

Cluster Analysis

Thomas Bosch

Technical University of Munich, Germany

05.11.2008

Abstract

The purpose of this research paper is to verify the hypothesis that the ANETT classification is based on logical criteria although these criteria are of subjective nature. To implement this, the author of this paper compared the outcomes of two different types of cluster analysis approaches using two variable sets with the result of the ANETT classification.

Keywords: Cluster analysis; Agglomerative hierarchical clustering; Single-link clustering; Ward method; K-means clustering

1. Introduction

In the year 2006 the study called ASPIS (Athens Study of Psychosis Proneness and Incidence of Schizophrenia) was accomplished (Stefanis, 2006).

The major issue of this survey is the ANETT classification dividing persons into one of the categories left handed, right handed or both hands. According to its own statement the ANETT classification bases upon subjective but logical criteria.

To proof this objectiveness the results of the ANETT classification of a sample dataset have to be compared with the outcomes of different cluster analysis approaches using two variable sets. If there can be considered similarities in the outputs it will be an indicator for the objectivity of the ANETT classification.

2. Description of the data

2.1. Individuals

The existent data is a sample of 80 individuals which is part of a larger sample of 2.500 airforce soldiers.

2.2. Variables

Within the scope of this research paper there will be relevant 25 variables.

Each airforce soldier had to answer 18 questions for the ASPIS survey (variables `des_1` to `des_18`). Nine of them for the use of the hands and three of them in each case for the utilization of feet, eyes and ears. The three possible answers were left, right and both coded by -1, 1 and 0.

The four variables hand, foot, eye and ear represent the arithmetic means of the previous variables of each body part. The range from -1 to 1 can be detected. Values close to -1 indicate left-handed individuals while values near to 1 display right handed persons.

The last variable named `anett` should classify each person as left, right or mixed handed.

2.3. Data import

The various calculations will be done with the statistic software R in its current version 2.7.2. All the R statements used to get necessary results can be read in the appendix of this paper.

The data is given in a SPSS format. For this reason you have to install an additional package in R. The package `foreign` provides functions for the data import from SPSS to R. Before using this package it has to be activated. Then the data can be read.

2.4. Data matrices

In order to build the dissimilarity matrices you have to create $(n \times p)$ data matrices with n individuals and p variables. In the framework of this research paper n will be 80.

The author used two different variable sets: `des_1` to `des_18` and `hand`, `foot`, `eye` and `ear`. In the first case p will be 18 and in second case p will be assigned to 4.

3. Cluster analysis approaches

Cluster analysis is used to allocate individuals to mutually exclusive, exhaustive groups. Individuals within a group are seen as similar (Chatfield, Collins, 1992). In this work there will be compared the ANETT classification with two different types of cluster analysis approaches: the agglomerative hierarchical clustering and the k-means clustering.

4. Agglomerative hierarchical clustering

4.1. Theoretical description of the cluster analysis approach

If you execute the agglomerative hierarchical clustering you will start with n groups of one individual and you will end with one group of n individuals. "At each stage in the process the method[...] fuse[s] individuals or groups of individuals which are closest (or most similar)" (Everitt, 2005, p. 116). "[...] A dendrogram [...] illustrates the fusions made a[t] each stage of the analysis" (Everitt, 2005, p. 117). In this diagram you can see the generated hierarchical tree and the appropriate threshold distances,

the distances in which the individuals or groups of individuals merge.

There are various methods available to compute the agglomerative hierarchical clustering of a multivariate dataset on dissimilarities: simple-link clustering, complete-link clustering, centroid-link clustering, average method, ward method, mcquitty method and median-link clustering. In the frame of this research paper there will be considered the two methods simple-link clustering and the ward method.

4.2. Description of the procedure

The first step is to generate the distance matrices from the data matrices. In order to calculate the dissimilarity matrices of the two variable sets there will be regarded three measures of dissimilarity in each agglomerative hierarchical clustering method namely the euclidean distance, the manhattan distance and the minkowski metric.

After computing the dissimilarity matrices you have to apply the clustering methods and to plot the resulting dendrograms.

The assignments of the individuals to the three different clusters (left, right and mixed) can be found by cutting the dendrograms. Theoretically, if you do not use a statistical program like R or SPSS, you have to move an imaginary horizontal line from the root of the hierarchical tree to the leaves as far as three vertical lines are cut through.

Now R provides the user with vectors containing the allocation of each individual to one of the three possible clusters. Altogether there can be compared thirteen vectors: One vector for the ANETT classification and twelve vectors including the results of the two applied agglomerative hierarchical clustering approaches. These methods are operated with three different measures of dissimilarity utilizing two amounts of variables.

These vectors were exported to an excel spreadsheet in order to verify the use of logical criteria in the ANETT classification.

The reader has the possibility to consult the appendix to look at the further calculations on the vectors resulting from the application of the different cluster analysis procedures in form of VBA statements.

The data, used for the comparison of the agglomerative hierarchical clustering approaches using different variable sets and the comparison of the agglomerative hierarchical clustering with the ANETT classification, is shown in the first table of the appendix.

4.3. Comparison of agglomerative hierarchical clustering approaches

It can be regarded that by applying either the ward method or the ANETT classification the individuals are approximately assigned in equal parts to the three clusters. If you execute the single-link clustering there will be almost all persons in the first cluster.

The differences between the different measures of dissimilarity and the variable sets can be disregarded in this case.

4.4. Comparison of agglomerative hierarchical clustering with ANETT classification

For each combination consisting of agglomerative hierarchical clustering method, measure of dissimilarity and variable set it was proofed how many individuals were assigned to the same cluster than in the ANETT classification. The respective percentage values were also calculated.

The problem is, that there is no matching between the clusters of the ANETT classification and the clusters generated by the agglomerative hierarchical clustering approaches. For this reason the author considered all the six cases of possible cluster equivalences.

The percent arithmetic means of these six cases are the results of this examination. In every combination 33.3 percent of all the individuals were allocated to the same cluster than in the ANETT classification.

It can be derived, that to build clusters in terms of the ANETT classification, the participants of the ASPIS survey although used subjective but quite logical criteria.

5. K-means clustering

5.1. Theoretical description of the cluster analysis approach

The k-means clustering partitions data into a specified number of groups k (Everitt, 2005, p. 122). "The most commonly used approach [...] is to find the partition of the n individuals into k groups, which minimizes the within-group sum of squares over all variables" (Everitt, 2005, p. 122). Within the limits of this research paper k will be three by reason there are three clusters left, right and mixed.

The k-means clustering algorithm consists of five steps:

1. Determination of k cluster centres
2. Each individual is assigned to that cluster which centre is closest to this individual
3. For each cluster its centre is calculated
4. Based on the new computed centres the individuals are allocated to the respective clusters
5. Repeat with step three if at least one individual was reassigned to another cluster

5.2. Description of the procedure

To assign each person to one of the three clusters you have to deliver the data matrix as argument to the R function kmeans. The condition is that this matrix do not has to contain any missing values. Since the data matrix including the first variable set des_1 to des_18 has missing values, this data matrix had to be transformed. By using the function substituteNA in R, the missing values were substituted for the means of each variable. In the data matrix, generated with the second amount of variables hand, foot, eye and ear, there were no missing values detected.

5.3. Comparison variable sets

There can be regarded no significant differences between the variable sets.

Considering the first variable set, the most individuals are assigned to the first cluster. If you examine the second amount of variables, you will see, that the most persons are allocated to the third cluster. Nevertheless, all three groups have almost the same count of airforce soldiers.

Assuming all six possible cases of cluster matchings between the ANETT classification and the k-means clustering, the two variable sets have the same percentage value of 33.3.

5.4. Comparison of k-means clustering with agglomerative hierarchical clustering

There is no important discrepancy between the k-means clustering and the ward method.

The only distinction, which can be made, considering the single-link clustering and the k-means clustering, is the quantity of individuals assigned to the three clusters. As described, applying the single-link method, there are almost all the individuals in the first cluster. If you use the k-means clustering, all the groups have nearly the same number of individuals.

The percentage of right cluster matching is identical.

5.5. Comparison of k-means clustering with ANETT classification

You can not detect any significant difference, regarding the allocation of the persons to the three groups. The distribution in the ANETT classification is just more balanced. The percent value of positive matching of the clusters between the k-means clustering and the ANETT classification is, like recognized in the agglomerative hierarchical clustering approaches, one third. That represents soever the incorporation of logical criteria in the model building of the ANETT classification.

The reader of this research paper has the possibility to have a look on the underlying data in table two to find in the appendix.

6. Conclusion

It has been verified that the members of the ASPIS survey included, despite of subjective thoughts, logical criteria, applying the ANETT classification.

Regarding the value 33.3 of positive cluster matchings between the ANETT classification, the two agglomerative hierarchical clustering approaches and the k-means clustering, the question of the objectiveness of the ANETT classification is clearly evidenced.

Furthermore, you can see that by adopting all the different cluster analysis approaches, the results are fundamentally the same.

Acknowledgment

This paper was supported by Dimitris Karlis, Assistant Professor at the Department of Statistics, Athens University of Economics and Business.

References

Chatfield C., Collins A.J. (1992). *Introduction to Multivariate Analysis*. Cambridge: Chapman & Hall.

Everitt B. S. (2005). *An R and S-PLUS Companion to Multivariate Analysis*. London: Springer.

Stefanis N.C., Vitoratou S., Smyrnis N., Constantinidis T., Evdokimidis I., Hatzimanolis I. et al. (2006). *Mixed handedness is associated with the Disorganization dimension of schizotypy in a young male population*. Athens: Schizophrenia research.

Appendix

Comparison of agglomerative hierarchical clustering approaches and comparison of agglomerative hierarchical clustering with ANETT classification

Table 1

Comparison of agglomerative hierarchical clustering approaches and comparison of agglomerative hierarchical clustering with ANETT classification

	1	2	3	4	5	6	7	8	9	10	11	12
13	78	76	78	78	78	78	26	15	26	33	17	33
14	1	3	1	1	1	1	8	22	8	24	14	24
15	1	1	1	1	1	1	46	43	46	23	49	23
16	20	22	20	19	17	19	43	39	43	28	41	28
17	0,3	0,3	0,3	0,2	0,2	0,2	0,5	0,5	0,5	0,4	0,5	0,4
18	20	20	20	18	17	18	28	16	28	9	30	9
19	0,3	0,3	0,3	0,2	0,2	0,2	0,4	0,2	0,4	0,1	0,4	0,1
20	29	27	29	30	32	30	51	58	51	50	44	50
21	0,4	0,3	0,4	0,4	0,4	0,4	0,6	0,7	0,6	0,6	0,5	0,6
22	31	31	31	30	31	30	23	29	23	35	33	35
23	0,4	0,4	0,4	0,4	0,4	0,4	0,3	0,4	0,3	0,4	0,4	0,4
24	29	27	29	31	32	31	14	12	14	17	6	17
25	0,4	0,3	0,4	0,4	0,4	0,4	0,2	0,2	0,2	0,2	0,1	0,2
26	31	33	31	32	31	32	1	6	1	21	6	21
27	0,4	0,4	0,4	0,4	0,4	0,4	0,0	0,1	0,0	0,3	0,1	0,3
28	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3	0,3

1: simple-link - euclidean - dev1dev18

2: simple-link - manhattan - dev1dev18

3: simple-link - minkowski - dev1dev18

4: simple-link - euclidean - means

5: simple-link - manhattan - means

6: simple-link - minkowski - means

7: ward - euclidean - dev1dev18

8: ward - manhattan - dev1dev18

9: ward - minkowski - dev1dev18

10: ward - euclidean - means

11: ward - manhattan - means

12: ward - minkowski - means

13: count individuals in cluster 1

14: count individuals in cluster 2

15: count individuals in cluster 3

16: count same cluster (assuptions: -1:1 | 0:2 | 1:3)

17: percent same cluster (assuptions: -1:1 | 0:2 | 1:3)

18: count same cluster (assuptions: -1:1 | 0:3 | 1:2)

19: percent same cluster (assuptions: -1:1 | 0:3 | 1:2)

20: count same cluster (assuptions: -1:2 | 0:1 | 1:3)
 21: percent same cluster (assuptions: -1:2 | 0:1 | 1:3)
 22: count same cluster (assuptions: -1:2 | 0:3 | 1:1)
 23: percent same cluster (assuptions: -1:2 | 0:3 | 1:1)
 24: count same cluster (assuptions: -1:3 | 0:1 | 1:2)
 25: percent same cluster (assuptions: -1:3 | 0:1 | 1:2)
 26: count same cluster (assuptions: -1:3 | 0:2 | 1:1)
 27: percent same cluster (assuptions: -1:3 | 0:2 | 1:1)
 28: percent same cluster (all assumptions)

Comparison of *k*-means clustering with ANETT classification

Table 2

Comparison of *k*-means clustering with ANETT classification

	1	2
3	42	17
4	14	22
5	24	41
6	11	53
7	0,14	0,66
8	8	27
9	0,10	0,34
10	15	40
11	0,19	0,50
12	41	20
13	0,51	0,25
14	28	7
15	0,35	0,09
16	57	13
17	0,71	0,16
18	0,33	0,33

1: k-means – des1des18
 2: k-means - means
 3: count individuals in cluster 1
 4: count individuals in cluster 2
 5: count individuals in cluster 3
 6: count same cluster (assuptions: -1:1 | 0:2 | 1:3)
 7: percent same cluster (assuptions: -1:1 | 0:2 | 1:3)
 8: count same cluster (assuptions: -1:1 | 0:3 | 1:2)
 9: percent same cluster (assuptions: -1:1 | 0:3 | 1:2)
 10: count same cluster (assuptions: -1:2 | 0:1 | 1:3)
 11: percent same cluster (assuptions: -1:2 | 0:1 | 1:3)
 12: count same cluster (assuptions: -1:2 | 0:3 | 1:1)
 13: percent same cluster (assuptions: -1:2 | 0:3 | 1:1)
 14: count same cluster (assuptions: -1:3 | 0:1 | 1:2)
 15: percent same cluster (assuptions: -1:3 | 0:1 | 1:2)
 16: count same cluster (assuptions: -1:3 | 0:2 | 1:1)
 17: percent same cluster (assuptions: -1:3 | 0:2 | 1:1)
 18: percent same cluster (all assumptions)

R statements

```
#---- install package necessary to read from spss ----
install.packages("foreign", dependencies = TRUE)
#----

#---- read data from spss ----
library(foreign)
data <- read.spss("D:/Studium/Erasmus/Courses/Multivariate
Techniques/project 1/example_for_mva_de3ioxeiria_80obs.sav")
```

```
#----

#---- data matrices ----
dataMatrixDev <-
matrix(c(data$des_1,data$des_2,data$des_3,data$des_4,data$des_5,dat
a$des_6,data$des_7,data$des_8,data$des_9,data$des_10,data$des_11,d
ata$des_12,data$des_13,data$des_14,data$des_15,data$des_16,data$de
s_17,data$des_18),80)

dataMatrixMeans <-
matrix(c(data$hand,data$foot,data$eye,data$ear),80)
#----

#---- creates an excel import file ----
write.table(data$anett, file = "excel import.csv")
#----

#---- dataMatrixDev ----

#---- set up plotting area to take three plots ----
par(mfrow=c(3,1))
#----

#---- Simple-link clustering ----

#---- euclidean ----
tree <- hclust(dist(dataMatrixDev,
method="euclidean"),method="single")
pclus(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#----

#---- manhattan ----
tree <- hclust(dist(dataMatrixDev,
method="manhattan"),method="single")
pclus(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#----

#---- minkowski ----
tree <- hclust(dist(dataMatrixDev,
method="minkowski"),method="single")
pclus(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#----

#---- Ward method ----

#---- euclidean ----
tree <- hclust(dist(dataMatrixDev,
method="euclidean"),method="ward")
pclus(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#----

#---- manhattan ----
tree <- hclust(dist(dataMatrixDev,
method="manhattan"),method="ward")
pclus(tree,hang=0.1,ylab="",xlab="",main="",sub="")
```

```

cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#----- minkowski -----
tree <- hclust(dist(dataMatrixDev,
method="minkowski"),method="ward")
pclusl(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#-----

#-----
#----- dataMatrixMeans -----

#----- simple-link clustering -----

#----- euclidean -----
tree <- hclust(dist(dataMatrixMeans,
method="euclidean"),method="single")
pclusl(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#----- manhattan -----
tree <- hclust(dist(dataMatrixMeans,
method="manhattan"),method="single")
pclusl(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#----- minkowski -----
tree <- hclust(dist(dataMatrixMeans,
method="minkowski"),method="single")
pclusl(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#-----

#----- Ward method -----

#----- euclidean -----
tree <- hclust(dist(dataMatrixMeans,
method="euclidean"),method="ward")
pclusl(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#----- manhattan -----
tree <- hclust(dist(dataMatrixMeans,
method="manhattan"),method="ward")
pclusl(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#----- minkowski -----

```

```

tree <- hclust(dist(dataMatrixMeans,
method="minkowski"),method="ward")
pclusl(tree,hang=0.1,ylab="",xlab="",main="",sub="")
cuttedtree <- cutree(tree,k=3)
write.table(cuttedtree, file = "excel import.csv")
#-----

#-----

#-----

#----- k-means clustering -----

#----- dataMatrixDEV -----

#----- substitute missing values from the data matrix through the means
of each variable -----
dataMatrixDevNoNA <- substituteNA(dataMatrixDev, type="mean")
#-----

kmeansDev <- kmeans(dataMatrixDevNoNA,3)
write.table(kmeansDev$cluster, file = "excel import.csv")
#-----

#----- dataMatrixMeans -----
kmeansMeans <- kmeans(dataMatrixMeans,3)
write.table(kmeansMeans$cluster, file = "excel import.csv")
#-----

#-----

```

VBA statements

```

Dim r, c, count As Integer
'----- annet classification - count individuals in cluster -1 -----
For r = 2 To 81
    If (Cells(r, 2).Value = -1) Then
        count = count + 1
    End If
Next
Cells(83, 2).Value = count
count = 0
'-----

'----- annet classification - count individuals in cluster 0 -----
For r = 2 To 81
    If (Cells(r, 2).Value = 0) Then
        count = count + 1
    End If
Next
Cells(84, 2).Value = count
count = 0
'-----

'----- annet classification - count individuals in cluster 1 -----
For r = 2 To 81
    If (Cells(r, 2).Value = 1) Then
        count = count + 1
    End If
Next
Cells(85, 2).Value = count
count = 0
'-----

```

```

'----- count individuals in cluster 1 -----
For c = 3 To 16
  For r = 2 To 81
    If (Cells(r, c).Value = 1) Then
      count = count + 1
    End If
  Next
  Cells(86, c).Value = count
  count = 0
Next
'-----

'----- count individuals in cluster 2 -----
For c = 3 To 16
  For r = 2 To 81
    If (Cells(r, c).Value = 2) Then
      count = count + 1
    End If
  Next
  Cells(87, c).Value = count
  count = 0
Next
'-----

'----- count individuals in cluster 3 -----
For c = 3 To 16
  For r = 2 To 81
    If (Cells(r, c).Value = 3) Then
      count = count + 1
    End If
  Next
  Cells(88, c).Value = count
  count = 0
Next
'-----

'----- count same cluster (assuptions: -1:1 | 0:2 | 1:3) -----
For c = 3 To 16
  For r = 2 To 81
    If ((Cells(r, 2).Value = -1 And Cells(r, c).Value = 1) Or _
      (Cells(r, 2).Value = 0 And Cells(r, c).Value = 2) Or _
      (Cells(r, 2).Value = 1 And Cells(r, c).Value = 3)) Then
      count = count + 1
    End If
  Next
  Cells(89, c).Value = count
  count = 0
Next
'-----

'----- percent same cluster (assuptions: -1:1 | 0:2 | 1:3) -----
For c = 3 To 16
  Cells(90, c).Value = Cells(89, c).Value / 80
Next
'-----

'----- count same cluster (assuptions: -1:1 | 0:3 | 1:2) -----
For c = 3 To 16
  For r = 2 To 81
    If ((Cells(r, 2).Value = -1 And Cells(r, c).Value = 1) Or _
      (Cells(r, 2).Value = 0 And Cells(r, c).Value = 3) Or _
      (Cells(r, 2).Value = 1 And Cells(r, c).Value = 2)) Then
      count = count + 1
    End If
  Next
  Cells(91, c).Value = count
  count = 0
Next
'-----

'----- percent same cluster (assuptions: -1:1 | 0:3 | 1:2) -----
For c = 3 To 16
  Cells(92, c).Value = Cells(91, c).Value / 80
Next
'-----

'----- count same cluster (assuptions: -1:2 | 0:1 | 1:3) -----
For c = 3 To 16
  For r = 2 To 81
    If ((Cells(r, 2).Value = -1 And Cells(r, c).Value = 2) Or _
      (Cells(r, 2).Value = 0 And Cells(r, c).Value = 1) Or _
      (Cells(r, 2).Value = 1 And Cells(r, c).Value = 3)) Then
      count = count + 1
    End If
  Next
  Cells(93, c).Value = count
  count = 0
Next
'-----

'----- percent same cluster (assuptions: -1:2 | 0:1 | 1:3) -----
For c = 3 To 16
  Cells(94, c).Value = Cells(93, c).Value / 80
Next
'-----

'----- count same cluster (assuptions: -1:2 | 0:3 | 1:1) -----
For c = 3 To 16
  For r = 2 To 81
    If ((Cells(r, 2).Value = -1 And Cells(r, c).Value = 2) Or _
      (Cells(r, 2).Value = 0 And Cells(r, c).Value = 3) Or _
      (Cells(r, 2).Value = 1 And Cells(r, c).Value = 1)) Then
      count = count + 1
    End If
  Next
  Cells(95, c).Value = count
  count = 0
Next
'-----

'----- percent same cluster (assuptions: -1:2 | 0:3 | 1:1) -----
For c = 3 To 16
  Cells(96, c).Value = Cells(95, c).Value / 80
Next
'-----

'----- count same cluster (assuptions: -1:3 | 0:1 | 1:2) -----
For c = 3 To 16
  For r = 2 To 81
    If ((Cells(r, 2).Value = -1 And Cells(r, c).Value = 3) Or _
      (Cells(r, 2).Value = 0 And Cells(r, c).Value = 1) Or _
      (Cells(r, 2).Value = 1 And Cells(r, c).Value = 2)) Then
      count = count + 1
    End If
  Next
  Cells(97, c).Value = count

```

```

    count = 0
Next
'-----

'----- percent same cluster (assuptions: -1:3 | 0:1 | 1:2) -----
For c = 3 To 16
    Cells(98, c).Value = Cells(97, c).Value / 80
Next
'-----

'----- count same cluster (assuptions: -1:3 | 0:2 | 1:1) -----
For c = 3 To 16
    For r = 2 To 81
        If ((Cells(r, 2).Value = -1 And Cells(r, c).Value = 3) Or _
            (Cells(r, 2).Value = 0 And Cells(r, c).Value = 2) Or _
            (Cells(r, 2).Value = 1 And Cells(r, c).Value = 1)) Then
            count = count + 1
        End If
    Next
Next

```

```

    Cells(99, c).Value = count
    count = 0
Next
'-----

'----- percent same cluster (assuptions: -1:3 | 0:2 | 1:1) -----
For c = 3 To 16
    Cells(100, c).Value = Cells(99, c).Value / 80
Next
'-----

'----- percent same cluster (all assuptions) -----
For c = 3 To 16
    Cells(101, c).Value = (Cells(89, c).Value + Cells(91, c).Value +
Cells(93, c).Value + Cells(95, c).Value + Cells(97, c).Value + Cells(99,
c).Value) / (6 * 80)
Next
'-----

```