it represents the same features seen in the 3D plot, even though the stress is slightly above 10%.

j) Unknown B1, Unknown B6, amide 16, amide 13.

Note: You are looking for the compounds (species scores to vegan) that are plotted in the direction of the Tenn samples. To get the species scores, you need to do the mds on the raw data matrix.

ordipointlabel(lipo.mds2)

k) Yes, the plot does.



amide 16 increases towards the locations of the Tenn samples.

ordisurf(lipo.mds2, lipo.m[,'amide\_16'])

l) Yes, the arrow appropriate summarizes the variation in amide 16. The arrow points in the direction of increasing amounts of amide 16, the contours from ordisurf are reasonably parallel and equi-distant.



2) Missouri River fish data

a) Bray-Curtis after standardizing to site total and Morisita-Horn are very highly correlated. Jaccard is much less correlated. BC and MH will give similar results; Jaccard may give very different results.

Note: the curvature in BC vs MH plot is not relevant (at least for nMDS) because nMDS is based on ranks, not linear correlation.

fish.bray <- vegdist(decostand(fish.m,’total’), ‘bray’)

fish.mh <- vegdist(fish.m, ‘horn’)

fish.j <- vegdist(decostand(fish.m, ‘pa’), ‘jacc’)

pairs(cbind(fish.bray, fish.mh, fish.j))



b) Yes. Standardizing by site total removes the dependence on total abundance. BC is most sensitive to differences in the abundant species. MH would have also been appropriate.

c) You would need 4 dimensions if you strictly demanded stress < 10%.

Stress for a 2D solution is 0.143, for a 3D solution is 0.102, for a 4D solution is ca 0.080.

fish.mds2 <- metaMDS(fish.bray, k=2)

fish.mds3 <- metaMDS(fish.bray, k=3, trymax=200)

fish.mds4 <- metaMDS(fish.bray, k=4, trymax=200)

Note: it was hard to find the best 3D solution and even harder to find the best 4D.

trymax= tells metaMDS to try even more starting positions.

d) gear is the only factor with visually obvious association with species composition.

The convex hulls for sites and habitats overlap each other, but those for some gears don’t overlap.

plot(fish.mds2,disp='sites')

ordihull(fish.mds2,fish.env$site,label=T)

title("Sites")

plot(fish.mds2,disp='sites')

ordihull(fish.mds2,fish.env$habitat,label=T)

title("Habitats")

plot(fish.mds2,disp='sites')

ordihull(fish.mds2,fish.env$gear,label=T)

title("Gears")





 

e) OcLH02 is very different from the others. The two species strongly associated with that sample are GZSD and BKCP.

Note: You are looking at for a point well separated from the others. The species most strongly associated with that sample are those with species scores closest to OcLH02.

f) When this HW was assigned, we knew about two ways to answer this question: visual assessment, e.g. using convex hulls and testing association between the ordination and the habitat using envfit. Either was appropriate.

In neither approach is there evidence of a difference.

Visually:



fish.lhmds <- metaMDS(decostand(fish.lh, ‘total’), autotransform=F, expand=F)

plot(fish.lhmds, disp=’sites’)

ordihull(fish.lhmds, fish.envlh$habitat)

The envfit test has a p-value of 0.97. (because this is permutation based, you may have a slightly different p-value).

No evidence of a difference among habitats.

envfit(fish.lhmds, fish.envlh$habitat)