**CPIS 342 Data Warehousing and Mining**

**Quiz#1 Fall 2019**

**Student Id:**

**Name:**

**Section:**

**Question #1:**

The calories for 10 brands of veggie dogs are given below. (Notice these data have been ordered from smallest to largest.)

40 45 **55** 57 60 70 81 **85** 110 155

1. Draw a basic boxplot for the calories per veggie dog. (1.5)

Median= (60 +70)/2 = 65

Q1= (index=.25 \* 10 = 3) =55

Q3= (index=.75 \* 10 = 8) = 85

IQR= Q3 – Q1= 85– 55 =30

IQR \* 1.5= 50 \* 1.5 = 45

Q1- 45=55 – 45 = 10

Q3+45 = 130

Five-number summary: 40, 55, 65, 85, 110

155

105

95

45

75

65

85

55

Q1

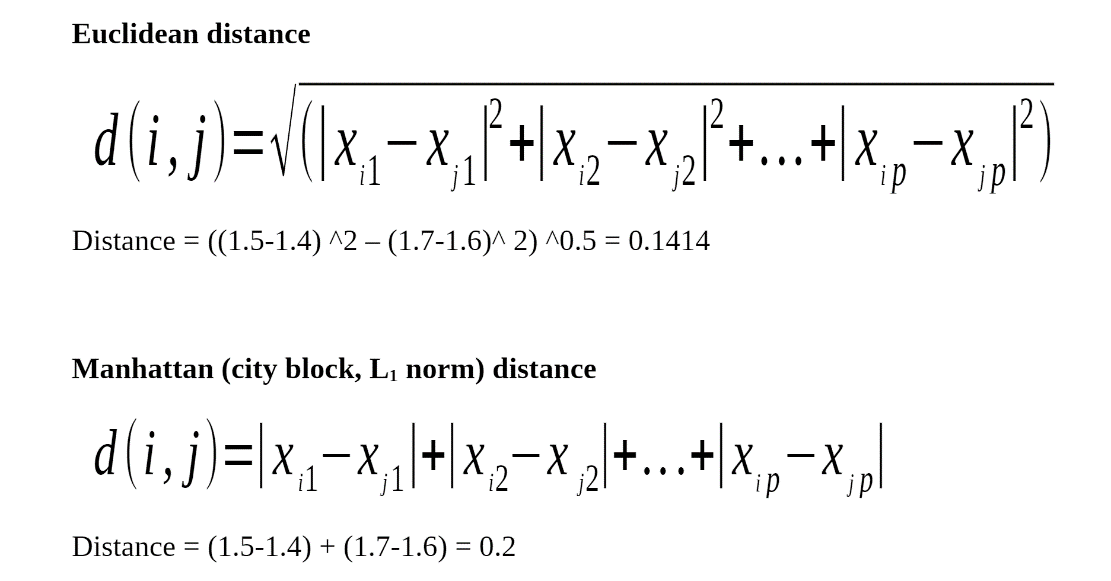
Q3

**Question #2:**

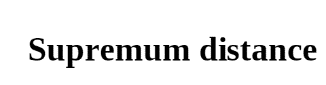
Suppose we have the following 2-D data set: (1.5)

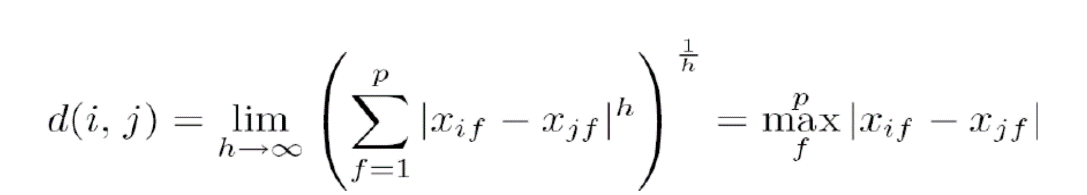
|  |  |  |
| --- | --- | --- |
|  | A1 | A2 |
| X1 | 1.5 | 1.7 |
| X2 | 1.4 | 1.6 |

Compute the (a) Euclidean distance, (b) Manhattan distance, (c) supremum distance, and (d) cosine similarity between the two objects. (e) Give an interpretation to each result.



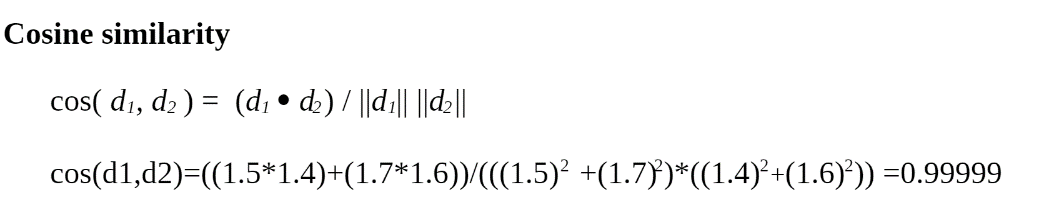
**near to be dissimilar or near to be similar**





= max (|1.5 – 1.4|, |1.7 – 1.6|) = .1

**near to be dissimilar or near to be similar**



**Near to be similar or near to be dissimilar**