Market Segmentation Based on the Dominant Movement Patterns of Tourists

Xia, Jianhong (Cecilia) and Evans, Fiona and Spilsbury, Katrina and Ciesielski, Vic and Arrowsmith, Colin and Wright, Graeme. 2010. Market segments based on the dominant movement patterns of tourists. Tourism Management 31 (4): pp. 464-469.

Abstract: This paper presents an innovative method for tourist market segmentation based on dominant movement patterns of tourists; that is, the travel sequences or patterns used by tourists most frequently. There are three steps to achieve this goal. In the first step, general log-linear models are adopted to identify the dominant movement patterns, while the second step is to discover the characteristics of the groups of tourists who travelled with these patterns. The Expectation-Maximisation algorithm was used to partition tourist segments in terms of socio-demographic and travel behavioural variables. The third step is to select target markets based upon the earlier analysis. These methods are applied to a sample of tourists, over the period of a week, on Phillip Island, Victoria, Australia. A significant outcome of this research is that it will assist tourism organisations to identify tourism market segments and develop better tour packages and more efficient marketing strategies aligned to the characteristics of the tourists.

Keyword: Tourist, Market segmentation, Expectation-Maximisation algorithm, Log-linear models, Movement pattern

1 Introduction

The movement of tourists is a complex process which can be modelled at a micro level as continuous processes with high resolution, such as in centimetres or at a macro level as discrete processes with low resolution, such as kilometres from one area to another. Tourist movement patterns are the theme of recurring or repeat movement sequences. This paper focuses on movement patterns at the macro level. Movement patterns represent the sequence of movements by tourists from one attraction site to another. The dominant movement patterns are the sequences or patterns that are used by tourists most frequently. These tourist movement patterns are vital to park managers or tour operators to understand the location of popular sites and the timing of visits. More importantly, an understanding of movement patterns can indicate how tourists combine attractions together and arrange their schedules.

Traditionally, tourism market segmentation was conducted to identify groups of tourists from an origin perspective, for example, analysis of the origin of tourists from China or Australia. Alternatively, analysis could be performed from a destination perspective, where the tourism market is segmentation-based on *a single* destination such as Melbourne. However, tourists usually visit several attractions during a trip. To understand the spatial combination of attractions and to clarify the characteristics of tourists who travel to these attractions will assist tourist organisations to design more appropriate and profitable tour packages.

The aim of this study was to develop a methodology to identify the characteristics of tourists who travel with dominant movement patterns. The first step involved identification of the dominant movement patterns based on the analysis of categorical data using general log-linear models. In the second step, the Expectation-Maximisation (EM) algorithm used in a mixture model framework was adopted to identify the characteristics of tourists who travel with dominant movement patterns. Tourists were divided into different segments based on similar socio-demographic characteristics and travel modes such as type of travel group, modes of transport or visit frequency. For example, the tourists in segment 1 who travel with pattern *A* may mostly be females travelling with their family by car. Finally, appropriate target markets can be determined from the socio-demographic data. Sections 2 and 3 of this paper address the first two steps, followed by a case study of tourists visiting Phillip Island Nature Park in Victoria, Australia.

2 Step one: identification of dominant movement patterns of tourists using general log-linear models

This section discusses the methodology used to identify the dominant movement patterns of tourists. Dominant movement patterns are defined as the combination of attractions used most frequently by tourists. The reasons for associations between attractions are diverse. Tourists could combine several attractions together as one trip because those attractions may be in close proximity, pairs of attractions might have been packaged by operators, the combination of these attractions fit to tourist desire or interests, or the park manager has developed promotion policies, for example, discounted entrance fees for combined attractions. For economic reasons, tourists might visit several attractions because they may want to save money or time to optimise their trips.

Generally, more than two attractions are visited by tourists during a trip, which means more than two categorical movement variables are concerned in tourist movement modelling. These multi-way contingency tables of frequency counts can be analysed using symmetrical general log-linear models under the assumption the counts follow a Poisson distribution (Goodman 1984; Agresti 1990). In order to fit tourist movement data to the general log-linear model, patterns of movement can be coded by a series of destination variables such that each represents a stage in the pattern of movement. For example, if a set of three attractions A, B and C were available for tourists to visit in a day, then for each tourist the variable Destination 1 includes the attraction first visited and variable Destination 2 would include the next attraction visited out of the two remaining attractions, and Destination 3 would include the last attraction visited. A log-linear model calculates the expected number of movement pattern counts for each destination combination as if there was no difference between the movement patterns. Chi-square goodness-of-fit tests are then used to compare these expected values to the observed counts. An association between two (or more) attractions can be considered dominant (significant) if the chi-square test *p*-value is <0.05. It may be observed, for example, that tourists who visit attraction A almost always immediately follow this with a visit to attraction C leading to a dominant movement pattern of AC in contrast to the less common movement pattern of AB.

3 Step two: tourist market segmentation methods

In this section, we discuss methods that can be used to generate tourism market segments for each of the significant movement patterns found in step 1. For instance, the significant movement patterns might be (OPERA-HOUSE, HARBOUR-BRIDGE) and (BONDI-BEACH, OPERA-HOUSE). Segmentation can be performed on tourist data, and the proportions of market segments matching the patterns (OPERA-HOUSE, HARBOUR-BRIDGE) and (BONDI-BEACH, OPERA-HOUSE) can be calculated and used to gain insight into the types of tourists that follow those movement patterns.

3.1 Determination of segmentation variables

This section describes a method to determine the tourist market segmentation variables. Generally tourist segments are divided in terms of geographic, socio-demographic, psychographic and travel behavioural variables (Wedel and Kamakura 2000). However, the determination of segmentation variables is dependent upon the objectives of tourism market segmentation. For example, Bigné and Andreu (2004) identified tourist emotion-based segments and analysed which segment was most satisfactory for leisure and tourism services based on socio-demographic variables, multiple-item scales of emotions, satisfaction and behavioural intentions. The objectives of market segmentation are summarised by Myers (1996) as follows:

- Identifying and characterising groups of tourists
- Focusing advertising efforts for greater impact
- Identifying likely targets for new tourist products
- Improving existing product/service design
- Looking for new product service opportunities
- Assessing the impact of a competitor’s new offering
- Establishing a better tourist attraction image

The objective of the tourism market segmentation in this paper belongs to the first category above – identifying the characteristics of groups of tourists who travelled with the same significant spatial movement pattern. Therefore, the geographic, socio-demographic and trip-related behavioural variables are used.

3.2 Tourist market segmentation methods

Market segmentation is a process of dividing a market into homogeneous subgroups. Tourists in the same group are similar to each other, and different from other groups, in the way they react to the market mix such as promotion or advertising and influencing spatial behaviours (Weinstein, 2004). Market segmentation has been extensively used as a valuable method to identify tourist market groups, select the target market and position the tourist market (Haley 1968; Chandra and Menezes 2001; Mykletun, Crotts et al. 2001; Kotler and Armstrong 2003; Lee, Morrison et al. 2005). Market segmentation is equivalent to data clustering and as such, algorithms developed in the area of clustering can be applied here.

Model-based clustering, also known as mixture modelling, assumes that each cluster has an underlying probability density function or model, and that the density for the mixture is a weighted sum of the cluster densities. That is, given observed data y_i for $i = 1,...,N$, where each y_i is a vector of M measured variables, the probability density function *p* for a mixture model composed of *K* clusters is given by

$$p(y \mid \theta) = \sum_{j=1}^{K} \alpha_j p_j \left(y \mid \theta_j \right) \tag{1}$$

where $p_j \left(y \mid \theta_j \right)$ is the probability density function for the *j*-th cluster and the mixture weights α_j represent the proportions of data belonging to the *j*-th cluster such that $\sum_{j=1}^{K} \alpha_j = 1$. The parameters θ of the mixture model comprise the parameters θ_j of the cluster densities as well as the mixture weights α_j. An advantage of model-based clustering is that it can be used for both numeric and categorical data. Conditional on cluster membership, numeric variables can be modelled by normal distributions and categorical variables can be modelled using multi-way cross-classification tables, with all variables assumed conditionally mutually independent.

The Expectation-Maximization (EM) algorithm for incomplete data (Dempster et al. 1977) can be used to perform maximum likelihood parameter estimation for mixture models. It applies the principle of maximum likelihood to find the model parameters, by iteratively repeating the Expectation (E) and Maximisation (M) steps after randomly initialising the mixture model parameters. The E and M steps are iterated until a desired convergence is achieved (Witten and Frank 2000).

At iteration *t*, given parameter estimates $\theta_j^{(t)}$ and $\alpha_j^{(t)}$, the E step calculates the probabilities w_{ij} that the *i*-th observation belongs to the *j*-th cluster using Bayes' rule:

$$w_{il}^{(t+1)} = \frac{\alpha_l^{(t)} p\left(y_i \mid \theta_l^{(t)}\right)}{\sum_{j=1}^{K} \alpha_j^{(t)} p\left(y_i \mid \theta_j^{(t)}\right)} .$$

The M step then fixes the cluster membership probabilities and maximises the likelihood to update the model parameters. The maximum likelihood estimates of the mixture weights are given by

$$\alpha_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} w_{ij}^{(t+1)} . \tag{2}$$

The weights are then incorporated into equation (1) and parameters $\theta_j^{(t+1)}$ are found by maximising the likelihood.

When the observed variables take numeric values, each cluster is assumed to follow a multivariate normal distribution with mean μ_j and covariance matrix Σ_j. Thus, the parameters of the probability density function for the *j*-th cluster are $\theta_j = \left\{ \mu_j, \Sigma_j \right\}$. The maximum likelihood estimates can be calculated explicitly at the M step of the EM algorithm, as follows:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{N} w_{ij}^{(t+1)} y_i}{\sum_{i=1}^{N} w_{ij}^{(t+1)}},$$

and

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{N} w_{ij}^{(t+1)} \left(y_i - \mu_j^{(t+1)} \right) \left(y_i - \mu_j^{(t+1)} \right)^T}{\sum_{i=1}^{N} w_{ij}^{(t+1)}}.$$

When the observed variables take categorical values, the cluster distributions are assumed to be multi-way contingency tables with all variables conditionally and mutually independent and the mixture modelling approach is equivalent to latent class analysis (Agresti 2002; Vermunt and Magidson 2002). Suppose that there are M categorical variables, each of which may take K_m possible values. Let y_{imk} be equal to one if observation *i* takes category *k* for the *m*-th variable. Let π_{mjk} be the cluster-conditional probability that an observation from cluster *j* takes the *k*-th category for the *m*-th variable. Thus, within each class, for each variable $\sum_{k=1}^{K_m} \pi_{mjk} = 1$. The probability density function for the *j*-th cluster is then fully parameterised by $\theta_j = \left\{ \pi_{mjk} \right\}$ for $i = 1,...N$ and $m = 1,...M$, and the probability density function for the mixture model is given by:

$$p(y \mid \theta) = \sum_{j=1}^{K} \alpha_j \prod_{m=1}^{M} \prod_{k=1}^{K_m} \left(\pi_{mjk} \right)^{y_{imk}} \qquad (3)$$

and the maximum likelihood estimates are calculated at the M step of the EM algorithm by:

$$\pi_{mj}^{(t+1)} = \frac{\sum_{i=1}^{N} y_{ik} w_{ij}^{(t+1)}}{\sum_{i=1}^{N} w_{ij}^{(t+1)}}, \qquad (4)$$

where $\pi_{mj}^{(t+1)}$ is the vector of length K_m of categorical probabilities for the *m*-th variable, conditional on cluster *j* and y_{ik} is the $N \times K_m$ matrix of observed outcomes y_{imk} on that variable.

There are two issues that arise when the EM algorithm is applied in practice. First, because the initialisation is random, the algorithm may converge to local likelihood maxima instead of the global maximum. This can be avoided by running the algorithm many times, and choosing the model results that correspond to the global maximum. Second, the EM algorithm can not be used to compute the number of clusters because the likelihood of a mixture model always increases as the number of clusters (and number of parameters) increases. To select the optimal number of clusters, the likelihood must be penalised as the number of model parameters increases. The Akaike information criterion (Akaike 1973) is commonly used for this purpose. It offsets the likelihood by the number of parameters, thereby maximising the ‘goodness of fit’ of the model while minimising the model complexity.

3.3 Selection of target markets

Target markets can be defined as market segments that are selected for a certain purpose, such as positioning a product or service based on certain criteria (Myers

1996). For example, in order to discover the most profitable market, mean tourist expenditure in each of the segments can be evaluated (Jang, et al. 2001). The aim of market segmentation in this section is to identify the socio-demographic characteristics of tourists and their travel behaviours for significant spatial movement patterns. There could be several market segments identified for a movement pattern. However, the segment with the heaviest users is selected as the target segment (Myers 1996). For example, if there are three different market segments identified for a movement pattern, the target market segment is the one with highest percentage of members or tourists.

4 Case study

4.1 Case study area

Phillip Island, located at the mouth of Westernport Bay, is 140 kilometres south-east of Melbourne, Australia. It covers approximately 10 000 hectares and is 26 km long and 9 km wide. The permanent population on the island is around 7 000, but currently almost 1.5 million visitors travel around the island each year (Phillip Island Internet Services 2005). There are a large number of natural features on the island, such as penguins, koalas, seals, shearwaters, mangroves, wetlands, sandy beaches and rugged rocky cliff faces. The major attractions are the Penguin Parade, the Koala Conservation Centre, Cowes, Churchill Island, Rhyll Inlet, Woolamai and the Nobbies (Figure 1), where visitors can experience wildlife in a natural environment. Each of the 10 identified major destinations are labelled A to J.

Figure 1 Map of Phillip Island (Phillip Island Nature Park 2005)

4.2 Movement data

Tourists' daily movement data for the Phillip Island Nature Park were collected via a self-administered questionnaire. The questionnaire was designed to ascertain basic visiting patterns and socio-demographic data including how tourists use Phillip Island, what they would like to see and how their satisfaction could be enhanced whilst at Phillip Island. The collected data made it possible to quantify and classify visitors based upon their socio-demographic background, group type and motivation for visit. This will allow park managers to develop more effective and efficient strategies for managing natural and recreation resources.

The questionnaire was designed to address three different areas. The first section aimed to acquire socio-demographic data or profiles of the tourists. The second aimed to collect information regarding travel mode, length of stay and with whom the tourist was travelling. The final section of questions gathered information on tourist movement. Here tourists were asked to write down their approximate arrival time and duration of stay at each attraction visited for the entire day. Tourists were also required to draw the route of travel to each attraction on a street map of Phillip Island. Eight hundred questionnaires were distributed from the 6-8th of March 2004 and from 17-20th Jan 2005 at the Phillip Island Nature Park Information Centre, Churchill Island, Koala Conservation Centre and Penguin Parade. Penguin Parade was the major sample site because more than 90% of tourists visit this attraction. Five hundred questionnaires were returned with 457 entered into the database. The other 44 questionnaires were incomplete and discarded.

4.3 Identification of statistically significant movement patterns of tourists on Phillip Island

In this case study, the movement of tourists visiting Phillip Island Nature Park and moving around nine attractions was considered. Based on the survey, tourists visited, at most, seven attractions during a day-long trip. Seven groups of movement patterns ranging from one-attraction movement patterns to seven-attraction movement patterns were classified. Statistically significant attraction associations were identified in each of the seven movement patterns groups using the General Log-linear Analysis tool (GENLOG) in Statistical Package for the Social Sciences (SPSS).

In the two-attraction movement pattern group, movement patterns, FG (Cowes to Penguin Parade), HG (Nobbies to Penguin Parade) and DG (Koala Conservation Centre to Penguin Parade) were identified as being strongly associated with each other with *p*-values less than 0.05 (Table 1). Therefore, these three movement patterns were considered dominant two–attraction patterns in the data set.

Table 1 Parameter estimates for two, three and four-attraction movement patterns

Table 1 shows the significant three-attraction movement patterns and four-attraction movement patterns, respectively. Because there are not many tourists visiting more than five attractions during their day trip on Phillip Island, the frequencies of major movement patterns composed of five, six and seven attractions were tabulated, instead of using log-linear analysis (See Table 2).

Table 2 Frequencies for five, six and seven-attraction spatial movement patterns (Code of attractions see Figure 1)

4.4 Tourism market segmentation using EM algorithm

This section describes how the collected geographic, socio-demographic and trip-related variables were used to identify the characteristics of tourists who travelled with the statistically significant movement patterns on Phillip Island.

We firstly segmented all tourists (457) in the dataset using the EM algorithm to provide a market segmentation of tourists who travelled to Phillip Island based on their general characteristics. The key variables used to distinguish tourist segments were ‘*Type of tourist*’, ‘*Type of group*’, ‘*Transport*’, ‘*Lifecycle*’, ‘*Return Visitor*’ and ‘*Gender*’. Because the variables were all categorical, the EM algorithm in this case is equivalent to latent class analysis (section 3.2). The probability density function for the mixture model is given by equation 3. The model parameters are the mixture weights, α_j, which are the proportions of tourists assigned to each of the *j* clusters (in this case, the market segments), and the probability vectors π_{mj} containing the probabilities that the m-th variable, i.e.,‘Type of tourist’ takes each of its possible categories, conditional on cluster *j*. The E step of the EM algorithm updates the mixture weights according to equation 2 and the M step updates the probability vectors according to equation 4. This is readily performed using the freely-available poLCA package, which is run from within the R statistical package (Linzer et al, 2007). The package was used to run the EM algorithm with the number of clusters varying from one to eight, with 500 random initialisations for each number of clusters. We minimised the Akaike information criterion (AIC) (Akaike 1973) to calculate the optimal number of clusters or classes of tourists, which was found to be six (figure 2), and simultaneously found the model parameters that maximised the likelihood.

Figure 2 Minimisation of the AIC identified six tourist segments.

The results are shown in table 3. In the first row of table 3, we see that the proportion, α_1, of visitors in the first market segment is 0.34. The probabilities, π_{m1}, that tourists in this market segment belong to each category of the six key variables, $m = 1, ..., 6$, are shown in the same row., When one category clearly dominates, the corresponding cell is shaded in grey. We see that the first segment is mainly composed of return, domestic visitors who are families with children aged six years or older, travelling by car. Since it is the largest market segment, the first segment can be considered the target market for Phillip Island. However, the second and third segments together comprise almost half of the market (proportions 0.25 and 0.17) and should not be ignored. The second segment mainly differs from the first by including people from a wider range of lifecycle stages. In contrast, the third segment is composed of first-time, international visitors, half of whom are travelling by tour bus or coach and who are from all stages of life. Phillip Island tour packages would ideally be targeted at this third market segment.

Table 3 Tourism market segments for the significant movement patterns

The same procedure was repeated for the dominant movement patterns 'G', 'FG', 'HG', 'DG', 'DFG', 'FHG', 'HFG', 'BFG' and 'DFHG' to identify the characteristics of tourist groups who travelled with these patterns (table 3). For patterns 'G' and

‘FG’, three clusters or market segments were identified by the EM algorithm and minimisation of the AIC. For the other movement patterns, the sample sizes were too small to perform clustering and therefore only single segments can be identified.

The market segments who travelled with different patterns did not necessarily have the same dominant segments as Phillip Island visitors in general. For example, the dominant sector of tourists following movement pattern ‘G’ tended to be first time, younger visitors to Phillip Island. However, the dominant sector of tourists travelling with movement pattern ‘FG’ corresponded reasonably well with the dominant sector for Phillip Island as a whole that was identified using all of the tourists.

Looking more closely at the market segments and patterns, we can see that high proportions of movement patterns ‘FG’ and ‘DFHG’ were travelled by older couples. Young single people travelled with movement patterns ‘G’, ‘HG’, ‘DG’, ‘DFG’ and ‘HFG’. Some patterns such as ‘DG’ and ‘DFG’ were used more often by international visitors and some patterns such as ‘FHG’ and ‘BFG’ were dominated by domestic visitors. Young international visitors preferred travelling in organised groups by tour bus. Young couples with no children travelled together by car or campervan. Most domestic visitors were returning to Phillip Island with their families, friends or relatives again. The majority of international visitors have visited Phillip Island only once.

5 Conclusions and future research

This paper presents a novel method to segment tourist markets associated with dominant movement patterns of tourists. Log-linear models were used to identify dominant movement patterns of tourists. Tourists who travelled with the same dominant movement pattern, were then divided into different segments using the EM algorithm based on the geographic, socio-demographic and trip-related behavioural variables. This method was specifically applied to develop tour packages for Phillip Island Nature Park.

A survey was conducted on Phillip Island Nature Park. Eight hundred questionnaires were distributed at four major attractions on the island. Tourists visiting the island who did not visit these attractions were not surveyed. However, the Park managers estimate that these would comprise less than 10% of tourists visiting the island and therefore it is unlikely that significant movement patterns were missed by the survey.

Log-linear models tested the significance of movement patterns and provided the threshold for dominant patterns using a *p*-value at the level of 0.05. The dominant movement patterns can not only be used to develop tour package but also to assist park managers in deciding how long to open an attraction and how the daily program of activities should be arranged for an attraction. One important issue for utilisation of log-linear models is sample size. The case study shows that if zero is recorded too often as the frequency of movement patterns, the expected frequencies of movement patterns will converge to zero during iterative fitting. Therefore, a large sample size is necessary (Kennedy 1992).

Small sample size is the major limitation of this case study. Ideally, the EM algorithm would be used to find market segments for each significant movement pattern. However, this could not be done using the Phillip Island data because of the small sample numbers of tourists travelling with each of the patterns. Instead, we have used the EM algorithm to find market segments using all of the available tourist data and then applied it to identify market segments for the movement patterns with sufficient data. This approach provides insight that may be used to assist the development of tour packages for Phillip Island, but with more available data the results may be improved and of greater detail.

Sometimes, expert experience or knowledge is needed to identify the characteristics of tourism market segments. Witten and Frank (2000) suggest that a supervised model, namely, the classification method, could be used to analyse the results of the EM algorithm.

The large number of daily monetary transactions (purchases) in the Phillip Island Nature Park suggests that in future these methods could be applied to tourist monetary transaction data. The tourist transaction data can record tourist ID numbers, the ID numbers of attractions that tourists visited, the dates of the transaction, and in-store tourist information such as tourist profiles. Therefore, based on these transaction records, tourist movement patterns could be identified and further market segments identified. However, this method is only suitable for developed attractions with their own information centre and EFTPOS (Electronic Funds Transfer at the Point of Sale) machines. Of course, the privacy issues regarding the use of transaction data also need to be considered.

Reference:

Agresti, A. (1990). *Categorical data analysis.* New York ; Brisbane, Wiley.

Agresti, A. (2002). *Categorical data analysis.* (2nd ed.) New York Wiley-Interscience.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In S. Kotz and N.L. Johnson (Eds), Breakthroughs in Statistics, Vol. 1: Foundations and basic Theory, pp 199-213. Springer-Verlag, New York, 1992. Originally published in the *Proceedings of the International Symposium of Information Theory*, 1973.

Allen, G. (1999). Cognitive Abilities in the Service of Wayfinding: A Functional Approach. *The Professional Geographer* **51**(4): 555-561.

Arrowsmith, C. and R. Inbakaran (2002). Estimating environmental resiliency for the Grampians National Park, Victoria, Australia: a quantitative approach. *Tourism Management* **23**(3): 295-309.

Bigné, J. E. and L. Andreu (2004). Emotions in segmentation: An Empirical Study. *Annals of Tourism Research* **31**(3): 682-696.

Bloom, J. Z. (2005). MARKET SEGMENTATION: A Neural Network Application. *Annals of Tourism Research* **32**(1): 93-111.

Calver, S. (1999). "Modelling Service Sector Characteristics Service Sector Market Segmentation II, http://apollo4.bournemouth.ac.uk/si/scalver/133segemnt.html, Retrieved 07/07/2005.

Chandra, S. and D. Menezes (2001). Applications of Multivariate Analysis in International Tourism Research: The Marketing Strategy Perspective of NTOs. *Journal of Economic and Social Research* **3**(1): 77-98.

Dempster, A. K, Laird, N. M. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39:1-38.

Devore, J. L. (2004). *Probability and statistics for engineering and the sciences* (6th ed.) Belmont, CA Thomson-Brooks/Cole,.

Emel, G. and Ç. Taşkın, Eds. (2005). Identifying Segments of a Domestic Tourism Market by Means of Data Mining. *Operations Research Proceedings 2005*. .

Goodman, L. A. (1984). *The analysis of cross-classified data having ordered categories*. Cambridge, Mass Harvard University Press.

Grizzle, J. E., C. F. Stammer, et al. (1969). Analysis of categorical data by linear models. *Biometrics* **25**: 489-504.

Haley, R. I. (1968). Benefit segmentation: A decision-oriented research tool. *Journal of Marketing* **32**(3): 30-35.

Han, J. and M. Kamber (2006). *Data mining: concepts and techniques* (2nd ed.) San Francisco, Morgan Kaufmann

Hosmer, D. W. and S. L. Lemeshow (1989). *Applied logistic regression*. New York, John Wiley.

Jang, S. C., A. M. Morrison, et al. (2001). Benefit segmentation of Japanese pleasure travelers to the USA and Canada: selecting target markets based on the profitability and risk of individual market segments. *Tourism Management* **23**(4): 367-378.

Jurowski, C. and A. Z. Reich (2000). An Explanation and Illustration of Cluster Analysis for Identifying Hospitality Market Segments. *Journal of Hospitality Tourism Research* **24**(1): 67-91.

Kennedy, J. J. (1992). *Analyzing qualitative data*: *log-linear analysis for behavioural research.* (2nd ed.) New York Praeger.

Kotler, P. and G. Armstrong (2003). *Principles of marketing*. (10th ed.) Upper Saddle River, NJ, Prentice-Hall.

Lee, G., A. M. Morrison, et al. (2006). The economic value portfolio matrix: A target market selection tool for destination marketing organizations. Tourism Management 27(4): 576-588.

Leiper, N. (1995). *Tourism management.* Collingwood, Vic, TAFE Publications.

Linzer, Drew A. and Jeffrey Lewis. 2007. poLCA: Polytomous Variable Latent Class Analysis. R package version 1.1. http://userwww.service.emory.edu/~dlinzer/poLCA.

McLaren, D. (2003). *Rethinking tourism and ecotravel* (2nd ed) Bloomfield, CT Kumarian Press.

Min, S.-H. and I. Han (2005). Detection of the customer time-variant pattern for improving recommender systems. *Expert Systems with Applications* **28**(2): 189-199.

Myers, J. H. (1996). *Segmentation and positioning for strategic marketing decisions.* Chicago American Marketing Association.

Mykletun, R. J., J. C. Crotts, et al. (2001). Positioning an island destination in the peripheral area of the Baltics: a flexible approach to market segmentation. *Tourism Management* **22**(5): 493-500.

Ogilvie, F. W. (1933). *The tourist movement: an economic study* London: Staples Press,.

Pearce, D. G. (1987). *Tourism today: a geographical analysis* (2nd ed.) Harlow, Essex, England : New York :, Longman Scientific & Technical ; Wiley.

Phillip Island Internet Services. (2005). Welcome to Phillip Island. http://www.phillipisland.net.au/trip/firehouse/firehouse.html Retrieved 09/05/05.

Phillip Island Nature Park. (2005). How to get to Phillip Island. www.Penguins.org.au.Retrieved 01/05/05.

Wedel, M. and W. A. Kamakura (2000). *Market Segmentation: Conceptual and Methodological Foundations*. Norwell, Kluwer Academic Publishers.

Weinstein, A. (2004). *Handbook of market segmentation: strategic targeting for business and technology firms* (3rd ed.) New York, The Haworth Press.

Vermunt and Magidson (2002) J.K. Vermunt and J. Magidson, *Latent Class Cluster Analysis*. In: J.A. Hagenaars and A.L. McCutcheon, Editors, Applied Latent Class Analysis, Cambridge University Press, Cambridge, UK. 89–106.

Witten, I. H. and E. Frank (2000). *Data mining: practical machine learning tools and techniques with Java implementations*. San Francisco, Calif., Morgan Kaufmann.

Xia, J. and C. Arrowsmith (2005). *Managing Scale Issues in Spatio-temporal Movement of Tourists Modelling*. MODSIM 05, Melbourne, Australia.

Yada, K., H. Motoda, et al. (2004). *Consumer Behavior Analysis by Graph Mining Technique*. 800-806.

Yan, X. and J. Han (2002). *gSpan: Graph-Based Substructure Pattern Mining*. Proceedings of the International Conference on Data Mining, Maebashi City, Japan.

Zhang, T., R. Ramakrishnan, et al. (1997). BIRCH: A New Data Clustering Algorithm and its Applications. *Data Mining and Knowledge Discovery* **1**: 141-182.

Table 1 Log-linear model parameter estimates for two, three and four-attraction movement patterns

Destination 1	Destination 2	Destination 3	Destination 4	Counts	Sig.
Cowes(F)	Penguin Parade(G)			68	0.001
Nobbies or Seal Rock(H)	Penguin Parade(G)			29	0.004
Koala Conservation Centre(D)	Penguin Parade(G)			14	0.019
Cowes(F)	Nobbies or Seal Rock(H)	Penguin Parade(G)		46	0.001
Koala Conservation Centre(D)	Cowes(F)	Penguin Parade(G)		27	0.005
Nobbies or Seal Rock(H)	Cowes(F)	Penguin Parade(G)		17	0.013
Koala Conservation Centre(D)	Nobbies or Seal Rock(H)	Penguin Parade(G)		12	0.026
Cape Woolamai(B)	Cowes(F)	Penguin Parade(G)		9	0.042
Koala Conservation Centre(D)	Cowes(F)	Nobbies or Seal Rock(H)	Penguin Parade(G)	20	0.009

Table 2 Frequencies of five, six and seven-attraction spatial movement patterns (Code of attractions see Figure 1)

Patterns	Number of attractions	Counts
CDFHG	5	6
BDFHG	5	2
DCGFH	5	2
JBJBJG	6	2
CBDEFHG	7	1
CFBDEHG	7	1

Table 3 Tourism market segments for the significant movement patterns

Movement patterns	Segment	α	Times to visit		Type of visitor		Transport			Lifecycle						Type of group			Gender	
			First time	Return	Domestic	Inter-national	Car	Tour bus or coach	Other	Middle to mature family	Mature single	Older couple	Older single	Young family	Young single	Alone	Organized group	With friends/ family	F	M
All Patterns	1 (146)*	0.34	0.23	0.77	1	0	0.93	0	0.07	0.7	0	0.05	0.04	0.19	0.01	0.01	0.04	0.95	0.61	0.39
	2 (115)	0.25	0.38	0.62	0.77	0.23	0.91	0	0.09	0.2	0.12	0.19	0.19	0	0.3	0	0	1	0.64	0.36
	3 (79)	0.17	1	0	0.08	0.92	0.48	0.5	0.02	0.29	0	0.15	0.07	0.07	0.43	0.09	0.09	0.83	0.46	0.54
	4 (49)	0.1	0.59	0.41	1	0	0.85	0	0.15	0.11	0.04	0.14	0.56	0.11	0.04	0.07	0.01	0.92	0.55	0.45
	5 (46)	0.1	0.97	0.03	0	1	0.74	0.04	0.22	0.06	0	0.21	0.63	0	0.1	0.03	0	0.97	0.65	0.35
	6 (22)	0.04	0.85	0.15	0.1	0.9	0.05	0.8	0.15	0.03	0.4	0.08	0	0	0.49	0.22	0.78	0	1	0
G	1 (42)	0.51	0.61	0.39	0.75	0.25	0.82	0.03	0.15	0.09	0.05	0.14	0.49	0	0.23	0.03	0	0.97	0.67	0.33
	2 (20)	0.26	1	0	0.22	0.78	0.45	0.46	0.09	0.44	0	0.11	0	0.05	0.4	0	0.14	0.86	0.64	0.36
	3 (19)	0.23	0.04	0.96	1	0	1	0	0	0.75	0	0.09	0	0.16	0	0.09	0	0.91	0.59	0.41
FG	1 (48)	0.69	0.33	0.67	1	0	1	0	0	0.52	0.08	0.13	0	0.15	0.12	0	0	1	0.65	0.35
	2 (17)	0.25	0.88	0.12	0.37	0.63	0.89	0	0.11	0.13	0.07	0.15	0.57	0	0.07	0.11	0	0.89	0.4	0.6
	3 (4)	0.06	1	0	0	1	0	0.75	0.25	0	0.75	0.25	0	0	0	0	1	0	1	0
HG	1 (29)	1	0.55	0.45	0.76	0.24	0.76	0.14	0.1	0.34	0	0.07	0.24	0.14	0.21	0.07	0.1	0.83	0.52	0.48
DG	1 (15)	1	0.73	0.27	0.4	0.6	0.87	0.13	0	0.47	0	0.13	0.13	0	0.27	0	0.14	0.86	0.6	0.4
DFG	1 (28)	1	0.61	0.39	0.43	0.57	0.5	0.46	0.04	0.32	0.07	0.07	0.18	0.11	0.25	0.14	0.18	0.68	0.57	0.43
FHG	1 (48)	1	0.35	0.65	0.74	0.26	0.77	0.06	0.17	0.46	0.04	0.15	0.17	0.04	0.15	0	0	1	0.57	0.43
HFG	1 (18)	1	0.67	0.33	0.44	0.56	0.56	0.33	0.11	0.17	0.06	0.17	0.17	0.06	0.39	0.06	0.22	0.72	0.61	0.39
BFG	1 ((9)	1	0.22	0.78	0.89	0.11	1	0	0	0.67	0.22	0	0	0.11	0	0	0	1	0.78	0.22
DFHG	1 ((20)	1	0.75	0.25	0.25	0.75	0.7	0.25	0.05	0.2	0.15	0.25	0.25	0.1	0.05	0.05	0.1	0.85	0.55	0.45

Note: each No. 1 segment for the movement patterns is considered to be the target market highlighted by red colour ; See the movement pattern codes in above section

* 1(146): 146 tourists in segment 1

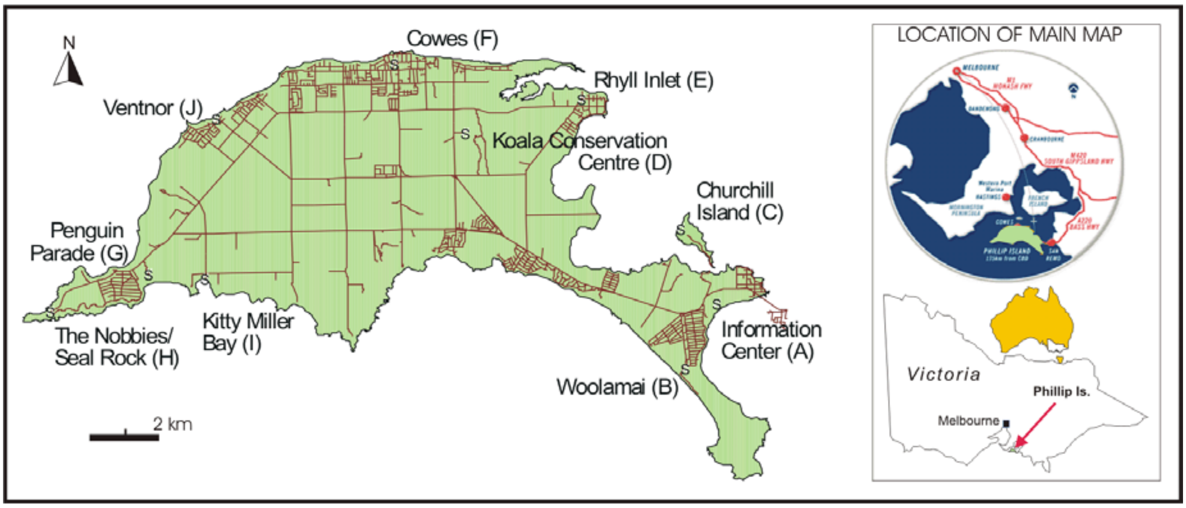

Figure 1 Map of Phillip Island (Phillip Island Nature Park 2005).

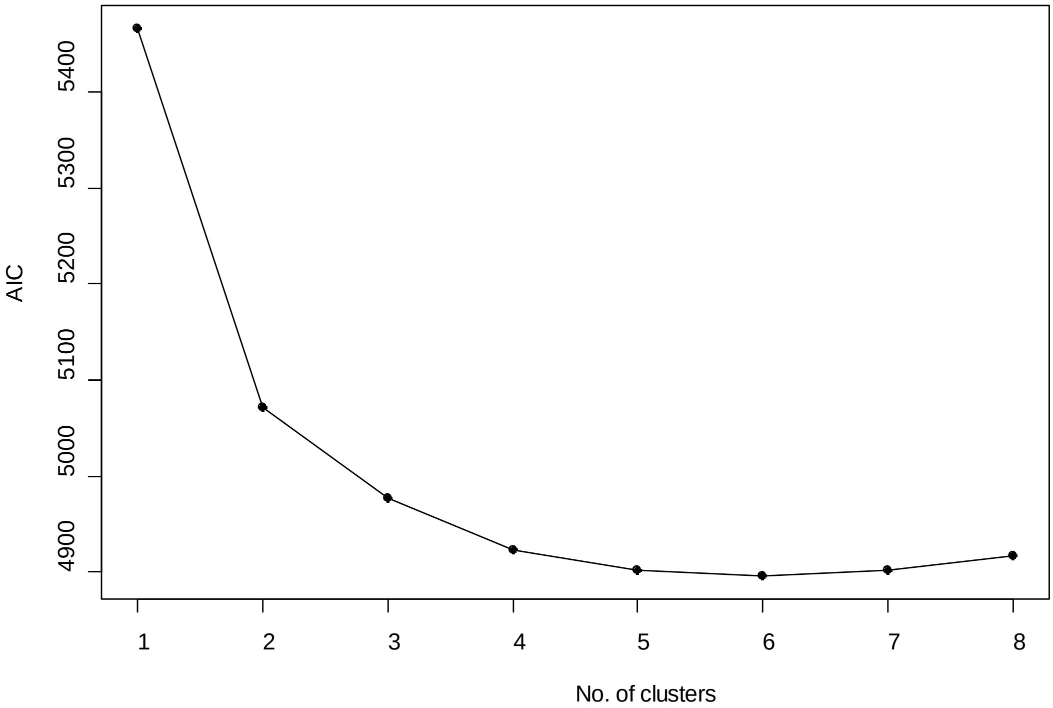

Figure 2. Minimisation of the AIC identified six groups of tourists.