Middle-East Journal of Scientific Research 13 (Mathematical Applications in Engineering): 116-123, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.13.mae.10003

Profile Analysis of Population Based on Age-Group in Sumatra

[1]Mustofa Usman, [2]Yusuf S. Barusman, [3]Faiz A.M. Elfaki and [3]Jamal I. Daoud

[1]Department of Mathematics, College of Sciences and Mathematics, Lampung University, Indonesia
[2]Department of Management, Faculty of Economic, University Bandar Lampung, Indonesia
[3]Department of Science, College of Engineering, IIUM, Malaysia

Abstract: The aims of this study is going to analyze the profile of the population in Sumatra based on the age-groups of the population. The Sumatra Island was divided into ten provinces. The profile analysis is going to use to compare the profile of the proportion of age-groups population among the ten provinces. The test of parallel for the profiles, some provinces are significantly different. The variations of the age-goup1 to age-group8 are more variation than the age-group9 to age-group16. The multiple comparison by Tukey's LSD methods also shows that there are significant difference of proportion of age-groups for some provinces.

Key words: Population . age-groups . profile analysis . Tukey's LSD method

INTRODUCTION

Profiles of a population are very important information to know the characteristic of a population in certain provinces or a country. The information can be used in many ways for government planning in the future. The profile analysis give a clear information and easy to use. Some studies of a profile analysis of a population in India [2] discussed the age and sex profile in India. In 2005, 36% of the population was below age 14 and 4% was age 65% or older. While in China the proportion of age 60 or older increase from 7% 1953 to more then 10% in 2000 [6].

In this study we will try to explore some characteristics of the population in Sumatra based on the age-groups in each district and province. In Sumatra there are 10 provinces with total 153 districts and age groups divided into 16 groups. The study will discuss and compares the profiles of age-groups of the ten provinces. The method of analysis to be used is profile analysis as was suggested and can be found such as in [3-5, 8, 10]. Profile analysis is a collection of statistical hypothesis testing procedures used to explore any possible similarities among the treatment effects. Profile analysis is relevant especially to the longitudinal data on a given response variable or in the situation where responses on several dependent variables are measured on the same experimental unit. A population profile is a plot of the components of the population mean vector versus the order in which the means are arranged [4]. In this study the groups will be the provinces (10 provinces), the unit will be the districts and the treatments will be the age-groups within the districts (there are 16 age-groups).

PROCEDURE AND TESTING THE PROFILE ANALYSIS

Ley Y_{ijk}, i=1,2,…, g; j=1,2,…, n_i ; k=1,2, …, p, be an observation (a response) in a repeated measures experiment, where i, j and k stand for province, districts and age-groups, respectively. And g=10 provinces, n_i is total number of districts in i-th province and p= 16 age-groups. Selected subject in the i-th province and j-th district and k-th age-group and its mean response are:

$$Y_{ijk} = [y_{ij1}, y_{ij2}, \ldots, y_{ijp}] \quad (1)$$

denotes the response vector for the j-th district within the i-th province and

$$\bar{Y}_{ijk} = [\bar{y}_{i1}, \bar{y}_{i2}, \ldots, \bar{y}_{ip}] \quad (2)$$

denotes the mean response vector for the i-th province group. Multivariate analysis of variance (MANOVA) is the natural choice for analyzing the type of data in the one way completely randomized design. Profile analysis (Repeated Measure MANOVA) suggested by many authors [4, 5, 8, 9] is most appropriate. MANOVA with repeated measures is used when the measure that is reapeted (e.g. across time) is a compound formed in the usual MANOVA fashion

Corresponding Author: Mustofa Usman, Department of Mathematics, Faculty of Sciences and Mathematics, Lampung University, Indonesia. Jln. Prof. Sumantribrojonegoro No.1, Lampung, Indonesia

across multiple dependent variables. Thus, it is different from having multiple repeated measures factors [9].

The hypothesis of equality of means response vectors sought in MANOVA which is given by

$$Ho: \mu_1 = \mu_2 = \ldots. = \mu_g \quad (3)$$

where $\mu_i = (\mu_{i1}, \mu_{i2}, \ldots., \mu_{ip})^T$, implies that the treatment have the same average effects. Though useful, mere acceptance or rejection of such a hypothesis does not provide adequate insight into the type of similarities and dissimilarities that may exist among the treatments. In order to gain more understanding, we can formulate the above hypothesis as three hypotheses to be tested sequentially and subjected to the acceptance of the hypothesis at the previous stage. Specifically, we can ask: Are the profile parallel? If so, are they coincidental? And finally, if so, are they all horizontal? [4].

The null hypothesis

Ho1: The g provinces profile parallel is equivalent to the null hypothesis

$$Ho1: \begin{bmatrix} \mu_{11} - \mu_{21} \\ \mu_{11} - \mu_{31} \\ \vdots \\ \mu_{11} - \mu_{g1} \end{bmatrix} = \begin{bmatrix} \mu_{12} - \mu_{22} \\ \mu_{12} - \mu_{32} \\ \vdots \\ \mu_{12} - \mu_{g2} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{1p} - \mu_{2p} \\ \mu_{1p} - \mu_{3p} \\ \vdots \\ \mu_{1p} - \mu_{gp} \end{bmatrix} \quad (4)$$

or

$$Ho1 : L_1 BM_1 = 0 \quad (5)$$

Where

$$L_1 = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ 1 & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 \end{bmatrix}_{(g-1)xg}$$

$$B = \begin{bmatrix} \mu_1^T \\ \mu_2^T \\ \vdots \\ \vdots \\ \mu_g^T \end{bmatrix}_{gxp} \quad \text{and } M_1 = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ -1 & 0 & \ldots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 \end{bmatrix}_{px(p-1)}$$

Thus, the hypothesis of parallel profiles can be formulated as a general linear hypothesis. The general linear hypothesis given in (5) is tested using the appropriate multivariate test. There are several different multivariate test statistics available for the test of parallel profile and generally they give equivalent results. Four common tests statistics are: Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace and Roy's Greatest Root [4]. Wilks' Lambda (Λ) is the most desirable because it can be converted exactly to an F-statistics [5, 8]. Reject Ho1 at α level if the ratio of generalized variance (Λ) is to small or if F-statistics is greater than the critical value.

If the hypothesis of parallel profile is not rejected, then we can test the second hypothesis: Are the profiles coincidental, given that profiles are parallel? The null hypothesis is

Ho2: The profiles coincidental, given that profiles are parallel.

or

$$Ho2 : L_2 BM_2 = 0 \quad (6)$$

Where

$$L_2 = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ 1 & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 \end{bmatrix}_{(g-1)xg} \quad \text{and } M_2 = 1_p$$

To test Ho2, also we use tests statistics Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace and Roy's Greatest Root (Khattree and Naik, 2005). Reject Ho2 at α level if the ratio of generalized variance (Λ) is to small or if F-statistics is greater than the critical value. But, if the hypothesis of parallel profile (Ho1) is rejected, then the null hypothesis (Ho2) of coincidental profiles has no meaning. [11] suggested when the parallel hypothesis is rejected, it may be best to analyze each age-groups separately.

If the hypothesis of coincidental is not rejected, then we can test the third hypothesis: Are the profiles horizontal? The null hypothesis is

Ho3: The profiles are horizontal.

or

$$Ho3: L_3BM_3 = 0 \quad (7)$$

Where

$$L_3 = (1\ 0\ 0\ \ldots\ 0)$$

and

$$M_3 = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ -1 & 0 & \ldots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 \end{bmatrix}_{px(p-1)}$$

To test Ho3, also we use tests statistics Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace and Roy's Greatest Root [4]. Reject Ho3 at α level if the ratio of generalized variance (Λ) is to small or if F-statistics is greater than the critical value.

DATA ANALYSIS AND RESULTS

The necessary data for this study were obtained from Indonesia Bureau of Statistics, based on data census of 2010. The data was taken from 10 Provinces in Sumatra, namely: Aceh (23 districts), Sumatra Utara, Medan, (33 districts), Padang (19 districts), Riau(12 districts), Jambi(12 districts), Bengkulu (10 districts), Lampung (16 districts), Bangka Belitung (7 districts)

and Kepri (7 districts). The population age-groups was divided into 16 groups, Age-group1(0-4 years-old), Age-group2 (5-9 years-old), Age-group3 (10-14years-old), Age-group4 (15-19years-old), Age-group5 (20-24years-old), Age- group6 (25-29 years-old), Age-group7(30-34 years-old), Age-group8(35-39 years-old), Age-group9 (40-44 years-old), Age-group10 (45-49years-old), Age-group11(50-54 years-old), Age-group12 (55-59 years-old), Age-group13 (60-64years-old), Age-group14(65-69 years-old), Age-group15 (70-74 years-old), Age-group16(75 years or older).

From box plot in each age group we can see some behavior of data as follow: the variation of population proportion in Aceh, Medan, Padang and Riau for age group1 is higher than the other provinces (Fig. 1(a)); the variation of population proportion in Aceh and Medan for age group2 and age group10 is higher than the other provinces (Fig. 1(b) and 1(j)); the variation of population proportion in Medan for age group9 is higher than the other provinces (Fig. 1(i)); the variation

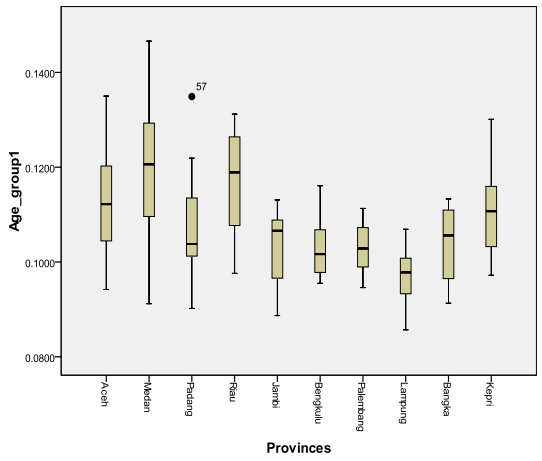

Fig. 1(a): Box plot age-group 1

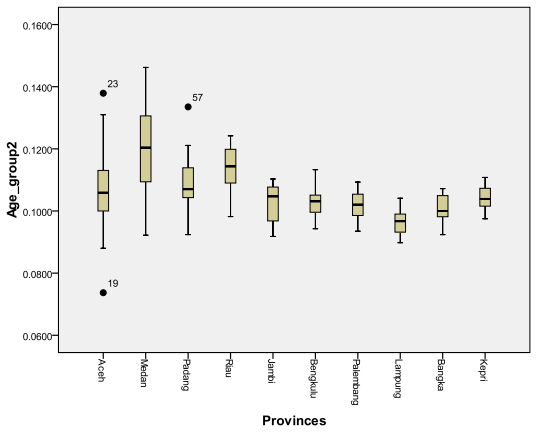

Fig. 1(b): Box plot age-group 2

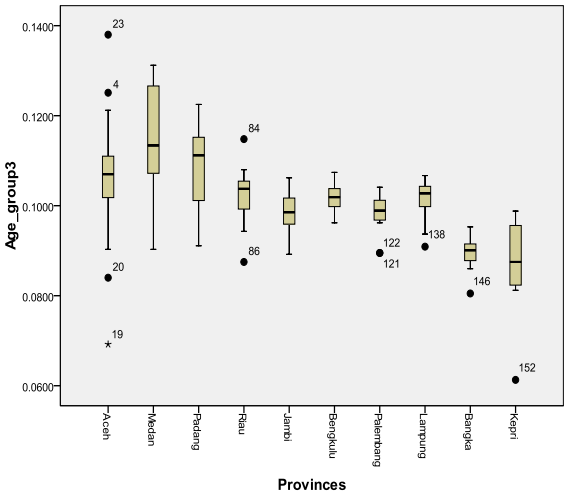

Fig. 1(c): Box plot age-group 3

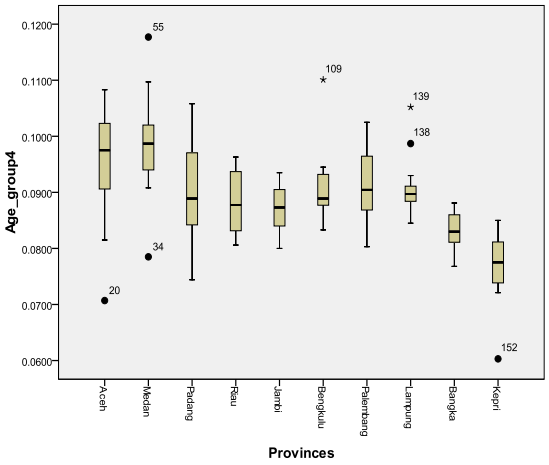

Fig. 1(d): Box plot age-group 4

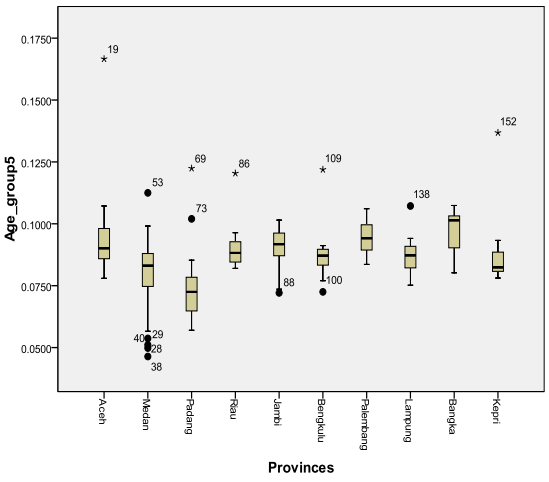

Fig. 1(e): Box plot age-group 5

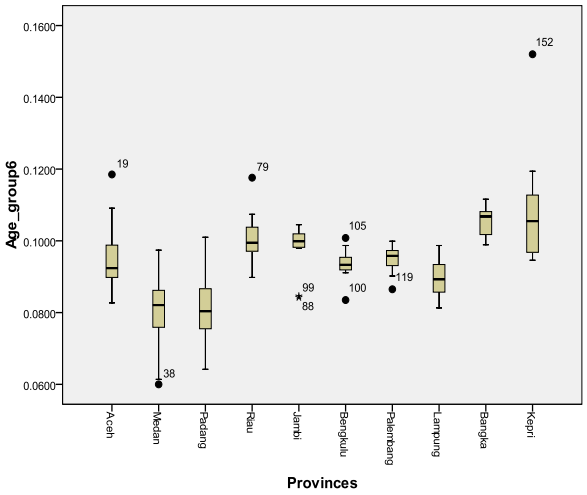

Fig. 1(f): Box plot age-group 6

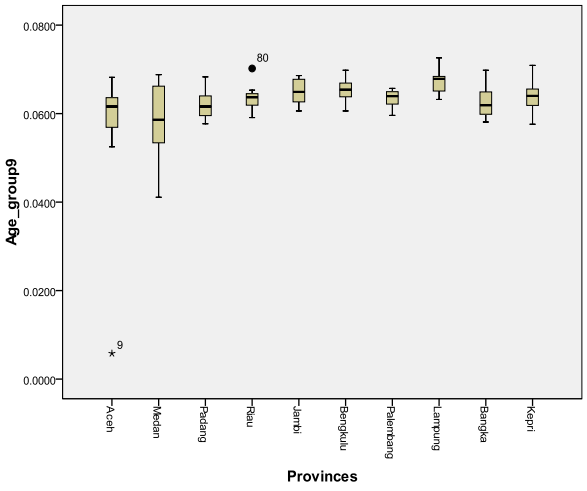

Fig. 1(i): Box plot age-group 9

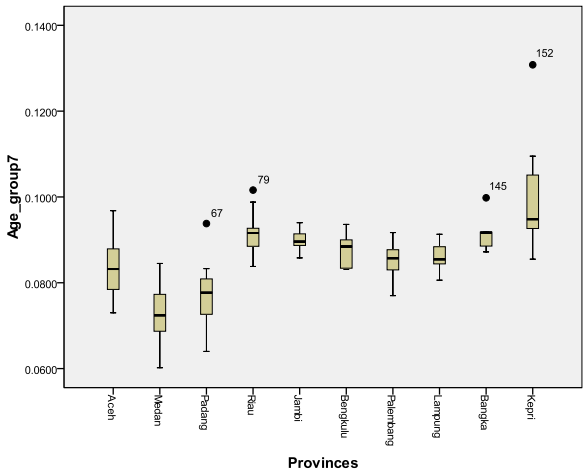

Fig. 1(g): Box plot age-group 7

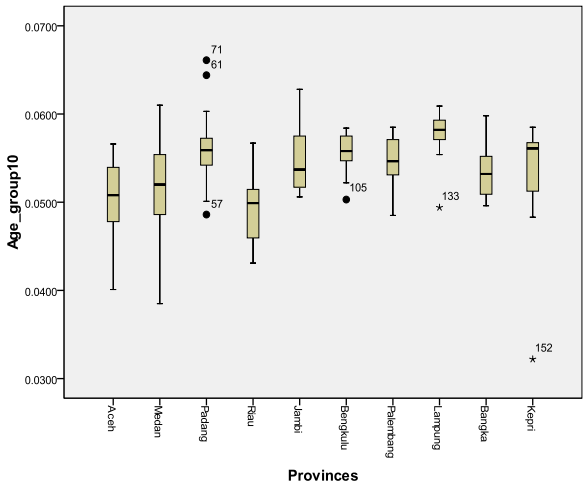

Fig. 1(j): Box plot age-group 10

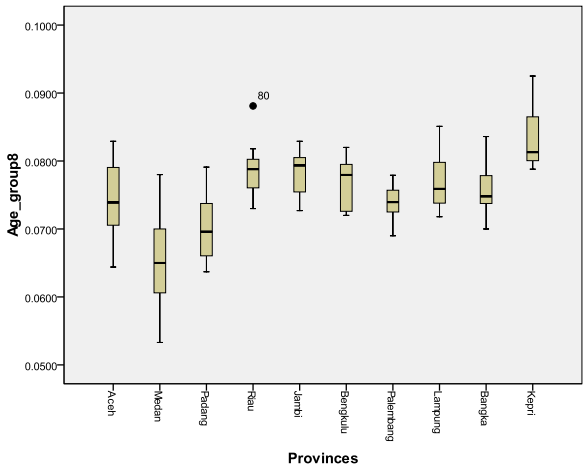

Fig. 1(h): Box plot age-group 8

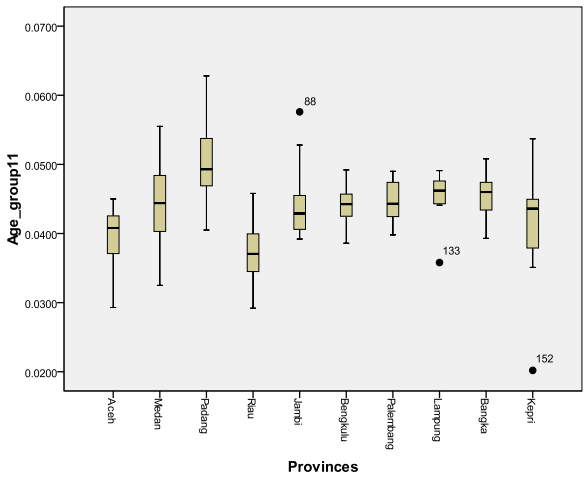

Fig. 1(k): Box plot age-group 11

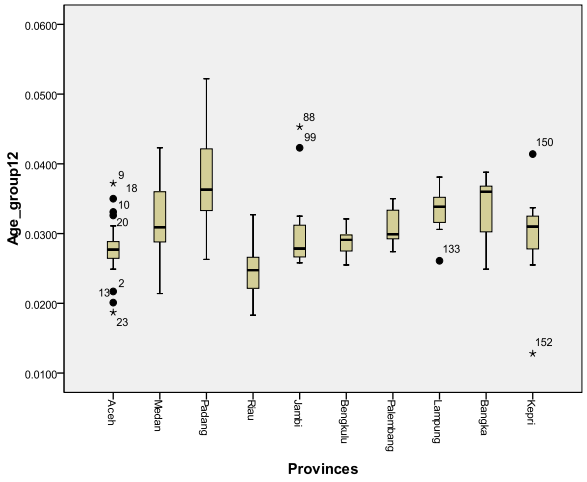

Fig. 1(l): Box plot age-group 12

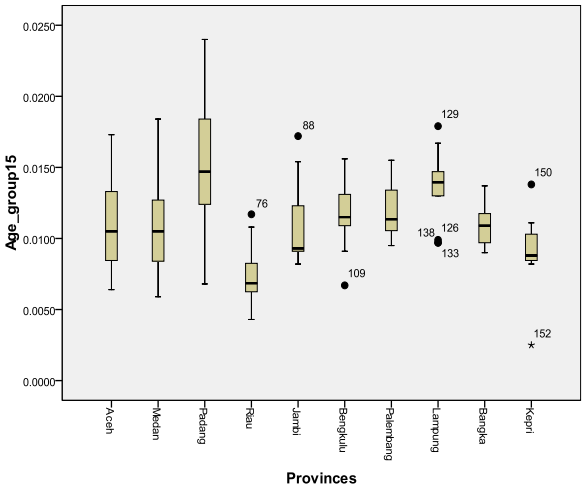

Fig. 1(o): Box plot age-group 15

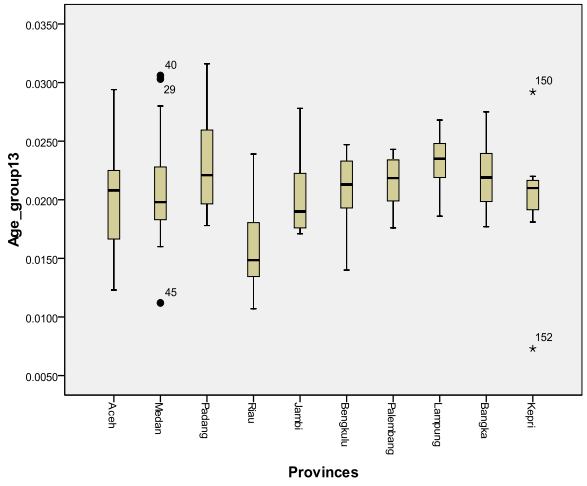

Fig. 1(m): Box plot age-group 13

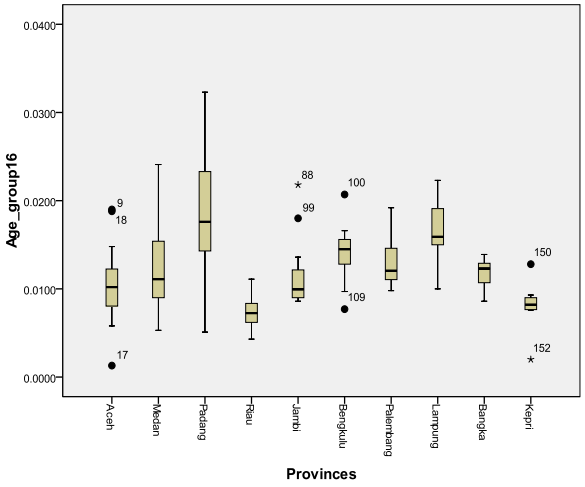

Fig. 1(p): Box plot age-group 16

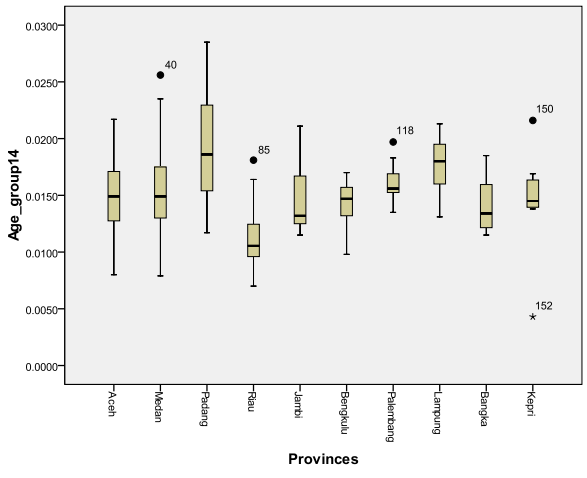

Fig. 1(n): Box plot age-group 14

of population proportion in Aceh, Medan and Padang for age group3, age group4, age group 6, age group8, age group11, age group 12, age group 13, age group 14, age group 15 age group 16 are higher than the other provinces (Fig. 1(c), 1(d), 1(f), 1(h), 1(k), 1(l), 1(m), 1(n), 1(o) and 1(p)).

Test for compound symmetry is not satisfied, multivariate procedure should be employed [11]. From the result of sphericity test by using Box's M statistics we have M test equal to 807.48 ($p < 0.000$). It can be said that variance-covariance matrix is not homogeneous. Therefore we used MANOVA repeated measure as suggested by [11]. Test based on the manova approach are free of sphericity assumption [7].

A plot for profile of age groups of population proportion for the 10 provinces given in Fig. 2. The figure shows that the profile even tough looks have a

Table 1: SAS output for testing profile parallel

			N = 63.5		
	S = 9	M = 2.5			
Statistics	value F	value	Num DF	Den DF	Pr>F
Wilks' Lambda	0.0096	6.16	135	1018.90	<0.0001
Pillai's Trace	3.1256	4.86	135	1233.00	<0.0001
Hotelling Lawley trace	7.8121	7.37	135	665.84	<0.0001
Roy's Greatest Root	3.0095	27.49	15	137.00	<0.0001

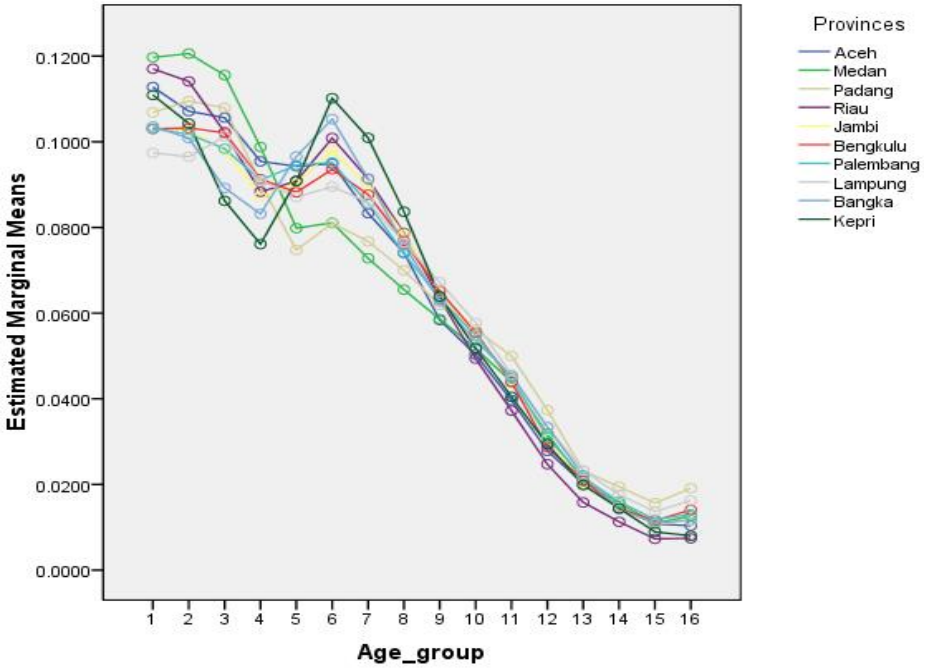

Fig. 2: The profile of age-groups proportion in 10 provinces in Sumatra

similar trend, but they look no parallel. The test statistics also for testing the hypothesis Ho1: The g provinces profile parallel,

from the SAS print out in Tabel 1 Wilks' Lambda (Λ) yields Λ=0.0096 or F= 6.16 with p-value <0.0001. Therefore we reject the null hypothesis of profile parallel. The other three multivariate tests are also in agreement with this conclusion (Table 1).

From Fig. 2, that the population proportion from age group 1 to age group 8 (or from 0 to 39 years-old) shows more variations compare to age group 9 to age group 16 (or from 40 to 75+ years old).

The result of analysis for testing the profile parallel was rejected (not parallel) this caused the other two hypothesis given in Equation (6) and (7) contingent on the tenability of the hypothesis profile parallel given in Equation (5), not to be tested. In view of this rejection, the null hypothesis of coincidental profile and the null hypothesis that the profile are horizontal has no meaning here [4].

From the multiple comparison test for each age-groups by using Tukey's LSD with alpha = 5%, the result given in Tabel 2. The result of multiple comparison for each group indicating significant different at alpha=5%. The maximum and minimum means proportion in each age group are as follow: In age-group1 (0-4 years-old) the highest proportion population is Medan 11.97% and the smallest proportion population is Lampung 9.74%, In age-group2 (5-9 years-old) the highest proportion population is Medan 12.06% and the smallest proportion population is Lampung 9.65%, In age-group3 (10-14 years-old) the highest proportion population is Medan 11.55% and the smallest proportion population is Kepri 8.62%, In age-group4 (15-19 years-old) the highest proportion population is Medan 9.87% and the smallest proportion population is Kepri 7.61%, In age-group5 (20-24 years-old) the highest proportion population is Bangka 9.66% and the smallest proportion population is Padang 7.47%, In

Tabel 2: Mean of each provinces and Tukey's LSD for each Age_groups (AG) (alpha = 5%)

Provinces	AG1	AG2	AG3	AG4	AG5	AG6	AG7	AG8
Aceh	0.1127bcd	0.1071ab	0.1056 cd	0.0954cd	0.0943b	0.0948bcd	0.0833bc	0.0741bc
Medan	0.1197d	0.1206d	0.1155d	0.0987d	0.0798ab	0.0810a	0.0728a	0.0654a
Padang	0.1068abcd	0.1095bcd	0.1079cd	0.0903bcd	0.0747a	0.0809a	0.0767ab	0.0699ab
Riau	0.1170cd	0.1141cd	0.1022c	0.0883bc	0.0908ab	0.1008cde	0.0913c	0.0787cd
Jambi	0.1031ab	0.1026abc	0.0980abc	0.0873bc	0.0901ab	0.0980bcd	0.0897c	0.0784cd
Bengkulu	0.1030ab	0.1032abc	0.1021c	0.0912bcd	0.0882ab	0.0935bc	0.0876c	0.0768c
Palembang	0.1030ab	0.1018abc	0.0983abc	0.0910bcd	0.0945b	0.0951bcd	0.0856c	0.0739bc
Lampung	0.0974a	0.0965a	0.1012bc	0.0909bcd	0.0870ab	0.0895ab	0.0858c	0.0765c
Bangka	0.1035abc	0.1008ab	0.0892ab	0.0831ab	0.0965b	0.1053de	0.0913c	0.0759de
Kepri	0.1109abcd	0.1042abc	0.0862a	0.0761a	0.0908ab	0.1101e	0.1009d	0.0836e
Means	0.1095	0.1083	0.1042	0.0917	0.0871	0.0915	0.0833	0.0733
Provinces	AG9	AG10	AG11	AG12	AG13	AG14	AG15	AG16
Aceh	0.0583a	0.0503ab	0.0396ab	0.0278ab	0.0201ab	0.0148abc	0.0107abc	0.0103ab
Medan	0.0586a	0.0517ab	0.0439abc	0.0318bc	0.0207ab	0.0156abc	0.0109abc	0.0126abc
Padang	0.0619ab	0.0561bc	0.0499c	0.0373c	0.0231b	0.0194c	0.0156d	0.0191d
Riau	0.0634ab	0.0493a	0.0372a	0.0247a	0.0158a	0.0112a	0.0073a	0.0074a
Jambi	0.0649ab	0.0547abc	0.0444bc	0.0306ab	0.0204ab	0.0147ab	0.0108abc	0.0115abc
Bengkulu	0.0651ab	0.0554bc	0.0439abc	0.0287ab	0.0209ab	0.0144ab	0.0115bcd	0.0140bcd
Palembang	0.0634ab	0.0547abc	0.0448bc	0.0310abc	0.0215b	0.0160bc	0.0118bcd	0.0130abc
Lampung	0.0672b	0.0577c	0.0456bc	0.0333bc	0.0232b	0.0176bc	0.0136cd	0.0162cd
Bangka	0.0627ab	0.0535abc	0.0453bc	0.0334bc	0.0221b	0.0142ab	0.0109abc	0.0117abc
Kepri	0.0639ab	0.0518abc	0.0404ab	0.0294ab	0.0198ab	0.0144ab	0.0089ab	0.0080a
Means	0.062	0.0533	0.0437	0.031	0.0209	0.0156	0.0115	0.0128

Note: Mean with the same letter not significantly different at alpha = 5%

age-group6 (25-29 years-old) the highest proportion population is Kepri 11.01% and the smallest proportion population is Padang 8.09%, In age-group7 (30-34 years-old) the highest proportion population is Kepri 10.09% and the smallest proportion population is Medan 7.28%, In age-group8 (35-39 years-old) the highest proportion population is Kepri 8.36% and the smallest proportion population is Medan 6.55%, In age-group9 (40-44 years-old) the highest proportion population is Lampung 6.72% and the smallest proportion population is Aceh 5.83%, In age-group10 (45-49 years-old) the highest proportion population is Lampung 5.77% and the smallest proportion population is Riau 4.94%, In age-group11 (50-54 years-old) the highest proportion population is Padang 4.99% and the smallest proportion population is Riau 3.72%, In age-group12 (55-59 years-old) the highest proportion population is Padang 3.73% and the smallest proportion population is Riau 2.47%, In age-group13 (60-64 years-old) the highest proportion population is Lampung 2.32% and the smallest proportion population is Riau 1.58%, In age-group14 (65-69 years-old) the highest proportion population is Padang 1.95% and the smallest proportion population is Riau 1.12%, In age-group15 (70-74 years-old) the highest proportion population is Padang 1.56% and the smallest proportion population is Riau 0.73% and In age-group16 (75+ years-old) the highest proportion population is Padang 1.91% and the smallest proportion population is Riau 0.74%.

REFERENCES

1. Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis. 2nd Edn. New York: John Wiley & Sons, Inc.
2. Carl Haub and O.P. Sharma, 2006. India Population Reality: Reconciling Change and Tradition. Population Bulletin, Vol: 61 (3).
3. Johnson, R.A. and D.W. Wichern, 2002. Applied Multivariate Statistical Analysis. 5th Edn. New Jersey: Prentice-Hall, Inc.
4. Khattree, R. and D.N. Naik, 2005. Applied Multivariate Statistics with SAS Software. 2nd edd. North Carolina: SAS Institute, Inc.
5. Morrison, D.F., 1976. Multivariate Statistical Methods. 2nd Edn. New York: McGraw-Hill Book Company.

6. Nancy, E. Riley, 2004. China's Population: New Trends and Challenges. Population Bulletin, Vol: 59 (3).
7. O'Brien, R.G. and M.K. Kaiser, 1985. MANOVA method for Analyzing Repeated Measures Designs: An Extensive Primer. Psychological Bulletin, Vol: 97 (2).
8. Srivastava, M.S., 2002. Methods of Multivariate Statistics. New York: John Wiley & Sons. Inc.
9. Tabacknick, B.G. and L.S. Fidel, 2007. Using Multivariate Statistics. 5th Edn. New Jersey: Prentice-Hall. Inc.
10. Timm, N.H., 2002. Applied Multivariate Analysis. New York: Springer-Verlag.
11. Timm, N.H., 1975. Multivariate Analysis with Application in Education and Psychology. California:Brooks/Cole.