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

Team Control Number
89236

Problem Chosen
B

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

**2018**
**MCM/ICM**
**Summary Sheet**
(Your team's summary should be included as the first page of your electronic submission.)
Type a summary of your results on this page. Do not include the name of your school, advisor, or team members on this page.

# Language Change and Mathematics


To part 1, we establish a model to predict the number of the given language speakers. The model consists of two sub-models: the Logistic prediction model for native speakers and the multiple regression model for second language speakers. For a given language, by the historical data of native speakers, we fit the logistic model parameter values. Then we use the Logistic model to predict the change of native speakers in the next 50 years. The forecast result shows that the number of native speakers will increase slowly and finally remain stable. Next, we take into account three indicators including economy, culture and science & technology that affect the change of second language speakers. We take advantage of multiple linear regression to analyze the relationship of the quantity of second language speakers and the three defined indicators. We predict the change trend of second language speakers and it is slightly fluctuating.

After that, we combine the prediction result with some other factors and model the distribution of various language speakers over time. Some languages in the list concentrate in the mother tongue, such as Portuguese, Russian and Punjabi. Other languages are widely distributed such as English, Spanish and Chinese. Then we study the global population and human migration patterns predicted in the future and discover that the geographic distributions of these languages change simultaneously. It proves that the change is almost consistent with the trend and direction of population migration by conducting a correlation analysis of population migration with second language speakers.

To part 2, according to our model result, we refer to the number and change trend of given language speakers, the level of development of corresponding to countries or regions. As to the short term versus the long term, we mainly focus on different aspects and propose diverse suggestions of the location and language of new international offices. In the short term, we mainly consider the number of speakers in a particular language. However, we pay more attention to the change trend of a given language speakers in the long term. Then, taking the development of global communications into account, we need to combine the client company's operating conditions with regional communications to save the company resources. It proves reasonable that we recommend opening less than six offices when the region's communications capabilities meet the company's business needs.

To part 3，we summarize the modeling result and furnish the suggestions to the Chief Operating Officer of the service company.

Besides, as to the first sub-model, we analyze the model sensitivity of the parameter $r$ (the growth rate of the number of native speakers). By changing the value of the parameter $r$, we find that as $r$ increases, the predicted growth rate increases rapidly, but the final stability value does not change. Then, we test our model by the historical data. The relative error of most calculations is within 10%. It indicates that our model is reasonable and effective.



**Keywords**: Language change; Multiple-regression; Logistic prediction model; Distribution fitting

# Contents

# 1 Introduction

## 1.1 Background

Language is an important factor of promoting the development of human civilization. Nowadays, there are approximately 6900 languages in the world, each language stands for a unique culture. Due to the rapid development of modern civilization, such as economic growth, the improvement of education quality, tourism development, people are not only speaking a single language any more, more and more people begin to learn and use a second language. This phenomenon is also leading to changes in the geographical distribution of languages and the number of users. Mandarin (incl. Standard Chinese), Spanish, English, Hindustani, Arabic, Bengali, Portuguese, Russian, Punjabi, and Japanese are used widely in the world. About half the global population claim one of the following ten languages as a native language.However,the number of users in some languages is shrinking.

## 1.2 Restatement of the Problem

Now, our team is hired by a large multinational service company, helping them to investigate trends of global languages and location options for new offices.

We need consider the following:

- After considering some influences and factors, based on projected trends, model the distribution of various language speakers over time.
- Predict what will happen to the numbers of native speakers and total language speakers in the next 50 years.
- Predict and explain that any of the languages in the current top-ten lists will be replaced by another language.
- In the next 50 years, whether the geographic distributions of these languages change over this same period of time.
- Located six suitable international offices and recommend one standard office language in each office.
- Take account of the client resources, select the appropriate number of international offices.

### 1.3 Our Work

To predict global language development trend and provide advice for client company, our solutions are as follows:

- Establish the population prediction model which contains two sub-models: the Logistic prediction model for native speakers and the multiple regression model for second language speakers.
- Take the prediction result and some other influence factors into account. Then we model the distribution of various language speakers over time on the map.
- Study the global population and human migration patterns predicted in the next 50 years. To describe the change of the geographic distributions, we conducted a correlation analysis of population migration with second language speakers.
- Give the suggestions about the location and language options depending on our model results. Meanwhile, we analyze the basis of choice and rationality.

## 2 Assumptions and Justifications

We make the following assumptions to approximate and simplify the problem.

- **The data we collect from online databases is accurate, reliable and mutually consistent.** Because our data sources are all websites of international organizations, it is reasonable to assume the high quality of their data.
- **Services provided by multinational service companies to people in all regions are equal and there is no discrimination in the region.** Ignoring the influence of regional discrimination is conducive to the company allround development. It is reasonable and works well for simplifying the model.
- **Ignore unpredictable or high-impact, low probability events such as asteroid collisions.** It would cause a catastrophic jump in evolutionary trends over time, and possibly render all languages extinct.

Besides these general assumptions, there are also hypotheses we make for the specific models. We will present and discuss them in specific model.

# 3 Symbol Description

Table 1: Notations

| Symbol | Meaning |
|---|---|
| $p$ | The number of the given native speakers |
| $r$ | The growth rate of the population |
| $p_k$ | The population after $k$ years |
| $p_0$ | The population of the current year |
| $p_m$ | The maximum number of the population depending on environment |
| $s$ | The population ratio coefficient |
| $y_i$ | The number of second language speakers |
| $x_1$ | Regional economic development level indicator |
| $x_2$ | Regional science and technology development level indicator |
| $x_3$ | Regional culture development level indicator |

# 4 The Model

For a certain language, the users are divided into two parts, the native speakers and the second language speakers.

In order to quantitatively describe the temporal and spatial variations of the number of speakers, we establish the model to predict the number of language users. The model consists of two sub-model models. The first sub-model is to predict the number of native speakers and the second sub-model is to predict the number of second language speakers.

## 4.1 Establish the Model

### 4.1.1 Logistic Model of Mother Tongue Population Prediction

**1.Prediction Model Establishment**

Considering the number of native speakers mainly depends on the number of residents in the region, we adopt the Logistic model which is suitable for population prediction.

Factors such as natural resources and environmental conditions impede the growth of the population, and the growth rate of the population will decrease after reach a certain amount of population growth.If the general population index growth model is adopted, it is obviously not practical for the long-term population forecast. Therefore, our team establish the Logistic model of the dynamic change of the growth rate of the population.

The block effect of environment and resources on population growth is main-

ly reflected in the change of population growth rate $r$, and $r$ is negatively correlated with population size $p$. $r(p)$ can be expressed as a function of $p$. To simplify the model,we assume that $r(p)$ is a linear function of $p$.

$$r(p) = r_0 - sp \tag{1}$$

Combined with simple population growth index model:

$$p_k = p_0(1+r)^k \tag{2}$$

Where: $p_0$ is the population of the current year;
$p_k$ is the population after $k$ years;
$r$ is the annual growth rate of the population (r remains the same).

When research population changes in a country or region, the value of population $p(t)$ at time $t$ is significant. We can treat the population $p(t)$ as a continuous differential function. Using calculus knowledge, we can get:

$$\frac{dp}{dt} = rp, p(0) = p_0 \tag{3}$$

Solving differential equations can obtain:

$$p(t) = p_0 e^{rt} \tag{4}$$

Substituting equation (1) into equation (3):

$$\frac{dp}{dt} = r(p)p, p(0) = p_0 \tag{5}$$

We introduce the maximum number $p_m$ that can be accommodated by natural resources and environmental conditions ,and pm is called population capacity. If $p = p_m$, $r(p_m)$=0,then substituting equation(1). The result $s = \frac{r}{p_m}$ can be gained, combined equation(5):

$$\frac{dp}{dt} = rp(1 - \frac{p}{p_m}), p(0) = p_0 \tag{6}$$

The $rp$ in the equation reflects the growth trend of the population ,and $1 - \frac{p}{p_m}$ reflects the blocking effect of the environment. These two factors jointly determine the trend of population change.

For the upper differential equation, the separation variable method can be used to solve the problem.

$$p(t) = \frac{p_m}{(1 + (p_m/p_0 - 1)e^{-rt}} \tag{7}$$

**2.Model Parameter Estimation**

It is necessary to estimate the parameters $r$ and $p_m$ in the above model by using the above model to predict the future population. We look up the historical

data of the number of native speakers in various languages. In order to estimate the parameters $r$ and $p_m$ in the model, we can further transform the formula (6) into :

$$\frac{dp/dt}{p} = r - sp, (s = r/p_m) \tag{8}$$

On the left side of equation (8), the value can be calculated by using the numerical differential of population data in the historical data.The right side parameter r and s are a linear relation, and we can use the historical data to do the linear fitting. The value of parameters that will be obtained in equation (7) can be used to predict the population in the future.

**3.Model Solution**

Our team look up the historical data of ten native speakers on the related websites (part of the data is missing).

We choose the Spanish mother tongue population data for parameter linear fitting,calculating $r = 0.169/(3years)$, $s = 0.0003$, $p_m = 563.3$ million(refer to Fig.1).

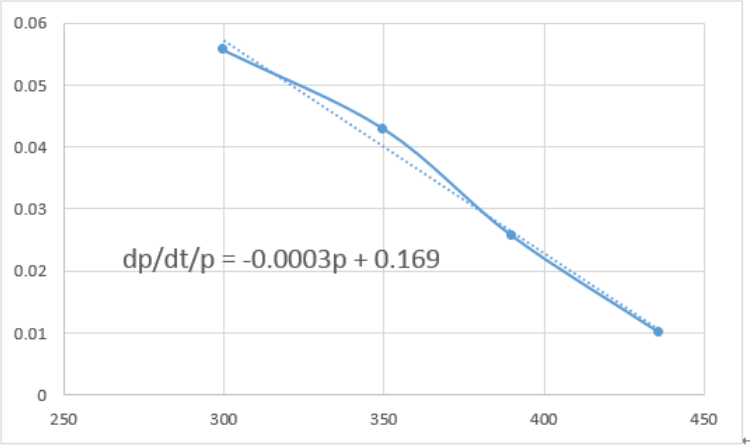


Figure 1: The parameters $r$,$p_m$ estimation

Substituting the values $r = 0.169/(3years)$ and $p_m = 563.3$ million into equation (7), the prediction equation of population growth is obtained:

$$p(t) = \frac{563.3}{(1 + (563.3 \div 436 - 1)e^{-0.169t}} \tag{9}$$

Next ,by using the equation (9), our team predict the number of native Spanish speaker in the future 50 years.

Table 2: Prediction the number of native Spanish speaker[1]

| Year | Population($10^6$) | Year | Population($10^6$) | Year | Population($10^6$) |
|---|---|---|---|---|---|
| 2020 | 326.43 | 2038 | 562.89 | 2056 | 586.75 |
| 2023 | 393.13 | 2041 | 572.29 | 2059 | 587.32 |
| 2026 | 450.04 | 2044 | 578.27 | 2062 | 587.67 |
| 2029 | 494.31 | 2047 | 582.03 | 2065 | 587.88 |
| 2032 | 526.35 | 2050 | 584.38 | 2068 | 588.02 |
| 2035 | 548.34 | 2053 | 585.84 | | |

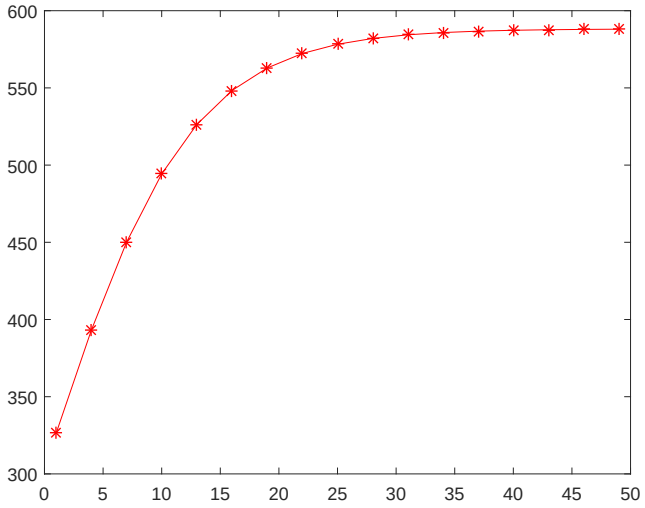


Figure 2: The Prediction of the number of native Spanish speakers

### 4.1.2 Multiple regression model

We apply the second sub-model to explore the regularity of population change, and population here refers to the people who regard one of the ten languages (Mandarin, Spanish, English, etc.) as the second language. The number of people attracted by the second language greatly depends on the factors such as government promotion, immigration, studying abroad and so on. Further speaking, whether a language can become a second language in other countries depends on the country's level of technological, economic and cultural development and its impact on other countries. Therefore, in order to find out the relationship between the number of second language and the above factors, we use analytic hierarchy process and multiple regression method to solve the problem. We propose AHP method to explore the relationship between economy, technology, culture and their respective sub-indicators, while multiple regression analysis is used to analyze the relationship between these three indexes and the final second language population. Specific solution process is as follows.

**1.Analytical Hierarchy Process**

First, establish a hierarchical analysis model , as shown in Fig.3.

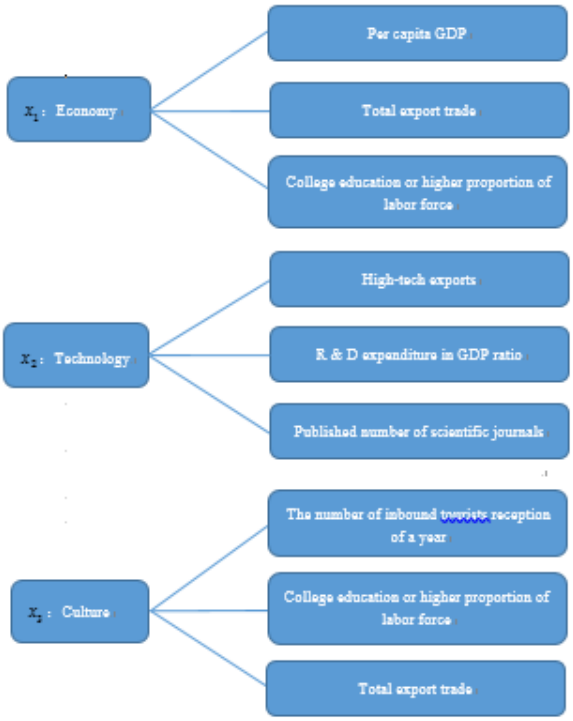


Figure 3: The impact indicators of second language speakers

Based on the calibration scale table, we can get the hierarchical relationship shown below.

**Economy:**

$$\begin{pmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{pmatrix}$$

**Science and Technology:**

$$\begin{pmatrix} 1 & 2 & 3 \\ 1/2 & 1 & 2 \\ 1/3 & 1/2 & 1 \end{pmatrix}$$

**Culture:**

$$\begin{pmatrix} 1 & 2 & 5 \\ 1/2 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{pmatrix}$$

On the analysis of economics, by using Matlab, the result can be tested by consistency. The weights are respectively 0.6370, 0.2573 and 0.1047.

In the same way, for science and technology and culture, each result accords with the consistency. Therefore, the relations between the economy, science and technology, culture and their respective indexes are as follows:
**Economy:** $x_1 = 0.6370x_{11} + 0.2573x_{12} + 0.1047x_{13}$
**Science and Technology:** $x_2 = 0.5396x_{21} + 0.2970x_{22} + 0.1634x_{23}$
**Culture:** $x_3 = 0.5816x_{31} + 0.0.3090x_{32} + 0.1095x_{33}$

**2.Multiple Regression Model Algorithm**
our final goal is to gain a relationship between people whose second language is one of the ten languages and the three indicators.

It is easy to find the value of these three indexes' sub-indexes. Therefore, we think that we can find rules from past data and use multiple regression models to solve problem. After processing the data, we get the economic, scientific and cultural values according to the relational expressions obtained by using the analytic hierarchy process. Next, we use the SPSS software for multiple linear regression to get linear regression coefficients, and finally get the corresponding expression. The results are as follows.

Taking Chinese as an example, we use SPSS to fit the multiple linear regression equation. For the people whose second language is Mandarin,the variation of their quantity is in accordance with the following formula.

$$y = 1.842x_1 - 0.742x_2 - 0.359x_3 \tag{10}$$

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 182.555 | 217.482 | | .839 | .433 |
| | economy | 4.610E-10 | .000 | 1.842 | .940 | .384 |
| | technology | -1.279E-9 | .000 | -.742 | -.396 | .706 |
| | culture | -3.932E-5 | .000 | -.359 | -.968 | .370 |

a. Dependent Variable: number

Figure 4: The coefficients of multiple linear regression

## 4.2 The results of the model

### 4.2.1 Model the distribution of various speakers

It is not hard to imagine that spatial distribution can take advantage of changes over time in the number of languages, so our team first try to model the number of people over time. In order to predict the distribution of various speakers, our team is committed to solving it by quantitatively describing the temporal and spatial variations of the number of speakers. Therefore, we elaborate on this issue from two aspects: time distribution and spatial distribution.

With the model just found out and data from previous years, we draw three maps based on changes of the number of people in previous years.

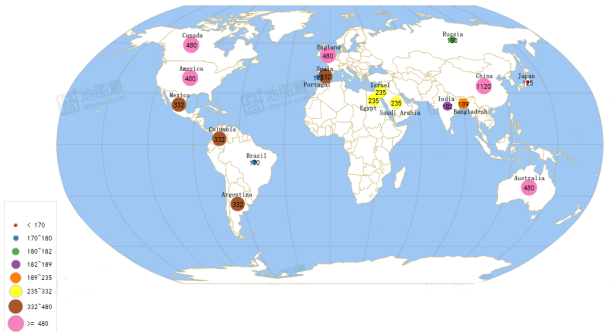

Figure 5: The distribution of various language speakers in 2010s

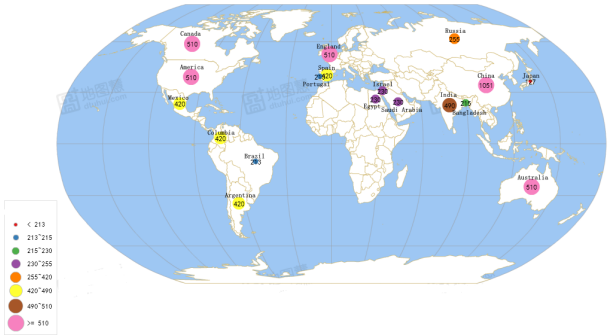

Figure 6: The distribution of various language speakers in 2020s

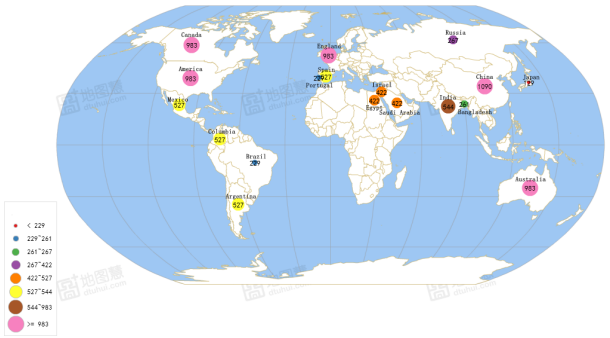

Figure 7: The distribution of various language speakers in 2030s

We can see from the three pictures(refer to Fig.5,6,7), changes for area in the use of major languages is not large. Mandarin has the largest amount of users all the time, which mostly gather in China. North America, Oceanic, England and other areas of English have a continuous growing population. Spanish (Spain, Mexico, Colombia, Argentina), Portuguese (Portugal, Brazil), Hindustani (India) users are on the rise.

### 4.2.2 Predict the change of given language speakers

Use the model result above, we have projected the demographic changes in various languages in the next 50 years by setting the economy, technology and culture indicators value. We get the following conclusion. Referring to the Fig.8,

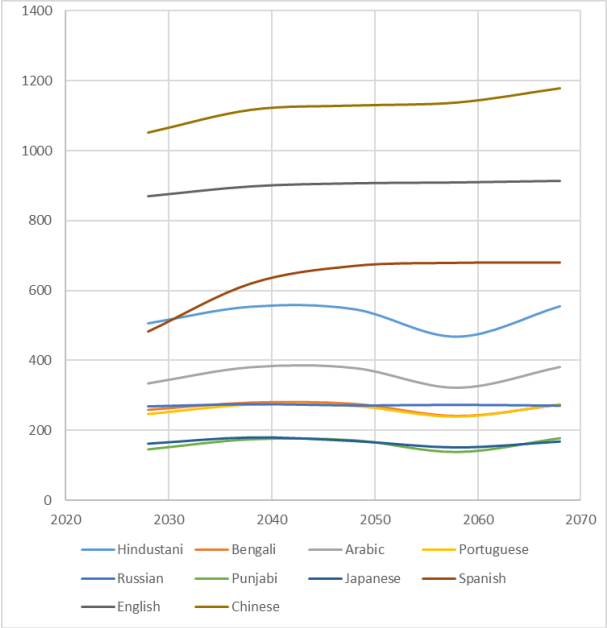


Figure 8: The change trend of the numbers of total language speakers in the next 50 years[7]

it is not difficult to find that in the next 50 years, almost given ten languages are maintaining a slow growth trend, and the relative amount is relatively stable. So the possibility that a language will be replaced by another may be tiny. However, a tiny difference can be found that different from other languages, about in the year of 2048, the population of Hindustani, Arabic, Portuguese and Punjabi show a downward trend, but immediately they rise again. If we go deeper into the analysis, this trend seems to be understandable. The countries that use these languages are almost developing countries having great potential for development, like India and Brazil. And in 2048, they may have undergone a phase of rapid growth and get into the painful period of economic restructuring. Economic and other aspects of growth temporarily slow down. After the completion of economic restructuring, the pace of development has picked up and the number of second language learners in other countries also rise again.

### 4.2.3 Describe the change of language geographic distributions

In view of the global population and population migration patterns, our team analyze the geographic distributions of these languages change.

In order to simplify the model, here we do the assumption:

**For a particular language, the geographical distribution changes only considered as the changes of second language population distribution.** Because the native speakers almost live in their country, we ignore the small amount of population change.

In order to study the relationship between the change of the future global population and the geographic distributions of these languages, we first predict the global population and population migration patterns in the next 50 years.We look up the data of the global population changes in recent years on the web(refer to Fig.9).

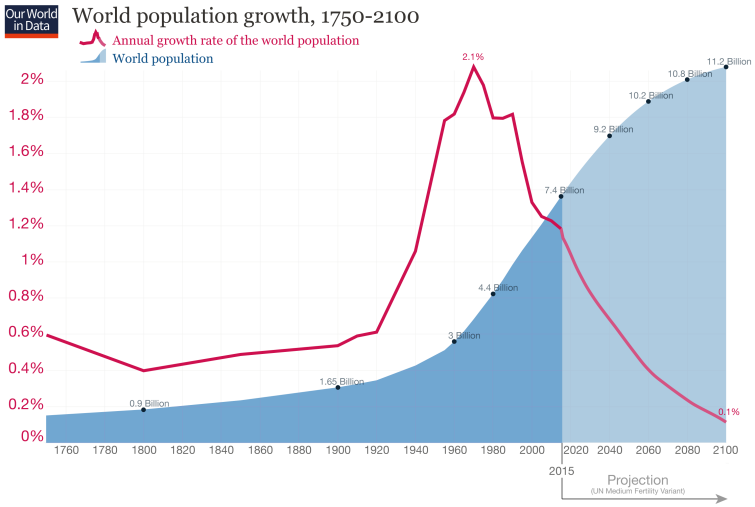


Figure 9: World Population Growth Trend[5]

From Fig.10, we can see that from 2007 to 2011ïijŇthe number of migrants showed a downward trend, and form 2011 to 2016,the number of migrants increased significantly. Considering the trend of globalization, the number of migrants still shows an upward trend in the future.

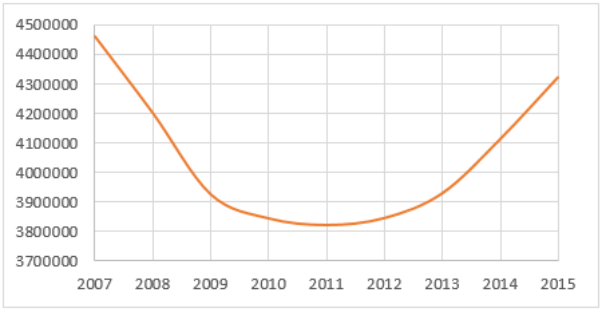


Figure 10: Total number of permanent immigrants into selected OECD countries[4]

For the population migration in a particular language countries, we look up the population migration data of 2007-2015 years and fit the trend of theirs de-

velopment. According to Fig.12, the migration population of most specific language countries remains basically stable, even though slight fluctuations exist. We can see that there are an upward trend in the future. Among them, Spanish countries have a significant downward trend in the number of their immigration. In addition, more immigrants migrate to English speaking countries, and these countries become the first choice of immigrants(refer to Fig.11).

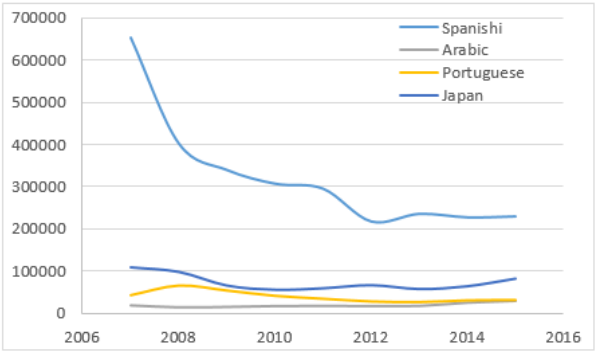


Figure 11: The number of permanent immigrants into the given language countries[6]

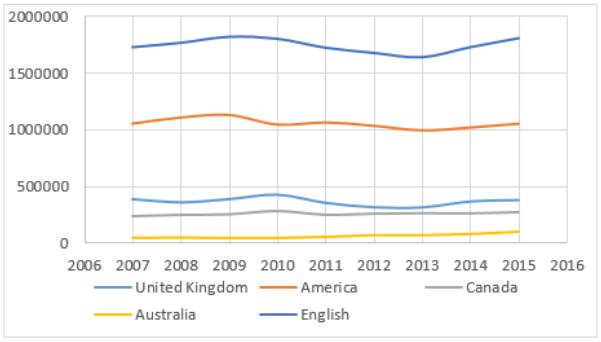


Figure 12: The number of permanent immigrants into the countries that speak English[6]

In order to describe the correlation between the geographical distribution of language and future population and its migration, we use the index of the correlation between the two of them. Depending on the prediction data, we use SPSS to analyze the relevance between migration population and second language speakers. Taking English for example, we first standardize the data and draw the trend of the two variables.

According to Fig.13, we discover that the trend of the two variables is basically the same. Thus, we can consider that the change in geographical distribution of languages is approximately consistent with the trend of global population migration. Combined with the data from several other languages, the changes of geographical distribution of languages are as follows.

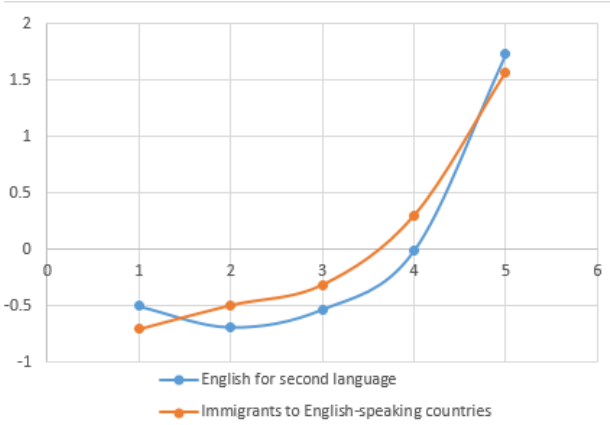


Figure 13: The relevance between migration population and second language speakers

- Some languages will have a small increase in their users and the distribution will be wider, such as English, Chinese.
- For other languages, such as Punjabi, Bengali, the speakers will reduced slightly. And the geographical distribution will also show the down trend.
- Meanwhile, taking Spanish and Arabic for example, the speakers and the geographical distribution will maintain stable basically.

## 4.3 Location and language options for new offices

Based on the above model, we comprehensively consider the influence of various factors on language development, and quantitatively describe the number of speakers in a particular language and the change of geographical distribution. Next, we will use the information obtained from the above model to determine the location of the multinational corporationâĂŹs offices and their language choice.

### 4.3.1 The Choice of Language

We use the established model to predict the tendencies in the number of speakers mentioned above. Considering that the development of multinational service companies depends on the number of customers, we regard the number of native speakers and second language speakers in a given language as a measure of language choice in establishing new offices.

In addition to offices already established in New York, USA and Shanghai, China, we intend to select six languages among the remaining eight languages except English and Mandarin, and then select the appropriate office location in the corresponding country.

Based on the results from the above model, we fit the trend of the number of speakers in the remaining 8 languages(refer to Fig.14). The emphasis on the choice of language, in short term and long term, is different. Because population changes are almost stable in the short term, there may be great changes in the long term. In the short-term analysis, we mainly focus on the number of speakers in a particular language at the present stage. The larger the number, the more the potential customers. Establishing an office in the language region should be able to run well. In the long-term analysis, we are more concerned about the trend of the number of speakers in particular languages. If the overall trend shows a steady upward trend, it indicates that the comprehensive strength of the language speaking region is improving, and the demand for multinational companies will also increase.

Therefore, in the short term, the six languages we choose are as follows:

**Spanish, Hindustani, Arabic, Russian, Portuguese, Bengali**

In the long term, the six languages we choose are as follows:

**Spanish, Hindustani, Arabic, Russian, Portuguese, Japanese**

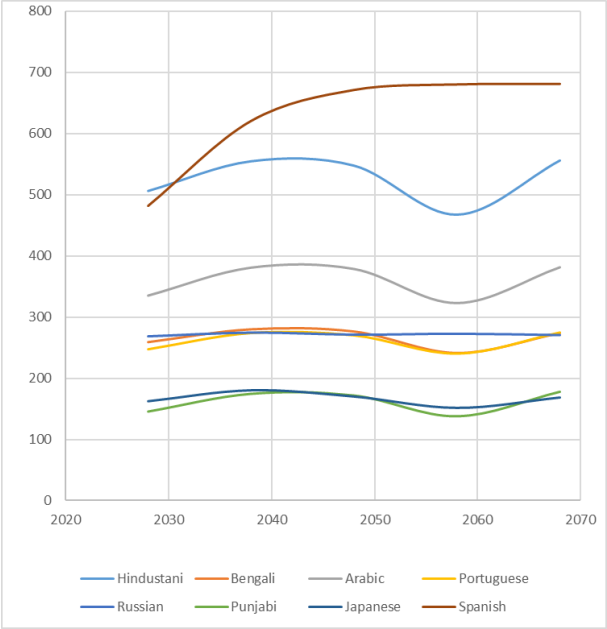


Figure 14: The number of the second language speakers[1]

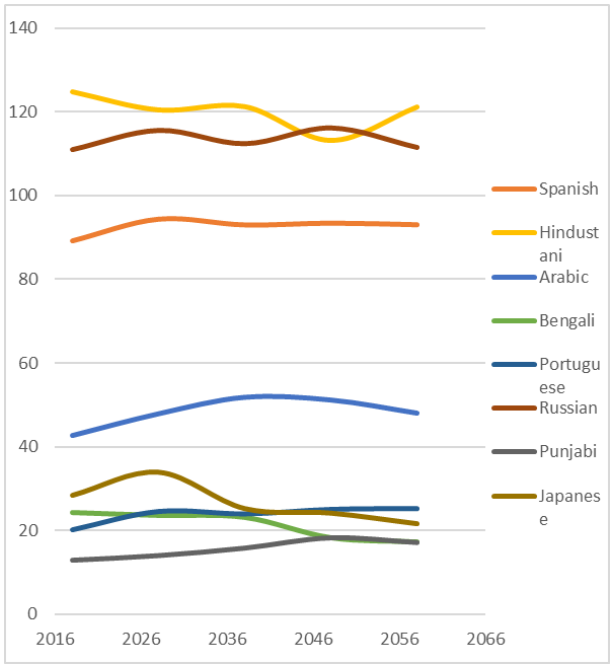


Figure 15: The number of the second language speakers[1]

### 4.3.2 Determination of the Position of the Office

In the selected six languages, we examine the comprehensive indicators of these languages speaking countries and choose a representative and influential country in each language speaking countries. After determining the specific country, analyze the influence and economic lever of the cities in the country.

After searching the corresponding information online, we take the report of the 2017 global city competition list as a long-term basis, while the short-term basis is the ranking of the global cities in 2018. We assume that the capital is most comprehensively powerful city if the country is not listed in the report. Therefore, our final advice to the customer company is as follows:
In the short term, the locations and language choice for the six new offices are shown in the following table3.

Table 3: The locations and language choice in the short term

| Areas | Mexico | Bombay | Dubai | Moscow | St.Paul | Tokyo |
|---|---|---|---|---|---|---|
| Language | Spanish | Hindustani | Arabic | Russian | Portuguese | Japanese |

In the long term, the locations and language choice for the six new offices are shown is the following table4.

Table 4: The locations and language choice in the long term

| Areas | Barcelona | Delhi | Doha | Moscow | Lisbon | Tokyo |
|---|---|---|---|---|---|---|
| Language | Spanish | Hindustani | Arabic | Russian | Portuguese | Japanese |

## 4.4 Suggestions for Opening Less Offices

With the development of Internet communication technology, long distance communication is more convenient. In an effort to save client company resources, our team suggests that the company open less than six international offices. In this way, the company can also get better benefits.

In order to make an accurate judgement of the number and location of the office, we need some additional information below:

- **Business types and service crowds of multinational service companies.**
- **The efficiency and operating cost of the International Office.**
- **The level of science and technology communication in a particular language peaking region.**

Considering the improvement of global communication capacity, the need for an international office in a particular language area depends on whether the multinational company can meet the needs of a particular population through network communications.

In some places where communication capacity is strong, traffic is convenient and the economy is developed, it is often possible to solve practical problems through network communication. Using network communication to do services can reduce some expenses, such as housing rent, staff employment fees and so on. It can save company resources and maximize profits.

The premise of reducing the international office is that it does not affect the normal operation of the company. We should first consider the companyâĂŹs business types and service crowds. If the business type does not support remote operations and needs the staff to work locally, the international office must be set. Secondly, when the work efficiency of the international office is low and the operating cost is high, the investment is disproportionate to the profit. It is not necessary to set up the international office. Finally, if the region in which the international office is located has a higher level of communication and science and technology, and the impact of other factors is negligible, the daily work can be done through Internet communication.

# 5 Evaluate of the Mode

## 5.1 Model Testing

To test our model whether it has good predictive effect, we collect data of the total number of the given language speakers, and compare them to our model calculation results.

We use our model of the given language speakers prediction to calculate total number of the top-five language speakers(refer to Tab.5). Then we looked up and

Table 5: The model calculation result of the total number of language speakers

| Year | Chinese | Spanish | English | Hindustani | Arabic |
|---|---|---|---|---|---|
| 2009 | 1052.7609 | 482.2287 | 870.0194 | 506.0991 | 334.8505 |
| 2011 | 1117.3611 | 620.7112 | 897.9833 | 554.4636 | 380.65298 |
| 2013 | 1129.5288 | 671.24658 | 906.6488 | 547.3034 | 378.4062 |
| 2015 | 1138.0366 | 680.14607 | 909.1407 | 467.4791 | 322.54236 |
| 2017 | 1178.3856 | 681.04036 | 913.4968 | 555.6699 | 381.1806 |

sorted the data of total number of the top-five language from 2009-2017 year(refer to Tab. 6).

Table 6: The historical data of the total number of language speakers[1]

| Year | Chinese | Spanish | English | Hindustani | Arabic |
|---|---|---|---|---|---|
| 2009 | 1001.8 | 490.2 | 810.2 | 535.2 | 319.2 |
| 2011 | 963.2 | 650.7 | 823.4 | 521.3 | 332.4 |
| 2013 | 1000.2 | 600.4 | 842.1 | 536.7 | 341.5 |
| 2015 | 1240.1 | 642.5 | 835.2 | 524.7 | 330.1 |
| 2017 | 1000.1 | 650.1 | 891.5 | 530.1 | 346.9 |

Compare the historical data with our model calculation results, we figure out the relative error(refer to Tab.7). By analyzing the relative error, the relative error of most calculations is within 10%. Considering the accuracy of the data, it is almost reasonable that our model can efficiently predict the change of the given language speakers.

Table 7: The relative error of model calculation result

| Year | Chinese | Spanish | English | Hindustani | Arabic |
|---|---|---|---|---|---|
| 2009 | 4.84% | -1.65% | 6.88% | -5.75% | 4.67% |
| 2011 | 13.79% | -4.83% | 8.31% | 5.98% | 12.68% |
| 2013 | 11.45% | 10.55% | 7.12% | 1.94% | 9.75% |
| 2015 | -8.97% | 5.53% | 8.13% | -12.24% | -2.34% |
| 2017 | 15.13% | 4.54% | 2.41% | 4.60% | 8.99% |

## 5.2 Sensitivity Analysis

In the first sub-model, we use the Logistic prediction analysis. And the model result depends on the parameters:$r$, $p_m$. Before, we used the historical data to fit the values of these two parameters.The population growth forecast by the model first rises and then stabilizes basically.

Our sensitivity analysis will be based on factor variation method, in order to see how the result of the first sub-model changes when the input parameters change. We will mainly analyze the parameter growth rate $r$.

We choose to predict the changes of native speakers in Chinese. We change the parameter $r$: $r = 0.16$, $r = 0.26$, $r = 0.36$. Calculation results are as follows.

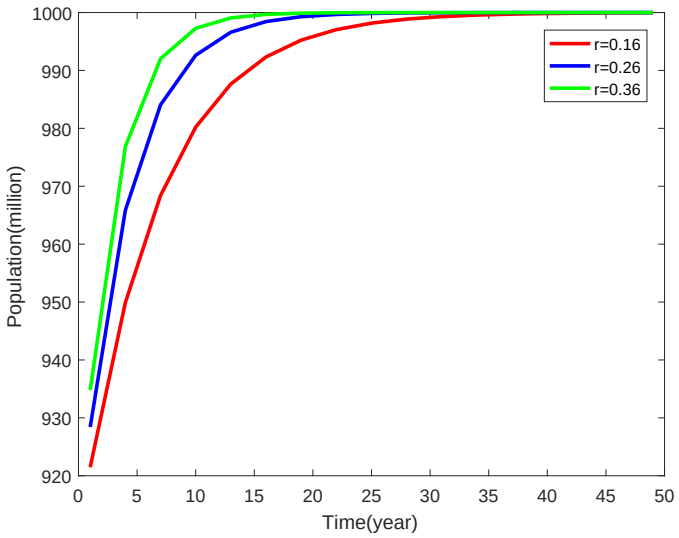


Figure 16: The sensitivity analysis of the parameter $r$

Referring to the Figue16, we find that the parameter $r$ mainly decide the speed of population growth. And the larger the value of the parameter $r$, the faster the population grows. But the quantity of population eventually tends to a maximum stable value. The size of this value has nothing to do with the parameter $r$.

# 6 Strengths and weaknesses

## 6.1 Strengths

- **Applies widely**
  Our model can be used for other language prediction, and it also solves the problem of the lack of data.

- **High accuracy**
  Our model uses a large number of data for experimental analysis, so the analysis results have credibility. Analysis and forecasts depend on actual data.

- **A certain degree of innovation**
  The traditional model can not solve the problem well, so we use the new prediction model to obtain more accurate results.

### 6.2 Weaknesses

- **Ignore other languages**
  We only consider the most widely used 10 languages for model establishment and forecast, ignore other languages.

- **Some results are not quantified**
  In the process of solving the problem, some results are affected by many factors and it is difficult to describe quantitatively. So we can only characterize these changes qualitatively.

## 7 Memo

To Chief Operating Officer:

Our team have been working on the investigation and analysis of global language trends, as well as choosing suitable locations for your new offices these days, and we achieved the desired result.

Combined with the data of the past 20 years, we conduct a series of mathematical analysis and forecast, we finally come to these conclusions.

1. At present, the ten most spoken languages in turn are Mandarin Chinese, Spanish, English, Hindustani, Arabic, Bengali, Portuguese, Russian, Punjabi and Japanese, by our forecast model, these 10 languages will still occupy the top ten list in the next 50 years. Not only that, the relative rankings between them will not change much either. This also greatly reduces the contingency of your company's new office location problem

2. As for the population migration, according to the historical data of population migration and our mathematic model, we estimate that among the countries where the top 10 languages are spoken, most of the immigrants of them remained stable with slight fluctuations, and the overall trend is showing an upward trend. Although, there will be some exceptions, for

example, Spanish immigrants have shown a clear downward trend. In addition, the migrants that move into English-speaking countries are of huge amount, which dominates the world trend of immigration.

Based on the above analysis, some of our suggestions on the location of your new office are as follows.

1. On condition that your company will establish 6 offices, our suggestions can be divided into two parts: one suggestion is for the short-term, we recommend that your company establish offices in Spanish, Hindustani, Arabic, Russian, Portuguese and Bengali speaking countries. Corresponding to this, we have also 6 cities to advise, they are Mexico city, Bombay, Dubai, Moscow, Sao Paulo and Tokyo; the other is long-term suggestion, the 6 languages suggested are Spanish, Hindustani, Arabic, Russian, Portuguese and Japanese. Corresponding to these languages, six cities are advised as well, they are Barcelona, Delhi, Doha, Moscow, Lisbon and Tokyo. This could maximize the potential customers.
2. If the number of offices is not limited, our suggestion is that it could be less than 6, this could be implemented only if the following indicators are suitable.
    - Your company business can be operated remotely and network communications could meet the needs of specific groups of people.
    - The efficiency and operating cost of the International Office.
    - The level of science and technology communication in a particular language-speaking region.

In some places with strong traffic capacity, convenient transportation and economically developed, it is often possible to solve practical problems through network communications. Using network communications services can reduce some costs, such as housing rent, employee employment fees and so on. It saves company resources and maximizes profits.

These are our suggestions. We hope that it can be of help to your company business.
Thank you.

# 8 Conclusions

In this paper, we aim to devise a model evaluating and prediction global language development trend. Meanwhile, we propose the method of location and language options for new international offices.Thus, we mainly develop a model of the given language speakers prediction. The model consists of two sub-model: the Logistic prediction model for native speakers and the multiple

regression model for second language speakers. Firstly,for a given language, we calculate depending on the data and get the change trend of native speakers by sub-model1. By the sub-model2, we obtain the number of second language speakers and the change trend.

Next, we combine the prediction result with some other factors and model the distribution of various language speakers over time. We discover that the geographic distributions of these languages change simultaneously. The change is almost consistent with the trend and direction of population migration. To prove it, we conducted a correlation analysis of population migration with second language speakers.

And then, we use the information obtained from the above model to determine the location of the multinational corporationâĂŹs offices and their language choice. As to the short term versus the long term, we mainly focus on different aspects and propose suggestions of the location and language of new international offices. In order to save the company resources, we recommend opening less than six offices when the region communications capabilities meet the company business needs.

# References

[1] a.http://www.vistawide.com/languages/top30languages.htm
b.http://research.omicsgroup.org/index.php/List of languages by total number of speakers
c.https://en.wikipedia.org/wiki/List of languages by number of native speakers
d.http://dailynewsdig.com/top-ten-spoken-languages-world-2014/
e.http://app.fanyi.baidu.com/static/passage/2016-08/2016-08-14/000/index.html

[2] http://www.imf.org/en/Countries

[3] http://www.latexstudio.net/

[4] http://www.chinatex.org/

[5] Max Roser and Esteban Ortiz-Ospina. ”World Population Growth” , April, 2017. https://ourworldindata.org/world-population-growth population-size

[6] International Migration Outlook 2017.OECD READ edition
http://www.keepeek.com/Digital-Asset-Management/oecd/

[7] https://data.worldbank.org.cn/indicator

[8] http://www.oecd.org/

[9] http://www.imf.org/en/Countries
http://www.un.org/zh/index.html

# Appendices

## Appendix A First appendix

Here are simulation programmes we used in our model as follow.

**Input matlab source:**

```
clear
clc
Y=[900 873  935 955 965 975 897]
T=[1997 2000 2004 2007 2010 2013 2017]
% Linear treatment
for t = 1:7,
   x(t)=exp(-0.16*t);
   y(t)=1/Y(t);
end
%  Computing regression equation: y'=a+bx',solving the values of a and b.
c=zeros(7,1)+1;
X=[c,x'];
B=inv(X'*X)*X'*y'
for i=1:7,
% Calculate the regression fitting value
    z(i)=B(1,1)+B(2,1)*x(i);
    z(i)=B(1,1)+B(2,1)*x(i);
    s(i)=y(i)-sum(y)/7;
    w(i)=z(i)-y(i);
end
S=s*s';
Q=w*w';
U=S-Q;
F=5*U/Q
for j=1:7,
    Y(j)=1/(B(1,1)+B(2,1)*exp(-0.16*j));
end
% The fitting curve of the output nonlinear regression model is presented.
plot(T,Y)
```