The Analysis of Factors Impacting on Population Aging in London

Programme: MSc Spatial Data Science and Visualisation

Department: Centre for Advanced Spatial Analysis

Student number: ucfnmzh

https://github.com/MengqingZhao/CASA0007_Assessment_ucfnmzh.git

- **Research question**

Population aging is the outcome of the development of social and economic at a certain stage. It is an important problem faced by many countries in the world and has become a global phenomenon. It refers to a dynamic process in which the number of young people in the total population decreases while the proportion of old people in the total population increases gradually. Generally speaking, population aging occurs first in developed countries. It is the inevitable result of rapid economic development, improvement of people's living standard, improvement of medical security system and scientific progress.

Similar to many other countries, age structure in the UK is changing towards elder ages. By 2050, one in four people in the UK is expected to be aged 65 or over, up from around one in five in 2019.The ageing structure of the UK population is driven by two factors, which are falling fertility rates and people living longer. Firstly, fertility rates are falling, with people having fewer and later children. At the same time, rising life expectancy means people are living longer and into older age. In addition, the population aged 65 and over is growing faster than any other age groups (Office for National Statistics, 2021). It brings a series of problems to the society, such as the shortage of labor and the increasing demand for pension, even impeding the development of social, political, and economic.

As a result, population aging is a serious problem that cannot be ignored and this paper is aimed to find the factors impacting on population aging. There are five factors selected, including GDP per capita, median income, population density, unemployment rate, and healthcare expenditure. London was chosen as the study area.

- **Literature Review**

At present, relevant scholars around the world have carried out a large number of analyses in different regions. Because some European countries and Japan have relatively serious population aging problem, there are a lot of academic research on the aging problem in Europe and Japan. European scholars (United Nations, 1956) were the first to study the influencing factors of population structure. Early studies on the effects of birth, death and migration on age structure have shown that declining birth rates have a much greater impact on population aging than declining death rates. Therefore, it can be concluded that one of the main reasons for the aggravation of the aging problem is the decrease of birth rate.

According to a number of studies, the direct factors of population aging are birth rate and death rate, but they are too macro and broad. In this paper, it is aimed to select some relatively detailed indexes for research. These two direct factors are determined by many indirect factors, so from these two aspects, in some studies, the indirect influencing factors of population aging have been also analyzed from the micro point of view. Research has found that public health expenditure and population density are

the most important factors affecting the proportion of the elderly population in China and both of them have positive effects on population aging, by establishing a multiple linear regression model (Chen, 2012). Based on a similar study by Yin (2015), it was shown that population density and the add values of tertiary industry are the main factors of the population aging in Anhui Province.

Based on the availability of data and reference to domestic and foreign literature, five factors were selected in this paper that can affect birth and death rates for quantitative analysis, which are GDP per capita, median income, population density, unemployment rate, and healthcare expenditure in London.

- **Presentation of Data**

Multiple datasets are collected to perform this analysis and the related part in each dataset has been extracted in Microsoft Excel. Here is the list of the extracted data from each dataset:

Dependent variable:

elder_proportion - The proportion of the population ages 65 and above in London between 2000-2017 (The World Bank, 2019), unit: % of total population.

Independent variables:

GDP - Gross domestic product (GDP) per capita in London between 2000-2017 (Office for National Statistics, 2019), unit: £ per capita.

median_income – The median personal incomes of tax payers by tax year in London between 2000-2017 (HM Revenue & Customs, 2018), unit: £ / year.

population_density – Population density in London between 2000-2017 (Greater London Authority, 2018), unit: population per square kilometre.

unemployment_rate - Unemployment rate in London between 2000-2017 (Office for National Statistics, 2020), unit: %.

healthcare_expenditure – Health care spending per capita in London between 2000-2017 (UK Public Spending, 2020), unit: £ per capita.

Summary Statistics:

	elder_ proportion	GDP	median_ income	population _density	unemployment _rate	healthcare_ expenditure
quantity	18	18	18	18	18	18
mean	16.71	41183.81	22416.67	5063.71	7.48	1901.55
median	16.31	42136.92	22550.00	5010.60	7.13	1860.59
min	15.89	29894.79	16600.00	4603.10	5.40	946.52
max	18.29	53388.10	28800.00	5663.60	9.88	2618.70
range	2.40	23493.30	12200.00	1060.50	4.48	1672.17
LQ	16.02	35521.29	19025.00	4741.50	6.92	1465.22
UQ	17.35	45683.52	25350.00	5344.18	8.50	2432.15
IQR	1.33	10162.23	6325.00	602.68	1.58	966.93
lo-tukey	14.02	20277.94	9537.50	3837.49	4.54	14.82

hi-tukey	19.35	60926.88	34837.50	6248.19	10.87	3882.55
lo-outliers	0	0	0	0	0	0
hi-outliers	0	0	0	0	0	0
variance	0.72	52114970.39	14336764.71	123341.74	1.56	339216.84
st. dev.	0.85	7219.07	3786.39	351.20	1.25	582.42

Table 1: Summary Statistics

According to Table 1, all data for each column are in Tukey fences and there is no outlier for each variable. The box plots are also shown in Figure 1.

Figure 1: Box plots of all variables

The box plots in Figure 1 also show that there is no outlier in this dataset.

The histogram of each variable is shown in Figure 2.

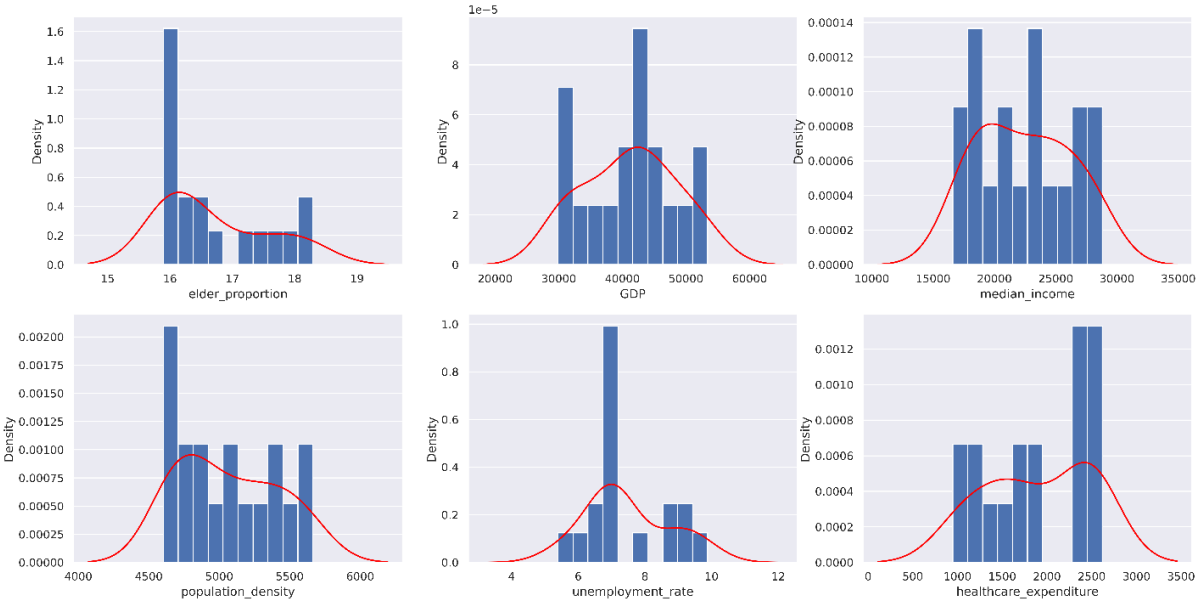

Figure 2: Histograms of all variables

According to the above six figures, because there are only 18 items in the dataset. This amount might be a little bit small and the above data distributions can be approximated as normally distributed. Therefore, there will be a linear relationship.

- **Explanation of methodology**

Because there are multiple factors, Ordinary Least Squares (OLS) Multiple Linear Regression was applied to analyze the correlation between the dependent variable and multiple independent variables. In this method, the above multiple factors, including GDP, median_income, population_density, unemployment_rate, and healthcare_expenditure, are combined to create a linear model for predicting the proportion of the aged population. Here is the target formula of the modelled relationship:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_0$$

Because there is no outlier in this dataset, the step of dealing with outliers is not need.
Before applying multiple linear regression, to make the model more stable and reliable, it is necessary to detect and eliminate multicollinearity between predictors by calculating Variance Inflation Factors (VIF). A threshold of 10 was used.
The class statsmodels.formula.api.ols was used in Python to fit a linear model by using Ordinary Least Squares.
After multiple linear regression, the modeled result should be checked. Tests of goodness of fit and significance are necessary by R-Squared and ANOVA F Test. If the R-Squared value of the regression model is high, the goodness of fit will be high. If p-value < 0.001, the linear relationship of this model will be significant.
Moreover, residual analysis can be used to detect non-linearity, unequal error variances, and outliers.

- **Presentation of results**

Here are the scatter plots of all predictor variables to the dependent variable:

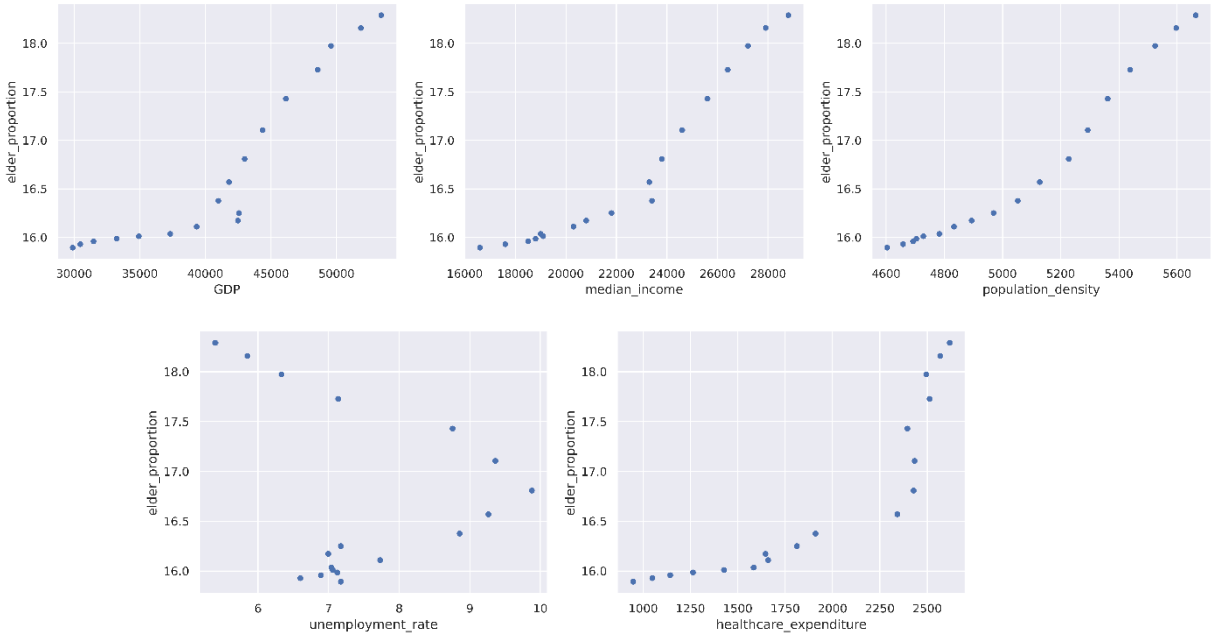

Figure 3: Scatter plots of all variables

Before applying multiple linear regression, it is necessary to detect and eliminate multicollinearity between predictors by calculating Variance Inflation Factors (VIF). A threshold of 10 was used.

features	VIF Factor
const	13001.7218
GDP	31.300871
median_income	105.637894
population_density	111.518479
unemployment_rate	5.083328
healthcare_expenditure	59.294003

Table 2: VIF before elimination

features	VIF Factor
const	77.657214
median_income	1.000443
unemployment_rate	1.000443

Table 3: VIF after elimination

In each round, if the highest VIF is larger than the threshold, remove the corresponding variable from the list. Therefore, population_density, healthcare_expenditure, and GDP are deleted in turn. As shown in Table 2 and 3, after variable selection, there are only two Predictor variables: 'median_income' and 'unemployment_rate'. This model becomes more stable and reliable.

Here is the result of OLS multiple linear regression:

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	12.9034	0.429	30.065	8.0546E-15	11.989	13.818
median_income	0.0002	1.32E-05	16.076	7.27062E-11	0	0
unemployment_rate	-0.1288	0.04	-3.212	0.006	-0.214	-0.043

Table 4: OLS regression results

According to Table 4, p-value of unemployment_rate is 0.006, larger than 0.001 and p-value of unemployment_rate is smaller than 0.001. It means that median_income factor has significant influence on electricity consumption but unemployment_rate does not. Additionally, the coefficient of median_income is positive. It means that median_income has a positive relation with elder_proportion.

Then, to evaluate the result, here is Regression Statistics.

R-squared	0.948
Adj. R-squared	0.941
F-statistic	135.5
Prob (F-statistic)	2.50E-10

Table 5: OLS regression statistics

According to Table 5, with the large R Square 0.948, the combined model explains about 94.8% of variance. It means that this model has high goodness of fit. The ANOVA F Test was used to check the significance of the model. From this table, we can find that p-value is much smaller than 0.001. It means that the linear relationship is significant.

Residual analysis can be used to detect non-linearity, unequal error variances, and outliers. Here are the regression plots for median_income and unemployment_rate.

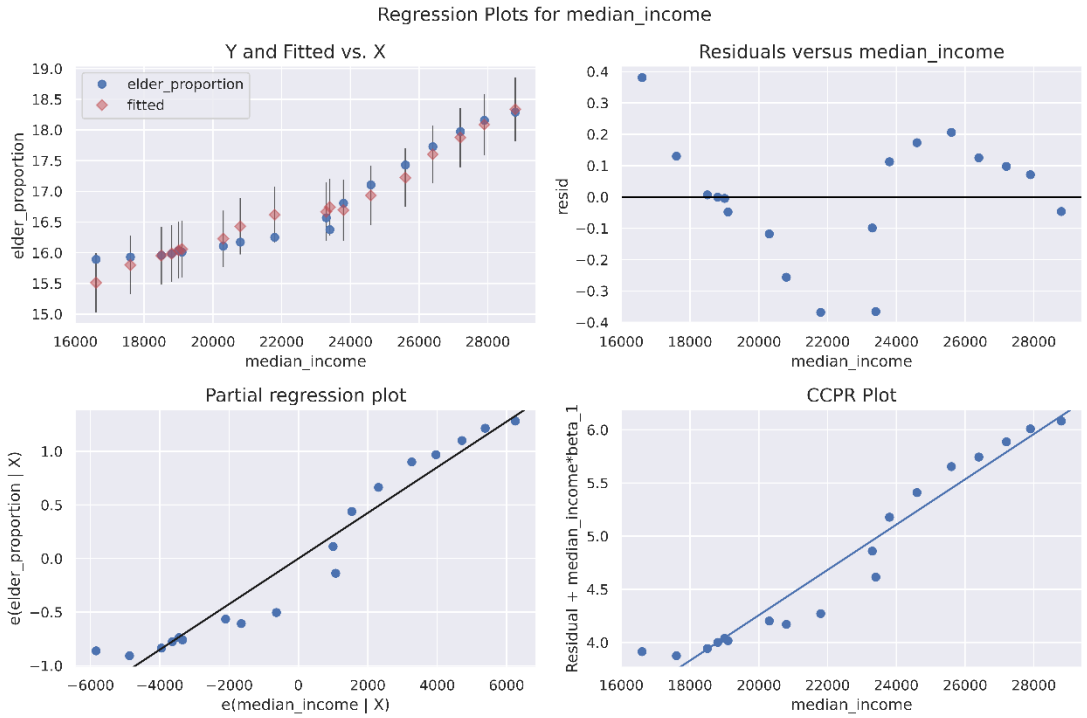

Figure 4: Regression plots for median_income

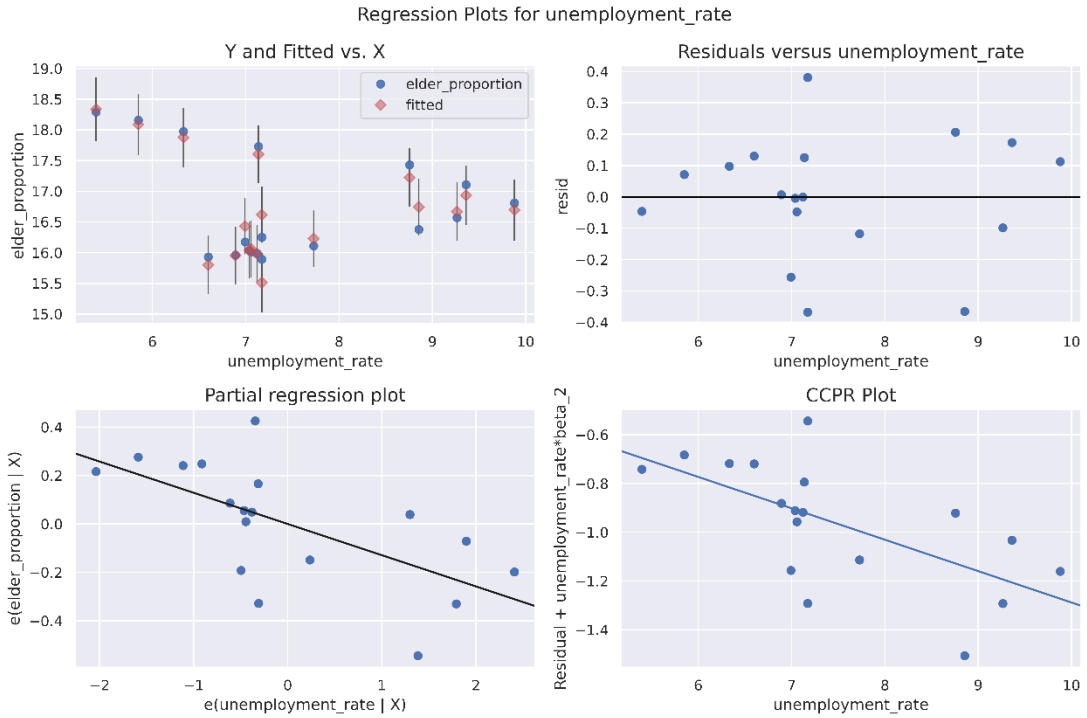

Figure 5: Regression plots for unemployment_rate

According to Figure 4, median_income may be not linearly related to elder_proportion because the residuals depart from 0 in a systematic manner. It would better to apply a non_linear model model between median_income and unemployment_rate.

According to Figure 5, there is a relationship between unemployment_rate and elder_proportion.

- **Discussion of results**

According to the above analysis, median income has positive correlation with the proportion of the elderly population with the most significant influence. Because of the multicollinearity between median income, GDP per capita, population density, and healthcare expenditure, these factors are positively related to the aged population.

When the median income or GDP per capita increases, it indicates that the quality of people's life in London is better, which is conducive to the extension of life expectancy.

The larger the population density is, the more people will live in a unit area. People's living space and resources will be limited, so the birth cost will be correspondingly higher, which will lead to the decline of the birth rate and the aggravation of the aging degree.

Health care expenditure reflects government investment in health services. The more money is invested, the more services and security people can enjoy, and the more conducive to the extension of human life.

- **Conclusions**

In this paper, five predictor factors are selected to establish a multiple linear regression model of population aging in London. The model has high goodness of fit and the linear relationship is significant. It also concluded that with the development of society and country, the aging of population is inevitable.

It should encourage people to have more children to increase the birth rate, which would have a positive effect on alleviating London's aging population. Meanwhile, the problem of supporting old people is extremely important. It is necessary to raise public awareness of the problem of ageing, appropriately extend the retirement age, improve the pension service system in London, and rationally plan the balance of income and expenditure of the pension fund.

(Word count: 1698 words)

References

Office for National Statistics (2021) *Overview of the UK population: January 2021* [Online]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/january2021#the-uks-population-is-ageing (Accessed: 19 January 2021).

United Nations (1956) 'The Aging of Populations and its Economic and Social Implications'. *New York: United Nations.* 7(6)

Chen, R. et al. (2012) 'A Research Based on Multiple Linear Regression Model about the Influence Factors of the Aging Population in China'. *Science & Technology Vision.* 2012(11), pp. 3-5

Yin, X. (2015) 'The Analysis of Factors Influencing Population Aging in Anhui Province', *Aging Research.* [Online] 2 (2), pp. 7 - 13. Available from: http://dx.doi.org/10.12677/AR.2015.22002 (Accessed: 19 January 2021)

The World Bank (2019) *Population ages 65 and above (% of total population) - United Kingdom.* [Online] Available from: https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS?locations=GB (Accessed: 19 January 2021)

Office for National Statistics (UK) (2019) *Gross domestic product (GDP) of London in the United Kingdom (UK) 2000-2018.* [Online] Available from: https://www.ons.gov.uk/file?uri=%2feconomy%2fgrossdomesticproductgdp%2fdatasets%2fregionalgrossdomesticproductallnutslevelregions%2f1998to2018/regionalgrossdomesticproductgdpallnutslevelregions.xlsx (Accessed: 19 January 2021)

HM Revenue & Customs (2018) Average Income of Tax Payers, Borough. [Online] Available from: https://data.london.gov.uk/dataset/average-income-tax-payers-borough (Accessed: 19 January 2021)

Greater London Authority (GLA) (2018) A Land Area and Population Density, Ward and Borough. [Online] Available from: https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough (Accessed: 19 January 2021)

Office for National Statistics (ONS) (2020) Unemployment Rate, Region. [Online] Available from: https://data.london.gov.uk/dataset/unemployment-rate-region (Accessed: 19 January 2021)

UK Public Spending (2020) United Kingdom Central Government and Local Authority Spending. [Online] Available from: https://www.ukpublicspending.co.uk/download_multi_year_1991_2021LOb_17c1li111mcn_10t (Accessed: 19 January 2021)