homework3

Jasmine Williamson

2024-10-14

Question 1) Given the following matrices A and B, perform matrix addition, subtraction, and multiplication. Verify if the matrices can be divided directly.

A =
1 2 3
4 5 6
7 8 9

B =
9 8 7
6 5 4
3 2 1

A + B = (straight across)
10 10 10
10 10 10
10 10 10

A - B = (straight across)
-8 -6 -4
-2 0 2
4 6 8

A x B = (mult each row in A by each col in B, sum)
30 24 18
84 69 54
138 114 90

The determinate of B is 0, so its inverse does not exist and division isnt possible.

Question 2) Calculate the determinant and rank of the matrix C:

C =
4 3 2
2 1 0
1 2 3

$$det(C) = C11(C22C33 - C23C32) - C12(C21C33 - C23C31) + C13(C21C32 - C22C31$$

det(C) = 0

rank = 0

Question 3) Find the eigenvalues and eigenvectors of the matrix D:

D =
5 4
1 2

Eigenvalues

$$det(D - \lambda I)$$

Answer: #1= 6, #2= 1

Using R function 'eigen':

```
## Eigenvalues:
```

```
## [1] 6 1
```

Eigenvectors

$$(D - \lambda I)v = 0$$

v1=
4
1

v2=
-1
1

Using R function 'eigen':

```
## Eigenvectors:
```

```
##           [,1]       [,2]
## [1,] 0.9701425 -0.7071068
## [2,] 0.2425356  0.7071068
```

Apparently the Eigenvectors given by the R code are different from the answers I calculated because R normalizes them.

Question 4) Describe the difference between "R-mode" and "Q-mode" analyses. What kinds of study questions could be answered with each? Which "mode" is most relevant to your data and study question(s)?

Q-mode analysis is used to understand relationships among objects (rows). These methods can be used for binary presence/absence data. I have aggregated my count data to presence/absence for occupancy modelling but I dont think I'll use that for any of these analyses (why would I when I have count data?).

It could be interesting to see how the sites cluster based on this data, and if sites with different treatments group together because of similar patterns. Possible Questions: How do treatment groups compare in terms of environmental data and salamander abundance? Do certain environmental conditions lead to distinct groupings with higher or lower salamander abundance?

R-mode analysis is used to understand relationships among descriptors (columns). You can look for patterns, correlation, covariance, etc within your variables.

With R-mode, I would look at how the environmental and salamander data relate to each other. Possible questions: How do environmental covariates correlate with salamander abundance? Does treatment type lead to significant differences in covariates? What is the combined effect of environmental covariates on salamander abundance?

I think a lot of these would be interesting to look at. I should probably focus on the R-mode methods and look at how environmental variables affect salamanders.

Question 5) Identify three different scenarios where you would choose to use Jaccard similarity, Bray-Curtis dissimilarity, and Euclidean distance matrices, respectively. Explain why each association matrix is appropriate for the given scenario.

1. Jaccard similarity would be used with binary data to assess the composition of species across sites. It is an asymmetric measure of similarity, which means it ignores double zeros and is not affected by the double zero problem.
2. Bray-Curtis dissimilarity would be used to compare communities in different sites using abundance data.
3. Euclidean distance matrices would be used to see how different the study sites are using numeric environmental data.

Question 6) Which association matrix(ces) is/are most appropriate for your data? Justify your choice.

Species data - I only have two, do I do any of this? *Species data generally uses asymmetrical dissimilarity matrix. For this I'm not sure if Bray-Curtis or Hellinger would be a better fit, but Bray-Curtis seems like the most used choice. I could ue Chi-square, since it is particularly useful for when rare species drive trends - but in my data, the generally "rare" species does not look rare because I targeted my surveys for that species.*

For environmental data: **Pearson's**: looking at correlations between environmental data and abundance (linear, continuous). could switch to kendall if not all of my variables are linear. **Covariance matrix**: measure how variables change together, could help identify the variables that are driving most of the change between sites/treatments

Question 7) In ecological data analysis, the "double zero problem" can affect the interpretation of similarity and distance matrices. Explain what the double zero problem is and how it can influence the results of multivariate analyses. Do you foresee this being an issue for your data? If so, how will you address this problem through the preparation of your data prior to analysis and/or selection of an association matrix?

The double zero problem is a situation where you have zeros for two sites and your analysis assumes some similarity between sites based on that commonality of no species present. This is not a safe assumption, because your site characteristics could fall on the high or low end of the range of habitability of your species, and could be very different. I have a lot of zeros in my abundance data and I honestly do not understand what to do about that, but I assume it will be an issue. I reduced my matrix to the site level rather than the plot level to reduce some zeros, and that has helped distributions a bit already.