For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

Team Control Number
70232

Problem Chosen
E

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

2017
MCM/ICM
Summary Sheet

Proper urban planning can ensure that people have access to equitable and sustainable homes, resources and jobs. Nowadays, the world is rapidly urbanizing and by 2050, 2.5 billion people will be added to the urban population. Thus, it is essential for us to accurately evaluate the smart growth degree of a city and correctly predict the development degree of a city in the future. In this way, a smart growth plan can be made to ensure the sound development of a city and support the blooming population in the future.

To access the smart growth level of a city, we develop a Smart Growth Degree Evaluation Model, which based on cluster analysis and discrimination analysis in multivariate statistical analysis. Considering the three E's of sustainability, we choose 7 factors to describe the development status of a city. We select 15 cities in different development status as training set and divide them into 4 groups based on PAM Cluster Method. So we can get 4 levels of smart growth and the metric by using the discrimination analysis. Then we choose Seattle and Chong Zhou as test set and verify them by using the Mahalanobis distances. It demonstrates the accuracy and validity of our model.

Next, we develop the Growth Rate Prediction Model to estimate the growth of the city based on linear regression method. In this way, we can get the predictive growth rate of each factor. Then, we develop the best growth plan according to the Plans' Effect Model and calculate the optimal solution by using linear programming.

The Population Regression Model develops the relationship between population and the 7 factors we choose. By using the Lasso shrinkage method, we can eliminate collinearity of the factors and get the accurate prediction number of population in 2050. It clearly shows that our plan can support this level of population growth.

Taking coherent steps to work out the best growth plan and evaluate its effect is our main strength. We build models to evaluate and predict the smart growth of a city, and work out feasible plans for urban planning. Additionally, the adequate official data ensures the accuracy of our models.

Contents

1. Introduction

It is projected that by 2050, 66 percent of the world's population will be urban—this will result in a projected 2.5 billion people being added to the urban population. As the world is rapidly urbanizing, to make our cities sustainable becomes an essential task.

In order to do so, we should know how to measure the growth of our cities and determine what aspects should we take into consideration to develop a long large, sustainable growing plan for different cities. In addition, once we come out with a matric, we can study the current situation of any city. We implement these using the cluster analysis and the discriminant analysis methods.

Next, we develop another model to estimate a city's growth. It uses the linear regression method to stimulate the trend of the development of a city. Then we use the smart growth degree evaluation model and the growth rate prediction model to indicate that our growth plans are quite successful and rank the individual items in the plans.

To solve the increasing population problem by 2050, we build a population regression model using the shrinkage method. We find our plans support the booming population well.

2 Smart Growth Degree Evaluation Model

In order to define a matric to measure the success of smart growth of a city, we must cluster and analyze different growth level of cities. Then we can find the weakness of a city's development and decide what plan should be taken.

2.1 Basic Data

In consideration of the three E's of sustainability and the ten principles of smart growth, we choose the following seven aspects as indicators for a city's smart growth degree.

• Unemployment rate (UR) • Number of pension insurance (NPI) • Number of health insurance (NHI)	Socially Equity
• Annual air quality days (AAQD)	Environment condition
• Housing price (HP) • GDP per capita (GPC)	Economics
• Number of College Students (NCS)	Education

Table 1. Classification of Seven Aspects

It is clear that the data items can be classified into four different groups. We want the items show smart growth degree of cities from all aspects. Therefore, we choose social, natural and economic condition as well as education standard.

Then, we choose 15 different cities to be subjects. We obtain the data from National Bureau of Statistics of the People's Republic of China to do cluster analysis and discriminant analysis. The datasheets are as follows:

	GPC(￥)	AAQD (day)	HP(￥)	UR (%)	NPI(10^4 people)	NHI(10^4 people)	NCS(10^4 people)
Bei Jing	92210.14	168	18833	1.3	1392.6	1604.3	60.5
Shang Hai	89444.22	278	16787	4.1	1457.4	1678.5	50.7
……	……	……	……	……	……	……	……
Fo Shan	104215.91	234	8484	2.35	344.65	272.94	4.67
Chong Qing	47688.04	246	5519	3.5	825.5	3256.8	69.2

Table 2. Statistics of 15 Cities in 2014

Statistics of 15 cities in 2015 are listed in the appendix. As our study subject is the general trend of smart growth, we must work out the rate of the items' change. With the data in 2014 and 2015, we calculate the growth rate of each item of 15 cities to measure their smart growth:

	GPC(￥)	AAQD (day)	HP(￥)	UR (%)	NPI(10^4 people)	NHI(10^4 people)	NCS(10^4 people)
Bei Jing	0.158	0.107	0.202	0.077	0.023	0.033	-0.002
Shang Hai	0.151	-0.094	0.248	-0.024	0.025	0.024	0.010

……	……	……	……	……	……	……	……
Fo Shan	0.045	0.038	0.012	0.004	0.104	0.018	0.058
Chong Qing	0.102	0.146	-0.006	0.029	0.029	0.003	0.036

Table 3. Growth Rate of the Items

The complete tables of table2 and table3 can be seen in the appendix.

2.2 Assumptions

- The data we collected is reliable.
- The data is showed in annual average, we ignore the fluctuations in a year.
- The items we choose can indicate the smart growth of one city.

2.3 Cluster Analysis

We use the statistics of the above 15 cities (Table 3) to do cluster analysis. With the Average silhouette method, we can estimate the dispersion degree of every sample. Notice that the higher average silhouette width, the better clustering result:

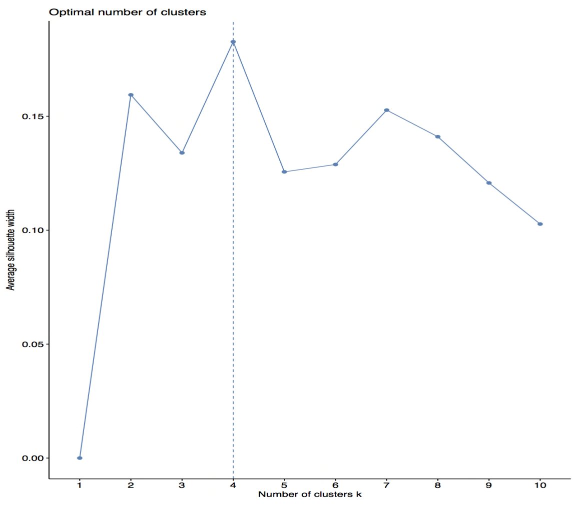

Figure 1. Average Silhouette

It can be seen from the figure above that the average silhouette width reaches the highest level when the number of clusters is four. Therefore, we divided the samples into four clusters using PAM clustering method.

After that, we use the PAM clustering method to divide the samples into different clusters and get the clustering result. To calculating the Manhattan distance as the cost between two points we use the following formula:

$$\text{cost }(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

Total cost is the summation of the cost of data object from its medoid in its cluster :

$$\text{Total cost} = \sum_{t=1} \sum_{i \in s_t} cost(x_i - y_t)$$

The aim is to minimize the total cost.
By using R, we can obtain the following figures, which show the average silhouette and clustering result.

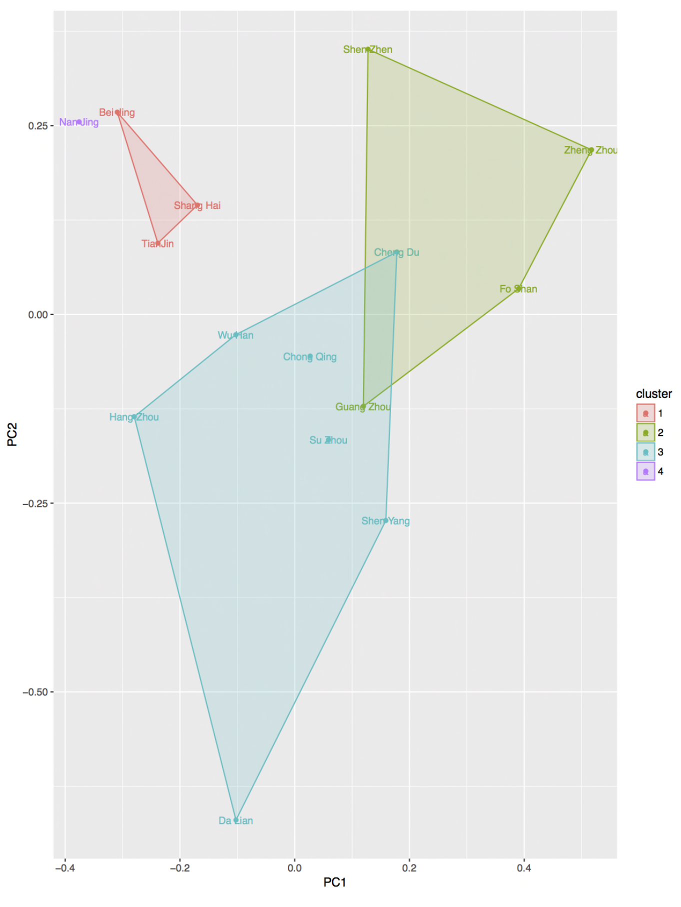

Figure 2. Clustering Result

From the next figure above, the clustering results are clear:

Clustering Result	**Cities**
A(G_1) Low level	Bei Jing, Shang Hai, Tian Jin
B(G_2) Middle level	Cheng Du, Wu Han, Chong Qing, Su Zhou …… Da Lian
C(G_3) High level	Shen Zhen, Fo shan, Guang Zhou, Zheng Zhou
D(G_4) Singular group	Nan Jing

Table 4. Clustering Result

If a city falls in class-C, its smart growth condition is quite good. The growth of cities in class-B is slower than those in class-C. We also find the cities in class-A is quite developed. However, the smart growth condition is not good because the growth rate of each item is low. Nan Jing alone falls in class-D because its particularity that its environment pollution growth rate is surprisingly high.

The results fit every city's condition quite well. The visualization result shows the cluster progress is effective.

2.4 Discriminant Analysis

With the clustering results, we can define the four different groups to be G_i, i=1,2,3,4. Their average mean are X_i , i=1,2,3,4.
Assumption: • The covariance matrix of each group is the same (S).

Then we will get the Mahalanobis distance between a sample X by calculating:

$$d^2 (X, G_i) = (X- X_i)'S^{-1}(X- X_i)$$

$$= X' S^{-1}X – 2(S^{-1}X_i)'X + X_i'S^{-1}X_i$$

Let

$$Y_i(X) = (S^{-1}X_i)'X - \frac{1}{2} X_i'S^{-1}X_i$$

and it is the linear discriminant function. It is obvious that if $Y_i(X)$ is a large number, then d^2 (X, G_i) is a small number, which means X is close to G_i.

We use the DISCRIM procedure in SAS and the following table shows the coefficients of the linear discriminant function.

Linear Discrimination Function for Group					
Variable	Label	G_1	G_2	G_3	G_4
Constant		-8783	-8173	-9257	-9486
GPC	GPC	889.18	802.9	830.3	750.9
AAQD	AAQD	771.85	764.23	775	809.3
HP	HP	614.27	589.6	607	567
UR	UR	-1234	-1232	-1272	-1296
NPI	NPI	6066	6080	6373	6449
NHI	NHI	2158	2234	2235	2702
NCS	NCS	7732	7698	7905	7664

Table 5. The DISCRIM Procedure

We can get the linear discriminant function of the four groups in 2.2:

$$Y_1(X) = -8783 + 889.2x_1 + 771.8x_2 + 614.3x_3 - 1234x_4 + 6066x_5 + 2158x_6 + 7732x_7$$

$$Y_2(X) = -8713 + 802.9x_1 + 764.2x_2 + 589.6x_3 - 1232x_4 + 6080x_5 + 2234x_6 + 7698x_7$$

$$Y_3(X) = -9257 + 830.3x_1 + 775x_2 + 607x_3 - 1272x_4 + 6373x_5 + 2235x_6 + 7905x_7$$

$$Y_4(X) = -9486 + 750.9x_1 + 809.3x_2 + 567x_3 - 1296x_4 + 6449x_5 + 2702x_6 + 7664x_7$$

By calculating the minimum value of Yi(X) we can determine which group does X belong to.

Data Item	x_i
GPC （￥）	x_1
AAQD (day)	x_2
HP （￥）	x_3
UR (%)	x_4
NPI (10^4people)	x_5
NHI (10^4people)	x_6
NCS (10^4people)	x_7

Table 6. The Meaning of x_i

2.5 Model Validity Test

We do the validity test with SAS and we choose two different cities as the test set: Chong Zhou, China and Seattle, USA. According to the HDI data and the development of these cities, their smart growth standard should fall in middle level.

By using the smart growth degree evaluation model in 2.3, we obtain the Mahalanobis distance of the cities to four level groups. Next, we can show the posterior probability bellow:

Obs	From Group	Classified into Group		G_1	G_2	G_3	G_4
1	1	1		0.9989	0.0000	0.0011	0.0000
2	1	1		0.9990	0.0000	0.0010	0.0000
…	…	…		…	…	…	…
14	3	3		0.0000	1.0000	0.0000	0.0000
15	2	1	*	0.6702	0.0001	0.3298	0.0000
16	.	2	*	0.0021	0.9979	0.0000	0.0000
17	.	2	*	0.2417	0.7583	0.0000	0.0000

Table 7. Posterior Probability

The results show that only the 15^{th} city's judgment result is wrong. Chong Zhou and Seattle are classified into G_2, which meets our expectation. Therefore, it ensures the validity of the smart growth degree evaluation model.

2.6 Sensitivity Analysis

To do sensitivity analysis, we slightly change the value of some items of Chong Zhou. For instance, we change NPI in 2014 from 13.7 to 13.90,change HP in 2014 from 5440 to 5430 and change UR in 2014 from 3.99% to 3.95%. The results all shows that this city's smart growth standard falls in middle level, same with the original result. The results of sensitivity analysis demonstrates that the prediction model is stable.

2.7 The Current Growth Plan Analysis

As mentioned in 2.4, we choose

- **Chong Zhou, China**
- **Seattle, USA**

as two mid-sized cities on two different continents to study. We obtain the current data of these two cities from Chong Zhou Bureau of Statistics and U.S. Census Bureau.

The data in the recent years of Chong Zhou are as follows.

	GPC	AAQD	HP	UR	NPI	NHI	NCS
2011	169.94	219	4000	3.75%	30.45	13.64	5.78
2012	183.4	223	4849	3.58%	29.77	13.68	5.88
2013	203.23	209	5440	3.99%	30.51	13.7	5.76
2014	226.1	216	5420	3.96%	31.34	14.02	5.71
2015	252	211	5509	3.88%	31.72	14.34	5.75

Table 8. Current Data of Chong Zhou

Then, we can calculate the growth rate of every item by year, the results are listed in table 9 bellow:

	GPC	AAQD	HP	UR	NPI	NHI	NCS
2012	0.079	0.018	0.212	-0.045	-0.022	0.003	0.017
2013	0.108	-0.063	0.122	0.115	0.025	0.001	-0.020
2014	0.113	0.033	-0.004	-0.008	0.027	0.023	-0.009
2015	0.115	-0.023	0.016	-0.020	0.012	0.023	0.007

Table 9. Growth rate of Chong Zhou

In the same way, we can calculate the growth rate of Seattle of every item by year, the results are listed in table 10 bellow:

	GPC	AAQD	HP	UR	NPI	NHI	NCS
2013	-0.025	0.000	0.120	-0.030	-0.016	0.012	-0.050
2014	0.008	0.000	-0.044	0.182	0.025	-0.003	0.033
2015	-0.009	0.000	0.045	-0.082	-0.016	-0.001	-0.034

Table 10. Growth rate of Seattle

Using the growth rates of current years in smart growth degree evaluation model, we find their smart growth standards both fall in the middle level G_2 (in middle smart growth level).

That means according to our metric, the current growth plans of the two cities are not bad, but the plans can be developed to be better ones (in high smart growth level).

3. The Growth Rate Prediction and Plans' Effect Model

To develop a growth plan feasible for a city, we should estimate its smart growth condition in the future. We study a city's smart growth from the same seven aspects mentioned in the smart growth degree evaluation model. We obtain the data, from 1999 to 2015 of Chong Zhou and Seattle from Chong Zhou Bureau of Statistics and U.S. Census Bureau.

In order to predict a city's smart growth condition, we should build models to estimate the growth rate of its items. Combining the prediction results with the plans we make for the city's development, we can finally know the success of our future smart growth plans with the help of the smart growth degree evaluation model.

3.1 Assumptions

- The data we collected is accurate.
- The growth is stable and there is no sudden occurrence like financial crisis. (This can ensure the accuracy of the prediction model.)
- The stochastic disturbance on economic growth satisfied Gauss-Markov model.

3.2 The Future Growth Plan

3.2.1 Smart Growth Plan for Seattle

As for Seattle, it is a livable city with wonderful natural environment. There are many big companies like Microsoft there, making the GPC of Seattle quite high. However, the growth rate of its GPC is low. At the same time, there are many social welfare investment funds, but the growth rate also remains low. These make data of Seattle quite different from our training set.

To know how every aspect contributes to the smart growth of Seattle, we use the corresponding coefficient in the metric we make. With the discriminant analysis method, we multiply the value of the items of Seattle with the corresponding coefficient and get the results:

Seattle	G_1	G_2	G_3	G_4
GPC	895.22	808.62	835.81	756.26
AAQD	771	764	775	809
HP	693.20	666.10	685.28	642.38
UR	－1057.71	－1056	－1090.29	－1110.86
NPI	6144.86	6159.04	6476.11	6532.84
NHI	2187.57	2264.61	2265.62	2739.02
NCS	7883.276	7848.61	8059.66	7813.95

Table10. Seattle's Items Contribution to the Smart Growth

Considering that Seattle falls in G_2, we should develop a plan to help it falls in G_3 (high level of smart growth). By analyzing the products of G_2 above, we can find Seattle has very high scores in NCS (7848.61) and NPI (6159.04). At the same time, the D-values of NCS and NPI between G_2 and G_3 are high. These mean that NCS, NPI contributes a lot to the smart growth of Seattle.

However, it has a low score in UR (－1056), which means, relatively, the improvements of UR investment can not contribute to the smart growth of Seattle. The possible reason for this is that this city is well developed and simply increasing capital investment on UR cannot actually make great difference in UR. As the of the dimension of last four items is the same, we can compare their efficacy directly.

The smart growth plans we make for Seattle are:

- To decrease the growth rate of investment on UR
- To increase the growth rate of investment on pension insurance
- To give More support for the development of Education

3.2.2 Smart Growth Plan for Chong Zhou

The smart growth of Chong Zhou falls in G_2 (in middle smart growth level). The situation of Chong Zhou's development is quite different with that of Seattle. Chong Zhou is a rising town and its investment on social welfare is now insufficient, which means it has great potential on the development of social welfare.

To know how every aspect contributes to the smart growth of Chong Zhou, we use the corresponding coefficient in the metric and multiply the value of the items of Chong Zhou with the corresponding coefficient like the progress in 3.2.1 and get the results:

Chong Zhou	G_1	G_2	G_3	G_4
GPC	990.84	925.08	894.98	837.03
AAQD	753.15	757.06	746.31	790.27
HP	624.08	616.97	599.69	578.34
UR	-1209.07	-1246.30	-1207.11	-1269.82
NPI	6139.55	6470.52	6153.72	6527.19
NHI	2207.26	2286.01	2284.99	2763.67
NCS	7786.16	7960.38	7751.93	7717.69

Table11. Chong Zhou's Items Contribution to the Smart Growth

Chong Zhou falls in G_2. By analyzing the products of G_2 above, we can find Chong Zhou has very high score in NPI (6470.52) and the D-value NPI between G_2 and G_3 is the highest (316.8). These mean that NPI contributes a lot to the smart growth of Chong Zhou and has great potential to be improved.

However, it has a low score in UR (-1246.30), which means, relatively, the improvements of UR investment cannot contribute much to the smart growth of Chong Zhou.

Considering the development condition of Chong Zhou, the smart growth plans we make for this rising city are:

- To decrease the growth rate of investment on UR
- To make great effort to develop the pension insurance

After making qualitative analysis of our plans, **we will be able to give out the quantitative results in section 3.4**. In order to evaluate the success of our plans, take Chong Zhou for example, we will give artificial constraint to the growth rate of UR and NPI. In addition, the reasonable growth rate of the other five item will be calculated with the method of linear regression.

3.3 GPC Growth Model

We first choose GPC in the seven items to develop its growth model. And if needed, the growth model of the other items can be built in the same way.

The PIC from 1999 to 2015 of Chong Zhou are shown in the figure bellow:

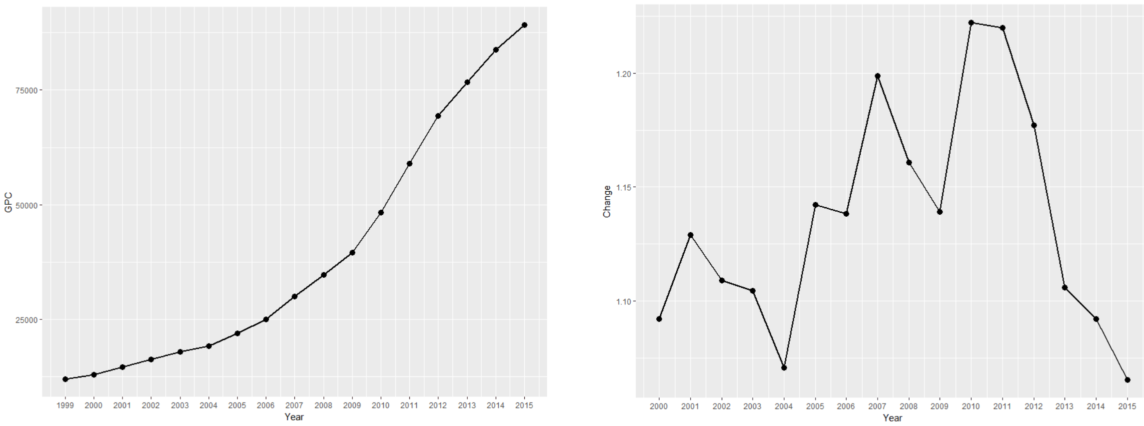

Figure 3. PIC and its growth rate of Chong Zhou from 1999 to 2015

From figure3 we can find the growth rate of PIC is high between 2005 and 2012. We choose the data above to be our training set and try to use linear regression to fit the data. Then we get the fitting function and visualization result:

$$Y = 5025 * X - 1005000$$

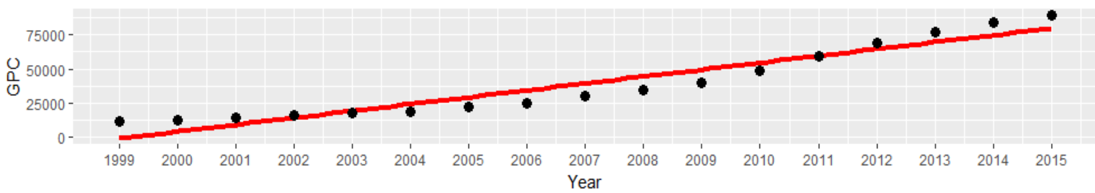

Figure 4. Linear Regression of PIC

Next, we draw the Q-Q plot to do the test of normality:

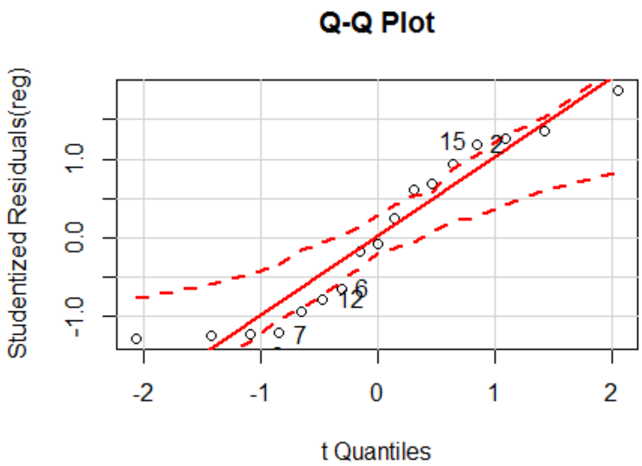

Figure 5. Test of Normality (1)

From the test of normality, we find there are six points do not obey normal distribution. Their numbers are 2, 6, 7, 8, 12, 15. Therefore, we must transform the data to make they obey normal distribution so that we can get more accurate fitting results. We choose Box-Cox transformation to solve this problem:

$$y^{\lambda} = \frac{y^{\lambda}-1}{\lambda} \qquad \lambda \neq 0 \qquad (1)$$

$$y^{\lambda} = log(y) \quad \lambda = 0$$

To determine the value of λ, we use the method of Maxima Likelihood Estimation. We definite the search range of λ is [-1.5,1.5] and choose 0.1 as the step width. Then we can get the ideal value of λ:

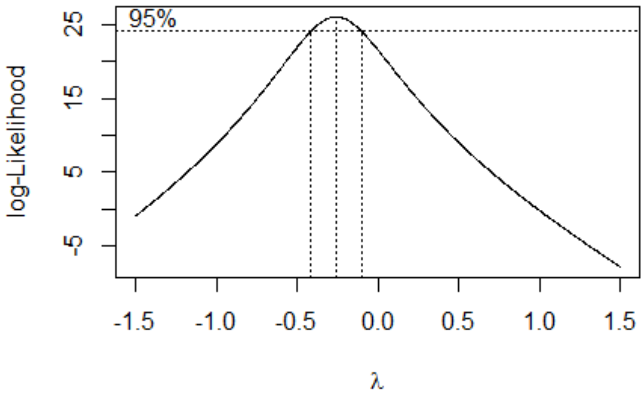

Figure 6. Principle of λ Choosing

From the picture above, we find that when we choose -0.25 as the value of λ, the likelihood reaches the highest level. Therefore, we use formula (1) to transform GCP and do the linear regression and get the fitting function:

$$(y^{\lambda}-1)/\lambda=0.001019x-16.75 \qquad (2)$$

The function fits the data well and we also draw the Q-Q plot to do the test of normality:

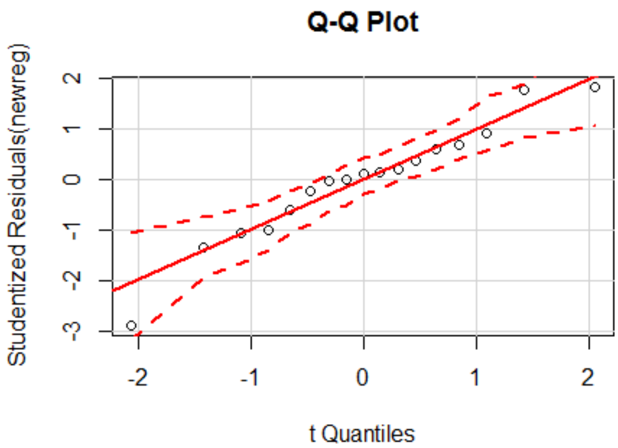

Figure 7. Test of Normality (2)

From the figure above, we know the data after transformation all fall between the two dotted lines, which means all data obey normal distribution. Therefore, we choose the fitting formula (2) as the GPC growth model.

The prediction results are as follows:

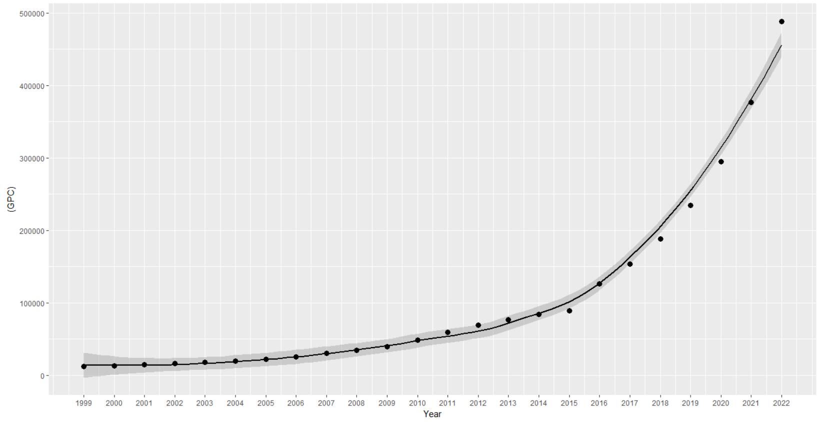

Figure 8. Final Regression Result of GPC

We can calculate the growth rate of GPC in the future if there is no policy intervention. This will help us evaluate the success of our future smart growth plans in the following section.

Year	2017	2018	2019	2020	2021	2022	...
Growth Rate	1.117	1.130	1.144	1.159	1.177	1.198	...

Table 12. Growth Rate of GPC

After developing the PIC growth model, we can develop the growth rate prediction model of other items in the same way.

We use R to analyze the validity of our model. The results are as follows.

	critical value	Number of points	Result
Outlier Points	2	1	Assumptions acceptable
High Leverage Points	0.25	0	Assumptions acceptable
Influential Points	0.308	0	Assumptions acceptable

Table 13. Model Validity Test

From the table above, we know that the number of points which fall above the critical value is 1, 0, 0. The result ensures the validity of the growth rate prediction model.

3.4 Plans' Effect Model and Potential Initiatives

3.4.1 Plans' Effect Model

Take the growth of Chong Zhou into consideration, according to the growth rate prediction model, we know the growth rate of each item in 2022:

Items	GPC	AAQD	HP	NHI	NCS
Growth Rate	1.1978	1.0638	1.1059	1.0628	1.0391

Table 14. Growth Rate in 2022

Then we must work out the optimum solution of UR's and NPI's growth rate. We firstly simplify the following discriminant function with the data above

$$Y_2(X) = -8713 + 802.9x_1 + 764.2x_2 + 589.6x_3 - 1232x_4 + 6080x_5 + 2234x_6 + 7698x_7$$

$$Y_3(X) = -9257 + 830.3x_1 + 775x_2 + 607x_3 - 1272x_4 + 6373x_5 + 2235x_6 + 7905x_7$$

and get the results:

$$Y2(X) = - 1232\, x_4 + 6080\, x_5 + 4086.925$$
$$Y3(X) =- 1272\, x_4 + 6373\, x_5 +3822.633$$

To insure that Chong Zhou falls in G_3 is to minimize:

$$Y2(X)- Y3(X)= 40\, x_4 - 293\, x_5+ 263.292$$

As x_4 and x_5 can reflect the growth rate of GCP (1.19778 = k in 2022) to some extent, we should limit their D-value. The object function is transformed to:

$$minimize\ (40\, x_4 - 293\, x_5 + 263.292 + a|\, x_4\text{-}k| + b|\, x_5\text{-}k|) \qquad X4 >0\ x5> 0$$

With the method of Maxima Likelihood Estimation, we calculate the optimum solution: a = 1.21, b = 1.16. So the growth rates of UR and NPI in 2022 are 0.958537535 and 1.0823483. And the growth rates of UR and NPI in 2014 are X4 0.97979798 and 1.01212508.

$$(0.958537535\text{-}0.97979798)/(2022\text{-}2014) = -\ 0.002657556 \qquad (decrease)$$
$$(1.0823483\text{-}1.01212508)/\ (2022\text{-}2014) = 0.0087779025 \qquad (increase)$$

The plan is to make the growth rate of investment on UR decrease 0.002657556 every year. At the same time, we should make the growth rate of investment on pension insurance increase 0.0087779025 every year. Using the same method we get the quantitative results of our plan for Seattle.

Putting the data in 2022 into the smart growth degree evaluation model, we find $Y_3 < Y_2$ so that both Chong Zhou and Seattle will fall in G_3. This result indicates that our plan successfully help these two cities reach high level of smart growth in the future.

3.4.2 Potential Initiatives

We define an initiative to be potential if its growth rate stays in a relatively high level. So we calculate the average of each item's growth rate over these years. With the growth rate prediction model and data in section 3.4.1, we get the following results:

Average Rate	GPC	AAQD	HP	UR	NPI	NHI	NCS
Chong Zhou	1.1036	0.9915	1.0867	1.0104	1.0105	1.0126	0.9988
Seattle	0.9913	1.0000	1.0405	1.0234	0.9979	1.0023	0.9832

Table 15. Average Growth Rate

From the table above, we can rank the initiatives by their potentiality:

Rank	1	2	3	4	5	6	7
Chong Zhou	GPC	HP	NHI	NPI	UR	NCS	AAQD
Seattle	HP	UR	NHI	AAQD	NPI	GPC	NCS

Table 16. Ranking

The comparison between the two city is clear in table 14. The most potential initiative of Seattle is HP, which means the growth rate of housing price may keep increasing in the future. The least potential initiative is NCS because the investment on education has already reach a relatively high level which makes its growth rate stay low.

As for Chong Zhou, the most potential initiative is GPC, which means the development of economy has great potentiality. The least potential initiative is AAQD. The reason for this is the development of Chemical works. This will pollute the environment over a long period of time.

4. Population Regression Model

To find the relationship between the population and the seven items, we build a population regression model using the shrinkage method. And then we use the growth rate prediction model to get the value of the seven factors in 2050. According the population regression model, we can calculate the population in 2050 and we find the growth rate is 43.985%. Thus, our plan supports the booming population well.

4.1 Assumptions

- In the next 50 years, social environment is relatively stable.
- The number of migration of international population is equal to emigration.

4.2 Population Regression Model

As for Chong Zhou, we use Lasso to make regression analysis:

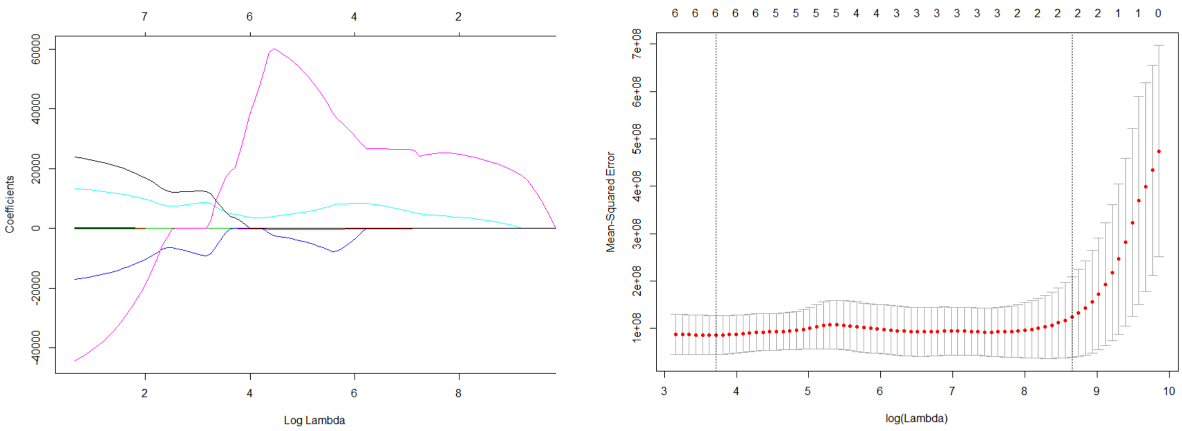

Figure 9. Lasso Regression

Based on L1-norm method, we can select variables and shrink regression coefficient simultaneously. While the log(λ) is equal to 10, the number of variables has been punished to 0. Then we can use the Cross-Validation to determine the best value of λ.

By using Cross-Validation method, we can get the best value of λ equal to 41.13242. Then we get the population regression model:

$$Population = 1791.992 + 0.02959355GPC + 4268.097NPI + 25251.77NHI$$

It shows that NPI and NHI can strongly affect the population growth, which is reasonable [1]. These two items can reflect the standard of social welfare, which can ensure the quality of life. We can also know that the three chosen items all positively related to the population growth. The fitting effect is shown in the figure below:

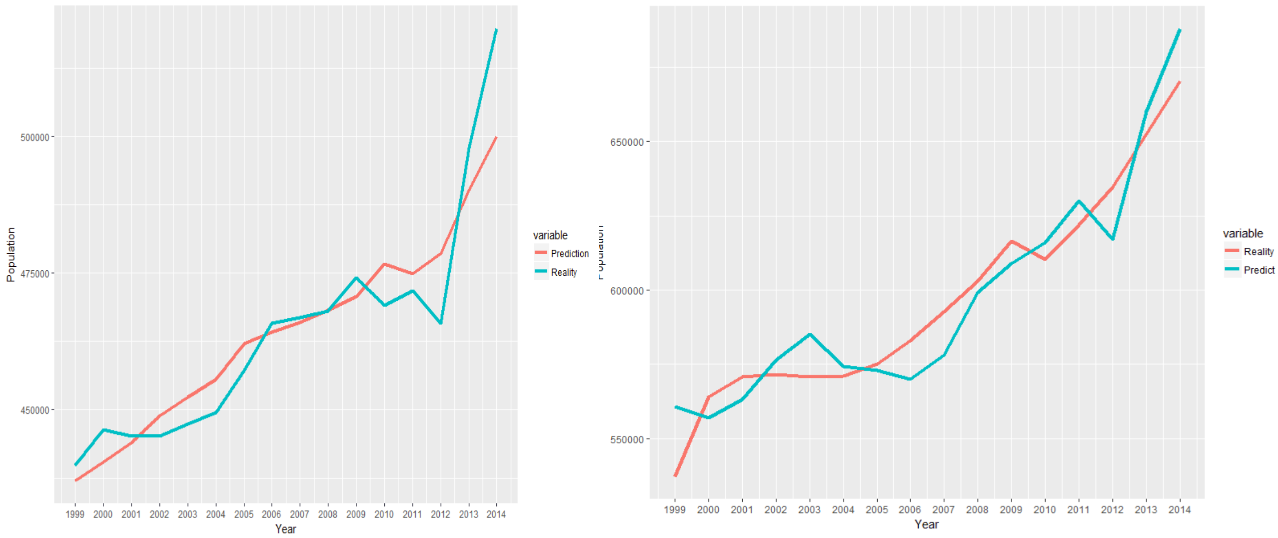

Figure 10. Fitting Results

The model fits our data well. Using the population regression model and the growth rate prediction model, we get the fitting result in 2050:

GPC	NPI	NHI	Population
157800	60.68	19.13	748516

Table 17. Regression Result

Compared with the population in 2014 (519857), the increase of predicted population in 2050 is 43.985%. Obviously, the growth rate is near 50%. So our plan supports the population blooming growth.

As for Seattle, we use the same way to predict the population in 2050, we get the population regression model:

$$Population = 2638.149 + 0.05278624GPC + 3145.823NPI + 29219.346NHI$$

The fitting effect is shown in figure 10 together with the fitting result of Chong Zhou.

We can see the model fits well. And we can calculate the growth rate in 2050 is 48.425%. So our plan for Seattle supports the population blooming growth.

5. Strength and Weakness of Models

5.1 Model 1: Smart Growth Degree Evaluation Model

Our model takes the advantage of cluster analysis and discriminant analysis to divide cities into groups and work out the low dimensional metric. So we get a relatively credible metric to evaluate a city's smart growth.

However, we cannot guarantee that the items we choose can indicate the smart growth of a city comprehensively.

5.2 Model 2 & 3: The Growth Rate Prediction and Plans' Effect Model

The prediction model fits the data well, so our prediction is approximately accurate. In addition, we calculate the optimal solution in the second model with the method of linear programming. Therefore, our growth plan is successful.

However, our assumptions are too optimized.

5.3 Model 4: Population Regression Model

We use the Lasso to reduce the number of the variables, so our model is greatly simplified.

The data we collected is a little insufficient for the prediction.

References

[1] Yang Zaigui, (2008) "Analysis of OLG Model on Submitting Rate of Endowment Insurance and Population Incearasing Rate"

Appendix

Model 1 Code in R:

```
library(vegan)
library(cluster)
library(factoextra)
library(ggfortify)
data = read.csv("E:/2.csv")
data = data[-8,]
dataset = scale(as.matrix(data[,-1]))
d = data[,-1]
data_scale = scale(d)
row.names(data_scale) = c('Bei Jing','Shang Hai','Guang Zhou','Cheng Du','Shen Zhen','TianJin', 'Hang Zhou','Wu Han','Shen Yang','Zheng Zhou','Nan Jing','Da Lian','Su Zhou','Fo Shan','Chong Qing')
data_pam = pam(data_scale,k = 4,stand = TRUE)
data_pam
autoplot(pam(data_scale,k = 4,stand = TRUE), data = data_scale,label = TRUE,label.size = 3, frame = TRUE)
library(cluster)
fviz_nbclust(dataset, kmeans, method = "silhouette")
```

Model 2&3 Code in R:

```
Per=c(11858.49527,12951.62883,14622.17861,16215.95331,17911.87739,19178.47025,21903.8817,24932.00363,29892.08633,34697.5089,39525.89991,48311.57528,58933.79192,69377.66411,76731.25527,83800,89264.46281)
Year = c(1999:2015)
data = data.frame(Year = Year,GPC = Per)
library(ggplot2)
library(ggthemes)
ggplot(data=data, aes(x=Year, y=GPC, group=1)) +
  geom_line(colour="black", size=1) +
  geom_point(colour="black", size=3, fill="white") +
  scale_x_continuous(breaks=seq(1999, 2015, 1))
Per_change = c(1:16)
for (i in 2:17)
  Per_change[i-1] = Per[i]/Per[i-1]
new_data = data.frame(Year = c(2000:2015),Change = Per_change)
ggplot(data=new_data, aes(x = Year, y = Change, group = 1)) +
  geom_line(colour="black", size=1) +
  geom_point(colour="black", size=3, fill="white") +
  scale_x_continuous(breaks=seq(2000, 2015, 1))
library(MASS)
```

```
library(car)
reg = lm(GPC~Year,data = data)
plot(Year,Per)
abline(reg)
ggplot(data=data, aes(x = Year, y = GPC, group = 1))+ geom_smooth(method = "lm", formula = y ~ x, se = FALSE,color = "red",size = 1.5) + geom_point(colour="black", size=3, fill="white")+ scale_x_continuous(breaks=seq(1999, 2015, 1))
qqPlot(reg,labels = row.names(data),id.method = "identify",simulate = TRUE,main = "Q-Q Plot")
t = boxcox(GPC ~ log(Year), data = data,lambda = seq(-1.5, 1.5, length = 1000))
lambda = -0.25
new_Per = (Per^lambda - 1) / lambda
newreg = lm(new_Per~Year,data = data)
qqPlot(newreg,labels = row.names(data),id.method = "identify",simulate = TRUE,main = "Q-Q Plot")
new = data.frame(Year= c(2016:2022))
GPC_trans = predict(newreg,newdata = new)
GPC_new = (GPC_trans * lambda + 1) ^ (1/lambda)
options(scipen=200)
all_GPC = c(Per,GPC_new)
all_Year = c(1999:2022)
all_data = data.frame(GPC = all_GPC,Year = all_Year)
ggplot(data=all_data, aes(x=Year, y=(GPC), group=1)) +
  geom_smooth(colour="black", size=1) +
  geom_point(colour="black", size=3, fill="white") +
  scale_x_continuous(breaks=seq(1999, 2022, 1))
```

Model 4 Code in R:

```
library(glmnet)
library(car)
library(caret)
library(ggplot2)
library(reshape2)
library(ggthemes)
library(reshape)
MSE = function(y, h)
{
  return(mean((y - h) ^ 2))
}
data = read.csv("E:/1.csv")
x = data[,-1:-2]
x = as.matrix(x)
y = data[,2]
```

```
y = as.matrix(y)
fit.lasso = glmnet(x,y,alpha = 1)
plot(fit.lasso,xvar = "lambda")
plot(fit.lasso,xvar = "dev")
cv.lasso=cv.glmnet(x,y,type.measure="mse")
plot(cv.lasso)
coef = coef(cv.lasso,s = cv.lasso$lambda.min)
lasso.mse = MSE(y,predict(fit.lasso,s = cv.lasso$lambda.min,newx = x))
a = rep(1,16)
p = data.frame(intercept = a,data[,-1:-2])
mat = as.matrix(p)
pred = mat%*%coef
pred = as.vector(pred)
pred1 = data.frame(Prediction = pred)
newdata = data.frame(pred1,data)
newdata = rename(newdata,c(Population="Reality"))
temp = melt(newdata[,1:3],id.vars = "Year")
ggplot(temp,aes(x=Year,y=value)) + geom_line(aes(color = variable),size = 1.5)
+ scale_x_continuous(breaks=seq(1999, 2014, 1)) + scale_y_continuous(name =
"Population")
```

	GPC	AAQD	HP	UR	NPI	NHI	NCS
Bei Jing	92210.14	168	18833	1.3	1392.6	1604.3	60.5
Shang Hai	89444.22	278	16787	4.1	1457.4	1678.5	50.7
Guang Zhou	129242.12	282	15719	2.26	1070.5	1054.7	101.9
Cheng Du	70337.68	216	7032	2.87	547.5	588.4	72.93
Shen Zhen	155051.61	348	24732	2.26	870.69	1157.83	8.77
Tian Jin	97609.44	175	9219	3.5	545.4	1023.6	50.6
Hang Zhou	104038.44	216	13896	1.84	663.45	840.21	47.47
Wu Han	98527.2	177	7951	3.15	380.91	594.98	96.21
Shen Yang	91909.89	190	6217	3.03	355.8	357.9	40
Zheng Zhou	73798.63	135	7571	1.37	176.7	163.8	78.3
Nan Jing	107730.4	188	11198	2.1	276.52	342.91	70.23
Da Lian	110263.57	276	9216	4.09	194.5	505.4	28.6
Su Zhou	130081.11	227	12319	2	462.91	593.78	20.72

Fo Shan	104215.91	234	8484	2.35	344.65	272.94	4.67
Chong Qing	47688.04	246	5519	3.5	825.5	3256.8	69.2

Table 2

2015	GPC	AAQD	HP	UR	NPI	NHI	NCS
Bei Jing	106751.25	186	22633	1.4	1424.2	1656.6	60.4
Shang Hai	102919.55	252	20949	4	1493.8	1719.2	51.2
Guang Zhou	138377.05	312	14612	2.2	1159.4	1052.6	104.3
Cheng Du	74862.42	211	6875	3.19	570.8	620.8	75.6
Shen Zhen	162381.97	340	33942	2.34	954.34	1213.16	9.01
Tian Jin	109032.71	216	10107	3.5	565.2	1054.1	51.3
Hang Zhou	113063.2	242	14422	1.74	668.65	870.71	47.56
Wu Han	105973.98	189	8556	3.08	407.56	602.35	95.68
Shen Yang	87854.47	207	6861	3.18	370.6	351.4	40.4
Zheng Zhou	78003.73	136	7537	1.6	198.1	164.9	82.4
Nan Jing	118313.68	231	11489	1.9	299.04	388.37	70.62
Da Lian	111358.2	270	8929	2.93	195.1	506.7	29
Su Zhou	136947.12	240	12267	2	468.41	609.9	21.48
Fo Shan	108887.98	243	8587	2.36	380.47	277.74	4.94
Chong Qing	52549.71	282	5486	3.6	849.3	3266.3	71.7

Table 3