

第三届“泰迪杯”

全国大学生数据挖掘竞赛

优秀作品

作品名称：基于数据挖掘技术的市财政收入分析预测模型

荣获奖项：二等奖

作品单位：甘肃农业大学

作品成员：王久旺 吴俊红 许丽雯

指导教师：赵有益

基于数据挖掘技术的广州市财政收入分析预测模型

摘要：通过挖掘、整理 1999-2014 年与广州市财政收入关联的经济、社会等指标统计数据，利用 SAS 和 SPSS 软件，分析了广州市财政收入的构成及其比例，建立了因素分析的回归模型，分析了在财政收入结构中，**企业所得税、印花税、罚没收入、消费税和增值税**是影响财政收入的主要方面，进一步分析识别出了影响财政总收入的关键因素是**企业所得税、增值税、第三产业、交通业**，其中，影响增值税的关键因素是**商品进口总值、工业增加值**，影响营业税的关键因素是**第三产业增加值、建筑业总产值、住宿和餐饮业零售额**，影响个人所得税的关键因素是**第二产业增加值**，影响企业所得税的关键因素是**商品进口总值、工业增加值**。

在因素分析基础上，构建了灰色预测和回归预测模型，预测结果误差 5%左右，其中广州市 2015 年财政总收入、增值税、营业税、企业所得税、个人所得税的预测区间分别是（29031430.82,31087829.18）、（2890442.273,2986829.727）、（1934437.078,2069874.922）、（1492866.117,1527933.883）、（576012.5216,592216.6784）。

通过对广州市近几年社会经济发展的人口、就业、环境、医疗卫生和社会福利、财政收入及支出等数据的挖掘，再用 R 软件等对影响财政收入的经济因素和非经济因素两方面进行了分析，明确了财政收入支出与经济、社会的相互关系，向广州市财政局提出财政收入和支出预算的平衡性、行业或部门财政支出分配的弹性、政策制定与监管落实的严肃性等方面的建议，为下一年的财政预算支出政策提供指导依据。

关键词：广州市财政收入；广州市财政支出；逐步回归；灰色预测模型

The thesis title

Abstract: Through mining and sorting the statistical data about social and economy associated with fiscal revenue of Guangzhou from 1994 to 2014. Using SAS and SPSS software, analyzed the composition and proportion of fiscal revenue in Guangzhou City, we set up a regression model of factor analysis. Draw a conclusion that the corporate income tax, stamp tax, incomes from fines and confiscations, the consumption tax and value-added tax are the major aspects on affecting the fiscal revenue.

Further analyzing and identifying the key factors affecting the total fiscal revenue are the enterprise income tax, value-added tax, the third industry, transportation industry. Among them, the key factors affecting the value-added tax are the total value of imports and industrial added quantity.

The key factors affecting the business tax are the third industrial added value, total output value of construction, accommodation and catering industry retail sales. The key factor affecting the personal income tax is the added value of the second industry. The key factors affecting the corporate income tax are the total value of imports, industrial added quantity.

On the basis of factor analysis, we constructed the grey forecasting model and regression prediction model, the prediction result error is about 5%.

Among them, the prediction intervals of Guangzhou 2015 total fiscal revenue, value added tax, business tax, enterprise income tax, personal income tax are (29031430.82, 31087829.18)、(2890442.273, 2986829.727)、(1934437.078, 2069874.922)、(1492866.117, 1527933.883)、(576012.5216, 592216.6784) .

By the population of Guangzhou city social and economic development in recent years, employment, environment, health care and social welfare, fiscal revenue and expenditure data of mining. Then using R software analyzed the two aspects of both economic and non-economic factors affecting the financial income. We cleared the relationship between the fiscal revenue expenditure and the economic and social.

Putting forward the advice to the Guangzhou Municipal Finance Bureau about the balance fiscal revenue and expenditure budget, industry or sector spending allocation of elasticity, policy and regulation to carry out the seriousness. It provides the basis for the next year's fiscal expenditure policy.

Key words: local fiscal revenue stepwise regression grey forecasting model

目 录

1. 研究目标.....	1
2. 分析方法与过程.....	4
2.1. 总体流程	4
2.2. 具体步骤	3
2.3. 结果分析	4
3. 结论.....	19
4. 参考文献.....	19

“泰迪杯”优秀作品

1. 挖掘目标

本次建模的目标是通过收集、整理 1999-2014 年的与广州市财政收入关联的经济、社会等指标的统计数据，分析广州市财政收入的来源、构成及其比例，建立因素分析回归模型，分析相关经济数据对广州市财政收入及其构成要素的相互关系，以此确定影响财政收入的关键因素；在因素分析基础上，利用数据挖掘技术，构建灰色预测模型和回归预测模型，对广州市 2015 年的财政总收入及各个类别收入进行科学合理的预测；同时，为尽量避免和克服年度预算收支规模确定的随意性和盲目性，结合广州市近几年社会经济发展的财政收入及支出等情况，正确处理财政与经济的相互关系，向广州市财政局提出几点财政收入和支出预算的建议，为下一年的财政预算支出政策提供指导依据。

2. 分析方法与过程

2.1. 总体流程

为了对问题有序有效分析研究，将建模过程和建模方法进行整体设计，具体见图 1 和图 2

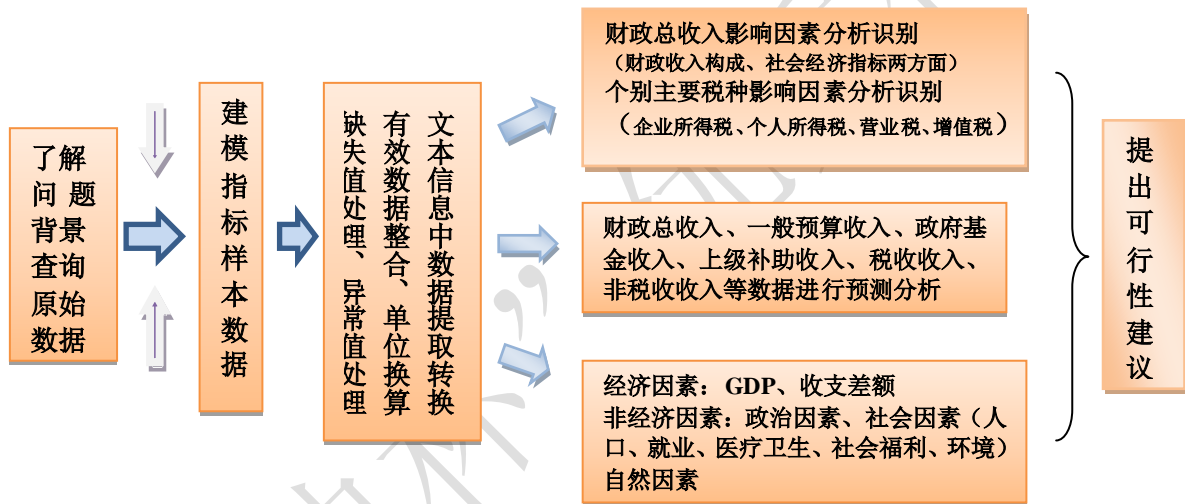


图 2.1.1 分析流程图

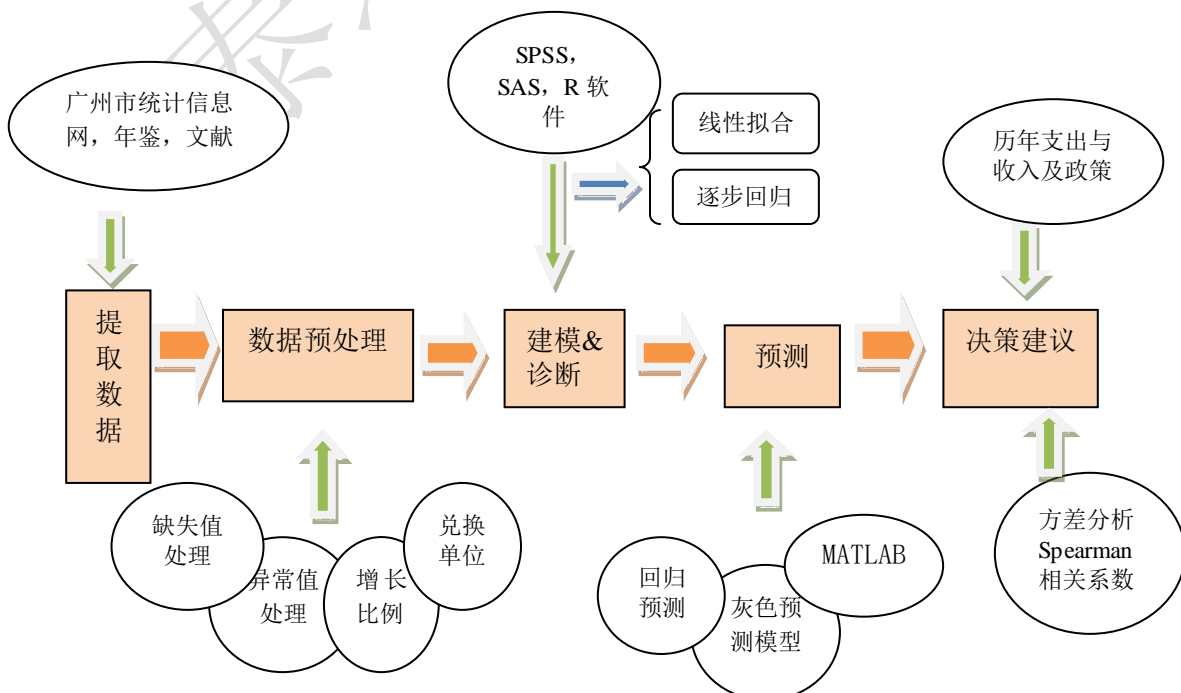


图 2.1.2 方法流程图

主要包括以下几个步骤：

步骤 1：通过广州市统计信息网、文献数据库、图书馆文献资料等多渠道了解问题背景，收集梳理相关经济社会统计统计数据，从“广州市政府预算支出”、“广州市国民经济和社会发展统计公报”等文本中提取、整理信息与数据。对个别缺失、异常数据进行预处理。对部分统计数据的单位进行换算处理，对个别指标进行整合计算。

步骤 2：建立回归模型，分析、识别影响广州市财政总收入、税收（包括增值税、营业税、企业所得税、个人所得税等）、非税收等收入的关键影响因素。

步骤 3：基于财政收入的因素分析，构建灰色预测模型、回归预测模型，对广州市 2015 年地方财政总收入、一般预算收入、营业税、增值税等指标进行预测与分析。

步骤 4：结合广州是社会经济发展，进一步挖掘近几年财政收入、预算的数据信息，挖掘影响财政收入的经济与非经济因素的人口、环境、就业等指标数据，并从财政收入和支出预算的角度进行分析，向广州市财政局提出建议，为下一年广州市的财政预算支出政策提供指导依据。

2.2. 具体步骤

结合上述总体流程图 1 和图 2，对广州市地方财政收入的主要影响因素分析、收入预测和财政收支分析等内容，对每一问题解决进行具体说明与解释。

步骤 1：数据的挖掘、抽取与整理

通过广州市信息统计网、“广州市 1999-2013 年预算报表”、百度文库、中国知网、广州市政府信息平台、“广州市国民经济和社会发展统计公报”等方面多渠道多方面搜集、挖掘和梳理数据。主要数据整理如下：

1999-2014 年广州市地方财政总收入相关指标的数据（见附件 1），1999-2014 年广州市地方财政收入的税收与非税收的数据（见附件 2），1999-2014 年广州市地方财政收入的税收相关指标的数据（见附件 3），1999-2014 年广州市地方财政收入的非税收相关指标的数据（见附件 4），1999-2013 年地方财政收入与 GDP、三大产业（见附件 5），1999-2013 年地方总收入与各个产业指标（见附件 7），1999-2013 年增值税的相关指标（见附件 10），1999-2013 年营业税相关指标（见附件 12），1999-2013 年企业所得税（见附件 14），1999-2013 年个人所得税相关指标（见附件 16），2015-2016

年相关指标的预测值（见附件 19），2008–2014 年 GDP、地方财政收入、支出增长率的数值（见附件 20），2007–2014 年地方财政收入与支出的差额（见附件 21），2007–2013 年地方财政支出的相关指标（见附件 22），2008–2013 年人口与财政收入、支出增长率的数据（见附件 24），2007–2013 年就业人口与财政收入、支出的增长率（见附件 25），2007–2013 年医疗卫生机构与财政收入、支出增长率（见附件 26），2007–2013 年社会福利与地方财政收入、支出的数据（见附件 27）

步骤 2：对数据进行预处理

1、缺失值处理：用插值和平滑法对个别缺少的数据进行填补。如：1999 和 2000 年企业所得税中规模以上工业企业盈亏相抵后的利润总额、1999–2001 年企业所得税中限额以上连锁店零售额、1999–2003 年增值税中批发零售业增加值，等。

2、异常值处理：从折线图或增长率分析判断某些指标值异常，严重影响回归分析模型和预测模型的精确性和可靠性，因此用一阶差分法对这些值进行处理。如：企业所得税中 1999 年建筑业企业利润总额、2012 年非税收收入，等。

3、指标单位的统一与换算：将所有指标单位统一和换算为“万元”，如以“美元”为单位的指标，按照汇率换算成人民币，并将亿元转化成万元。如：增值税中商品进口总值单位为“美元”，按照每年美元与人民币的汇率换算为人民币，以“亿元”为单位的“城镇居民储蓄存款余额”，等；

4、指标的整合：某些量纲、单位不统一的指标进行整合。如，在个人所得税影响因素分析中，“城镇单位职工年平均工资”与“城镇非私营单位从业人员数”整合为“城镇职工年总收入”，等。

5、增长率计算：根据问题分析需求，计算出相关指标的增长率。如：财政收入与财政支出的差额与增长率，医疗卫生机构与财政收入与支出增长率，等。

步骤 3：财政收入关键影响因素分析

为了对财政收入的构成进行科学合理分析，将其相关指标进行分类整理，具体见图 2.1.3。

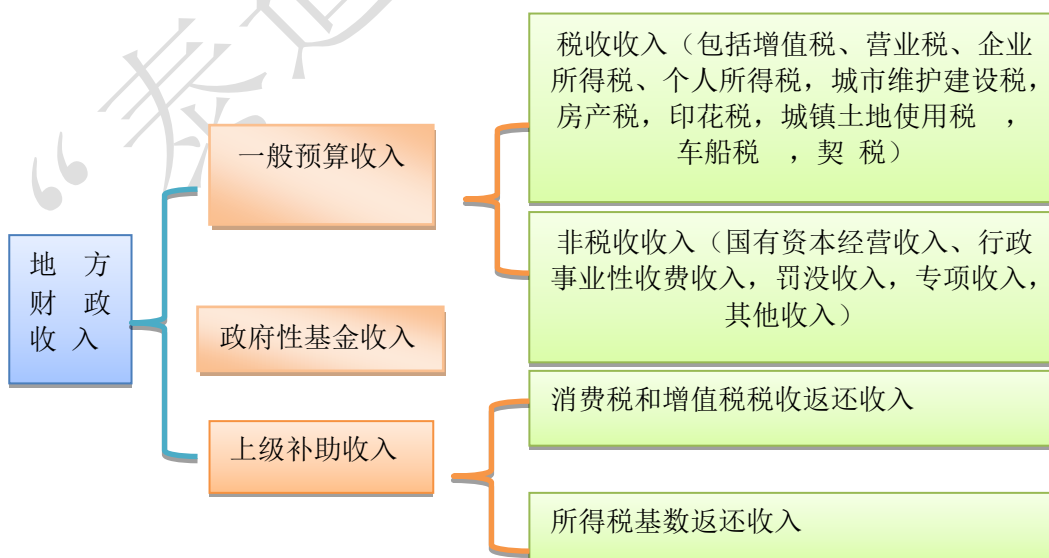


图 2.1.3 财政收入的指标构成

1、财政收入相关数据分析

用 Excel 对广州财政收入与一般预算收入、政府性基金收入和上级补助收入作折线图、一般预算中对税收和非税收作折线图、对税收收入中的指标作折线图、非税收收入中的组成指标作折线图、用 R 软件对 1999-2014 年广州市一般预算收入中税收收入和非税收收入的比例作饼图，并对其构成比例和增长率进行分析说明。

2、建立数学模型，识别关键影响因素

(1) **影响财政总收入关键因素的识别。**首先，由附件 1-4 的数据，以财政总收入为因变量，以增值税、营业税、企业所得税、个人所得税等所有财政收入构成指标为自变量，建立逐步回归模型，用 SAS 软件进行计算分析，拟合出最优的回归方程，以此说明财政收入构成的主要部分。

然后，由附件 5 和附件 7 的数据，以财政总收入为因变量，筛选对其影响较大的第一、二、三产业产值，工业、农业、商业、交通运输业、建筑业等产值 8 个指标作为自变量，分别建立逐步回归模型，用 SPSS 软件进行计算分析、检验，识别其关键影响因素。

(2) **影响增值税关键因素的识别。**由附件 10 的数据，以增值税为因变量，筛选对其影响较大的商品进口总值、地区生产总值、工业增加值、批发零售业零售额、工业增加值占 GDP、批发零售业增加值 6 个指标作为自变量，建立逐步回归模型，用 SAS 软件进行计算分析、检验，识别其关键影响因素。

(3) **影响营业税关键因素的识别。**由附件 12 的数据，以营业税为因变量，筛选对其影响较大的公路货运量、公路客运量、建筑业增加值、第三产业增加值、全社会房地产开发投资额等 9 个因素作为自变量，建立逐步回归模型，用 SAS 软件进行计算分析、检验，识别其关键影响因素。

(4) **影响企业所得税关键因素的识别。**由附件 14 的数据，以企业所得税为因变量，筛选对其影响较大的第二产业增加值、第三产业增加值、全社会固定资产投资额、城市商品零售价格指数（1978=100）、规模以上工业企业盈亏相抵后的利润总额、规模以上国有及国有控股工业企业企业亏损面、建筑业总产值等 9 个因素作为自变量，建立逐步回归模型，用 SAS 软件进行计算分析、检验，识别其关键影响因素。

(5) **影响个人所得税关键因素的识别。**由附件 16 的数据，以个人所得税为因变量，筛选对其影响较大的城市居民年人均可支配收入、城镇单位职工年平均总工资、城镇居民储蓄存款余额、地区生产总值、第二产业增加值 5 个因素作为自变量，建立逐步回归模型，用 SAS 软件进行计算分析、检验，识别其关键影响因素。

步骤 4：财政收入及各个类别收入的预测

在步骤 3 中问题解决的基础上，利用附件 1-4 的数据，根据 1999-2014 年相关指标数据，构建灰色预测模型和回归预测模型，对 2015 年广州市地方财政总收入，以及一般预算收入、营业税、增值税等各个类别进行预测，并进行误差分析。

然后，为降低预测误差，将灰色预测与回归预测值进行加权处理（权重分别为 0.5）。

最后，给出 2015 年广州市地方财政总收入，以及一般预算收入、营业税、增值税等各个指标的预测区间。

步骤 5：结合对社会经济发展和财政收入支出的分析，提出建议

1、经济因素对财政收入支出的分析

以 GDP 为主要经济因素，进行其与财政收入支出的分析，数据见见附件 20 和附件 21。

2、非经济因素对财政收入支出的分析

- (1) 对人口与财政收入、支出增长率的分析，数据见附件 24。
- (2) 对就业情况与财政收入、支出的增长率的分析，数据见附件 25。
- (3) 对医疗卫生机构与财政收入、支出增长率的分析，数据见附件 26。
- (4) 对社会福利与地方财政收入、支出的分析，数据见附件 27。
- (5) 其他因素对地方财政收入、支出的分析，如，政策倾向性、自然灾害，等

3、历年财政收、支情况分析

结合 1999-2014 年广州市地方财政收入与支出的数据，分析广州市地方财政收支均衡状况。结合 2015 年财政预算和收入的预测值，进行收支状况分析。

4、综合分析，提出建议

通过对广州市近几年社会经济发展的人口、就业、环境、医疗卫生和社会福利、财政收入及支出等数据的挖掘，对影响财政收入的经济因素和非经济因素两方面的分析，明确财政收入支出与经济、社会的相互关系，从广州市财政局提出财政收入和支出预算的平衡性、行业或部门财政支出分配的弹性、“营改增”等政策制定与监管落实的严肃性等方面提出科学合理化的建议。

2.3. 结果分析

1、用 Excel 和 R 用软件进行相关指标源数据分析

①用 Excel 对财政收入与一般预算收入、政府性基金收入和上级补助收入做柱状图分析

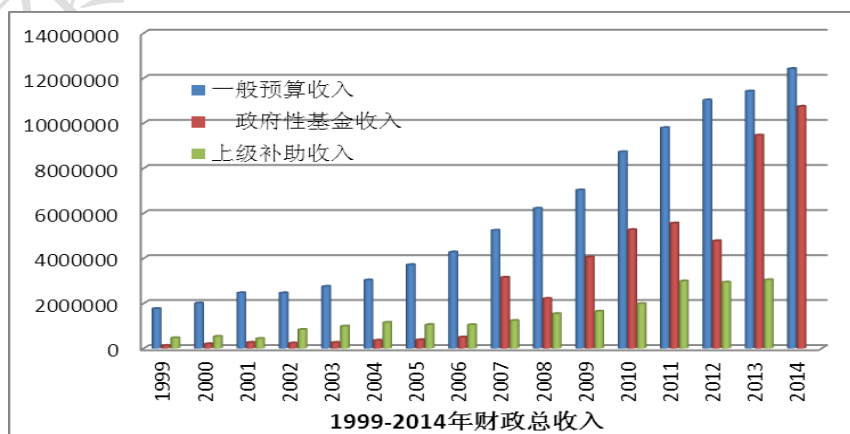


图 2.3.1 1999-2014 年财政收入构成

地方财政总收入是由一般预算收入、政府性基金预算、上级补助收入构成。从 1999-2014 年财政总收入条形图上可以看出一般预算收入的变化曲线明显高于政府性基金收入，高于上级补助收入曲线；一般预算曲线在 2003 年以前变化较平缓，2003 年以后增长趋势较大，一直到 2012 年变化趋势变得缓慢。这是由于从 2004 年广州市政府才开始组织财政收入。加强税收征管，加大协税护税力度，做到应收尽收；进一步加强非税收入管理^[1]。政府性基金收入增长趋势较平缓，直到 2007 年增长的幅度特别大，2008 年至 2011 年持续加速增长，2011 年到 2012 年出现通幅度下降，2013 年到 2014 年又开始快速增加，这是因为在 2011 年 7 月 5 日广州市政府为了适应社会主义市场经济发展要求规范广州市调节基金征集使用管理，印发《广州市价格调节基金管理规定》，使政府性基金收入的受到了限制上级补助收入是由中央政府根据国情返还，保持稳定增减。

②用 Excel 对一般预算收入中税收和非税收做折线图和比例饼图。

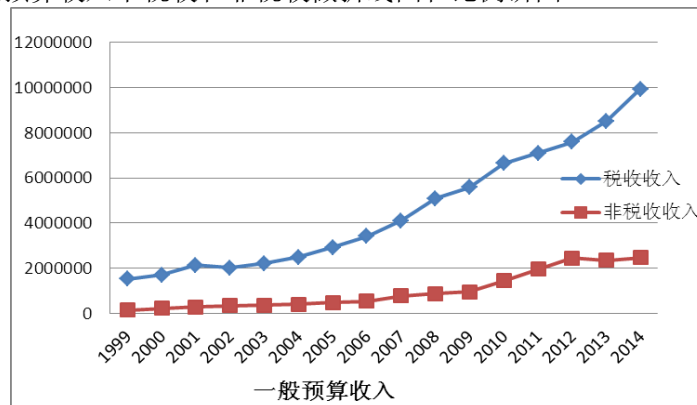


图 2.3.2 税收和非税收构成情况折线图



图 2.3.3 税收和非税收构成饼图

目前，我国财政收入一词有广义和狭义之分。狭义的财政收入一般仅包含税收和一般预算非税收入的预算内收入^[2]。在一般预算收入趋势图里，税收收入的增长趋势持续增高。非税收入的趋势变化较为平缓；税收收入的增长曲线远远高于非税收入，这就意味着税收收入对一般预算的影响远远大于非税收入的影响。图 2.3.3 是通过 R 软件做的 1999-2014 年税收与非税收的饼形图，从图中明显可以看出，税收收入都大于非税收入。结合图 2.3.2 与图 2.3.3 可以得出，税收收入对一般预算的影响远远大于非税收入的影响，及在税收收入与非税收入之间，税收收入为影响一般预算收入、地方财政收入的主要因素。

③对税收收入的相关指标做折线图

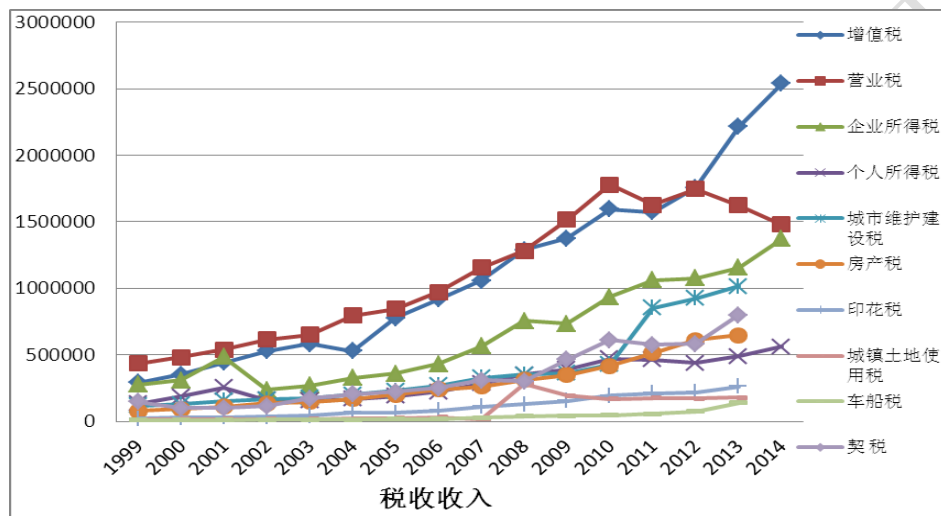


图 2.3.4 财政收入中各税种的变化情况

税收收入是由增值税、营业税、企业所得税、个人所得税等多个部门组成^[2]。从税收收入趋势图中可以大致看出大部分税种都呈现增长趋势，尤其明显的为增值税。其中，营业税在 2011 年以前持续上升，到 2011-2014 年呈现递减趋势，这主要是由于 2011 年经国务院批准，财政部、国家税务总局联合下发营业税改增值税试点方案。自 2012 年 8 月 1 日起至年底，国务院将扩大“营改增”试点至 10 省市，广东省包括在内。进入 2015 年下半年后，中国或将全面告别营业税^[3]。

2、影响财政收入的关键因素分析

(1) 影响财政总收入关键因素的分析与识别

从影响指标看，将其分为两类情况进行分析。

由附件5的数据，用spss软件对地方财政收入 y 和国内生产总值 x_1 、第一产业 x_2 、第二产业 x_3 、第三产业 x_4 ，进行逐步回归，将软件运行结果（见附件6）进行整理，得到表2.3.3.

表2.3.3 影响财政总收入关键因素的方差分析表

模型	平方和	df	均方	F 值	P 值
回归	5.160E14	1	5.160E14	539.920	.000a
残差	1.243E13	13	9.558E11		
总计	5.285E14	14			

模型	B	标准误差	标准系数	t 值	P 值
1 (常量)	-1145149.257	461884.823		-2.479	.002
x4	.212	.009	.988	23.236	.000

经过逐步回归和上表结果可以看出，影响地方财政收入的指标的系数都为正，属于正相关，与客观实际相符合。当在显著性水平为 $\alpha = 0.05$ ， F 值为 539.920， P 值远远小于 0.05，由此表明，回归方程拟合程度比较高，显著性明显。根据 T 检验得到参数 x_4 的 P 值远远小于 0.05，得到最优的回归方程为

$$\hat{y} = -1145149 + 0.212x_4 \tag{1}$$

从而得到，第三产业 x_4 对税收的影响最大。

同时，由附件 7 中指标数据，利用 SPSS 软件对地方财政收入 y 和农业 x_1 、工业 x_2 、商业 x_3 、交通业 x_4 、建筑业 x_5 ，作逐步回归，将软件运行结果（见附件 8）进行整理，得到表 2.3.4。

表 2.3.4 影响财政总收入关键因素的方差分析表

模型	平方和	df	均方	F 值	P 值
回归	5.197E14	1	5.197E14	770.167	.000a
残差	8.772E12	13	6.748E11		
总计	5.285E14	14			

模型	B	标准误差	标准系数	t 值	P 值
1 (常量)	-1549495.158	399405.357		-3.880	.002
x_4	.320	.012	.988	27.752	.000

经过逐步回归和上表结果可以看出，影响地方财政收入的指标的系数都为正，属于正相关，与客观实际相符合。在显著性水平为 $\alpha = 0.05$ ，方程的 F 值为 770167， P 值远远小于 0.05。得到最优的回归方程为

$$\hat{y} = -1549495.158 + 0.320x_4 \tag{2}$$

从回归方程可以看出，交通业 x_4 对税收的影响较大。

综合上述分析，可以得到，第三产业和交通业两个指标对地方财政收入的影响最大，认为它的最关键影响因素。

从财政收入的构成来看，将地方财政收入与所有构成要素进行分析

①用 SAS 软件对地方财政收入 y 和增值税 x_1 、营业税 x_2 、企业所得税 x_3 、个人所得税 x_4 、城市维护建设税 x_5 、房产税 x_6 、印花税 x_7 、城镇土地使用税 x_8 、车船税 x_9 、契税 x_{10} 、国有资本经营收入 x_{11} 、行政事业性收费收入 x_{12} 、罚没收入 x_{13} 、专项收入 x_{14} 、其他收入 x_{15} 、消费税和增值税 x_{16} 、上级补助收入 x_{17} 、税收返还收入 x_{18} 、所得税基数返还收入 x_{19} ，作逐步回归，将软件运行结果（见附件 9）进行整理，得到表 2.3.5。

表 2.3.5 影响财政总收入构成要素的方差分析表

模型	平方和	df	均方	F 值	Pr > F
回归	5.275842E14	4	1.318961E14	1493.10	<.0001
残差	8.833728E11	10	88337284366		
总计	5.284676E14	14			

模型	B	标准误差	II 型 SS	F 值	Pr > F
Intercept	-2021786	316564	3.603221E12	40.79	<.0001

x3	2.47683	1.12437	4.28662E11	4.85	0.0522
x7	37.13999	5.17531	4.549401E12	51.50	<.0001
x13	10.84101	3.38581	9.056487E11	10.25	0.0095
x16	0.83846	0.09728	6.562257E12	74.29	<.0001

从经济意义的角度，影响地方财政收入最大的四个指标的系数都为正相关，符合客观实际。

假设显著性水平为 $\alpha = 0.05$ ，方程的 F 值为 1493.10， P 值远远小于 0.05，由此看出，回归方程拟合程度比较高，显著性明显。根据 T 检验得到参数 x_7, x_{13}, x_{16} 的 P 值都远远小于 0.05， x_3 的 P 值为 0.0522，稍稍大于 0.05，而数据在实际统计中会有少数误差，所以 x_3 的拟合度还可以。得到最优的回归方程为

$$\hat{y} = -2021786 + 2.47683x_3 + 37.13999x_7 + 10.384101x_{13} + 0.83846x_{16} \quad (3)$$

由回归方程可以看出企业所得税 x_3 、印花税 x_7 、罚没收入 x_{13} 、消费税和增值税 x_{16} 与财政收入的拟合度较高。

因此，在财政收入结构中，企业所得税、印花税、罚没收入、消费税和增值税是影响财政收入的主要方面。

(2) 影响增值税关键因素的识别。

由附件 10 的数据，利用 SAS 软件对增值税 y 和商品进口总值 x_1 、地区生产总值 x_2 、工业增加值 x_3 、批发零售业零售额 x_4 、工业增加值占 GDP x_5 、批发零售业增加值 x_6 ，作逐步回归，将软件运行结果（见附件 11）进行整理，得到表 2.3.6。

表 2.3.6 影响增值税关键因素的方差分析表

模型	平方和	df	均方	F 值	Pr > F
回归	4.771117E12	2	2.385558E12	343.45	<.0001
残差	83351519987	12	6945959999		
总计	4.854468E12	14			

模型	B	标准误差	II 型 SS	F 值	Pr > F
Intercept	60505	56873	7861579718	1.13	0.3083
x1	-0.0157637	0.007445	31156143146	4.49	0.0557
x3	0.05519	0.00591	6.051809E11	87.13	<.0001

从经济意义的角度，影响增值税最大的两个指标的系数符合客观实际。在显著性水平为 $\alpha = 0.05$ ，方程的 F 值为 343.45， P 值远远小于 0.05，由此看出，回归方程拟合程度比较高，显著性明显。根据 T 检验得到参数 x_1, x_3 的 P 值都远远小于 0.05， x_1 的 P 值为 0.0557，稍稍大于 0.05，而数据在实际统计中会有少数误差，所以 x_3 的拟合度还可以。得到最优的回归方程为

$$\hat{y} = 60505 + -0.01575x_1 + 0.05519x_3 \quad (4)$$

由回归方程可以看出，商品进口总值 x_1 、工业增加量 x_3 对增值税的影响较大。

综合上述分析，可以得到，商品进口总值和工业增加值两个指标对增值税的影响最大，认为是增值税的最关键影响因素。

(3) 影响营业税关键因素的识别

由附件 12 的数据, 利用 SAS 软件对营业税 y 和公路货运量 x_1 、公路客运量 x_2 、建筑业增加值 x_3 、第三产业增加值 x_4 、全社会房地产开发投资额 x_5 、全社会住宅投资额 x_6 、建筑业总产值 x_7 、住宿和餐饮业零售额 x_8 、限额以上餐饮业主营业务收入 x_9 , 作逐步回归, 将软件运行结果 (见附件 13) 进行整理, 得到表 2.3.7.

表 2.3.7 影响增值税关键因素的方差分析表

模型	平方和	df	均方	F 值	Pr > F
回归	6.474036E12	3	2.158012E12	932.02	<.0001
残差	25469667772	11	2315424343		
总计	6.499505E12	14			

模型	B	标准误差	II 型 SS	F 值	Pr > F
Intercept	55670	31412	7272504397	3.14	0.1040
x4	0.04923	0.00499	2.252746E11	97.29	<.0001
x7	-0.06026	0.01484	38176583067	16.49	0.0019
x8	0.05869	0.01870	22800821647	9.85	0.0094

从经济意义的角度, 影响营业税最大的三个指标的系数符合客观实际。

在显著性水平为 $\alpha = 0.05$, 方程的 F 值为 932.02, P 值远远小于 0.05, 由此看出, 回归方程拟合程度比较高, 显著性明显。根据 T 检验得到参数 x_4, x_7, x_8 , 的 P 值都远远小于 0.05。得到最优的回归方程为

$$\hat{y} = 55670 + 0.04923x_4 + -0.06026x_7 + -0.05869x_8 \quad (5)$$

由回归方程综合可以看出, 第三产业增加值 x_4 、建筑业总产值 x_7 、住宿和餐饮业零售额 x_8 对营业税的影响较大, 认为是营业税的最关键影响因素。

(4) 影响企业所得税关键因素的识别

由附件 14 的数据, 利用 SAS 软件对企业所得税 y 和第二产业增加值 x_1 、第三产业增加值 x_2 、全社会固定资产投资额 x_3 、城市商品零售价格指数 (1978=100) x_4 、规模以上工业企业盈亏相抵后的利润总额 x_5 、规模以上国有及国有控股工业企业企业亏损面 x_6 、建筑业总产值 x_7 、建筑业企业利润总额 x_8 、限额以上连锁店(公司)零售额 x_9 , 作逐步回归, 将软件运行结果 (见附件 15) 进行整理, 得到表 2.3.8.

表 2.3.8 影响增值税关键因素的方差分析表

模型	平方和	df	均方	F 值	Pr > F
回归	1.462542E12	2	7.312712E11	287.90	<.0001
残差	30480601021	12	2540050085		
总计	1.493023E12	14			

模型	B	标准误差	II 型 SS	F 值	Pr > F
Intercept	-1293604	448417	21138865924	8.32	0.0137
x4	2666.98347	805.04931	27876481739	10.97	0.0062
x8	0.79489	0.12469	1.032227E11	40.64	<.0001

从经济意义的角度，影响企业所得税最大的两个指标的系数都为正相关，符合客观实际。在显著性水平为 $\alpha = 0.05$ ，方程的 F 值为 287.9， P 值远远小于 0.05，由此看出，回归方程拟合程度比较高，显著性明显。根据 T 检验得到参数 x_4, x_8 的 P 值都远远小于 0.05。得到最优的回归方程为

$$\hat{y} = -1293604 + 2666.98347x_4 + 0.79489x_8 \quad (6)$$

由回归方程可以看出，商品进口总值 x_1 、工业增加值 x_3 对企业所得税的影响较大，认为是企业所得税的最关键影响因素。

(5) 影响个人所得税关键因素的识别

由附件 16 的数据，利用 SAS 软件对个人所得税 y 和城市居民年人均可支配收入 x_1 、城镇单位职工年平均总工资 x_2 、城镇居民储蓄存款余额 x_3 、地区生产总值 x_4 、第二产业增加值 x_5 ，总体作逐步回归，将软件运行结果（见附件 17）进行整理，得到表 2.3.9。

表 2.3.9 影响增值税关键因素的方差分析表

模型	平方和	df	均方	F 值	Pr > F
回归	507713473	1	507713473	795.36	<.0001
残差	3191710	5	638342		
总计	510905183	6			
模型	B	标准误差	II 型 SS	F 值	Pr > F
Intercept	6453.47872	646.88751	63530708	99.52	0.0002
x5	0.00068938	0.00002444	507713473	795.36	<.0001

从经济意义的角度，影响个人所得税最大的一个指标的系数为正数，符合客观实际。

在显著性水平为 $\alpha = 0.05$ ，方程的 F 值为 932.02， P 值远远小于 0.05，由此看出，回归方程拟合程度比较高，显著性明显。根据 T 检验得到参数 x_5 的 P 值都远远小于 0.05。得到最优的回归方程为

$$\hat{y} = 6453.47872 + 0.0068938x_5 \quad (7)$$

由回归方程可以看出，第二产业增加值 x_5 个人所得税的影响较大，认为是个人所得税的最关键影响因素。

3、财政收入及各个类别收入的预测

结合需求 1 各指标影响因素的分析，构建灰色预测模型和回归预测模型，对财政总收入、税收收入、非税收入等指标进行预测。

(1) 建立灰色预测模型

设原始灰色序列为 $X_0^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)]$ ，对原始灰色序列作 1 次累加生成新序列 $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ ，其中， $x^{(1)}(i) = \sum_{k=1}^i x^{(0)}(k)$ ，则灰色 GM(1, 1) 模型为

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \quad (8)$$

式中, $x^{(1)}(t)$ 是 t 时刻 1 次累加值, a 称为发展灰数, b 称为内生控制灰数。则式 (8) 的解为

$$x^{(1)}(t) = (x^{(0)}(1) - \frac{b}{a})e^{-a(t-1)} + \frac{b}{a} \quad (9)$$

参数的估值 $(a, b)^T = (\hat{B}\hat{B})^{-1}\hat{B}^T Y$, 且

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -[x^{(1)}(2) + x^{(1)}(1)]/2 & 1 \\ -[x^{(1)}(3) + x^{(1)}(2)]/2 & 1 \\ \dots & \dots \\ -[x^{(1)}(n) + x^{(1)}(n-1)]/2 & 1 \end{bmatrix}$$

由此得到模型还原值为

$$\hat{x}^{(0)}(k+1) = (\hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)), \quad k = 1, 2, \dots, n \quad (10)$$

根据式 (10), 得到原序列对应的预测序列值 $\hat{X}^{(0)} = [\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(n)]$

利用附件 1-附件 4 数据, 对 1999-2014 年个指标序列, 进行灰色预测, 所有模型预测参数与预测结果见表 2.3.10。

表 2.3.10 2015 年财政收入各指标灰色模型预测值

指标	a	b	预测值	相对平均误差%
财政总收入	-0.179978222	1553284.591	30833134.35	6.70%
一般预算收入	-0.133188839	1754691.424	15688854.71	4.75%
政府性基金收入	-0.223460452	1751381.8	13150536.93	1.95%
上级补助收入	-0.139783539	420508.1007	4241597.474	0.73%
税收收入	-0.127101108	1411172.458	11510896.3	2.21%
非税收收入	-0.178034767	191264.7186	2889034.491	7.61%
增值税	-0.129898385	346172.6908	2875592.945	3.21%
营业税	-0.079772008	570533.3129	2084179.366	6.55%
企业所得税	-0.123904581	192645.3795	1550247.237	2.04%
个人所得税	-0.094643471	131067.945	623391.2313	2.62%
城市维护建设税	-0.200275649	63796.35337	1412035.322	9.61%
房产税	-0.152038937	73927.087	905644.816	0.55%
印花税	-0.163036738	29063.09701	407406.0416	9.18%
契 税	-0.155946548	75292.99531	1099476.248	5.85%
国有资本经营收入	-0.126845866	215846.3006	616351.0238	1.62%
行政事业性收费收入	-0.193151244	255810.5316	1299120.97	8.81%
罚没收入	-0.07222718	154786.7597	286636.3597	1.16%
专项收入	-0.215648859	124053.1092	824568.4133	5.22%

由预测结果可以看出, 该模型预测结果中, 个别指标预测误差较大, 如, 城市维护建设税的误差为 9.61%, 印花税的误差为 9.18%, 行政事业性收费收入误差为 8.81%, 其余指标预测值均在 5% 左右, 总体来说, 预测结果比较可靠。

(2) 建立回归预测模型

根据财政收入中各指标的变化趋势, 构建回归模型进行预测。对国有资本经营收入 营业税 个人所得税、行政事业性收费收入 罚没收入指标建立线性回归模型, 模型结构为: $y = at + b$, 并进行参数检验与分析。对财政总收入、一般预算收入、政府性基金收入、上级补助收入、税收收入、非税收收入、专项收入、增值税、企业所得税、城市维护建设税、房产税、印花税、契税指标建立指数回归模型, 模型结构为: $y = ae^{bt}$, 并进行参数检验与分析。其中, 预测时 $t=2015$

利用附件 1-附件 4 数据, 对 1999-2014 年个指标序列, 进行回归预测, 所有模型预测参数与预

测结果见表 2.3.11。

表 2.3.11 2015 年财政收入各指标回归预测值

指标	模型结构	a	b	2015 年预测值	相对平均误差%
财政总收入		0.179969896	-345.446717	29286125.36	0.14%
一般预算收入		0.139530567	-264.5787853	15796617.19	0.00%
政府性基金收入		0.195156619	-376.9341706	12072217.07	0.66%
上级补助收入		0.138650038	-264.1237322	4223162.284	0.16%
税收收入		0.129466231	-244.6145508	11523553.99	0.02%
非税收入		0.184865462	-357.4758928	3361884.546	0.10%
增值税		0.139806906	-266.7962332	3001679.826	0.07%
营业税		96983.09265	-193500798.6	1920133.1	0.21%
企业所得税		0.11574112	-219.0172084	1470552.208	0.28%
个人所得税		27940.45294	-55755174.7	544837.975	0.15%
城市维护建设税		0.163772628	-315.9606433	1177188.671	0.71%
房产税		0.152456146	-293.4746701	912974.7181	0.01%
印花税		0.186112604	-362.0301313	436596.7148	0.19%
契 税		0.15055871	-289.5328286	1027841.768	0.22%
国有资本经营收入		70732.92857	-141800907.7	333086.2143	0.80%
行政事业性收费收入		108326.2143	-217216430.3	1060891.5	0.31%
罚没收入		13835.10714	-27605916.36	271824.5357	0.21%
专项收入		0.20183773	-393.1525618	792461.4126	0.24%

(3) 预测值处理

为降低预测误差，将灰色预测与回归预测值进行加权处理（权重分别为 0.5），最后给出 2015 年广州市地方财政总收入，以及一般预算收入、营业税、增值税等各个指标的预测区间。

表 2.3.12 2015 年财政收入各指标最终预测区间

指标	2015 年平均预测值	相对平均误差%	预测区间下界	预测区间上界
财政总收入	30059630	3.42%	29031430.82	31087829.18
一般预算收入	15742736	2.38%	15368680.38	16116791.62
政府性基金收入	12611377	1.31%	12446739.15	12776014.85
上级补助收入	4232380	0.45%	4213480.322	4251279.678
税收收入	11517225	1.11%	11389049.06	11645400.94
非税收入	3125460	3.85%	3005000.068	3245919.932
增值税	2938636	1.64%	2890442.273	2986829.727
营业税	2002156	3.38%	1934437.078	2069874.922
企业所得税	1510400	1.16%	1492866.117	1527933.883
个人所得税	584114.6	1.39%	576012.5216	592216.6784
城市维护建设税	1294612	5.16%	1227806.417	1361417.583
房产税	909309.8	0.28%	906747.2597	911872.3403
印花税	422001.4	4.68%	402234.9375	441767.8625
契 税	1063659	3.03%	1031379.849	1095938.151
国有资本经营收入	474718.6	1.21%	468974.8108	480462.3892
行政事业性收费收入	1180006	4.56%	1126207.137	1233804.863
罚没收入	279230.4	0.69%	277314.9105	281145.8895
专项收入	808514.9	2.73%	786435.9278	830593.8722

注：表中标注为蓝色是主要的预测指标值。

4 近几年社会经济发展和财政收入支出的分析

地方财政收入是国家财政收入的重要组成部分，是区域国民经济的综合反映，也是市场经济国家的政府进行宏观调控的基础。改革开放以来，广州省的广州市发展迅速，是“千年商埠”，商业中心之一。当然在地区生产总值飞跃的同时，也意味着财政收入的增收。财政收入是地区经济的宏观调控，财政收入的合理支配，是促进地区经济稳步增长的关键。对广州市地方财政从经济因素和非经济因素两个方面来分析。

1、经济因素对财政收入支出的分析

从经济因素方面的 GDP 与地方财政支出（见图 2.5.1）及地方财政收入的关系和地方财政支出的差值（见 2.5.2）两方面来分析。数据见见附件 20 和附件 21。

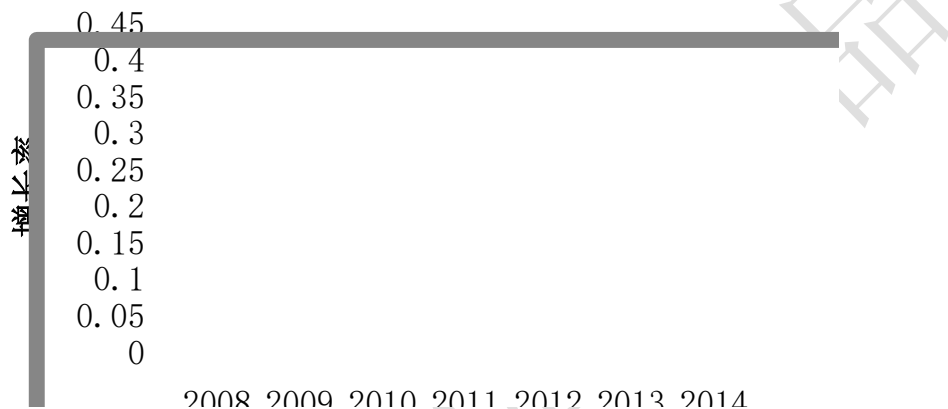


图 2.5.1 GDP、地方财政支出及地方财政收入关系图

考虑 GDP、地方财政支出及地方财政收入（数据见附件 20）三者在今年的增长，通过 GDP 的增长变化可以反映出广州市的经济实力和市场模型。从总体是可以看出，GDP 的变化趋势与地方财政收入基本保证一致，且 GDP 波动较小，地方收入和支出的波动较大。2012 年以前国内生产总值带动着地方财政收入和支出的变动，并且 2009-2012 年之间支出明显大于收入，GDP 处于下降趋势且变动较大，及此时的广州市政府处于收支并不平衡。

2012 年以前广州市政府处于 2012 以后三者之间的变化趋势一致，地方财政收入大于地方财政支出，国内生产总值变化较为稳定。

利用 R 语言中计算三者之间的相关性（见程序代码 1）

```
> cor(a$国内生产增长率,a$地方财政支出增长率,method="spearman")
[1] 0.75
> cor(a$国内生产增长率,a$地方财政收入增长率,method="spearman")
[1] 0.07142857
> cor(a$地方财政支出增长率,a$地方财政收入增长率,method="spearman")
[1] 0.4642857
```

从数据中可以得出，国内生产总值与地方财政支出的斯皮尔曼相关系数为 0.75，拟合度最高，次之是地方财政支出与地方财政收入为 0.46，最低的是国内生产总值与地方财政收入，为 0.07，拟合程度最低。

2、非经济因素对财政收入支出的分析

社会因素主要有人口（利用人口增长百分比分析）、就业（利用从业人数增长率代替分析）、医疗卫生（利用医疗卫生机构增长率代替分析）、社会保障和社会救济（利用福利财政性支出代替分析）、以及自然环境五个方面。

(1) 对人口与财政收入、支出增长率的分析

人口的发展对经济的发展是有其重要的反作用的，因为人社会生活中的主体，也是社会生产行为的主体，没有一定的人口，就没有社会，没有社会生活，也没有社会生产^[4]。数据见附件 24



图 2.5.2 人口与财政收入、支出增长率关系

从图中可以看出，广州市近几年人口变动不大，但从图中可以看出人口发展与财政收支是反作用关系，2011 年与 2013 年人口增长率上升时，财政收支增长率都处于下降趋势

```
> cor.test(a$地方财政支出增长率,a$人口增比,method="spearman",use="complete.obs")
Spearman's rank correlation rho
data: a$地方财政支出增长率 and a$人口增加百分比
S = 48, p-value = 0.4972
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.3714286

> cor.test(a$地方财政收入增长率,a$人口增加百分比,method="spearman",use="complete.obs")
Spearman's rank correlation rho
data: a$地方财政收入增长率 and a$人口增加百分比
S = 66, p-value = 0.03333
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.8857143
```

通过 R 软件证实了人口对财政支出的反作用关系，斯皮尔曼相关系数都为负值。而且人口对财政收入的影响很大，相关系数高达-0.8857143。

(2) 对就业情况与财政收入、支出的增长率的分析

就业（利用从业人数增长率代替分析，数据见附件 25），见图 2.5.3。

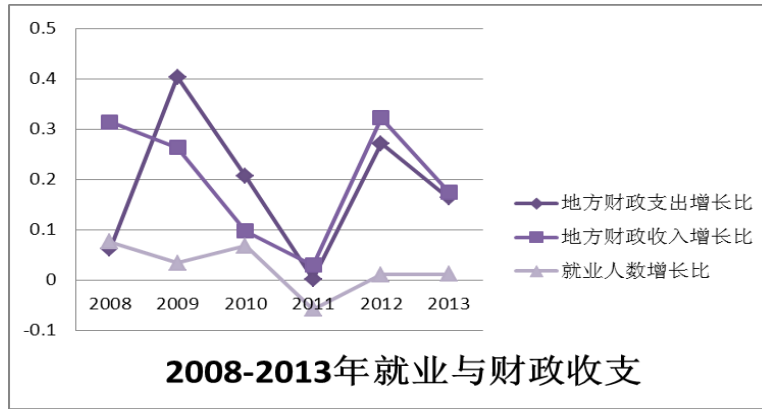


图 2.5.3 就业因素与地方财政收支

就业波动比人口要更加明显，从折线图中可以看出，就业因素与地方财政收支为正相关、我国拥有丰富的劳动力资源，同时也是世界上就业压力最大的国家之一，每年包括数量巨大的大学毕业生、下岗、失业和农民的剩余劳动力都需要在劳动力市场进行就业。因此，在未来的相当长的一段时期内，就业问题无意识政府面临的重要目标之一^[5]

用 R 软件进行检验（见代码 4）

```
> attach(a)
> cor.test(a$地方财政支出增长率,a$就业人数增长率,method="spearman")
```

Spearman's rank correlation rho

```
data: a$地方财政支出增长率 and a$就业人数增长率
S = 30, p-value = 0.8028
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.1428571
```

```
> cor.test(a$地方财政收入增长率,a$就业人数增长率,method="spearman")
```

Spearman's rank correlation rho

```
data: a$地方财政收入增长率 and a$就业人数增长率
S = 26, p-value = 0.6583
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.2571429
```

就业对于财政收支的影响正如折线图中所显示的为正相关，而且就业对于财政收入的影响略大于支出。

(3) 对医疗卫生机构与财政收入、支出增长率的分析

医疗卫生（利用医疗卫生机构增长率代替分析，数据见附件 26），见图 2.5.4。

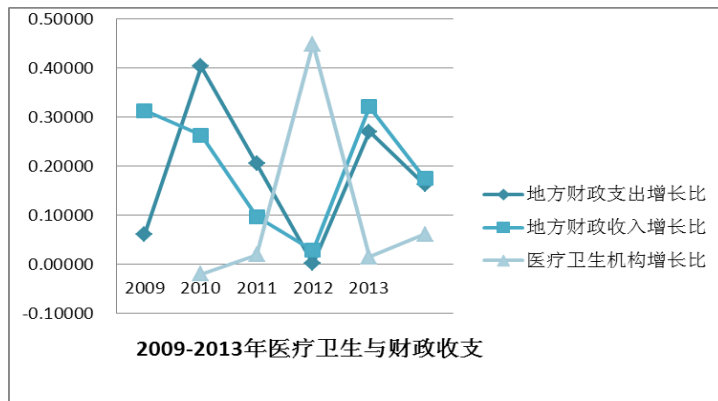


图 2.5.4 医疗卫生机构与财政收入、支出增长率

用 R 软件做检验（命令见代码 5）

```
> cor.test(a$地方财政支出增长率, a$医疗卫生机构增长率, method="spearman")
Spearman's rank correlation rho
data: a$地方财政支出增长率 and a$医疗卫生机构增长率
S = 40, p-value = 0.01667
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-1
```

```
> cor.test(a$地方财政收入增长率, a$医疗卫生机构增长率, method="spearman")
Spearman's rank correlation rho
data: a$地方财政收入增长率 and a$医疗卫生机构增长率
S = 36, p-value = 0.1333
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.8
```

地方财政支出与医疗卫生的相关系数达到了-1，而地方财政收入与医疗卫生的相关系数达到了-0.8，三者之间都为负相关

(4) 对社会福利与地方财政收入、支出的分析

社会保障和社会救济（利用福利财政性支出代替分析数据见附件 27），见图 2.5.5。

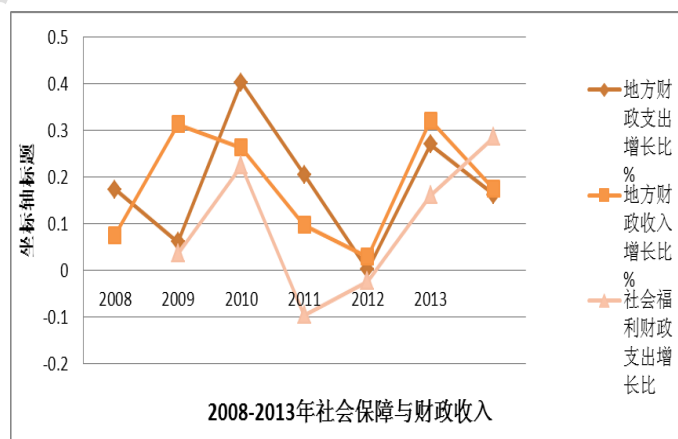


图 2.5.5 社会福利与地方财政收入、支出

用 R 软件做相关性检验（命令见代码 6）

```
> cor.test(a$地方财政支出增长率, a$社会福利财政支出增长率, method="spearman")
Spearman's rank correlation rho
data: a$地方财政支出增长率 and a$社会福利财政支出增长率
S = 22, p-value = 0.4972
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.3714286

> cor.test(a$地方财政收入增长率, a$社会福利财政支出增长率, method="spearman")

Spearman's rank correlation rho

data: a$地方财政收入增长率 and a$社会福利财政支出增长率
S = 20, p-value = 0.4194
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.4285714
```

通过 R 语言中的斯皮尔曼相关系数可以看出，地方财政收支与社会福利都呈现正相关。

(5) 其他因素对地方财政收入、支出的分析，

2013 年全年因自然灾害原因导致我市受灾人口共计 63902 人，紧急转移安置人口 7155 人，倒塌房屋间数 1934 间。全年农作物受灾面积 7.17 千公顷，农作物绝收面积 1.77 千公顷。发生各类地质灾害共 87 起，因灾死亡 7 人，直接经济损失 7.45 亿元。据统计，2014 年广州因自然灾害造成 26 万人口受灾，直接经济损失高达 13.4 亿元。由此可见自然灾害对财政收入与支出的影响是非常大的。

还有政策倾向性等因素。

3、历年财政收、支情况分析

结合 1999-2014 年广州市地方财政收入与支出的数据，分析广州市地方财政收支均衡状况。

结合 2015 年财政预算和收入的预测值，进行收支状况分析。

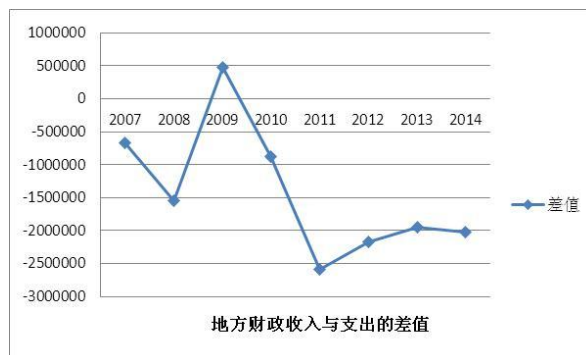


图 2.5.6 历年财政收、支情况分析

此图为 2007-2014 年地方财政收入减去地方财政支出的差值分布图。从图中可以看出曲线分布除了在 2009 年为正值，支出大于收入，其余均处于坐标轴以下，及收入大于支出。曲线在 2011

年之前波动较大，在 2011 年以后波动较小。

通过 R 软件对差值做单样本方差分析（见代码 2）

```
> t.test(b$差值, mu=0)

One Sample t-test

data: b$差值
t = -3.9947, df = 7, p-value = 0.005225
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -2256236.1 -578330.7
sample estimates:
mean of x
-1417283
```

从结果总可以看出，差值的均值处于-1417283，收支并不平衡，且 95%的置信区间比较大，与折线图结果相符。用 spss 软件对地方财政支出 y 和公共财政支出 x_1 、一般公共服务 x_2 、国防 x_3 、公共安全 x_4 、教育 x_5 、科学技术 x_6 、文化体育与传媒 x_7 、社会保障和就业 x_8 、医疗卫生 x_9 、环境保护 x_{10} 、城乡社区事务 x_{11} 、农林水事务 x_{12} 、交通运输 x_{13} 、资源勘探电力信息等事务 x_{14} 、工业商业金融等事务 x_{15} 、商业服务业等事务 x_{16} 、粮油物资储备管理等事务 x_{17} 、金融监管等事务支出 x_{18} 、地震灾后恢复重建支出 x_{19} 、住房保障支出 x_{20} 、其他支出 x_{21} 、政府性基金支出 x_{22} 、附:上解上级支出 x_{23} ，总体（数据见附件 22）做逐步回归（见附件 23），得到最后的方差分析表（表 2.2.5）。

表 2.3.13 2015 年财政收入各指标最终预测区间

模型	平方和	df	均方	F 值	P 值
回归	5.197E14	1	5.197E14	770.167	.000a
残差	8.772E12	13	6.748E11		
总计	5.285E14	14			

模型	B	标准误差	标准系数	F 值	P 值
1 (常量)	6068596.495	.000	.887	.000	.000
x12	20.063	.000	.158	.000	.000
x9	3.091	.000	-.095	.000	.000
x21	-1.901	.000	.001	.000	.000
x8	.014	.000	.000	.000	.000
x13	.001	.000	.000	.000	.000

从经济意义的角度，影响地方财政收入最大的四个指标的系数都为正，符合客观实际。假设显著性水平为 α ，方程的F值为770.167，P值远远小于0.05，由此看出，回归方程拟合程度比较高，显著性明显。根据T检验得到参数 β 的P值都远远小于0.05。得到最优的回归方程为

$$\hat{y} = 6068596.495 + 20.063x_{12} + 3.091x_9 - 1.901x_{21} + 0.14x_8 + 0.001x_{13}$$

回归方程可以看出农林水事务 x_{12} 、医疗卫生 x_9 、其他支出 x_{21} 、社会保障和就业 x_8 、交通运输 x_{13} 对税收的影响较大。

4、综合分析，提出建议

通过对广州市近几年社会经济发展的人口、就业、环境、医疗卫生和社会福利、财政收入及支出等数据的挖掘，对影响财政收入的经济因素和非经济因素两方面的分析，明确财政收入支出与经济、社会的相互关系，从广州市财政局提出财政收入和支出预算的平衡性、行业或部门财政支出分配的弹性、“营改增”等政策制定与监管落实的严肃性等方面提出科学合理化的建议。

综上所述因素分析，提出几点建议：

1. GDP 的发展状况较大程度影响着财政收入的情况。而近几年财政收入与支出的差额一直为负值，财政收入长期小于支出，严重影响了国民经济的发展，政府应该健全税收制度，狠抓企业或个人偷税漏税的现象，加强企业及个人对税收的管理，与此同时，应该增加税收的弹性系数。当财政收入大于支出的时候，应该多投资给教育，医疗卫生，社会保障等行业，使经济均衡稳步发展。
2. 人口的增长与财政收入成反比，在城市化进程中，政府应该适当的控制人口的数量。
3. 就业人数对财政收支的影响是成正比例关系，适当的增加就业人数可以提高财政收入。
4. 医疗卫生行业对财政收入起到牵引作用，应改善医疗保障体制，培养专业的技术人员，增加对医疗卫生方面的支出。
5. 发生自然灾害后，财政支出会大规模扩大，事后难以降到原来水平，所以在之后的几年内，政府为了缓解自然灾害造成的损失，应该继续加大对预防自然灾害等的建设。
6. 正确处理政府与市场的关系，进一步转换政府职能，坚决实行政企分开；加快行政审批制度改革；加强经济调节和市场监管职能的同时，重视社会管理和公共服务只能；梳理以人为本的管理思想，坚持全面协调和可持续的发展观。统筹兼顾经济和社会、城市和农村、人和自然的协调发展，全面提高人民的物质文化水平和健康水平^[1]。
7. 采取财政贴息、财政补贴和税收支出等投入资金，以少量的财政资金引导更多的社会资金投入，并引导社会资金的使用方向，既发挥所谓的“四两拨千斤”的效应。

3. 结论

通过挖掘、整理 1999-2014 年与广州市财政收入关联的经济、社会等指标统计数据，利用 SAS 和 SPSS 软件，分析了广州市财政收入的构成及其比例，建立了因素分析的回归模型，分析了在财政收入结构中影响财政收入的主要方面，进一步分析识别出了影响财政总收入的关键因素是企业所得税、增值税、第三产业、交通业，其中，影响增值税的关键因素是商品进口总值、工业增加量，影响营业税的关键因素是第三产业增加值、建筑业总产值、住宿和餐饮业零售额，影响个人所得税的关键因素是第二产业增加值，影响企业所得税的关键因素是商品进口总值、工业增加量。回归方程的拟合度是非常高的。

在因素分析基础上,构建了灰色预测和回归预测模型,对广州市 2015 年财政总收入、增值税、营业税、企业所得税、个人所得税的进行了预测,对预测结果进行加权处理,结果误差最大 5%左右,表明模型可靠,符合广州的实际情况。

通过对广州市近几年社会经济发展的人口、就业、环境、医疗卫生和社会福利、财政收入及支出等数据的挖掘,再用 R 软件等对影响财政收入的经济因素和非经济因素两方面进行了分析,明确了财政收入支出与经济、社会的相互关系,向广州市财政局提出财政收入和支出预算的平衡性、行业或部门财政支出分配的弹性、政策制定与监管落实的严肃性等方面的建议,为下一年的财政预算支出政策提供指导依据。

总体来说,通过对该问题中数据比较全面的挖掘,有效的提取,科学的整理,其预测模型和数据处理真实可靠,很好地解决了基于数据挖掘技术的广州市的财政收入分析预测问题。

4. 参考文献

- [1] 吕泽楷,莫其盛,李伟.新世纪以来广州市财政收入与支出的演变分析[J].华南师范大学增城学院.
<http://3y.uu456.com/bp-c42f4008763231126edb11e4-1.html> 三亿文库 [2015.5.21]
- [2] 陈共.财政学(第七版).中国人民大学出版社[M].北京,2012.
- [3] 营业税改增值税.百度百科
http://baike.baidu.com/link?url=fPBn0XvpssIv_6Uhx4VEJyakwJGyfZpt6_E0Cgc38uJ_pGSs06XTeZ169aFNtN8QNB-aea34uKkyvoUpzfkU2YFFMNRJK4L4NrqsLrsU5t5HHik62KBePW9jhoMDv4M7ZZSszSW28S1n_ukKCbdUL9xw4iDQrcfidGGQkf5S0IGO
- [4] 王秀元,李亚峰,陶精林.论人口与经济的关系[J].
http://wenku.baidu.com/link?url=D_I00Ka4RjwD83PVXOGejtviU9RDex6amCoIDTFqn-BsFc30cP2cJ3QKfIm61kyHIUC3IxFvAsoScGN76x73wOW53sdrglkPIt0NLJ2Aaa
- [5] 陈宪,黄健柏,董晶晶.财政收支对农民工就业能力影响研究[J].中南大学学报,1998,22(1):34-37.

附页

代码 1:

```
> a<-read.csv("C:\\Documents and Settings\\Administrator\\ \\增长率.csv")
```

```
> a
```

	年份	国内生产增长率	地方财政支出增长率	地方财政收入增长率
1	2008	0.16572938	0.174024975	0.07636666
2	2009	0.10266574	0.061688274	0.31373703
3	2010	0.17619096	0.403648112	0.26316289
4	2011	0.15585377	0.205887976	0.09718501
5	2012	0.09077760	0.001987846	0.02901477
6	2013	0.13791610	0.270794427	0.32187334
7	2014	0.08344477	0.162948095	0.17516213

代码 2:

```
> b<-read.csv("C:\\Documents and Settings\\Administrator\\ \\差值.csv")
```

```
> b
```

	年份	差值
1	2007	-666910
2	2008	-1547947
3	2009	481688
4	2010	-879985
5	2011	-2582093
6	2012	-2172325
7	2013	-1953695
8	2014	-2017000

代码 3:

```
> a<-read.csv("C:\\Documents and Settings\\Administrator\\ \\人口.csv")
```

```
> a
```

	年份	人口总数. 人.	增加人口数	人口增加百分比	地方财政支出增长率	地方财政收入增长率
1	2007	2382491	NA	NA	0.174024975	0.07636666
2	2008	2425582	43091	0.018	0.061688274	0.31373703
3	2009	2474396	48814	0.020	0.403648112	0.26316289
4	2010	2526804	52408	0.021	0.205887976	0.09718501
5	2011	2595686	68882	0.027	0.001987846	0.02901477
6	2012	2646091	50405	0.019	0.270794427	0.32187334
7	2013	2706068	59977	0.022	0.162948095	0.17516213

代码 4:

```
> a<-read.csv("C:\\Documents and Settings\\Administrator\\桌面\\2.csv")
```

```
> a
```

	年份	地方财政支出增长率	地方财政收入增长率	就业人数增长率	就业人数. 人.
1	2008	0.06169	0.31374	0.07597	7145396
2	2009	0.40365	0.26316	0.03381	7386999
3	2010	0.20589	0.09719	0.06824	7891061
4	2011	0.00199	0.02901	-0.05821	7431755
5	2012	0.27079	0.32187	0.01093	7512997
6	2013	0.16295	0.17516	0.01149	7599295

代码 5:

```
> a<-read.csv("C:\\Documents and Settings\\Administrator\\ \\2.csv")
> a
  年份 地方财政支出增长率 地方财政收入增长率 医疗卫生机构增长率 医疗卫生机构
1 2009          0.40365          0.26316          -0.01968174          2341
2 2010          0.20589          0.09719          0.01964972          2387
3 2011          0.00199          0.02901          0.44909929          3459
4 2012          0.27079          0.32187          0.01503325          3511
5 2013          0.16295          0.17516          0.06209057          3729
```

代码 6:

```
> a<-read.csv("C:\\Documents and Settings\\Administrator\\桌面\\2.csv")
> a
  年份 地方财政支出增长率. 地方财政收入增长率. 社会福利财政支出增长率
1 2007          0.174024975          0.07636666          NA
2 2008          0.061688274          0.31373703          0.03626833
3 2009          0.403648112          0.26316289          0.22475986
4 2010          0.205887976          0.09718501          -0.09603271
5 2011          0.001987846          0.02901477          -0.02298214
6 2012          0.270794427          0.32187334          0.16149450
7 2013          0.162948095          0.17516213          0.28601057
  社会福利财政性支出
1          18887
2          19572
3          23971
4          21669
5          21171
6          24590
7          31623
```