Cluster Analysis Workshop

Dr Robert Mayer
Centre for Air Transport Management
Cranfield University

In this workshop you are provided with information from an air passenger survey of over 500 passengers at a UK airport. Based on the data collected and a cluster analysis that was conducted, you are require to discuss the clusters and the clustering process. In particular, you should work on the following aspects:

- Discuss the clustering approach (How did the researchers conduct the cluster analysis?).
- Comment on the number of clusters.
- Describe the characteristics of each cluster.
- Develop a “label” for each cluster (i.e. a cluster name that attracts attention).
- What are differences between the clusters?
- Discuss the choice of variables used in the cluster analysis? How useful are they? What variables would you have chosen?
- How can airlines benefit from segmenting passengers according to environmental attitudes? How can airlines use this information?

The Survey

In the survey, passengers were asked to rate eight statements (seven attitudinal and one behavioural) on a 5-point Likert scale, where 1 = "*Strongly agree*", 2 = "*Agree*", 3 = "*Neither agree nor disagree*", 4 = "*Disagree*" and 5 = "*Strongly disagree*":

- Air travel is essential to the UK economy and the country's continuing prosperity.
- Air travel is a significant contributor to climate change.
- Some airlines do more for the environment than others.
- Passengers should pay more to fly because of the negative environmental aspects of aviation.
- Airlines that have higher CO_2 emissions should pay higher taxes and charges to operate at UK airports.
- The effectiveness of reducing airlines' environmental impact: Airlines using newer aircraft.
- The effectiveness of reducing airlines' environmental impact: Airlines having a positive attitude towards the environment.
- I always look for the cheapest flights.

Besides this information, the 543 respondents provided also demographic information as well as information on attitudes towards green airline marketing and their travel behaviour.

Using a two-stage clustering process, where hierarchical clustering was applied to identify the appropriate number of clusters which then fed into *k*-means clustering. Ward's method was chosen as the applied agglomerative algorithm, using squared Euclidean distance as distance measurements. From the initial hierarchical clustering, preliminary results were generated, which helped to select the number of clusters, obtain centroids for each cluster and identify outliers.

The Cluster Analysis

Based on the two-stage cluster analysis, following cluster centres have been identified:

Variable	Cluster				
	1	2	3	4	5
1 Air travel is essential to the UK economy and the country's continuing prosperity	1.70	1.71	2.06	1.45	1.71
2 Air travel is a significant contributor to climate change	1.81	2.64	2.51	3.18	1.92
3 Passengers should pay more to fly because of the negative environmental aspects of aviation	1.88	3.81	2.98	4.34	3.59
4 Some airlines do more for the environment than others	2.38	3.08	2.46	2.46	2.19
5 Airlines that have higher CO2 emissions should pay higher taxes and charges to operate at UK airports	1.98	3.28	2.32	3.08	2.04
6 Airlines using newer aircraft	1.75	2.49	2.05	1.24	1.77
7 Airlines having a positive attitude towards the environment	1.91	2.75	2.08	1.79	1.62
8 I always look for the cheapest flights	1.66	1.45	3.33	1.50	1.41

Table 1: Final Cluster Centres (k-means Clustering)

The results from the data analysis also show the distances between the cluster centres, i.e. how similar, dissimilar the different clusters are.

Cluster	1	2	3	4	5
1		2.807	2.213	3.084	1.764
2	2.807		2.514	1.880	2.165
3	2.213	2.514		2.711	2.232
4	3.084	1.880	2.711		1.924
5	1.764	2.165	2.232	1.924	

Table 2: Distances between Clusters

External Validity and Socio-Demographic Data

Once the cluster analysis was performed, other variables (i.e. those that were not used in the analysis) where used to identify any differences between the clusters and validate the results from the cluster analysis.

	1	2	3	4	5	Total
Gender						
Female	51.5%	50.5%	64.5%	34.2%	56.1%	50.5%
Male	48.5%	49.5%	35.5%	65.8%	43.9%	49.5%
Age band						
18-34	43.4%	39.8%	21.0%	38.6%	55.4%	42.6%
35-54	33.3%	40.7%	27.4%	31.6%	22.9%	30.7%
> 54	23.2%	19.4%	51.6%	29.8%	21.7%	26.7%
Occupation						
Self-employed	15.5%	15.1%	4.9%	11.6%	9.3%	11.6%
Employed full time	43.3%	37.7%	42.6%	49.1%	48.3%	44.8%
Employed part time	11.3%	6.6%	14.8%	6.3%	9.3%	9.1%
Looking after home or family	2.1%	4.7%	0.0%	2.7%	0.7%	2.1%
Permanently retired from work	13.4%	17.0%	32.8%	17.0%	14.6%	17.5%
Unemployed and seeking work	1.0%	3.8%	1.6%	1.8%	1.3%	1.9%
In education	13.4%	12.3%	3.3%	11.6%	15.2%	12.1%
Government work or training scheme	0.0%	0.9%	0.0%	0.0%	0.7%	0.4%
Unable to work	0.0%	1.9%	0.0%	0.0%	0.7%	0.6%
Personal Income (per annum)						
< £ 20,001	29.0%	53.7%	44.7%	27.4%	44.7%	39.7%
£ 20,001-40,000	43.5%	26.9%	31.6%	42.5%	35.0%	36.3%
£ 40,001-60,000	14.5%	14.9%	7.9%	16.4%	15.5%	14.6%
> £60,000	13.0%	4.5%	15.8%	13.7%	4.9%	9.4%
Personal income not provided	30.3%	38.5%	39.7%	36.5%	34.4%	35.5%

Table 3: Socio-demographic Characteristics by Cluster

Travel Behaviour by Cluster

Passengers also provided information on their travel behavior. The following table shows the results per cluster regarding passengers' airport access mode.

	1	2	3	4	5	Total
Own car	27.3%	28.3%	28.6%	34.2%	23.6%	28.0%
Public transport (Train, Bus)	19.2%	22.6%	14.3%	17.5%	29.9%	22.1%
Lift from family or friend	24.2%	26.4%	27.0%	24.6%	25.5%	25.4%
Taxi	23.2%	17.0%	27.0%	14.0%	15.3%	18.2%
Other	6.1%	5.7%	3.2%	9.6%	5.7%	6.3%

Table 4: Airport Access by Cluster

Table 5 shows the number of flights undertaken in the twelve months leading up the survey by segment, as well as changes in the past two years and planned future changes for the following two years.

	1	2	3	4	5	Total
Number of return flights over the last 12 months for leisure purposes						
2 or fewer	44.2%	38.3%	41.4%	35.5%	35.0%	38.1%
3-4	30.5%	34.6%	44.8%	29.1%	33.1%	33.4%
5 or more	25.3%	27.1%	13.8%	35.5%	31.8%	28.5%
Changes in the number of flights for leisure/personal reasons over the last 2 years						
Increase	27.8%	33.0%	28.6%	42.0%	50.6%	38.5%
About the same	61.9%	55.7%	50.8%	50.0%	40.3%	50.6%
Decrease	10.3%	11.3%	20.6%	8.0%	9.1%	10.9%
Plans to change the number of flights for leisure/personal reasons over the next 2 years						
Increase	22.6%	21.4%	23.3%	26.2%	26.5%	24.3%
About the same	66.7%	76.7%	65.0%	68.2%	63.6%	67.9%
Decrease	10.8%	1.9%	11.7%	5.6%	9.9%	7.8%

Table 5: Air Travel Behaviour by Cluster

Use of Carbon Off-setting by Cluster

In the questionnaire, passengers were also asked about Carbon Off-Setting Schemes. Table 6 shows the responses by cluster.

	1	2	3	4	5	Total
Have you heard of “Carbon Off-setting” Schemes?						
Yes	56.6%	41.3%	48.4%	57.9%	46.5%	49.9%
No	43.4%	58.7%	51.6%	42.1%	53.5%	50.1%
Have you used “Carbon Off-Setting” for today’s flight?						
Yes	7.0%	4.4%	12.9%	1.5%	4.2%	5.2%
No	80.7%	73.3%	58.1%	83.6%	83.1%	78.2%
I don't know	12.3%	22.2%	29.0%	14.9%	12.7%	16.6%
How likely is it that you will use “Carbon Off-Setting” for future leisure/personal trips						
Likely	33.3%	6.7%	29.0%	6.1%	26.8%	20.0%
I don't know yet	40.4%	46.7%	54.8%	33.3%	49.3%	43.7%
Unlikely	26.3%	46.7%	16.1%	60.6%	23.9%	36.3%

Table 6: Approach to Carbon Off-Setting by Cluster