|  |
| --- |
| [Type the company name] |
| STAT 445 Assignment 8 |
| Kun Yang 301178299 |
|  |
|  |
|  |

|  |
| --- |
|  |

**Problem 12.19**

**> dist\_mat<-as.matrix(read.csv("T12-13.csv",header=F))**

**> dist\_mat[upper.tri(dist\_mat)]=as.integer(0)**

**> dist\_mat=dist\_mat+t(dist\_mat)**

**> rownames(dist\_mat)=c('P1980918','P1931131','P1550960','P1530987','P1361024','P1351005','P1340945','P1311137','P1301062')**

**> colnames(dist\_mat)=c('P1980918','P1931131','P1550960','P1530987','P1361024','P1351005','P1340945','P1311137','P1301062')**

**> dist\_mat**

P1980918 P1931131 P1550960 P1530987 P1361024 P1351005 P1340945 P1311137 P1301062

P1980918 0.000 2.202 1.004 1.108 1.122 0.914 0.914 2.056 1.608

P1931131 2.202 0.000 2.025 1.943 1.870 2.070 2.186 2.055 1.722

P1550960 1.004 2.025 0.000 0.233 0.719 0.719 0.452 1.986 1.358

P1530987 1.108 1.943 0.233 0.000 0.541 0.679 0.681 1.990 1.168

P1361024 1.122 1.870 0.719 0.541 0.000 0.539 1.102 1.963 0.681

P1351005 0.914 2.070 0.719 0.679 0.539 0.000 0.916 2.056 1.005

P1340945 0.914 2.186 0.452 0.681 1.102 0.916 0.000 2.027 1.719

P1311137 2.056 2.055 1.986 1.990 1.963 2.056 2.027 0.000 1.991

P1301062 1.608 1.722 1.358 1.168 0.681 1.005 1.719 1.991 0.000

**### Multidimensional scaling with 3 dimensions**

**> (nMDS3=isoMDS(d=dist\_mat,k = 3))**

initial value 9.082563

iter 5 value 4.395974

iter 10 value 0.848282

iter 15 value 0.363813

iter 20 value 0.073085

final value 0.005127

converged

$points

[,1] [,2] [,3]

P1980918 1.25573619 -0.4982819 -0.2577593

P1931131 -3.02742976 1.6219218 0.4510898

P1550960 0.83286023 0.1117732 0.2913323

P1530987 0.65551562 0.3109390 0.1647135

P1361024 0.53852356 0.3240394 -0.3206008

P1351005 0.93936056 0.1528150 -0.4313809

P1340945 1.08230924 -0.1559933 0.4077919

P1311137 -2.36419642 -2.4960068 0.4311717

P1301062 0.08732078 0.6287934 -0.7363581

$stress

[1] 0.005126824

**### Multidimensional scaling with 4 dimensions**

**> (nMDS4=isoMDS(d=dist\_mat,k = 4))**

initial value 0.905054

iter 5 value 0.057620

iter 10 value 0.026707

final value 0.008832

converged

$points

[,1] [,2] [,3] [,4]

P1980918 0.4992704 -0.2840371 0.24506069 0.62504100

P1931131 -1.3225297 0.6767956 0.61536939 0.04767958

P1550960 0.4546521 -0.1167087 0.24563492 -0.29669942

P1530987 0.3968788 0.1085906 0.01844162 -0.32916207

P1361024 0.2130110 0.2877619 -0.30542894 -0.08179472

P1351005 0.5490105 0.1526313 -0.24341621 0.22136464

P1340945 0.5755191 -0.3296088 0.43808297 -0.18762960

P1311137 -1.1337425 -1.1201990 -0.30474728 -0.05085409

P1301062 -0.2320698 0.6247742 -0.70899716 0.05205468

$stress

[1] 0.008832013

**### Multidimensional scaling with 5 dimensions**

**> (nMDS5=isoMDS(d=dist\_mat,k = 5))**

initial value 0.002672

final value 0.000000

converged

$points

[,1] [,2] [,3] [,4] [,5]

P1980918 0.5119010 -0.27797662 0.2421046 0.67644342 0.118956423

P1931131 -1.3184960 0.69177870 0.6229927 0.04985327 -0.023615843

P1550960 0.4696575 -0.07075632 0.1855302 -0.30157380 0.058979348

P1530987 0.3875460 0.08742269 0.0492339 -0.34361942 0.101767639

P1361024 0.2340266 0.29527907 -0.3249103 -0.05204965 0.121534837

P1351005 0.4688497 0.13734912 -0.2187626 0.13932145 -0.281320804

P1340945 0.5812702 -0.34886752 0.4570202 -0.17853821 -0.101485287

P1311137 -1.1180751 -1.12218943 -0.3159596 -0.05212138 -0.005024992

P1301062 -0.2166798 0.60796030 -0.6972491 0.06228432 0.010208679

$stress

[1] 8.786885e-14

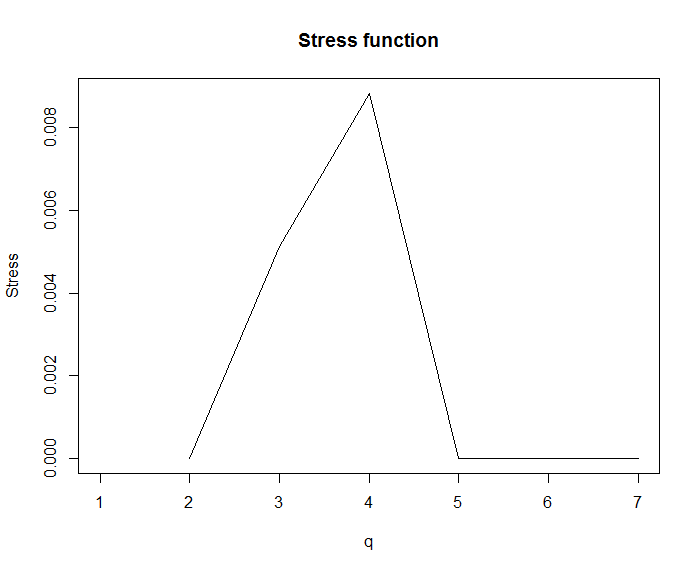
***### Plot minimum stress(q) versus q.***

**stress\_vector=rep(NA,7)**

**stress\_vector[2]=nMDS5$stress**

**for (i in (3:7)){ stress\_vector[i]=isoMDS(d=dist\_mat,k = i)$stress}**

**plot(stress\_vector,type="l",xlab="q",ylab="Stress",main="Stress function")**



***The minimum stress of configuration for q=5-dimentional solution is almost equal 0.***

***So q=5 is the best method.***

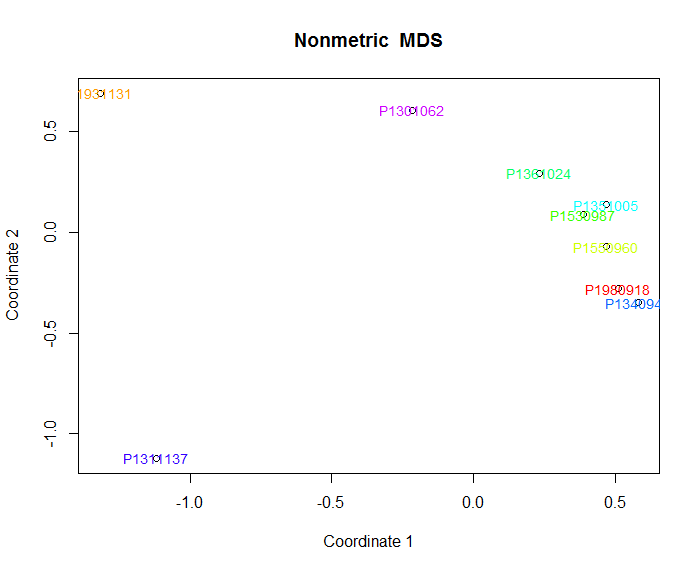
**### Locate the sites in 2 dimensions (first 2 principal component) using the coordinate for q=5-dimensional solution.**

**> x=nMDS5$points[,1]**

**> y=nMDS5$points[,2]**

**> plot(x,y, xlab="Coordinate 1", ylab="Coordinate 2", main="Nonmetric MDS")**

**> text(x, y, labels = rownames(dist\_mat), cex=.9,col=rainbow(10))**



**This method shows that the time of a site can be tracked by a certain cycling pattern.**

***2. Consider the following set of three variables:***

𝑋1 = the height of a car window opening produced in a carefully monitored production plant

𝑋2 = the width of the same window opening

𝑋3 = the weight of the door frame surrounding the above window

Would you expect the vector to have an approximate multivariate normal distribution? Explain your reasoning in at most three sentences.

Yes, I expect the vector to have an approximate multivariate normal distribution.

We can know that each of the variables fit the central limit theorem, they are all normally distributed.

In this car window case, we know the vector composed of elements (X1, X2, X3) that are each normally distributed.

***3. Let Y = the number of lodgepole pine seedlings (young trees) found in a randomly located 1-meter by 1-meter plot in a recently burned-over stand of lodgepole pine trees.***

***a. Which of the following distributions would be the least inappropriate distribution to use to model this distribution?***

i. Binomial ***ii. Poisson***  iii. Normal

***b. Provide one reason why you would anticipate that the assumptions underlying this distribution would be violated.***

**The number of lodgepole pine seedlings in the 1-meter by 1-meter plot may change when we choose another non-overlapping 1-meter by 1-meter plot. So it violates the assumption that the probability of an event within a certain area does not change over different area. And the independence also may be violated.**

***c. If you choose either of the two discrete distributions above, would you expect this assumption violation to lead to over- or under-dispersion relative to the formula for that distribution? Explain your reasoning in at most two sentences.***

**I would like to expect this assumption violation to lead to over-dispersion.**

**The violation in question (b) may cause extra variation and makes the observed variances are proportionally enlarged to the expected variance under the poisson assumption. So it makes the variance larger than the mean.**