ที่ ศธ 0520.204/วสถ 116/2550

ภาควิชาสถิติ คณะวิทยาศาสตร์
มหาวิทยาลัยศิลปากร
วิทยาเขตพระราชวังสนามจันทร์
นครปฐม 73000

9 เมษายน 2550

เรื่อง ตอบรับบทความเข้าร่วมประชุมวิชาการสถิติและสถิติประยุกต์ประจำปี 2550

เรียน คุณนันท์นภัส พรุเพชรแก้ว

คณะกรรมการฝ่ายอำนวยการจัดงานและประสานงาน การจัดการประชุมวิชาการสถิติและสถิติประยุกต์ประจำปี 2550 โดย ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร ร่วมกับเครือข่ายการวิจัยสถิติศาสตร์และสมาคมสถิติแห่งประเทศไทย ระหว่างวันที่ 24-25 พฤษภาคม 2550 ณ โรงแรมโนโวเทลทิพย์วิมานรีสอร์ท แอนด์สปา อ.ชะอำ จ.เพชรบุรี มีความยินดีที่จะแจ้งให้ทราบว่าบทความเรื่องต่อไปนี้ได้รับการพิจารณาว่าเหมาะสมที่จะนำเสนอในการประชุมวิชาการฯ ดังกล่าว

"Logistic Regression for Modeling Life Tables for Southern Thailand"
โดย Nannapat Pruphetkaew, Phattrawan Tongkumchum and Chamnein Choonpradub

จึงเรียนมาเพื่อทราบและดำเนินการต่อไปโดยเฉพาะการจองที่พักและการลงทะเบียนเข้าร่วมประชุมฯ คณะกรรมการฯ ขอขอบพระคุณที่ท่านให้สนใจส่งบทความเข้าร่วมในการประชุมวิชาการ และหวังว่าจะได้รับความร่วมมือจากท่านอีกในโอกาสต่อไป

ขอแสดงความนับถือ

กมลชนก พานิชการ

(ผู้ช่วยศาสตราจารย์ ดร.กมลชนก พานิชการ)
หัวหน้าภาควิชาสถิติ
ประธานกรรมการการประชุมวิชาการสถิติและสถิติประยุกต์ประจำปี 2550

สำนักงานภาควิชาสถิติ
โทรศัพท์ 034-272924
โทรสาร 034-252275
http://stat.sc.su.ac.th

การประชุมวิชาการสถิติและสถิติประยุกต์ ประจำปี 2550

35 ปี คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

วันที่ 24 - 25 พฤษภาคม 2550
ณ โนโวเทลทิพย์วิมานรีสอร์ทแอนด์สปา
ชะอำ จังหวัดเพชรบุรี

เนื่องในโอกาสครบรอบ 35 ปีคณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร
และเครือข่ายการวิจัยสถิติศาสตร์
ร่วมกับ
สมาคมสถิติแห่งประเทศไทย

ISBN 978-974-641-159-2

LOGISTIC REGRESSION FOR MODELING LIFE TABLES FOR SOUTHERN THAILAND

Nannapat Pruphetkaew
Phattrawan Tongkumchum
Chamnein Choonpradub
Department of Mathematics and Computer Science,
Prince of Songkla University, Pattani

Abstract

The objective of this study was to construct model life tables for the 14 provinces in Southern Thailand. The data comprise the number of deaths in the year 2000 and the mid-year corresponding populations at risk classified by gender, 5-year age group and province. We first estimated age-specific death rates for males and females in each province. Logistic regression was then used for modeling a set of life tables. It was found that six provinces (Narathiwat, Yala, Pattani, Satun, Krabi and Phuket) could be fitted reasonably well with a common model for males and for females, and the remaining eight could be grouped together in this way giving a different model. The main difference between these models (labeled "south" and "north" respectively) is that for each gender the "south" model is smoother than that of the corresponding "north" model. The "north" curves show peaks at age 25 years whereas the "south" curves increase monotonically.

Key words model life table, mortality, logistic regression, southern Thailand

1. Introduction

Population forecasting in Thailand is hampered by the unavailability of accurate mortality data. Although population data are collected from individual households every ten years, in contrast to other countries in South East Asia the data collection form does not enquire about recent mortality. Moreover, statistics based on death certificates are known to undercount mortality, possibly by as much as 15% (Prasartkul and Vapattanawong 2006).

The 2000 Population and Housing Census of Thailand (National Statistical Office 2002) revealed substantial differences in growth rates from 1990 to 2000 between the 14 provinces in the Southern Region. For example, Phuket Province grew by 50% during this decade, whereas the increase in Nakhon Sri Thammarat was just 9%. There were also substantial differences in growth rates between the Muslim and non-Muslim populations in the southern provinces. For example, the Muslim population in Pattani province (78% in 1990) increased by 19% in the decade from 1990 to 2000 while at the same time the non-Muslim population increased by only 3%, whereas the Muslim population of Krabi (36% in 1990) also increased by 19% during the decade but the non-Muslim population increased by 23%.

In this paper our objective is to construct model life tables for the 14 provinces in Southern Thailand, based on mortality statistics by gender, 5-year age group and province provided by the Ministry of Public Health's Bureau of Policy and Strategy (2002).

2. Materials and Methods

The basic data comprise the numbers of deaths in the year 2000 and the corresponding (mid-year) populations at risk classified by gender, 5-year age group and province.

Life Table

The method for constructing a life table l_x for x in (0, 5, ..., 85) by gender and province (Pollard et al. 1974) is described as follows.

Denote the number of deaths and the population at risk in age group (x, x+5) by D_x and P_x, respectively. The age-specific death rate is $M_x = D_x/P_x$. The probability of dying between ages x and x+5 is $q_x = 5M_x/(1+5M_x/2)$ for x < 85 and $q_{85} = 1$. Now define $l_0 = 100{,}000$ and

$$l_{x+5} = (1 - q_x)\, l_x \qquad (1)$$

for each value of x.

Logistic Regression

Logistic regression provides an appropriate statistical method for modelling a set of life tables. Since males and females have essentially different life tables (Intachat et al 2005) we fit separate models for the two sexes. In this method, the outcome is the binary event denoting the death or survival of a male or female at risk in a specific demographic group indexed by 5-year age group and province. The risk of death M_{xj} to such a person in age group (x, x+5), and province j is defined in terms of its logit as

$$ln\{M_{xj}/(1-M_{xj})\} = a_x + b_j\,, \qquad (2)$$

where a_x is an age effect and b_j is a province effect. To avoid overparametrisation we can force the province effects to have zero mean, i.e., $\sum b_j = 0$.

The model life table for province j is now obtained by substituting the values of M_{xj} given by Equation (2) into Equation (1).

Asymptotic results using statistical theory provide estimates based on maximum likelihood fitting of the model, together with confidence intervals and p-values for testing relevant null hypotheses (Kleinbaum and Klein 2002).

Goodness-of-fit of model

For each cell corresponding to a combination of nominal determinants, the Pearson residual is defined as

$$z = \frac{M - \hat{M}}{\sqrt{\hat{M}(1-\hat{M})/n}}, \qquad (3)$$

where M is the proportion of outcomes observed in the cell, $\hat{M}$ is the corresponding probability given by the model, and n is the total number of cases in the cell. The goodness-of-fit of the model can be assessed visually by plotting these z-values against corresponding normal scores. The fit is adequate if the points in this plot are close to a straight line with unit slope. A p-value for the goodness-of-fit is obtained by subtracting the deviance associated with the saturated model from the model deviance and comparing this difference R_g with a chi-squared distribution having degrees of freedom equal to n_g❑□ m, where n_g is the number of cells and m is the number of parameters in the model.

3. Preliminary Analysis

Table 1 gives the male and female life tables obtained by applying the method described in Section 2 to the aggregated data for the year 2000 from all 14 provinces in the Southern Region of Thailand. Life table for "the north" is also given in the table.

	males				females			
x	D	P	l	l (north)	D	P	l	l (north)
0	766	364435	100000	100000	599	343532	100000	100000
5	221	394376	98955	99165	146	372369	99132	99282
10	217	397676	98678	98919	133	380638	98938	99103
15	539	391618	98409	98681	198	382400	98765	98960
20	805	349995	97734	98008	295	342789	98510	98735
25	1466	338153	96616	96804	486	354058	98087	98302
30	1579	333236	94545	94589	486	349292	97416	97572
35	1291	315919	92331	92279	509	327193	96741	96881
40	1013	267023	90463	90423	426	278914	95991	96173
45	1037	211532	88764	88799	504	220386	95261	95500
50	949	160092	86614	86610	569	168551	94178	94582
55	1196	135046	84084	84191	597	139516	92601	93139
60	1582	121082	80442	80722	955	133395	90641	91346
65	1911	91158	75353	75845	1290	102468	87453	88599
70	2086	67854	67848	69032	1588	78388	82117	84226
75	1878	37569	58163	60274	1609	45651	74200	77726
80	1660	20885	45241	48039	1805	28992	62183	66795
85	1985	15494	30242	33106	3082	25799	45433	50568

Table 1: Life tables for Southern Thailand and the north in 2000.

4. Model Life Tables

We first fitted the logistic regression model described by Equation (1) to the data from all 14 provinces in the Southern Region. The number of parameters in this model (m) is 31 corresponding to the constant plus 17 age group parameters and 13 province parameters and the number of cells (n_g) is 252 corresponding to the product of 18 age groups and 14 provinces, so the number of degrees of freedom for assessing the goodness-of-fit of the model is 221. The residual deviances based on these grouped data are 553.78 for the males and 474.00 for the females, indicating a poor fit in each case.

Next we looked for more homogeneous subgroups of provinces, and found that six provinces (the four southernmost – Narathiwat, Yala, Pattani and Satun – together with Krabi and Phuket) could be fitted reasonably well with common models for males and for females, and the remaining eight could also be grouped together in this way. We label these models as "south" and "north", respectively.

Figure 1 shows plots of the mortality curves based on the four fitted model life tables. As expected, the male mortality is higher than that for females at all ages (although the curves must converge at age 85 because the mortality at this age encompasses all higher ages). The main difference between the "north" and "south" mortality curves is that for each gender the curve for the "south" model is smoother than that for the corresponding "north" model. The "north" curves show peaks at age 25 years whereas the "south" curves increase monotonically.

5. Conclusion and Discussion

It should be noted that our mortality analysis is done by province. From this study we found that subgroups of six provinces can be fitted with common models for males and females, and the remaining eight can also be grouped together with a different model. The main difference between these models (labeled "south" and "north" respectively) is that the "south" model is smoother than that of the "north" model. The "north" curves show peaks at age 25 years whereas the "south" curves increase monotonically. It refers to the place where death occurs, which was not necessarily the place of residence. Our death rates of males and females were 5.5 and 3.7 respectively, whereas the corresponding rates for the whole country were 7.0 and 4.9 (Rukumnnuaykit 2006). This could be due to hospital deaths in other major cities around the country including Bangkok metropolis. Although this study was limited to the quality of death data, the finding from this study suggested some ideas on the pattern of death rate among micro level.

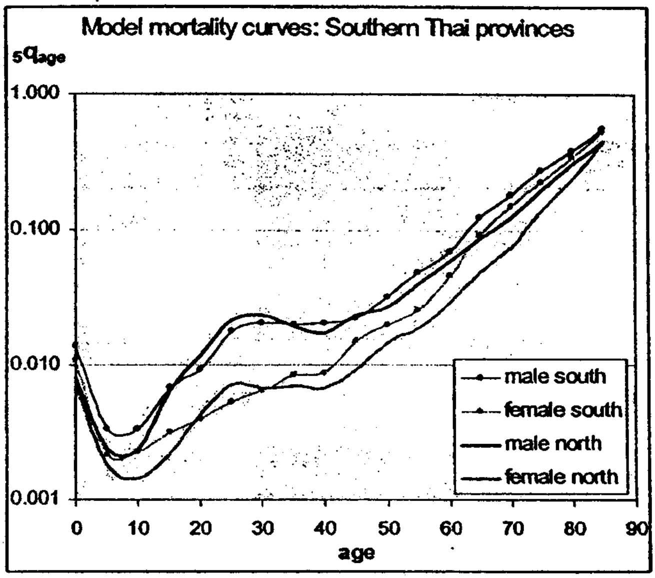

Figure 1: Age-specific mortality curves based on fitted logistic models

Acknowledgement

We would like to thank Prof. Don McNeil from Macquarie University, Australia for his suggestion.

References

Bureau of Policy and Strategy. Thailand Health Profile 1999-2000. Express Transportation Organization, Ministry of Public Health, Thailand, 2002.

Intachat, N., Sawangdee, Y., Entwisle, B. and Podhisita, C. "Cause of Death in Thailand: Gender Differential Perspective." *J Pop & Soc Stud* 13(2):47-69, 2005.

Kleinbaum, D.G. and Klein, M. Logistic Regression: A Self-Learning Text (2nd ed) Springer-Verlag. New York, 2002.

National Statistical Office. The 2000 Population and Housing Census, Southern Region. (http://webhost.nso.go.th:9999/nso/project/search/index.jsp)

Pollard AH, Yusuf F and Pollard GN. Demographic Techniques. Pergamon Press: Australia, 1974.

Prasartkul, P. and Vapattanawong, P. The completeness of death registration in Thailand: Evidence from demographic surveillance system of the Kanchanaburi project. World Health & Population 2006. (http://www.longwoods.com/product.php?productid=18054&cat=413&page=2)

Rukumnuaykit, P. "Mortality and Causes of Death in Thailand: Evidence From the Survey of Population Change and Death Registration." *Asia-Pacific Pop J* 21(2):67-84, 2006.