**Week 2 Assignment**

 Bookmark this page

Install these R packages before you work on this assignment:

• NbClust

• mclust

It is wise to have [MarketSegmentation.nb.html](https://courses.edx.org/asset-v1:ColumbiaX+BAMM.104x+1T2020+type@asset+block/MarketSegmentation.nb.html) handy because you will be asked to replicate the class learning on Market Segmentation using a new dataset.

The dataset ( [SelfStatedData.csv](https://courses.edx.org/assets/courseware/v1/e0d2fdd35ea0e96b5b7517a168f4a1b9/asset-v1:ColumbiaX+BAMM.104x+1T2020+type@asset+block/SelfStatedData.csv) ) contains self-stated preferences for five attributes concerning tablets. These preferences were obtained from a survey on tablets administered to 50 individuals. The five attributes are:

• Screen: Screen size

• Cell: Cell Connectivity (can the tablet connect to the cellular network?)

• OS: Operating System (does the individual care about the tablet’s operating system?)

• Battery: Battery Life (does the individual care about the battery life?)

• Price: Does the consumer care about the price of the tablet?

The preferences were measured on 7-point scale (1: attribute not important, 7: attribute very important to the individual).

**For the analysis of this dataset, please follow the R codes available in**[**MarketSegmentation.nb.htm**](https://courses.edx.org/asset-v1:ColumbiaX+BAMM.104x+1T2020+type@asset+block/Market_Segmentation.nb.html)[**l**](https://studio.edx.org/container/courses.edx.org/static/segmentation.pptx)**.**

Before you run any analysis, please set your seed to 1990 (set.seed(1990)), for each block.

In this assignment, you will be required to perform some analysis covered in class in order to respond to the questions based on the output. In order to help you understand the task, please read the following sample question.

Sample Question: Scale the data using the R function scale() and compute the distance matrix using the  R function dist() with the parameter: method = “euclidean”. What is the distance between individuals 1 and 3?

* + 1. 2.746685
    2. 3.804537
    3. 1.713049
    4. 2.774723

Solution:

We observe that in[MarketSegmentation.nb.html](https://courses.edx.org/asset-v1:ColumbiaX+BAMM.104x+1T2020+type@asset+block/MarketSegmentation.nb.html), we obtain the required information with the following code:

std\_seg\_data <- scale(seg\_data[,c("Trendy", "Styling", "Reliability", "Sportiness", "Performance", "Comfort")])

dist <- dist(std\_seg\_data, method = "euclidean")   
as.matrix(dist)[1:5,1:5]

Thus, in order to answer the sample question, we do the appropriate modifications and use the following code:

set.seed(1990)  
std\_SelfStatedData <- scale(SelfStatedData)

dist <- dist(std\_SelfStatedData, method = "euclidean")   
as.matrix(dist)[1:5,1:5]

Note that I started with set.seed(1990).

The output is as follows:

1 2 3 4 5

1 0.000000 2.746685 3.804537 1.713049 2.774723

2 2.746685 0.000000 3.228845 3.089934 1.081441

3 3.804537 3.228845 0.000000 3.601860 2.629999  
4 1.713049 3.089934 3.601860 0.000000 3.064781

5 2.774723 1.081441 2.629999 3.064781 0.000000

Thus, the correct answer is 3.804537 (answer 2).

Question 1

1 point possible (graded)

Perform a hierarchical clustering analysis using the R function hclust() with the parameter: method= “ward.D2”. How many clusters would you retain if you cut the dendogram at the distance 6?

2

3

4

5

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 2

1 point possible (graded)

Consider the four-cluster solution. Which cluster cares the most about the attribute Cell Connectivity?

Cluster 1

Cluster 2

Cluster 3

Cluster 4

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 3

1 point possible (graded)

Use the R function NbClust() with the following parameters: min.nc=2, max.nc=15, index="all", method="ward.D2". According to the majority rule, what is the best number of clusters for this dataset?

Note that, in contrast with the notebook, you do not need to remove variables in this case.

15

4

7

2

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 4

1 point possible (graded)

Consider the two-cluster solution. Use the R aggregate() function to compute the attribute preferences by cluster. Which are the two most important attributes for cluster 1?

Price and Battery

Price and Screen

OS and Price

OS and Battery

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 5

1 point possible (graded)

For the following questions, please begin your code by set.seed(1990). Use the R function kmeans() to run a three-cluster analysis. Use the following parameters: iter.max=100, nstart=100. What are the sizes of the resulting three clusters?

23, 22, 5

9, 4, 37

17, 17, 16

12, 25, 13

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 6

0.0/1.0 point (graded)

What is the most important attribute for cluster with the second largest size?

OS

Screen

Price

Cell Connectivity

unanswered

Submit

You have used 0 of 2 attemptsSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 7

1 point possible (graded)

Use the R function NbClust() with the following parameters: min.nc=2, max.nc=15, index="all", method="kmeans". According to the majority rule, what is the best number of clusters for this dataset?

2

3

4

5

unanswered

Submit

You have used 0 of 2 attemptsSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 8

1 point possible (graded)

For the following questions, please begin your code by set.seed(1990). Apply the R function Mclust() on the data. What is the best model suggested by this method?

Note that we use the full description of the model. The answer returned by R may differ slightly. In this case, the package documentation will help you match both descriptions.

Spherical, varying volume, equal shape model with 2 components

Diagonal, equal volume and shape model with 3 components

Ellipsoidal, equal volume, shape, and orientation with 2 components

Spherical Multivariate Normal with 1 component

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 9

1 point possible (graded)

Let’s now focus on the diagonal, equal volume and shape class of models (EEI). Use the R function Mclust() with the following parameter modelNames ="EEI". How many components (i.e., clusters) does this analysis suggest?

1

2

3

4

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer

Question 10

1 point possible (graded)

Consider the analysis in question 9 above. Extract the resulting cluster membership from the previous output using $classification. To interpret the clusters, use the R aggregate() function to compute the attribute preferences by cluster. Which are the two most important attributes for cluster 1?

Price and Battery

Price and Screen

OS and Price

OS and Battery

unanswered

Submit

You have used 0 of 1 attemptSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

SaveSave Your Answer ResetReset Your Answer