

# 实验二 一元非线性回归模型及多元线性回归模型的建立及其显著性检验

## 实验报告

日期：2025 年 10 月 10 日 班级：地信班 姓名：许愿 学号：109090000000

实验项目	应用 SPSS 拟合回归模型、检验及预测	成绩	
实验目的	巩固一元非线性、多元线性回归分析的基本原理及方法步骤；用 SPSS 软件拟合一元非线性回归模型及多元线性回归模型并对模型进行显著性检验，同时进行预测。		
主要仪器及软件	硬件：计算机 软件：SPSS		

实验内容	<p>1. 一元非线性回归分析：中国人口增长趋势建模与预测</p> <p>(1) 研究问题与数据准备</p> <p>人口是地理学研究的核心要素之一，其数量、结构和分布的动态演变深刻影响着区域社会经济发展、资源环境承载力以及国家战略规划。与许多匀速变化的物理过程不同，人口系统的演变是一个典型的非线性过程。中国的现代人口发展历程尤为典型，清晰地展现了非线性特征。从建国初期的快速增长，到 20 世纪 70 年代末开始实施严格的计划生育政策，导致总和生育率急剧下降，人口增速明显放缓。进入 21 世纪，随着老龄化趋势的加剧和政策的逐步调整，人口增长进一步减速，并最终在 2021 年达到峰值后，于 2022 年首次出现负增长。这一从增长到平稳再到下降的完整“S”型曲线的后半段特征，用简单的线性模型是无法捕捉的。因此，本研究旨在采用能够描述曲线趋势的非线性模型（如多项式回归）来探究中国人口随时间变化的规律。</p> <p>本研究采用的数据为 1990 年至 2023 年中华人民共和国年末总人口。数据来源于中华人民共和国国家统计局发布的年度统计公报及官方数据库。为便于在 SPSS 中进行回归分析，将自变量“年份”进行编码处理，记为时间变量 <math>t</math>，其中 1990 年对应 <math>t=1</math>，1991 年对应 <math>t=2</math>，以此类推，2023 年对应 <math>t=34</math>。因变量为“年末总人口”，单位为“万人”。整理后的数据如下表所示。</p>					
	年份	时间编码(t)	年末总人口(万人)	年份	时间编码(t)	年末总人口(万人)
	1990	1	114333	2007	18	132129
	1991	2	115823	2008	19	132802
	1992	3	117171	2009	20	133450
	1993	4	118517	2010	21	134091
	1994	5	119850	2011	22	134735
	1995	6	121121	2012	23	135404
	1996	7	122389	2013	24	136072
	1997	8	123626	2014	25	136782
	1998	9	124761	2015	26	137462
	1999	10	125786	2016	27	138271
	2000	11	126743	2017	28	139008
	2001	12	127627	2018	29	139538

	2002	13	128453	2019	30	140005
	2003	14	129227	2020	31	141212
	2004	15	129988	2021	32	141260
	2005	16	130756	2022	33	141175
	2006	17	131448	2023	34	140967

表 1：1990–2023 年中国年末总人口数据

## (2) 数据输入与可视化

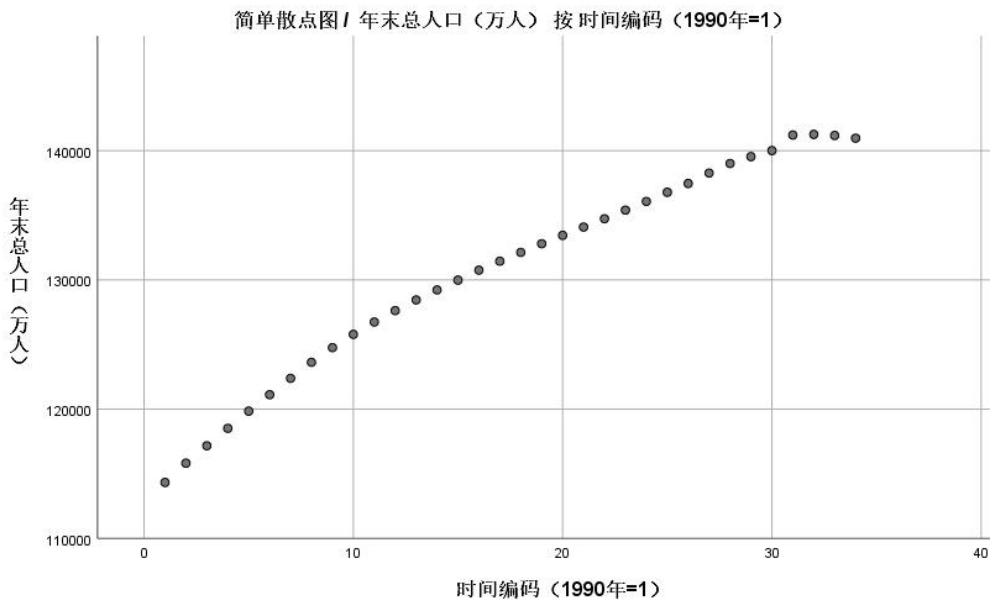
启动 SPSS，在“变量视图”中定义变量 t（时间编码）和 Population（年末总人口）。在“数据视图”中输入表 1 的 34 行数据。

	名称	类型	宽度	小数位数	标签	值	缺失	列	对齐	测量	角色
1	t	数字	8	0	时间编码（1990年=1）	无	无	8	右	未知	输入
2	Population	数字	8	0	年末总人口（万人）	无	无	8	右	未知	输入
3											

	t	Population	变量	变量	变量	变量	变量	变量
1	1	114333						
2	2	115823						
3	3	117171						
4	4	118517						
5	5	119850						
6	6	121121						
7	7	122389						
8	8	123626						
9	9	124761						
10	10	125786						
11	11	126743						
12	12	127627						
13	13	128453						
14	14	129227						
15	15	129988						
16	16	130756						
17	17	131448						
18	18	132129						
19	19	132802						
20	20	133450						
21	21	134091						
22	22	134735						
23	23	135404						
24	24	136072						
25	25	136782						
26	26	137462						
27	27	138271						
28	28	139008						
29	29	139538						
30	30	140005						
31	31	141212						
32	32	141260						
33	33	141175						
34	34	140967						
35								
36								
37								
38								

通过【图形】 – 【图表构建器】绘制 t 为 X 轴、Population 为 Y 轴的简单散点图。



### (3) 曲线估计与模型拟合

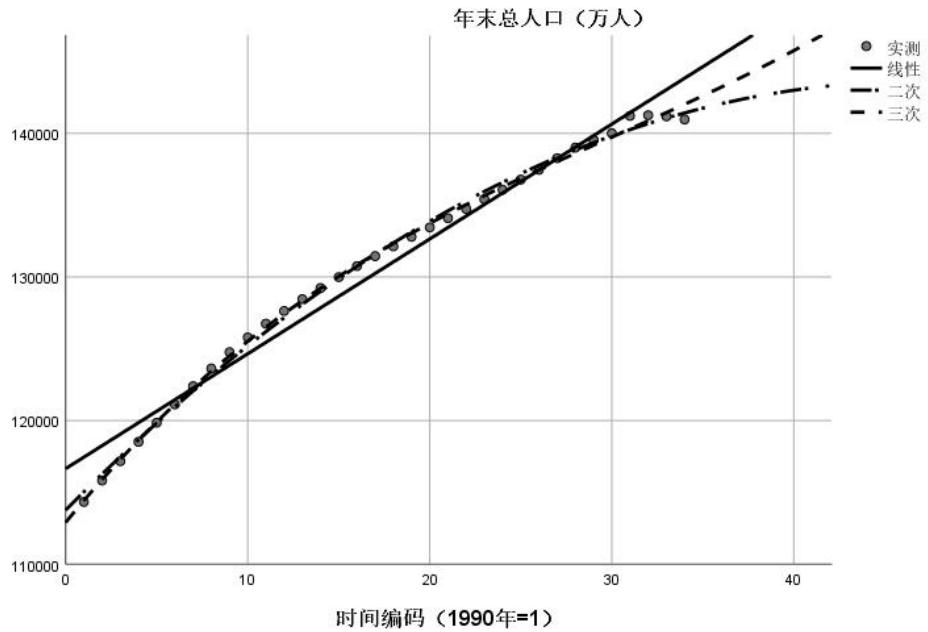
在主界面的菜单栏中选择【分析】 - 【回归】 - 【曲线估算】，将 Population 选入“因变量”，t 选入“自变量”；在“模型”区域，同时勾选线性、二次、三次。



点击“确定”后，SPSS 输出的信息如下：

模型摘要和参数估算值									
因变量: 年末总人口 (万人)									
方程	模型摘要				显著性	常量	参数估算值		
	R 方	F	自由度 1	自由度 2			b1	b2	b3
线性	.975	1253.203	1	32	.000	116653.759	799.587		
二次	.997	5590.383	2	31	.000	113767.018	1280.710	-13.746	
三次	.998	6263.398	3	30	.000	112908.742	1555.278	-33.077	.368

自变量为 时间编码 (1990年=1)。



- I. 线性模型调整后 R 方为 0.975；二次模型调整后 R 方为 0.997；三次模型调整后 R 方为 0.998。基于调整后 R 方最大化原则，选择三次模型为最优模型。
- II. 对于线性、二次和三次模型，其显著性值均为 0.000，远小于 0.05，表明三个模型在统计上都是高度显著的。
- III. 参数估算值表提供了构建回归方程所需的系数。对于最优的三次模型：

$$\text{常量} = 112908.742, \quad t = 1555.278, \quad t^2 = -33.077, \quad t^3 = 0.368$$

#### (4) 模型建立与预测

根据系数表的输出结果，建立三次多项式回归方程：

$$Population = 112908.742 + 1555.278t - 33.077t^2 + 0.368t^3$$

利用该方程预测 2025 年人口 ( $t = 36$ )：

$$\begin{aligned}
Population &= 112908.742 + 1555.278 \cdot 36 - 33.077 \cdot 36^2 + 0.368 \cdot 36^3 \\
&= 112908.742 + 55990.008 - 42867.792 + 17169.408 \\
&= 143200.366(\text{万人})
\end{aligned}$$

利用该方程预测 2030 年人口 ( $t = 41$ ):

$$\begin{aligned}
Population &= 112908.742 + 1555.278 \cdot 41 - 33.077 \cdot 41^2 + 0.368 \cdot 41^3 \\
&= 112908.742 + 63766.398 - 55602.437 + 25362.928 \\
&= 146435.631(\text{万人})
\end{aligned}$$

## 2. 多元线性回归分析：中国省级经济发展影响因素探究

### 1. 研究问题与数据准备

区域经济发展水平（以人均 GDP 为因变量 Y）受到资本、劳动力、技术、产业结构等多种因素驱动。本研究基于理论和数据可得性，选取以下自变量进行探究：

- $x_1$  城镇化率(%)：衡量人口与产业集聚程度，是现代经济增长的关键动力。
- $x_2$  人均固定资产投资(元/人)：反映资本形成与投入水平。
- $x_3$  第三产业增加值占比(%)：反映产业结构高级化程度。

本研究数据以 2022 年为截面，覆盖中国大陆 31 个省级行政区。数据主要来源于《中国统计年鉴 2023》及各省份 2022 年国民经济和社会发展统计公报。

地区	人均 GDP( $y$ )/元	城镇化率( $x_1$ )/%	人均固投( $x_2$ )/元	第三产业占比( $x_3$ )/%
北京	190126	87.6	45785	83.8
天津	118801	85.1	46521	60.9
河北	56888	61.7	45312	49.3
山西	73675	64.5	29013	54.1
内蒙古	96496	69.5	40040	50.8
辽宁	68509	73	15876	56.4
吉林	55024	64	22521	55.4
黑龙	50883	66	17929	57

	江			
上海	179370	89.3	40662	74.1
江苏	144468	74.4	59714	52.8
浙江	118830	73.4	59374	55.8
安徽	73687	60.2	59453	47.9
福建	126845	70.3	50821	46.1
江西	71001	62.3	59578	48.7
山东	85974	64.6	36006	52.1
河南	62071	57.1	46506	47.3
湖北	92170	64.7	56213	54.7
湖南	73498	61.4	50069	52.5
广东	101796	74.8	32103	54.7
广西	52215	55.7	51101	51.5
海南	66815	61.5	40847	61.6
重庆	90676	70.9	45544	52.7
四川	67785	58.4	43635	52.1
贵州	52348	56.1	41258	50.8
云南	61736	52.1	35836	55.6
西藏	58269	37.6	60515	55.4
陕西	82881	64.6	43075	46.4
甘肃	44986	53.7	28246	55.5
青海	60776	62	44268	46
宁夏	69925	67.2	41269	46.9
新疆	68525	57.8	42055	50.8

表 2: 2022 年中国 31 个省级行政区主要社会经济指标

## 2. 数据输入

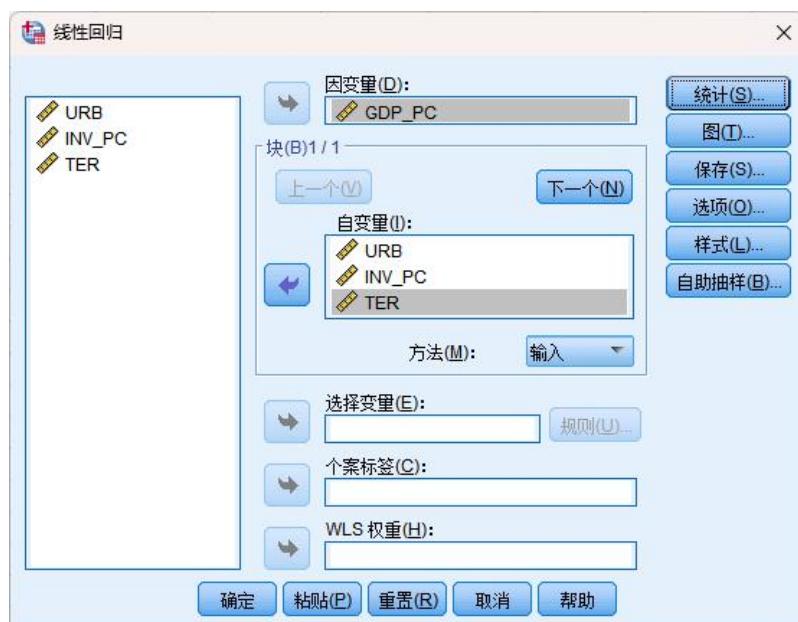
打开 SPSS，在新数据文件中定义  $GDP\_PC$  ( $y$ )、 $URB$  ( $x_1$ )、 $INV\_PC$  ( $x_2$ )、 $TER$  ( $x_3$ ) 四个变量，分别表示人均 GDP、城镇化率、人均固投、第三产业占比，再根据数据精度设置不同的小数位数。完成后，将表 2 的 31 行数据

输入“数据视图”。如图所示。

	GDP_PC	URB	INV_PC	TER	变量	变量
1	190126	87.6	45785	83.8		
2	118801	85.1	46521	60.9		
3	56888	61.7	45312	49.3		
4	73675	64.5	29013	54.1		
5	96496	69.5	40040	50.8		
6	68509	73.0	15876	56.4		
7	55024	64.0	22521	55.4		
8	50883	66.0	17929	57.0		
9	179370	89.3	40662	74.1		
10	144468	74.4	59714	52.8		
11	118830	73.4	59374	55.8		
12	73687	60.2	59453	47.9		
13	126845	70.3	50821	46.1		
14	71001	62.3	59578	48.7		
15	85974	64.6	36006	52.1		
16	62071	57.1	46506	47.3		
17	92170	64.7	56213	54.7		
18	73498	61.4	50069	52.5		
19	101796	74.8	32103	54.7		
20	52215	55.7	51101	51.5		
21	66815	61.5	40847	61.6		
22	90676	70.9	45544	52.7		
23	67785	58.4	43635	52.1		
24	52348	56.1	41258	50.8		
25	61736	52.1	35836	55.6		
26	58269	37.6	60515	55.4		
27	82881	64.6	43075	46.4		
28	44986	53.7	28246	55.5		
29	60776	62.0	44268	46.0		
30	69925	67.2	41269	46.9		
31	68525	57.8	42055	50.8		
32						

### 3. 多元线性回归分析

在菜单栏中选择【分析】 - 【回归】 - 【线性】，将 GDP\_PC 选入“因变量”，将 URB, INV\_PC, TER 选入“自变量”。在“统计”子对话框中，勾选“共线性诊断”。运行分析，得到初始模型输出。





在 SPSS 的菜单栏中选择【分析】-【相关】-【双变量】，对 GDP\_PC、URB、INV\_PC、TER 进行皮尔逊相关分析。

相关性

		GDP_PC	URB	INV_PC	TER
GDP_PC	皮尔逊相关性	1	.818**	.272	.646**
	Sig. (双尾)		.000	.138	.000
	个案数	31	31	31	31
URB	皮尔逊相关性	.818**	1	-.098	.556**
	Sig. (双尾)	.000		.600	.001
	个案数	31	31	31	31
INV_PC	皮尔逊相关性	.272	-.098	1	-.147
	Sig. (双尾)	.138	.600		.430
	个案数	31	31	31	31
TER	皮尔逊相关性	.646**	.556**	-.147	1
	Sig. (双尾)	.000	.001	.430	
	个案数	31	31	31	31

\*\*. 在 0.01 级别 (双尾)，相关性显著。

- I. GDP\_PC 与 URB 的相关系数为 0.818。
- II. GDP\_PC 与 TER 的相关系数为 0.646。
- III. GDP\_PC 与 INV\_PC 的相关系数为 0.272。
- IV. URB 与 TER 的相关系数为 0.556。
- V. URB 与 INV\_PC 的相关系数为 -0.098。
- VI. TER 与 INV\_PC 的相关系数为 -0.147。

根据最终模型的系数表输出结果，建立回归方程：

- 非标准化回归方程：

$$\hat{Y} = -197551.746 + 2293.638 \cdot X_1 + 1.173 \cdot X_2 + 1514.327 \cdot X_3$$

- 标准化回归方程：

$$Z_Y = 0.673 \cdot Z_{X1} + 0.387 \cdot Z_{X2} + 0.329 \cdot Z_{X3}$$

运行结果与分析	<p>1. 一元非线性回归分析总结</p> <p>通过对 1990–2023 年中国人口数据的曲线估计分析，比较了线性、二次和三次模型的拟合效果。结果显示，三次多项式回归模型的拟合优度最高（调整后 <math>R^2 = 0.998</math>），能够最精确地描述中国人口增长由加速到放缓，再到平台期乃至出现拐点的复杂非线性过程。</p> <p>最终确定的最优回归方程为：</p> $\text{Population} = 112908.742 + 1555.278t - 33.077t^2 + 0.368t^3$ <p>基于此模型，预测 2025 年中国年末总人口约为 14.320 亿人，2030 年约为 14.644 亿人。</p> <p>2. 多元线性回归分析总结</p> <p>本研究通过构建多元线性回归模型，探究了影响 2022 年中国省级人均 GDP 差异的驱动因素。初始模型包含了城镇化率、人均固定资产投资和第三产业占比三个自变量。最终发现，三个因素均是解释中国省级经济发展水平的三个核心且统计上高度显著的因素。模型的调整后 <math>R^2</math> 为 0.854，说明这三个变量共同解释了省级人均 GDP 差异的 85.4%。URB、INV_PC、TER 的 VIF 值分别为 1.447、1.023、1.465，这表明模型不存在严重的多重共线性问题。</p> <p>最终确定的回归方程为：</p> <ul style="list-style-type: none"> <li>- 非标准化回归方程：</li> </ul> $\hat{Y} = -197551.746 + 2293.638 \cdot X_1 + 1.173 \cdot X_2 + 1514.327 \cdot X_3$ <ul style="list-style-type: none"> <li>- 标准化回归方程：</li> </ul> $Z_Y = 0.673 \cdot Z_{X1} + 0.387 \cdot Z_{X2} + 0.329 \cdot Z_{X3}$
---------	---