A COMPARISON BETWEEN SINGLE LINKAGE AND COMPLETE LINKAGE IN AGGLOMERATIVE HIERARCHICAL CLUSTER ANALYSIS FOR IDENTIFYING TOURISTS SEGMENTS

NOOR RASHIDAH R.[1], SABRI A.[1] AND SAFIEK M.[2]

[1]*Department of Mathematics, Faculty of Science & Technology,*
[2]*Department of Management & Marketing, Faculty of Managment & Economics,*
Universiti Malaysia Terengganu, Kuala Terengganu, Terengganu, Malaysia.

noor_rashidah84@yahoo.com, sba@umt.edu.my, safiek@umt.edu.my

ABSTRACT: Cluster Analysis is a multivariate method in statistics. Agglomerative Hierarchical Cluster Analysis is one of approaches in Cluster Analysis. There are two linkage methods in Agglomerative Hierarchical Cluster Analysis which are Single Linkage and Complete Linkage. The purpose of this study is to compare between Single Linkage and Complete Linkage in Agglomerative Hierarchical Cluster Analysis. The comparison of performances between these linkage methods was shown by using Kruskal-Wallis test. The result of the comparison used for segmenting tourists of Kapas Island. The statistical software SPSS has been applied to analyze data of this research. The result from Kruskal-Wallis test shows Complete Linkage is more useful in identifying tourists segments.

ABSTRAK: Analisis Gugusan ialah satu kaedah multivariat dalam bidang statistik. Analisis Gugusan Aglomeratif Berhierarki ialah satu daripada pendekatan dalam Analisis Gugusan. Ada terdapat dua kaedah rantaian dalam Analisis Gugusan Aglomeratif Berhierarki iaitu Rantain Tunggal dan Rantaian Lengkap. Tujuan kajian ini ialah untuk mencari perbandingan antara Rantaian Tunggal dengan Rantaian Lengkap dalam Analisis Gugusan Aglomeratif Berhierarki. Perbandingan prestasi antara dua rantaian tersebut dibuat menggunakan Ujian Kruskal-Wallis. Keputusan perbandingan tersebut digunakan untuk meruas pelancong di Pulau Kapas. Perisian statistic SPSS telah digunakan bagi menganalisa data kajian. Keputusan Ujian Kruskal-Wallis menunjukkan Rantaian Lengkap adalah lebih berguna untuk mengenalpasti segmen pelancong.

KEYWORDS: *agglomerative hierarchical cluster analysis; single linkage; complete linkage; Kruskal-Wallis test; tourists*

1. INTRODUCTION

In statistics area, there are some methods available to gather observations. Some methods have been developed to divide a sample of observations into some smaller groups. One of the methods is Cluster Analysis. This method involves sorting observations into different groups based on their similarity. Cluster Analysis also refers as a collection of statistical methods that identifies groups of sample that show similar characteristics.

There are many approaches in Cluster Analysis. One of the approaches is Agglomerative Hierarchical Cluster Analysis. The first step need to be considered in this approach is computation of similarity among cases or observation. The similarities among

cases were considered as distance in Agglomerative Hierarchical Cluster Analysis. Euclidean Distance Measure will apply to compute the distance among cases in this study. The cases that have same similarities will be set in the same clusters or groups. The distance among clusters can be compute using Single Linkage or Complete Linkage methods. Single Linkage is a method that focused on minimum distances or nearest neighbor between clusters meanwhile Complete Linkage concentrates on maximum distance or furthest neighbor between clusters.

This research compares the efficiency of Single and Complete Linkage in Agglomerative Hierarchical Cluster Analysis. This comparison based on evaluation of the output for both linkage methods. Kruskal-Wallis is a method that will apply in this research to contrast the performances between Single Linkage and Complete Linkage. Kruskal-Wallis test is a non-parametric test used to make comparison between independent groups of sampled data. The objectives of this research are:

   a. To compare performances of Single Linkage and Complete Linkage in Agglomerative Hierarchical Cluster Analysis.

   b. To assign groups or clusters of tourists those visit Kapas Island, Terengganu.

Cluster Analysis is a multivariate data analysis method that groups similar objects together. Agglomerative Hierarchical Cluster Analysis is a method of Cluster Analysis. The method is initially seeking for the similarities between different points by using Euclidean distance measure. The similarities between different clusters are calculated using Single Linkage and Complete Linkage methods. Therefore, the comparison between these linkage methods by using Kruskal-Wallis test will be performed in determining the clusters of Kapas Island tourists. It is difficult to assign groups of these tourists since they come from various backgrounds. This problem is solved using Agglomerative Hierarchical Cluster Analysis.

Cluster Analysis is widely used family of multivariate techniques for grouping individuals, objects or behaviors into similar clusters [1]. The flexibility of cluster analysis to accommodate wide range of applications makes it one of the most useful tools for understanding the natural structures among observations [1]. In tourism research, for example, cluster analysis is often used to identify market segments in order to improve the effectiveness of marketing efforts These segments may be based on a variety of variables including demographic characteristics (such as age, income, gender and location) and trip characteristics (such as trip length, purpose, group size and benefits) [1]. Reference [2] stated hierarchical cluster analysis is a set of statistical techniques that is particularly useful for separating a set of objects into constituent group or clusters which minimize variation between members of the same groups without making assumptions about the number of groups or the group structure.

2. MATERIALS AND METHODS

2.1 Research Site and Instrument

The selection site for this research is Kapas Island. This island is located at Marang, Terengganu. The sample size preferred for Hierarchical Cluster Analysis is not more than 200 samples [3]. Reference [4] mentioned large data sets can be problems with Agglomerative Hierarchical Cluster Analysis. An alternative to Agglomerative Hierarchical Cluster Analysis for more than 200 data is given by various forms of nonhierarchical Cluster Analysis [4]. The sample of this research was 200 respondents

included local and international tourists that visit Kapas Island in July until September 2009. They have been chosen by using snowball sampling technique. It was one of the non-probability sampling techniques. By using this technique, the local and international tourists in this research have been chosen randomly.

A questionnaire was distributed to the sample of this research. Ten separate visitor surveys were carried out at Kapas Island. The mode of survey delivery for this research was self-administered questionnaire. The surveys were based on a 7-page questionnaire. There are three sections in this questionnaire. The sections are Section A, B and C. In Section A, it was included questions about the respondents' demographic profiles. Section B included questions about details of visit. There are 10 questions in this section. The items in this section are frequency of their visit to Kapas Island, the purpose to visit Kapas Island and so on. Section C is contained items of visitor satisfactions of Kapas Island. The respondents need to answer 24 questions about their characteristics of visit in Kapas Island. Likert Scale has been used in Section C.

2.2 Agglomerative Hierarchical Cluster Analysis

In this method, clustering of each observations or objects begins in separate clusters. Next, the clusters of the object or observation that are close together are merged to create one large cluster. The general formula for Agglomerative Hierarchical Cluster Analysis as follows [5]:

$$d_{k\to(r,s)} = \alpha_r d_{k\to r} + \alpha_s d_{k\to s} + \beta d_{r\to s} + \gamma \tag{1}$$

where

α_r = system parameter corresponds with cluster r
α_s = system parameter corresponds with cluster s
β = system parameter
γ = system parameter
$d_{k\to r}$ = distance between cluster k to cluster r
$d_{k\to s}$ = distance between cluster k to cluster s
$d_{r\to s}$ = distance between cluster r to cluster s

The value for all parameters as in Table 1 will be used for simplification of (1). Reference [3] has recommended the following constraints of parameter values to simplify (1).

$$\alpha_r + \alpha_s + \beta = 1; \alpha_r = \alpha_s; \gamma = -\tfrac{1}{2}; \beta < 1;$$

When $\alpha_r = \alpha_s$, hence

$$\alpha_r + \alpha_s + \beta = 1$$
$$\alpha_r + \alpha_r + \beta = 1$$
$$2\alpha_r + \beta = 1$$
$$2\alpha_r = 1 - \beta$$
$$\alpha_r = (1-\beta)/2$$

Here, $\alpha_r = (1-\beta)/2$. Next, a value of β needs to be selected. It is suggested that $\beta = 0$ since $0<1$. If a small value of β has been use such as $\beta = -0.5$ or $\beta = 0.5$, it becomes

$$\alpha_r + \alpha_s + \beta = 1$$
$$\tfrac{1}{2} + \tfrac{1}{2} + (-0.5) \neq 1$$

Or

$$\alpha_r + \alpha_s + \beta = 1$$
$$\tfrac{1}{2} + \tfrac{1}{2} + (0.5) \neq 1$$

Table 1: Value of Parameters.

Parameter	Complete Linkage	Single Linkage
α_r	½	½
α_s	½	½
β	0	0
γ	½	-½

2.3 Complete Linkage

There are some steps in getting the model or formula of Complete Linkage by using model of Agglomerative Hierarchical Cluster Analysis. By using all the values of parameter for Complete Linkage as in Table I into (1), it becomes

$$d_{k\to(r,s)} = \tfrac{1}{2}\, d_{k\to r} + \tfrac{1}{2}\, d_{k\to s} + \tfrac{1}{2}\, |d_{k\to r^-} \tag{2}$$

If $\quad d_{k\to r} > d_{k\to s} \quad$ then

$$|\, d_{k\to r} - d_{k\to s}\,| = d_{k\to r} - d_{k\to s} \tag{3}$$

Subsequently, (3) needs to be substituted into (2). Therefore, (2) reduces to become as follows:

$$
\begin{aligned}
d_{k\to(r,s)} &= \tfrac{1}{2}\, d_{k\to r} + \tfrac{1}{2}\, d_{k\to s} + \tfrac{1}{2}\, |d_{k\to r} - d_{k\to s}| \\
&= \tfrac{1}{2}\, d_{k\to r} + \tfrac{1}{2}\, d_{k\to s} + \tfrac{1}{2}\, d_{k\to r} - \tfrac{1}{2} d_{k\to s} \\
&= \tfrac{1}{2}\, d_{k\to r} + \tfrac{1}{2}\, d_{k\to r} \\
&= d_{k\to r}
\end{aligned}
$$

On the other hand, if $d_{k\to r} < d_{k\to s}$, then

$$|\, d_{k\to r} - d_{k\to s}\,| = d_{k\to s} - d_{k\to r} \tag{4}$$

By using all the value of parameters for Complete Linkage as in Table I, (2) reduces as follows:

$$
\begin{aligned}
d_{k\to(r,s)} &= \tfrac{1}{2}\, d_{k\to r} + \tfrac{1}{2}\, d_{k\to s} + \tfrac{1}{2}\, |d_{k\to r} - d_{k\to s}| \\
&= \tfrac{1}{2}\, d_{k\to r} + \tfrac{1}{2}\, d_{k\to s} + \tfrac{1}{2}\, d_{k\to s} - \tfrac{1}{2}\, d_{k\to r} \\
&= \tfrac{1}{2}\, d_{k\to s} + \tfrac{1}{2}\, d_{k\to s} \\
&= d_{k\to s}
\end{aligned}
$$

Since dk→r and dk→s is symmetric, the model of Complete Linkage approach can be written a follows:

$$d_{k\to(r,s)} = \max\,[(d_{k\to r}),\,(d_{k\to s})]$$

2.4 Single Linkage

There are some steps that need to follow to get the formula or model of Single Linkage. The first step that needs to follow is substitution the value of parameters for Single Linkage approach as in Table 1 into (2).

$$d_{k\to(r,s)} = \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}\,d_{k\to s} + (-\tfrac{1}{2})\,|d_{k\to r} - d_{k\to s}|$$

$$d_{k\to(r,s)} = \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}\,d_{k\to s} - \tfrac{1}{2}\,|d_{k\to r} - d_{k\to s}|$$

(5)

Since the condition of (3) is compulsory for the (5), substitutions of (3) into (5) need to be performed. Hence,

$$d_{k\to(r,s)} = \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}\,d_{k\to s} - \tfrac{1}{2}\,(d_{k\to r} - d_{k\to s})$$
$$= \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}\,d_{k\to s} - \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}d_{k\to s}$$
$$= \tfrac{1}{2}\,d_{k\to s} + \tfrac{1}{2}\,d_{k\to s}$$
$$= d_{k\to s}$$

When the substitution of (4) into (5) is done, the following equation will exist.

$$d_{k\to(r,s)} = \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}\,d_{k\to s} - \tfrac{1}{2}\,(d_{k\to s} - d_{k\to r})$$
$$= \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}\,d_{k\to s} - \tfrac{1}{2}\,d_{k\to s} + \tfrac{1}{2}\,d_{k\to r}$$
$$= \tfrac{1}{2}\,d_{k\to r} + \tfrac{1}{2}\,d_{k\to r}$$
$$= d_{k\to r}$$

Since dk→r and dk→s is symmetric, the model of Single Linkage approach can be written a follows:

$$d_{k\to(r,s)} = \min\,[(d_{k\to r}),\,(d_{k\to s})]$$

2.5 Kruskal-Wallis Test

Kruskal-Wallis test is one of statistical tests in nonparametric statistic. Comparative studies frequently involve the simultaneous comparison not just of two but of three or more treatments or conditions [7]. Kruskal-Wallis test used to compare between Single Linkage and Complete Linkage in this research.

The first procedure in Kruskal-Wallis test is ranking all the observations in the combined sample. Data values are grouped and need to be ranked. Next, compute the sum of the ranks for each cluster. The formula of sum of the ranks, $\sum r_i$ is given as follows:

$$\sum r_i = \frac{r_{i1} + r_{i2} + \cdots + r_{in\,i}}{n_i}$$

where

n_i = number of subjects to the ith treatment.

r_{il} = rank in the 1st treatment group.

r_{i2} = rank in the 2nd treatment group

r_{ini} = rank in the n_i treatment group

The Kruskal-Wallis test is applied after the computation the sum of the ranks. The assumptions for this test are all samples are random samples from their respective population and the measurement scale is at least ordinal. The Kruskal-Wallis test statistic is given by:

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{\overline{r_i}^2}{n_i} - 3(N+1)$$

where

N = total of respondents

n = total of respondents in each cluster

$\overline{r}$ = total of rank

The null hypothesis of Kruskal-Wallis test for a population is they have the same means. This hypothesis can be written in terms of the respective treatment effects as:

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_d$$
$$H_1 : \text{at least two } \theta s \text{ differ}$$

3. RESULTS AND DISCUSSION

Ordinal data gathered from research respondents usually not normal distribution, therefore it needs to be analyzed using nonparametric tests [8]. The purpose of normality test is to check whether all the variables that will be applied is not normally distributed since Kruskal-Wallis test is one of the approaches in nonparametric statistic.

3.1 Normality Test for Ordinal Data

The assumption is all the variables for ordinal data are qualitative. The hypotheses for this test are as follows:

$$H_0 : \text{The sample comes from a normal distribution}$$
$$H_1 : \text{The sample does not come from a normal distribution}$$

Based on Table 2, from the Kolmogorov-Smimov test it can be conclude that since the significant value (p-value) for all variables are $0.000 < 0.05$, all the variables are not normally distributed. Reference [8] stated for Kolmogorov-Smirnov and Shapiro-Wilk tests, data are normal distribution if both of them are not significant, which Sig. > 0.05. Here, there is enough evidence at the 5% level of significance that significant values (Sig.) for all variables are 0.000 which is less than 0.05. Hence, it can reject H_0 from the above hypotheses. It can be accepted that the sample does not come from normal distribution.

3.2 Determination Number of Clusters

The formula of Rule of Thumb has been used to determine the number of clusters. The formula as follows:

$$k = \sqrt{(n/2)}$$

where n = number of object

Since the tourists of Kapas Island be the object that need to build clusters among them, here n=200. Hence

$$k = \sqrt{(200/2)} = 10$$

It shows that the number of clusters that need to be built in this research is ten clusters by using the formula of Rule of Thumb.

Table 2: Test of normality.

	Kolmogorov-Smirnov(a)		
	Statistic	*df*	*Sig.*
WEATHER	.247	200	.000
CLEANLIN	.283	200	.000
SCENERY	.296	200	.000
ATMOSPHE	.271	200	.000
SAFETY	.192	200	.000
FRIENDLI	.188	200	.000
ACCOMODA	.224	200	.000
LOTSTOSE	.203	200	.000
TRANSP_A	.297	200	.000
ACCOMO_A	.248	200	.000
ACCOMO_B	.251	200	.000
AVAILABI	.265	200	.000
ENTERTAI	.246	200	.000
AVAILIBI	.206	200	.000
ACCOMO_C	.204	200	.000
COSTOFAC	.229	200	.000
SOUVENIR	.191	200	.000
SIGNAGE	.251	200	.000
OVERALLC	.178	200	.000
TASTEOFF	.230	200	.000
PRICEOFF	.199	200	.000
VARIETYO	.212	200	.000
FRIENDLY	.243	200	.000
APPEARAN	.237	200	.000

3.3 Data Analysis of Agglomerative Hierarchical Cluster Analysis Using Single Linkage

When using Single Linkage in Agglomerative Hierarchical Cluster Analysis the members of ten clusters are as presented in Table 3.

Table 3 shows the number of members for each cluster when applied Single Linkage in Agglomerative Hierarchical Cluster Analysis. Cluster 1 had the majority members which 191 members while Cluster 2, 3, 4, 5, 6, 7, 8, 9 and 10 only had one member in their cluster.

Table 3:

Total of members of ten clusters when applying Single Linkage in Agglomerative
Hierarchical Cluster Analysis.

Cluster	Total of members	Percentage
1	191	95.5
2	1	0.5
3	1	0.5
4	1	0.5
5	1	0.5
6	1	0.5
7	1	0.5
8	1	0.5
9	1	0.5
10	1	0.5
Total	200	100

3.4 Data Analysis of Agglomerative Hierarchical Cluster Analysis Using Complete Linkage

When using Complete Linkage in Agglomerative Hierarchical Cluster Analysis the
members of ten clusters are as follows:

Table 4 shows the number of members for each cluster when applied Complete Linkage in
Agglomerative Hierarchical Cluster Analysis. Cluster 1 had the majority members which
61 members while Cluster 10 only had the minority members which only two members in
its cluster.

Table 4:
Total of members of ten clusters when applying Complete Linkage in Agglomerative
Hierarchical Cluster Analysis

Cluster	Total of members	Percentage
1	61	30.5
2	44	22
3	4	2
4	6	3
5	18	9
6	12	6
7	30	15
8	15	7.5
9	8	4
10	2	1

Total	200	100

Table 4 shows the number of members for each cluster when applied Complete Linkage in Agglomerative Hierarchical Cluster Analysis. Cluster 1 had the majority members which 61 members while Cluster 10 only had the minority members which only two members in its cluster.

3.5 Calculation of Kruskal Wallis test from application of Single Linkage in Agglomerative Hierarchical Cluster Analysis

The responds for each respondent on Likert scale questions in this research have been total up. The data will be analyzed using Kruskal Wallis test. Hypotheses of research are as follows:

H0 : Ten clusters of tourists have same satisfaction value about Kapas Island

H1 : Ten clusters of tourists have different satisfaction value of Kapas Island.

Step 1: Arrangement of positions and rank for ordinal scale score

The positions of respondents have been arranged in ascending order which start f rom 1 until 200. It is because there were 200 respondents. Based on the positions, the ranks of respondents have identified. There were some respondents that have same satisfaction value with other respondents. i.e respondents 16, 139 and 183. So their rank was $\frac{8 + 9 + 10}{3} = 9.0$

Step 2: Calculation total of ranks for each cluster.

Total of ranks for each cluster is shown as follows:

Table 5: Total of rank of rank for ten clusters (SINGLE linkage).

Cluster	Total of Rank
1	19714.5
2	19.0
3	25.0
4	25.0
5	39.5
6	47.5
7	45.0
8	52.0
9	16.5
10	72.5

According to Table 5, it shows Cluster 1 has the highest total of rank (19714.5) meanwhile Cluster 9 has the lowest total of rank (16.5).

Step 3: Calculation of estimation Kruskal Wallis, KW.

By using the formula as follows, the estimation value of KW can be determined.

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{\overline{r_i}^2}{n_i} - 3(N+1)$$

$$KW = \frac{12}{200(200+1)} \left[\frac{(19714.5)^2}{191} + \frac{(19)^2}{1} + \frac{(25)^2}{1} + \frac{(52)^2}{1} + \frac{(39.5)^2}{1} + \frac{(47.5)^2}{1} + \frac{(45)^2}{1} + \frac{(52)^2}{1} + \frac{(16.5)^2}{1} + \frac{(72.5)^2}{1} + \right] - 3(200+1)$$

$$= \frac{12}{40200} \left[\begin{array}{l} 2034877.0170 + 361 + 625 + 2704 + 1560.25 + \\ 2256.25 + 2025 + 2704 + 272.25 + 5256.25 \end{array} \right] - 603$$

$$= 0.0003[2052641.017] - 603$$

$$= 12.7923$$

Step 4: Finding critical value of KW

The degree of freedom is *df*=*k*-1=10-1=9. By referring Table of critical value for Chi Square on *df*=9 and significant level, *p*=0.005, the critical value of KW is 16.92.

Step 5: Making decision for Kruskal Wallis test

The comparison value of estimation KW and critical value of KW shows that the estimation value KW (12.7923) lower than critical value of KW (16.92). Therefore it is accepted that hypothesis null, H_0 which stated ten clusters of respondents or tourists have same satisfaction value about Kapas Island.

3.6 Calculation of Kruskal Wallis test from Application of Complete Linkage in Agglomerative Hierarchical Cluster Analysis

Step 1 and Step 2 for calculation of Kruskal Wallis for the data of clusters that exist when applying Complete Linkage in Agglomerative Hierarchical Cluster Analysis is same as analysis for Single Linkage in Agglomerative Hierarchical Cluster Analysis.

Table 6: Total of rank and mean of rank for ten clusters (Complete linkage)

Cluster	Total of rank	Mean of rank
1	10037	164.5410
2	3498.5	79.5114
3	184	46.0
4	103	17.1667
5	1661	92.2778
6	131.5	10.9583
7	3400	113.3333
8	693	46.20
9	347.5	43.4375
10	5	2.5

According to Table 6, it shows Cluster 1 has the highest total of rank (10037) meanwhile Cluster 10 has the lowest total of rank (5).

Step 3: Calculation of estimation value of Kruskal Wallis, KW

By using the formula as follows, the estimation value of KW when applied Complete Linkage in Agglomerative Hierarchical Cluster Analysis can be calculated.

$$KW = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{\overline{r_i}^2}{n_i} - 3(N+1)$$

$$= \frac{12}{200(201)}\left[\begin{array}{l}\frac{(10037)^2}{61} + \frac{(3498.5)^2}{44} + \frac{(184)^2}{4} + \\ \frac{(103)^2}{6} + \frac{(1661)^2}{18} + \frac{(131.5)^2}{12} + \\ \frac{(3400)^2}{30} + \frac{(693)^2}{15} + \frac{(347.5)^2}{8} + \frac{(5)^2}{2}\end{array}\right] - 3(200+1)$$

$$= 0.0003\left[\begin{array}{l}1651497.8525 + 278170.5057 + 8464 + 1768.1667 + \\ 153273.3889 + 1441.0208 + 385333.3333 + \\ 30745.176 + 15094.5313 + 12.5\end{array}\right] - 603$$

$$= 0.0003[2525800.4752.] - 603$$

$$= 154.7401$$

Step 4: Finding the critical value of Kruskal Wallis, KW

The degree of freedom is similar as in case of Single linkage. The critical value of KW can be found by referring the Table of critical value for Chi Square. By referring the table, on *df*=9 and *p*=0.05, the critical value of KW is 16.92.

Step 5: Making decision for Kruskal Wallis test

By referring Step 3 and Step 4 for this case, it shows the estimation value of KW (154.7401) is higher than estimation value of KW (16.92). Therefore the null hypothesis was rejected. It can be concluded that ten clusters of respondents that occurred after applying Complete Linkage in Agglomerative Hierarchical Cluster Analysis have the different satisfaction value of Kapas Island. It shows that Cluster 1 have the highest satisfaction value of Kapas Island compare than other clusters (mean of rank for Cluster 1=164.5410, mean of rank for Cluster 2=79.5114, mean of rank for Cluster 3=46.0 mean of rank for Cluster 4=17.1667, mean of rank for Cluster 5=92.2778, mean of rank for Cluster 6=10.9583, mean of rank for Cluster 7=113.3333, mean of rank for Cluster 8=46.20, mean of rank for Cluster 9=43.4375, mean of rank for Cluster 10=2.5).

4. CONCLUSION

This study shows that the application of Complete Linkage approach in Agglomerative Hierarchical Cluster Analysis is more useful compare than Single Linkage approach in segmenting tourists of Kapas Island. It is because the result from the application of Complete Linkage in Agglomerative Hierarchical Cluster Analysis shows the difference of satisfaction value between ten clusters of tourists. If the clusters had same

satisfaction value of Kapas Island, it means the clusters had no difference among them and there is no occurred clusters among tourists.

ACKNOWLEDGEMENTS

REFERENCES

[1]   D.R. Fesenmaier, and J. Jeng, Cluster Analysis. In Jafar, J., Encyclopedia of Tourism,. London and New York: Routledge Taylor & Francis Group, 2003, vol.1, p.85.

[2]   S.C. Leung, W.K. Fung, and K.H. Wong, "The identification of credit card encodes by hierarchical cluster analysis of the jitters if magnetic stirpes," Science & Justice Journal, vol. 1, pp.85, 2003.

[3]   M.R. Yaacob, SPSS For Business And Social Science Students Version 14 For Windows, 1st ed. Kota Bharu, Kelantan: Pustaka Aman Press, 2008.

[4]   P.K. Hopke, and L.Kaufman, "The use of sampling to cluster large data sets," Chemometrics & Intelligent Laboratory System, vol. 8, pp 195-204, 1990.

[5]   K. Teknomo. (2009, August 8). Hierarchical clustering tutorial.Retrieved from http://people.revoledu.com/kardi/tutorial/clustering/index.html [Retrieved August 8, 2009]

[6]   G.N. Lance, and W.T. William,"A general theory of classifacatory sorting strategies," Computer Journal, vol. 6, pp 373-380, 1967.

[7]   E.L. Lehmann, Nonparametric Statistical Methods Based On Ranks, New York: Springer.

[8]   C.Y. Piaw, Asas Statistik Penyelidikan: Analisis Data Skala Ordinal Dan Skala Nominal, Malaysia: McGraw Hill.