A Study on the Association of Income with Mortality and Life Expectancy in Small Area

19th International Longevity Risk and Capital Markets Solutions Conference

September 16, 2024

Hsin-Chung Wang, Aletheia University

Jack C. Yue, National Chengchi University

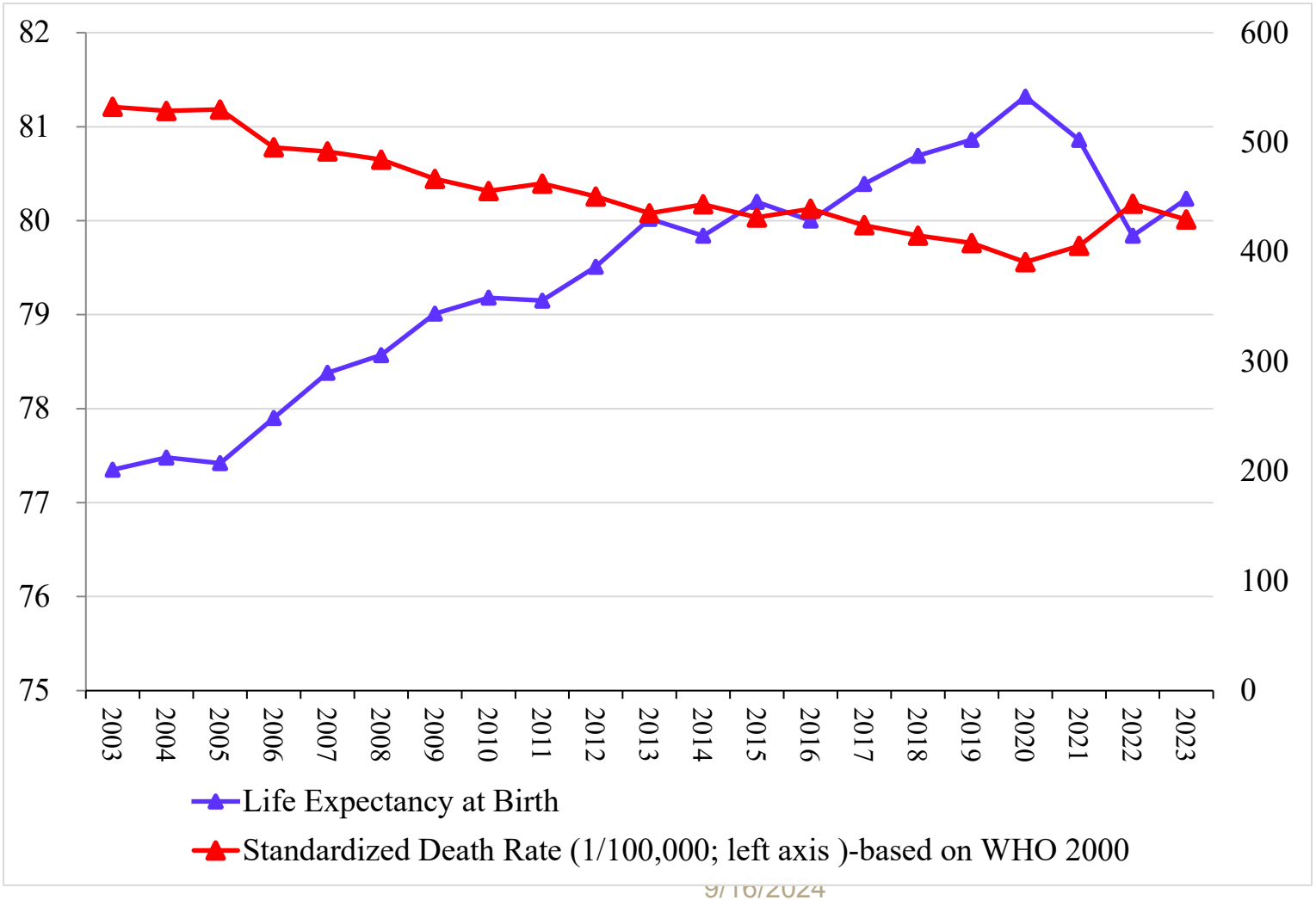
Outline

- Motivation
- Methodology
- Exploratory Data Analysis (EDA)
- Life Expectancy Estimate

1. Multiple regression analysis
2. Traditional life table construction method
3. Haybittle (1998) Method

- Clustering (22 Counties and 368 Townships)
- Reference Population Choice
- Conclusion

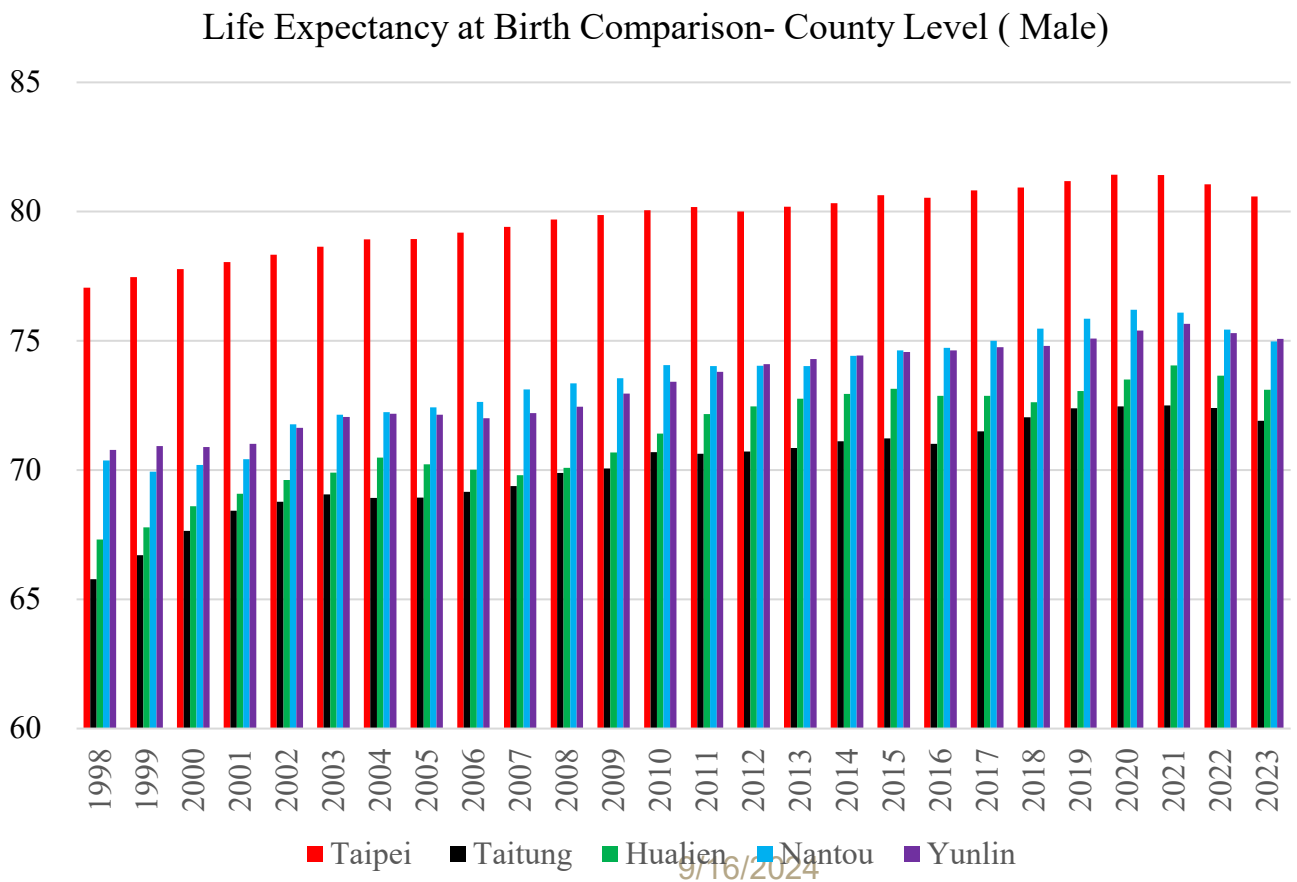
Taiwan's standardized mortality rate and life expectancy at birth have continued to grow, except for the impact of COVID-19 in 2021-22.



- Geographic inequalities in life expectancy among Taiwan's Townships/Districts
  - Regional variation in life expectancy with the degree of urban development.

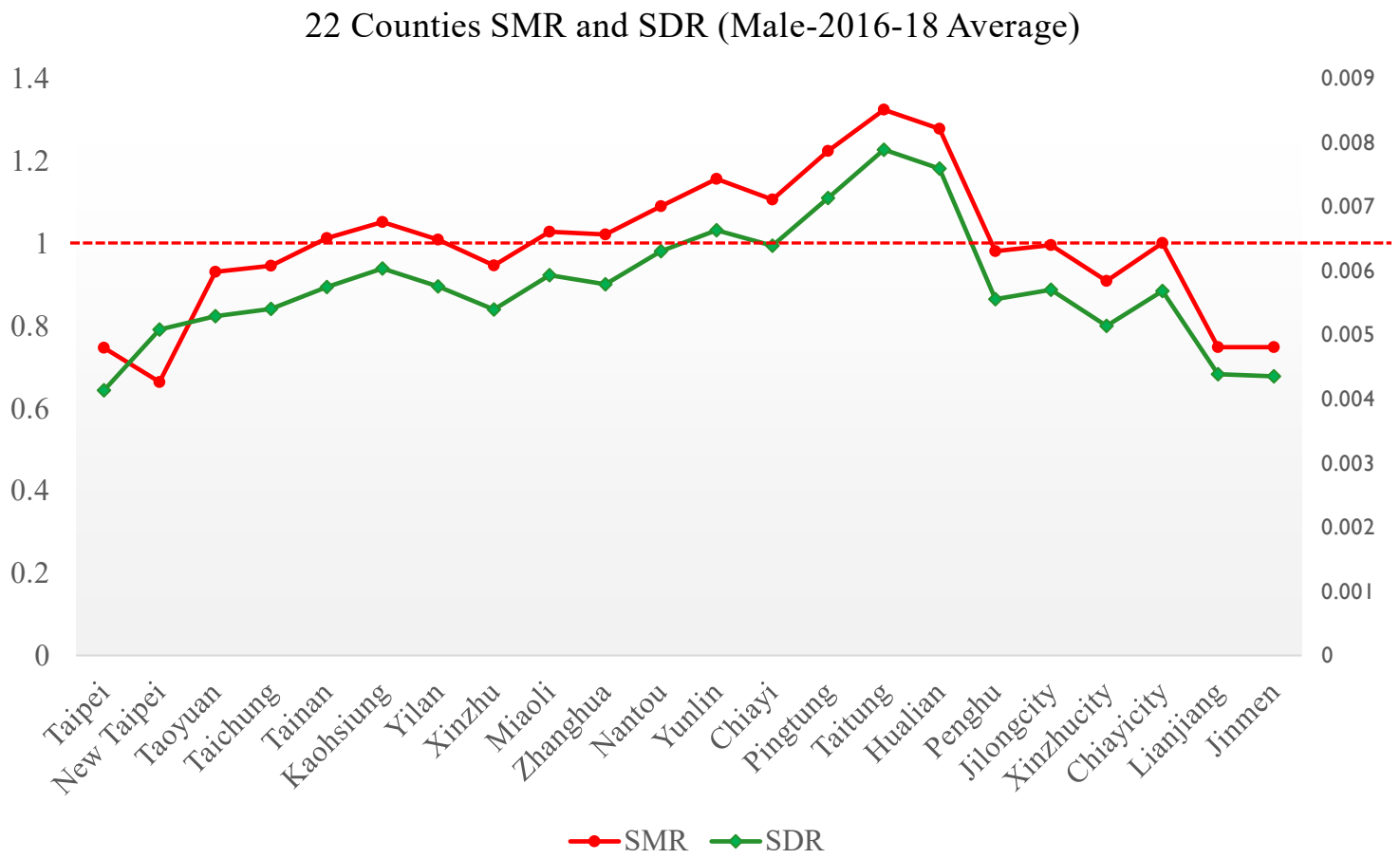
The life expectancy at birth in Taipei City, Taitung County, Hualien County, Nantou County, and Yunlin County from 1998 to 2023.

➔ For Example: The life expectancy at birth for males in Taipei City has remained at least 8 years higher than that of males in Taitung County for 26 years.



- Out of all 22 counties, 12 have higher mortality rates than the overall Taiwan population.

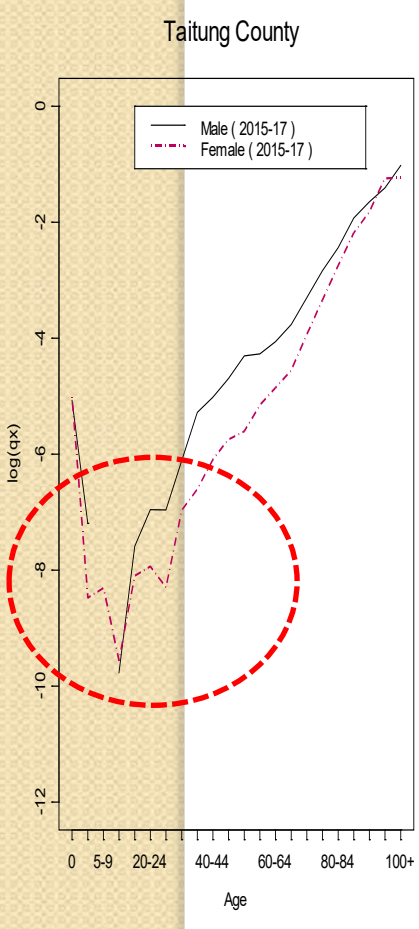
*SMR value less than 1 indicates a mortality rate lower than that of the overall Taiwan population.
*SDR: Standardized Death Rate (right axis)-based on WHO 2000

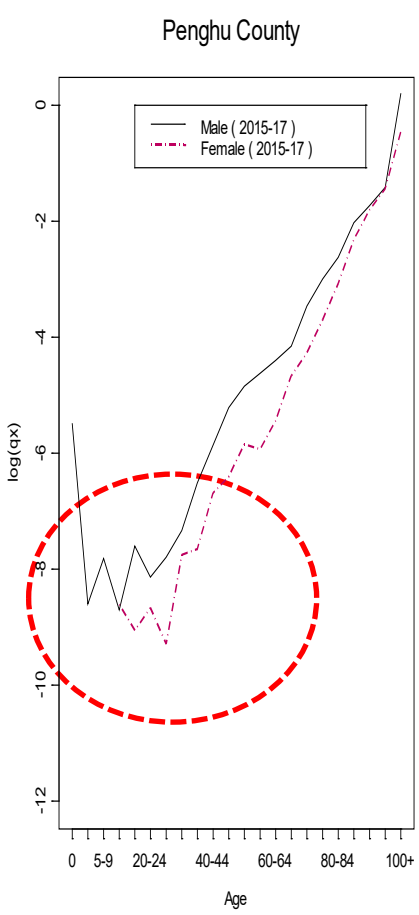


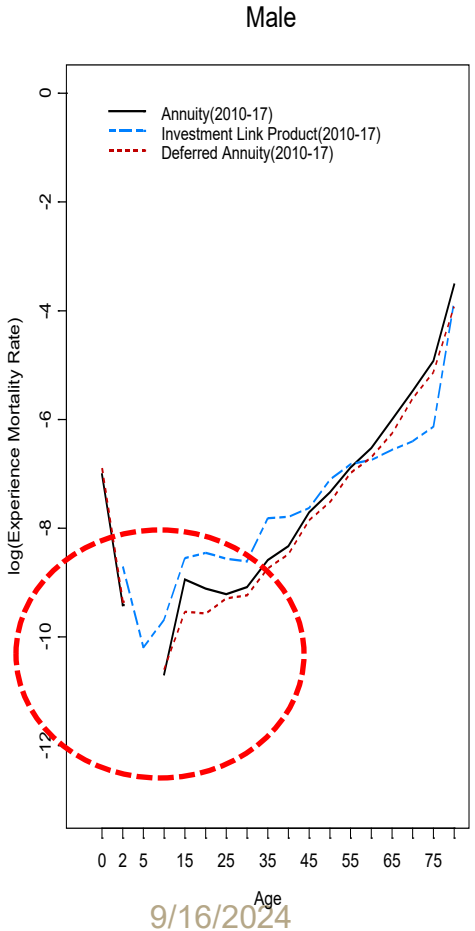
Motivation

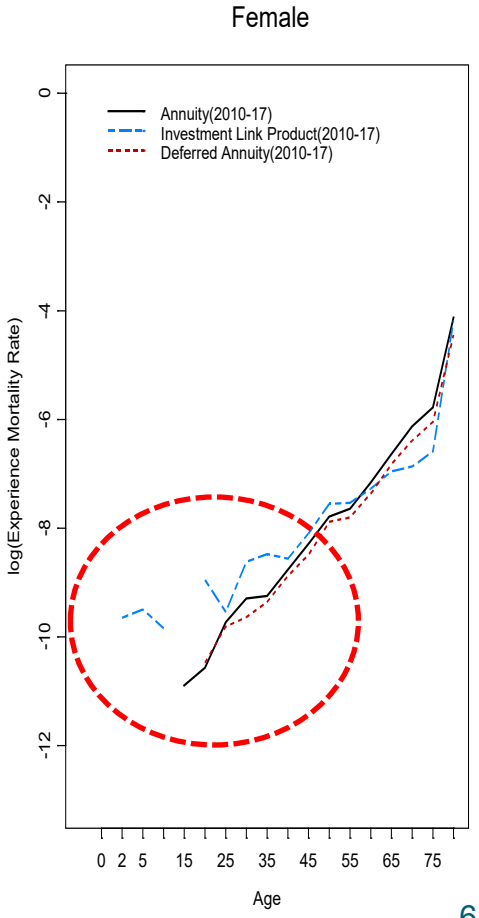
- Similar challenges in fitting townships age-specific mortality curves, the incidence experience rates for Taiwan life insurance policies, and in other related field.

--When the number of deaths (occurrences) in adjacent age groups is zero, it causes the age-specific mortality curve to become unstable and exhibit significant fluctuations.









Motivation

- Issues and Importance of Life Expectancy, Mortality Rate, and Incidence Rate Predictions for Small Populations
- Wilson (2018) found that regional demographic data are often short time-series and have high fluctuation caused by the small number of deaths, particularly among younger age groups.
- Alexander et al. (2017) pointed out that reliable estimates of mortality rates for small populations are crucial when studying health inequalities in a country..
- Orwat-Acedańska, A. (2024) pointed out that for the development and evaluation of regional healthcare and urban planning are important such as the aging process is increasing

The county/townships-level population sizes are often small, and their observed age-specific mortality rates fluctuate significantly across consecutive ages.

Motivation

- How to solve the missing data problem ?

Methods :

1.Combining data across years or age groups

2. Replace death rates from larger geographic units, substituting zero cells with small positive values

3.Standard regression analyses that model relationships between death rates and population demographic and socioeconomic characteristics

4.Graduation techniques that smooth crude mortality rates

5. With principal component methods that capture typical age patterns of mortality to estimate small-area death rates

6. Tracts with acceptable age patterns of mortality served as models from which to borrow information for tracts with missing age-specific death counts.

Methodology

- However, neighboring townships or regions might possess significantly different socioeconomic characteristics, and their age-specific occurrence rate curves may also differ.

In this study, we calculate life expectancy by three methods:

1. We check 8 socioeconomic indicators such as income, unemployment rate, and labor force participation rate to fit multiple linear regression and predict regional life expectancy.

2. We first classify many small populations using clustering methods based on standardized death rate (SDR), or standardized mortality ratio (SMR) and adopted heterogeneity index (Lee, 2003) to choose the reasonably reference populations. Then, we use

1. Traditional life table construction method

2.Haybittle (1998) Method, it estimates life expectancy directly from SMR of small areas.

To calculate life expectancy.

Methodology

- Data

We use data from 22 counties and 368 townships in Taiwan (2007-2018)

EX: Male Population (2016-18 Average)

There are 65 Townships /Districts with populations between 5,000 and 10,000 and 58 Townships / Districts populations less than 5,000.

Population	0-5,000	5,001-50,000	50,000-100,000	100,000-300,000	300,000 +
Number of Counties	0	1	2	10	9
Number of Townships/Disticts	58	241	43	26	--

Note: The division of administrative regions in Taiwan is mainly based on governance convenience, including electoral population, to determine the scope of government authority, responsibility attribution, jurisdiction of residents, and fiscal acquisition.

Methodology

Mortality Indexes:

- Standardized Death Rate(SDR)

$$SDR = \frac{\sum(m_x \times P_x^S)}{\sum P_x^S}$$

-m_x age-specific death rate of study population for age group x

-P_x^S No. of persons of standard population for age group x

Note: World (WHO 2000-2025) Standard - Standard Populations for SDR

Methodology

- The standard mortality ratio (SMR)

$$\text{SMR} = \frac{\sum_x d_x}{\sum_x n_x \cdot u_x^*}$$

- d_x is the death number of small area at age of x
- n_x is the population of small area at age of x
- u_x^* is the mortality rate of standard population at age of x.

- The SMR is calculated using the same age structure
- The smaller the SMR, is usually the larger the life expectancy is.

.

Note: We can aggregate the Taiwan historical mortality data for the target population (age-wise), treating the aggregation as the reference population for SMR.

Methodology

- The Whittaker method is to minimize the following objective function, i.e., weighted sum of fit function F and smoothness function S:

$$\mathrm{M} = \mathrm{F} + hS = \sum_{x=1}^{\mathrm{n}} w_x (v_x - u_x)^2 + h \sum_{\mathrm{x}=1}^{\mathrm{n-z}} (\Delta^z v_x)^2$$

- u_x and v_x are observed and graduated mortality rates for age x
- w_x is the weight for age x
- n is population size
- h and z are the parameters to be decided.
- Δ^z is the z time difference.

The Whittaker ratio is an extended version and we plug into the mortality ratio of small and reference populations for graduation, instead of mortality rates.

Methodology

- The partial SMR: is one way to deal with estimating mortality rates of small populations, by adding information from other (large) population to correct possible bias

$$v_x = u_x^* \times \exp\left(\frac{d_x \times \hat{h}^2 \times \log(d_x / e_x) + (1 - d_x / \sum d_x) \times \log(\text{SMR})}{d_x \times \hat{h}^2 + (1 - d_x / \sum d_x)} \right)$$

- u_x^* and v_x are observed and graduated mortality rates for age x

- d_x and e_x where and are the observed and expected numbers of deaths at age x for the small population

- $\hat{h}^2$ is the estimate of parameter h^2 for measuring the heterogeneity (in mortality rates) between the small population and large population.

--Lee, W. (2003), "A Partial SMR Approach to Smoothing Age-specific Rates." *Annals of Epidemiology* 13(2), 89-99.

Methodology

- $\hat{h}^2$ was the estimated value of the heterogeneity parameter h^2 via:

$$\hat{h}^2 = \max\left(\frac{\sum\left((d_x - e_x \times SMR)^2 - \sum d_x\right)}{SMR^2 \times \sum e_x^2} \;,\; 0 \right)$$

- Higher value implies more heterogeneity in the rate ratios of the study and the standard populations (or stated differently, more dissimilarity in shape between the two age curves)
- Lower value, less rate-ratio heterogeneity (more similarity between the two curves).
- When it is zero, i.e. no rate-ratio heterogeneity detected in the data

➔We use it as a criterion for choosing a suitable reference population.

Methodology: Life Expectancy Estimate

- Methods

1. Multiple Linear Regression and Life table Method
2. Haybittle (1998) Method

$$e_x^S \cong e_x^R - {\log(SMR)}/{K}$$

$wher\ e_x^S, e_x^R$ are the life expectancy at age x for the study population and reference population respectively. K is a constant estimate form the observed age-specific mortality rate on the Gompertz function, $\mu_x = Be^{kx}$.

Note: The abridged life table for Taiwan's counties and cities is created by combining three years of population data for each region, with old-age mortality rates extrapolated using the Gompertz method.

Reference

Haybittle JL. (1998) The use of the Gompertz function to relate changes in life expectancy to the standardized mortality ratio. *Int J Epidemiol*. 27: 885– 889

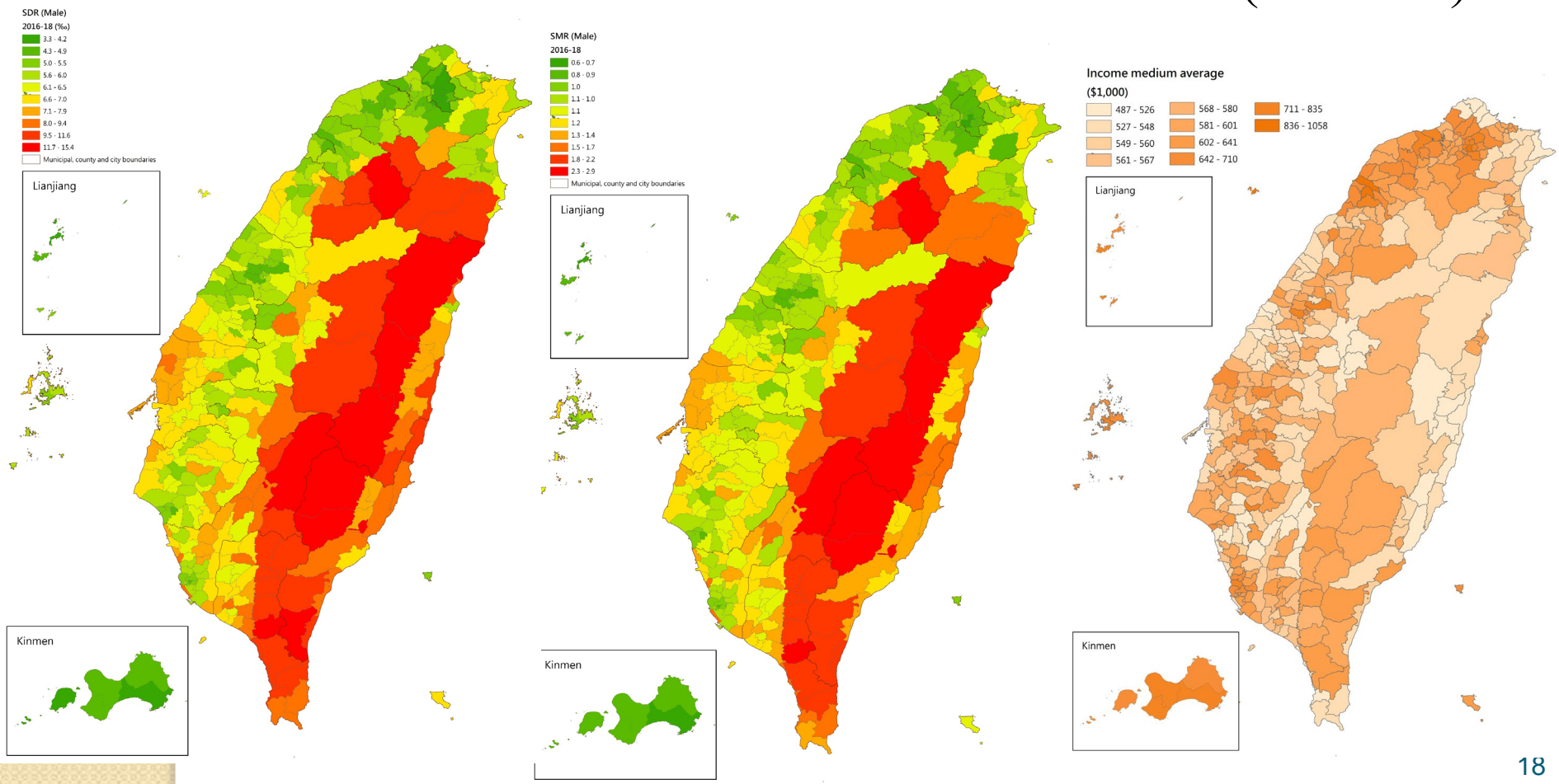
Lai D, Guo F, Hardy RJ. (2000) Standardized mortality ratio and life expectancy: a comparative study of Chinese mortality. *Int J Epidemiol*. 29:852–5.

Multiple regression analysis

Exploratory data analysis (EDA)

- SDR, SMR, and Income

Significant clustering phenomena are observed in geographical space concerning the standardized death rate, standard death ratio, and income among both men and women. EX: Male (2016-18)

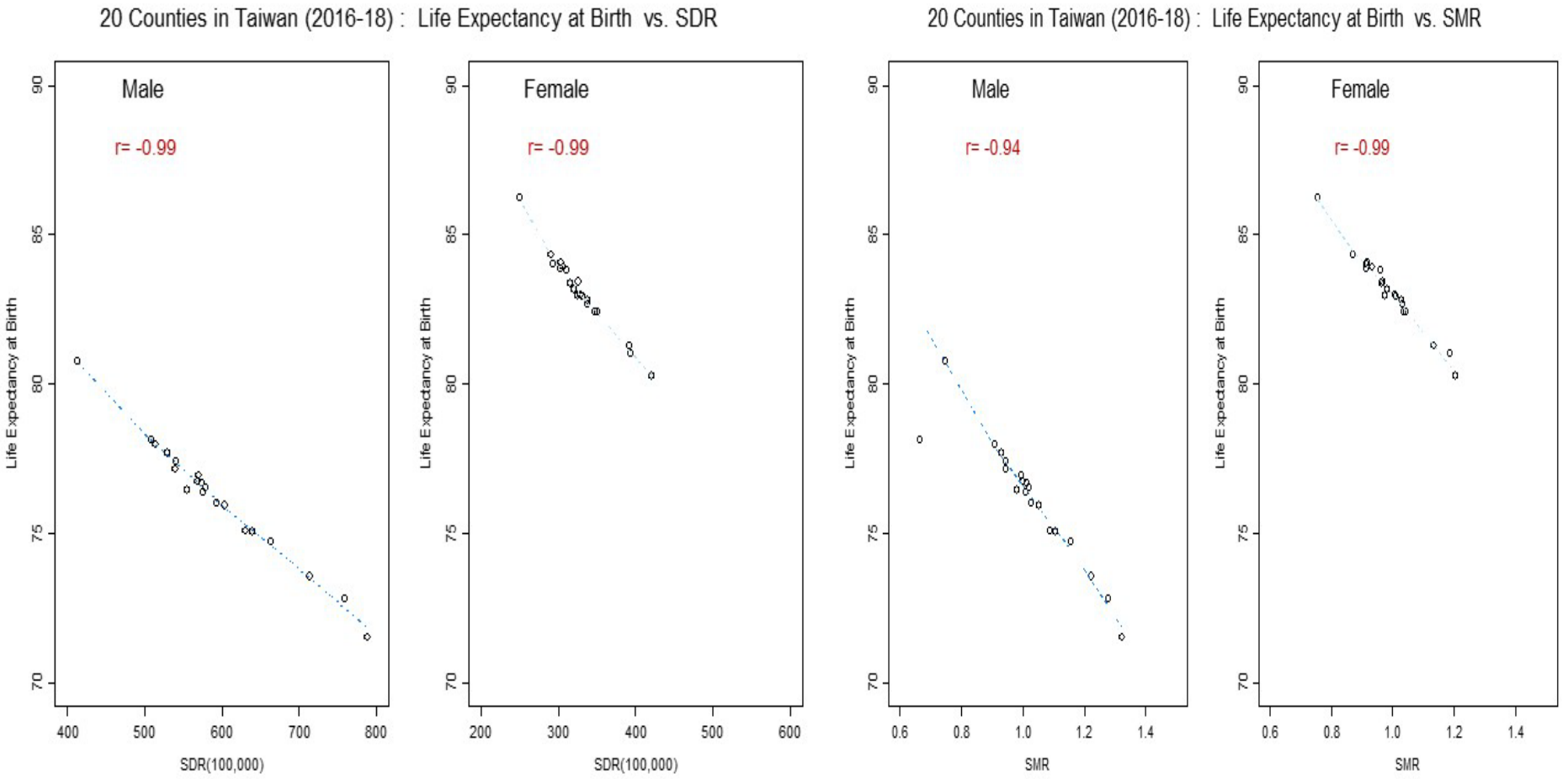


Exploratory data analysis (EDA)

- SDR, SMR, Income Vs. Life Expectancy (LE) for 20 Counties

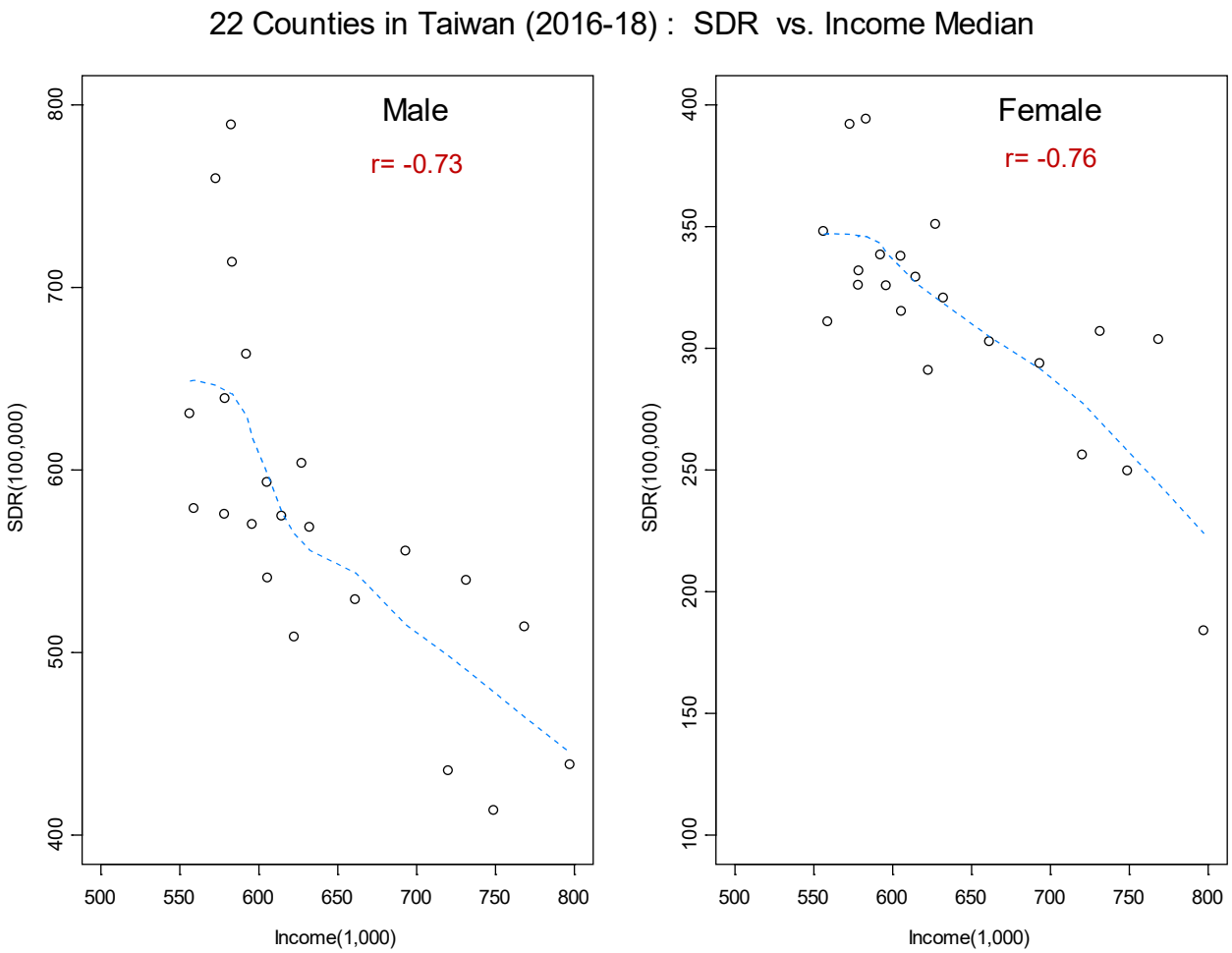
  - LE has a strong negative correlation with SDR and SMR.

Note: LE of Lianjiang and Jinmen counties have not been announced.



Exploratory data analysis (EDA)

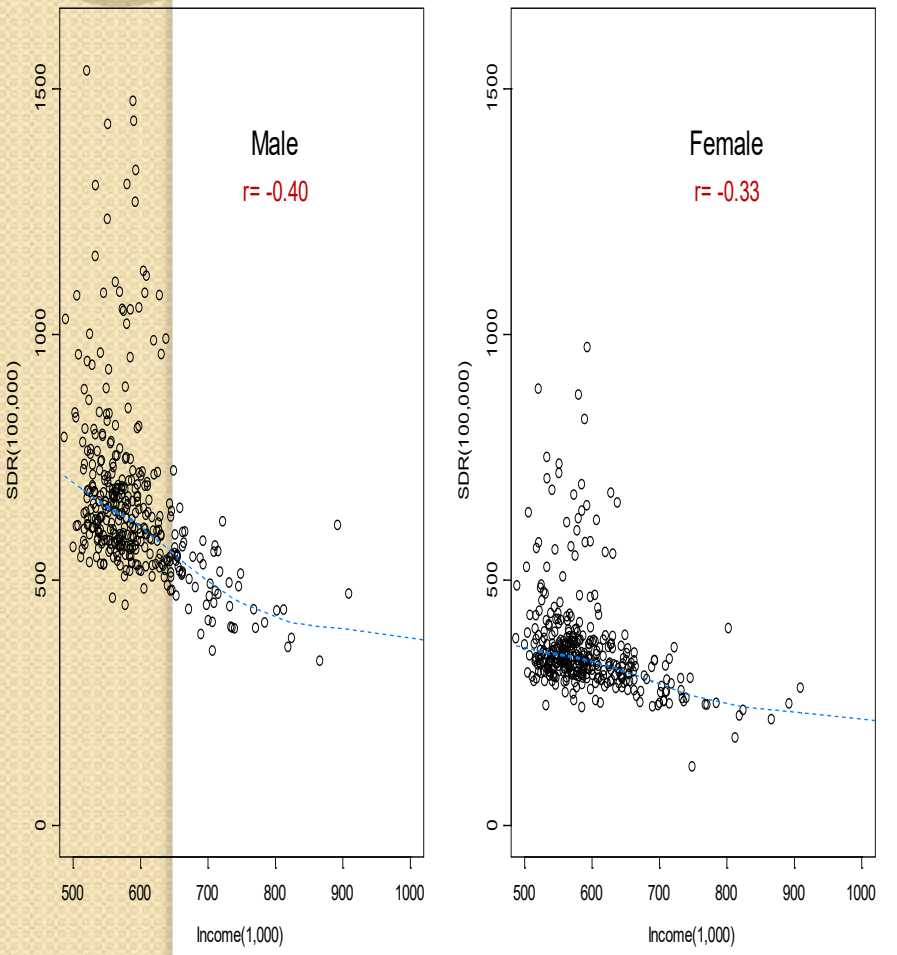
- SDR, SMR Vs. Income for 22 Counties
  - SDR of 22 counties in Taiwan is negative correlation with income.



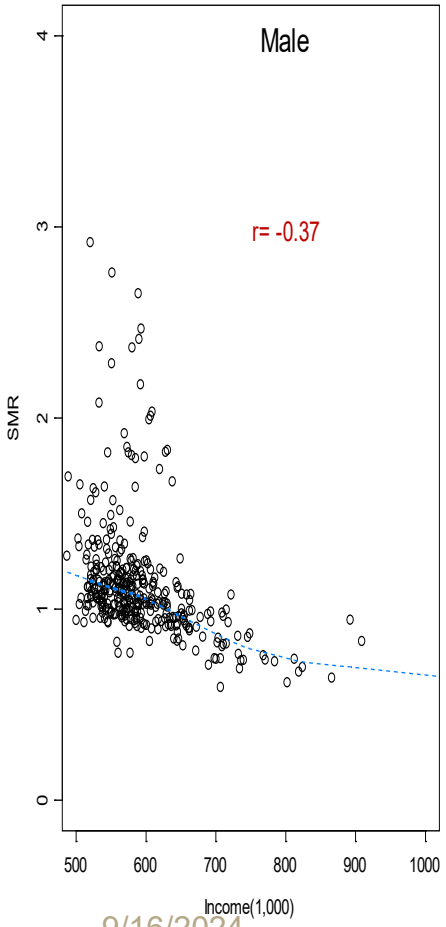
Exploratory data analysis (EDA)

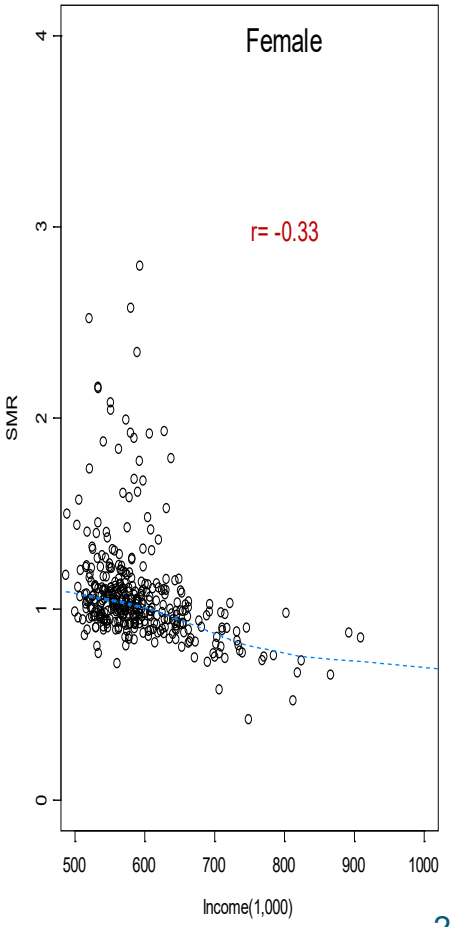
- SDR, SMR Vs. Income for 368 Counties
  - SDR and SMR of 368 Townships in Taiwan is negative correlation with income.

368 Townships in Taiwan : SDR vs. Income Median



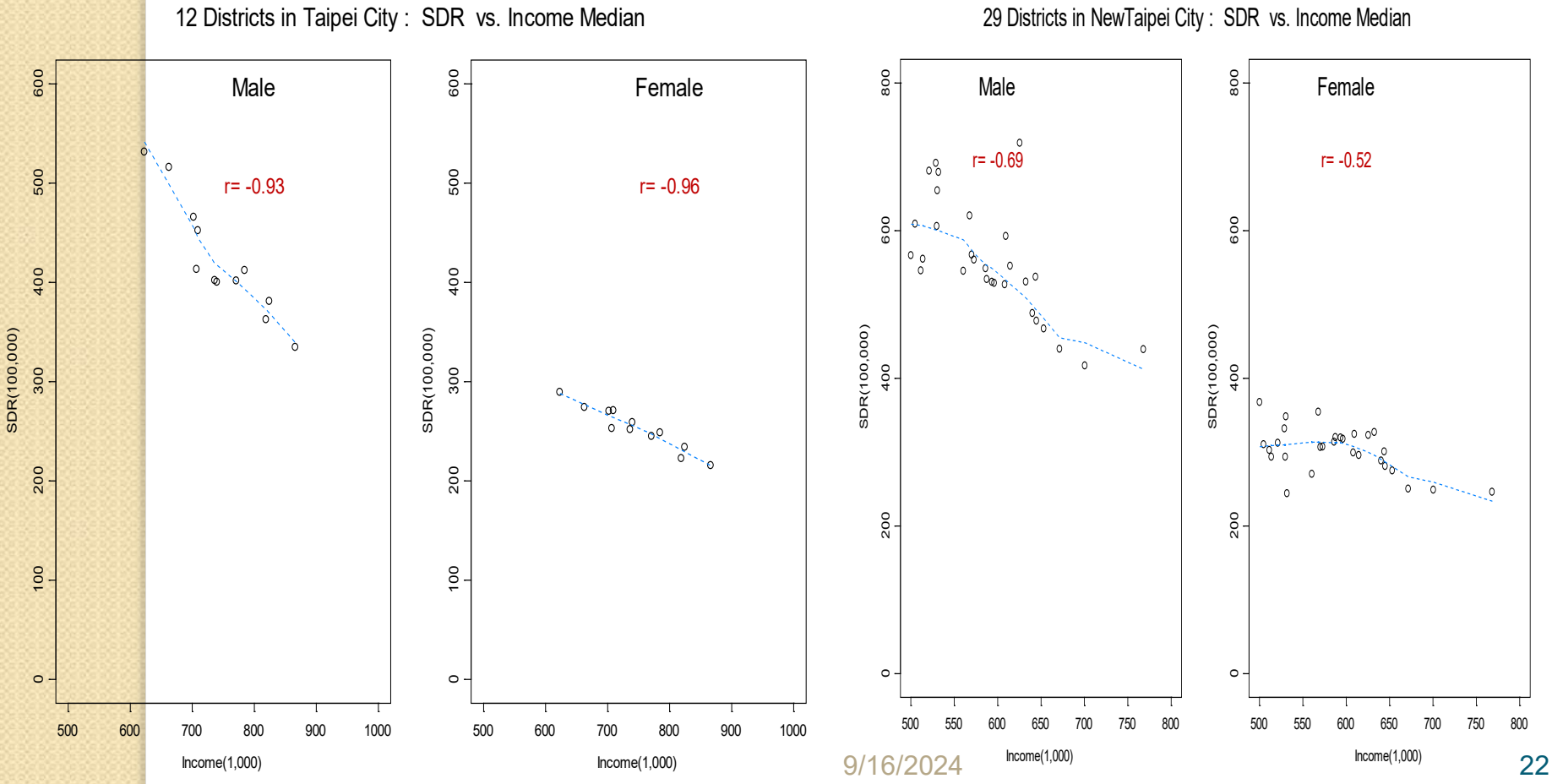
368 Townships in Taiwan : SMR vs. Income Median





Exploratory data analysis (EDA)

- SDR Vs. Income for high degree of urban development
  - SDR of 12 districts in Taipei City is strongly negative correlation with income.
  - SDR of 29 districts in New Taipei City is also negative correlation with income.



Exploratory data analysis (EDA)

- Socioeconomic Indicators Vs. Life Expectancy (LE) for 20 Counties

8 Socioeconomic Indicators:

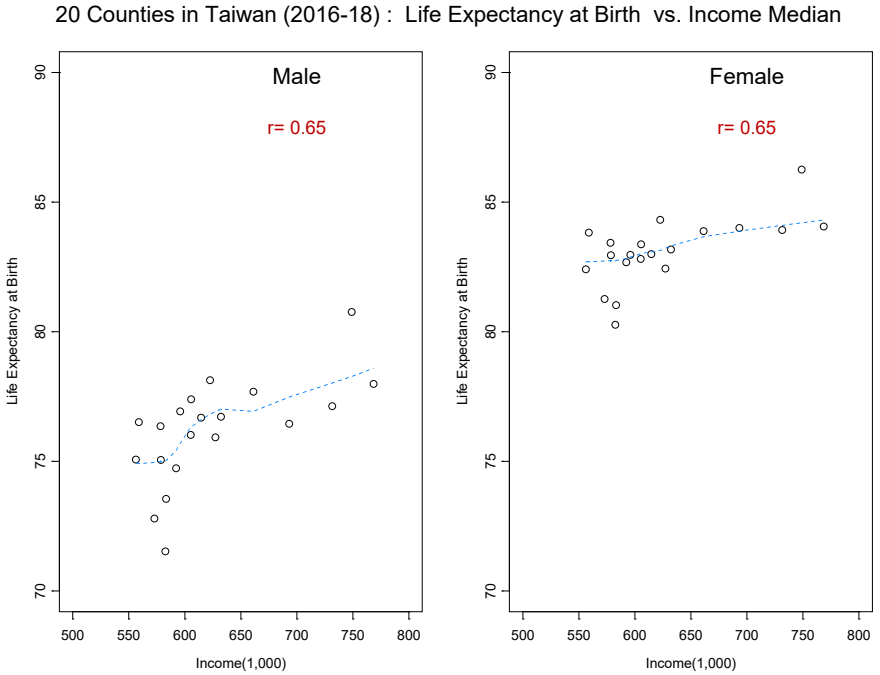
- Medical Institution Numbers
- Doctor Numbers
- Monthly Personal Consumption Expenditures (MPCI)
- Median Disposable Personal Income (DI)
- Individual Income
- Population Density
- Agricultural Laborers Proportion
- The Deprivation Index of Areas

Note: The deprivation index of areas was developed using the proportion of agricultural laborers and the proportion with no schooling among 12−21 year olds.

Note: Data from Statistics Yearbook of Practicing Physicians and Health Care Organizations in Taiwan, Individual Income Tax Return, The Report on the Survey of Personal Income Distribution in Taiwan Area, Population and Housing Census, 2010.

Exploratory data analysis (EDA)

- Socioeconomic Indicators
  - LE has positive correlation with, Doctor Numbers, MPCI, DI, Individual Income, Population Density, and negative correlation with Agricultural Laborers Proportion, the Deprivation Index of Areas.
  - The correlation between Medical Institution Numbers and LE is not significant.



		Medical Institution Numbers	Doctor Numbers	MPCI	DI	Individual Income	Population Density	Agricultural Laborers Proportion	The Deprivation Index of Areas
Male	Correlation Coefficient	0.5360	0.6142	0.7230	0.7517	0.6474	0.7124	-0.7237	-0.7441
	P-Value	0.0149	0.0040	0.0003	0.0001	0.0020	0.0004	0.0003	0.0002
Female	Correlation Coefficient	0.4215	0.5321	0.6250	0.6746	0.6512	0.6738	-0.5746	-0.5701
	P-Value	0.0640	0.0157	0.0032	0.0011	0.0019	0.0011	0.0081	0.0087

The significance threshold was set at .05

Life Expectancy Prediction

- Multiple regression analysis for 7 socioeconomic and SDR indicators
  - We use the training set (2007-16 Data from 20 counties) to fit model and find the regression coefficients.
  - Each socioeconomic indictor as the explanatory variable, and LE at birth as the response variable.

Ex: We use the parameter values of the previous year and 2018 indicator value to predict 2018 LE at birth for 20 counties. Multiple R-Squared of SDR variable attains 0.95.

	Male		Female	
	Multiple R-Squared	MAPE(%)	Multiple R-Squared	MAPE(%)
Doctor Numbers	0.395	1.680	0.298	1.065
MPCI	0.483	1.684	0.364	0.956
DI	0.537	1.658	0.408	0.911
Individual Income	0.364	1.769	0.364	1.011
Population Density	0.512	1.520	0.440	0.901
Agricultural Laborers Proportion	0.517	1.421	0.327	0.962
The Deprivation Index of Areas	0.542	1.422	0.316	0.985
SDR	0.964	0.367	0.951	0.304

Life Expectancy Prediction

- Multiple regression analysis for 7 socioeconomic and SDR indicators

  - We use stepwise regression analysis to find the explanatory variables .

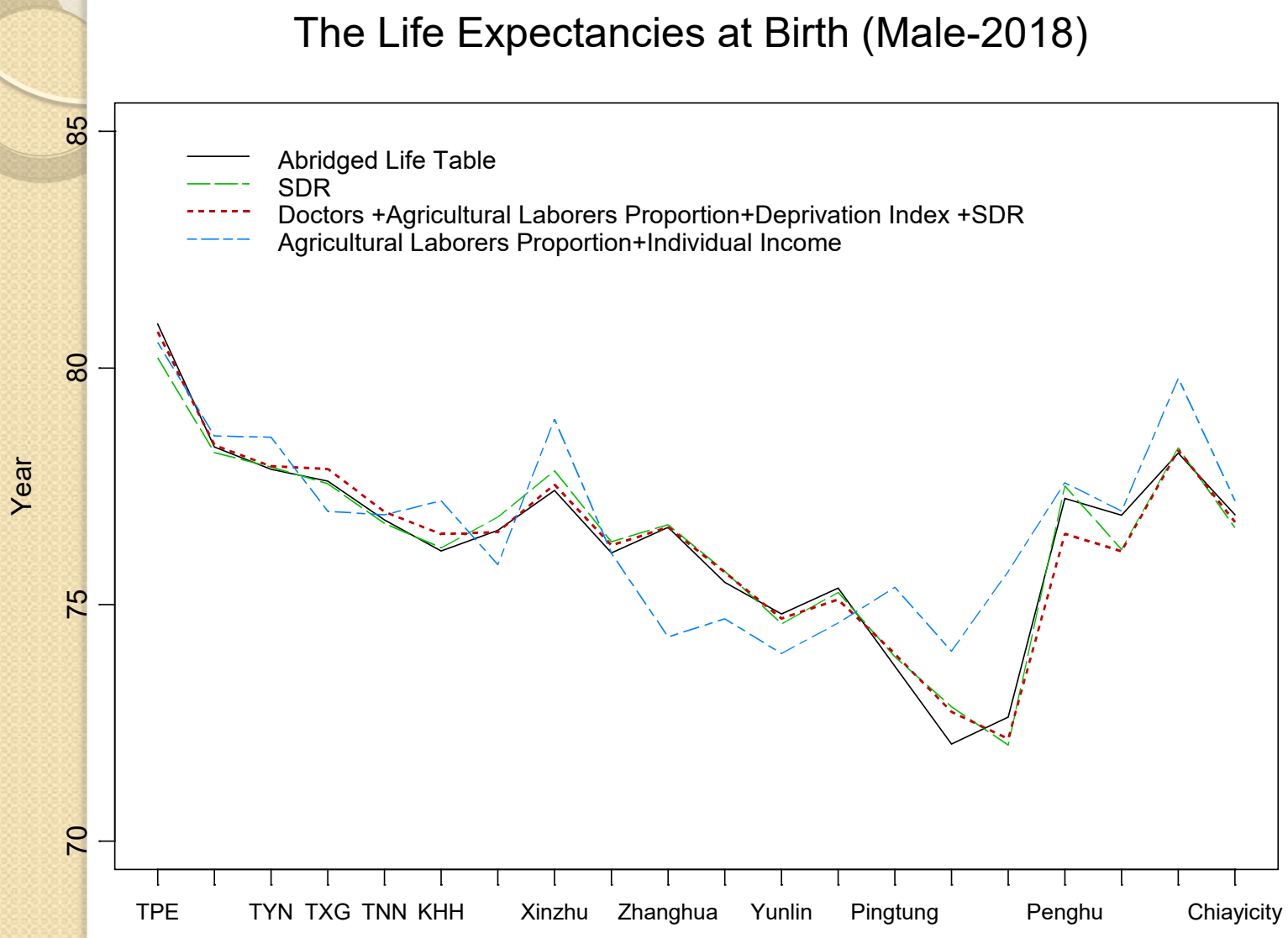
Male: The Doctor Numbers, Agricultural Laborers Proportion, The Deprivation Index of Areas, and SDR as the explanatory variables. The multiple R-Squared attains 0.99.

Female: The SDR as the explanatory variables. The multiple R-Squared attains 0.95.

	Male		Female	
	Multiple R-Squared	MAPE(%)	Multiple R-Squared	MAPE(%)
SDR	0.964	0.367	0.951	0.304
Doctor Numbers+Agricultural Laborers Proportion+The Deprivation Index of Areas+SDR	0.986	0.334	0.943	0.248
Agricultural Laborers Proportion+ Individual Income	0.715	1.260	0.608	0.856

Life Expectancy Prediction

- Life Expectancy at Birth for 20 Counties:



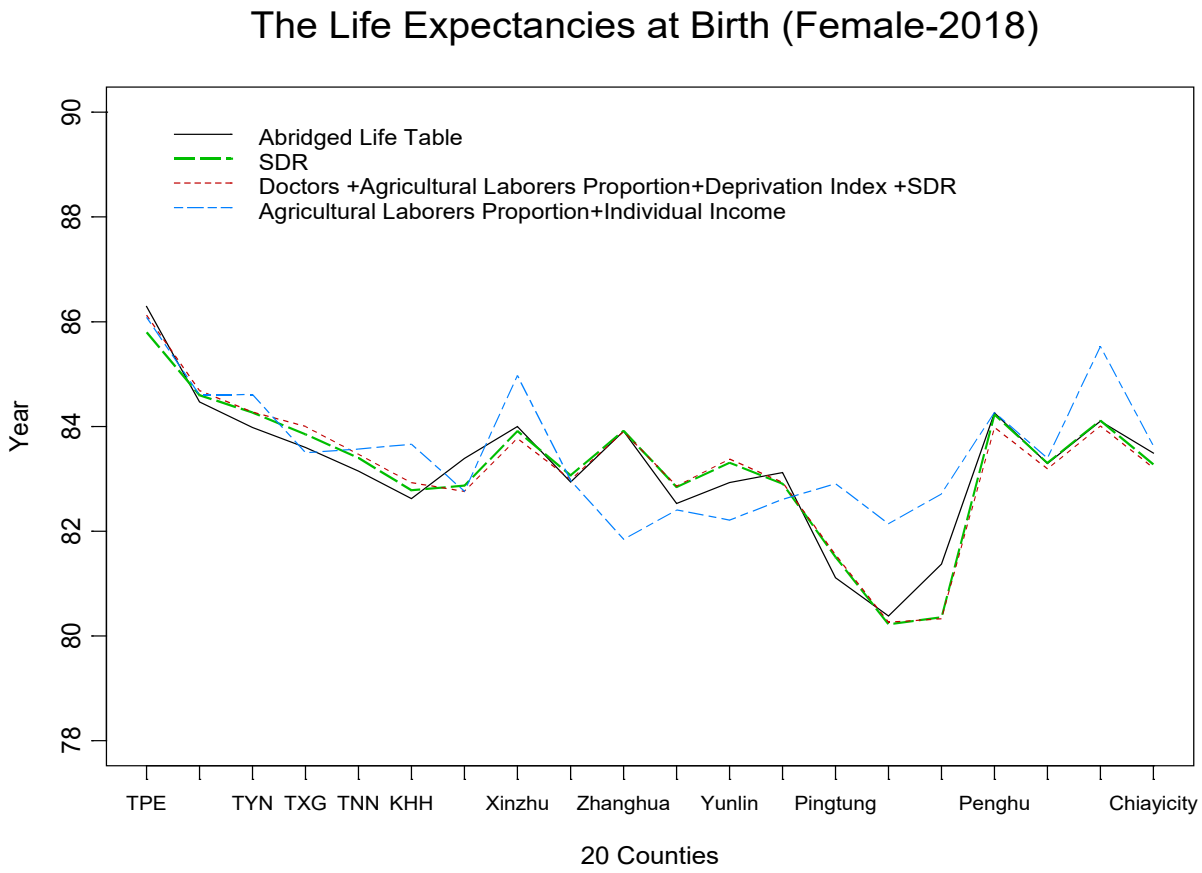
Life Expectancy Prediction

- Life Expectancy at Birth for 20 Counties:

➢ The SDR seems to be used to direct estimate life expectancy.

➢ We can add some socioeconomic indictors to enhance prediction accuracy of LE.

Note: Estimates of the life expectancy for other ages can be deduced in the same way.



Traditional life table construction method and Haybittle (1998) Method

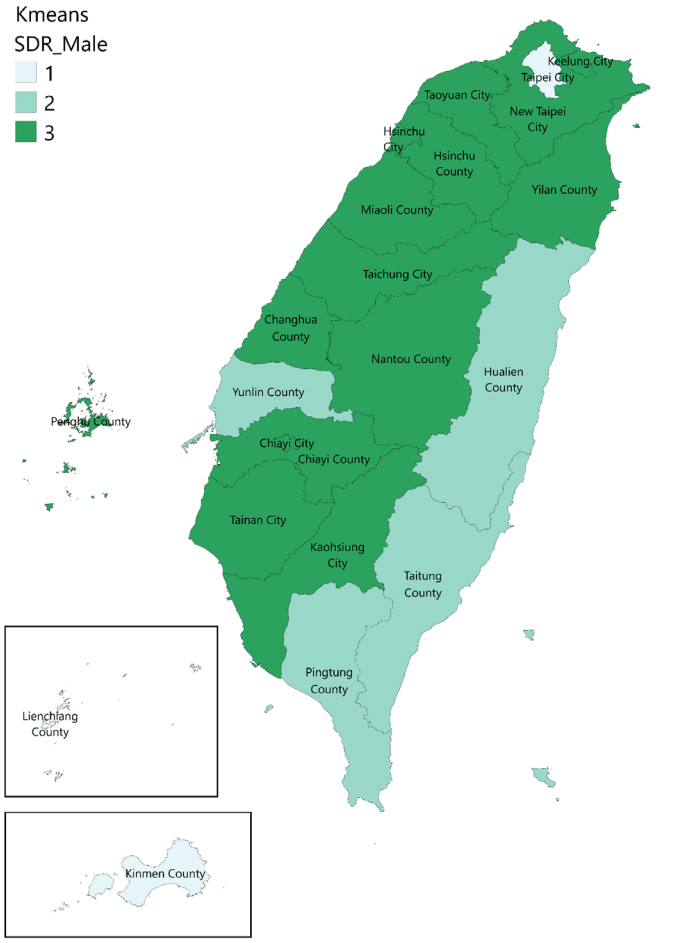
Classification

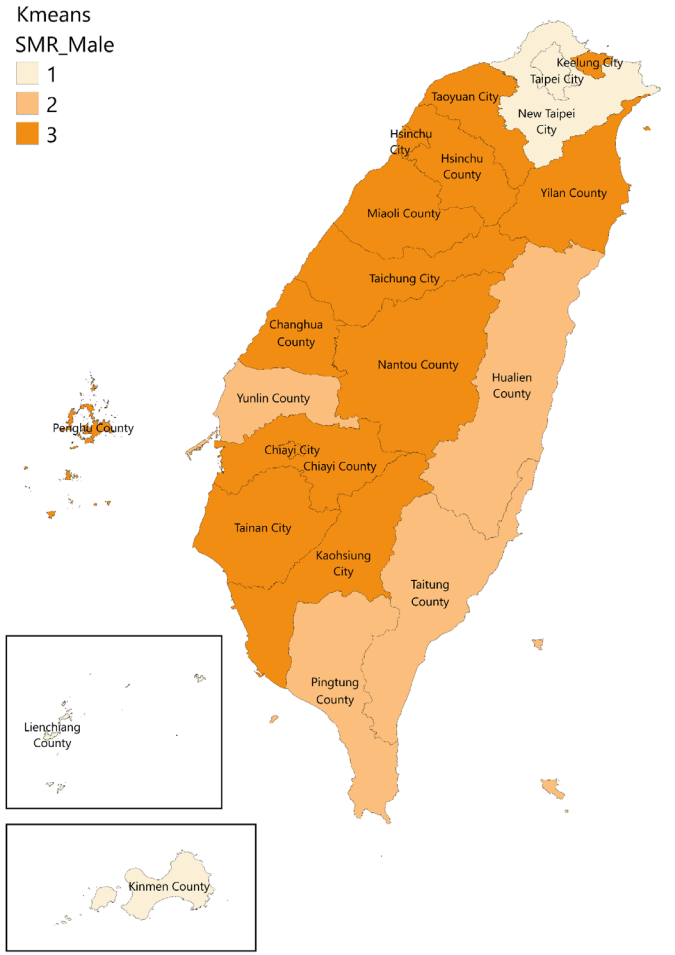
- K-means and Hierarchical clustering: Classified to 3Groups (Male's SDR and SMR)
  - There are similar classification results for SDR and SMR.
  - Ex: Taipei City, Lienchiang County, Kinmen County are in the same group➔ We can use the group as the reference population to estimate life expectancy for Kinmen County.
  - Ex: Pingtung County, Taitung County, Hualien County are in the same group➔ We can use the group as the reference population for one of these three counties to estimate life expectancy.

Classification

- K-means clustering: Classified to 3Groups (Male's SDR and SMR)

There are similar classification results, except that when we use SMR as the indicator, New Taipei City is also classified into the same category as Lianjiang County, Kinmen County, and Taipei City.

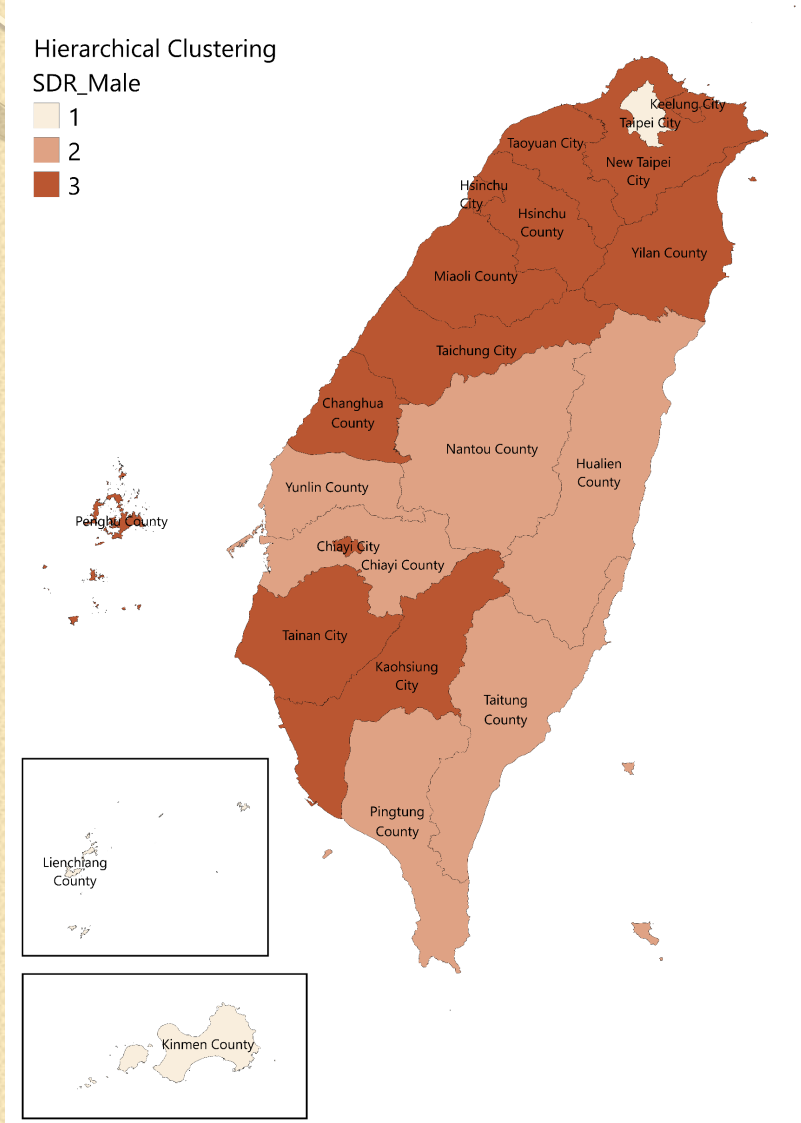


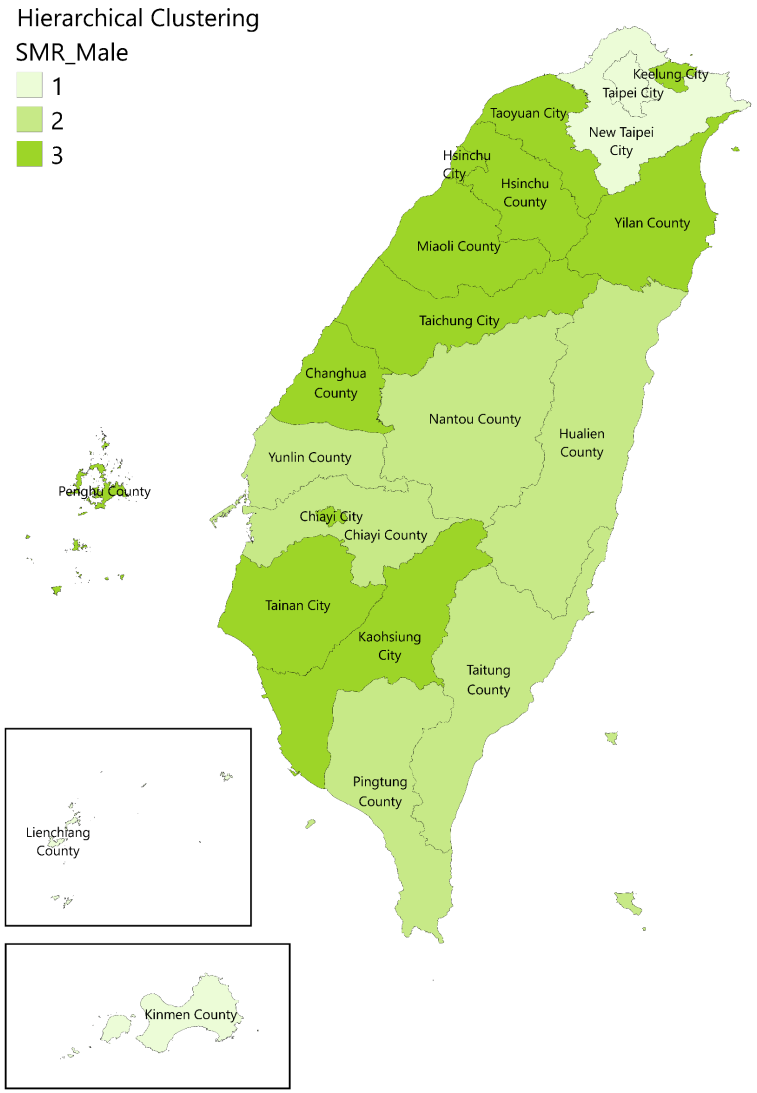


Classification

- Hierarchical clustering: Classified to 3Groups (Male's SDR and SMR)

There are similar classification results.





Classification

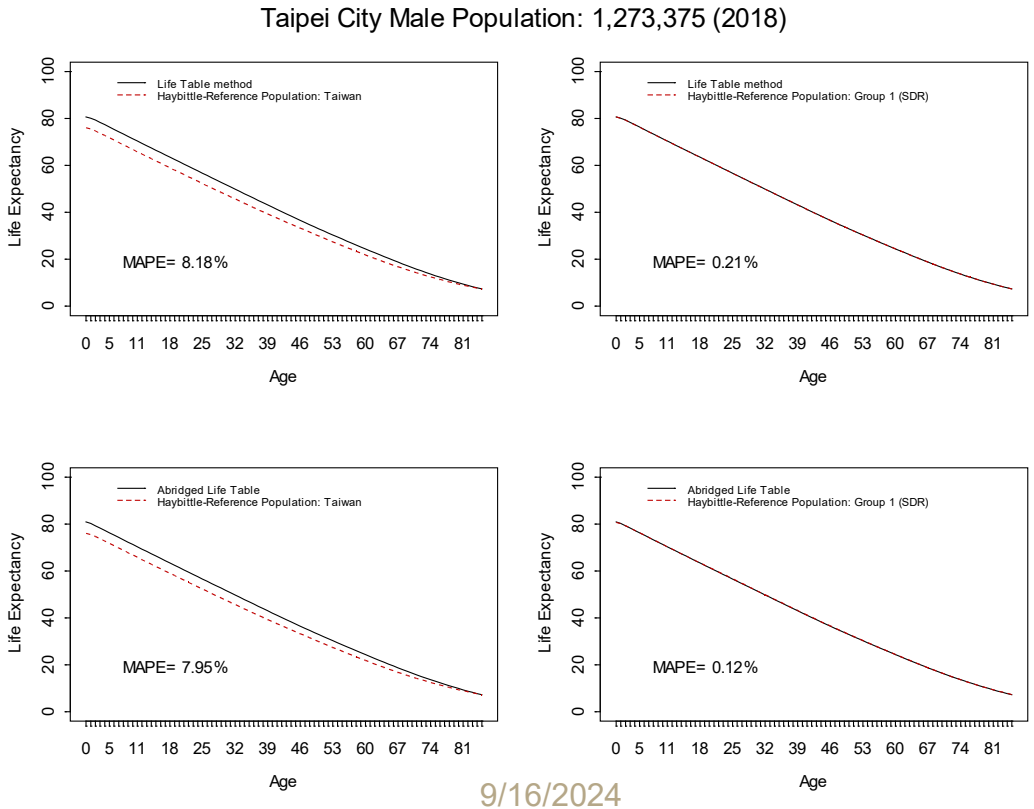
- K-means and Hierarchical clustering: Classified to 3Groups

	Male			
	SDR		SMR	
	Kmeans3	Hierarchical Clustering3	Kmeans3	Hierarchical Clustering3
Taipei City	○	○	○	○
New Taipei City	◎	◎	○	○
Taoyuan City	◎	◎	◎	◎
Taichung City	◎	◎	◎	◎
Tainan City	◎	◎	◎	◎
Kaohsiung City	◎	◎	◎	◎
Yilan County	◎	◎	◎	◎
Hsinchu County	◎	◎	◎	◎
Miaoli County	◎	◎	◎	◎
Changhua County	◎	◎	◎	◎
Nantou County	◎	△	◎	△
Yunlin County	△	△	△	△
Chiayi County	◎	△	◎	△
Pingtung County	△	△	△	△
Taitung County	△	△	△	△
Hualien County	△	△	△	△
Penghu County	◎	◎	◎	◎
Keelung City	◎	◎	◎	◎
Hsinchu City	◎	◎	◎	◎
Chiayi City	◎	◎	◎	◎
Lienchiang County	○	○	○	○
Kinmen County	○	○	○	○

Comparison and Validation

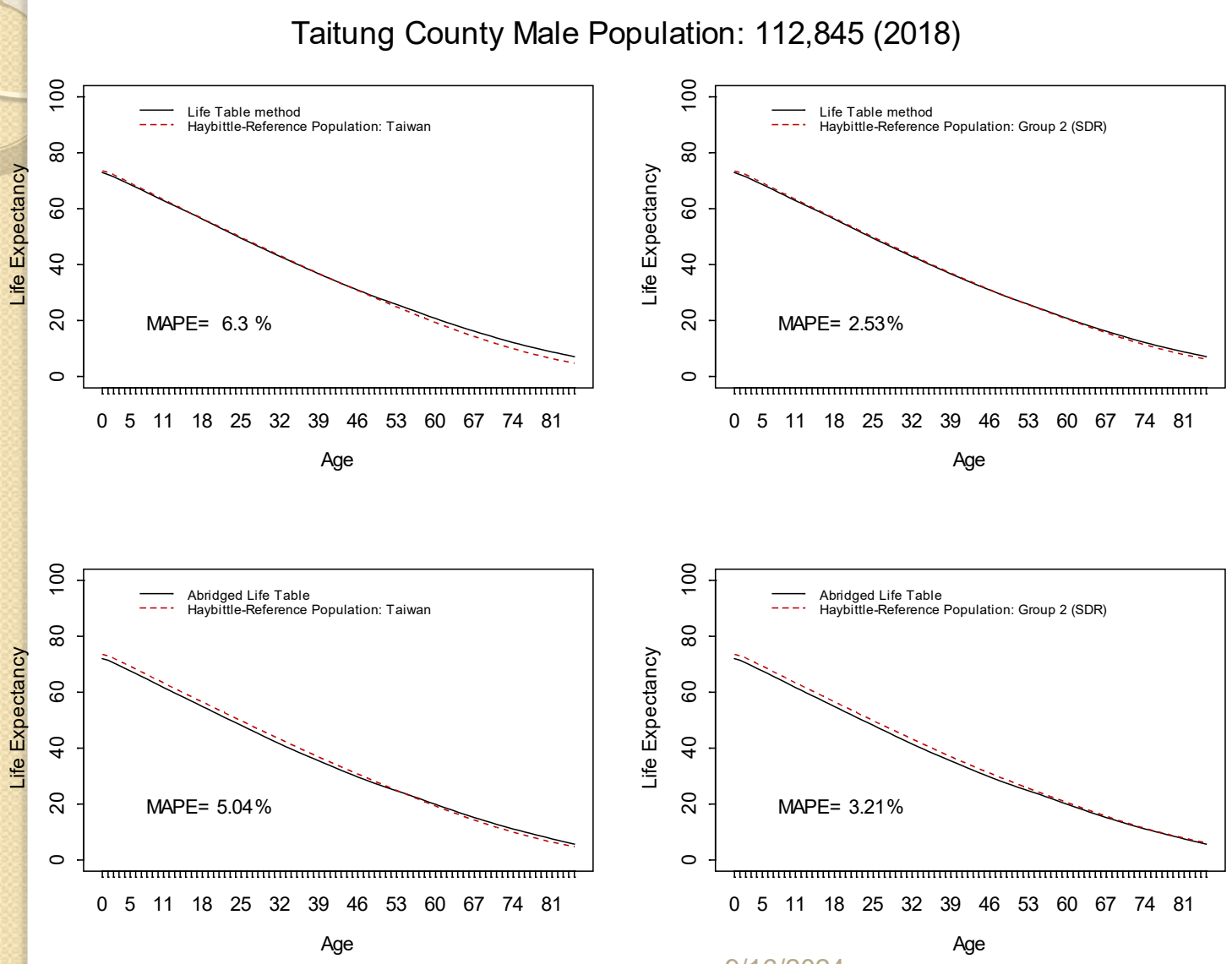
- Taipei City:
  - When we use a suitable reference population, the Haybittle (1998) Method's estimated life expectancy results are not much different from those of the Life Table Method (top right picture)
  - When compared with the abridged life table, the Haybittle (1998) Method also has good estimation results (bottom of the right picture)

➔Next four slides show similar conclusion for the other counties with population 50,000-170,000



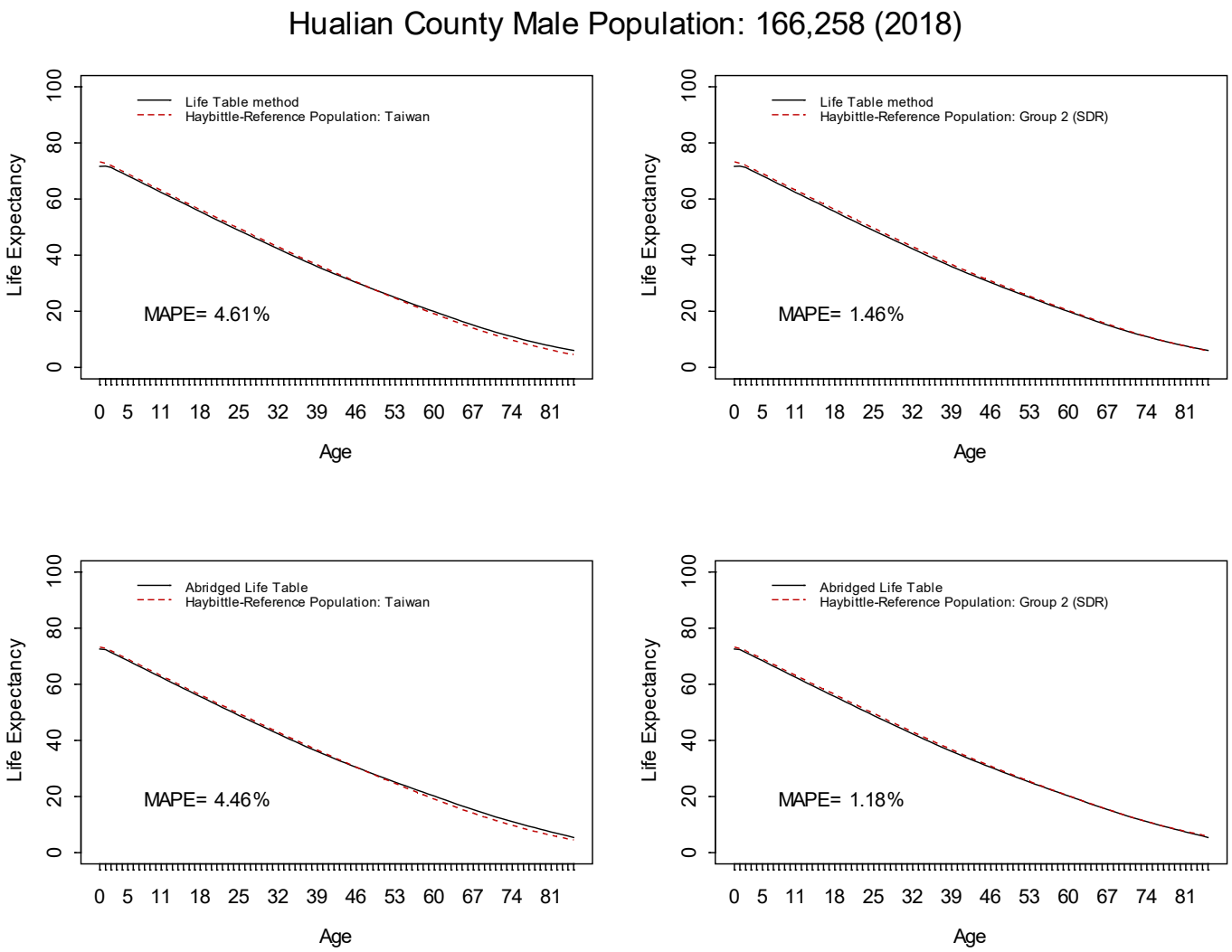
Comparison and Validation

- Taitung County



Comparison and Validation

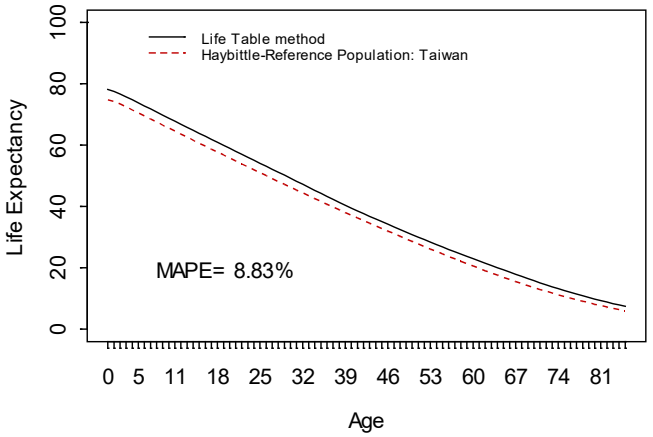
- Hualian County

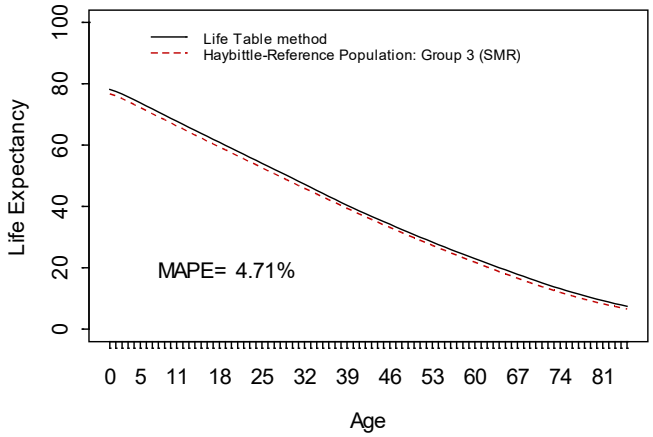


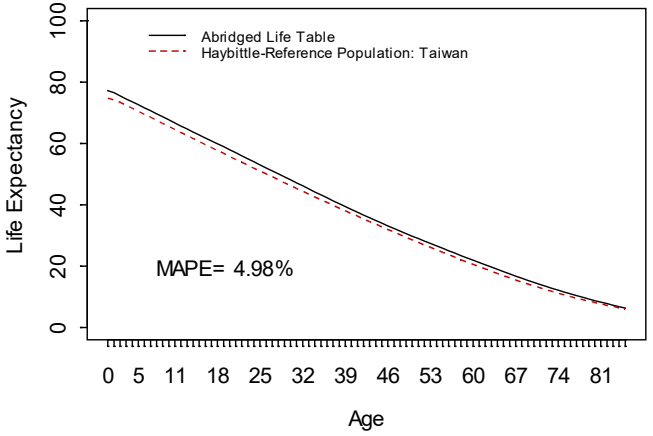
Comparison and Validation

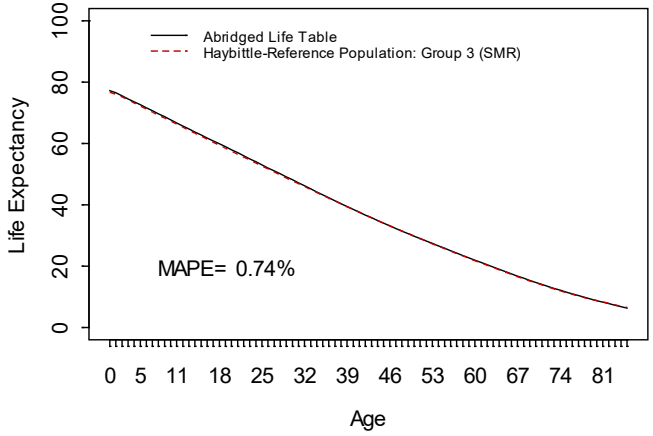
- Penghu County

Penghu County Male Population: 53,807 (2018)





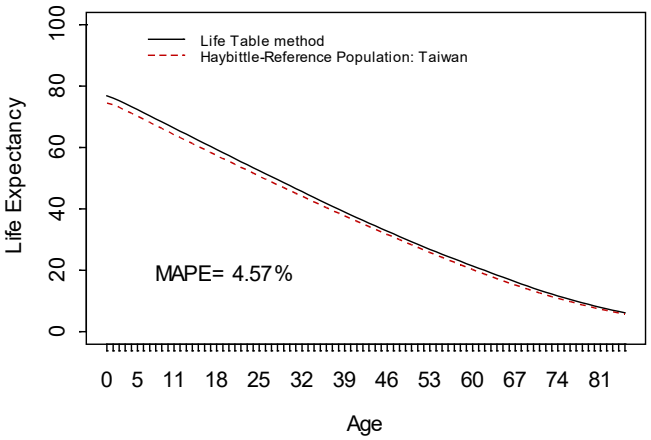


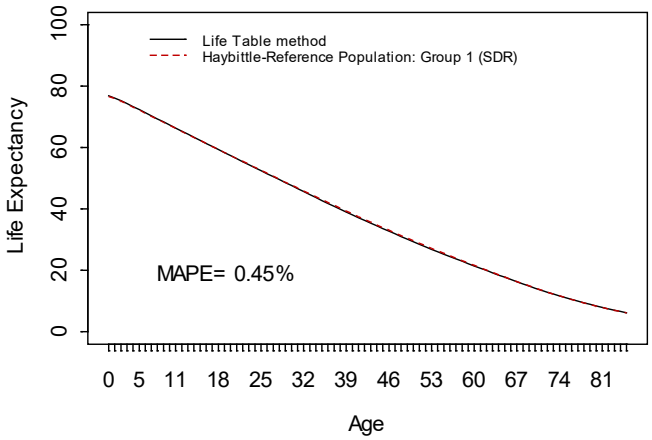


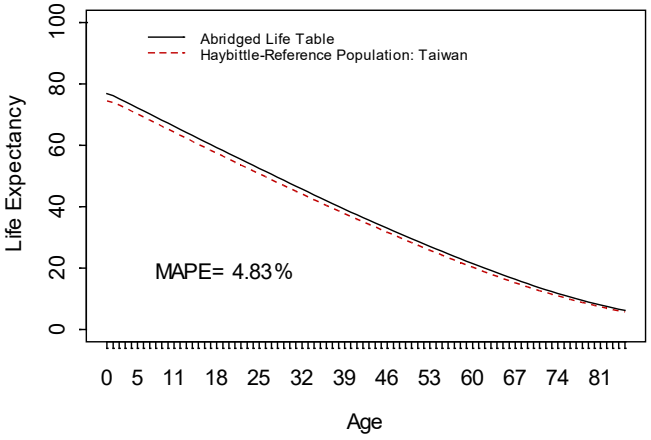
Comparison and Validation

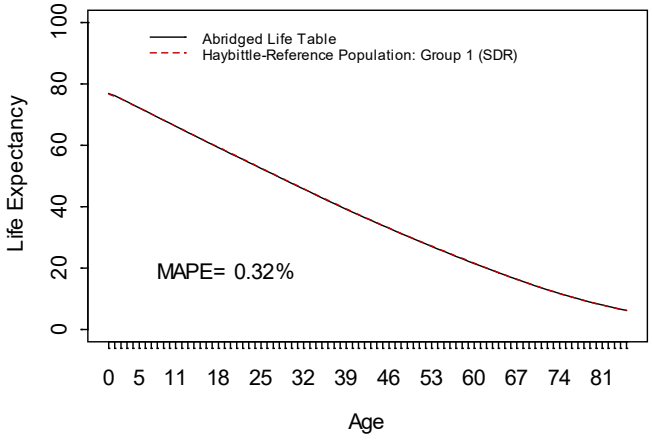
- ChiayiCity

ChiayiCity Male Population: 130,416 (2018)



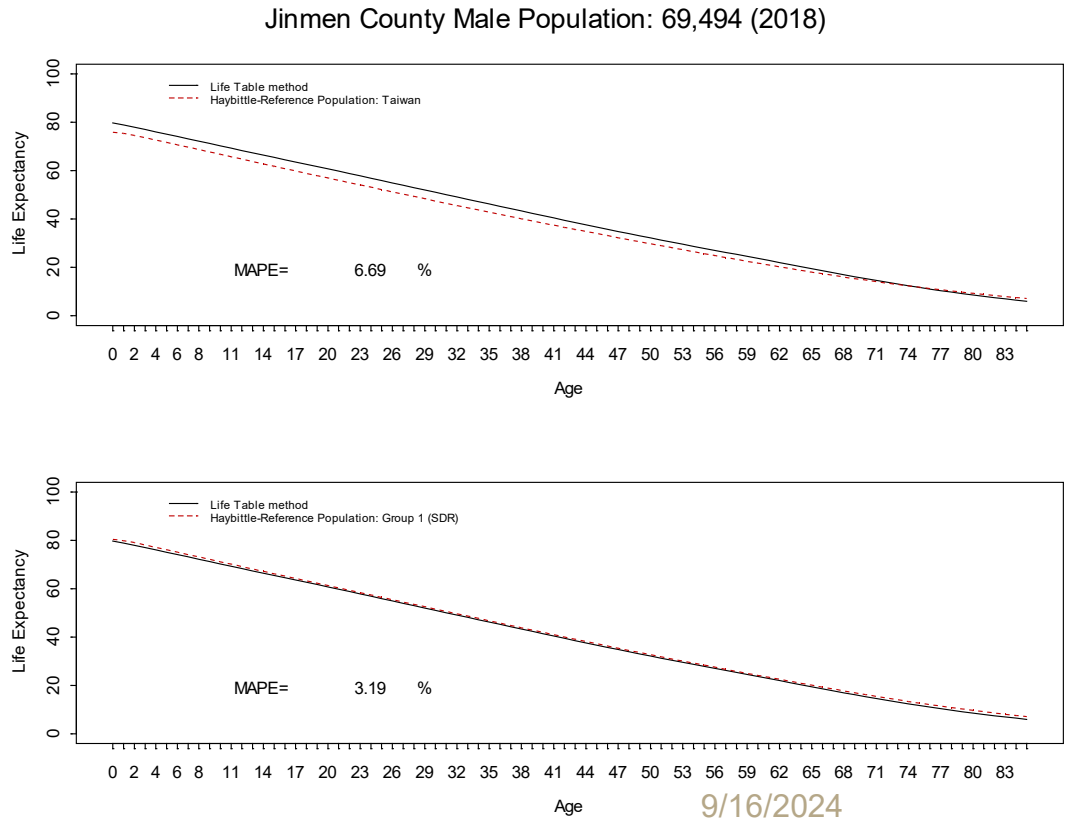






Application

- Jinmen County: (The government has not published abridged life table)
  - We use the group classified by K-means with SDR as the indicator as the reference population for estimating the life expectancy in Kinmen County.
  - Then the estimation error using the Haybittle method and the life table method is half that of using Taiwan as a whole as the reference population.

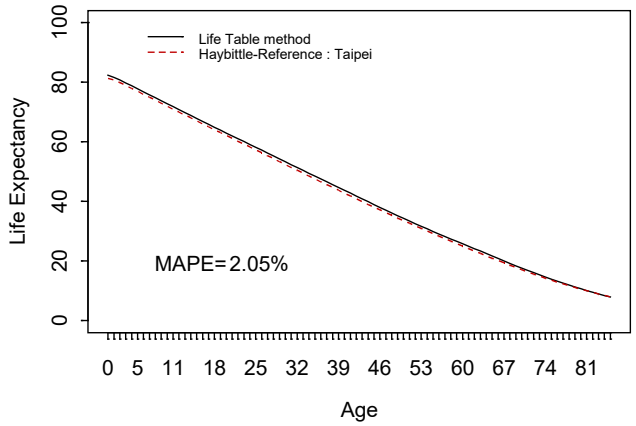


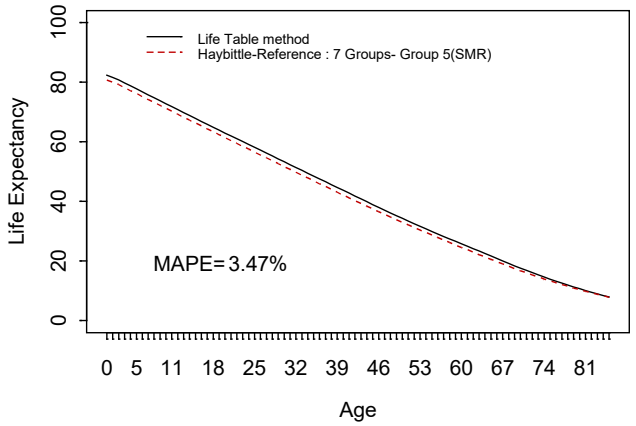
Application to 368 Townships

Life Expectancy Estimate

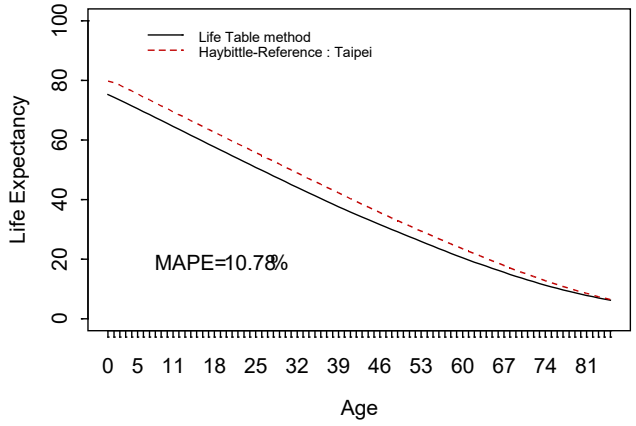
- The choice of the reference population affects the difference between the two methods of estimating the life expectancy, especially for townships with small populations. For example, the better reference population for Wanhua District is group 5 out of cluster 7 groups classification.

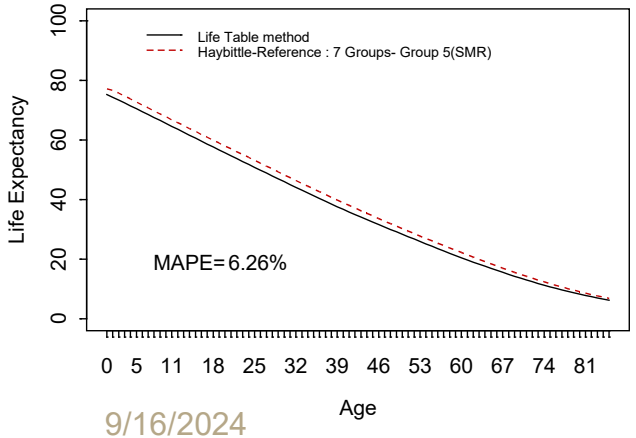
Songshan District Male Population: 96,393 (2018)





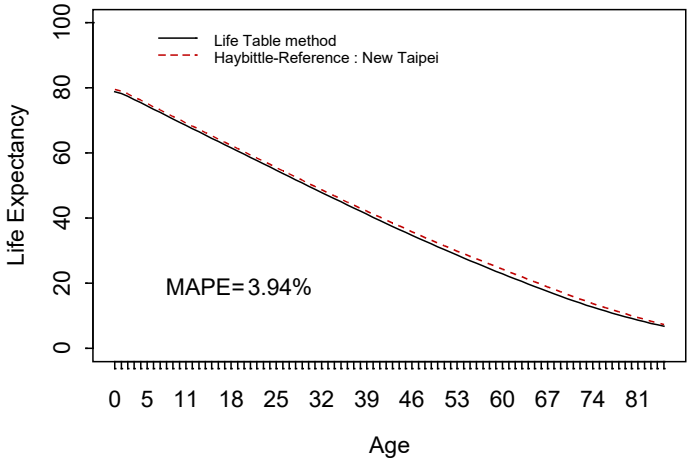
Wanhua District Male Population: 92,969 (2018)

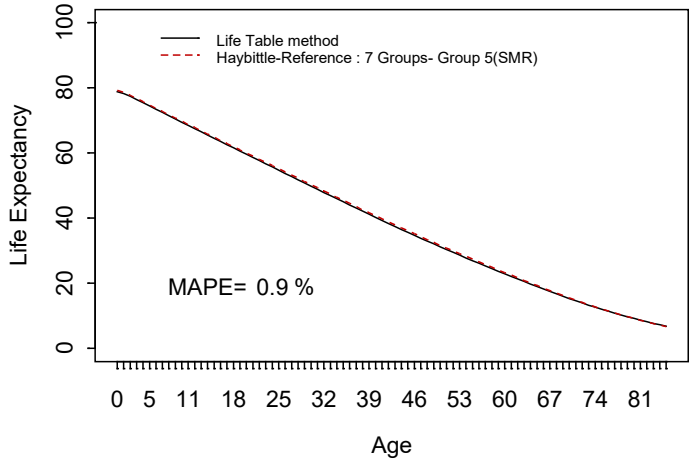




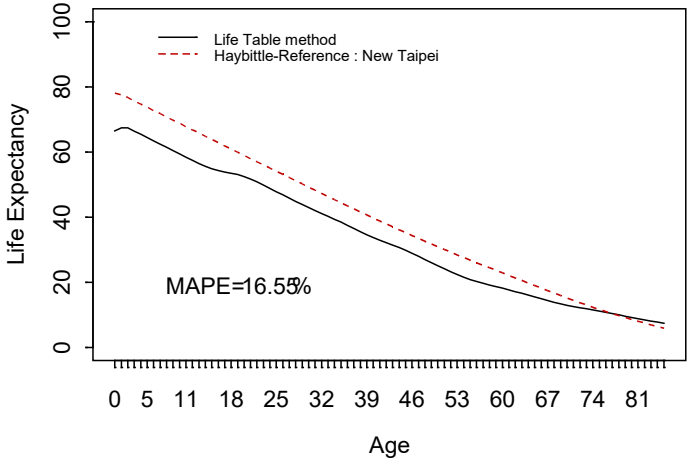
Life Expectancy Estimate

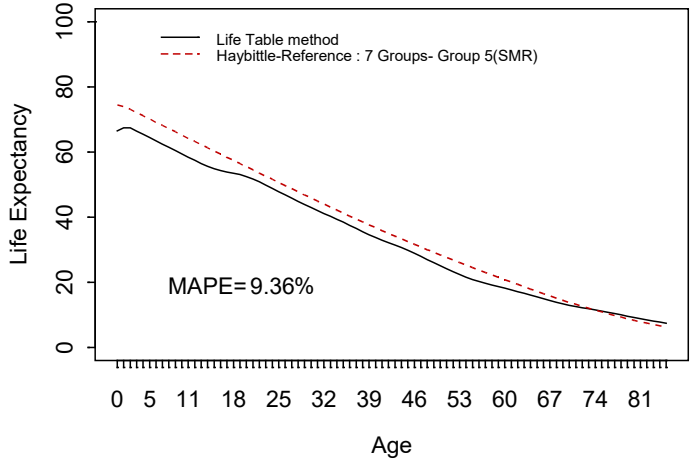
Banqiao District Male Population: 270,394 (2018)





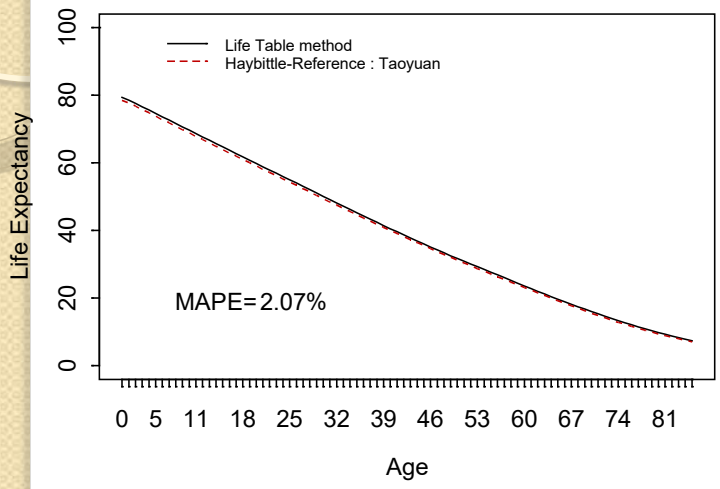
Shiding District Male Population: 4,220 (2018)

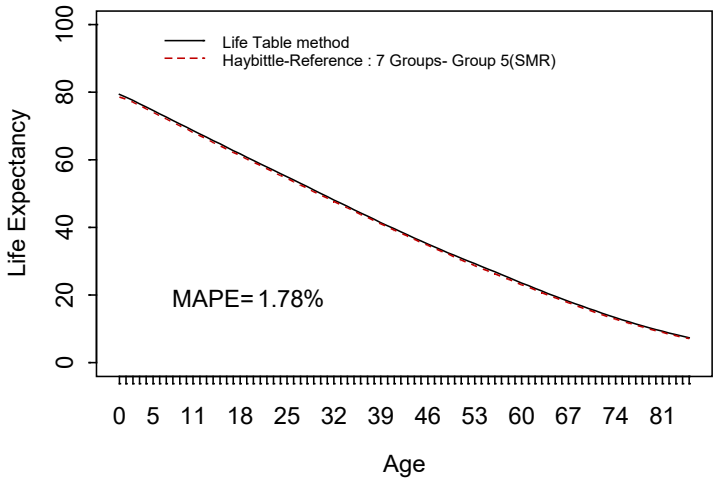




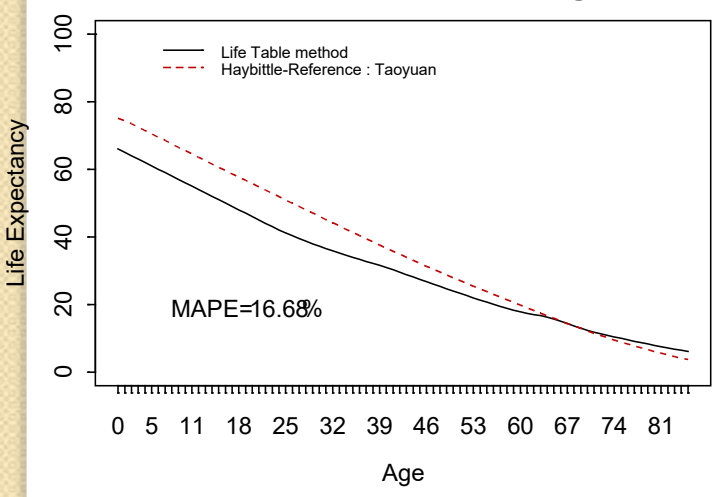
Life Expectancy Estimate

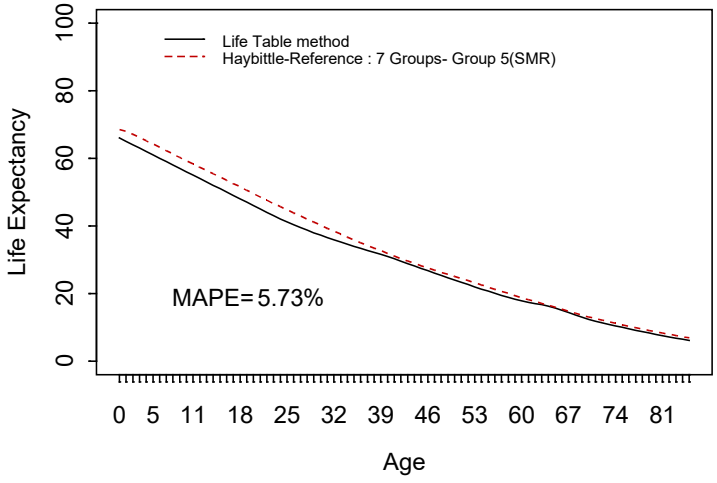
Zhongli District Male Population: 20,2613 (2018)





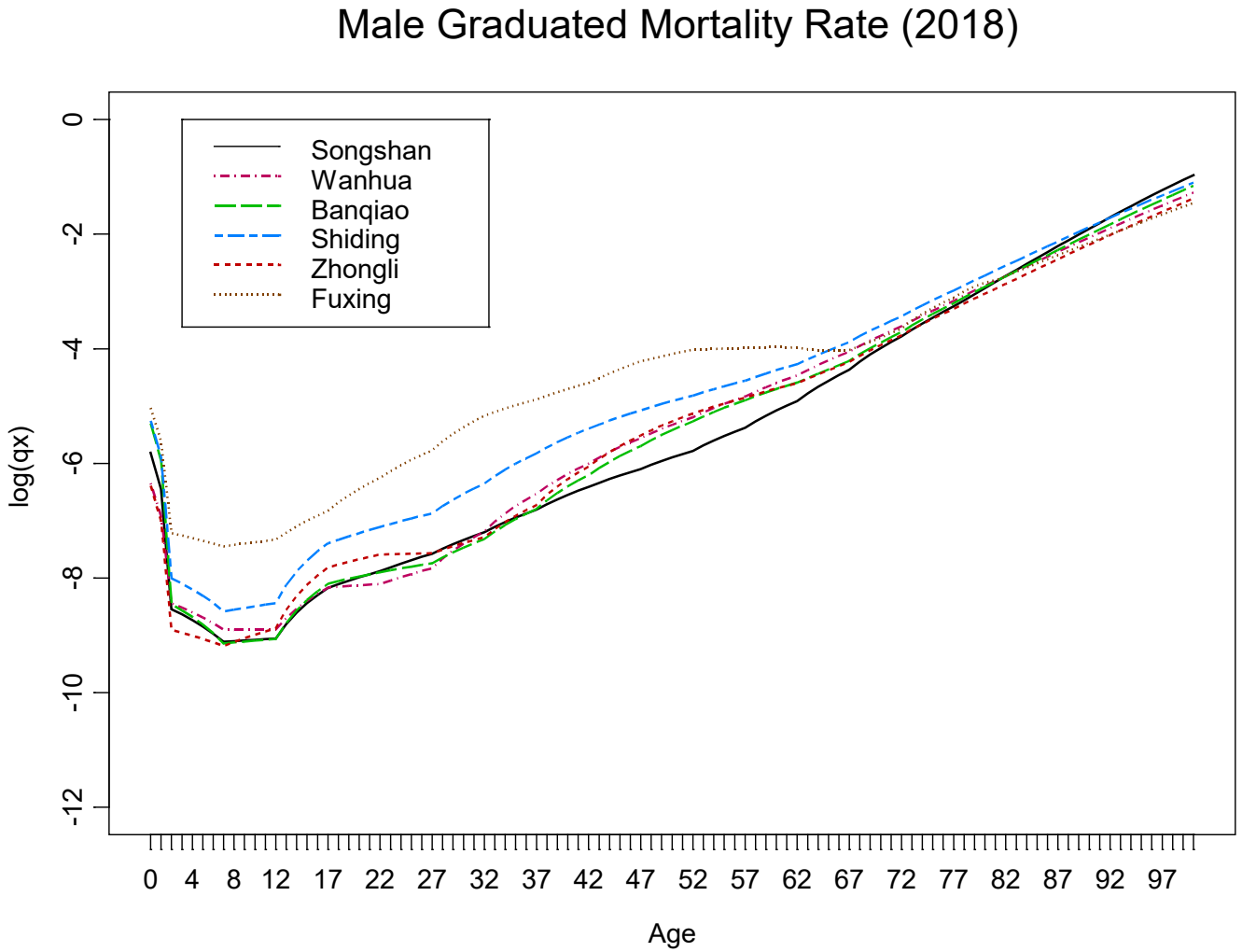
Fuxing District Male Population: 6,498 (2018)





Construct Districts/ Townships Life tables

- In practice, to construct a single-age life table, we calculate the single-age death rate using both interpolation and Gompertz extrapolation methods after applying Whittaker ratio or PSMR graduation.



Constructing Districts/ Townships Life Tables

- Comparison the Life expectancy at Birth of 2018 abridged life table with District/Township

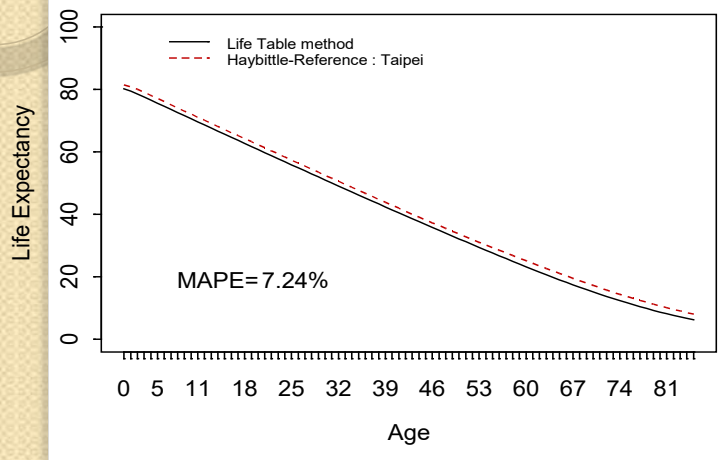
Ex: The life expectancy at birth of Wanhua District is (78.34 Year) less than Taipei City Abridged Life Table (80.93 Year) and Songshan District (80.2 Year)

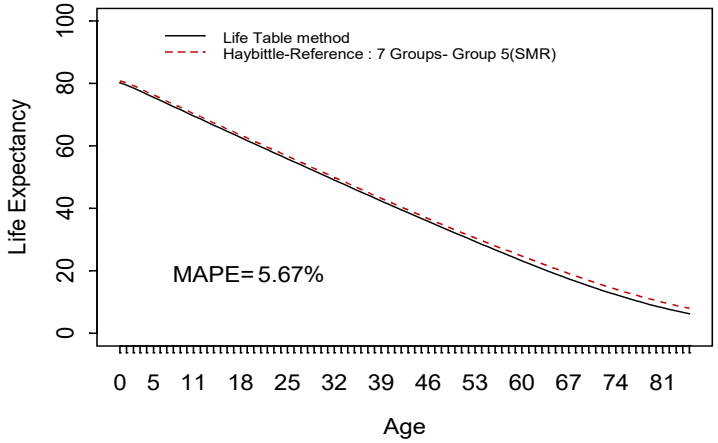
	Life expectancy at Birth
Taipei City 2018 abridged life table	80.93
Taipei Songshan District	80.20
Taipei Wanhua District	78.34
NewTaipei City 2018 abridged life table	78.33
NewTaipei Banqiao District	78.72
NewTaipei Shiding District	74.30
Taoyuan City 2018 abridged life table	77.86
Taoyuan Zhongli District	79.43
Taoyuan Fuxing District	68.36
Taichung City 2018 abridged life table	77.61
Taichung Xitun District	79.00
Taichung Heping District	75.61
Tainan City 2018 abridged life table	76.79
Tainan Anping District	76.61
Tainan Beimen District	70.12
Kaohsiung City 2018 abridged life table	76.13
Kaohsiung Xinxing District	79.69
Kaohsiung Maolin District	60.27

Life Expectancy Estimate

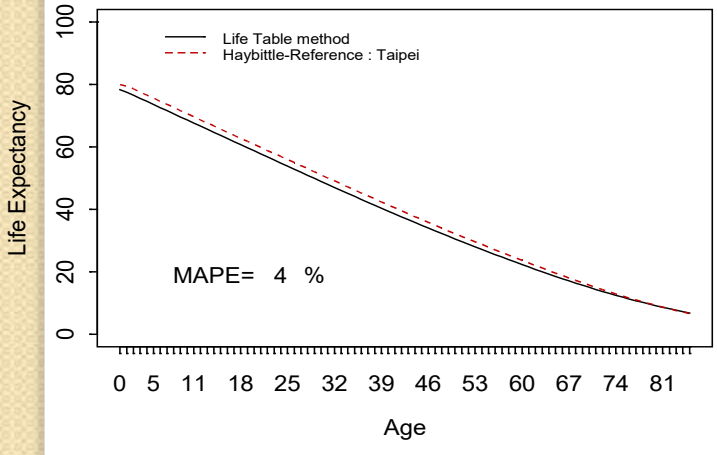
- After apply graduation method. There are similar results to those shown in the previous slides.

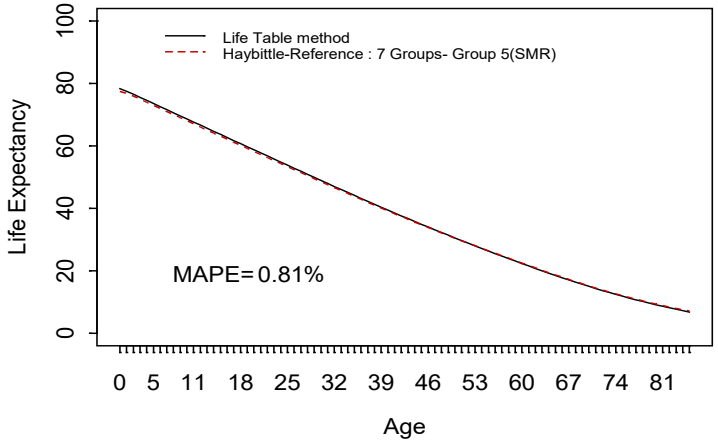
Songshan District Male Population: 96,393 (2018)





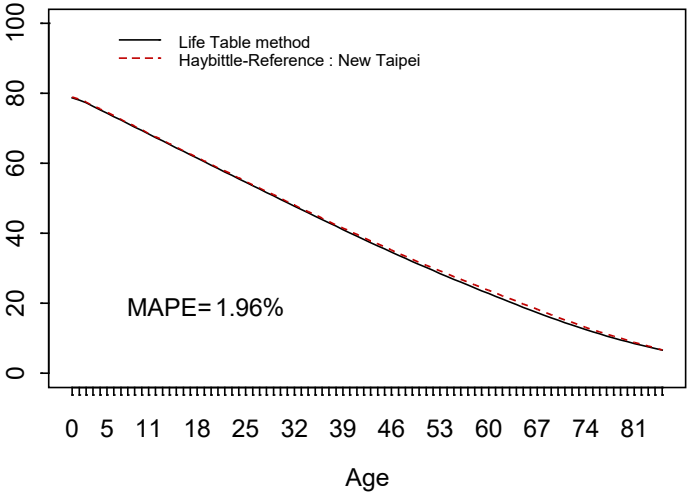
Wanhua District Male Population: 92,969 (2018)

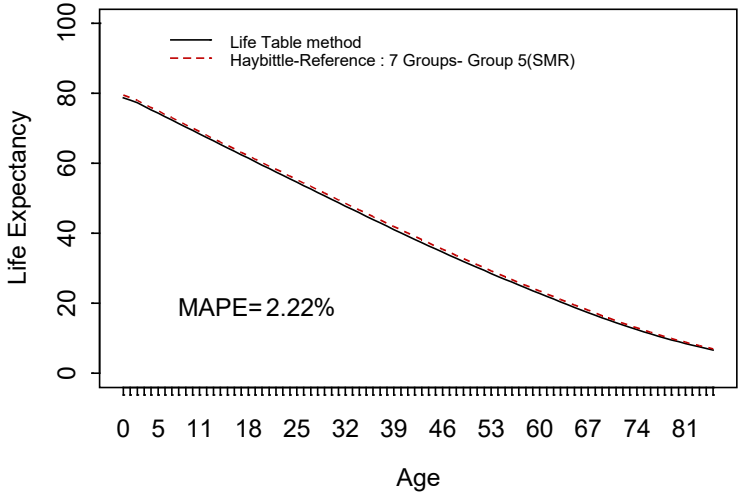




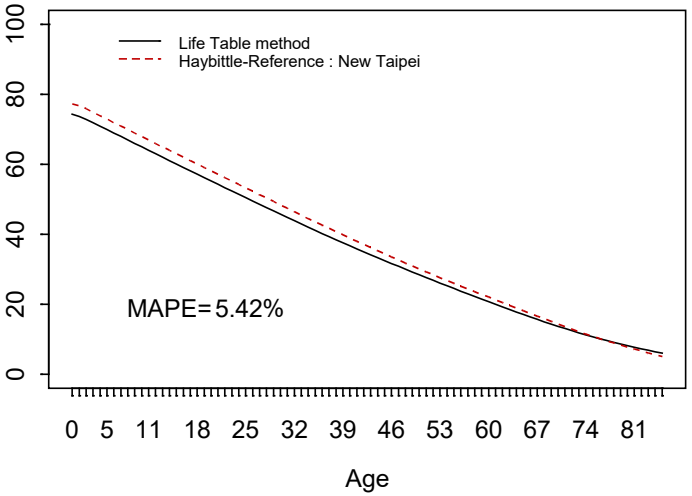
Life Expectancy Estimate

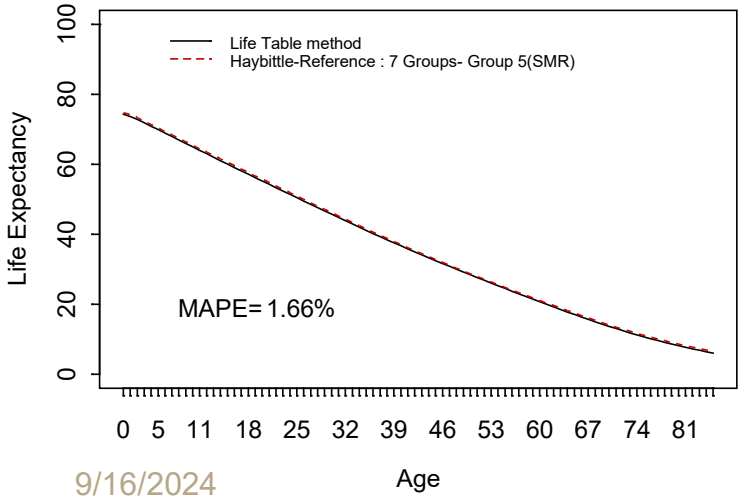
Banqiao District Male Population: 270,394 (2018)





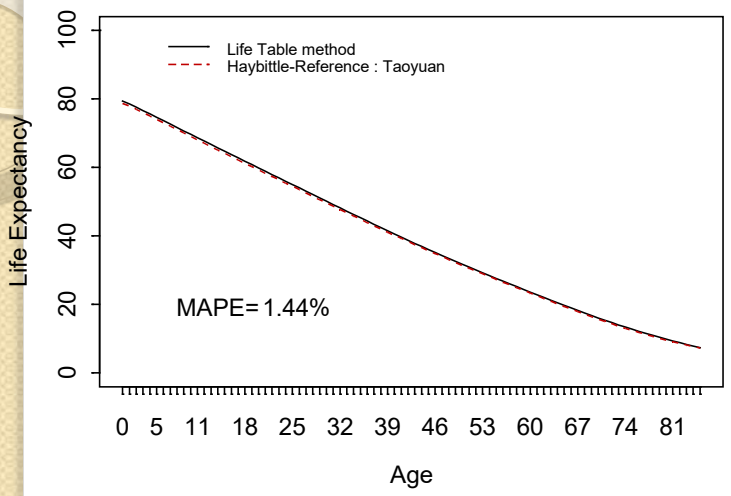
Shiding District Male Population: 4,220 (2018)

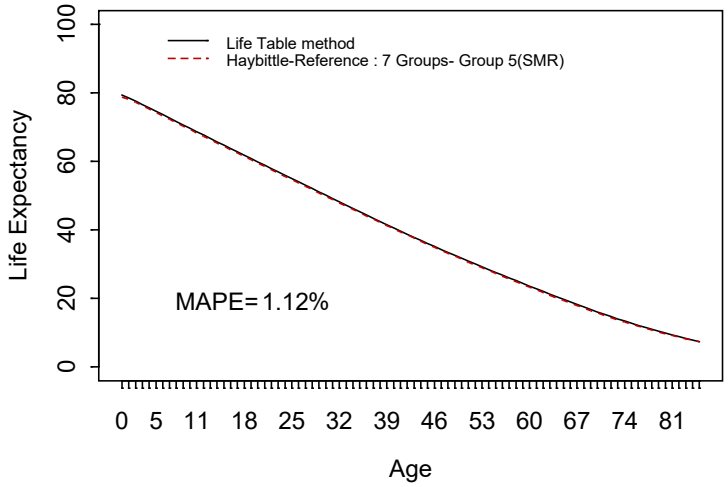




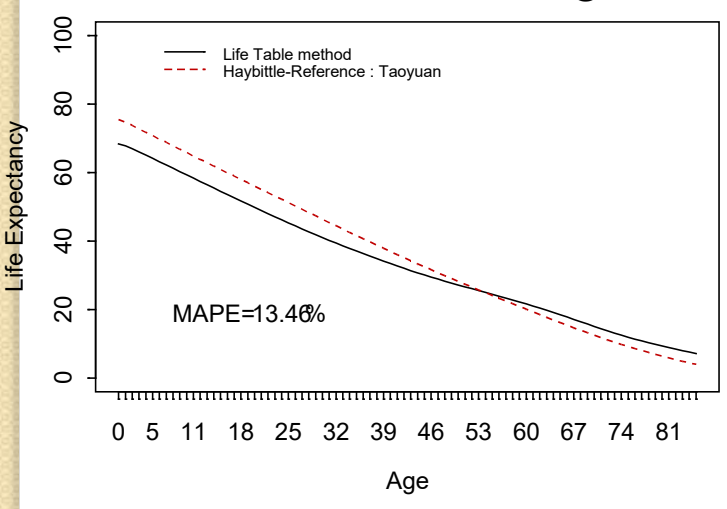
Life Expectancy Estimate

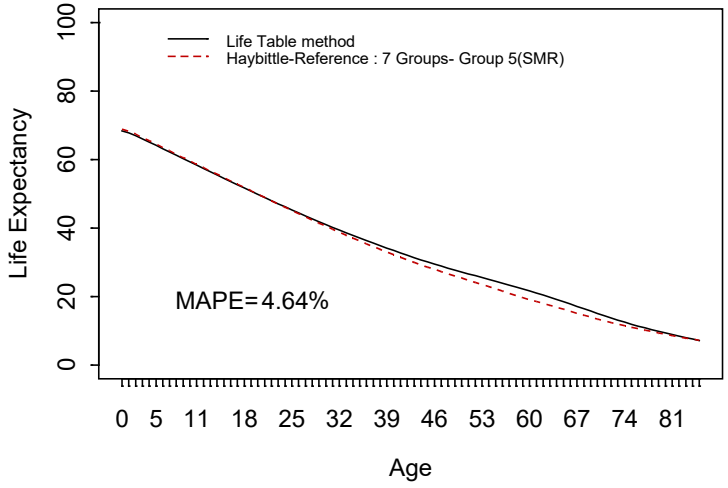
Zhongli District Male Population: 20,2613 (2018)





Fuxing District Male Population: 6,498 (2018)





Conclusion

- The idea of increasing sample size can be used in small area estimations.
- We propose using SMR or SDR as the indexes, and K-means is an appropriate clustering method for our analysis.
- Lee (2003) heterogeneity index could be used to determine whether the reference populations is suitable for small populations.
- If there are too many age groups with a death count of 0, the estimation error will be larger. Ex: In Jinmen Wuqiu Township, there are only 347 male residents, and on average(2007-18), there are only 1.6 deaths per year.
- Whitter ratio method is suitable for small population than PSMR
- We found that selecting a homogeneous reference population can reduce estimation errors
- We propose Haybittle (1998) Method, using the SMR into a linear model through suitable reference populations, is a easy and feasible alternative for estimating life expectancy of small areas.
- Utilize the SDR/SMR directly or combine socioeconomic indicators also can predict life expectancy.

Thank you for your attention.