2. (2.0 points)

An actuary wants to cluster five Workers' Compensation classes based on excess ratios at two limits: 500,000 and 1,000,000. The actuary decides to use a weighted k-means algorithm with two clusters. Given the following:

Class	On-Leveled Earned Premium ($ Thousands)	Normalized Excess Ratio at 500,000 Limit	Normalized Excess Ratio at 1,000,000 Limit	Initial Cluster
1	6,500	0.240	0.080	A
2	5,000	0.350	0.200	A
3	4,000	0.210	0.080	A
4	3,000	0.110	0.030	A
5	5,000	0.180	0.070	B

- Distance will be measured using the L^2 (Euclidean) norm.
- At the start of the algorithm, the actuary randomly assigns each class to a cluster.

a. (1.75 points)

Determine the cluster for each class after the first iteration of the weighted k-means algorithm.

b. (0.25 point)

Briefly describe one advantage of using the L^1 measure rather than L^2 when computing clusters.

CONTINUED ON NEXT PAGE

QUESTION 2

TOTAL POINT VALUE: 2 | **LEARNING OBJECTIVE(S): A1**

SAMPLE ANSWERS

Part a: 1.75 points

Determine centroid of each initial cluster

$R_A(500K) = (6500*0.24+5000*0.35+4000*0.21+3000*0.11)/(6500+5000+4000+3000) = 0.242$

$R_B(1M) = (6500*0.08+5000*0.2+4000*0.08+3000*0.03)/(6500+5000+4000+3000) = 0.104$

$R_B(500K) = 0.18$

$R_B(1M) = 0.07$

Now determine Euclidean distance between each class and centroids of A & B, and assign class to cluster with smaller distance.

Class	Distance to R_A	Distance to R_B	New Cluster
1	0.024 = [(0.24-0.242)^2 + (0.08-0.104)^2] ^ (1/2)	0.061 = [(0.24-0.18)^2 + (0.08-0.07)^2] ^ (1/2)	A
2	0.144	0.214	A
3	0.040	0.032	B
4	0.151	0.081	B
5	0.07	0	B

Part b: 0.25 point

- Many small errors would have the same effect as one large error which results in outliers having less of an impact on the result.
- L^1 minimizes the relative error $PLR*|R_C(L)-R_{HG}(L)|$; L2 does not necessarily.
- L^1 minimizes the relative error when calculating the premium.
- L^1 minimizes the relative error when calculating the excess ratio.
- The unit of L^1 is the same as expected loss costs, i.e. in dollars, whereas L^2 has unit in $dollar^2$

EXAMINER'S REPORT

Candidates were expected to know how to perform the first iteration of a clustering analysis – determine centroid coordinates, calculate L^2 distance from each class to each centroid and assign each class to the closest centroid. Many candidates struggled with the calculation of the L^2 distance measure.

Part a

Candidates were expected to calculate the centroid coordinates (500K, 1M) of each cluster as they were assigned in the question. With those centroids, candidates were then expected to calculate the Euclidean distance (L^2) of each class from each of the two clusters. Finally, candidates needed to assign each class to a cluster based on the smaller of the two distances.

Common mistakes included:

- Taking a straight average rather than a weighted average of the centroid A excess ratios
- Taking an average of the 500K and 1M excess ratios for each class individually and using that as the centroid coordinate
- Only calculating centroids and distances for one limit
- Using an incorrect distance formula (such as L^1)
- Calculating separate distances and assigning separate clusters for each limit
- Not showing complete supporting work on how clusters were assigned

Part b

Candidates were expected to state that:

- The L^2 measure gives undue weight to outliers, or
- The L^1 measure minimizes the relative error in excess premium (or excess ratios).

Common mistakes included:

- Stating that L^1 is easier to calculate/requires less computation
- Stating that L^1 is easier to explain
- Stating that L^1 considers negative/positive direction of the distance