# 数学实验作业二

|  |  |  |  |
| --- | --- | --- | --- |
| **姓名** | 刘兆宇 | **班级** | 大数据管理与应用4班 |
| **学号** | 202330860601 | **时间** | 2024.03.27 |

1. **实验目的和内容**
2. 引入部分和数列及相关数列，作图并观察图像变化，从而讨论调和级数 的变化规律。
3. 根据表格数据分析1990-2010年间我国人口增长的规律，并以此预测2011和2012年的人口数量，然后与实际人口数量作对比评价预测模型的优劣，并对我国人口政策提出建议。
4. **实验过程**
5. **讨论调和级数 的变化规律**
6. 画出部分和数列变化的折线图，观察变化规律；
   * + 1. 问题分析：

根据无穷级数的有关概念，部分和 ，故首先需要**定义求解调和级数部分和的函数**，可以通过两种方法实现：

1. 使用内置函数：[cumsum(A)](https://www.mathworks.com/help/releases/R2023b/matlab/ref/cumsum.html?overload=cumsum%20false&doclanguage=zh-CN&nocookie=true&prodfilter=ML%20SL%20BL%20C2%20VP%20CT%20CF%20DA%20DB%20DF%20DD%20DH%20NN%20ET%20EC%20IT%20FI%20FL%20GD%20GC%20IA%20IP%20OT%20IC%20MG%20ME%20CO%20MJ%20MR%20TE%20DX%20MT%20OP%20DM%20PD%20PW%20PM%20RL%20RQ%20RK%20SQ%20SG%20SE%20VV%20CI%20RT%20SK%20SD%20CV%20SO%20DV%20WT%20FA%20PL%20XP%20SR%20SZ%20EL%20SF%20ST%20SM%20ZC%20ID%20TA%20WA%20WB&docviewer=helpbrowser&docrelease=R2023b&s_cid=pl_webdoc&loginurl=https%3A%2F%2F127.0.0.1%3A31515%2Ftoolbox%2Fmatlab%2Flogin%2Fweb%2Findex.html%3Fsnc%3DFONQCH%26external%3Dtrue%26channel%3D__mlfpmc__&searchsource=mw&snc=FGBVKM&container=jshelpbrowser#d126e278777) 可以返回向量A的累积和；
2. 使用循环语句：for循环累加每一项 ，存储即可；

调用函数之后对所得到的向量plot绘图即可。

* + - 1. 伪代码：

设置数列项数 N;

S\_n = 调用函数(参数N)

% 绘制部分和数列

创建图形窗口;

plot绘图Sn

坐标轴 标题设置

函数 harmonic\_partial\_sum\_1(N) % 方法①内置函数

返回 cumsum(1 / (1~N))

函数 harmonic\_partial\_sum\_2(N) % 方法②循环实现

初始化 S\_n 为长度 N 的零数组

partial\_sum = 0

对于 n = 1~N

partial\_sum += 1 / n

S\_n[n] = partial\_sum

返回 S\_n

* + - 1. 代码：

N = 1000; % 设置数列项数

S\_n = harmonic\_partial\_sum\_1(N); % 调用函数计算部分和数列

%% 绘制部分和数列{S\_n}折线图

figure;

plot(1:N, S\_n, 'b-', 'LineWidth', 2)

grid on; grid minor;

xlabel('${n}$', 'Interpreter', 'latex')

ylabel('${S\_n}$', 'Interpreter', 'latex')

title('部分和数列');

% 法1：使用内置函数

function S\_n = harmonic\_partial\_sum\_1(N)

S\_n = cumsum(1./(1:N));

end

% 法2：使用循环实现

function S\_n = harmonic\_partial\_sum\_2(N)

S\_n = zeros(1, N);

partial\_sum = 0;

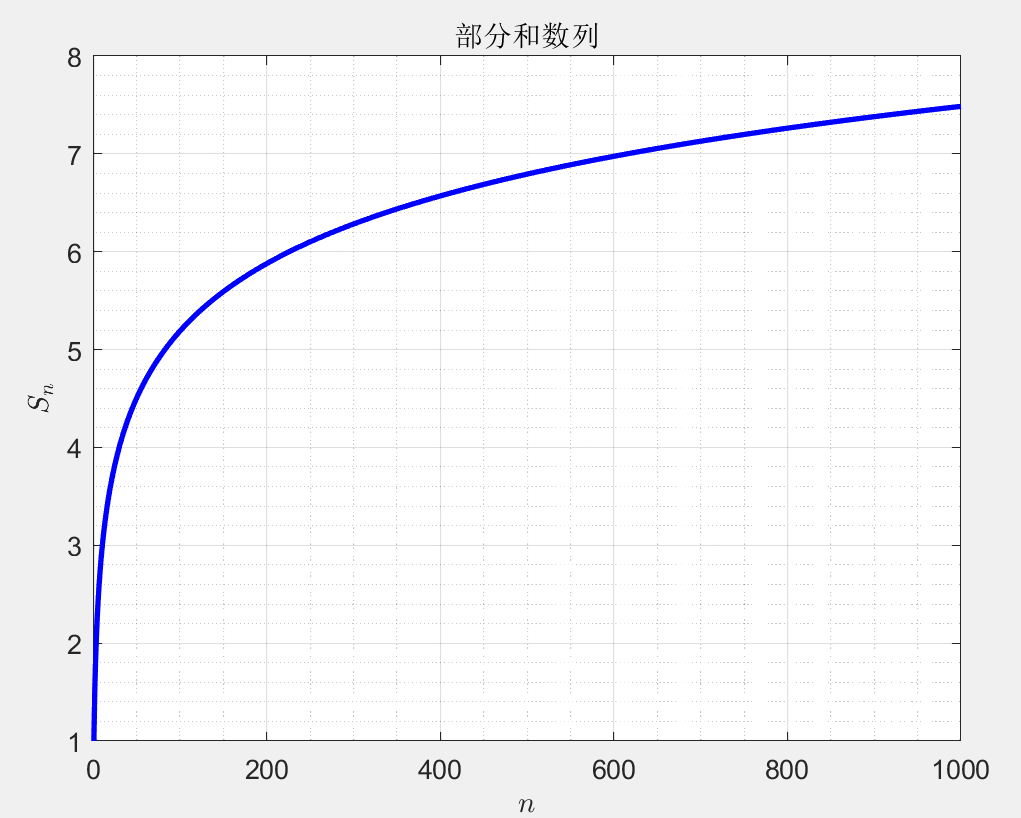
for n = 1:N

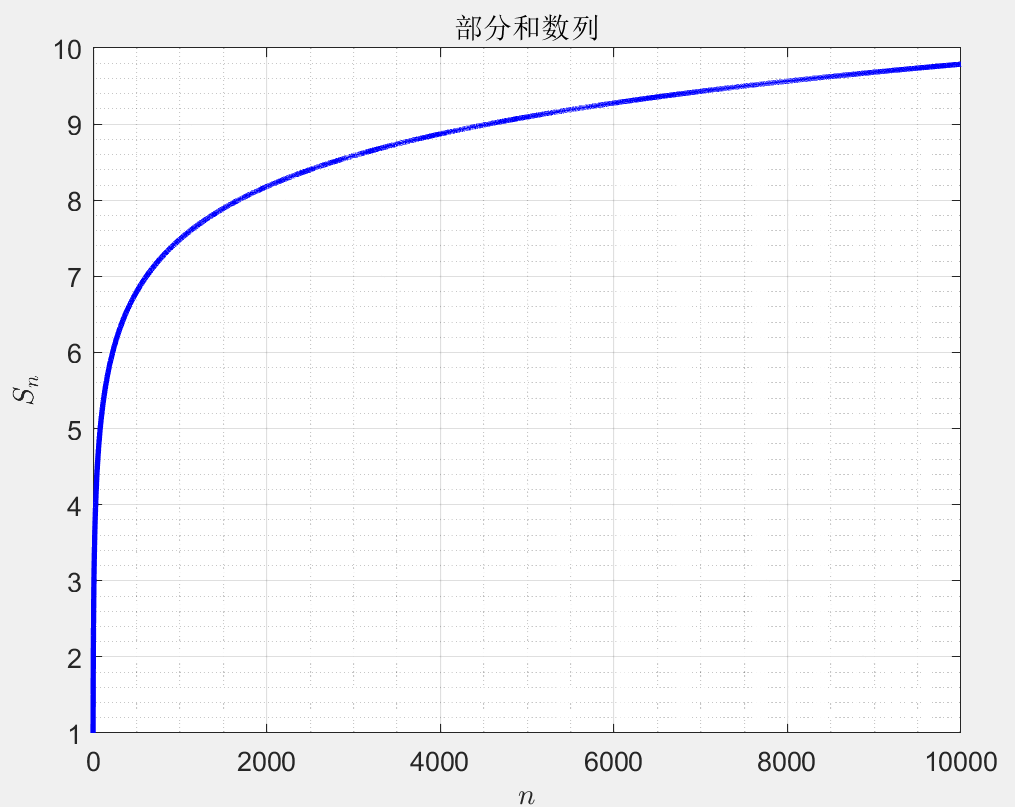
partial\_sum = partial\_sum + 1/n;

S\_n(n) = partial\_sum;

end

end

* + - 1. 运行结果



* + - 1. 结果分析：

从上述两张折线图（；）中，可以观察到调和级数部分和有以下变化规律：

1. 随着的增加而增加，在折线图左侧（较小的区域）曲线斜率较大，增长较快，而在右侧（较大的区域）曲线趋于平坦，表明曲线增长速率逐渐减慢，这说明每个额外项对总和的贡献越来越小；
2. 尽管增长速率减慢，但部分和并没有趋向于一个固定的极限值，可以从图像上初步判断：调和级数是发散的。
3. 引入数列，作图观察其变化，猜测是否有极限；
   * + 1. 问题分析：

调用部分和求解函数，将传入参数修改为2N即可得到，但在作差的时候应注意是中的**偶数项**与逐项作差，得到。

* + - 1. 代码：（伪代码较简单）

N = 1000; % 设置数列项数

S\_n = harmonic\_partial\_sum(N);

S\_2n = harmonic\_partial\_sum(2\*N);

% 绘制 H\_n 的折线图

figure;

H\_n = S\_2n(2:2:end) - S\_n;

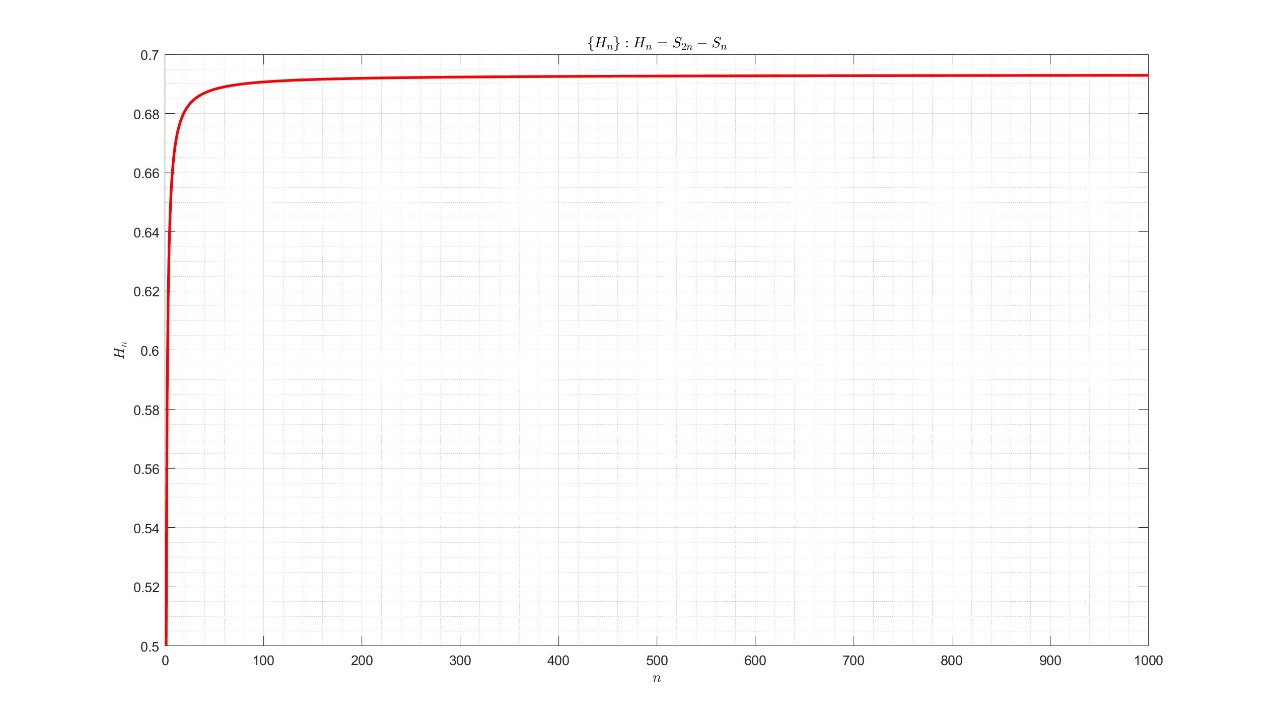
plot(1:N, H\_n, 'r-', 'LineWidth', 2);

grid on; grid minor;

xlabel('${n}$', 'Interpreter', 'latex');

ylabel('${H\_n}$', 'Interpreter', 'latex');

title('${\{H\_n\}:H\_n = S\_{2n}-S\_n}$', 'Interpreter', 'latex');

* + - 1. 运行结果：
      2. 结果分析：

观察图像发现，随着的增加，的值迅速上升并很快趋于平稳，逼近一个固定的值，当时约为0.6926，当时约为0.6931，故猜想数列**存在极限**。

数学求解证明过程如下：

故猜想成立，数列极限存在，极限值为。

1. 引入数列，作图观察其变化，寻找恰当的函数拟合；
   * + 1. 问题分析：

上述问题中已经求解出，选择其偶数项即为。

基本数学事实表明，调和级数的增长与对数增长类似，故可尝试使用对数函数进行拟合，拟合形式为 ，使用polyfit函数提取相应参数，计算拟合曲线即可。

* + - 1. 伪代码：

计算 S\_2n(1~2N的调和级数)

提取 G\_n(取S\_2n的偶数项)

plot绘制图表

对 G\_n 进行拟合:

拟合 G\_n 至对数曲线

polyfit(log(1:N), G\_n, 1);

获取拟合参数 a, b

计算拟合曲线 G\_fit:

对于 i 从 1 到 N:

G\_fit[i] = a \* log(i) + b

hold on

绘制拟合曲线图像

hold off

* + - 1. 代码：

%% 绘制{G\_n}:G(n)=S(2n)并函数拟合

G\_n = S\_2n(2:2:end);

% 绘制 G\_n 的折线图

figure;

plot(1:N, G\_n, 'bo', 'LineWidth', 2, 'MarkerSize', 4);

grid on; grid minor;

xlabel('${n}$', 'Interpreter', 'latex');

ylabel('${G\_n}$', 'Interpreter', 'latex');

title('${\{G\_n\}:G\_n = S\_{2n}}$', 'Interpreter', 'latex');

% 使用对数函数进行拟合

coefficients = polyfit(log(1:N), G\_n, 1);

a = coefficients(1);

b = coefficients(2);

% 计算拟合值

G\_fit = a \* log(1:N) + b;

% 绘制拟合曲线

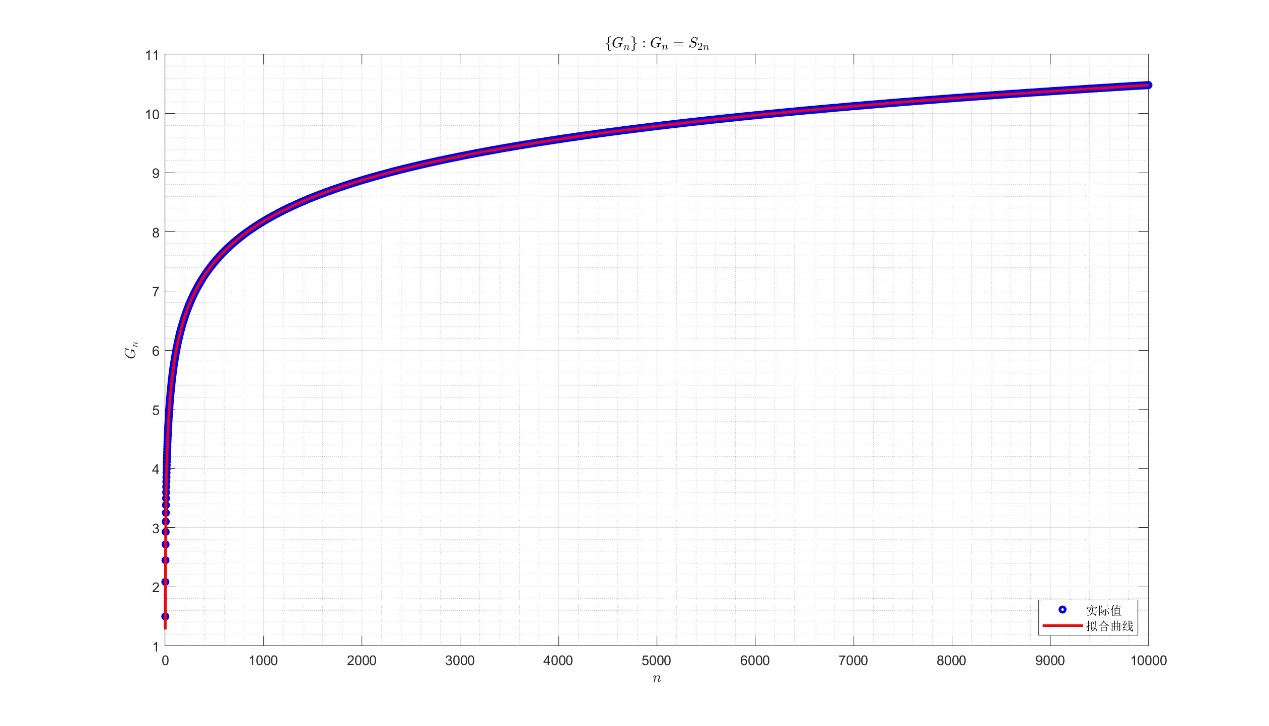
hold on;

plot(1:N, G\_fit, 'r-', 'LineWidth', 2);

legend('实际值', '拟合曲线', 'Location', 'southeast');

hold off;

* + - 1. 运行结果：



* + - 1. 结果分析：

图中实际数据点以蓝色圆圈表示，拟合曲线用红色表示。观察图像发现，数列的变化规律与(1)中部分和数列的变化规律一致，同时对数曲线拟合与实际数据点非常吻合，说明函数拟合非常成功，这也验证了调和级数的对数增长性质。

1. 讨论部分和数列{Sn}的变化规律.

① **递增性**：因为每一项都是正数，随着的增加而单调递增；

② **增长速度**：虽然随着的增加而递增，但它的增长速度逐渐放缓。对于较小的，数列增长较快；对于较大的，每个新增项对总和的贡献越来越小，即增量递减。

③ **发散性**：尽管的增长放缓，但数列并不收敛于某个有限值，即是一个发散的数列。

④ **对数关系**：数学上证明和绘图加以验证，调和级数的部分和与自然对数函数呈近似线性关系，即可以通过加上一个常数（欧拉常数γ）来近似，也即对于足够大的，有。

1. **人口增长规律分析及预测模型的构建与评估**
   * + 1. 问题分析

**表 1990-2010 年我国的人口数量(万)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 年份 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 |
| 数量 | 114333 | 115823 | 117171 | 118517 | 119850 | 121121 | 122389 | 123626 |
| 年份 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
| 数量 | 124761 | 125786 | 126743 | 127627 | 128453 | 129227 | 129988 | 130756 |
| 年份 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |  |
| 数量 | 131448 | 132129 | 132802 | 133450 | 134091 | 134916 | 135922 |  |

1. 首先需要提取表格数据，根据表格人口数据描点连线作图，即可得到人口统计年鉴图，从而可以初步**观察**到人口数量的变化规律；
2. 为了更准确地分析人口增长规律，可以**定量计算**人口的年均增长量和增长率（相邻年份之间），在MATLAB中，可以使用uitable函数来创建一个在图形用户界面（GUI）窗口中显示的表格，来呈现数据；
3. 为了更好的呈现数据结果，可以采用plot将**表格数据**（增长量和增长率）**可视化**，由于增长量和增长率的数据范围差异较大，所以采用在两个不同的y轴上分别绘制。其中yyaxis left 和 yyaxis right 函数允许为同一个x轴指定两个不同的y轴，使得两组数据可以在同一个图表上比较而不会因为数值范围的差异而影响可读性；
4. 接下来要根据已有数据和规律（1990-2010年）建立人口数量预测模型，从而预测2011和2012年的人口数据：常见的人口模型有Malthusian Model（马尔萨斯模型）、Logistic增长模型（S型曲线）等，但由于我国人口增长受国家政策影响较大，指数特性增长不显然（从年鉴图中也可观察到更偏向于线性），故而可以采用polyfit**多项式拟合**，在本次实验中，尝试一阶拟合（线性回归）和高阶多项式拟合，根据预测曲线计算出2011和2012年的预测人口数据，并在[中国统计年鉴 - 国家统计局 (stats.gov.cn)](https://www.stats.gov.cn/sj/ndsj/)上查询这两年的实际人口数据，从而观察比较拟合效果，并同时提防数据过拟合的问题；
5. 为了定量评估模型的拟合效果，还需要计算预测模型的准确性。为此，引入了3个模型评价指标：均方误差(MSE)、决定系数(R2)、平均相对误差百分比(ARE%)，由此可以定量比较各阶多项式的拟合效果。具体计算公式如下：
6. 最后根据上述结果和总计的规律，尝试对我国人口政策提出建议。
   * + 1. 代码：

% 数据提取

years = 1990:2010;

population = [114333, 115823, 117171, 118517, 119850, 121121, 122389, 123626, ...

124761, 125786, 126743, 127627, 128453, 129227, 129988, 130756, ...

131448, 132129, 132802, 133450, 134091];

%% 人口统计年鉴

figure;

plot(years, population, 'bo-','LineWidth', 1, 'MarkerSize', 5, 'MarkerFaceColor', 'b');

title('\bf人口统计年鉴', 'FontSize', 14);

xlabel('年份', 'FontSize', 12);

ylabel('人口数量 (万)', 'FontSize', 12);

xticks(1990:1:2010); % 设置x轴刻度为每年一个刻度

grid on; grid minor;

%% 定量分析人口增长规律（绘制表格）

% 计算每年的增长量

growth\_amount = diff(population);

% 计算每年的增长率

growth\_rate = (growth\_amount ./ population(1:end-1)) \* 100;

% 从第二年开始计算增长量

adjusted\_years = years(2:end);

% 创建表格数据

growth\_data = table(adjusted\_years', growth\_amount', growth\_rate', ...

'VariableNames', {'年份', '增长人口数量 (万)', '年增长率 (%)'});

% 将表格数据转换为单元数组

cell\_data = table2cell(growth\_data);

% 估计uitable大小

table\_height = 22 \* (size(cell\_data, 1) + 1);

% 创建figure窗口

fig = figure('Name', 'Population Growth Table', 'NumberTitle', 'off', ...

'Position', [100, 100, 520, table\_height + 50]);

% 显示表格

t = uitable('Parent', fig, ...

'Data', cell\_data, ...

'ColumnName', growth\_data.Properties.VariableNames, ...

'RowName', [], ...

'ColumnWidth', {80, 125, 80}, ...

'Position', [20, 20, 480, table\_height]);

%% 绘制增长量和增长率（可视化）

figure;

% 创建左侧 y 轴用于绘制增长量

yyaxis left;

plot(adjusted\_years, growth\_amount, 'b-o', 'LineWidth', 1.5, 'MarkerSize', 5, 'MarkerFaceColor', 'b');

ylabel('增长人口数量 (万)');

% 创建右侧 y 轴用于绘制增长率

yyaxis right;

plot(adjusted\_years, growth\_rate, 'r-o', 'LineWidth', 1.5, 'MarkerSize', 5, 'MarkerFaceColor', 'r');

ylabel('年增长率 (%)');

% 设置通用图表属性

title('\bf人口增长规律分析');

xlabel('年份');

xticks(1990:1:2010);

legend({'增长人口数量 (万)', '年增长率 (%)'}, 'Location', 'northeast');

grid on; grid minor;

%% 模型：线性回归预测

predicted\_years = 2011:2012;

poly\_order = 1; % 线性回归

p = polyfit(years, population, poly\_order); % 线性回归拟合

% 使用拟合模型预测2011年和2012年的人口数量

predicted\_2011 = polyval(p, 2011);

predicted\_2012 = polyval(p, 2012);

predicted\_population = [predicted\_2011, predicted\_2012];

% 输出预测结果

fprintf('Predicted population for 2011: %f\n', predicted\_2011);

fprintf('Predicted population for 2012: %f\n', predicted\_2012);

% 将预测值绘制在图表上

extended\_years = 1990:2012;

% 计算预测人口数

extended\_population = polyval(p, extended\_years);

% 实际人口数量

actual\_2011 = 134916;

actual\_2012 = 135922;

actual\_population = [actual\_2011, actual\_2012];

% 绘制原始数据和预测线

figure;

% 实际人口数据

plot(years, population, 'bo', 'MarkerSize', 5, 'MarkerFaceColor', 'b');

hold on;

% 预测线

plot(extended\_years, extended\_population, 'r-', 'LineWidth', 1.3);

plot(predicted\_years, predicted\_population, 'gs', 'MarkerSize', 7.5, 'MarkerFaceColor', 'k');

plot(predicted\_years, actual\_population, 'bo', 'MarkerSize', 5.5, 'MarkerFaceColor','r');

hold off;

% 设置图表的标题和标签

title('\bf人口数量预测', 'FontSize', 14);

xlabel('年份', 'FontSize', 12);

ylabel('人口数量 (万)', 'FontSize', 12);

legend('实际人口数量', '线性预测', '预测人口数量', '实际人口数量(2011 & 2012)', 'Location', 'northwest');

xticks(1990:1:2012);

grid on; grid minor;

%% 模型：多项式拟合预测

poly\_order = 2; % 拟合的多项式阶数

p = polyfit(years, population, poly\_order);

% 使用拟合模型预测2011年和2012年的人口数量

predicted\_2011 = polyval(p, 2011);

predicted\_2012 = polyval(p, 2012);

predicted\_population = [predicted\_2011, predicted\_2012];

% 高阶多项式需要更多的点来平滑曲线

fine\_grained\_years = 1990:0.1:2012; % 使用更细的年份间隔来生成平滑的曲线

extended\_population = polyval(p, fine\_grained\_years);

% 绘制原始数据和预测线

figure;

plot(years, population, 'bo', 'MarkerSize', 5, 'MarkerFaceColor', 'b');

hold on;

plot(fine\_grained\_years, extended\_population, 'r-', 'LineWidth', 1.3);

plot(predicted\_years, predicted\_population, 'gs', 'MarkerSize', 7.5, 'MarkerFaceColor', 'k');

plot(predicted\_years, actual\_population, 'bo', 'MarkerSize', 5.5, 'MarkerFaceColor','r');

hold off;

% 设置图表的标题和标签

title('\bf人口数量高阶多项式预测', 'FontSize', 14);

xlabel('年份', 'FontSize', 12);

ylabel('人口数量 (万)', 'FontSize', 12);

legend('实际人口数量', ['多项式预测(' num2str(poly\_order) '阶)'], '预测人口数量', '实际人口数量(2011 & 2012)', 'Location', 'northwest');

xticks(1990:1:2012);

grid on; grid minor;

%% 模型评估

% 计算残差

residuals = [actual\_2011 - predicted\_2011, actual\_2012 - predicted\_2012];

% 计算均方误差 MSE

MSE = mean(residuals.^2);

% 计算总平方和 SST 和残差平方和 SSE

SST = sum((population - mean(population)).^2) + sum(([actual\_2011, actual\_2012] - mean([population, actual\_2011, actual\_2012])).^2);

SSE = sum(residuals.^2);

% 计算决定系数

R2 = 1 - (SSE / SST);

% 输出模型评估结果

fprintf('均方误差 (MSE): %f\n', MSE);

fprintf('决定系数 (R^2): %f\n', R2);

% 计算每个预测年份的相对误差百分比

relative\_errors = abs(residuals) ./ [actual\_2011, actual\_2012] \* 100;

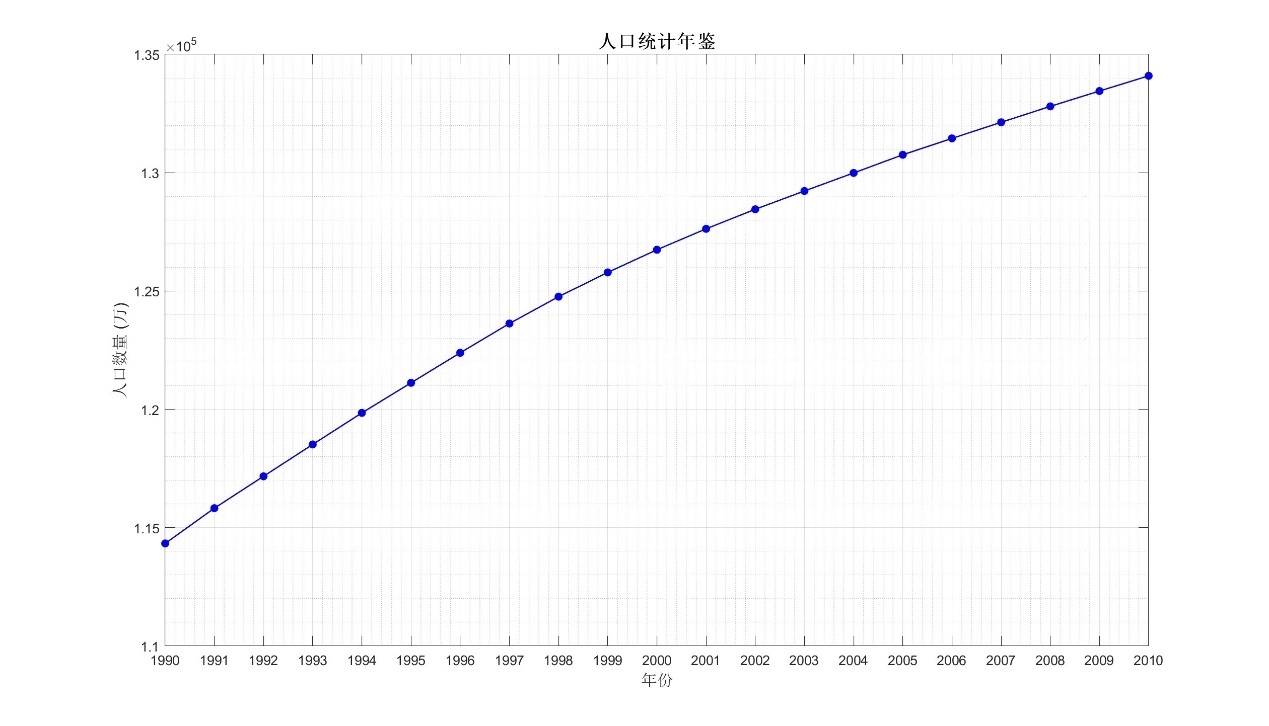
% 计算平均相对误差百分比ASE%

average\_relative\_error = mean(relative\_errors);

fprintf('平均相对误差百分比: %f%%\n', average\_relative\_error);

* + - 1. 运行程序与结果分析：

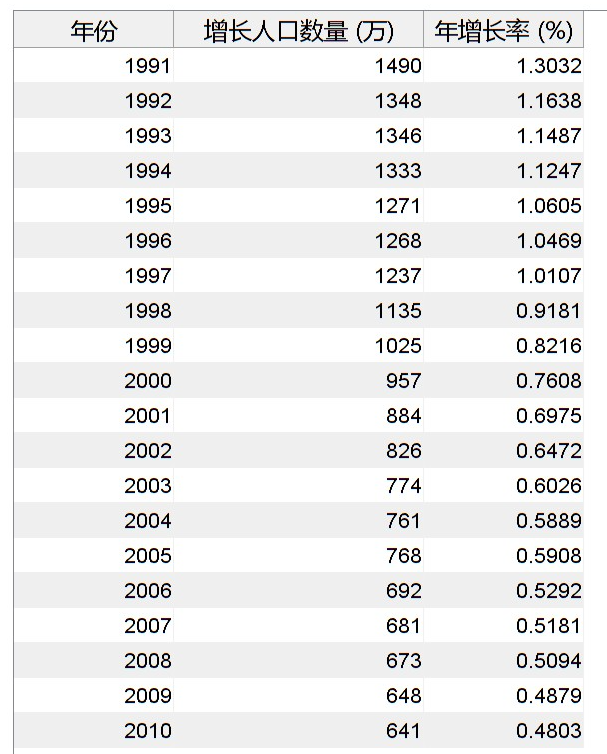
**3.1 人口增长规律**



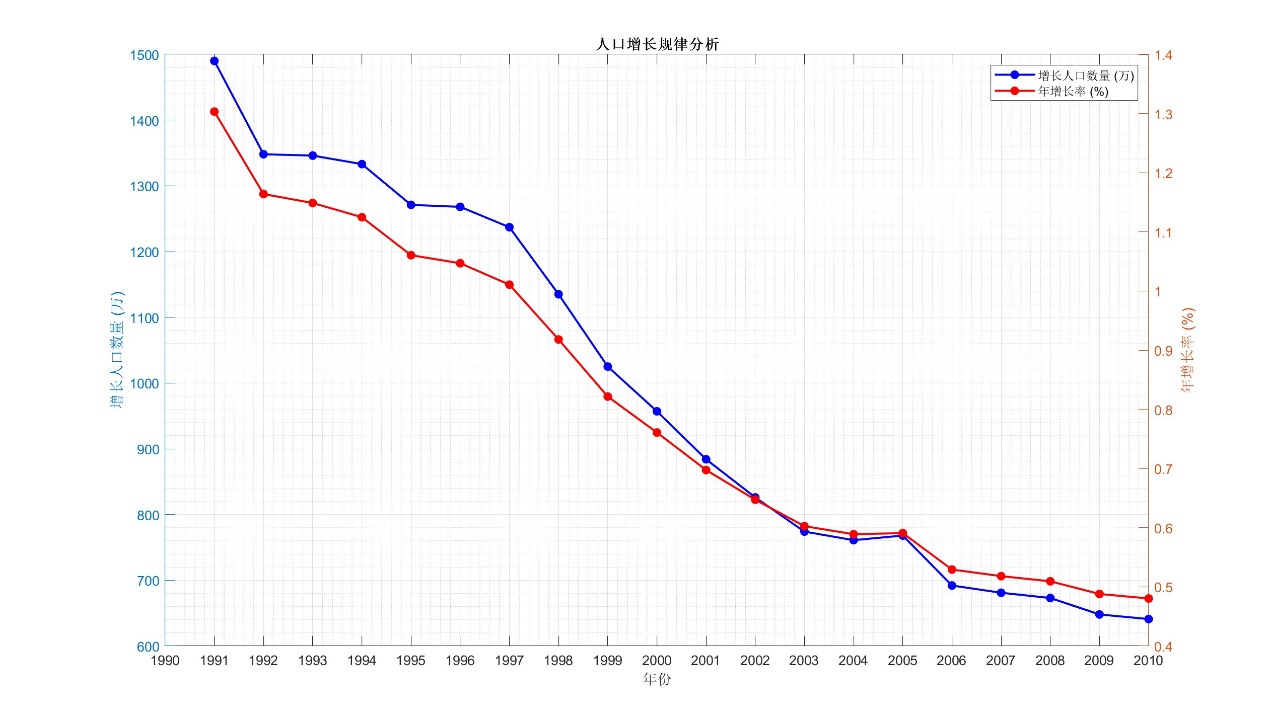
观察上图发现，人口增长显示出以下特点：

1. 线性增长趋势：图标上的，点近似沿一条直线排列，表明人口增长可能接近线性，即1980-2010年每年增加的人口数量接近。
2. 增长速度：从整体上看，人口增长速度相对稳定，没有明显的突增或突降。

**3.2 定量计算年均增长量和增长率（表格）**

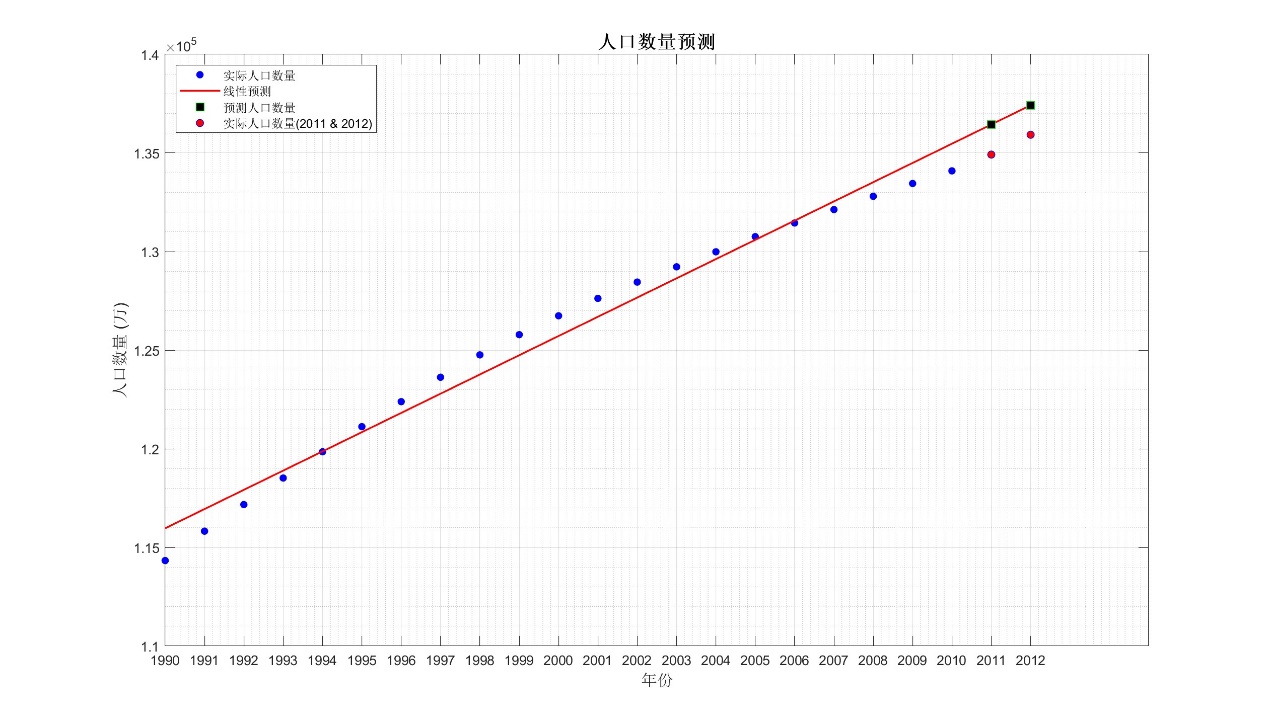


**3.3 年均增长量和增长率数据的绘图可视化**

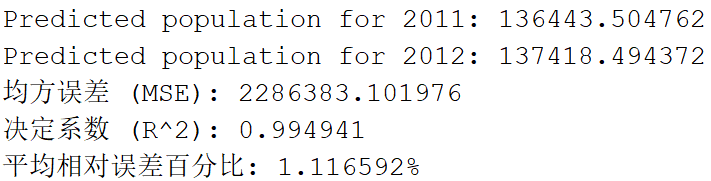


上图中，左侧的y轴以蓝色线条和圆点表示增长量，右侧的y轴以红色线条和圆点表示增长率。观察图像发现，1991-2000年我国处于人口较高增长阶段，随着计划生育政策进一步的全面阐述，高出生率稳步下降；从2000年之后，曲线逐渐趋缓，人口平稳增长。

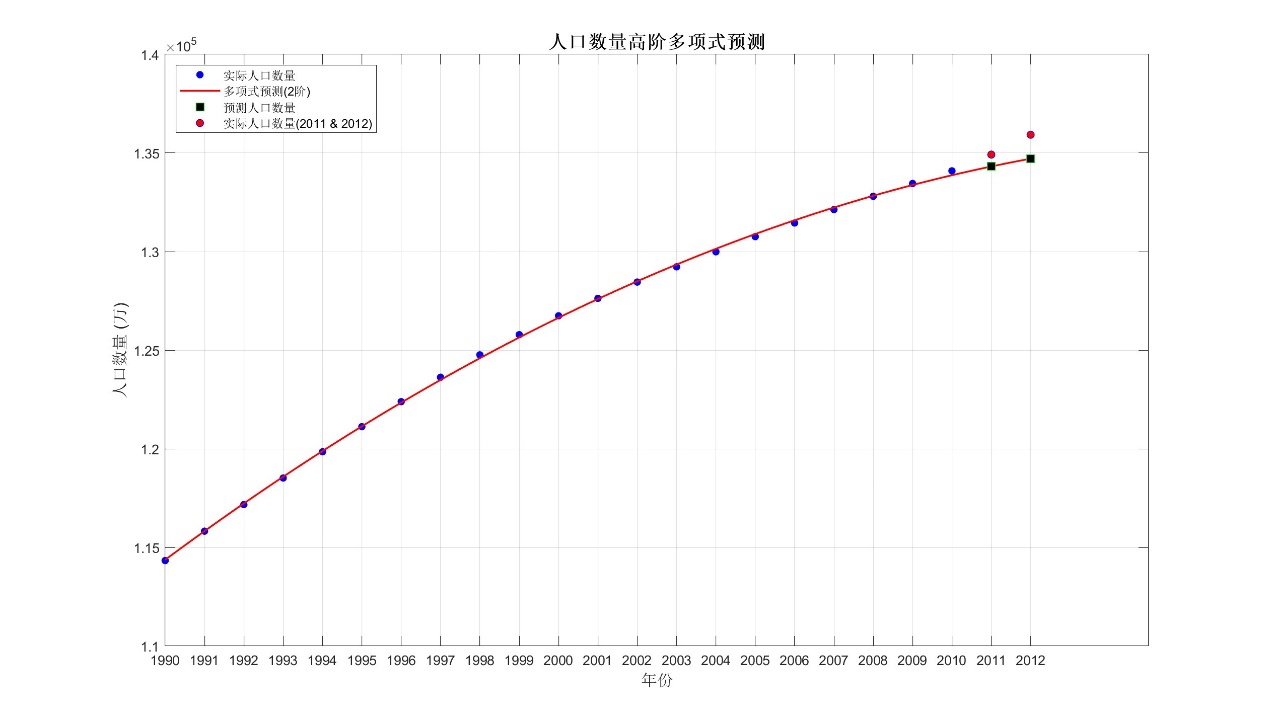
**3.4 人口数量预测模型及其评估：线性回归**

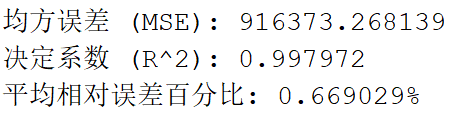


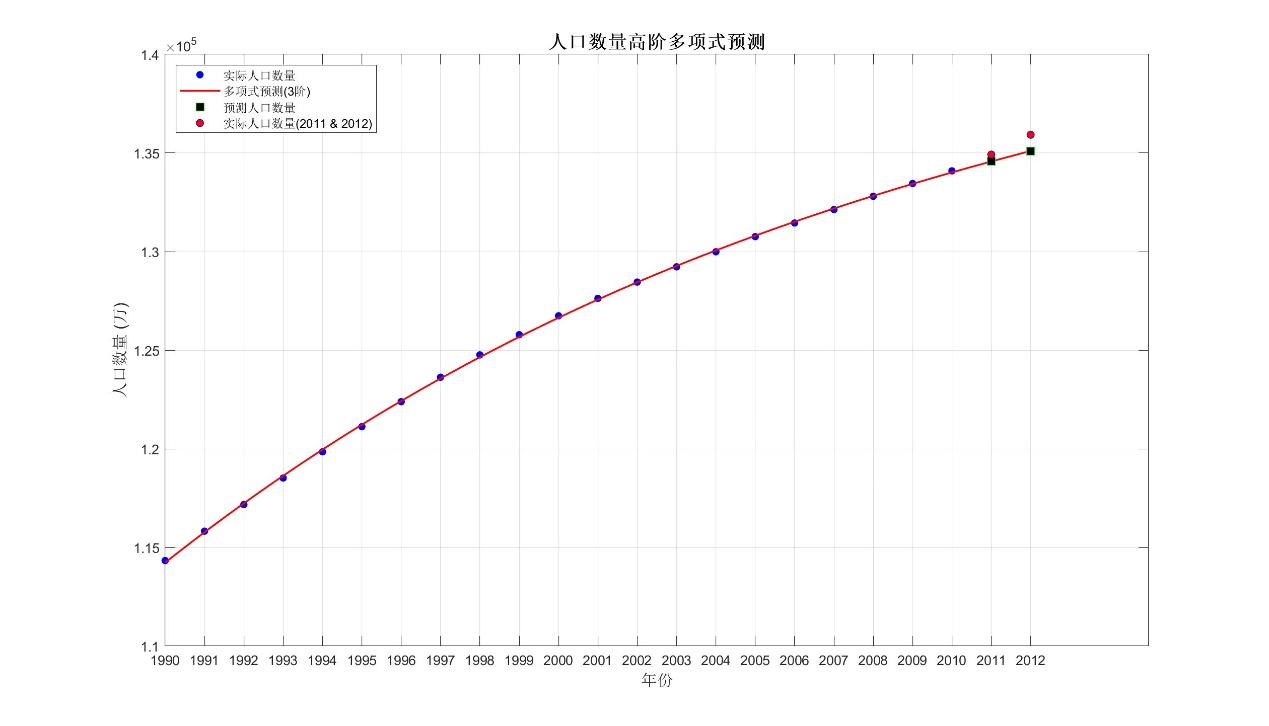
由于人口随时间的增长看起来是相对均匀的，没有大的波动，故线性关系可能是一个合适的模型，但观察图像发现，部分数据点与线性预测直线还是有一定差距，对于2011和2012年的预测准确度不符合预期，从下方的定量指标也表明，虽然决定系数(R2)与1较为接近，模型较好的捕获了数据的变化趋势，但平均相对误差百分比(ASE%)仍然不如设想中的小，故接下来尝试增大多项式拟合阶数，继续实验。

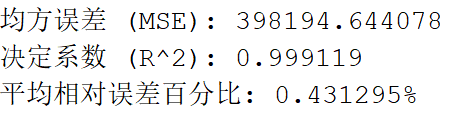
（线性回归）

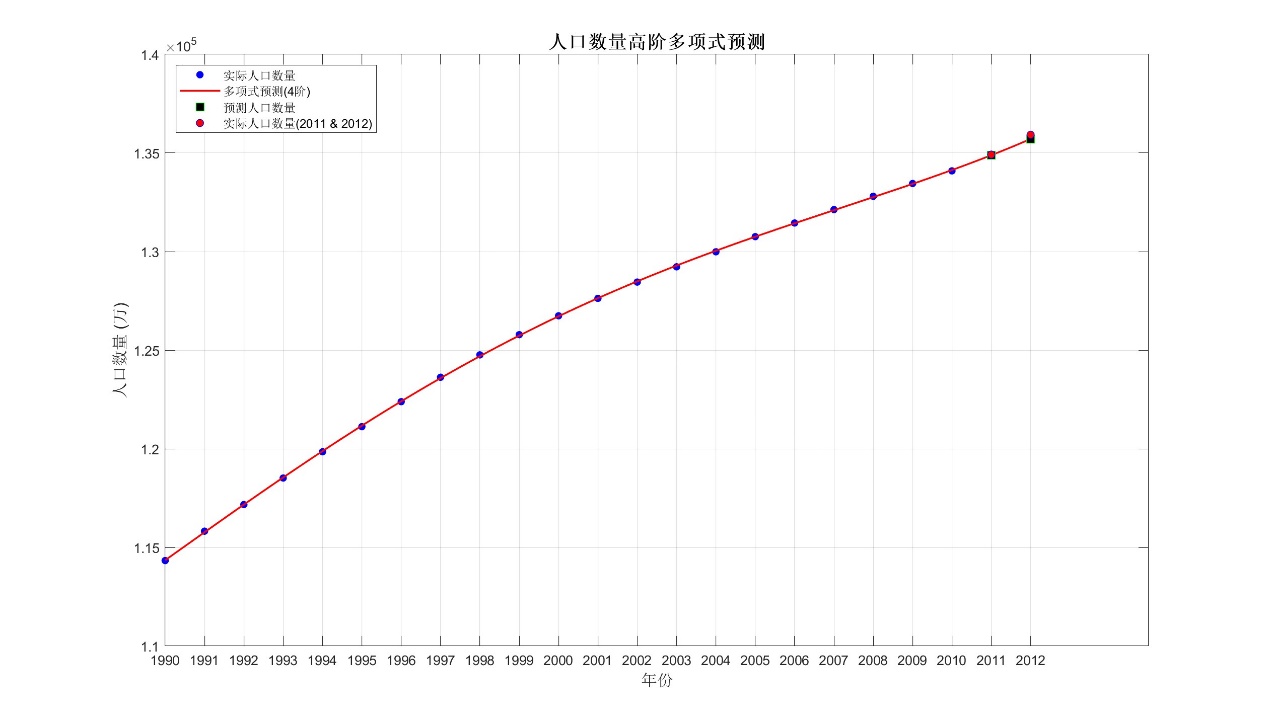
**3.5 人口数量预测模型及其评估：高阶多项式拟合**

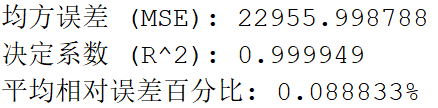


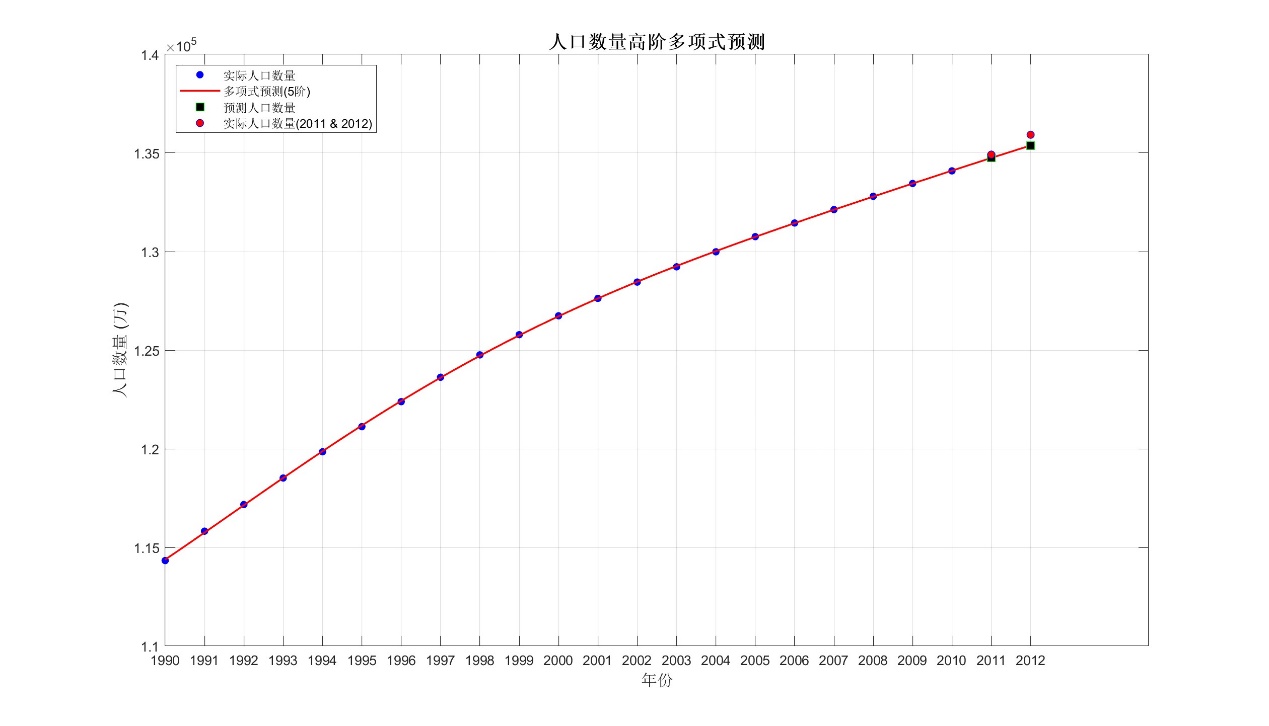
（2阶多项式拟合）

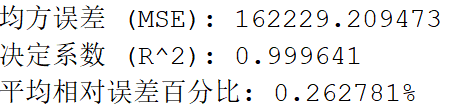


（3阶多项式拟合）



（4阶多项式拟合）



（5阶多项式拟合）

将上述5次结果（1阶~5阶）的模型评估**指标结果汇总**为如下表格：

degrees = 1:5; % 多项式的阶数

MSE\_values = [2286383.101976, 916373.268139, 398194.644078, 22955.998788, 162229.209473];

R2\_values = [0.994941, 0.997972, 0.999119, 0.999949, 0.999641];

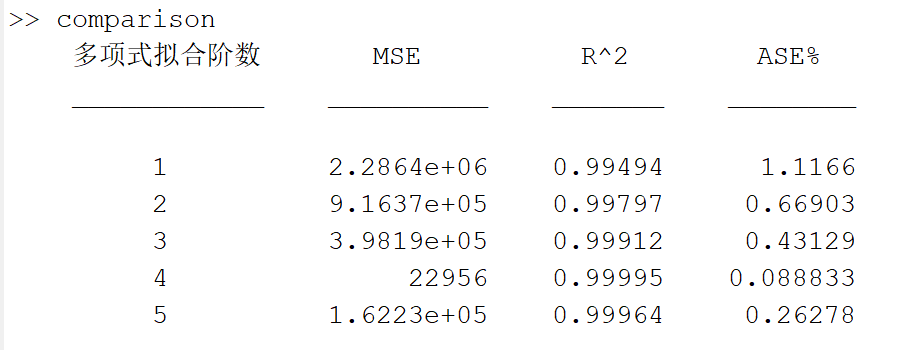
ASE\_values = [1.116592, 0.669029, 0.431295, 0.088833, 0.262781];

% 创建表格

T = table(degrees', MSE\_values', R2\_values', ASE\_values', ...

'VariableNames', {'多项式拟合阶数', 'MSE', 'R^2', 'ASE%'});

disp(T);



以上数据表明：

① 对于一阶模型（线性）：

* MSE相对较高，表明预测值和实际值之间的平均误差较大。
* R2值很高，意味着模型能够解释大部分的数据趋势，但这可能不足以反映模型的预测准确性。
* ASE%相对较高，这也表明模型在实际预测中的平均误差较大。

② 对于二阶模型：

* MSE显著降低，表明预测精度有所提高。
* R2值略有提高，说明模型对数据变异的解释能力增强。
* ASE%也有所下降，意味着平均预测误差减少。

③ 对于三阶模型：

* MSE进一步下降，预测精度进一步提高。
* R2值更接近1，说明模型拟合得非常好。
* ASE%进一步降低，表明误差减小。

④ 对于**四阶模型**：

* 在实验中MSE最低，显示出最佳的预测精度。
* R2值非常接近1，几乎完美地解释了数据的趋势。
* ASE%非常低，预测准确性最高。

⑤ 对于五阶模型：

* MSE**上升**，相对于四阶模型，这可能表示模型开始出现**过拟合**。
* R2仍然非常高，但没有四阶模型高。
* ASE%上升，虽然仍然较低，但是比四阶模型的预测误差大。

**3.6 模型评估总结**

① 随着多项式阶数的增加，模型复杂度也随之增加。在二阶到四阶模型中，性能明显提升，表明添加额外的非线性项有助于捕捉数据的更多特征。

② 在实验中，**四阶模型**提供了最优的MSE和ASE%，且R2接近完美，可以认为它是本次实验模型中最合适的人口预测模型选择。

③ 在具体实践中，最佳模型的选择不仅要考虑拟合统计量，还要考虑模型的解释性、复杂度以及在未见数据上的泛化能力。鉴于本次实验中只预测了2011和2012两年的数据，对于2013~2023年间的人口数量数据并没有考虑，且未考虑社会环境等人文因素的影响，故虽然四阶模型在本次的统计指标上表现最佳，但仍需要进一步的数据验证。

* + - 1. 人口政策

坚持综合决策：切实将人口融入经济社会政策，在经济社会发展战略规划计划、经济结构战略性调整、投资项目和生产力布局、城乡区域关系协调、可持续发展等重大决策中，充分考虑人口因素，不断健全人口与发展综合决策机制。

突出以人为本：坚持以人民为中心的发展思想，优先投资于人的全面发展，建立健全面向全人群、覆盖全生命周期的人口政策体系，促进共同参与、共享发展，增强人民群众获得感和幸福感。

强化正向调节：尊重人口规律，顺应经济社会发展要求和群众根本利益，完善服务保障政策，将生育水平调控并维持在适度区间，推动人口结构优化调整、人口素质不断提升、人口流动更加有序，持续增强人口资源禀赋。

加强风险防范：加强超前谋划和战略预判，重视把握人口各要素之间，以及人口与经济社会、资源环境等外部要素间的相互关系，提早防范和综合应对潜在的人口系统内安全问题和系统间的安全挑战，切实保障人口安全。

深化改革创新。积极转变人口调控理念和方法，统筹推进生育政策、计划生育服务管理制度、家庭发展支持体系和治理机制综合改革，完善人口预测预报预警机制，健全重大决策人口影响评估制度[1]。

1. **参考文献**
2. 国务院关于印发国家人口发展规划(2016—2030年)的通知[J].中华人民共和国国务院公报,2017(06):24-35.