Exercises: Similarity and summary statistics

August 30, 2016

Task 1: Similarity

Compute the similarity between object 1 and objects 2–4 using the Jaccard coefficient and the cosine similarity.

(Note: for Jaccard similarity, all non-zero values should be treated as 1s.)

Object	attr 1	attr 2	attr 3	attr 4	attr 5	attr 6
1	1	2	0	0	0	0
2	1	0	1	1	1	0
3	3	0	0	0	0	3
4	2	4	0	0	0	0

Table 1: Sample data set.

Solution

Objects	Jaccard	Cosine
1 vs. 2		
1 vs. 3		
1 vs. 4		

Table 2: Similarity between object 1 and objects 2–4.

Task 2: Summary statistics

Compute summary statistics for the following values: 17, 5, 3, 9, 49, 53, 11.

Round the results to two decimal places.

Solution

70th percentile	
Range	
Median	
Mean	
Variance	
Absolute Average Deviation	

Table 3: Summary statistics.