

Test of Ordered Multivariate Discrete Selection Model for Average Life Expectancy

Jiwei Liu

College of Quantitative and Technological Economics, University of Chinese Academy of Social Sciences, Beijing, China

Email: liujiwei@ucass.edu.cn

How to cite this paper: Liu, J.W. (2022)

Test of Ordered Multivariate Discrete Selection Model for Average Life Expectancy. *Journal of Applied Mathematics and Physics*, 10, 261-269.

<https://doi.org/10.4236/jamp.2022.102020>

Received: January 12, 2022

Accepted: February 7, 2022

Published: February 10, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

At present, there are significant regional differences in average life expectancy among countries in the world. Not only is there a great disparity in average life expectancy, but also the gender difference is positive and negative, and is distributed in a bipolar distribution of “long life in rich countries and short life in poor countries”. This paper analyzes the factors affecting the life grade by using the ordered multivariate discrete selection model and combined with the average life expectancy data of countries all over the world in 2017. The test results show that: 1) The growth of per capita GDP, elderly dependency ratio and the proportion of people using at least basic drinking water services can effectively improve the level of life expectancy; 2) The birth rate has an inhibitory effect on the average life expectancy; 3) Through model comparison, probit model is more suitable for the analysis of this kind of problems than logit model, and the properties of the obtained model are better.

Keywords

Average Life Expectancy, Multivariate Discrete, Ordered Model, Life Grade Prediction

1. Introduction

Average life expectancy is a comprehensive index to measure the social and economic development level and medical and health service level of a country or region. It refers to the average number of years that can survive after the exact age of X under a certain age-specific mortality level. It is an index to measure the health level of residents of a country, nationality and region. It can reflect the quality of life of a society. Social and economic conditions and the level of health care limit people's life expectancy. Therefore, the length of human life span va-

ries greatly in different societies and different periods; At the same time, due to personal differences such as physique, genetic factors and living conditions, everyone's life span varies greatly. It is an important part of the United Nations Millennium Development Goals and the basis for calculating international comparative indicators such as population quality index (PQLI) and human development index (HDI) [1] [2] [3]. This indicator is closely related to gender, age and race, so it often needs to be calculated separately [4]. Average life expectancy is the most commonly used indicator of life expectancy. It shows the average number of years that the newborn population is expected to survive. It is an important indicator to measure the health status of the population. The length of service life is restricted by two aspects [5]. On the one hand, socio-economic conditions and health care level limit people's life span, so the length of life span varies greatly in different societies and different periods. On the other hand, due to personal differences such as physique, genetic factors and living conditions, everyone's life span is very different. Therefore, although it is difficult to predict the life expectancy of a specific person, it can be calculated and informed by scientific methods that under a certain death level, the average number of years that each person is expected to survive at birth. This is the average life expectancy of the population.

The current conventional method does not consider economic and social factors in the selection of indicators. Therefore, this paper selects indicators: total, GDP per capita, Current Health Expenditure per capita, Age Dependency Ratio, Birth Rate, crud, Hospital Beds, Population Density, and People Using at least basic drinking water services as independent variables to improve the prediction accuracy. At the same time, AA is used as the prediction model in order to obtain better analysis results.

2. Introduction of Ordered Multivariate Discrete Selection Model

The econometric model used in the empirical test of this paper is the ordered multivariate discrete choice model, which is an econometric model suitable for microanalysis [6] [7] [8]. There are two differences between the ordered multivariate discrete selection model and the binary discrete selection model: one is that there are three or more discrete choices; Second, discrete selection has a certain order. The following is a brief introduction to the ordered multivariate discrete selection model (hereinafter referred to as the ordered model). In the ordered model, the observed value y as the explained variable represents the ordered result or classification result, and its value is an integer. Explanatory variable X_i is a variety of factors that may affect the order of explained variables. X_i can be a set of multiple explanatory variables, that is, vectors [9].

The general form of ordered model is:

$$y_i^* = x_i' \beta + \varepsilon \quad (1)$$

where y_i^* is the hidden variable, x_i' is the set of explanatory variables, β is

the parameter to be estimated, and ε is the random disturbance. The implicit variable has no observed value, which can be regarded as utility according to economic theory, and its size can be measured by numerical value; Secondly, the explained variable (implicit variable) under this method can be explained linearly with explanatory variables, so it has reasonable practical meaning. After sorting y_i , according to its different thresholds, the probability of each y_i observation value is determined by the following formula:

$$\begin{aligned}
 y_i &= \begin{cases} 0 & \text{if } y_i^* \leq \gamma_1 \\ 1 & \text{if } \gamma_1 \leq y_i^* \leq \gamma_2 \\ \vdots & \\ n & \text{if } \gamma_n \leq y_i^* \end{cases} \\
 P(y_i = 0 | x_i, \beta, \gamma) &= F(\gamma_1 - x_i' \beta) \\
 P(y_i = 1 | x_i, \beta, \gamma) &= F(\gamma_2 - x_i' \beta) - F(\gamma_1 - x_i' \beta) \\
 P(y_i = 2 | x_i, \beta, \gamma) &= F(\gamma_3 - x_i' \beta) - F(\gamma_2 - x_i' \beta) \\
 &\vdots \\
 P(y_i = n | x_i, \beta, \gamma) &= 1 - F(\gamma_n - x_i' \beta)
 \end{aligned} \tag{2}$$

where F is the cumulative distribution function of random disturbances in Equation (1). If probit model is selected, F is the standard normal distribution function [10]. If you choose the logit model, F is the logical distribution function. From this, we can see that what the ordered model estimates is actually the probability that each observation y falls into different intervals. Secondly, γ is the limit point estimated together with the regression coefficient. By maximizing the following log likelihood function, γ and the regression coefficient results can be obtained.

$$\begin{aligned}
 L(\beta, \gamma) &= \sum_{y_i=0} \log(P(y_i = 0 | x_i, \beta, \gamma)) + \sum_{y_i=1} \log(P(y_i = 1 | x_i, \beta, \gamma)) \\
 &\quad + \dots + \sum_{y_i=n} \log(P(y_i = n | x_i, \beta, \gamma))
 \end{aligned} \tag{3}$$

3. Analysis of Regional Differences in Average Life Expectancy among Countries in the World

The data used in this paper are from World Bank WDI (World Development Indicator) Database

(<https://datatopics.worldbank.org/world-development-indicators/#archives>).

The data of 119 countries (regions) in 2017 were collected, including the following 8 variables. The variables are shown in **Table 1**.

Where, defined life level (LD):

- Level 5 (high life)—over 80 years old, $80 < \text{LIFE}$;
- Level 4 (medium and high life)—76 - 80 years old, $75 < \text{LIFE} \leq 80$;
- Level 3 (medium life)—71 - 75 years old, $70 < \text{LIFE} \leq 75$;
- Level 2 (middle and low life)—66 - 70 years old, $65 < \text{LIFE} \leq 70$;
- Level 1 (low life)—under 65 years old. $\text{LIFE} \leq 65$.

Table 1. Variables of average life expectancy of countries in the world in 2017.

Variables	Description
LIFE	Life expectancy at birth, total (years)
GDPP	GDP per capita (current US\$)
HEP	Current Health Expenditure per capita (current US\$)
ADR	Age Dependency Ratio, old (% of working-age population)
BR	Birth Rate, crude (per 1000 people)
HB	Hospital Beds (per 1000 people)
PD	Population Density (people per sq. km of land area)
PUW	People Using at least basic drinking water services (% of population)

Note: Since the latest data of some indicators and some regions are up to 2017, the data of 2017 is selected in this paper.

3.1. Basic Statistics of Data

From the WDI Database, in 2017, the world average life expectancy was 72 years. Japan's highest age is 84, Nigeria's lowest age is 54, and China's age is 76, ordered 52nd among 119 countries and regions. The world average life expectancy grades are artificially divided, so it is difficult to intuitively judge the significance of indicators and the rationality of scores, and the ordered model can be easily solved. According to the quantitative calculation results, we can not only get the significance level of each index, but also get the influence degree of each index on the rating, which is conducive to the objective evaluation of the credit rating system and point out the direction for the index adjustment and score adjustment. Considering the factors affecting life expectancy, the Spearman correlation coefficient test was conducted for the LD variable of life grade with per capita GDP, per capita health expenditure, elderly dependency ratio, birth rate, number of hospital beds, population density and people using at least basic drinking water services (accounting for % of the population). The results are shown in **Table 2**.

According to the results in **Table 2**, the variable LD has strong correlation with GDPP, HEP, ADR, BR, HB, PD and PUW, and passes the sig significance level. Therefore, LD can be selected as the explained variable, and GDPP, HEP, ADR, BR, HB, PD and PUW can be selected as the explained variable.

Perform collinearity test on the explained variables and check VIF. The results are shown in **Table 3**.

The data results in **Table 3** show that there is no obvious collinearity among explanatory variables, which is the premise to obtain better regression results.

3.2. Analysis of Influencing Factors of Average Life Expectancy based on Ordered Model

In order to study the influencing factors of regional differences in average life expectancy around the world, LD is used as the explained variable, GDPP, HEP,

Table 2. Spearman correlation coefficient test.

LD	GDPP	HEP	ADR	BR	HB	PD	PUW
Correlation coefficient	0.82	0.83	0.68	-0.71	0.40	0.19	0.70
Sig.	0.00	0.00	0.00	0.00	0.00	0.04	0.00

Table 3. Collinearity test of explanatory variables.

	Tolerance	VIF
GDPP	0.135	7.391
HEP	0.127	7.844
ADR	0.283	3.534
BR	0.201	4.977
HB	0.577	1.733
PD	0.866	1.154
PUW	0.357	2.804

ADR, BR, HB, PD and PUW are selected as the explained variables, and a multivariate ordered discrete model (order) is established. Assuming that the error term is normal distribution (probit) and logical distribution (logit), regression is carried out respectively. Using the econometric software Eviews, the ordered multivariate discrete selection model can be easily estimated, and the results are shown in **Table 4** and **Table 5**.

The model regression results under two different distribution assumptions show that the explanatory variables of HEP, HB and PD are not significant. That is, the coefficient value of these three variables is very likely to be zero. After removing these three explanatory variables, the new estimation results are shown in **Table 6** and **Table 7**.

Comparing the fitting results before and after, after removing the three variables with insignificant interpretation effect, the coefficient estimates of the other variables change slightly, which shows that HEP, HB and PD have little impact on the ordered model and can be eliminated. Because the Hannan Quinn criterion, SC, AIC and log likelihood of probit model are smaller, and the P value corresponding to LR statistic and AVG. log likelihood are larger, the probit model established in this paper is better [11].

The fitting diagram and residual diagram of the estimation results of the new ordered model are shown in **Figure 1**.

The standard equation form of the ordered model is as follows:

$$I_LD = 0.00006 \cdot GDPP + 0.04662 \cdot ADR - 0.06636 \cdot BR + 0.05254 \cdot PUW \quad (4)$$

where I_LD in Equation (1) is the implicit variable of the ordered model, that is, $x'_i\beta$ in theoretical models (1) and (2). Linear interpretation is made by the explanatory variables on the right of the equal sign of the equation. In the next 10 equations, $cnorm$ represents the normal distribution function. The parameters

Table 4. Multivariate ordered discrete model under different distributions.

Variables	Probit				Logit			
	Coefficient	SE	Z	P	Coefficient	SE	Z	P
GDPP	0.00	0.00	2.95	0.00	0.00	0.00	2.29	0.02
HEP	0.00	0.00	-0.36	0.72	0.00	0.00	-0.09	0.93
ADR	0.07	0.02	2.97	0.00	0.12	0.05	2.57	0.01
BR	-0.07	0.04	-1.97	0.05	-0.13	0.06	-2.02	0.04
HB	-0.11	0.06	-1.82	0.07	-0.20	0.11	-1.77	0.08
PD	0.00	0.00	1.28	0.20	0.00	0.00	1.22	0.22
PUW	0.05	0.02	2.42	0.02	0.09	0.04	2.44	0.01
Limit Points								
LIMIT_2: C(8)	1.77	2.26	0.78	0.44	2.89	3.90	0.74	0.46
LIMIT_3: C(9)	2.93	2.31	1.27	0.20	5.08	4.01	1.27	0.21
LIMIT_4: C(10)	5.09	2.31	2.20	0.03	8.80	4.04	2.18	0.03
LIMIT_5: C(11)	7.17	2.34	3.06	0.00	12.46	4.11	3.03	0.00

Table 5. Model performance test.

	Probit		Logit				
Pseudo R-squared	0.50	Akaike info criterion	1.64	Pseudo R-squared	0.49	Akaike info criterion	1.65
Schwarz criterion	1.89	Log likelihood	-86.46	Schwarz criterion	1.91	Log likelihood	-87.20
Hannan-Quinn criter.	1.74	Restr. log likelihood	-171.69	Hannan-Quinn criter.	1.75	Restr. log likelihood	-171.69
LR statistic	170.45	Avg. log likelihood	-0.73	LR statistic	168.97	Avg. log likelihood	-0.73
Prob(LR statistic)	0.00			Prob(LR statistic)	0.00		

Table 6. Multivariate ordered discrete model under different distributions.

Variables	Probit				Logit			
	Coefficient	SE	Z	P	Coefficient	SE	Z	P
GDPP	0.00006	0.00	5.05	0.00	0.00010	0.00	4.57	0.00
ADR	0.04662	0.02	2.46	0.01	0.08531	0.04	2.40	0.02
BR	-0.06636	0.03	-2.01	0.04	-0.11405	0.06	-1.98	0.05
PUW	0.05254	0.02	2.58	0.01	0.09381	0.04	2.62	0.01
Limit Points								
LIMIT_2: C(5)	2.10	2.21	0.95	0.34	3.84	3.76	1.02	0.31
LIMIT_3: C(6)	3.28	2.26	1.45	0.15	6.02	3.90	1.54	0.12
LIMIT_4: C(7)	5.37	2.27	2.37	0.02	9.64	3.93	2.45	0.01
LIMIT_5: C(8)	7.35	2.29	3.21	0.00	13.14	4.00	3.29	0.00

Table 7. Model performance test.

	Probit		Logit	
Pseudo R-squared	0.48	Akaike info criterion 1.63	Pseudo R-squared	0.48 Akaike info criterion 1.64
Schwarz criterion	1.82	Log likelihood -89.13	Schwarz criterion	1.83 Log likelihood -89.79
Hannan-Quinn criter.	1.71	Restr. log likelihood -171.69	Hannan-Quinn criter.	1.72 Restr. log likelihood -171.69
LR statistic	165.11	Avg. log likelihood -0.75	LR statistic	163.80 Avg. log likelihood -0.75
Prob(LR statistic)	0.00		Prob(LR statistic)	0.00

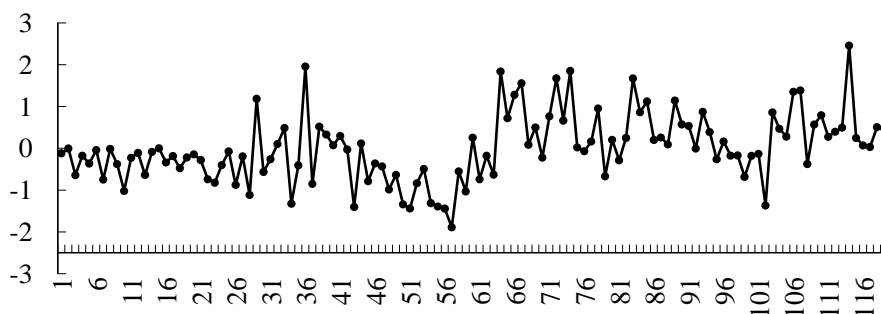


Figure 1. Residual diagram after HEP, HB and PD are removed.

Table 8. Effects of variables on life expectancy.

Variable	Coefficient	Effect
PUW	0.05254	Favorable
ADR	0.04662	Favorable
GDPP	0.00006	Favorable
BR	-0.06636	Unfavorable

in the equation give the boundary points that divide the normal distribution into five intervals, which are determined by LD_1 to LD_10 these 10 equations give probability distributions with ratings ranging from 1 to 5:

$$\begin{aligned}
 LD_1 &= @CNORM(2.10 - I_LD) \\
 LD_2 &= @CNORM(3.28 - I_LD) - @CNORM(2.10 - I_LD) \\
 LD_3 &= @CNORM(5.37 - I_LD) - @CNORM(3.28 - I_LD) \\
 LD_4 &= @CNORM(7.35 - I_LD) - @CNORM(5.37 - I_LD) \\
 LD_5 &= 1 - @CNORM(7.35 - I_LD)
 \end{aligned}
 \tag{5}$$

From the probit model, the regression results of each explanatory variable show that the GDPP, ADR and PUW coefficients are positive, that is, they can improve the average life expectancy. The BR coefficient is negative, which means that the average life expectancy can be reduced. The results are sorted from high to bottom (and from promotion to inhibition) as shown in **Table 8**.

The greater the PUW, ADR and GDPP, the greater the improvement of average life expectancy, and the improvement effect of PUW is the most obvious (its

coefficient is 0.05254). The absolute value of each variable coefficient reflects the impact of each index in the credit rating system. BR is negative, but it has the greatest impact on the average life expectancy.

4. Conclusions

Based on the average life expectancy data of countries around the world in 2017, the factors affecting the life grade are analyzed. This paper draws the following conclusions: probit model has better model superiority than logit model in this field. Per capita health expenditure, the number of hospital beds and population density have no significant impact on average life expectancy; The per capita GDP, the elderly dependency ratio, the population ratio using at least basic drinking water services and the birth rate have passed the significance test, which has a significant effect on the average life expectancy. Among them, the per capita GDP, the elderly dependency ratio and the population ratio using at least basic drinking water services have a positive effect on improving the life expectancy of the population. However, the increase in the birth rate will lead to the decline of the average life expectancy.

Combined with the above conclusions, China pays more attention to the per capita problem while increasing the overall economic growth. Any figure other than 1.3 billion will become very small. There is a long way to go to improve the per capita GDP. We need to vigorously develop the real economy, strengthen and optimize the real economy, and build a “double cycle” new development pattern. At the same time, on the issue of aging, improve the composition structure of the working population and the elderly care service system, give further play to the institutional advantages of the whole country in capital construction and other links, and expand and strengthen the infrastructure represented by new infrastructure; In terms of population policy, we should change the direction of population policy in different development periods of the demographic dividend period, in order to achieve a stable population structure and reduce the social burden.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Gou, X.T. (2013) A Quantitative Analysis of International Differences in Life Expectancy and Their Influencing Factors. *Journal of Nanjing College for Population Programme Management*, **29**, 31-36.
- [2] Xu, J. (2010) Analysis of Chinese Population Policies' Effect. Ph.D. Thesis, Jilin University, Changchun.
- [3] Wei, J., Ni, X.M. and He, A.C. (2018) Study on the Relationship between Population Policy and Economic Growth in the Context of Aging. *Study on the Relationship between Population Policy and Economic Growth in the Context of Aging*, **38**,

337-350.

- [4] Wang, X.L. (2012) Economic Influencing Factors of World Population Life Expectancy and Its Enlightenment to China. *The 9th Postgraduate Academic Forum*, Beijing, 21 June 2016, 191-197.
- [5] Yang, J.J. and Zhang, A.X. (2013) The Effects of Age Structure of Population and the Old-Age Insurance System Transition on Residents' Savings Rates. *Social Sciences in China*, **8**, 49-68.
- [6] Shen, L.F. and Zhang, W.W. (2021) Research on Factors Affecting the Low-Carbon Behaviors of Urban Residents Based on Discrete Choice Model. *Journal of Engineering Management*, **1**, 95-100.
- [7] Zhang, W.W. (2021) Research on Low-Carbon Behaviors of Urban Residents Based on Discrete Choice Model, Ph.D. Thesis, Central South University of Forestry and Technology, Changsha.
- [8] Chen, S.Q. (2021) Comprehensive Evaluation Analysis of Accounting Firms Based on Ranking Selection Model. *Economic Research Guide*, **32**, 136-138.
- [9] Wang, H. and Sheng, L.S. (2006) Testing the Client Credit Grading System Using Econometric Model. *The Journal of Quantitative & Technical Economics*, **23**, 138-147.
- [10] Smith, T.E. and Lesage, J.P. (2004) A Bayesian Probit Model with Spatial Dependencies. *Advances in Econometrics*, **18**, 127-160.
[https://doi.org/10.1016/S0731-9053\(04\)18004-3](https://doi.org/10.1016/S0731-9053(04)18004-3)
- [11] Akaike, H. (1974) A New Look at the Statistical Model Identification: *IEEE Transactions on Automatic Control*, **19**, 716-723.
<https://doi.org/10.1109/TAC.1974.1100705>