Question 1 of 12: Conduct k-means clustering with K=2, using k-Means. (Not W-SimpleKMeans). Use factory settings for all other parameters. Drag arrows between both of the output ports of k-Means to the results at the right. Look at the Centroid Table within the Cluster Model tab. Which attributes have the biggest difference between cluster_0 and cluster_1?

- (•) A) A and F
- ( ) B) C and E
- ( ) C) B and D
- ( ) D) C and F

Question 2 of 12: Now look at the Plot view in the ExampleSet tab. Set the X axis to be the first answer from Question 1 (A), and set the Y axis to be the second answer from Question 1 (F). Now set the color column to cluster. There are seven major groupings ("lumps") in this data. How many of them are red?

1

Question 3 of 12: What did k-Means do here?

- ( ) A) It split the lumps in the data approximately evenly in clusters
- ( ) B) It did a median split on two key variables
- (•) C) It found the least central lump in the data and made it a cluster
- ( ) D) It found the most central lump in the data and made it a cluster

Question 4 of 12: Now re-run k-Means with k=7. Did k-Means find the 7 lumps in the data that you saw earlier?

- ( ) A) Yes
- (•) B) No

Question 5 of 12: Plot each of the other variables against each other (not including the variables in question 1). Does there appear to be meaningful structure in any of these variables?

- ( ) A) Yes
- (•) B) No

Question 6 of 12: Filter out all of the variables except the ones in question 1, and re-run k-Means using just these two variables, with k=7. Are all seven of the seven data lumps now more or less incorporated into seven reasonable clusters?

- ( ) A) Yes
- (•) B) No

Question 7 of 12: What happened?

A) It looks the same as when k= 2

B) One region of space without a lump got a cluster, and two lumps got a single cluster

C) Several clusters were devoted to regions of space without a lump

D) Two regions of space without lumps got clusters, and three lumps got a single cluster

Question 8 of 12: For fun, you might want to try playing with different values of k, and the other parameters within k-Means. When you're ready to move on, type the number 0.

0

Question 9 of 12: Try running Expectation Maximization Clustering with k=7. Look at the cluster probabilities for each cluster in Plot View. Which clusters are focused on a single data lump? (As opposed to including lots of outliers?)

A) All of them but a single outlier cluster

B) All but clusters 2, 4 and 6

C) All but cluster 1 and cluster 5

D) All but cluster 3 and cluster 4

Question 10 of 12: For fun, you might notice that outliers close to the top-right cluster still got placed into an outlier cluster (cluster_5_probability), but with lower certainty. This is the power of having centers and radii. When you're ready to move on, type the number 0.

0

Question 11 of 12: When you're done looking, try running Agglomerative Clustering. Look at the Dendrogram. Nifty, huh?

A) Yes, that is nifty.

B) I dispute the value of this question as assessment.

Question 12 of 12: OK, fine. Squint really hard and look at the top-right of the dendogram. You'll see at the very top fork, that a branch goes down the right side. How many nodes are in this branch? (e.g. how many data points end up in this branch). Note that an immediate

2