Second Year Project II

By Shuvani Choudhury and Anh Tran

Question 1: Pre-processing

Methods Used:

- One-Way Tabulation
- N-Way Tabulation
- Excel
- Create new sheet by Specifying the Frequency/Range...

Results Found:

- Clusters (1,1) = Lane 1
- Clusters (1,6) = Lane 2

Question 2: Testing

Methods/Tests Used:

- Estimate Equation (Least Squares)
- Wald Test
- Jarque Bera Test

- LM Test
- Q-Test
- Likelihood Ratio Test
- Akaike Information Criterion (AIC)

Question 2: Testing (Contd.)

- White Test
- Estimate Equation (Robust Least Squares)

- Forecast

Results:

- U(t) = 46.3948 + 0.3162*U(t-1) - 0.1287*U(t-2) + 0.1576*U(t-3) - 0.0046*U(t-4) - 0.1010*U(t-5) + 0.1251*U(t-6) + 0.1168*U(t-7) - 0.1118*U(t-8) + 0.0349*U(t-9) + 0.0078*U(t-10) - 0.0100*U(t-11) + 0.0798*U(t-12) - 0.1241*U(t-13) + 0.1038*U(t-14) - 0.0467*U(t-15) + 1.4942*(@WEEKDAY=2) - 21.4596*(@WEEKDAY=3) + 1.0064*(@WEEKDAY=4) - 20.0365*(@WEEKDAY=5)

Question 2: Testing (Contd.)

- U(t) = 57.0012 - 0.0218*U(t-1) - 0.0357U(t- 2) + 0.0010*U(t-3) + 0.0438*U(t-4) + 0.0439U(t-5) + 2.2656*(@WEEKDAY=2) - 1.3117*(@WEEKDAY=3) + 0.9816*(@WEEKDAY=4) + 4.9891*(@WEEKDAY=5)
- *V(t) = 57.2526 + 58.3742*(@WEEKDAY=2) - 4.1702*(@WEEKDAY=3) + 73.2439*(@WEEKDAY=4) - 20.1628*(@WEEKDAY=5)
- *V(t) = 80.4066 + 11.2566*(@WEEKDAY=2) + 28.2842*(@WEEKDAY=3) + 2.3225*(@WEEKDAY=4) + 119.0351*(@WEEKDAY=5)
- *W(t) = 3663.4284 - 619.9506*(@WEEKDAY=2) + 753.4660*(@WEEKDAY=3) - 18.6422*(@WEEKDAY=4) + 877.6078*(@WEEKDAY=5)
- W(t) = 3908.2041 + 0.0918*W(t-1) - 0.1536*W(t-2) + 0.2163*W(t-3) - 0.0568*W(t-4) + 0.2710*W(t-5)

Question 3: Forecasts

Date	Number of Shipments (1)	Number of Shipments (2)	Gross Volume (1)	Gross Volume (2)	Gross Weight (1)	Gross Weight (2)
10/16/2017	76.1074	59.6025	57.2526	80.4066	3663.428	5695.253
10/17/2017	76.3347	66.9714	115.6268	91.6632	3043.478	5400.956
10/18/2017	51.3366	60.3709	53.0824	108.6907	4416.894	5898.426
10/19/2017	72.4791	61.3272	130.4965	82.7290	3644.786	5678.467
10/20/2017	59.7489	67.2448	37.0898	199.4417	4541.036	6020.416

Question 4-6:

PREPROCESSING

- Aggregate the shipment data (weight, volume, number of shipments) for a given origin-destination pair.
- Ignore pickup dates.
- New dataset: **4011** shipments
- Choose relevant variables: Origin Country, Origin City, OriginCluster, OriginClusterLat, OriginClusterLong, OriginLat, OriginLong, Dest Country, Dest City, DestinationCluster, DestinationClusterLat, DestinationClusterLong, DestLat DestLong, TR Gross Weight (KG), Nb of Ship Units, TR Gross Volume (M3)

Question 4: Performance of initial clustering

CLASSES

- Shipment
- Cluster
- Location
- ShipmentList
- Clustering

Other classes

- HaversineDistance

Question 4: Performance of initial clustering (Contd.)

Origin	Destination
Total distance from locations to clusters: 5428.961960619172	Total distance from locations to clusters: 4874.904502514321
Average cluster wrt total distance: Cluster902 • Value of average cluster: 8.313240107222832 • Average value: 8.390976755207383	Average cluster wrt total distance: Cluster684 • Value of average cluster: 6.123313380001216 • Average value: 6.298326230638661

Question 4: Performance of initial clustering (Contd.)

Average cluster wrt total weight: Cluster76 • Value of average cluster: 76387.0 • Average value: 72602.05645131374 Average cluster wrt total volume: Cluster550 • Value of average cluster: 1706.556 • Average value: 1638.9395069551786 Average cluster wrt total quantity: Cluster996 • Value of average cluster: 364 • Average value: 364.548686244204	Average cluster wrt total weight: Cluster547 • Value of average cluster: 61364.0 • Average value: 60689.31592248065 Average cluster wrt total volume: Cluster10 • Value of average cluster: 1464.9469999999994 • Average value: 1370.0179082687332 Average cluster wrt total quantity: Cluster547 • Value of average cluster: 310 • Average value: 304.7325581395349

Question 4: Performance of initial clustering (Contd.)

Top 3 total origin distances: 1. Cluster2 (3342.5071723061615) 2. Cluster8 (503.9226266971474) 3. Cluster3 (278.97624112065745)	Top 3 total destination distances: 1. Cluster2 (2328.4887185208154) 2. Cluster8 (667.1922833921053) 3. Cluster441 (199.28990420469611)
Top 3 total origin weights: 1. Cluster2 (6262006.481000002) 2. Cluster902 (5195074.104) 3. Cluster15 (3061947.3500000006)	Top 3 total destination weights: 1. Cluster2 (7966273.728) 2. Cluster906 (5519990.334) 3. Cluster546 (3219875.7010000004)
Top 3 total origin volumes: 1. Cluster443 (122109.77900000085) 2. Cluster19 (106330.67500000008) 3. Cluster15 (104512.2429999997)	Top 3 total destination volumes: 1. Cluster2 (244201.52699999948) 2. Cluster443 (149982.4760000009) 3. Cluster19 (106808.47200000002)
Top 3 total origin quantities: 1. Cluster2 (53327) 2. Cluster1 (24848) 3. Cluster15 (11964)	Top 3 total destination quantities: 1. Cluster2 (56656) 2. Cluster1 (39947) 3. Cluster546 (22109)

Question 4: Performance of initial clustering (Contd.)

Bottom 3 total origin distances: 1. Cluster827 (0.0) 2. Cluster637 (0.0) 3. Cluster596 (0.0)	Bottom 3 total destination distances: 1. Cluster19 (0.0) 2. Cluster15 (0.0) 3. Cluster17 (0.0)
Bottom 3 total origin weights: 1. Cluster606 (0.22) 2. Cluster304 (0.5) 3. Cluster526 (0.5)	Bottom 3 total destination weights: 1. Cluster719 (0.25) 2. Cluster586 (0.5) 3. Cluster1024 (0.5)
Bottom 3 total origin volumes: 1. Cluster886 (0.0) 2. Cluster526 (0.0) 3. Cluster195 (0.001)	Bottom 3 total destination volumes: 1. Cluster586 (0.0) 2. Cluster856 (0.001) 3. Cluster771 (0.001)
Bottom 3 total origin quantities: 1. Cluster719 (1) 2. Cluster601 (1) 3. Cluster607 (1)	Bottom 3 total destination quantities: 1. Cluster748 (1) 2. Cluster858 (1) 3. Cluster812 (1)

Question 5: k-Means Clustering

HOW IT WORKS

1. Randomly assign shipments to k clusters (We did this sequentially).
2. Calculate location of cluster centroids
3. Assign shipments to nearest centroids
4. Repeat step 2-3 until the shipment assignment does not change

RESULTS FOR SEVERAL VALUES OF k

General not as good as those of the initial clustering.

Possible reason: Initial clustering is more holistic

Question 5: k-Means Clustering (Contd.)

K = 5:

- Total distance from locations to clusters: 720861.7421654129
- Total distance from locations to clusters: 911800.9561343524

Metrics are higher for this algorithm.

However, k-Means clustering gives a good general idea about the shipments.

Question 6: Alternatives

SEGMENTATION

- Divide up the dataset by countries
- Group 1: shipments whose origin & destinations are the same (country-specific)
- Group 2: cross-border shipments
- Perform k-Means Clustering on each data subset

Conclusion

- Timeseries Analysis: statistical tests, forecasts
- Cluster Analysis: metric evaluation, clustering algorithm
- Improvements